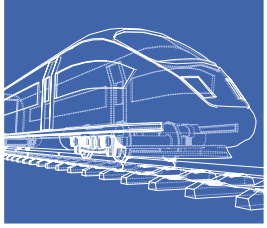
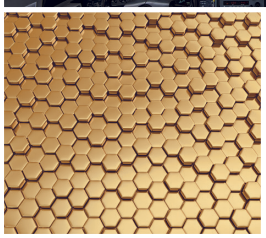
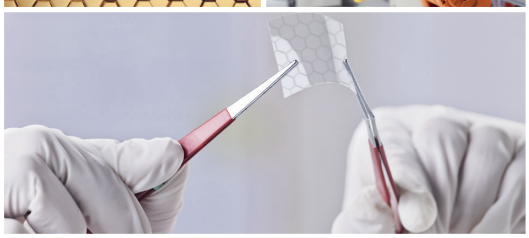










Advances in Transdisciplinary Engineering series

volume 22

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering

2022 NMMVE

EDITED BY
Jinyang Xu
Yukui Cai
Mohamed El Mansori

IOS Press

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering

New materials are constantly being developed which may improve or transform many aspects of our lives, and nowhere is this more exciting than in the fields of vehicle and machinery technology.
This book presents the proceedings of the 2022 International Conference on New Materials, Machinery and Vehicle Engineering (NMMVE 2022), held as a virtual event due to the COVID-19 pandemic and travel restrictions, from 18 – 20 March 2022.

NMMVE 2022 provides an international forum for researchers and engineers to present and discuss recent advances, new techniques, and applications in the fields of new materials, machinery and vehicle engineering, and attracts academics, scientists, engineers, postgraduates, and other professionals from a wide range of universities and institutions. A total of 121 submissions were received, from which 48 were accepted for inclusion in the conference and proceeding after a rigorous, standard single-blind reviewing process. The papers are grouped into 3 sections: machinery (30 papers); new materials (11 papers); and vehicle engineering (7 papers).

Providing an overview of the latest developments in these fields, the book will be of interest to all those wishing to know more about new materials and machine and vehicle engineering.

ISBN 978-1-64368-270-9 (print)
ISBN 978-1-64368-271-6 (online)
ISSN 2352-751X (print)
ISSN 2352-7528 (online)

PROCEEDINGS OF THE 1ST INTERNATIONAL CONFERENCE ON NEW MATERIALS, MACHINERY AND VEHICLE ENGINEERING

Advances in Transdisciplinary Engineering

Advances in Transdisciplinary Engineering (ATDE) is a peer-reviewed book series covering the developments in the key application areas in product quality, production efficiency and overall customer satisfaction.

ATDE will focus on theoretical, experimental and case history-based research, and its application in engineering practice. The series will include proceedings and edited volumes of interest to researchers in academia, as well as professional engineers working in industry.

Volume 22

Recently published in this series

Vol. 21. A.H.C. Ng, A. Syberfeldt, D. Högberg and M. Holm (Eds.), SPS2022 – Proceedings of the 10th Swedish Production Symposium
Vol. 20. C.-H. Chen (Ed.), Applied Mathematics, Modeling and Computer Simulation – Proceedings of AMMCS 2021
Vol. 19. M. Yang, J.C.G. Lanzinha and X. Bao (Eds.), Hydraulic and Civil Engineering Technology VI – Proceedings of the 6th Interna-tional Technical Conference on Frontiers of HCET 2021
Vol. 18. D. Bhattacharjee (Ed.), Computer Methods in Medicine and Health Care – Proceedings of the CMMHC 2021 Workshop
Vol. 17. D. Dobrotă and C. Cheng (Eds.), Proceedings of the 2nd International Conference on Green Energy, Environment and Sustainable Development (GEESD2021)
Vol. 16. L. Newnes, S. Lattanzio, B.R. Moser, J. Stjepandić and N. Wognum (Eds.), Transdisciplinary Engineering for Resilience: Responding to System Disruptions – Proceedings of the 28th ISTE International Conference on Transdisciplinary Engineering, July 5 – July 9, 2021
Vol. 15. M. Shafik and K. Case (Eds.), Advances in Manufacturing Technology XXXIV – Proceedings of the 18th International Conference on Manufacturing Research, incorporating the 35th National Conference on Manufacturing Research, 7–10 September 2021, University of Derby, Derby, UK

ISSN 2352-751X (print)
ISSN 2352-7528 (online)

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering

Edited by

Jinyang Xu

Shanghai Jiao Tong University

Yukui Cai

Shandong University

and

Mohamed El Mansori

Arts et Métiers ParisTech

IOS Press

Amsterdam • Berlin • Washington, DC

ISBN 978-1-64368-270-9 (print)
ISBN 978-1-64368-271-6 (online)
Library of Congress Control Number: 2022935873
doi: 10.3233/ATDE22

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

For book sales in the USA and Canada:
IOS Press, Inc.
6751 Tepper Drive
Clifton, VA 20124
USA
Tel.: +1 703 830 6300
Fax: +1 703 830 2300
sales@iospress.com

Preface

The 2022 International Conference on New Materials, Machinery and Vehicle Engineering (NMMVE 2022) was held successfully on March 18, 2022. Because of the COVID-19 pandemic and the strict traveling restrictions, the event had to take place virtually eventually. However, this did not prevent our participants from enjoying great keynote speeches and presentations.

NMMVE 2022 has attracted a great number of academics, scientists, engineers, postgraduates, and other professionals from a large range of universities and institutions. We aim to provide a high-level international forum for researchers and engineers to present and discuss recent advances, new techniques, and applications in the field of New Materials, Machinery and Vehicle Engineering.

A total of 121 submissions were received, among which 48 have been accepted to be included in the conference proceedings. We followed the standard single-blind reviewing process to evaluate papers. Each paper was manually assigned to reviewers after an initial paper bidding process. We would like to express our sincere gratitude to all the experts for their professional and detailed review comments, which have contributed to the high standard of this collection. We also thank the 27 universities and research institutes who took part in our program and special thanks to Zhenzhou University of Light Industry and Shandong University for their sponsorship.

We would like to thank the Steering Committee for their advice and guidance. We would also like to thank all the authors for their contributions to making NMMVE an exciting conference and call for their contributions to the future events of NMMVE. We also owe thanks to the PC and external reviewers for their hard work in reviewing and shepherding. Finally, we would like to thank IOS Press for the support of their publication services.

We hope you all enjoy your experience at NMMVE 2022!

Prof. Jinyang Xu, Shanghai Jiao Tong University
Prof. Yukui Cai, Shandong University
Prof. Mohamed El Mansori, Arts et Métiers ParisTech
NMMVE 2022

This page intentionally left blank

Conference Committee List

General Chairs
Prof. Jianlei Cui, Xi'an Jiaotong University, China
Prof. Yukui Cai, Shandong University, China
Prof. Peng Zhao, Zhejiang University, China
Prof. Muhammad Zubair lqbal, Institute of Biomedical Materials at Zhejiang Sci-Tech University, China

Program Chairs
Prof. Jie Teng, Hunan University, China
Prof. Wenjun Wang, Xi'an Jiaotong University, China
Prof. Chunbao Liu, Jilin University, China
Prof. Jing Wei, Chongqing University, China
Assoc. Prof. Qinghua Wang, Southeast University, China

Honorary Advisor
Dr. Sean Monkman, Carbon Cure Technologies, Canada
Ms. Shana Kelley, KPFF, USA

Technical Chairs
Prof. Jinyang Xu, Shanghai Jiao Tong University, China
Prof. Mohamed El Mansori, Arts et Métiers ParisTech, France
Prof. Xiaofei Sun, Xi'an Jiaotong University, China
Prof. Liangwen Wang, Zhengzhou University of Light Industry, China
Prof. QingjuTang, Heilongjiang University of Science and Technology, China
Assoc. Prof. Fuhao Mo, Hunan University, China
Assoc. Prof. Qiang Gao, Southeast University, China

Publication Chairs
Prof. Hao Tong, Tsinghua University, China
Prof. Shiwei Zhang, South China University of Technology, China

Technical Program Committees
Prof. Weimin Huang, Nanyang Technological University, Singapore
Prof. Gagik H. Torosyan, National Polytechnic University of Armenia, Armenia
Prof. Tiesong Lin, Harbin Institute of Technology, China
Prof. Muhammed Bashir, Mu'azu Ahmadu Bello University, Zaria, Nigeria
Prof. Wang Xiao Dong, North China Electric Power University, China
Prof. Xuegong Yu, Zhejiang University, China
Prof. Hongyong Jiang, China University of Geosciences, China
Prof. Ming-Yang Chen, Jiangsu University, China
Prof. Kim Jonghyun, Chongqing University, China

Contents

New Materials

Machinery

This page intentionally left blank

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 3
J. Xu et al. (Eds.)

doi:10.3233/ATDE220206

Research on Infrared Large Field of View Mosaic Based on Improved ORB Algorithm

Hu ZHOU[a], Lin CHAI[a,1] and Lizuo JIN[a]
[a]*School of automation, Southeast University, Nanjing, China*

Abstract. In view of the current narrow field of view and low resolution of infrared images, the SIFT and SUFT algorithms have a long time to stitch images, a real-time infrared large field of view stitching algorithm based on Oriented FAST and Rotated BRIEF (ORB) is proposed. The algorithm makes full use of the priori information of the positional relationship between images, and adopts a matching algorithm based on a two-way matching strategy that is universal for infrared images in a variety of application scenarios, and finally introduces an adaptive weighted fusion algorithm that fades in and out. On the basis of ensuring the quality of image stitching, improving the running speed of the algorithm can realize fast and accurate stitching of infrared images with a large field of view. The experimental results show that under the same level of matching rate, the algorithm in this article has obvious advantages in terms of speed.

Keywords. Large field of view infrared image stitching, ORB, prior overlap area, two-way matching, adaptive weighting

1. Introduction

Infrared image large-field stitching is the process of stitching two or more infrared images with a narrow field of view into a seamless and complete infrared image by extracting the overlapping parts of the image. In various fields such as night reconnaissance, navigation, security monitoring, and synthetic medical images, the large field of view stitching of infrared images has a wide range of applications. However, for application scenarios that require infrared images with high resolution and wide field of view, mos[1]t of the current infrared imaging equipment cannot meet the requirements and need to be completed with the help of image stitching software.

Infrared image stitching mainly includes image enhancement, image registration and image fusion. The key of the whole process is image registration. So far, there are three types of algorithms of image registration: methods based on image information, methods based on frequency domain correlation, and methods based on feature correlation. Among them, image stitching based on feature correlation is popular because of its robustness. LOWE proposed the SIFT algorithm [1] in 2004, and BAY proposed the SURF algorithm [2] in 2006. As a classic algorithm in image registration,

[1] Corresponding Author, Lin CHAI, School of automation, Southeast University, Nanjing, China; E-mail: chailin1@seu.edu.cn.

it can effectively extract feature point pairs and has a high registration rate. However, the calculation time of the algorithm is relatively long, which cannot meet the real-time requirements of large-field stitching of infrared images. In response to this situation, many scholars have carried out related studies and made improvements to SIFT and SURF. Literature [3] combines SURF and K-d tree to increase the average processing time of images; Literature [4] combines ROI with SIFT, which effectively increases the speed of image stitching; Literature [5] combines SURF with two-way adaptive threshold to improve the quality and speed of image stitching. The acquisition speed of common video acquisition equipment is 25FPS to 50FPS, and the stitching technology is required to reach a flat speed. It can be found from the above methods that the improvement of SURF and SIFT has improved the stitching speed and quality to a certain extent, but there is still a gap between the speed and the real-time requirements. In the literature [6], comparing the five algorithms of SIFT, SURF, FREAK, BRISK, ORB, it can be found that the ORB algorithm has the fastest image registration speed, but there is a situation where the feature points are incorrectly matched, and the registration is completed at the same time. There are obvious gaps in the image. In view of the above situation, an improved ORB algorithm is proposed in this paper, which uses the prior information of the positions between images, and adopts a universal two-way matching algorithm, and finally introduces an area-based fade-in and fade-out adaptive weighting algorithm. On the basis of ensuring the quality of image stitching, improve the speed of image stitching.

2. Traditional orb image stitching algorithm

2.1. OFAST (Oriented FAST) feature point detection

In an infrared image, select a circular area with a radius of 3, and there are 16 pixels on the circumference of this circular area. The step of FAST algorithm to determine whether a pixel point p is a candidate point: first set a threshold t, if there are more than or equal to 12 of the 16 pixels around the candidate point, the pixel difference between p and pi exceeds the threshold t, then the point p is called the candidate feature point.

Non-maximum suppression is performed on the selected feature points to improve the robustness of the algorithm. In order to make the algorithm directional invariance, the gray centroid method is introduced.

2.2. BRIEF feature point description

The method of steered BRIEF [8] feature point descriptor is: select a square (9×9) area centered on the feature point in one of the images to be stitched; Randomly select n pairs of feature points (p, q) that obey $(0, S^2 / 25)$ Gaussian distribution in the area, and compare the pixel values $f(p)$ and $f(q)$ of p and q:

$$\tau(p, x, y) = \begin{cases} 1 & f(p) < f(q) \\ 0 & f(p) \geq f(q) \end{cases} \tag{1}$$

Randomly select n (n = 256 in this paper) pairs of points from the feature points.

$$D = \begin{pmatrix} x_1 & x_2 & \cdots & x_{256} \\ y_1 & y_2 & \cdots & y_{256} \end{pmatrix} \tag{2}$$

Rotate the image to be stitched into an angle θ, and the corresponding 256 point pairs will generate a new matrix:

$$D_\theta = R_\theta D \tag{3}$$

Where, $R_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$.

Repeat the above steps to form a new binary code string consisting of 0, 1. The descriptor of feature point is the binary code.

3. Improved orb algorithm

The image stitching algorithm flow in this article is shown in Figure 1. In contrast to the traditional ORB algorithm, the improvement of this article is mainly to obtain the prior position relationship and two-way matching measures to greatly reduce the mismatch points, and on this basis, the fusion effect is improved through the adaptive weighted fusion method, and finally real-time and accurate Infrared panoramic image stitching.

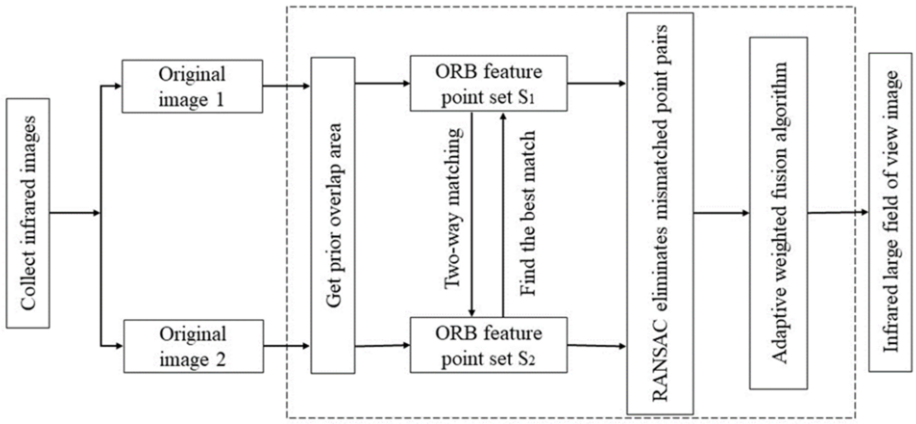


Figure 1. Flow chart based on improved ORB infrared stitching algorithm.

3.1. Prior overlap area

The traditional ORB algorithm directly detects feature points from two complete images; as shown in Figure 2, it is meaningless to detect feature points in the blank areas (non-overlapping areas) on both sides, so the algorithm in this paper is improved

from this perspective to avoid Feature point detection in non-overlapping areas improves efficiency.

As shown in Figure 2, the four corners of image 1 to be stitched are ABCD, and the four corners of image 2 to be stitched are EFGH. Use quadrilateral A 'B' C' D' to represent the prior overlap area. Extracting feature points in overlapping area will save the ORB algorithm's time and shorten the running time of the algorithm.

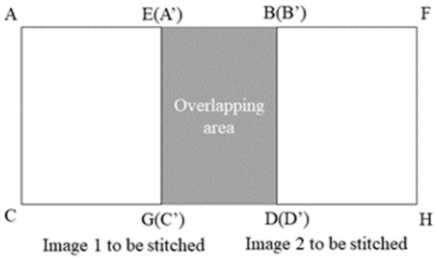


Figure 2. Mosaic diagram.

According to the built-in parameters of the acquisition device, the a prior [10] overlapping area of the two images can be effectively obtained. The overlapping area can be calculated with a 3×3 matrix W:

$$W = \begin{bmatrix} w_0 & w_1 & w_2 \\ w_3 & w_4 & w_5 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

In matrix W, w_0 and w_1 reflect the rotation transformation of the image, w_3 and w_4 reflect the expansion and contraction transformation of the image, w_2 reflects the translation distance of the image on the x-axis, and w_5 reflects the translation distance of the image on the y-axis. The steps to calculate the prior overlap area are as follows: Firstly, calculate the coordinates of E, F, G, and H according to the matrix W; secondly, obtain 4 straight line equations, representing EF, FH, AB, and BD respectively; Then get the intersection coordinate $A^{`}$ of $AB$ and $EG$, the intersection coordinate $B^{`}$ of BD and AB, the intersection coordinate C^{r} of EG and GH, and the intersection coordinate D' of GH and BD; finally get the a priori overlapping area quadrilateral $A'B'C'D'$.

3.2. Two-way matching of feature points

When the traditional ORB algorithm performs feature point matching, the Hamming distance is calculated according to the feature descriptor composed of a binary code string of 0,1. Suppose the feature descriptor of any feature point p of the reference image, and the feature descriptor of any feature point q of the image to be registered. Then the Hamming distance formula between two feature points is as follows:

$$dis(x_1, x_2) = \sum x_1[i] \oplus x_2[i] \tag{5}$$

From equation (5), the closest distance point and secondary distance point of each feature point on the image to be registered corresponding to the reference image can be calculated and stored in the candidate matching point set P. After that, it is screened by formula (6).

$$\frac{\text{closest}}{\text{second closest}} < T \tag{6}$$

In the formula, T is the ratio threshold, usually 0.5~0.7, and T = 0.6 in this article. Firstly, select the feature points of the reference image, and then find the matching points in the image to be registered according to the similarity measurement criterion. This one-way matching method is prone to mismatch point pairs.

The so-called two-way matching strategy [11] is to use the feature points between the reference image and the image to be registered to find their matching points in the corresponding image through the similarity measurement criteria, and then compare the two matching point pairs. The same part in the set of two matching points is the matching result. The formula for obtaining the same matching point pair is as follows:

$$P_{same} = P_1 \cap P_2 \tag{7}$$

The improved ORB adopts a matching algorithm based on a two-way matching strategy, making it universally applicable to infrared images in a variety of application scenarios, and reducing mismatched point pairs to a certain extent.

3.3. RANSAC

After the feature point matching is completed, there will be mismatches. Therefore, in order to improve the registration rate of feature points, it is necessary to introduce the RANSAC algorithm to eliminate mismatched point pairs. The main task of the RANSAC algorithm [12,13] is to find an optimal homography matrix H to ensure the maximum number of feature points applicable to the matrix. H represents the conversion relationship between two images to be stitched, including translation, rotation, and scaling between images. The transformation relationship between the two images is as follows:

$$A' = HA \tag{8}$$

In the formula, A and A' are two infrared images to be stitched.

Feature points of adjacent images are respectively expressed as (x', y') and (x, y), then:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{9}$$

In the formula, h_{33}=1.

For a pair of feature points, (x', y') and (x, y) are substituted into the formula. If H is obtained as the optimal matrix, the following cost function needs to be minimized:

$$W = \sum_{i=1} (x_i^{'} - \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}})^2 + (x_i^{'} - \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}})^2 \tag{10}$$

3.4. Adaptive weighted image fusion

Due to the gray level difference between the two images to be stitched, the infrared image after ORB registration has a gap in the stitching place. To provide the observer with a more intuitive scene image, the gap in the stitching place needs to be eliminated. At the same time, considering the real-time requirements of infrared image stitching technology, it is not suitable to use a more complex algorithm, so this paper adopts the fade-in and fade-out weighting algorithm [14]. The process of fade-in and fade-out weighted [15] image fusion is to make the overlapping parts of two images realize the natural process of transition from one image to another. It is necessary to overlap the two infrared images f1 and f2. The pixel values of the regions are merged into a stitched image according to a certain proportion. The weighting rules are as follows:

$$f = \begin{cases} f_1, & (x,y) \in f_1 \\ w_1 f_1 + w_2 f_2, & (x,y) \in (f_1 \cap f_2) \\ f_2, & (x,y) \in f_2 \end{cases} \tag{11}$$

Where, w_1+w_2=1.

The method is to take a certain weighted value for the pixel values of the two images respectively, and calculate their sum. However, the weights and usually need to be set in advance and adjusted many times. It is difficult to find an optimal weight. This paper adopts the adaptive weights based on area, as shown in Figure 3.

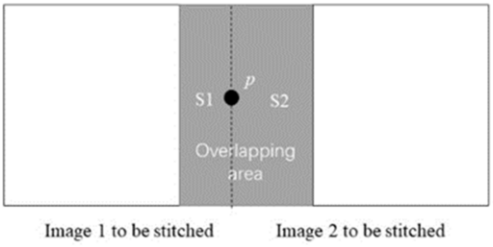


Figure 3. Mosaic diagram.

Suppose a pixel p in the overlap area. The overlap area is divided by the vertical dotted line by p. The left side of the area is S_1, and the right side of the area is S_2. Then the values of w_1 and w_2 are as follows:

$$w_1 = \frac{S_1}{S_1 + S_2} \tag{12}$$

$$w_2 = \frac{S_2}{S_1 + S_2} \tag{13}$$

This method can effectively eliminate the gaps in the image stitching and smooth the image; at the same time, it solves the problem of weight adaptation and has better reusability.

Compared with the traditional ORB algorithm, the algorithm in this paper reduces the time required for feature point detection through a priori overlapping area; secondly, it uses the combination of two-way matching and RANSAC to improve the universality and the feature point matching accuracy; finally, an adaptive weighting algorithm is used to smooth image stitching and eliminate gaps.

4. Results and discussion

In order to test and verify the feasibility of the improved algorithm, the experimental simulation is based on the 64-bit win10 operating system, the CPU is 2.70GHz AMD processor, the memory is 8GB PC platform, and the experimental programming environment is MATLAB-2020a. The following is an analysis of the experimental results of the algorithm for specific infrared images.

4.1. Getting prior overlap area

Figure 4. (a)Left and right images to be stitched.

Figure 4. (b)Left and right images to be stitched.

It can be found from Figure 4 that each group of adjacent images to be stitched has an obvious overlap area, and this article obtains the prior overlap area of the image according to the built-in parameters of the image acquisition device. The result is shown in Figure 5, which is helpful to improve the speed of feature point detection.

Figure 5. Priori overlap area of the image to be stitched.

4.2. Detecting feature points

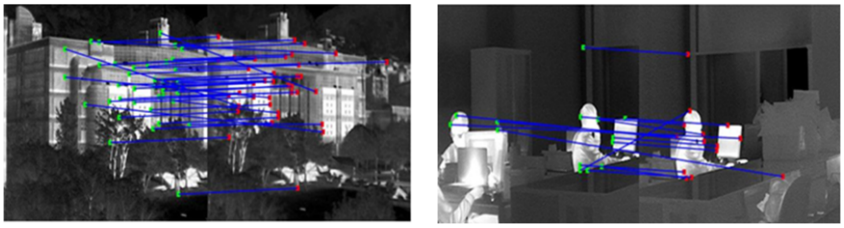


Figure 6. ORB registration effect.

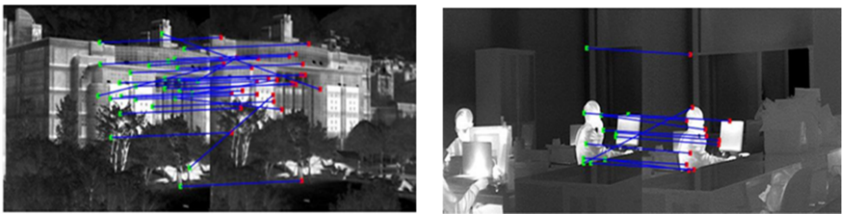


Figure 7. The registration effect of the improved algorithm.

It is shown in the Figure 6 and 7 that the traditional ORB algorithm searches for feature points in the entire image, and detects more pixels in non-overlapping areas, which causes an increase in time-consuming and is likely to cause subsequent mismatched point pairs; and In contrast, the feature points detected by the algorithm in this paper are all in the prior overlap area, which greatly reduces the number of pixels that need to be detected, and reduces the time for feature point detection to a certain extent. At the same time, the influence of the feature points in the non-overlapping area on subsequent matching is avoided.

In order to reflect the running speed of the improved algorithm, compare and analyze the following algorithms: SIFT, SURF, ORB and the improved algorithm. The result is shown in Table 1.

Table 1. Feature point detection time comparison.

Image	SIFT	SURF	ORB	Improved algorithm
(a)	2.034s	0.323s	0.024s	0.013s
(b)	1.893s	0.196s	0.017s	0.011s

Table 1 is the average value obtained by multiple experiments. The calculation found that the detection time of SIFT algorithm is 1.964s, the detection time of SURF algorithm is 0.259s, the detection time of ORB algorithm is 0.021s, and the average detection speed after optimization in this paper is 0.012s. The improved algorithm is 1.75 times faster than the ORB algorithm. It shows that the algorithm in this paper has speed superiority in the detection of feature points.

4.3. Matching feature points

Because the BRIEF algorithm describes feature points with binary codes, the mismatch rate of feature points is relatively high, and further screening is needed. This article combines two-way matching and RANSAC algorithm. There are two steps. The first step: two-way selection when matching the feature points to improve the accuracy; the second step: the error feature points are eliminated after the matching is completed. The comparison between the original registration effect Figure 8 and the optimized registration effect Figure 9 is as follows:

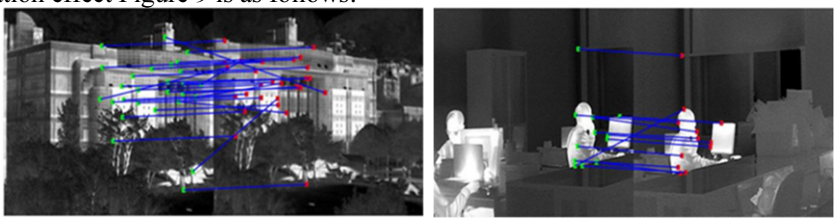


Figure 8. Original registration effect.

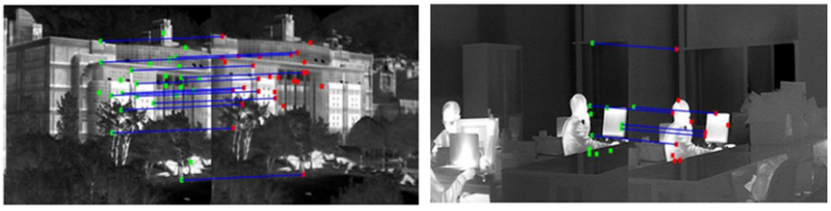


Figure 9. The registration effect of the algorithm in this paper.

From the comparison of Figure 8 and 9, it can be found that the combination of two-way matching and RANSAC algorithm can significantly optimize the matching effect of feature point pairs and decrease the number of wrong matching point pairs.

In order to evaluate the feature point matching effect of various methods, compare and analyze the following algorithms: SIFT, SURF, ORB and the improved algorithm. The result is shown in Table 2.

Table 2. Image registration accuracy comparison.

Image	SIFT	SURF	ORB	Improved algorithm
(a)	96.7%	92.3%	85.9%	91.2%
(b)	95.9%	93.7%	88.7%	92.2%

Through the comparison of the four algorithms, the average registration rate of SIFT is 96.3%, the average registration rate of SURF is 93.0%, the average registration rate of ORB is 87.3%, and the average registration rate of the improved algorithm is 91.7%. Although there is a certain gap with the SIFT and SURF algorithms, it is increased by about 4.4% compared with the ORB algorithm, indicating that the improved algorithm has higher accuracy in feature point matching.

4.4. Image fusion

Figure 10. (a)Effect of stitching before and after improvement.

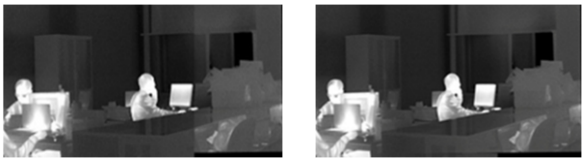


Figure 10. (b)Effect of stitching before and after improvement.

Figure 11. (a)The effect of stitching place before and after improvement.

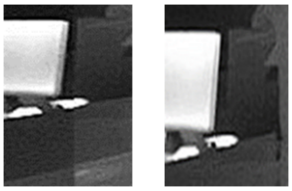


Figure 11. (b)The effect of stitching place before and after improvement.

It can be seen from Figures 10 and 11 that there are obvious gaps in the stitching of the registered images in this paper. Therefore, this paper uses the adaptive weighting method of fade-in and fade-out on this basis to fuse the images, which can effectively smooth the image. The gaps at the stitching place get a seamless stitched image.

The image stitching algorithm in this paper uses a priori position information, combines the feature point two-way matching with the RANSAC algorithm, and adopts the adaptive weighted fusion method of fade-in and fade-out, which can cut down the mismatch rate to less than 8%, and the stitching time is less than 30ms, achieving the effect of seamless stitching.

5. Conclusion

This paper proposes an infrared large-field stitching algorithm based on an improved ORB. Firstly, obtain the prior overlap area according to the built-in parameters of the image acquisition device, and perform feature point detection in the overlap area, which greatly saves the time of feature point detection; at the same time, two-way matching and the combination of RANSAC algorithm are used to improve the

registration rate of feature points; finally, an area-based weighting algorithm of fade-in and fade-out is adopted to fuse the images. Compared to the ORB algorithm, on the basis of keeping the algorithm steps relatively simple, seamless stitching of infrared images with a large field of view can be completed more quickly and accurately. Through the image results obtained after the experiment, it can be found that the algorithm proposed in this paper can stitch infrared images with a large field of view more quickly, has a higher matching rate, and can achieve seamless stitching.

References

[1] David G. L. Distinctive Image Features from Scale-invariant Keypoints. International Journal of Computer Vision. (S0920-5691), 2004; 60(2): pp 91-100.
[2] Herbert B, Tinne T, and Luc V G. SURF: Speeded up Robust Features. Proc of European Conference on Computer Vision, 2006; pp 404-417
[3] Xia M, Yahao L. UAV infrared image mosaic in photovoltaic array fault detection.Acta Solar Energy. 2020; 41(03): pp 262-269.
[4] Shaosheng D, Li Y. Research on high-precision infrared panoramic stitching algorithm based on ROI. Semiconductor Optoelectronics. 2020; 41(04): pp 572-577.
[5] Hongmei X, Qiang L, Lei X, He Q. Infrared image stitching based on SURF and two-way adaptive threshold registration. Aviation Weaponry; 2018; (06): pp 84-89.
[6] Chunbao S, Dongqing Y, Yunpeng L. Compare SIFT, SURF, BRISK, ORB, FREAK Algorithms from Multiple Angles. Beijing Surveying and Mapping. 2014; (4): pp 23-26.
[7] Deepak G V. Features from Accelerated Segment Test (FAST).
[8] Ethan R, Vincent R, Kurt K, et al. ORB: an efficient alternative to SIFT or SURF// IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain. November 6-13 2011; IEEE; 2011.
[9] Zhizhong S, Qi W, Fei C, et al. Improvement of Steer BREIF Feature Point Description Algorithm in ORB. Electronic Technology and Software Engineering. 2017; 1: pp 76-76.
[10] Jiguang C. Research on UAV Farmland Thermal Infrared Remote Sensing Image Mosaic Method Based on Overlap Priority. Northwest Sci-Tech University of Agriculture and Forestry. 2021.
[11] Lei X, Xinyu H, Yawei Y, Xiaoqin X. Infrared image mosaic based on improved SURF algorithm.Internet of Things Technology. 2020; 10(06): pp 48-51.
[12] Yixia C, Quansen S, Huanyu X, et al. Remote Sensing Image Matching Method Combining SURF Algorithm and RANSAC Algorithm. Computer Science and Exploration. 2012; 6(9): pp 822-828.
[13] Qing C, Bin Z, Jinglin S. Feature Image Matching Method Based on SIFT and RANSAC. Journal of East China University of Science and Technology. 2012; 38(6): pp 747-751.
[14] Lan Z, Bowen A, Fang C. An Infrared Image Stitching Algorithm Based on Feature Point Matching. Computer Applications and Software. 2015; (9): pp 195-196.
[15] Pinggai Z, Lisheng W, Shengwen Z. Research on Panoramic Parking Assist System Based on Image Fusion. Journal of Anhui Engineering University. 2015; (5): pp 58-62.

doi:10.3233/ATDE220207

Design and Laboratory Test of Vibration Excitation Device for Jujube Harvester

Zhiyuan ZHANG[a,b], Xianfei WANG[a,b] and Jingbin LI[a,b,1]
[a]*College of Mechanical and Electrical Engineering, Shihezi University, Shihezi,* 832003, *China*
[b]*Xinjiang Production and Construction Corps Key Laboratory of Modern Agricultural Machinery, Shihezi,* 832003*, China*

Abstract. In order to improve the working efficiency of excitation device of jujube harvester, an excitation device was designed by using eccentric block excitation mechanism. Through the combination of theoretical analysis and virtual simulation, the mass of eccentric block *m*, the motor speed *n*, the spring preload *f* were determined as test factors, and the angular acceleration *α* and amplitude *A* were used as evaluation indexes to carry out the combined test of orthogonal rotating center with three factors and five levels. First, this paper shows the maximum instantaneous angular acceleration *α* and the maximum amplitude *A* in space of the marking point of shift lever through 3D high-speed camera technology. Then, the Design-Expert v8.0.6.1 software was used for analysis of variance of experimental results, established the mathematical regression model of evaluation index and various relevant factors, and analyzed the influence of significant factors on evaluation index and optimized the test parameters. Final, the optimal parameters were determined as follows: eccentric mass *m* =233 g, motor speed *n* =1080 r/min, spring preload *f* =35 N. According to the combination of optimal parameters, the results shown that under the optimal combination of parameters, the average amplitude was *A*=46.73 mm, the average angular acceleration was *α*=11.72 rad/s^2, and the minimum inertia force generated by shell vibration was *F*=9.87 N. It could be seen that the excitation device satisfies the requirements for jujube harvesting. The study may provide theoretical basis and technical reference for the improvement of the excitation system of the jujube harvester.

Keywords. Jujube, harvester, motion simulation, amplitude, frequency

1. Introduction

Xinjiang is located in the hinterland of the Eurasian continent, which is a good place to breed high-quality jujube for its arid climate and abundant sunshine [1-3]. Statistics have shown that the total growing area of Xinjiang jujube reached 500,000 hm2 and 3,264,000 tons of jujube were produced by the end of 2017, making Xinjiang the biggest producer of jujube in China [4-5]. The booming of jujube in Xinjiang stimulated the market for jujube machine [6-7]. At present, Xinjiang jujube is mainly harvested by manual work. The low-efficiency and high-cost harvest method greatly restricted the sustainable development of Xinjiang jujube industry [8-9]. Hence, it is

[1]Corresponding Author, Jingbin LI, College of Mechanical Electrical Engineering, Shihezi University, Shihezi, China; Email: ljb8095@163.com.

urgent to realize mechanization to improve the efficiency and reduce the labor cost in jujube harvesting.

Jujube is an endemic species in China. The growing area and yield of jujube in China account for 99% of the whole world. So there are no mature jujube harvesting machine from other countries. But they have mature prototypes for other fruits e.g. citrus, mango, and apricot, etc. For example, Kececioglu developed a vibration harvester for olive harvesting. The harvester adopted eccentric block type excitation device to realize the harvesting. Researches have shown that the best harvesting happens at 20-28 Hz frequency and 10s vibrating at 20-30 mm amplitude [10]; Donald and Stephen designed an apple harvester which shakes the tree trunk with an eccentric block type excitation device. Researches showed that the average classification rate of the harvester is 87% when working, and the average vibration harvest rate is more than 95% [11]; Parameswarakumar and Gupta designed an inertial type slider-crank mango harvester which adopted branch shaking method for mango harvesting. Researches showed that the highest harvest rate can be achieved when the frequency is at 11-13 Hz and the amplitude is 76-102 mm. At this time, the damage to the fruit tree is also the smallest [12]. Wang Changqin and Xu Linyun et al. studied a shaking type fruit harvester, which adopted symmetrically placed eccentric block for the excitation device. The test on harvesting walnut showed that the average harvesting rate reached 89.5%-92.6% when the excitation frequency is 19-20 Hz, and no destructive damage on the trunk from the clamping device was found [13-14]; A 4YS-24 type jujube harvester was developed in the Institute of Machinery Equipment, Xinjiang Academy of Agricultural and Reclamation Science. This harvester also adopted the shaking harvest mode. The vibration exciter used an eccentric block to generate omnidirectional vibration. Tests showed that the harvest clean rate is as high as 91.4% for jujube trees with trunk diameter between 80-200 mm and growing row spacing of more than 4 m [15-16]; based on the dwarf and close planting mode of Xinjiang jujube, College of Mechanical and Electrical Engineering of Shihezi University developed a crown vibration-based all-hydraulic and self-propelled jujube harvester using slider-crank mechanism as the excitation source. The initial amplitude of the vibration dial is 15-20 mm, the excitation frequency is 15-20 Hz, and the harvest rate is 90% [17].

Above all, there are mainly two types of fruit harvester excitation device: eccentric action type excitation device, and fixed route slider-crank excitation device[18]. The first type generates inertial centrifugal force through the uniform circular movement of the eccentric block and makes reciprocated excitations under the effect of the inertial centrifugal force. The device is featured with simple structure, high vibration frequency and large exciting force. But eccentric block is required and needs to be changed when adjusting amplitude in this mechanism. Therefore, the device is mainly used for large harvesting machinery [19-20]; the later type called the slider-crank excitation device rotates the crank to drive the output rod and makes straight reciprocating movement. This device can realize a wide route but cannot make adjustment according to the characteristics of the harvesting target, which may harm the fruit tree [21-22].

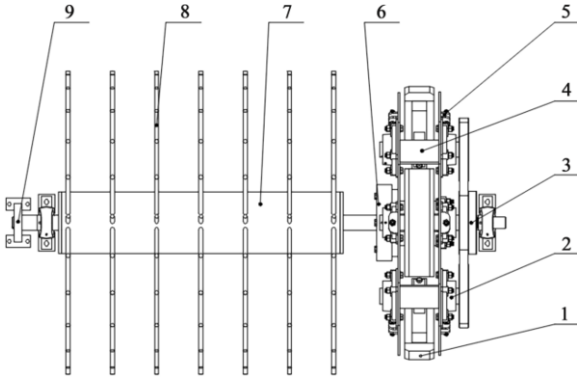
Based on previous researches on the effect of jujube tree excitation transfer and considering the high cost and low efficiency in Xinjiang jujube harvesting, this paper designed an excitation device for large jujube harvester. The excitation device used three eccentric blocks for the source. The paper mainly studied the eccentric mass, speed and the effect of damping device's spring preload on the amplitude and angular acceleration, and laboratory tests were also conducted. The research can provide

theoretical basis and technical support for the design of excitation device of jujube harvester.

2. Structure and working principle

2.1. Structure

The eccentric block type vibrator mainly consists of excitation device, rotary roller and damping device. The excitation device is made up of shell, timing pulley, timing belt, revolution axis, rotation axis, and three evenly placed eccentric blocks with equivalent mass; the rotary roller is composed by roller assembly and dial; the damping device includes the spring, damper belt and damper wheel; the excitation device and rotary roller are fixed by flange, the rotary roller and the damping device are fixed and connected by flat key, and the vibrator is fixed to the rack by two vertical plummer blocks. Fig. 1 showed the schematic of the vibrator.



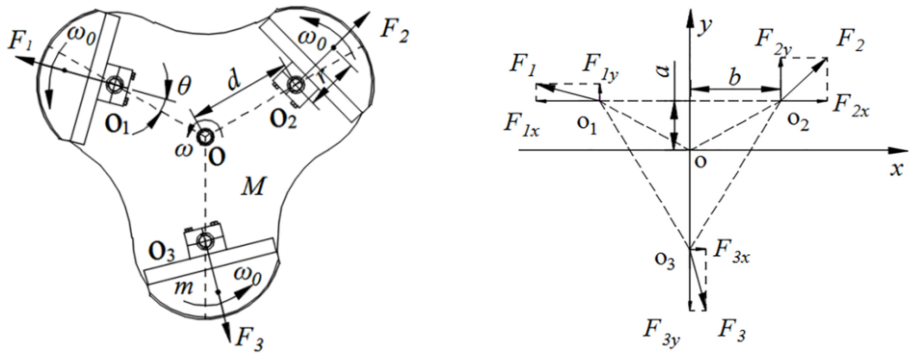
1. Eccentric block 2. Bearing seat 3. Synchronous pulley 4. Housing 5. Adjusting sheet 6. Flange 7. Rotating drum 8. Paddle 9. Damping device

Figure 1. The schematic of the vibrator.

2.2. Working principle

The eccentric block type vibrator, which is driven by hydraulic motor, is the major part of a jujube harvester. When working, the hydraulic motor drives the revolution axis and the timing pulley on the revolution axis to rotate; then the timing pulley will drive the three evenly placed eccentric blocks to rotate at a constant angular velocity; the rotating eccentric blocks will produce centrifugal force and further generate couple moment to drive the shell and rotary roller to make reciprocated rotation. Therefore, the device can adjust the motor speed to control the amplitude and frequency of the rotary roller and realize a high-frequency and small-amplitude harvesting of jujube.

3. Motion and vibration analysis of excitation device



a Excitation device motion analysis diagram

b Inertial centrifugal force decomposition diagram

Figure 2. Excitation device motion analysis.

The power source of the excitation device comes from three synchronously rotated eccentric blocks (angular velocity=ω_0) which generate a centrifugal force F on the shell through the rotation axis O_1, O_2 and O_3, then the couple on the shell will actuate the vibration of the rotary roller around the revolution axis O. In order to simplify the excitation device, the rotary roller and the shell were rolled up into one. The motion model is shown as Fig. 2.

Fig. 2a is the simplified model of the excitation device and Fig. 2b is the force diagram of the excitation device; k is the elastic coefficient; c is the damping coefficient; m is the mass of the eccentric, g; M is the mass of the shell and rotary roller device (hereafter referred to as shell), g; d is the distance between the eccentric axis's rotation axis and the revolution axis, mm; r is the distance between eccentric mass center and rotation axis, mm; R is the radius of the shell, mm; ω_0 is the eccentric block rotation angular velocity, rad/s; ω is the angular velocity of the vibration of the shell and rotary roller device, rad/s; θ is the angle that the eccentric block rotates around the rotation axis, rad; φ is the rotation angle of the eccentric axis's rotation axis around the revolution axis，rad; F is the inertial centrifugal force generated by the rotation of the eccentric block, N.

3.1. Excitation device motion analysis

When the excitation device is working, the eccentric blocks rotate at the angular velocity of ω_0 at time t; the rotation angle of the eccentric blocks is θ, and the rotation angle of the shell is φ, which means,

$$\theta=\omega_0 \cdot t \tag{1}$$

$$\varphi=\omega \cdot t \tag{2}$$

where, t is the time, s; F is the inertial centrifugal force produced by the rotation of the eccentric blocks, N; then

$$F = m \cdot r \cdot \omega_0^{\,2} \tag{3}$$

The three working eccentric blocks have the same inertial centrifugal force F because they have the same mass, geometric dimensioning and rotating angular velocity, then,

$$F_1 = F_2 = F_3 = F \tag{4}$$

where, F_1, F_2 and F_3 represent the inertial centrifugal forces of three eccentric blocks rotating the rotation axis O_1, O_2 and O_3, respectively. When the relative position of the three eccentric blocks with the shell is the same as the Fig. 2a, the Cartesian coordinate system O-xy was established. The inertial centrifugal force decomposition diagram of this system is shown as Fig. 2b. At this time, F_1, F_2 and F_3 can be decomposed into F_{1x}, F_{2x} and F_{3x} in the x-axis direction and F_{1y}, F_{2y} and F_{3y} in the y-axis direction. Then we sum the component forces in both the x-axis and y-axis:

$$\begin{cases} F_x = F_{1x} + F_{2x} + F_{3x} = 0 \\ F_y = F_{1y} + F_{2y} + F_{3y} = 0 \end{cases} \tag{5}$$

According to (5), the joint inertial centrifugal force produced by the eccentric blocks is 0 N; then we calculated the couples on the revolution axis for the composition forces of each inertial centrifugal force respectively (ω direction is the positive direction), and we can get:

$$\sum M_O(F) = F_{1x} \cdot a - F_{1y} \cdot b - F_{2x} \cdot a + F_{2y} \cdot b + F_{3x} \cdot d \tag{6}$$

where, a and b stand for the distance, m; according to the triangle relationship in Fig. 2b, we can get:

$$a = \frac{1}{2} \cdot d \ , \quad b = \frac{\sqrt{3}}{2} \cdot d \tag{7}$$

Combine (4), (6) and (7):

$$\sum M_O(F) = \frac{5}{2} \cdot d \cdot F \cdot \sin\theta \tag{8}$$

The analysis of (8) showed that the motion law of the shell satisfies the simple harmonic motion under the effect of driving couple moment. When $\theta \in (0, \pi)$, the driving couple moment drives the shell in the anti-clockwise direction to enable the shell's anti-clockwise acceleration vibration; when $\theta \in (\pi, 2\pi)$, the driving couple moment drives the shell in the clockwise direction to enable the shell's anti-clockwise reduction vibration.

3.2. Vibration analysis of excitation device

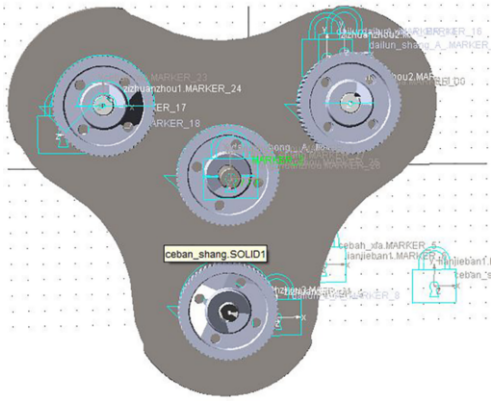
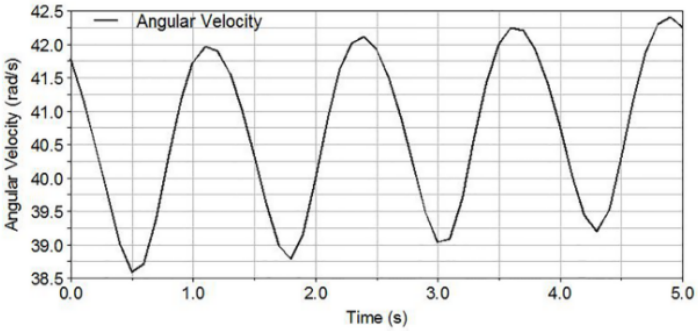
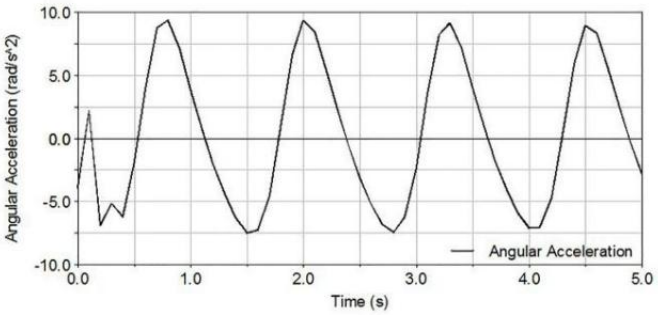


Figure 3. Vibrator simulation analysis

In order to verify the correctness of the theoretical analysis, the simplified model is drawn by SolidWorks software as the simulation model, saved as *.x_t format and imported into ADAMS software. The male rotating shaft is connected with the ground as a fixed pair, and a rotating pair is added between the male rotating shaft and the shell, and between the shell and their respective rotating shafts. The motion simulation of the device is carried out, as shown in Fig. 3. The simulation mainly analyzes the motion state of the simplified excitation device shell, where M=100 kg, m=10 kg, d=0.28 m, r=0.12 m, ω_0=3 rad/s, set the simulation time to 5 s, and obtain the time-varying curves of angular velocity and angular acceleration of shell and synchronous pulley, as shown in Fig. 4.



a Curve of angular velocity of shell motion with time



b Curve of angular acceleration of shell motion with time

Figure 4. The simulation of angular velocity and angular acceleration with time.

From the angular velocity and angular acceleration curves of ADAMS software simulation diagram, it can be seen that in the time period of 1.2 s, the motion curve is a regular simple harmonic curve, and the angular velocity increases with the increase of angular acceleration, but when the angular acceleration is maximum, the angular velocity does not reach the maximum value. In Fig. 4, it can be seen from the angular velocity diagram that the angular velocity value at the wave crest is gradually increasing because the vibration exciter damping device is ignored in the simulation; The stiffness coefficient and damping coefficient of the damping device have obvious influence on the diagonal velocity curve and angular acceleration curve, which proves the rationality of the vibration exciter design of the damping device. Therefore, the curve obtained from the motion analysis of ADAMS software is basically consistent with the curve derived from the theoretical formula.

3.3. Vibration analysis of excitation device

Apply conservation law of mechanical energy to the system:

$$W = T + U + V \tag{9}$$

where, W is the couple work of the inertial centrifugal force, J; T is the kinetic energy, J; U is the elastic potential energy, J; V is the Rayleigh dissipates energy, J.

Then the couple work of the inertial centrifugal force W is:

$$W = \sum M_O(F) \cdot \beta = 5/2 \cdot d \cdot F \cdot \sin\theta \cdot \varphi \tag{10}$$

The kinetic energy T of the system produced during vibration is:

$$T = 3/2 \cdot J_0 \cdot \omega^2 + 1/2 \cdot J_k \cdot \omega^2 \tag{11}$$

where, J_0 and J_k are the moments of inertia of the eccentric block and the shell, kg·m^2; the moments of inertia J_0 and J_k of the eccentric block and the shell on the revolution axis O are:

$$\begin{cases} J_0 = m \cdot r^2 + m \cdot R_c^2 \\ J_k = 1/2 \cdot M \cdot R^2 \end{cases}$$

Where, R_c is the distance from the eccentric mass center to the revolution O, mm;

$$R_c = \sqrt{d^2 + r^2 + 2 \cdot d \cdot r \cos\theta}$$

The spring preload of the damping device is f, N; then $f = k \cdot \Delta x$

The elastic potential energy U produced from the damping spring during the vibration of the system is:

$$\mathrm{U} = 1/2 \cdot k \cdot (x + \Delta x)^2 + 1/2 \cdot k \cdot (x - \Delta x)^2 = k \cdot \Delta x^2 = f \cdot \Delta x \tag{12}$$

where x and Δx are the original length and elongation of the damping spring, m; k is the spring stiffness factor, N/m; $\Delta x = \varphi \cdot R$.

During the vibration, the Rayleigh dissipates energy V produced from the damping device is:

$$\mathrm{V} = 1/2 \cdot c \cdot \dot{\varphi}^2 = 1/2 \cdot c \cdot \omega^2 \tag{13}$$

After combining the above formulas and ignoring the Rayleigh dissipates energy V, we can get the angular acceleration of the shell α:

$$\alpha = \frac{2 \cdot f \cdot R + 5 \cdot d \cdot m \cdot r \cdot \omega_0^2 \cdot \sin(\omega_0 \cdot t)}{3J_0 + J_k} \tag{14}$$

The accelerated velocity of the shell vibration is a, m/s^2, and:

$$a = \alpha \cdot R \tag{15}$$

According to (14), the shell can make reciprocating movement during the vibration; tests showed that the maximum breaking force of jujube handle is 6 N [3-4]. By combining (15) and $F = Ma$ (M is the shell mass, g; a is the acceleration of the shell vibration, m/s2) we can calculate the inertia force applied on the shell from the stimulation of the eccentric blocks and judge whether the eccentric device test-bed can satisfy the work requirement.

4. Test materials and methods

4.1. Test materials

Test device: eccentric block excitation device; test field: Xinjiang Production and Construction Corps Key Laboratory of Modern Agricultural Machinery, Shihezi University.

4.2. Test methods

This test studied the influencing factors to the vibration effect based on the structure and changing of working parameters of the excitation device.

4.2.1. Determination of influencing factors

According to the motion analysis, structure parameters and working parameters of the excitation device, this test selected three major influencing factors (eccentric mass *m*, damping spring preload *f*, and motor speed *n*) that influence the vibration effect of the excitation device.

(1) Eccentric mass

Eccentric mass has a direction effect on the working effect of the excitation device. According to (16) and (17), higher eccentric mass will cause larger inertia couple moment and angular acceleration and create better vibration effect of the excitation device. The study set the eccentric mass *m* at 198-242 g according to the jujube harvesting requirement, and the working condition and reliability of the testing apparatus.

$$\sum M_O(F) = 5/2 \cdot d \cdot F \cdot \sin\theta = 5/2 \cdot d \cdot m \cdot r \cdot {\omega_0}^2 \cdot \sin\theta \tag{16}$$

$$\alpha = \frac{2 \cdot f \cdot R + 5 \cdot d \cdot m \cdot r \cdot {\omega_0}^2 \cdot \sin\left(\omega_0 \cdot t\right)}{3J_0 + J_k} \tag{17}$$

(2) Motor speed

Motor speed *n* has a direct impact on the angular velocity *ω0* of three eccentric blocks rotating their own rotation axis. Formula (17) showed that the maximum angular acceleration *α* has a direction connection with *ω0*; the relationship between motor speed *n* and the angular acceleration of the eccentric block is: $n = \pi\omega_0 / 30$, where *n* is the speed, r/min; *ω0* is the angular velocity, rad/s; based on previous researches [4], the motor speed can be set at 960-1080 r/min when the resonant frequency of the sample jujube tree is 13-18 Hz.

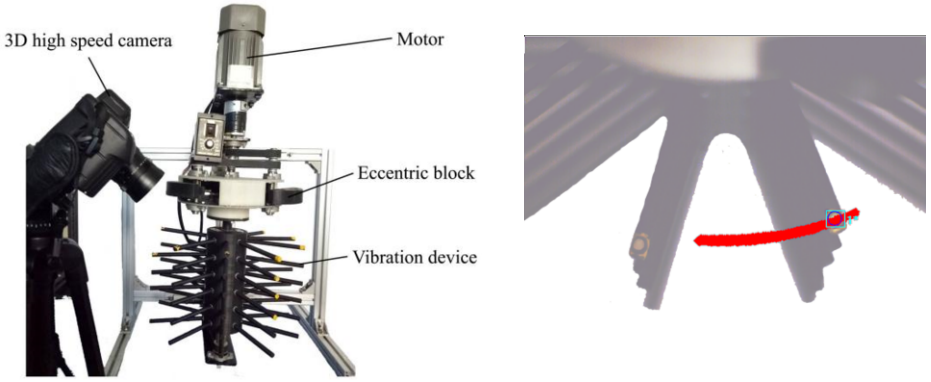
(3) Spring preload

The damping spring preload *f* has a large impact on the dial vibration of the excitation device. The larger the preload, the smaller the amplitude will be. According to (17), the larger the *f*, the larger the angular acceleration *α* will be; previous pre-test

showed that the best effect can be achieved when the damping spring preload f =25-45 N.

4.2.2. Determination of response indexes

This test employed 3D high-speed photograph technology and 3D ProAnalyst software to measure the response indexes of maximum amplitude A and maximum angular acceleration α of the top dial point of the excitation device under different working parameters. The spatial motion trail, amplitude and angular acceleration of the test-bed and dial point are shown as Fig. 5.



a. The test-bed of excitation device b. The motion trajectory of marker point

Figure 5. The motion analysis of device.

(1) Maximum amplitude

The amplitude of the excitation device has a large impact on the working effect of the vibrator. The larger the amplitude, the better the vibration effect will be. With dial's maximum amplitude A as the response index, we used 3D ProAnalyst to analyze the dial's spatial motion trail which was photographed by 3D high-speed camera, and then we confirmed the maximum amplitude A of the dial in this section by statistical analysis.

(2) Maximum angular acceleration

According to (17), the maximum angular acceleration α of the dial is closely related to the eccentric mass m, damping spring preload f, and motor speed n. With dial's maximum angular acceleration α as the response index, we used 3D ProAnalyst to analyze the dial's spatial motion trail which was photographed by 3D high-speed camera and we got the maximum angular acceleration α.

When using 3D ProAnalyst to conduct the test analysis, we selected a section of stable vibration (about 500 frames) as the test data collection samples, and confirmed the maximum amplitude A and maximum angular acceleration of the dial in this section by statistical analysis.

4.2.3. Design of test

This test used Design-Expert V8.0.6.1 to make an optimal three-factor and five-level orthogonal rotation center combination test. The eccentric mass m, motor speed n, and

damping spring preload f were also studied. The test factors and levels are shown as Table 1.

Table 1. Factors and levels of exciter test.

Level	Factor		
	Eccentric mass m/g	Motor speed n/(r/min)	Spring preload f/N
Upper asterisk arm(1.682)	242	1121	52
Upper level(1)	233	1080	45
Zero level(0)	220	1020	35
Lower level(-1)	207	960	25
Lower asterisk arm(-1.682)	198	919	18

5. Results and analysis

5.1. Regression analysis of test results

This test is divided into 20 groups, and each group was repeated five times. The average of the five tests was taken as the final test result, shown as Table 2.

Table 2. Experiment design and results.

No.	Factors and levels			Response index	
	X_1/g	X_2/(r/min)	X_3/N	Y_1/mm	Y_2/(rad/s^2)
1	207	960	25	109.72	7.93
2	233	960	25	179.77	9.79
3	207	1080	25	61.01	9.4
4	233	1080	25	112.89	10.15
5	207	960	45	92.67	9.01
6	233	960	45	156.98	10.95
7	207	1080	45	55.01	10.56
8	233	1080	45	29.56	12.84
9	198	1020	35	50.42	8.77
10	242	1020	35	158.85	12.15
11	220	919	35	145.21	8.78
12	220	1121	35	56.19	11.03
13	220	1020	18	98.9	8.63
14	220	1020	52	49.65	11.19
15	220	1020	35	74.85	9.83
16	220	1020	35	76.47	10.24
17	220	1020	35	66.97	10.11
18	220	1020	35	76.8	9.62
19	220	1020	35	61.24	10.22
20	220	1020	35	59.95	10.12

Note: X_1 is eccentric mass, g; X_2 is motor speed, r/min; X_3 is spring preload, N; Y_1 is amplitude, mm; Y_2 is angular acceleration, m/s^2.

Then a variance analysis was conducted using Design-Expert V8.0.6.1, and the coding regression mathematical models, which took dial's maximum amplitude A and maximum angular acceleration α as the response functions respectively and each influencing factor as the independent variable, were established, as shown in Formula (18) and Formula (19).

$$Y_1 = 71.51 + 25.13 \cdot X_1 - 31.51 \cdot X_2 - 15.52 \cdot X_3 - 13.49 \cdot X_1 \cdot X_2 - 10.38 \cdot X_1 \cdot X_3 + 13.25 \cdot X_1^2 + 11.86 \cdot X_2^2 \quad (18)$$

$$Y_2 = 9.96 + 0.92 \cdot X_1 + 0.66 \cdot X_2 + 0.76 \cdot X_3 + 0.2 \cdot X_1 \cdot X_3 + 0.2 \cdot X_2 \cdot X_3 + 0.16 \cdot X_1^2 \quad (19)$$

where, X_1, X_2 and X_3 stand for the eccentric mass m, motor speed n and spring preload f.

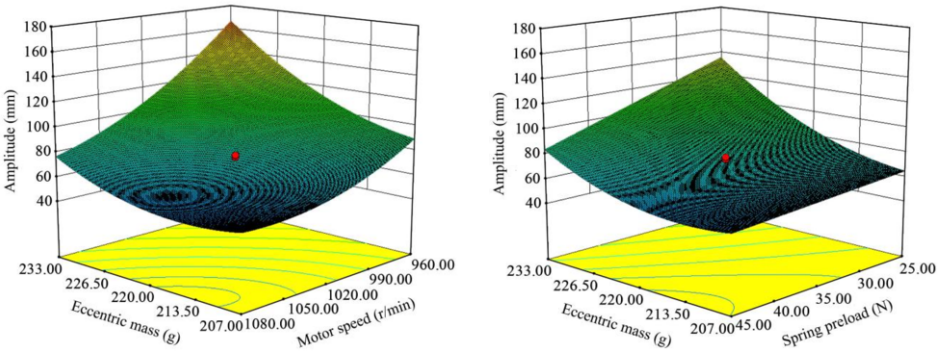
The regression analysis results in Table 3 showed that the equation evaluation indexes of amplitude A and P value of angular acceleration α are all smaller than 0.01, which means the result of the regression equation model in (17) and (18) is significant. Meanwhile, the determination coefficients R^2 are 0.9485 and 0.9342 respectively, which indicated that the two models can fit over 93% of the test results. Therefore, the above two regression equations have a good fitting relationship with the real situations, which is of practical significance.

Table 3. Variance analysis of regression models.

Experimental index	Difference source	Sum of squares	Degree of freedom	Mean square	*F* value	*P* value
Amplitude/mm	Model	32020.23	7	4574.32	25.19	< 0.0001(**)
	Residual	2179.10	12	181.59		
	Lake of Fit	1882.86	7	268.98	4.54	0.0575
	Purea Error	296.24	5	59.25		
Angular Acceleration/(m/s^2)	Model	26.40119	6	4.400199	74.43519	< 0.0001(**)
	Residual	0.768488	13	0.059114		
	Lake of Fit	0.465955	8	0.058244	0.962611	0.5433
	Purea Error	0.302533	5	0.060507		

Note: P<0.01(highly significant,**);P<0.05(significant,*); P>0.05(not significant)

5.2. Influence of test factors on amplitude



a. Effect of eccentric mass and motor speed on amplitude, while spring preload is 35 N

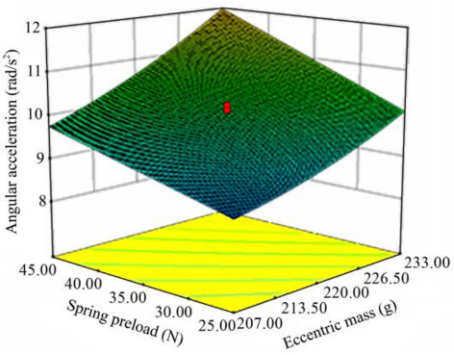
b. Effect of eccentric mass and spring preload on amplitude, while motor speed is 1020 r/min

Figure 6. Effect of experiment factors on amplitude.

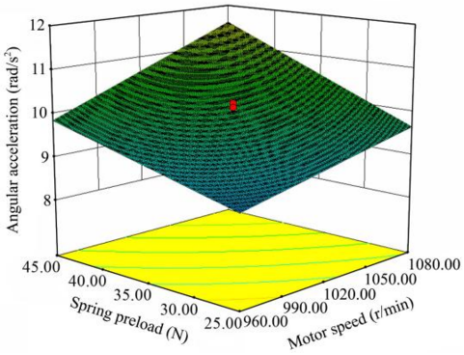
The 3D response surface diagram of eccentric mass *m*, motor speed *n* and spring preload *f* on the amplitude *A* is shown as Fig. 6a-6b. In Fig. 6a, when the spring preload *f* is at the 0 level (*f* =35 N), the law of the interaction influencing the amplitude *A* is as follows: the amplitude *A* increased slowly with the eccentric mass *m* while decreased rapidly with the motor speed *n*, showing that, under test level, the effect on the amplitude *A* from motor speed *n* is more significant than that from eccentric mass *m*. In Fig. 6b, when the motor speed *n* is at the 0 level (*n* =1020 r/min), the law of the interaction influencing the amplitude *A* is as follows: the amplitude *A* increased rapidly with the eccentric mass *m* while decreased slowly with the spring preload *f*, showing that, under test level, the effect on the amplitude *A* from eccentric mass *m* is more significant than that from spring preload *f*.

Therefore, the order of the significance of interaction influencing the amplitude *A* under test level is: motor speed *n*>eccentric mass *m*> spring preload *f*.

5.3. Effect of test factors on angular acceleration



a. Effect of spring preload and eccentric mass on angular acceleration, while motor speed is 1020 r/min



b. Effect of spring preload and motor speed on angular acceleration, while eccentric mass is 220 g

Figure 7. Effect of test factors on angular acceleration.

The 3D response surface diagram of eccentric mass *m*, motor speed *n* and spring preload *f* on the angular acceleration *α* is shown as Fig. 7a-7b. In Fig. 7a, when the motor speed *n* is at the 0 level (*n*=1020 r/min), the law of the interaction influencing the angular acceleration *α* is as follows: the angular acceleration *α* increased rapidly with the spring preload *f* and the eccentric mass *m,* while the interaction between eccentric mass *m* and angular acceleration *α* is more obvious. This shows that, under test level, the effect on the angular acceleration *α* from eccentric mass *m* is more significant than that from spring preload *f*. In Fig. 7b, when the eccentric mass *m*, is at the 0 level (*m*=220 g), the law of the interaction influencing the angular acceleration *α* is as follows: the angular acceleration *α* increased with the spring preload *f* and the motor speed *n*, but the change of the motor speed *n* is larger than the change of the spring preload *f* on the angular acceleration *α*, which shows that, under test level, the effect on the angular acceleration *α* from motor speed *n* is more significant than that from spring preload *f*.

Therefore, under test level, the effect from the eccentric mass m on the angular acceleration α is the most significant, followed by motor speed n and spring preload f.

5.4. Parameters optimization and verification

5.4.1. Parameters optimization

In order to further optimize the structural parameters and working parameters of jujube harvest vibrator and get the optimal combination of parameters, this paper made an optimization on the test device based on the requirement of high-frequency and small-amplitude jujube harvesting [31-32]. Design-Expert V8.0.6.1 was used to make an optimal analysis on the 3-index quadratic regression model. The constraints include: 1) Objective function: A[min]; α[max]; 2) Influencing factor constraints: $m\in$[-1,1] (eccentric mass 207-233 g); $n\in$[-1,1] (motor speed 960-1080 r/min); $f\in$[-1,1] (spring preload 25-45 N). After optimization, the Design-Expert V8.0.6.1 was used to choose the combination with the highest satisfaction degree as the optimal parameter combination: eccentric mass m=233 g, motor speed n=1080 r/min, spring preload f=35 N, model predicted amplitude A=50 mm, and angular acceleration α=13 rad/s^2.

5.4.2. Test verification

The test was conducted on the test-bed of the excitation device. Parameter combination: eccentric mass m=233 g, motor speed n=1080 r/min, spring preload f=35 N. The test showed that the test-bed shell mass M=3000 g, shell radius R=0.3 m, and the vibration acceleration of the shell can be obtained by $a=\alpha\cdot R$. The test was conducted five times and the average of the five tests was taken as the final test result (Table 4).

Table 4. Measurement results of test-bed testing.

	Max	Min	Avg.	S.D.	C.V.
Amplitude/mm	51.21	42.32	46.73	3.63	7.78%
Angular acceleration/(rad/s^2)	12.33	10.96	11.72	0.61	5.13%
Acceleration/(m/s^2)	3.70	3.29	3.51	0.18	5.13%

The test-bed testing using the optimal parameter combination is shown in Table 4. As shown, the average amplitude A=46.73 mm, and the average angular acceleration α=11.72 rad/s^2, which were lower than expected. The reason is due to the processing errors of the test-bed; through $F=Ma$ (M is the shell mass, and a is the vibration acceleration of the shell), we calculated the minimum inertial force imposed on the shell from the eccentric block excitation F=9.87 N, which is larger than the maximum breaking force of the jujube handle (6 N), showing that this excitation device can satisfy the requirement of jujube harvesting.

6. Conclusion

(1) This paper designed a excitation device used in large jujube harvester to deal with the current high-cost and low-efficiency harvesting of Xinjiang jujube. The excitation device is powered by three eccentric blocks, which can realize a continual work of the harvester;

(2) Based on the kinematic analysis of the excitation device, we worked out the relationship between the angular acceleration α and the eccentric mass m, motor speed n, and spring preload f;

(3) Through response surface tests, we analyzed the effect trend of eccentric mass m, motor speed n and spring preload f on the amplitude A and angular acceleration α, and established a quadratic multiple response model of amplitude A and angular acceleration α on the three factor levels; the order of the significance of each test factors influencing the amplitude A is: motor speed n, eccentric mass m, and spring preload f, successively; under test level, the effect from the eccentric mass m on the angular acceleration α is the most significant, followed by motor speed n and spring preload f;

(4) Design-Expert V8.0.6.1 is used to optimize the test results, and the optimal parameter combination is: eccentric mass m=233 g, motor speed n=1080 r/min, spring preload f =35 N;

(5) According to the test based on the optimal parameter combination, the average amplitude A=46.73 mm, average angular acceleration α=11.72 rad/s^2, and the minimum iner.tial force imposed on the shell from the eccentric block excitation F=9.87 N, which showed that this excitation device can satisfy the requirement of jujube harvesting.

References

[1] Rendong S, Tingting Q, Qqinbo Z, Anyan C, Wei Z, Shenduo L, Caiqi H. Design of Jujube Harvester Based on Hydraulic Driven. Journal of Qingdao Agricultural University (Natural Science). 2020 Nov; 37(3): pp 230-234.

[2] Yuanhang Z, Za K, ChengSong L, Huiming Z, Wei F, Jian C. Research and development of fruit collection techniques for landing fruit trees. Journal of A gricultural Mechanization Research. 2017 Jan; 39(1): pp 256-263.

[3] Fengkui Z, Fufeng Y, Zzhongjie L, Hong Z, Haipeng L, Ping L, Chaojun Z. Design and field testing of the air-suction machine for picking up Chinese jujube fruits. Journal of Fruit Science. 2019 Dec; 37(2): pp 278-285.

[4] Wei F, Zhiyuan Z, Yudong L, Junbing P, Jian C, Kai D, Huiming Z. Simulation Experiment in Lab on Force Transfer Effect of Jujube under Excitation. Transactions of the CSAE. 2017 Sep; 33(17): pp 65-72.

[5] Xiaocheng G, Qian'e L. Jujube Cultivation Techniques. Yangling, Shaanxi: Northwest A&F University Press; 2005.

[6] Can H, Bing L, Shulin H, Xiaokang Y, Xufeng W. Research status and development countermeasures on harvesting machinery of jujube in Xinjiang. Journal of Chinese Agricultural Mechanization. 2016 Jul; 37(7): pp 222-240.

[7] Gaokun S. Design and experimental study of pneumatic red dates harvesting machinery. Alar: Tarim University; 2014 Jun.

[8] Yaou Z, Za K, Chengsong L, Wei F, Lihong W, Linlin L. Design of Dwarf and Close Planting Jujube Harvester Pickup Device. Journal of Agricultural Mechanization Research. 2016 Apr; 38(4): pp 71-75.

[9] Junbing P, Yudong L, Kai D, Zhiyunan Z, Wei F. Design and Motion Simulation of Pneumatic Device for Picking up Ground Jujube. Journal of Agricultural Mechanization Research. 2018 Feb; 40(09): pp 139-143.

[10] Kececioglu. Atalet kuvvet tipli sarsici ile zeytin hasadi imkanlari uzerine bir arastirma (Research on olive harvesting possibilities with an inertia type shaker). Izmir, Turkey: Department of Agricultural Machinery, Agricultural Faculty, Ege University; 1975.

[11] Donald Peterson L, Stephen Miller S. Advances in Mechanical Harvesting of Fresh Market Quality Apples. Journal of Agricultural Engineering Research.1989 Jan; 42(01): pp 43-50.

[12] Parameswarakumar M, Gupta CP. Design Parameters for Vibratory Mango Harvesting System. Transactions of the ASAE. 1991 Jan; 34(1): pp 14-20.

[13] Changqin W, Linyun X, Hongping Z, Yeming C, Hua C. Development and experiment of eccentric-type vibratory harvester for forest-fruits. Transactions of the CSAE. 2012 Aug; 28(16): pp 10-16.
[14] Changqin W. Design and Experimental Study of the Eccentric-type Forest-fruit Vibratory Harvester. Nanjing: Nanjing Forestry University; 2012 Jun.
[15] Zhihui T, Xiangjin M, Congju S, Shouxing J, Yan Z, Xuan Z, Chaoming Q. Design and Experimental Investigation of Mechanical Vibration Tree Fruits and Nuts Harveste. Journal of Agricultural Mechanization Research. 2010 Aug; 32(8): pp 65-69.
[16] Zhihui T, Xiangjin M, Congju S, Shouxing J, Yan Z, Xuan Z, Qiwei C, Liqaing Y. Development of 4YS-24 type jujube harvester. Xinjiang Agricultural Mechanization; 2010 Feb; (1): pp 30-32.
[17] Rong H, Za K, Wei F, Lihong W, Hongyin Y, Yu S, Shizhe C. Design and Kinematic Analysis of Excitation Device of Dwarf and Close Planting Jujube Harvest Machine. Journal of Agricultural Mechanization Research. 2014 Feb; 36(02): pp 64-67.
[18] Luhong T, Yimunisha A, Ying D. Power harvesting of fruit trees in foreign countries. Xinjiang Agricultural Mechanization. 2004 Jun; (3): pp 54-56.
[19] Li GY. Selection and Calculation of Vibration Parameters of Vibratory Seed Seeder. Journal of Nanjing Forestry University (Nat. Sci. Ed.). 1984 Apr; 8(1): pp 101-114.
[20] Lenker DH, Hedden SL. Optimum shaking action for citrus fruit harvesting. Transactions of the ASAE. 1968 Jan; 11(3): pp 347-349.
[21] Peterson DL, Wolford SD. Mechanical harvester for fresh market quality stemless sweet cherries. Transactions of the ASAE. 2001 Jan; 44(3): pp 481-485.
[22] Liang Y,Yin LQ. Experimental Research on Harvesting Seabuckthorn Vibrator. The Global Seabuckthorn Research and Development. 2009 Dec; 6(4): pp 32-38.

doi:10.3233/ATDE220208

Collision Simulation of Inner Brace Grasping Manipulator During Operation Based on Grasping Impact Velocity Variation

Liangwen WANG[a,1], Yangguang KONG[a], Tianyun HE[b], Hongwei HAO[b], Ruolan WANG[c], Zhigang ZHANG[a] and Weiwei ZHANG[a]
[a] *School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Henan Provincial Key Laboratory of Intelligent Manufacturing of Mechanical Equipment, Zhengzhou 450002, China*
[b] *Henan Huanghe Whirlwind Co. LTD, Changge 461500, Henan, China*
[c] *School of international education, Zhengzhou University of Light Industry, Zhengzhou 450002, China*

Abstract. For the needs of gripping and transferring and assembling thin-walled fragile parts in industrial production, an internally supported manipulator configuration with finger-palm synergy features for thin-walled brittle cylindrical inner workpieces is proposed. Due to the poor impact resistance and low tensile strength of brittle materials, they are easily broken during the manipulator operation. In order to find the internal brace gripper finger configuration and stiffness matching for the operation of thin-walled fragile parts, and to explore the contact-collision law of the gripping process, the finite element model of the mechanical finger end parts was established by the integrated modeling method of Hypermesh and other software, and the change of the internal force of the mechanical finger contact with the workpiece when the gripping impact speed changes was studied. The corresponding constraints, loads and contact types are applied to the finite element model by LS-prepost software, and post-processing is performed to calculate the stress and strain clouds during the contact collision of the fragile parts. The simulation results show that the stress on the fragile part increases linearly with the increase of the impact speed of the manipulator: Under the speed of 4mm/ms, the stress increases linearly and slowly. When the speed goes from 0.5mm/ms to 4mm/ms, the stress increases about 8 times in X, XY direction and about 14 times in Maximum Principal, Y direction. Above the velocity of 4mm/ms the stress increases sharply and the model is destroyed. The results of the study establish the basis for optimizing the manipulator's operating process.

Keywords. Internal support manipulator, thin-walled fragile parts, impact velocity, integrated modeling, impact simulation

[1] Corresponding Author, Liangwen Wang, School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Henan Provincial Key Laboratory of Intelligent Manufacturing of Mechanical Equipment, Zhengzhou 450002, China; E-mail: w_liangwen@sina.com.

1. Introduction

In food packaging production, processing of non-metallic materials and various logistics lines, there is still a great demand for the use of manipulators to grip thin-walled fragile workpieces [1, 2]. During manipulation of fragile parts using manipulators, the permissible strength of the fragile parts limits the contact stress between the manipulator and the fragile workpiece. It is important to analyze and simulate the contact and collision of manipulators operating on thin-walled fragile workpieces.

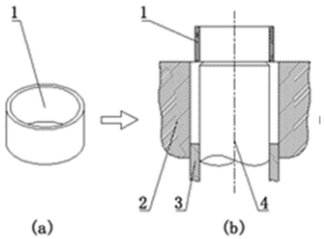
The key component of a robotic system to achieve a specific function is the end-effector [3]. Many researchers have discussed the gripping of manipulators [4], such as: Sintov et al, who studied the strategy of lateral grasping [5], and Vincent Babin et al, who improved the design of robotic grasping can grasp large and thin objects [6].

Studying the process of gripping fragile parts by robots and carrying out impact simulation of the gripping process is crucial to improve the structure of the robots. Similar studies conducted on the fragile parts operation process contain: Saurabh Rathod et al. [7] numerically analyzed the ballistic performance of ceramic-metal composites, Bresciani et al. [8] simulated the impact of projectiles on alumina ceramic tiles, and Wenjin Hu et al. studied the collision of FRP ships with ice at different velocities [9]. However, in the existing studies, there are fewer studies involving the contact touching of manipulators gripping thin-walled fragile workpieces during operation.

In our study, an internally braced robotic configuration with finger-palm synergistic features for thin-walled brittle cylindrical inner workpiece operations is proposed for the needs of gripping transfer and assembly of thin-walled fragile parts in industrial production. In order to explore the contact-collision law of the gripping process of the manipulator, the finite element model of the mechanical finger end parts is established by using the integrated modeling method of Hypermesh and other software, and the stress and strain clouds during the contact-collision of fragile parts are calculated, etc. The research process establishes the basis for improving the finger structure and operation process.

2. Overall structure of the manipulator

This paper studies the loading and assembly robot for a specific production line, whose operation process is shown in Figure 1.

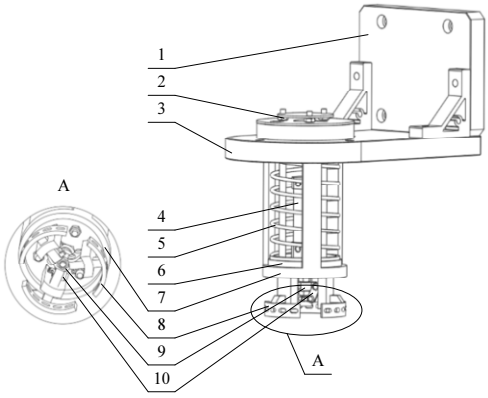


1. Inner wall workpiece 2. Outer mold 3. Bottom mold 4. Punch mold

Figure 1. The typical production process.

A special fragile ring-shaped workpiece (Fig. 1(a)) is transferred above the mold (Fig. 1(b)) and assembled on the convex die of the mold cavity. When a robot is used to mount the gripped workpiece on the convex die of the die cavity, it is necessary to complete the workpiece set on the convex die between the narrow gap between the outer wall of the central convex die of the molding die and the inner wall of the outer die, and its space avoidance is difficult to implement. Therefore, the gripping structure of external clamping is difficult to meet the requirements of this work, and the use of an internally supported gripping manipulator is a realistic choice. Considering the above-mentioned operational characteristics and technical requirements, the workpiece assembly process must avoid the convex die in the center of the molding cavity, so it is necessary to use a finger system retraction and palm downward pressure mode of operation.

The design of the mechanical finger is the key of the robot gripping fragile parts, to determine whether the instantaneous impact will exceed the tensile strength of fragile parts and cause them to break.

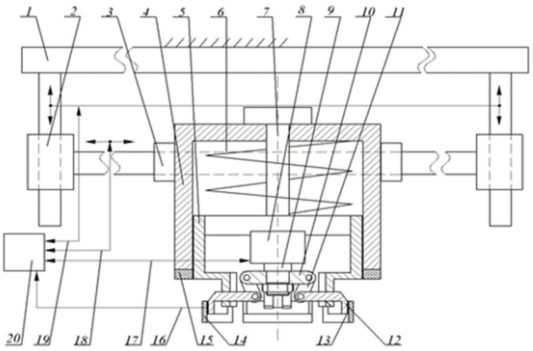


1. Rear connecting plate 2. Positioning nut 3. Connecting plate 4. Drive cylinder 5. Spring 6. Finger seat 7. Cylindrical palm 8. Finger body 9. Finger root 10. Finger joint

Figure 2. The manipulator structure with the inner brace grasping.

According to the way of force when gripping by the internal support robot, an internal support curved three-claw finger structure with a coating was designed as the main execution structure for fragile parts gripping and assembly. The related structure is shown in Figure 2.

In the design, the inner support type three fingers are nested in the inner cylindrical surface of the barrel-shaped palm, and the workpiece is assembled by the joint finger-palm downward pressure in a narrow space. The guide bar structure and the elastic structure are set to ensure that the fingers of the manipulator can return to the original position smoothly after the assembly is completed [10, 11]. The robot is mounted on the horizontal and vertical arrangement of the motion module according to the structural and functional requirements of the robot. The conformation formed is shown in Figure 3.



1. Equipment rack 2. Vertical movement module 3. Horizontal movement module 4. Cylindrical palm 5. Finger seat 6. Elastic element 7. Finger seat guide post 8. Finger driver 9. Drive piston 10. Finger root 11. Finger joint 12. Arc-shaped finger body 13. Pressure sensor 14. Soft finger 15. Palm cushion 16. Detection and feedback 17. Finger opening control 18. Horizontal motion control 19. Vertical motion control 20. Manipulator control system

Figure 3. The configuration of manipulator system.

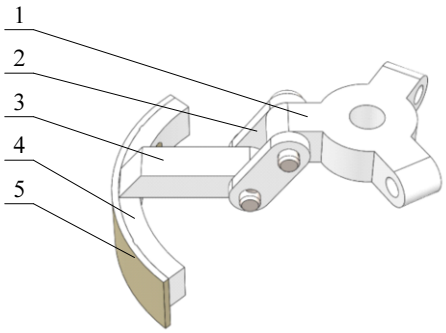
3. Manipulator grasping fragile parts contact impact simulation related technology

Different mechanical finger structures and gripping parameters will create different force transfer with the contact object. Here, Hyperworks, SolidWorks, and LS-prepost are used for integrated modeling, and LS-DYNA software is used to study the contact simulation of a robotic gripping workpiece.

There are two ways to build the finite element model of the mechanical finger frontal collision fragile part: (1) Create a 3D or 2D simulation model using LS-Prepost software, then build a geometric model by meshing, and finally perform a series of post-processing on it; (2) Support a variety of CAD software geometric model input, and then import it to Hypermesh pre-processing software to complete the meshing and other work. In this research, the process of mechanical finger collision fragile parts is selected to be processed by LS-DYNA software.

The structure of a typical manipulator drive system is shown in Figure 4, consists of cylinder-piston rod. The linear telescopic movement of the cylinder output is transformed into the opening and contracting action of the robot after the transfer of the finger root, finger joint and finger body to realize the grasping and placing operation of fragile parts.

Based on SolidWorks, Hypermesh, LS-DYNA integrated modeling, the finite element model of the inner-supported gripping mechanical finger impacting fragile parts was established, as shown in Figure 5. Among them, the inner-supported mechanical hand adopts the geometric structure of combining rigid fingers and soft fingers, and the fragile parts form a Rigid-Flexible-Brittle fragile transfer model. The inner-supported mechanical finger body material is aluminum alloy, and the mechanical finger end is covered with rubber.



1. Finger root 2. Finger joint 3. Finger body 4. Arc finger 5. Finger muscles

Figure 4. Manipulator cylinder connection structure.

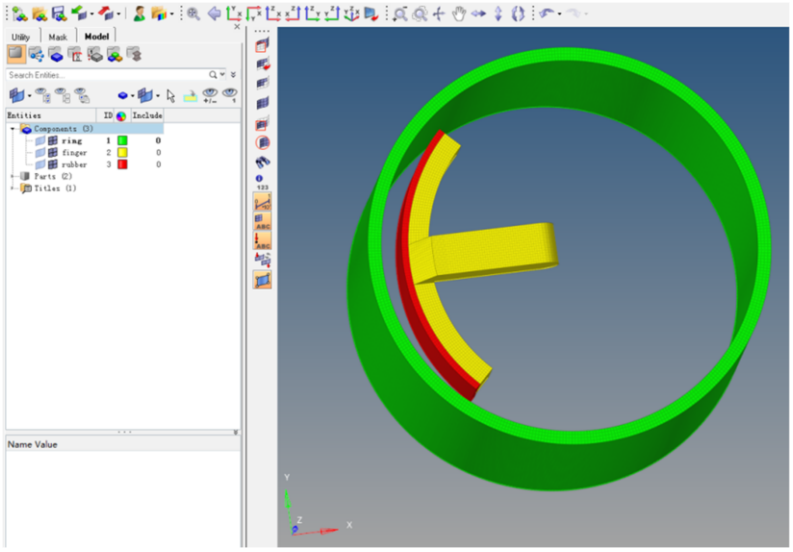


Figure 5. Finite element model of mechanical finger impacting fragile parts.

In general, for brittle materials, the compressive resistance is high, but the tensile resistance is weak. Due to the tensile waves generated by the free surface, which can lead to a large number of fragments, the compressive and shear strength is much higher than the tensile strength. Maximum tensile stress theory (first strength theory) points out that the main cause of fracture of brittle materials subjected to tension is the maximum tensile stress, regardless of the stress situation, the maximum tensile stress reaches a certain limit value, then fracture will occur. That is, one-way stretching as long as $\sigma_1(\sigma_2=\sigma_3=0)$ reaches the ultimate strength, fracture will occur, fracture conditions are: maximum tensile stress limit strength σ_1 > ultimate strength σ_b .

4. Collision Simulation Analysis of Fragile Parts Grabbing by Inner Supporting Robot Fingers

During Rigid-Flexible-Brittle fragile part operations, to study the force, deformation, energy transfer law of high-speed contact collision process between Rigid-Flexible-Brittle fragile Part, and the change law of different configurations of Rigid-Flexible finger system of stiffness and fragile parts contact. With a common brittle material like glass rings as fragile parts to be grasped, its inner radius is 27.75mm, outer radius is 30.75mm, and the depth of the ring is 35mm. The basic parameters of the model are: the basic speed of the mechanical finger gripping impact is 1mm/ms, the thickness of the soft finger is 1mm, the position of the mechanical finger is flush with the upper surface of the fragile part, and the initial distance between the finger and the fragile ring is 3mm.

The material parameters of the glass are obtained in the experiment by Johnson-Holmquist [12] ceramic material model. In the simulation analysis, we focus on the effect of the variation of the impact velocity on the magnitude of the stress on the fragile parts. Some of the parameters are shown in Table 1.

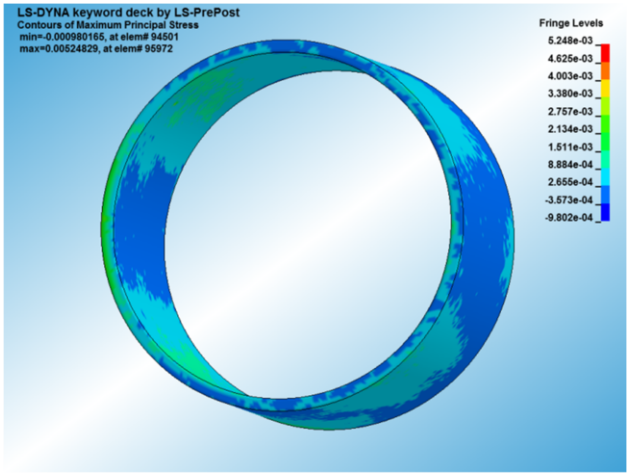
Table 1. Material parameters of glass.

Parameter Name	Parameter Value
Density	2.53e-6
Shear modulus	30.4
Reference strain rate	1
Maximum tensile strength	0.15

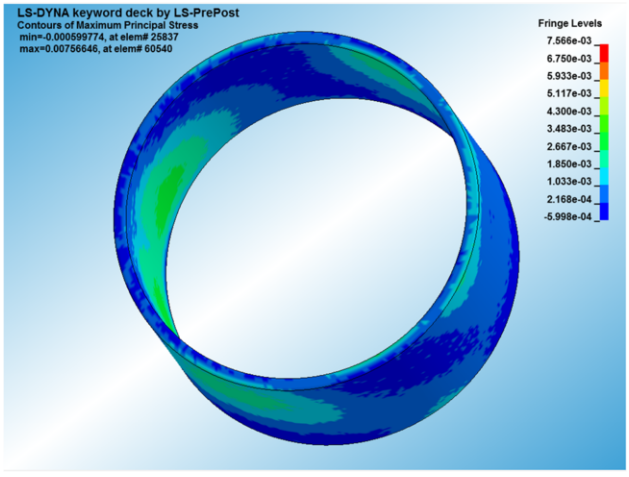
Based on the basic parameters, the finite element model is formed by choosing to vary the impact velocity. Select 10 groups of impact speed, respectively for 0.5mm/ms, 1mm/ms, 1.5mm/ms, 2mm/ms, 2.5mm/ms, 3mm/ms, 3.5mm/ms, 4mm/ms, 4.5mm/ms, 5mm/ms. Figure 6 shows the stress clouds of 3 groups models during the collision, they are the Maximum Principal stresses during the collision for 0.5mm/ms, 1mm/ms, 1.5mm/ms respectively. where: the positive direction of X is the direction of the finger grasping impact perpendicular to the surface of the workpiece, and the positive direction of Y is the direction along the outer surface of the workpiece facing upward.

Since the X-direction of the fragile part is subjected to the frontal impact of the circular-shaped manipulator finger, the X-direction is subjected to maximum compressive stresses. And the Y-direction is subjected to maximum tensile stresses. The results are shown in Table 2, the stresses is lowest in model 1 and highest in model 8 when no damage occurs in the model. And the impact speed is above 4mm/ms, the model will be destroyed.

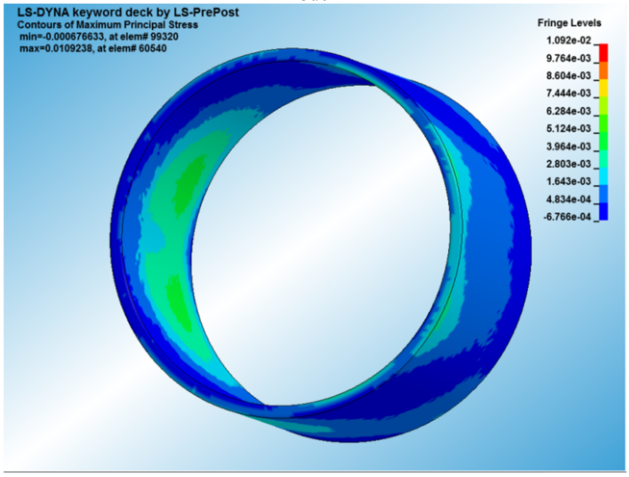
The stress change curve of the glass fragile parts in the X, Y, XY and other directions throughout the process is shown in Figure 7.



Model 1



Model 2



Model 3

Figure 6. Maximum Principal stresses map of fragile parts.

Table 2. Results of stress analysis.

Model	Velocity (mm/ms)	Maximum Principal (Gpa)	X- stresses (Gpa)	Y- stresses (Gpa)	XY- stresses (Gpa)
1	0.5	5.248e-3	2.681e-3	2.929e-3	1.749e-3
2	1	7.566e-3	3.77e-3	5.227e-3	2.126e-3
3	1.5	1.092e-2	5.367e-3	7.491e-3	2.866e-3
4	2	1.48e-2	6.95e-3	1.018e-2	3.827e-3
5	2.5	1.828e-2	8.558e-3	1.228e-2	4.769e-3
6	3	2.216e-2	1.037e-2	1.478e-2	5.659e-3
7	3.5	2.768e-2	1.242e-2	2.707e-2	7.834e-3
8	4	3.95e-2	2.185e-2	3.942e-2	1.626e-2
9	4.5	1.044	1.055	1.354	2.574e-1
10	5	1.015	1.031	1.154	1.995e-1

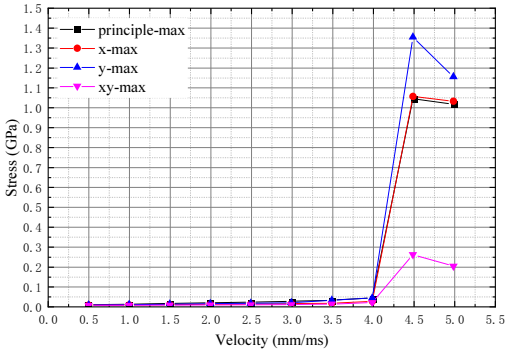


Figure 7. Stress variation diagram of fragile parts.

5. Conclusion

For ordinary glass, it has a higher compressive strength of 600~1200Mpa and a smaller tensile strength of 300~1000Mpa. By analyzing the impact velocity test index, the following conclusions are drawn:

In the same conditions, the higher the impact speed of the manipulator, the higher the impact on fragile parts. Under the speed of 4mm/ms, the stress increases linearly and slowly. When the speed goes from 0.5mm/ms to 4mm/ms, the stress increases about 8 times in X, XY direction and about 14 times in Maximum Principal, Y direction. Above the velocity of 4mm/ms the stress increases sharply and the model is destroyed.

We use Rigid-Flexible-Brittle fragile transfer model, which can significantly improve the gripping reliability. When designing the motion parameters of the manipulator, we need to select the operating speed of the manipulator when grasping according to the allowable stress conditions of the fragile parts.

Funding

This project was supported by National Natural Science Foundation of China (Grant No.52075500, 52005453), the Key Science and Technology Research Project of the

Colleges and Universities of Henan Province, China (21A460034), and Key Scientific Project of Henan Province (211110220200), China.

References

[1] Piazza C, Grioli G, Catalano M G. A century of robotic hands. Annual Review of Control, Robotics, and Autonomous Systems. 2019 May; 2(1): pp 1-32.
[2] Psomopoulou E, Karashima D, Doulgeri Z. Stable pinching by controlling finger relative orientation of robotic fingers with rolling soft tips. Robotica. 2018 January; 36(2): pp 204–224.
[3] Shibo C, Zhicheng T, Weiwei W. Multi-fingered dexterous hands: From simplicity to complexity and simplifying complex applications. Journal of Mechanical Engineering 2021 March; 57: pp 1-14.
[4] Máximo AR and Raúl S. Finding locally optimum force-closure grasps. Robotics and Computer-Integrated Manufacturing. 2008 June; 25(3): pp 536-544.
[5] Sintov A and Shapiro A. Dynamic regrasping by in-hand orienting of grasped objects using non-dexterous robotic grippers. Robotics and Computer-Integrated Manufacturing. 2018 April; 50: pp 114–131.
[6] Vincent B, David SO, Clément G. Stable and repeatable grasping of flat objects on hard surfaces using passive and epicyclic mechanisms. Robotics and Computer Integrated Manufacturing. 2019 February; 55: pp 1–10.
[7] Rathod S, Tiwari G, Chougale D. Ballistic performance of ceramic–metal composite structures. Materials Today: Proceedings. 2020; 41(5): pp 1125-1129.
[8] Bresciani L M, Manes A, Romano T A. Numerical modelling to reproduce fragmentation of a tungsten heavy alloy projectile impacting a ceramic tile: Adaptive solid mesh to the SPH technique and the cohesive law. International Journal of Impact Engineering. 2016; 87(2): pp 3-13.
[9] Wenjin H, Baoyu N, Xiaolong B. Impact performance of a glass fiber reinforced plastic ship with ice floes based on the nonlinear FEM. Journal of Vibration and Shock. 2018; 37(14): pp 262-268,276.
[10] Liangwen W, Tuanhui W, Yalin M. The bionic configuration of loading flexible manipulator based the inner bracing grab method. 2018 International Conference on Robots & Intelligent System (ICRIS). IEEE Computer Society; 2018; pp 513-516.
[11] Liangwen W, Weiwei Z, Yalin M. Realization of structure and spring parameter optimization of a manipulator for the pressing down assembling of the workpiece with cylindrical inner wall. Journal of Mechanical Transmission. 2021; 45(01): pp 146-151.
[12] Zhang X and Hao H. Laboratory test and numerical simulation of laminated glass window response to impact and blast loads. The 9th international conference on shock and impact loads on structures. 2011; pp 705-714.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 39
J. Xu et al. (Eds.)

doi:10.3233/ATDE220209

Facial Dyskinesia Recognition Based on Deep Learning

Xinwen HAN[a], Lizuo JIN[a] and Jun YAN[b, 1]
[a] *School of Automation, Southeast University, Nanjing, China*
[b] *Department of Geriatric Neurology, Affliated Brain Hospital of Nanjing Medical University, Nanjing, China*

Abstract. Facial dyskinesia has small movement range and short duration, thus the recognition effect is not ideal. To improve the recognition accuracy of facial movement disorders, a recognition method combining deep 3D Convolutional Networks (C3D) and Long Short-Term Memory (LSTM) is proposed. First, face detection and face alignment on original videos are performed, then the eye area based on facial landmarks is cropped. Second, C3D is used to extract spatio-temporal features of videos. Then LSTM further processes temporal features. Finally, softmax classifier is used to recognize and classify types of facial dyskinesia. According to experiment results, the approach we proposed can obtain a high accuracy rate.

Keywords. facial dyskinesia, C3D, temporal features, LSTM

1. Introduction

Facial dyskinesia refers to the loss of voluntary movement ability of facial muscles, showing some involuntary movement in a relaxed state. Facial dyskinesia is a kind of rare nervous system disease, which mostly occurs in the middle-aged and elderly. Clinically, it is manifested as blepharospasm and oromandibular dystonia, that is, involuntary twitch of muscles in the eye, perioral and mandibular areas. Blepharospasm is the most common first symptom[1-2], which is caused by excessive contraction of orbicularis oculi muscle and adjacent muscles. We mainly study the recognition of blepharospasm facial dyskinesia. At present, there are very few researches related to facial dyskinesia recognition, so we start research from related fields. Facial dyskinesia recognition can be understood as a special kind of action recognition. In traditional action recognition methods, researchers usually use manual design to extract relevant features. This method leads to poor generalization ability and low recognition accuracy in practical applications. Compared with traditional methods, deep learning can automatically perform feature learning and the recognition accuracy is higher. Therefore, the feature extraction method based on deep learning technology gradually replaced the method of manual feature extraction.

[1] Jun YAN: Department of Geriatric Neurology, Affliated Brain Hospital of Nanjing Medical University, Nanjing, China, 2519002414@qq.com

Convolutional neural networks was first used in action recognition task in [3], but the disadvantage of this approach is that it ignores the time dimension information in videos, which is actually the same as 2D image classification model. With the rapid development of deep learning, researchers have done a lot of research on temporal relationships. Two-stream Convolutional Networks was proposed for action recognition in videos[4]. In the network, two parallel convolutional networks have been applied to learn spatial-temporal domain features of video respectively. The output vectors of softmax layer of the two networks are merged to achive classification results afterwards. Three-dimensional convolutional neural network algorithm was first proposed for human action recognition[5] using 3D convolution kernel to learn deep spatiotemporal features. The authors use C3D for large-scale video classification[6], $3 \times 3 \times 3$ convolution kernels are used in both convolution layer and pooling layer, thus achieving high efficiency and accuracy on action recognition. Donahue and Hendricks proposed LRCN model[7]. The network uses CNN to process visual input and uses LSTM to handle temporal features, thus realizing action classification.

To capture the temporal changes of actions in videos more efficiently and make full use of spatial and temporal information for facial dyskinesia recognition, we consider using 3D ConvNets to extract spatiotemporal features and LSTM to further process temporal information between video frames.

2. Facial Dyskinesia Recognition Method Based on Deep Learning

Facial dyskinesia recognition we proposed consists of three parts. Firstly, face detection and face alignment are performed on original videos to obtain facial image sequence, and the region of interest of the eye is segmented. Second, the spatio-temporal features of facial movement are extracted by C3D network and LSTM network. Finally, types of facial dyskinesia are recognized and classified by softmax classifier. The overall strcture of the model is presented in figure 1.

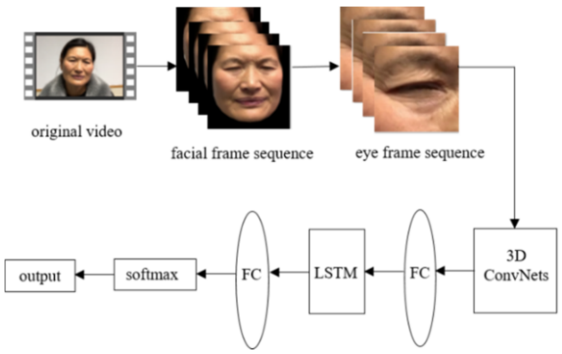


Figure 1. Framework of facial dyskinesia recognition method based on deep learning

2.1. Video Preprocessing

Since original videos contain a lot of irrelevant information, it will cause certain influence to efficiency and accuracy of the model recognition. Therefore, face detection and face alignment on video is performed firstly, and the ROI of the eye is obtained by

clipping according to the position coordinates of facial landmarks. Then, the cropped eye video is divided into frames and the resolution of a single image is uniformly adjusted to 112 pixels × 112 pixels.

In this paper we use Convolutional Experts Constrained Local Model (CE-CLM)[8] to detect key points of the face. The algorithm mainly includes two parts: the computation of probabilistic response maps based on convolutional experts network and update of parameters based on Point Distribution Model (PDM). The algorithm uses CEN to calculate response maps. An area of n×n pixels around the current estimated landmark is used as the input of the model and the appearance of different landmarks is modeled through the hybrid expert layer. Alignment probability response map of the landmark on a single pixel is used to help accurately locate the coordinates of each feature point. PDM is used to calculate the position of landmarks and adjust the shape of face. In addition, Multi-task Cascaded Convolutional Neural Network detector is used to initialize CE-CLM[9,10]. Figure 2 visualizes the 68 facial landmarks obtained by CE-CLM algorithm.

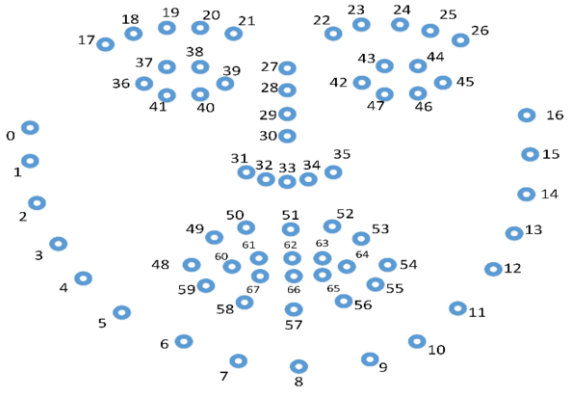
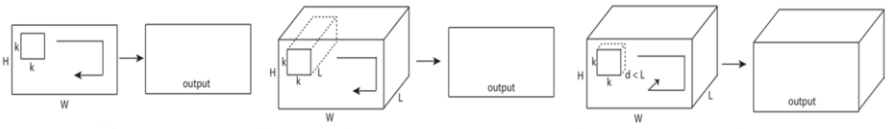


Figure 2. Location of 68 facial landmarks

Since we mainly study eye movement disorders, we first need to obtain eye image sequence. Each eye contains position coordinates of six landmarks, denoted as $p(x_i, y_i)$, where $i = 36,37, \ldots ,46,47$, respectively represent the positions of the left corner of the eye, the right corner of the eye, the upper eyelid and the lower eyelid. We take the center point of the line connecting the left and right corners of eye landmarks as the center of cutting area, and set the size of cutting area as 120 pixels × 120 pixels to get the eye frame sequence of each video.

2.2. Structure of 3D Convolutional Neural Network

Convolution operation is applied to extract spatial features in 2D CNN, while 3D CNN adds a time dimension to 2D convolutional neural networks. Using 3D convolution kernel to calculate multiple consecutive frames of images can capture important motion information from frame to frame and better model the time information. Figure 3 illustrates the difference. The output of applying 2D convolution on an image or multiple images is an image. Therefore, 2D convolutional neural networks lose the information in time dimension of input images. Applying 3D convolution on a video sequence leads to another sequence, thus preserving time information well.



(a) 2D convolution operation (b) 2D convolution operation on consecutive frames (c) 3D convolution operation

Figure 3. different convolution operations

We use 3D convolution and 3D pooling operation in C3D network. For simplicity, 3D convolution kernel and 3D pooling kernel have a size of d × k × k, where d and k are the depth aned the size of the kernel respectively. 3D CNN has 5 convolution layers, 3 max-pooling layers, 2 fully connected layers and a softmax output layer. The numbers of five convolutional layers filters are 32, 64, 64, 128 and 128, respectively. Table 1 presents the parameters of C3D network. Size of all convolution kernels is $3 \times 3 \times 3$ with stride of 1. Size of all max-pooling layers is $2 \times 2 \times 2$ with stride of 2 except for pooling1. The kernel size of first pooling layer is $1 \times 2 \times 2$ with the purpose of preventing fusing time information too early. The output of two fully connected layers is 128. In addition, we also define video clips size as c × l × h × w, where c and l indicate the number of video channels and frames, h and w represent height and width of video frames, respectively. Videos are cut into non-repeated 16 frames and input to the network. The size of input clips is uniformly adjusted to $3 \times 16 \times 112 \times 112$.

Table 1. 3D convolutional neural network structure parameters

Layer	Filters	Kernals	Stride	Output
Input	-	-	-	3×16×112×112
Conv1	32	3×3×3	1	32×16×112×112
Pooling1	-	1×2×2	2	32×8×56×56
Conv2a	64	3×3×3	1	64×8×56×56
Conv2b	64	3×3×3	1	64×8×56×56
Pooling2	-	2×2×2	2	64×4×28×28
Conv3a	128	3×3×3	1	128×4×28×28
Conv3b	128	3×3×3	1	128×4×28×28
Pooling3	-	2×2×2	2	128×2×14×14
Fc4	-	-	-	128
Fc5	-	-	-	128

2.3. Long Short-Term Memory Network

Recurrent neural network can well be used to perform time series feature extraction due to the hidden layer. However, if input sequence is too long, gradient explosion and gradient disappearance will occur.

LSTM is derived from RNN. By introducing a gate mechanism, the feature information of long time interval in the sequence can be extracted. LSTM uses forget gate, input gate and output gate to control current memory unit status C_t. Forget gate is defined as Eq.(1).

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + b_f) \tag{1}$$

where W_{xf}, W_{hf} are the weight vectors from input layer and hidden layer to forget gate, respectively. x_t is the input sequence at time t, h_{t-1} is the vector of hidden layer at time t-1, b_f is offset term of forget gate, σ is sigmoid function. Input gate is defined as Eq.(2).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{2}$$

where W_{xi}, W_{hi} are the weight vectors from input layer and hidden layer to input gate , respectively. b_i is offset term of input gate.

Forget gate and input gate work together to update current memory unit status C_t, which is given by

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3}$$

$$\tilde{C}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{4}$$

where $\tilde{C}_t$ is the output unit state of input gate. W_{xc}, W_{hc} are the weight vectors from input layer and hidden layer to cell state, respectively. b_c is offset term of cell state.

The output h_t of the hidden layer is determined by O_t and C_t, which is given by

$$h_t = O_t * \tanh(C_t) \tag{5}$$

$$O_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{6}$$

where W_{xo}, W_{ho} are the weight vectors from input layer and hidden layer to output gate, respectively. b_o is offset term of output gate.

2.4. Overall Network Structure

We propose a fusion network combining 3D convolutional neural network and LSTM network to classify facial dyskinesia. Figure 4 presents the overall network structure. The input of the network is preprocessed eye image sequence. C3D is used to capture spatial-temporal features of video frames and the feature vector output by the convolutional network is input into LSTM network to further learn temporal features.

We use LSTM network of two-layer structure. The number of hidden layer units is 128 and 64 respectively, connecting the output of the LSTM to fully connected layer. Finally, softmax classifier is used to obtain classification results. We add dropout layer behind fully connected layer to avoid overfitting and set its probability to 0.5. Besides, the network uses cross entropy as loss function.

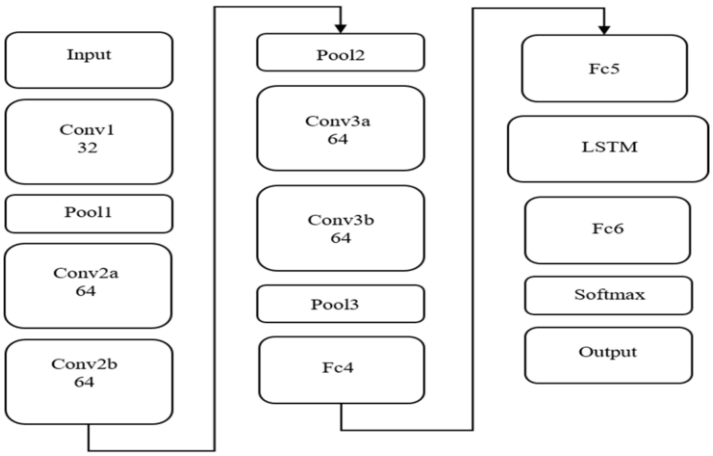


Figure 4. Overall network structure

3. Experiment and Analysis

3.1. Facial Dyskinesia Dataset

Since there is no publicly available dataset related to facial dyskinesia, we use self-collected facial video data of patients with facial dyskinesia and healthy controls. Our dataset includes 235 subjects, aged between 20 and 70. There are 344 videos in total with a resolution of 1920 × 1080. Frame rate is 30 fps. Table 2 presents the distribution of dataset.

Table 2. Distribution of facial dyskinesia dataset

Type of disease	Number of samples	Proportion
blepharospasm with severe eyelid closure	36	10.5%
blepharospasm with partial eyelid closure	154	44.8%
blepharospasm without eyelid closure	49	14.2%
normal	105	30.5%

On training, we unify the number of video frames for each sample to 240 frames. All samples were collected in a state of natural relaxation, sufficient illumination and strict experimental environment. There was no irrelevant facial motion noise during video shooting. Each sample in database is diagnosed and classified by clinical medical experts and divided into four categories: blepharospasm with severe eyelid closure, blepharospasm with partial eyelid closure, blepharospasm without eyelid closure and normal.

3.2. Experimental Settings

The experiment was operated on ubantu18.04 bit operating system, using tensorflow deep learning framework based on Python and NVIDIA GeForce GTX 1080ti graphics card for acceleration. The dataset is divided into two parts with a ratio of 7:3: training set and testing set. On training, the video frames are normalized and adjusted size to 112 × 112. Training is done by SGD and Adam optimization algorithm. Initial learning rate is 0.01 update by exponential attenuation method. Besides, L2 regularization is used to avoid overfitting. The optimization is stopped at 100 epochs.

3.3. Analysis of experimental results

Figure 5 records the change of average recognition rate of fusion model proposed under the four disease types. The abscissa (epoch) indicates the number of iterations and the ordinate (acc) indicates the recognition accuracy, where, train_ acc curve represents the recognition accuracy on training set, val_ acc curve represents the recognition accuracy on testing set. It can be seen that the accuracy increases rapidly within 20 epochs, reaches the peak after 90 epochs, and then fluctuates stably at about 95%.

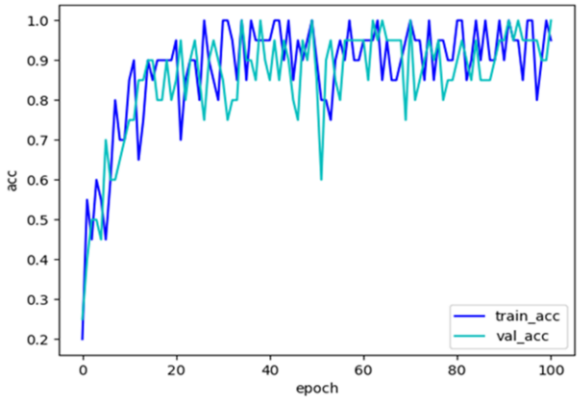


Figure 5. Accuracy curve of C3D+LSTM

In order to analyze the recognition effect of C3D and LSTM fusion model on each disease type, figure 6 gives the confusion matrix of the test set in which A, B, C and D represent "blepharospasm with severe eyelid closure", "blepharospasm without eyelid closure", "blepharospasm with partial eyelid closure" and "normal" respectively. Our method has a high recognition accuracy of 97% for blepharospasm with partial eyelid closure and normal, while the recognition rate of blepharospasm with severe eyelid closure is not high. We found that the manifestations of blepharospasm with partial eyelid closure and normal were significantly different from the other two diseases, so that the model performs well on these two disease types. On the other hand, the symptoms of blepharospasm with severe eyelid closure is similar to those of blepharospasm with partial eyelid closure. It is easy to be confused in recognition so that the accuracy is not high.

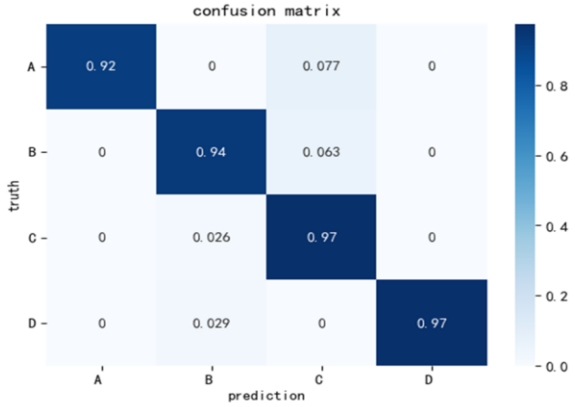


Figure 6. Confusion matrix of C3D+LSTM

We compare the average recognition rate of C3D + LSTM model with the other two methods(C3D, LSTM) in Table 3. C3D + LSTM fusion network significantly outperfoms C3D network and LSTM network on accurracy by 27.85% and 15.42% respectively. Extracting spatio-temporal features through C3D network and further learning temporal features through LSTM network can take full advantage of spatio-temporal information of video frame sequence so as to better model the motion information. This method greatly improves the accuracy of facial dyskinesia recognition in videos.

Table 3. Results of different networks

Method	Average recognition rate
C3D	67.30%
LSTM	79.73%
C3D+LSTM	95.15%

4. Conclusion

In this work we proposed a recognition method combining C3D network and LSTM network for addressing the problem of small action amplitude and short duration of facial dyskinesia. We obtained facial target area by preprocessing the original video and used C3D network to learn spatio-temporal features and LSTM network to further process temporal information. We showed that the fusion model can better model appearance and motion information in videos. This study provides a meaningful reference for the diagnosis and treatment of facial dyskinesia in clinical medicine.

Acknowledgments

This work is supported by the Special Funds of the Jiangsu Provincial Key Research and Development Projects, Grant/Award Number: BE2019612; Jiangsu Provincial Cadre Health Research Projects, Grant/Award Number: BJ17006.

References

[1] Guiling C, Min Y, Jinting H. Clinical research progress of meige syndrome. Chinese Journal of Laboratory Diagnosis. 2018 Mar; 22(3): pp 558-561.
[2] Stone J, Hoeritzauer I, Tesolin L, Carson A. Functional movement disorders of the face: A historical review and case series. Journal of the Neurological Sciences. 2018 Dec; 395: pp 35-40.
[3] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale video classification with convolutional neural networks. In: Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH. New York, NY: IEEE Press; 2014; pp 1725-1732.
[4] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos. In: Ghahramani Z, Welling, M, Cortes C, Lawrence ND, Weinberger KQ, editors. Proceedings of the 28th Conference on Neural Information Processing Systems. 2014 Dec; pp 08-13.
[5] Ji SW, Xu W, Yang M, Yu K. 3D convolutional neural networks for human action recognition. IEEE T Pattern Anal. 2013 Jan; 35(1): pp 221-231.
[6] Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning spatiotemporal features with 3D convolutional networks. Preceedings of IEEE International Conference on Computer Vision. Santiago, CHILE. New York, NY: IEEE Press; 2015. pp 4489-4497.
[7] Donahue J, Hendricks LA, Rohrbach M, Venugopalan S, Guadarrama S, Saenko K, Darrell T. Long-term recurrent convolutional networks for visual recognition and description. IEEE T Pattern Anal. 2017 Apr;39(4): pp 677-691.
[8] Zadeh A, Baltrušaitis T, Morency LP. Convolutional experts constrained local model for facial landmark detection. Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, HI. New York, NY: IEEE Press; 2017; pp 2051-2059.
[9] Zhang KP, Zhang ZP, Li ZF, Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE SPL. 2016 Oct; 23(10): pp 1499-1503.
[10] Baltrusaitis T, Zadeh A, Lim YC, Morency LP. OpenFace 2.0: facial behavior analysis toolkit. Proceedings of the 13th EEE International Conference on Automatic Face & Gesture Recognition. Xi an, PEOPLES R CHINA. New York, NY: IEEE Press; 2018; pp 59-66.

doi:10.3233/ATDE220210

Location Accuracy Evaluation of Subpixel Edge Detection Algorithm

Kunzi WANG [a], Mengxi YU [a], Liming XU [a,1], Lun SHI [a], Xiaobing FENG [a] and Dejin HU [a]
[a] *School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China*

Abstract. The edge detection algorithm based on sub-pixel can effectively improve the positioning accuracy of contour edges, but the quantitative evaluation method of sub-pixel positioning accuracy needs to be further studied. In this paper, a subpixel edge detection methodology supported by Zernike moment is initial analyzed. A triangle-based simulation method and a calibration-plate--based experiment method are proposed to judge the accuracy of the subpixel algorithm. The findings indicate that the proposed simulation method can quickly and quantitatively evaluate the accuracy of the sub-pixel detection algorithm. The effectiveness of the sub-pixel detection algorithm in improving the accuracy of edge detection is also verified.

Keywords. Subpixel, Zernike moments, edge detection, positioning accuracy, simulation

1. Introduction

In the imaging system, a photoreceptor cell not only receives the light hitting its own photo surface, but also receives the light hitting the adjacent photoreceptor. Especially for the edge points, the different reflection characteristics of the object and the background, and the integral effect of the sensor, cause the response of the sensor to step edges changing gradually. Thus, the edge can be characterized as a gray-scale distribution in the image [1]. The sub-pixel position can be determined as the edge point where the first derivative of the gray distribution reaches the peak, or the second-order derivative is zero. Because the image resolution is restricted by the hardware conditions of the vision system, sub-pixel positioning of the measurement edge has become a key in improving the detection accuracy of the vision system.

Main subpixel detection algorithms are as follows. The first is the interpolation technique [2], which interpolates the gray value or derivative to increase information and realize edge detection. This method has high computing efficiency but low positioning accuracy. The second is the fitting method [3] which simulates the edge gray change by function fitting, such as edge fitting by the Sigmoid function. This method is of strong anti-interference, but the solving speed will be affected when the model becomes complex. The third is the moment method [4]. According to the principle that the moment characteristics of an object remain unchanged before and

[1] Corresponding Author, Liming XU, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, Shanghai, China; E-mail: limingxu@sjtu.edu.cn.

after imaging, the moment method can be applied in sub-pixel positioning of targets such as edges and corners. This method takes into account both accuracy and speed, and is suitable for rapid and precise positioning of contour edges.

There are mainly two methods of simulation and actual image verification for the accuracy evaluation of sub-pixel detection algorithms. The operation speed and accuracy of edge detection algorithms was compared by using the actual image [5]. Zhou presented a computer simulation methodology to verify accuracy of the sub-pixel edge detection algorithm [6]. However, few studies evaluate the detection effect of sub-pixel algorithms quantitatively. Therefore, this paper proposes a triangular simulation method and a calibration board test method for quantitatively evaluating the positioning accuracy of the sub-pixel boundary detection algorithmic program supported by Zernike moment operator, and the pixel boundary detection algorithm of Canny operator comparatively.

2. Principle of algorithmic program supported by Zernike moment

The Zernike moment operator [7] calculates four parameters of pixels through the orthogonal moments of the image, and then judges the edge point. Figure 1 is a desired model for sub-pixel margin detection of an graphic . The k is grayscale difference, and l is the vertical distance from the middle of the circle to L. The φ is the angle between L and y axis. When the angle is 0, it becomes Figure 1(b) [8].

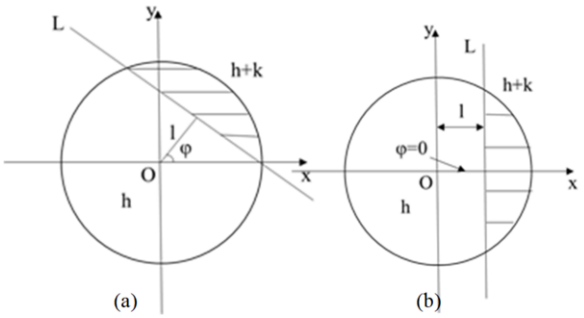


Figure 1. Desired model of subpixel margin detection

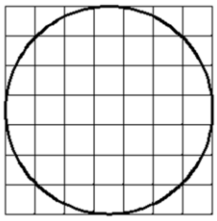


Figure 2. Zernike 7×7 template.

First, calculate the 7×7 convolution template as shown in Figure 2.

In the Zernike moment definition formula $A_{mn} = \frac{n+1}{\pi}\iint_{x^2+y^2\le 1} f(x,y)V^*_{\mathrm{nm}}(\rho,\ \theta)\ dxdy$. Let $f(x,y)=1$ and denote the template of Zernike moment A_{nm} as M_{nm}, then there is:

$$M_{nm} = \iint_{x^2+y^2\le 1} V^*{}_{nm}(\rho,\theta)dxdy \tag{1}$$

$$V^*{}_{nm}(\rho,\theta) = \sum \frac{(-1)^{s}(n-s)!\rho^{n-2s}}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!} e^{jm\theta} \tag{2}$$

Let C denote the area of the unit circle shown in Figure 2, and S_{ij} denote the area of the small square of row i and column j. Then the coefficient of current position of the template M_{nm} is:

$$M_{nm-ij} = \iint_{C \cap S_{ij}} V^*_{nm}(\rho, \theta) dxdy \tag{3}$$

The sub-pixel edge coordinates of the Zernike moment operator can be finally calculated [9]:

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \frac{N}{2} l \begin{bmatrix} cos\,\varphi \\ sin\,\varphi \end{bmatrix} \tag{4}$$

3. Triangular Simulation Evaluation Method

For estimating the positioning precision of sub-pixel algorithm, several simulation experiments are designed to study the positioning accuracy of pixel and sub-pixel edge extraction.

First, a standard white image of isosceles right-angled triangle is created based on Matlab as shown in Figure 3. The length of right-angle side of the triangle is of 100 pixels. It locates in the center of a black standard image with a size of 200×200 pixels, OXY is the image coordinate system.

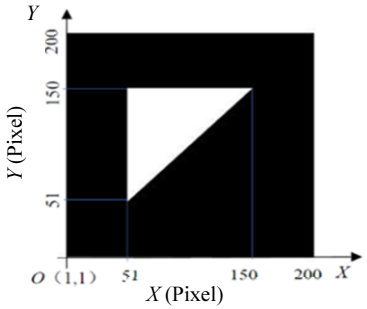
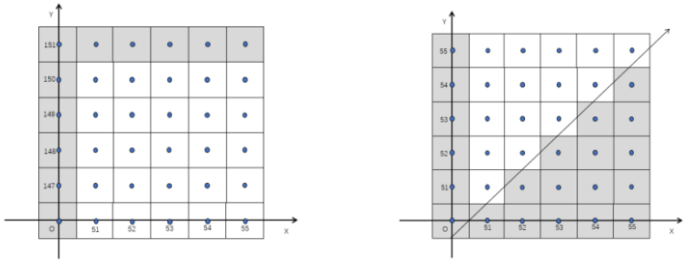


Figure 3. Image for triangular simulation.

Figure 4(a) is an enlarged view of the edge of right side of the triangle, and Figure 4(b) is an enlarged view of the hypotenuse edge of the triangle. The solid dots represent the coordinate positions of the pixels, and the sides of the square represent the boundaries of the pixels. It can be found that the theoretical right side edge of the triangle is on the boundary between the pixels. It does not pass through the center of the pixel, but passes through the midpoint of the line connecting the centers of two adjacent pixels.



(a) Enlarged view of right-side edge. (b) Enlarged view of the hypotenuse edge.

Figure 4. Schematic diagram of the distribution of pixels on the edge of a triangle.

According to the created triangle, the x coordinate of the left right side of the

triangle is x=50.5 pixel, and the y coordinate of the upper right side is y=150.5 pixel. The intercept of the hypotenuse on the y axis is -0.5, and the slope is 1. The hypotenuse can be expressed as:

$$y=x-0.5 \tag{5}$$

The boundary of the triangle is edge extracted and the position is calculated by Canny operator [10] and Zernike moment based sub-pixel algorithm, respectively. The extracted information are shown in Table 1.

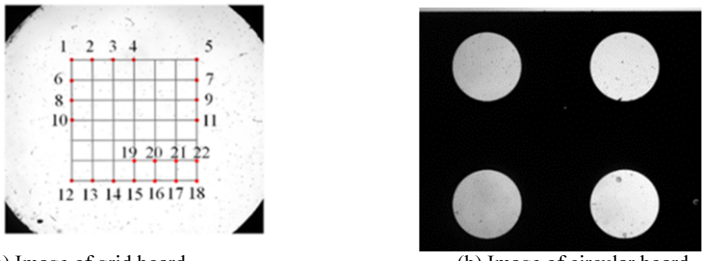
Table 1. Triangle boundary position determined based on two algorithms (unit: pixel).

Methods	Left boundary	Upper boundary	Hypotenuse
Pixel method	x= 50	y=151	y=0.9998x+0.52
Zernike moment method	x= 50.5335	y=150.4665	y=1.0002x-0.5260
theoretical value	x= 50.5	y=150.5	y=x-0.5

Analyzing the results, it can be obtained that the coordinate values of two right side of the triangle extracted by the Zernike moment method differ from the theoretical coordinates by 0.0335 pixels. The position of the right side extracted by Canny operator is 0.5 pixels off the theoretical value. The sub-pixel method can subdivide pixels by 1/0.0335=29.85 times. The slope of the hypotenuse extracted from either of the two edge extraction methods is different from the theoretical value by 0.0002, which can be ignored. Thus, it can be calculated that the normal distance between the hypotenuse extracted by Zernike moment method or by Canny operator method and the theoretical hypotenuse is 0.0184 pixels or 0.721 pixels, respectively. The sub-pixel subdivision ability reaches 1/0.0184=54.34 times. Therefore, the subdivision capability supported by Zernike moments algorithm can reach 30 times or more in the simulation.

4. Experimental evaluation of positioning accuracy

To verify availability of the sub-pixel edge detection algorithm for the actual image edge detection, experiments are designed and conducted by grid calibration boards with grid length of 1mm and accuracy of ±1μm. The vision system consists of a 5 megapixels CCD camera, a telecentric lens and a parallel backlight. The image of calibration board is shown in Figure 5.



(a) Image of grid board. (b) Image of circular board.

Figure 5. Image of calibration board.

The calibration board is calibrated first. The pixel-level calibration is supported by Canny operator and the subpixel-level calibration is supported by Zernike moment. The nominal length of measured line is 6000 μm.

Table 2. Measurement results of grid calibration board.

Measured line	Pixel length 1 (pixels)	Calibration 1 (μm/pixels)	Sub pixel length 2 (pixels)	Calibration 2 (μm/pixels)
Line 1-5	860.02	6.977	860.73	6.971
Line 6-7	859.76	6.979	860.68	6.971
Line 8-9	859.68	6.979	860.53	6.972
Line 10-11	859.35	6.982	860.14	6.975
Line 1-12	859.72	6.979	860.50	6.975
Line 2-13	859.56	6.982	860.47	6.973
Line 3-14	859.37	6.980	860.43	6.973
Line 4-15	859.49	6.981	860.35	6.974
Average	859.62	6.980	860.50	6.973

The calibration result at pixel level is 6.980μm/pixel, and the calibration result at sub-pixel level is 6.973μm/pixel. The calibration results are used in pixel level algorithm and sub-pixel algorithm respectively.

Three sets of line pairs 15-16 and 19-20, 16-17 and 20-21, 17-18 and 21-22 are intercepted from the lower right corner of the calibration board. Since the width of the grid line cannot be ignored, the pixel and sub pixel algorithms are used for calculating the distance between the outer sides of the line pair, inner sides of the line pair, and finally the center distance of the two straight lines. The laboratory finding are shown as follows.

Table 3. Distance measurement of straight-line pair (unit: μm).

Algorithm	Outer distance of line pair	Inner distance of line pair	Center distance of line pair
Canny operator	1045.413	962.503	1003.958
Zernike moment	1043.344	960.160	1001.514
Nominal value	-	-	1000.0±1

The deviation between the center distance of the line pair obtained by the sub-pixel algorithm and the nominal distance is 1.514μm. The distance deviation obtained by the Canny operator is 3.958μm. Since the sub-pixel calibration result is 6.973μm/pixel, the precision of the sub-pixel technique is significantly improved. The difference between the deviation and the nominal error is 0.514μm, which is about 1/14 of the full pixel resolution.

Finally, a circular calibration board with a diameter of 3mm and a nominal error of ±1μm is used for the verification test. The image of the calibration board is shown in Figure 4(b), and the calibration results based on the grid calibration board are used for measurement.

The two methods were used to extract the contour of the round hole in the calibration board. The average values were calculated.

Table 4. Measurement of hole parameters of the calibration board (unit: μm).

Method	Roundness	Radius
Canny operator	14.118	1507.524
Zernike moment	8.239	1502.365

The roundness measured by the sub-pixel technique is significantly smaller than by pixel method. Besides, difference between the radius measured by the sub-pixel method and the theoretical radius is 2.365μm, which is 1.365μm off from the nominal

error and significantly smaller than the radius deviation of 7.524 μm measured by the pixel method. However, there still exists a certain measurement error, which probably caused by the imaging quality and manufacturing precision of calibration board, and calibration error, et al.

5. Conclusions

This paper mainly studies a simulation and test method for evaluating the positioning accuracy of a sub-pixel edge measurement algorithm. The simulation findings illustrate that resolution error of sub-pixel edge detection algorithm can reach about 1/30 pixel supported by Zernike moment. The test findings indicate that resolution error by using the sub-pixel algorithm is not more than 1/14 pixel. The positioning error mainly comes from the image quality, manufacturing accuracy of the calibration board. The proposed method is feasible and convenient for evaluating the positioning accuracy of two algorithms, and can have long-term application to evaluate positioning accuracy of other sub-pixel edge detection algorithms.

Acknowledgment

Funding for this study was provided by the National Natural Science Foundation of China (No. 52075331).

References

[1] Ahmad I, Moon I, Shin S J. Color-to-grayscale algorithms effect on edge detection—A comparative study. 2018 International Conference on Electronics, Information, and Communication (ICEIC). IEEE; 2018; pp. 1-4.
[2] Sun Q, Hou Y, Tan Q. A subpixel edge detection method based on an arctangent edge model. Optik, 2016; 127(14): pp. 5702-5710.
[3] Xu G S. Sub-pixel edge detection based on curve fitting. 2009 Second International Conference on Information and Computing Science. IEEE; 2009; 2: pp. 373-375.
[4] Duan Z, Wang N, Fu J, et al. High precision edge detection algorithm for mechanical parts. Measurement Science Review. 2018; 18(2): pp. 65-71.
[5] You Y. The sub-pixel edge detection based on improved Zernike moment for brain CT image. DEStech Transactions on Computer Science and Engineering. 2016.
[6] Zhou F, Zhao J, Ye T, et al. Fast star centroid extraction algorithm with sub-pixel accuracy based on FPGA [J]. Journal of Real-Time Image Processing. 2016; 12(3): pp. 613-622.
[7] Yu W, Ma Y, Wu X, et al. Research of improved subpixel edge detection algorithm using Zernike moments. 2015 Chinese Automation Congress (CAC). IEEE; 2015; pp. 712-716.
[8] Xie X, Ge S, Xie M, et al. An improved industrial sub-pixel edge detection algorithm based on coarse and precise location. Journal of Ambient Intelligence and Humanized Computing. 2020; 11(5): pp. 2061-2070.
[9] Huang C, Jin W, Xu Q, et al. Sub-Pixel Edge Detection Algorithm Based on Canny–Zernike Moment Method. Journal of Circuits, Systems and Computers. 2020; 29(15): p. 2050238.
[10] Song R, Zhang Z, Liu H. Edge connection based Canny edge detection algorithm. Pattern Recognition and Image Analysis. 2017; 27(4): pp. 740-747.

doi:10.3233/ATDE220211

A Simple Method for Position Analysis of Stephenson-III Spherical Six Bar Mechanism

Quanwei ZONG [a], Hua LU [a,1], Zhuoya AN [a] and Fudong ZHANG [a]
[a] *College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou, China*

Abstract. Aiming at the position analysis of Stephenson III spherical six bar mechanism, a simple method to solve its input-output equation is given. The Stephenson III spherical six bar mechanism is regarded as composed of basic spherical four-bar chain and spherical two-bar group. The basic coordinate system and branch coordinate system are established respectively. The coordinates of each hinge point are solved with the help of geometric principle and displacement rotation theory. Based on the motion constraints of the basic spherical four-bar chain and the coupling constraints with the spherical two-bar group, the constraint equations of the spherical six bar mechanism are established by using spherical trigonometry. The constraint equations are simplified and eliminated by Sylvester's resultant elimination method and triangular transformation formula, and then the constraint equations of the mechanism are obtained.

Keywords. Location analysis, Displacement rotation theory, Sylvester resultant elimination method, Spherical trigonometry

1. Introduction

The research content of mechanism kinematics can be divided into motion analysis and motion synthesis. Kinematics analysis includes position analysis, velocity analysis and acceleration analysis. Position analysis is the most basic task of motion analysis [1,2]. Based on the constraint relationship between mechanism components, the establishment of constraint equations and the forward and inverse solution of mechanism position can not only lay a solid foundation for the analysis of mechanism speed and acceleration, but also lay a solid foundation for the solution of mechanism workspace. Then we can more systematically and deeply study all kinds of existing mechanisms, master the law and scope of their motion transformation, and lay a good foundation for designing new and better mechanisms or machines.

The spherical surface has the characteristics of both planar mechanism and spatial mechanism, it has always been highly concerned by scholars at home and abroad. Based on spherical trigonometry, On the basis of a new 4-DOF spherical parallel mechanism proposed by Liu et al.[3], the positive and negative kinematics of the spherical parallel mechanism was solved, and its good kinematic characteristics were

[1] QuanWei Zong, College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou, China; E-mail: 17839194960@163.com.

verified. Hernandez s et al. [4] comprehensively designed the spherical mechanical wrist based on the analysis of Stephenson spherical six bar mechanism based on displacement rotation theory. This paper based on the geometric principle, displacement rotation theory and spherical trigonometry, the input-output equations of Stephenson-III spherical six bar mechanism are solved by analytical method, The solution process is simple, easy to understand and program.

2. Establishment of constraint equations of spherical six bar mechanism.

In the analysis and synthesis of spherical mechanism, in order to facilitate description and calculation, it is necessary to use spherical coordinates to study spherical mechanism. The position of any point on the sphere can be expressed by coordinate parameters $P(r,\alpha,\beta)$ [5],as

$$\begin{cases} P_x = \cos\beta\cos\alpha \\ P_y = \cos\beta\sin\alpha \\ P_z = \sin\beta \end{cases} \tag{1}$$

The variation ranges of r , α and β are $-\infty \le r \le \infty$, $-180° \le \alpha \le 180°$ and $-90° \le \beta \le 90°$ respectively. As shown in Figure 1.

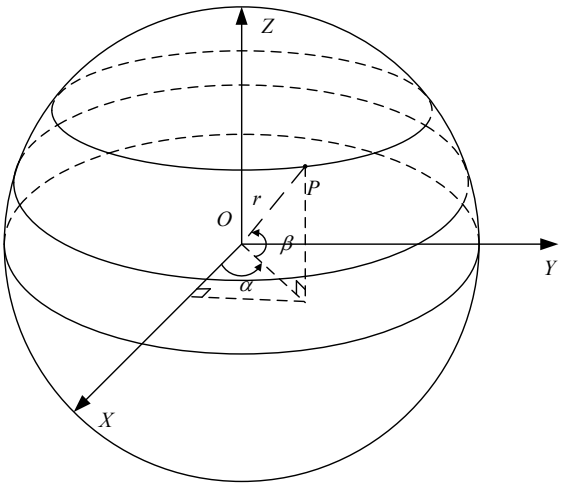


Figure 1. Spherical coordinate representation of spatial points

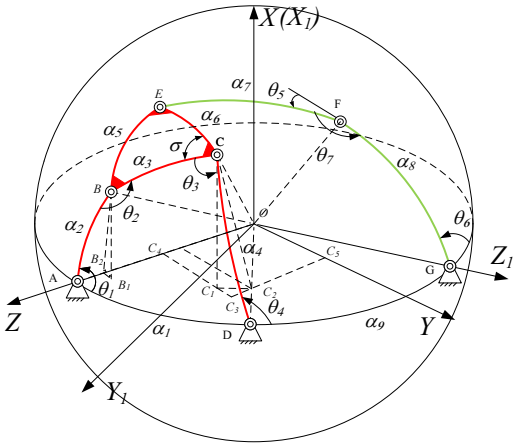


Figure 2. Schematic diagram of Stephenson-III spherical six bar mechanism

Figure 2 is the schematic diagram of Stephenson-III spherical six bar mechanism. In order to keep the generality of the research, it is assumed that the spherical six bar mechanism is located on the unit sphere. O is the center of the ball, A、D、G are the fixed hinge point, B、C、E、F are the movable hinge point. In the schematic diagram of Stephenson-III spherical six bar mechanism, AB is the input lever, CD and FG are output rods, BCE and EF are the connecting rod, AD and DG are the rack. The length of each rod is expressed by its ball center angle $\alpha_k\ (k=1,2,\cdots,9)$, Enter dihedral angle $\angle BAD=\theta_1$,Output dihedral angle 1 is $\angle CDG=\theta_4$, The output dihedral angle 2 is the complement of $\angle FGD$ and is recorded as θ_6. The spherical six bar mechanism is regarded as a basic spherical four-bar chain $ABCD$ and spherical two-bar group EFG, and the basic coordinate system $OXYZ$ and branch coordinate system $OX_1Y_1Z_1$ are established respectively, In the basic coordinate system $OXYZ$, the X axis is perpendicular to the plane composed of OA and OD,The Z axis coincides with the OA axis, and the Y axis is in the plane composed of OA and OD. In the branch coordinate system, the X_1 axis coincides with the X axis of the basic coordinate system, the Z_1 axis coincides with the OG axis, and the Y_1 axis is determined according to the right-hand rule [6].

Note that $\sin(\theta_i)=s\theta_i$, $\cos(\theta_i)=c\theta_i(i=1,2,\cdots 6)$, $\sin(\alpha_i)=s\alpha_i$, $\cos(\alpha_i)=c\alpha_i(i=1,2,\cdots 9)$ in the basic spherical four-bar chain, according to the motion relationship of spherical members and the condition of constant rod length, the constraint equation is

$$\overrightarrow{OB}\bullet\overrightarrow{OC}=c\alpha_3 \tag{2}$$

According to the constraint coupling between spherical four-bar mechanism and spherical two-bar group, the constraint equation between them can be obtained as

$$\begin{cases} \overrightarrow{OC} \bullet \overrightarrow{OE} = c\alpha_6 \\ \overrightarrow{OB} \bullet \overrightarrow{OE} = c\alpha_3 c\alpha_6 - s\alpha_3 s\alpha_6 c\sigma \end{cases} \tag{3}$$

The constraint equations of Stephenson-III spherical six bar mechanism can be obtained by combining equation (2) and equation group (3), as follows:

$$\begin{cases} \overrightarrow{OB} \bullet \overrightarrow{OC} = c\alpha_3 \\ \overrightarrow{OC} \bullet \overrightarrow{OE} = c\alpha_6 \\ \overrightarrow{OB} \bullet \overrightarrow{OE} = c\alpha_3 c\alpha_6 - s\alpha_3 s\alpha_6 c\sigma \end{cases} \tag{4}$$

3. Solution of constraint equations of spherical six bar mechanism

In the basic spherical four-bar chain, it is easy to know that the coordinates of fixed hinge point A and point D are $(0,0,1)^T$ and $[0, s\alpha_1, c\alpha_1]^T$ respectively. According to the geometric principle $BB_2 = OB \bullet s\alpha_2$, $\angle BB_2B_1 = \angle BAD = \theta_1$, $BB_1 = BB_2 \bullet s\theta_1$, $B_2B_1 = BB_2 \bullet c\theta_1$, At the same time, due to $\angle CC_2C_1 = \angle CDA = \pi - \theta_4$, $OB = OC = 1$, we can know that $C_1C_4 = OC_5 - C_1C_3 = c\alpha_4 s\alpha_1 + s\alpha_4 c\theta_4 c\alpha_1$, $CC_2 = s\alpha_4$, $CC_1 = s\alpha_4 s\theta_4$, $C_3C_5 = C_2C_5 + C_2C_3 = c\alpha_4 c\alpha_1 - s\alpha_4 c\theta_4 s\alpha_1$, To sum up, the coordinates of each hinge point in the basic spherical four-bar chain are

$$\begin{cases} \mathbf{A} = [0,0,1]^T \\ \mathbf{B} = [s\alpha_2 s\theta_1, s\alpha_2 c\theta_1, c\alpha_2]^T \\ \mathbf{C} = [s\alpha_4 s\theta_4, s\alpha_1 c\alpha_4 + c\alpha_1 s\alpha_4 c\theta_4, c\alpha_1 c\alpha_4 - s\alpha_1 s\alpha_4 c\theta_4]^T \\ \mathbf{D} = [0, s\alpha_1, c\alpha_1]^T \end{cases} \tag{5}$$

Referring to the literature[7], the fixed-point transformation of the spherical coordinate system can be obtained by rotating around three (fixed) coordinate axes, and the coordinate transformation matrices around the X axis and Z axis are

$$\mathbf{R}(X, \alpha_i) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c\alpha_i & -s\alpha_i \\ 0 & s\alpha_i & c\alpha_i \end{pmatrix}, \mathbf{R}(Z, \theta_i) = \begin{pmatrix} c\theta_i & -s\theta_i & 0 \\ s\theta_i & c\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$

Therefore, the coordinate transformation from the base coordinate system $OXYZ$ to the branch coordinate system $OX_1Y_1Z_1$ is the counterclockwise rotation angle $(\alpha_1 + \alpha_9)$ around the X axis, and its coordinate transformation matrix is

$$\mathbf{R}(X,(\alpha_1+\alpha_9))=\begin{pmatrix}1 & 0 & 0\\ 0 & c(\alpha_1+\alpha_9) & -s(\alpha_1+\alpha_9)\\ 0 & s(\alpha_1+\alpha_9) & c(\alpha_1+\alpha_9)\end{pmatrix} \tag{7}$$

Further, it can be obtained that the coordinate of the fixed hinge point G in the basic coordinate system $OXYZ$ is

$$\mathbf{G}=\mathbf{R}(X,(\alpha_1+\alpha_9))\bullet\mathbf{G}_1=\begin{pmatrix}1 & 0 & 0\\ 0 & c(\alpha_1+\alpha_9) & -s(\alpha_1+\alpha_9)\\ 0 & s(\alpha_1+\alpha_9) & c(\alpha_1+\alpha_9)\end{pmatrix}\bullet\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}=\begin{bmatrix}0\\ -s\alpha_1c\alpha_9-s\alpha_9c\alpha_1\\ c\alpha_1c\alpha_9-s\alpha_1s\alpha_9\end{bmatrix} \tag{8}$$

Therefore, based on the displacement rotation theory [8], it can be seen that the coordinates of movable hinge points F and E in the basic coordinate system $OXYZ$ can be expressed as

$$\begin{aligned}\mathbf{F}&=\mathbf{R}(Z,\theta_6)\bullet\mathbf{R}(X,\alpha_8)\bullet\mathbf{G}\\ \mathbf{E}&=\mathbf{R}(Z,\theta_5)\bullet\mathbf{R}(X,\alpha_7)\bullet\mathbf{F}\end{aligned} \tag{9}$$

It can be obtained by calculating in the mathematical software maple. The coordinates of hinge points E , F and G in the basic coordinate system $OXYZ$ are

$$\begin{cases}\mathbf{E}=\left[a_1c\theta_5s\theta_6+a_2s\theta_5c\theta_6-a_3s\theta_5,\ a_4s\theta_5s\theta_6-a_5c\theta_5c\theta_6+a_6c\theta_5,\ a_7c\theta_6-a_8\right]\\ \mathbf{F}=\left[a_9\bullet s\theta_6,\ a_{10}\bullet c\theta_6,\ a_{11}\right]^T\\ \mathbf{G}=\left[\ 0,-s\alpha_1c\alpha_9-s\alpha_9c\alpha_1,c\alpha_1c\alpha_9-s\alpha_1s\alpha_9\right]^T\end{cases} \tag{10}$$

In equation (10)

$$\begin{aligned}a_1&=c\alpha_8c\alpha_9s\alpha_1+c\alpha_8s\alpha_9c\alpha_1-s\alpha_8s\alpha_9s\alpha_1+s\alpha_8c\alpha_9s\alpha_1\\ a_2&=c\alpha_7c\alpha_8c\alpha_9s\alpha_1+c\alpha_7c\alpha_8s\alpha_9c\alpha_1-c\alpha_7s\alpha_8s\alpha_9s\alpha_1+c\alpha_7s\alpha_8c\alpha_9c\alpha_1\\ a_3&=s\alpha_7s\alpha_8c\alpha_9s\alpha_1-s\alpha_7s\alpha_8s\alpha_9c\alpha_1-s\alpha_7c\alpha_8s\alpha_9s\alpha_1+s\alpha_7c\alpha_8c\alpha_9c\alpha_1\\ a_4&=c\alpha_8c\alpha_9s\alpha_1+c\alpha_8s\alpha_9c\alpha_1-s\alpha_8s\alpha_9s\alpha_1+s\alpha_8c\alpha_9c\alpha_1\\ a_5&=c\alpha_7c\alpha_8c\alpha_9s\alpha_1-c\alpha_7c\alpha_8s\alpha_9c\alpha_1+c\alpha_7s\alpha_8s\alpha_9s\alpha_1-c\alpha_7s\alpha_8s\alpha_9c\alpha_1\\ a_6&=s\alpha_7s\alpha_8c\alpha_9s\alpha_1+s\alpha_7s\alpha_8s\alpha_9c\alpha_1+s\alpha_7c\alpha_8s\alpha_9s\alpha_1\end{aligned} \tag{11}$$

$$\begin{aligned}a_7&=s\alpha_7s\alpha_8s\alpha_9s\alpha_1-s\alpha_7c\alpha_8c\alpha_9s\alpha_1-s\alpha_7c\alpha_8s\alpha_9c\alpha_1-s\alpha_7s\alpha_8c\alpha_9c\alpha_1\\ a_8&=c\alpha_7s\alpha_8c\alpha_9s\alpha_1-c\alpha_7s\alpha_8s\alpha_9c\alpha_1-c\alpha_7c\alpha_8s\alpha_9s\alpha_1+c\alpha_7c\alpha_8c\alpha_9c\alpha_1\\ a_9&=c\alpha_8c\alpha_9s\alpha_1+c\alpha_8s\alpha_9s\alpha_1-s\alpha_8s\alpha_9s\alpha_1+s\alpha_8c\alpha_9s\alpha_1\\ a_{10}&=s\alpha_8s\alpha_9s\alpha_1-s\alpha_8c\alpha_9s\alpha_1-c\alpha_8s\alpha_9s\alpha_1-c\alpha_8c\alpha_9s\alpha_1\\ a_{11}&=c\alpha_8c\alpha_9c\alpha_1-c\alpha_8s\alpha_9s\alpha_1-s\alpha_8s\alpha_9c\alpha_1-s\alpha_8c\alpha_9s\alpha_1\end{aligned} \tag{12}$$

By substituting equation (5) and equation (10) into equation (4), the constraint equations of spherical six bar mechanism can be obtained as

$$\begin{cases} f_1c_4 + f_2s_4 + f_3 = 0 \\ g_1s_5s_6 + g_2c_5c_6 + g_3s_5c_6 + g_4c_5s_6 + g_5c_6 + g_6c_5 + g_7s_5 + g_8 = 0 \\ k_1s_5s_6 + k_2c_5c_6 + k_3s_5c_6 + k_4c_5s_6 - k_5c_5 + k_6s_5 + k_7c_6 + k_8 = 0 \end{cases} \tag{13}$$

In equation (13) $f_i\left(i=1,2,3\right), g_i, k_i\left(i=1,2,\cdots,8\right)$ are constant containing mechanism structure parameters and input parameters.

Make $\tan\left(\theta_4/2\right)=x_4, \tan\left(\theta_5/2\right)=x_5, \tan\left(\theta_6/2\right)=x_6$, we can get $s_4=2x_4/\left(1+x_4^{\,2}\right)$, $c_4=\left(1-x_4^{\,2}\right)/\left(1+x_4^{\,2}\right)$, $s_5=2x_5/\left(1+x_5^{\,2}\right)$, $c_5=\left(1-x_5^{\,2}\right)/\left(1+x_5^{\,2}\right)$, $s_6=2x_6/\left(1+x_6^{\,2}\right), c_6=\left(1-x_6^{\,2}\right)/\left(1+x_6^{\,2}\right)$. Substituting $s_4, c_4, s_5, c_5, s_6, c_6$ into equation (13) and simplify it, the input-output equations (14) of spherical six bar mechanism can be obtained as

$$\begin{cases} a_{11}x_4^{\,2} + a_{12}x_4 + a_{13} = 0 \\ a_{21}x_6^{\,2} + a_{22}x_6 + a_{23} = 0 \\ a_{31}x_6^{\,2} + a_{32}x_6 + a_{33} = 0 \end{cases} \tag{14}$$

In equation (14)

$$\begin{aligned} &a_{11}=g_3-g_1 \quad a_{21}=x_5^{\,2}(k_2+k_5-k_7+k_8)+x_5(2k_6-2k_3)+(k_8-k_2-k_5-k_7) \\ &a_{12}=2g_2 \quad\;\; ,a_{22}=-2k_4x_5^{\,2}+4k_1x_5+2k_4 \\ &a_{13}=g_1+g_3 \quad a_{23}=x_5^{\,2}(k_5+k_7+k_8-k_2)+x_5(2k_3+2k_6)+(k_2-k_5+k_7-k_8) \end{aligned} \tag{15}$$

$$\begin{aligned} &a_{31}=x_5^{\,2}(f_2-f_5-f_6+f_8)+x_5(2f_7-2f_3)+(f_8+f_6-f_5-f_2) \\ &a_{32}=-2f_4x_5^{\,2}+4f_1x_5+2f_4 \\ &a_{33}=x_5^{\,2}(f_5+f_8-f_2-f_6)+x_5(2f_3+2f_7)+(f_2+f_5+f_7+f_8) \end{aligned} \tag{16}$$

Stephenson-III spherical six bar mechanism has two output angles θ_4 and θ_6. x_4 can be obtained from the first equation of equation group (14) as

$$x_4=\frac{-a_{12}\pm\sqrt{a_{12}^{\,2}-4a_{11}a_{13}}}{2a_{11}} \tag{17}$$

And since $\tan\left(\theta_4/2\right)=x_4$,we can get

$$\theta_4=\arctan\left(\frac{-a_{12}\pm\sqrt{a_{12}^{\ 2}-4a_{11}a_{13}}}{2a_{11}}\right) \tag{18}$$

The sign in equation (18) corresponds to two circuits (assembly configuration) of the basic spherical four-bar chain. The second and third equations of equation group (14) can be solved by Sylvester's resultant elimination method as

$$\begin{aligned}&a_{21}^{\ 2}a_{33}^{\ 2}-a_{21}a_{22}a_{32}a_{33}-2a_{21}a_{23}a_{31}a_{33}\\&+a_{21}a_{23}a_{32}^{\ 2}+a_{22}^{\ 2}a_{31}a_{33}-a_{22}a_{23}a_{31}a_{32}\\&+a_{23}^{\ 2}a_{31}^{\ 2}=0\end{aligned} \tag{19}$$

By substituting $a_{21},a_{22},a_{23},a_{31},a_{32},a_{33}$ into equation (19) and simplifying it, we can get

$$d_1x_5^{\ 8}+d_2x_5^{\ 7}+d_3x_5^{\ 6}+d_4x_5^{\ 5}+d_5x_5^{\ 4}+d_6x_5^{\ 3}+d_7x_5^{\ 2}+d_8x_5+d_9=0 \tag{20}$$

In equation (19), $d_i\left(i=1,2,\cdots,9\right)$ is only a function of x_4 . therefore, by substituting the structural parameters of the mechanism into equations (17) and (20), 8 solutions of the positive position solution x_5 of the mechanism can be obtained. Because of $\tan\left(\theta_5/2\right)=x_5$, 8 solutions of the joint angle θ_5 can be obtained accordingly. It can be seen that there are at most 8 circuits (assembly configuration) of Stephenson-III spherical six bar mechanism. At the same time, it can also be obtained that the output angle θ_6 is

$$\theta_6=2\arctan\left(\frac{a_{21}a_{33}-a_{31}a_{23}}{a_{31}a_{22}-a_{21}a_{32}}\right) \tag{21}$$

4. Conclusion

Referring to the division method of Stephenson-III planar six bar mechanism, Stephenson-III spherical six bar mechanism is regarded as composed of basic spherical four-bar chain and spherical two-bar group. On this basis, the position analysis of Stephenson III spherical six bar mechanism is carried out, its input-output equations are obtained, and the correctness of the equations is verified by loop analysis, It not only lays a solid foundation for the kinematic analysis of this type of mechanism, but also contributes to the research of its dynamic analysis, and provides a theoretical basis for the better application of Stephenson III spherical six bar mechanism.

References

[1] Ying Z, Qineng H, Shimin W, Qizheng L. Based on times 1P5R series manipulator inverse kinematics analysis of the matrix. Journal of Beijing university of posts and telecommunications. 2021; pp. 1-61.
[2] Yanqiang C, Haibo L, Yuru L. Application status of mathematical methods for position analysis of mechanism in China. Mechanical research and application. 2014; 27(05): pp. 190-194.
[3] Chenglei L, Jianjun Z, Jianye N. Kinematic Performance of 4-DOF Generalized Spherical Parallel Mechanism for Ankle Rehabilitation.. Journal of mechanical engineering. 2021; 57(21): pp. 45-54.
[4] Hernandez S, Bai S, Angeles J. The Design of a Chain of Spherical Stephenson Mechanisms for a Gearless Robotic Pitch-Roll Wrist. ASME. J. Mech. Des. March 2006; 128(2): pp. 422-429.
[5] Tong Y, Jianyou H, Lairong Y. Function synthesis of four point spherical 4R mechanism based on solution domain. Journal of agricultural machinery. 2012; 43(10): pp. 200-206.
[6] Suixian Y, Yanfang L, Yan S. Position analysis method of spherical mechanism with four-bar closed chain. Journal of agricultural machinery. 2013; 44(08): pp. 262-267.
[7] Jianyou H, Tong Y. Higher Institutions. Beijing: China Machine Press. May 2015.
[8] Qixian Z. Analysis and synthesis of spatial mechanisms: Volume I. Beijing: Machinery Industry Press; 1984.

doi:10.3233/ATDE220212

Design and Research of Restaurant Intelligent Cleaning Robot

Li SONG [a,1], Tao ZHANG [a], Fu CHEN [a], Shiwei ZHAI [a] and Chongshu SUN [b]
[a] *Yantai Institute of Science and Technology, No. 34 Xianjing West Road, Penglai District, Yantai City, Shandong Province, 265600, China*
[b] *Penglai Jutal Offshore Engineering Heavy Industries Co., Ltd. No. 5 Harbin Road, Penglai District, Yantai City, Shandong Province, 265600, China*

Abstract. The restaurant intelligent cleaning robot, belongs to the field of robot technology, including driving the base, control the body and vision system. Mechanical arms are installed on both sides of the top of the machine body. One of the bottom end of the robot arm is installed with clamping claws, and the other mechanical arm bottom is installed with Clean hands for desktop cleaning. Driving wheels are symmetrically arranged on both sides of the bottom surface of the base, the body internal top installed vacuuming assembly, vacuuming components of the vacuum port through the pipe to connect the vacuum cover. Combined with lidar and other sensors, it can achieve automatic identification, automatic cleaning, multi-use, flexible movement. It also can reduce the probability of collision tables and chairs, the use of higher reliability, more convenient.

Keywords. Robot, Cleaning, Automatic recognition

1. Introduction

With the rapid development of science and technology, people pay more and more attention to the working environment and work efficiency. We note that some places such as canteens and restaurants often need to hire cleaning aunts to do repetitive and mechanical work such as sweeping the floor and cleaning the table to maintain hygiene. This not only wastes time and energy, but also needs to pay a lot of money. For the catering industry, what attracts people's attention most is the service robot. The appearance of robots is very attractive. They can not only replace waiters, but also win the smiles of customers and bring laughter. More importantly, nowadays, the labor cost is higher and higher, and the use of robots can save costs.

However, the robot function used in the existing restaurant is too single, which can only complete the simple meal delivery work, can not well meet the needs of the restaurant, and has poor mobility between the narrow dining tables, which is easy to bump into tables and chairs..

The purpose of restaurant intelligent cleaning robot is to provide a convenient restaurant intelligent cleaning robot. It can combine the functions of dust collection, floor mopping and table cleaning, realize automatic identification, automatic cleaning,

[1] Li SONH , Yantai Institute of Science and Technology, No. 34 Xianjing West Road, Penglai District, Yantai City, Shandong Province, 265600, China; E-mail: songlisun@163.com.

multi-purpose one machine, flexible action at the same time, reduce the probability of collision with tables and chairs, and have higher reliability and convenience.

By studying the mechanical model of restaurant intelligent cleaning robot, this paper establishes the mechanical model, and realizes the automation of restaurant cleaning through the combination of a variety of sensors to meet the needs of robot cleaning in complex environment.

2. Overall structure design of robot

2.1. Overall design parameters

The main body is the basic part of the restaurant intelligent cleaning robot, the installation basis of other assembly components, and one of the main components in motion. The main conditions considered in the design are: (1) the strength and stiffness design of the main body must meet the requirements of the robot itself; (2) Reduce the weight of the main body as much as possible to improve the effective bearing weight; Under the condition of ensuring sufficient rigidity of the main body; (3) Reduce the center of gravity of the main body as much as possible, so as to improve the anti overturning ability of the intelligent cleaning robot of the whole restaurant and ensure the better and stable operation of the intelligent cleaning robot of the restaurant.

In terms of shape, we use a taller and thinner body, combined with the omni-directional movement of mcnamu wheel, so that the robot can shuttle between the dining tables more flexibly.

The fuselage is made of ABS material, which has high strength, good toughness, insulation and corrosion resistance. It is the best choice for the fuselage shell.

The robot design parameters is shown in Table 1.

Table 1. Robot design parameters

Projects	Parameter	Remarks
Overall dimensions (mm)	480×480×1350	
Steering mode	Omnidirectional steering	Can realize in situ steering
Number of steering wheels	2	Horizontal steering wheel
Steering wheel power (Kw)	4	
Number of angle wheels	2	
Load mass (kg)	3000	
Readiness quality (kg)	1500	Including battery
Full load maximum speed (m/s)	0.5	
Control mode	Manual remote control, automatic	
Protection grades	IP54	
Suspension mode	Drum damper	
Battery	Lithium iron phosphate	
Voltage (v)	48	

2.2. Robot work planning

When the robot recognizes that there is food residue and other garbage on a table, it will automatically carry out path planning. Go to the table and clean it with the manipulator with a rag on the right. At the same time, open the cover plate of the robot's dustbin and collect the garbage into the dustbin brought by the robot. If there are cups and other tableware on the table, it can also be clamped by the manipulator on the left manipulator.

When the robot needs to clean the ground, it will start the vacuum cleaner to suck in the dust and garbage. At the same time, the circular mop starts to rotate. A drip device is installed in front of the mop. When dragging the ground, water droplets and detergent will drop to assist in cleaning. After cleaning, the robot puts down the rear tail scraper, which is composed of two layers. The first layer in the front row is similar to sponge material, which can suck residual detergent, etc., and the second layer is rubber material, which can wipe water stains dry.

2.3. Overall structure design

Among them, four groups of driving wheels are evenly arranged on the bottom surface of the driving base, and the driving wheel is mcnamu wheel, which can realize flexible omni-directional movement in a narrow area, make the device move more flexibly, reduce the probability of collision with tables and chairs, and have higher use reliability.

Figure 1. Restaurant intelligent cleaning robot

According to the cleaning requirements of modern restaurants, the overall structure layout is analyzed to meet the cleaning requirements of various restaurants. At the same time, the mechanical model of automatic restaurant intelligent cleaning robot is established. On this basis, the motion control model and software motion simulation are constructed.

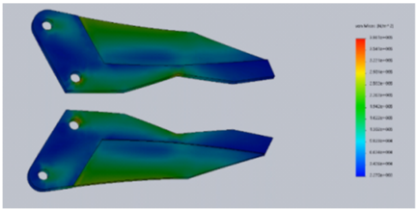


Figure 2. Finite element analysis of gripper

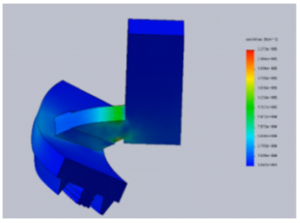


Figure 3. Finite element analysis of scraper hand

The overall structure of the restaurant intelligent cleaning robot is shown in Figure 4:

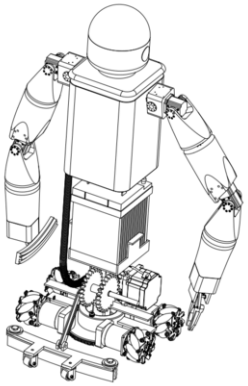


Figure 4. The main frame of restaurant intelligent cleaning robot

The restaurant intelligent cleaning robot comprises a driving base, a control body and a recognition head body arranged on the top of the control body. A mechanical arm is arranged on both sides of the top of the control body, in which the bottom end of one mechanical arm is movably installed with a clamping claw, and the bottom end of the other mechanical arm is movably installed with a cleaning hand for desktop cleaning. Drive wheels are symmetrically arranged on both sides of the bottom surface of the drive base, the top inside the control body is fixedly installed on the dust suction assembly, the dust suction port of the dust suction assembly is connected with a dust suction cover through a pipe, and the dust suction cover is fixedly embedded on one side of the bottom surface of the drive base. A battery assembly is arranged in the inner center of the control body, a cleaning motor is installed in the inner center of the driving base, a cleaning disc is fixedly installed at the power output end of the cleaning motor, the cleaning disc is movably embedded in the center of the bottom surface of the driving base, and a mop mechanism for ground mopping is arranged in the driving base. The manipulator is a five axis manipulator.

Four groups of driving wheels are evenly arranged on the bottom surface of the driving base, and the driving wheels are mcnamu wheels.

The cleaning hand includes a connecting bracket at the bottom end of one of the mechanical arms. A rag cotton body is installed in the center of the bottom surface of the connecting support, a wiper plate is installed on the inner side of the bottom surface of the connecting support, and a cleaning plate is installed on the outer side of the bottom surface of the connecting support.

The mop mechanism includes a mop motor driving the interior of the base. The power output end of the drag washing motor is provided with a driving gear, and one side inside the driving base is provided with a driven wheel group meshing with the driving gear. The driven wheel set is rotationally connected with a movable connecting rod on the side away from the drive gear, which extends to the outside of the drive base and is slidably connected with the drive base. The bottom end of the movable connecting rod is rotatably connected with a mop support, and a mop cotton body is installed in the center of the bottom surface of the mop support. The center of the bottom surface of the mop support is equipped with a scraper plate on the side away from the drive base.

Both sides of the top surface of the towing support are symmetrically connected with side guide wheels, and the towing support is symmetrically connected with the limited position support wheel through the support on the side away from the drive

base. The driving base is provided with a liquid spraying device on the side close to the movable connecting rod, which is used to spray detergent.

3. Design of traveling mechanism

As shown in Figure 5, the robot uses mcnamu wheel to replace the traditional universal wheel. The four mcnamu wheels are independently controlled by four motors. It is an omnidirectional wheel with special structure. The shape of mcnamu wheel is similar to a helical gear, and the teeth are drum shaped rollers that can rotate. The roller axis is formed with the wheel axis α Degrees. The roller has three degrees of freedom. When the roller rotates around itself, it can also rotate around the axle and around the point where the roller contacts the ground. In this way, the whole wheel body has three degrees of freedom: 1. Rotation around the axle. 2. Translation along the vertical direction of the roller axis 3. Rotation of the contact point between the roller and the ground. In this way, the driving wheel has active driving ability in one direction and can move freely in the other direction. Many small rollers are distributed around the circumference of the wheel, and the contour of the roller coincides with the theoretical circumference of the wheel, and the roller can rotate freely. When the motor drives the wheel to rotate, the whole wheel moves in the direction perpendicular to the drive shaft in a normal way. At this time, the rollers around the wheel will rotate freely along their respective axes. The wheat wheel can be divided into left rotation and right rotation according to the deflection of the roller. Compared with universal wheel, wheat wheel has the advantages of flexibility, accuracy and high efficiency. It is a controllable universal wheel. The tire arrangement of this structure adopts Abba form, and the tail scraper driven wheel is a universal wheel.

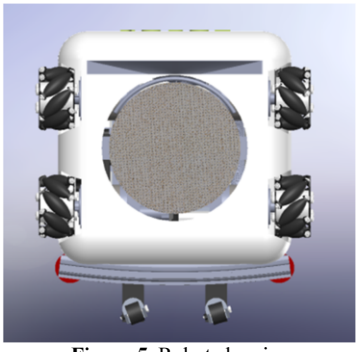


Figure 5. Robot chassis

4. The kinematic analysis

4.1. Calculation of axial velocity of wheat wheel

As shown in figure 7 and figure 8, the movement speed of wheel axis is:

$$\vec{\mathrm{V}} = \vec{\mathrm{V}}_t + \vec{w} \times \vec{r} \tag{1}$$

Calculate the components of X and Y axes respectively:

$$\begin{cases} v_x = v_{t_x} - w \cdot r_y \\ v_y = v_{t_y} - w \cdot r_x \end{cases} \tag{2}$$

Similarly, the axial speeds of the other three wheels can be calculated.

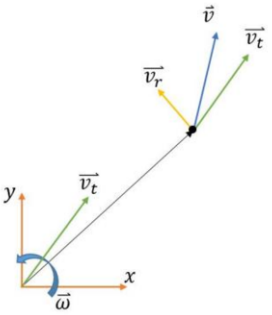


Figure 6. Diagonal analysis of wheat wheel

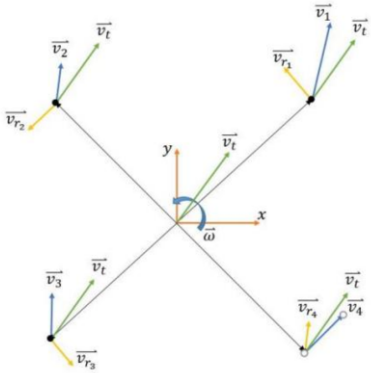


Figure 7. Overall analysis of wheat wheel

4.2. Calculation of axial velocity of wheat wheel

Through the wheel axis speed, the speed perpendicular to the roller and the speed along the roller direction can be decomposed.

Since the speed perpendicular to the roller direction will be converted into the rotation of the roller itself, it can be ignored, as is shown in figure 9.

Thus:

$$\vec{v}_{\Pi} = \vec{v} \cdot \hat{u} = \left(v_x \hat{l} + v_y \hat{j}\right) \cdot \left(-\frac{1}{\sqrt{2}} \hat{l} + \frac{1}{\sqrt{2}} \hat{j}\right) = -\frac{1}{\sqrt{2}} v_x + \frac{1}{\sqrt{2}} v_y \tag{3}$$

Calculation of wheel speed:

$$\vec{v}_w = \frac{v_{\Pi}}{\cos 45^{\circ}} = \sqrt{2}\left(-\frac{1}{\sqrt{2}} v_x + \frac{1}{\sqrt{2}} v_y\right) = -v_x + v_y \tag{4}$$

Defined by a and B as shown in the figure below, we can get:

$$\begin{cases} v_x = v_{t_x} - w \cdot b \\ v_y = v_{t_y} - w \cdot a \end{cases} \tag{5}$$

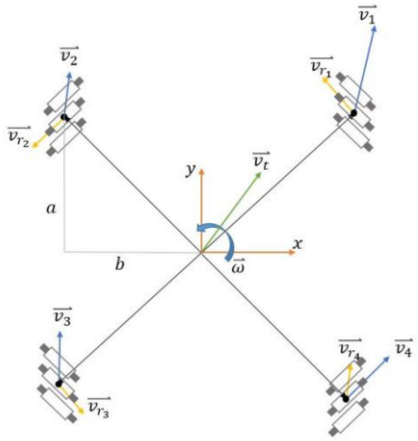


Figure 8. Analysis of moving speed of wheat wheel

To sum up, the rotational speed of the four wheels can be calculated as follows:

$$\begin{cases} v_{w_1} = v_{t_y} - v_{t_x} + w(a+b) \\ v_{w_2} = v_{t_y} + v_{t_x} - w(a+b) \\ v_{w_3} = v_{t_y} - v_{t_x} - w(a+b) \\ v_{w_4} = v_{t_y} + v_{t_x} + w(a+b) \end{cases} \tag{6}$$

Similarly, the equations of the forward kinematics model can be solved by reverse operation:

$$\begin{cases} v_{t_x} = \dfrac{1}{4}\left(-v_{w_1} + v_{w_2} - v_{w_3} + v_{w_4}\right) \\ v_{t_y} = \dfrac{1}{4}\left(v_{w_1} + v_{w_2} + v_{w_3} + v_{w_4}\right) \\ w = \dfrac{1}{4(a+b)}\left(v_{w_1} - v_{w_2} - v_{w_3} + v_{w_4}\right) \end{cases} \tag{7}$$

Write the inverse kinematics model as a matrix:

$$\begin{bmatrix} v_{w_1} \\ v_{w_2} \\ v_{w_3} \\ v_{w_4} \end{bmatrix} = \begin{bmatrix} v_{t_y} \\ v_{t_y} \\ v_{t_y} \\ v_{t_y} \end{bmatrix} + \begin{bmatrix} -v_{t_x} \\ +v_{t_x} \\ -v_{t_x} \\ +v_{t_x} \end{bmatrix} + \begin{bmatrix} +w(a+b) \\ -w(a+b) \\ -w(a+b) \\ +w(a+b) \end{bmatrix} \tag{8}$$

It can be concluded that the rotational speeds of the four wheels required for the current movement can be obtained by simply adding the four wheel speeds of the wheat wheel chassis in X translation, y translation and rotation. The reason is that the omnidirectional moving chassis composed of wheat wheels is a pure linear system.

5. Motion control

For 4-way motor driving wheat wheel, closed-loop control is adopted to ensure consistent rotation speed of wheat wheel, as is shown in figure 10.

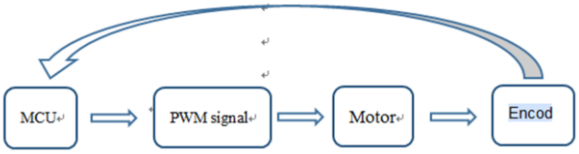


Figure 9. Closed-loop control

When the sensor recognizes the target, the single chip microcomputer carries out image processing and recognition, then outputs pulse signals to the bottom motor to control the robot to approach the target, and then outputs the angle signal to control the steering gear to control the movement of the 5-axis manipulator, as is shown in figure 11.

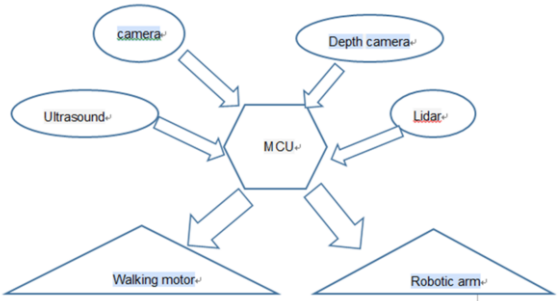


Figure 10. Control flow

6. Cleaning mechanism

In this design, the clamping manipulator, cleaning manipulator, vacuum cleaner, cleaning plate and other cleaning mechanisms are combined to realize various scene cleaning, such as ground cleaning + desktop cleaning. The hand design of the cleaning manipulator adopts a three-layer structure. The first layer is hard to scrape off large pieces of garbage such as bones, the middle layer is a rag to wipe away stains, and the last layer of rubber is used to scrape off residual oil stains. The tail scraper is composed of two layers. The first layer in the front row is similar to sponge material, which can suck residual detergent, etc., and the second layer is rubber material, which can wipe water stains dry. The mechanism controlling the lifting of tail scraper is a group of motor + reduction gear (reduction ratio 1:5) + connecting rod. The motor outputs power to the gear, and the connecting rod is fixed on the inner side of the gear root circle. When the gear rotates, it can drive the connecting rod to make a circular motion around the shaft. At the limit hole at the robot shell, the connecting rod is equivalent to a lever, and the limit hole is the fulcrum. At this time, when the motor rotates, the lifting control of the tail scraper can be realized.

In terms of dust collection, the vacuum cleaner motor is installed above the dustbin, and the robot shell is provided with an air outlet. When the moving impeller rotates at a high speed of 20000-30000 rpm, a negative pressure will be generated at the air inlet to form a vacuum, and then form suction.

Figure 11. Vacuum cleaner motor

7. Machine vision

If the robot wants to have the ability of path planning, lidar is indispensable. Lidar is like the sonar system on bats. It is a sensor that can detect the precise position of objects. It mainly emits laser signals to the target, and then calculates the distance according to the time difference of signals reflected from the object, Then the angle of the laser is emitted to determine the angle of the object and the transmitter, so as to obtain the relative position of the object and the transmitter. The lattice data in two-dimensional space is obtained through continuous scanning, which can help the machine with slam technology

People realize the functions of autonomous positioning, map construction and path planning.

Figure 12. Lidar

Principle of triangular ranging:

The principle of triangulation method is shown in the figure below. The laser first emits the laser, and the laser will be reflected after irradiating the object. The reflected light is received by the linear CCD. Since the laser is separated from the detector for a certain distance, objects with different distances will be imaged at different positions on the CCD according to the optical path. According to the trigonometric formula, the distance of the measured object can be deduced.

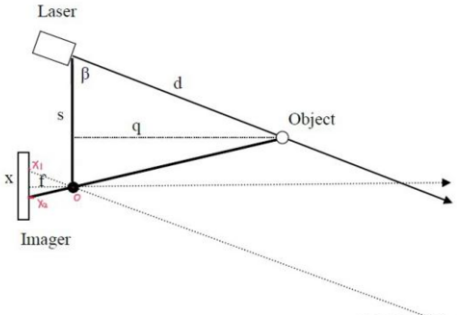


Figure 13. Triangular ranging method

Depth camera, also known as 3D camera, can detect the depth of field distance of shooting space through this camera, which is also the biggest difference from ordinary camera. Through the data obtained by the depth camera, we can accurately know the distance from each point in the image to the camera. In this way, combined with the (x, y) coordinates of the point in the 2D image, we can obtain the three-dimensional spatial coordinates of each point in the image. The real scene can be restored through three-dimensional coordinates to realize applications such as scene modeling.

The ultrasonic module is selected for obstacle avoidance because the ultrasonic module obtains the distance information (s = 340 * t / 2) by transmitting ultrasonic and calculating time. S is the distance and t is the time from sending to receiving, so the light will not affect its obstacle avoidance effect.

Combined with OpenCV, the camera first imports the color pictures taken by our camera into the C++ development environment, and converts the original image into gray image (using the cvtcolor function in openCV) to realize garbage recognition.

Slam, combined with sensors such as lidar camera, models the environment, and optimizes filtering in the later stage to realize mapping, positioning and path planning.

Figure 14. Triangular ranging method

8. Conclusion

The purpose of this paper is to study the robot technology based on restaurant intelligent cleaning, carry out the structure design of the robot body, omni-directional movement analysis and motion simulation system simulation, and through theoretical analysis, the results show that the restaurant intelligent cleaning robot with four servo motors controlling four mcnamu wheels separately has better omni-directional movement performance, It has good vertical and horizontal straight travel and in-situ (zero radius) steering functions, which can meet the flexible omnidirectional movement of the restaurant in a narrow area, make the device move more flexibly, reduce the probability of collision with tables and chairs, have higher reliability, and can better meet the needs of the restaurant. At the same time, information acquisition, image processing, environment modeling, algorithm optimization and fusion are carried out on the premise of the combination of lidar, depth camera, ultrasonic module, camera and other sensors. The results show that the robot with multiple sensors for environment perception has excellent path planning and autonomous obstacle avoidance ability!

References

[1] Boyan L. Intelligent restaurant service robot based on ARM[J].Science and Technology & Innovation,2019(22): pp. 20-23. p. 28.
[2] Hao W, Lu F, Kui Z, Xinyuan Z. Research and Application of Path Planning of Robot Vacuum Based[J]China Academic Journal Electronic Publishing House, 2021(04): pp. 104-106.
[3] Xiaopeng L and Xin W and Bin F. Modeling and control of a novel facade cleaning robot with four-ducted fan drive[J]. International Journal of Advanced Robotic Systems, 2021, 18(3) : pp. 124-126.
[4] Madridano Á et al. Trajectory planning for multi-robot systems: Methods and applications[J]. Expert Systems With Applications, 2021, p. 173.
[5] Yizhi L and Habibnezhad M and Jebelli H Brainwave-driven human-robot collaboration in construction[J]. Automation in Construction, 2021, p. 124.
[6] Zhijun H. Simulation of football based on PID controller and BP neural network[J]. Microprocessors and Microsystems, 2021, p. 81.
[7] Pei C, Based on stm32 intelligent restaurant service system robot design[J]. INFORMATION & COMMUNICATIONS, 2017, p. 9.
[8] Dongdong H, Li C. Design and Implementation of Full-intelligent Trackless Restaurant Service Robot Based on EXP-LM3S811[J]. Computer & Digital Engineering, 2020, p. 1.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 73
J. Xu et al. (Eds.)

doi:10.3233/ATDE220213

Local Wave Shape Control Technology in Tandem Cold Rolling

Yaxing LIU[a,d], Rongsheng SUN[a, e] Qing GU[a], Wei WANG[a,] Zhenhua BAI[a,b,1] and Yanyan ZHANG[a,c]
[a]*National Engineering Research Center for Equipment and Technology of Cold Strip Rolling, Qinhuangdao, Hebei, China*
[b]*State Key Laboratory of Metastable Materials Science and Technology, Yanshan University, Qinhuangdao, Hebei, China*
[c]*Tangshan yangbang iron and Steel Technology Research Institute Co., Ltd, Tangshan, Hebei, China*
[d]*Engineering Training Center, Yanshan University, Qinhuangdao, Hebei, China*
[e]*Ansteel Iron & Steel Research Institutes, Anshan, Liaoning, China*

Abstract. In view of the local wave shape problem of the rolling of cold continuous rolling mill, uneven distribution of emulsion along the width of strip steel, or uneven heating of the roll caused by the blockage of nozzle, the influence mechanism of the horizontal distribution of emulsion on the temperature field and hot roll shape of the working roll is analyzed. The influence mechanism of emulsion transverse distribution on strip shape is also analyzed. The strip shape distribution before and after emulsion adjustment is simulated. On this basis, the automatic prediction model of emulsion point cooling is established to ensure that the strip shape rolled by tandem cold mill does not have obvious shape defects and that the emulsion cooling capacity of each nozzle is uniform. This work provides technical guidance for the control of local wave shape on site. The local wave shape of finished strip can be significantly reduced after applied to the site.

Keywords. cold rolling, local wave shape, emulsion, point cooling

1. Introduction

To further improve and control the shape of a plate and meet the user's requirements on the quality and accuracy of this shape, many scholars have conducted in-depth research on the control of plate shape[1-4]. Bending rolls, tilting rolls, and other adjustment means achieve the purpose of controlling the plate shape by continuously changing the shape of the load roll gap[5,6]. However, the control effect on the local wave shape is poor. At present, the roll point cooling is an important process to control the local wave shape of the strip[7,8].

Bohacek Jan[9] proposed a novel method of cooling rolls in hot rolling. This method uses a combination of solid jet nozzles and specially shaped deflecting vanes. Shao J[10] put forward a coupled heat transfer model of work rolls. An aluminum strip is established to obtain the multi-parameter coupled subsection cooling regulation

[1] Corresponding Author, Zhenhua BAI, National Engineering Research Center for Equipment and Technology of Cold Strip Rolling, Qinhuangdao, Hebei, China; E-mail: bai_zhenhua@aliyun.com

characteristics of work rolls in aluminum cold rolling. Daron[11] optimized the nozzle arrangement, thereby significantly reducing fuel consumption and securing efficient fuel injection. Wang P[12] suggested that complex flatness deviation will emerge during rolling of thin cold-rolled strip. To improve the flatness of cold rolled strips, shortage of the old selective roll cooling control system and its effect on flatness is analyzed. Zhao Xing-tao[13], Liang Xun-guo[14], and others started with the deviation between the target setting value of the plate shape and the actual value on site. They used it as a control quantity and combined on-site control experience to improve the plate shape by fuzzy optimization of the emulsion nozzle parameters. However, the relevant literature is more about improving the shape of the roll through the hot roll shape and heat exchange control model. The modeling is slow, the solution time is long, and there is no direct cooling control of the emulsion point of the local shape. Therefore, the field research now focuses on how to establish the relationship between the point cooling of the emulsion and plate shape distribution and realize automatic control. This article is launched under this background.

2. Analysis of local wave shape control mechanism of tandem cold rolling mill

In cold tandem rolling, the strip needs to be continuously rolled through multiple stands. The shape of the loaded roll gap is the shape of the exit section of the strip, and the loaded roll gap is given in Formula (1). As the rolling speed fluctuates and the rolling process parameters such as tension and rolling force change, the friction heat and deformation heat generated between the roll and the strip are serious. They affect the thermal crown of the roll and further affect the shape of the loaded roll gap. When the emulsion is unevenly distributed along the width of the strip or the nozzle is blocked, the rollers are heated unevenly, causing uneven distribution of the roll gap along the transverse direction. This results in a partial wave shape running through the longitudinal direction, as shown in Figure 1.

$$\begin{cases} h_i = h_{n+1} - f_{wli}^{s} - f_{wli}^{x} - 2K'[\mathrm{q}_{z(n+1)} - \mathrm{q}_{zi}] + \dfrac{\Delta D_{wi}^{s} + \Delta D_{wi}^{x} + \Delta D_{wj}}{2}, 1 \le i \le n \\ h_i = h_{n+1} - f_{wri}^{s} - f_{wri}^{x} - 2K'[\mathrm{q}_{z(n+1)} - \mathrm{q}_{zi}] + \dfrac{\Delta D_{wi}^{s} + \Delta D_{wi}^{x} + \Delta D_{wj}}{2}, \quad n+2 \le i \le 2n+1 \end{cases} \tag{1}$$

where K' is flattening coefficient between work roll and rolled piece; f_{wri}^{s}, f_{wri}^{x}, f_{wli}^{s}, f_{wli}^{x} is respectively the value of the upper right, lower right, upper left, and lower left of the work roll deflection; q_{zj} is the rolling pressure at section j of work roll; ΔD_{wi}^{s}, ΔD_{wi}^{x} is the hot roll type of upper and lower work rolls, respectively; and ΔD_{wj} is convexity value caused by local cooling.

Figure 1. Emulsion cooling causes local crown changes of rolls

The unstable local heat exchange of the work rolls during the cold tandem rolling results in plate shape defects, ultimately affecting the quality of the strip. The existence of the point cooling system is particularly important. The rolling mill cooling system is aimed at the thermal deformation of the roll caused by the relative movement between the roll and the steel strip. It cools the surface of the roll and the steel strip through the flow of emulsion. As the roll is too long, to ensure the cooling effect so that each position of the roll and the strip can be fully cooled, the emulsion needs to be cooled in sections along the length of the roll, that is, point cooling. The emergence of the point cooling system of the rolling mill has alleviated the influence of local heat on the plate shape.

The existence of the cooling system of the rolling mill point has greatly improved the quality of the strip shape, and the strip is in the rolling process. Upon learning that there is always a deviation in the quality of the plate shape at a fixed position, that the point cooling system at the relevant position has a high probability of failure and blocking can be preliminarily analyzed.

3. Development of local wave shape control model for tandem cold rolling mill

3.1. Influence mechanism of emulsion lateral distribution on the temperature field and hot roll shape of work roll

The working principle of the point cooling system of the rolling mill is mainly to divide the part in contact between the roll and the steel strip in horizontal sections. The coolant flow rate of each section is different. When a certain section of the roll has a relatively large thermal crown owing to the lack of heat dissipation in time, by observing the shape meter at the exit of the fifth stand, that the shape of the position is not good is apparent. The emulsion flow rate at this position changes, taking away more heat, reducing the thermal crown of the roll, and improving the plate shape.

When considering the influence of the emulsion on the temperature field and hot roll profile of the work roll, the essence is to consider how much heat the emulsion takes away during rolling, that is, the size of the heat transfer coefficient. The heat transfer coefficient is a function of the temperature of the emulsion. Changing the temperature of the emulsion changes the heat transfer coefficient of the emulsion. When the emulsion temperature is relatively low, during rolling, the emulsion can take away more heat, making the rolling temperature drop faster. In view of the lateral distribution of the emulsion, the emulsion flow rate of each nozzle may vary, and the resulting heat transfer coefficient also differs. The calculation of the heat transfer coefficient is shown in Formula (2).

$$K_i = w_i^{c_1} T_S^{c_2} \exp\{c_3(1-\frac{C}{100}) + c_4 \frac{C}{100}\} \times 1.163 \tag{2}$$

where K_i is the heat transfer coefficient of the ith nozzle emulsion; w_i is the emulsification flow density of the ith nozzle, $L/(\min \cdot m^2)$; T_s is the strip temperature, °C; C is the emulsion concentration,%; and $c_1 = 0.264$, $c_2 = 0.212$, $c_3 = 9.45$, $c_4 = 9.73$.

Formula (2) shows that the emulsion temperature is usually lower than the strip temperature. In addition, the strip temperature, emulsion flow rate, and heat transfer

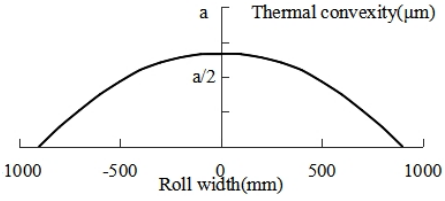
coefficient increase according to a certain exponential law. The concentration of the emulsion is reflected in the heat transfer coefficient expression in the form of an influence coefficient.

The specific influence of the heat transfer coefficient of the emulsion on the temperature field and hot roll profile is explored according to the relevant data of the tandem cold rolling mill and by consulting relevant literature. That the larger the heat transfer coefficient of the coolant, the more heat is taken away during rolling, and the lower the temperature of the strip steel are predicted. The temperature field under different heat transfer coefficients of the emulsion is compared, and the temperature field change under different heat transfer coefficients is obtained when the temperature of the emulsion is 55℃ as shown in Table 1.

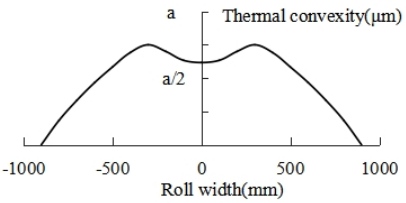
Table 1. Temperature field of simultaneous interpreting of different heat transfer coefficients under emulsion temperature of 55 °C

Heat transfer coefficiet		2300	3500	4600
F1	Entrance	45	45	45
	Exit	96.4	96.4	96.4
F2	Entrance	76.5	70.3	66.2
	Exit	148.5	142.2	138.1
F3	Entrance	103.4	87.1	77.3
	Exit	170.3	154	144.2
F4	Entrance	114.8	91.4	78.9
	Exit	169.8	146.5	134.0
F5	Entrance	114.6	88.7	76.3
	Exit	176.9	151.1	138.6
Coiling temperature		155.2	132.9	130.2

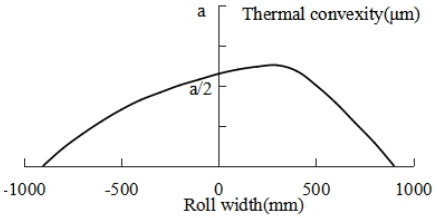
Through field calculation, the value of 3500$J/(s{\cdot}m2{\cdot}K)$ is the most suitable for rolling. The obtained temperature field is also the most suitable for rolling.



a) Ideal distribution of emulsion



b) Excessive cooling of the emulsion in the middle



c) Right part of the work roll indicating that the cooling is insufficient

Figure 2. Influence of emulsion transverse distribution of point cooling system on hot roll profile

As shown in Figure 2 a) is the ideal roll thermal crown when the roll is symmetrically cooled by the point cooling system of the rolling mill. For the edge wave shape defects, the shape of the rolled strip at this time is good. Figure 2 b) is the uneven flow control of the emulsion in the cooling system, resulting in a large flow of emulsion in the middle of the work roll, and excessive cooling between the rolls, which is reflected in the plate shape. It can easily cause edge waves, but it has a better compensation effect for the defects of the plate shape of the middle wave. Figure 2 c) shows that the flow rate of the emulsion on the right side of the work roll is insufficient. The heat of the work roll cannot be taken away in time, which can cause the roll to cool slowly. This is reflected in the plate shape, and unilateral waves can be easily formed. Similarly, the thermal convexity of the work roll under this cooling effect has a better effect on improving the single-sided corrugated shape defects.

In summary, the point cooling system of the rolling mill exerts a good adjustment effect on the hot roll profile of the work roll. When the emulsion nozzles are distributed along the transverse direction of the work roll and when the emulsion concentration, temperature, and flow rate of each nozzle can be adjusted independently, it has different cooling effects for various positions of the work roll. The hot roller shape can present a variety of curves, which plays an important role in reducing the local wave shape that exists along the longitudinal direction.

3.2. Mechanism of the influence of emulsion lateral distribution on plate shape

The lateral distribution of emulsion can affect the temperature field and thermal crown of the work roll. The change of thermal crown can also have a certain impact on the roll shape. To this end, this study calculates based on the flatness model described in the literature [15], as shown in Formula (3). The strip shape is mainly affected by the front tension, and the front tension is related to the hot roll shape of the work roll. The hot roll shape curve is reflected on the strip steel and affects the strip shape. When analyzing the influence of the lateral distribution of the emulsion on the plate shape, Figure 2 is taken as an example to perform uneven cooling at each position of the work roll. The unevenly cooled roll shape is then inputted into the plate shape simulation software to simulate the plate shape. The influence of the lateral distribution of the emulsion on the plate shape can be obtained.

$$f(\mathrm{x}_i) = -\frac{1-\nu^2}{E} \cdot \Delta\sigma_{1i} \cdot 10^5 I \tag{3}$$

where $f(\mathrm{x}_i)$ is the horizontal distribution value of plate shape, I, and $\Delta\sigma_{1i}$ is the transverse distribution value of the front tension.

Regardless of the influence of other factors on the plate shape, when the work roll is a straight line, the obtained plate shape is the most ideal. However, many factors affect the plate shape, so the cooling of each position of the work roll is not average.

When the cooling system in the middle part of the work roll is blocked, the cooling of the middle part is small, and the two ends of the work roll are sufficiently cooled. This kind of thermal convexity causes the middle of the work roll to bulge, which may cause the rolled plate shape to have local plate shape defects such as middle waves. At the same time, this type of roller has a certain effect on improving edge wave defects. To verify the guess, the roll shape is inputted into the shape simulation software. The

shape data at this time are calculated, and the shape curve corresponding to the shape of the emulsion before adjustment as shown in Figure 4 is drawn.

As depicted in the curve of the plate shape before the emulsion adjustment in Figure 3, when the cooling effect of the emulsion on the middle of the work roll is less than that on both sides, the resulting plate shape is likely to form a medium wave. To improve the plate shape defect, the hot roll shape must be adjusted, that is, the emulsion should be used to change the cooling condition of the roll. For the point cooling system, at this time, according to Formula (2), the emulsion flow in the middle of the work roll must be increased or the emulsion temperature should be reduced so that the heat transfer coefficient of the emulsion becomes larger and more heat can be taken away. The shape value can thus be improved. After adjusting the emulsion flow rate, the hot roll shape is adjusted accordingly. The simulation is performed again to obtain the plate shape distribution curve after the emulsion adjustment as shown in Figure 3.

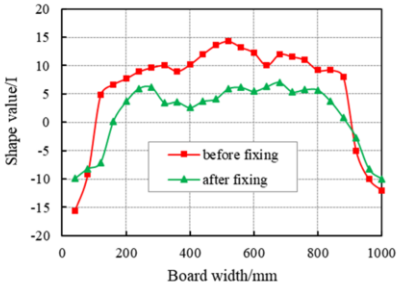


Figure 3. Profile curve of work roll before and after emulsion adjustment

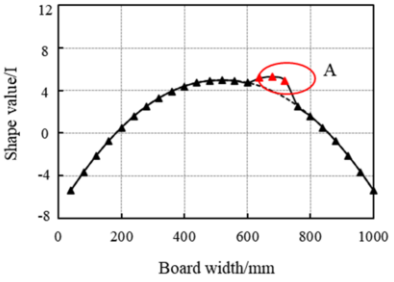


Figure 4. Shape of a steel when the emulsion nozzle is blocked

For the point cooling system, multiple emulsion nozzles are distributed laterally along the work roll to achieve cooling at different positions and obtain a high-quality plate shape. When the emulsion nozzle is blocked in a certain position, the cooling of the work roll covered by the nozzle is affected, and the cooling becomes insufficient, causing the change of the hot roll shape. Taking the rolled strip steel when the emulsion nozzle is blocked as an example, the relevant average shape curve is drawn according to the hot roll shape, as shown in Figure 4. In area A in the figure, the emulsion nozzle at this location is blocked. During the processing of the strip, the elongation of the strip at this location is too large, resulting in plate shape defects always existing at this location. Taking the average shape as an example, the shape data of the strip in the A zone is significantly higher than that of the nearby shape. When the nozzles in a certain area of the point cooling system are blocked, the shape defects are very serious.

3.3. Establishment of automatic prediction model for emulsion point cooling

According to the point cooling system of a five-stand tandem cold rolling mill in a steel mill, the nozzles are arranged horizontally along the work roll, and the emulsion flow of each nozzle corresponds to a length of the work roll. To ensure the uniform cooling of each position of the strip steel and avoid the occurrence of plate defects, a set of automatic prediction model of emulsion flow point cooling must be established.

The prediction principles of the automatic prediction model for emulsion flow point cooling are as follows. For a certain location area during rolling due to the long-

term flatness defect that resulted from the hot roll shape problem, which causes an excessively large difference between the flatness of the area and the adjacent position, the emulsion flow rate is predicted through the corresponding control model. There are two main objectives: (1) To ensure that the strip shape rolled by the cold tandem mill does not have obvious shape defects and (2) to ensure that the emulsion cooling capacity of each nozzle is uniform. In addition, it should also meet the constraints that the emulsion flow rate calculated by the control system is less than the maximum allowed emulsion flow rate and that the overall emulsion flow rate is a fixed value. For this reason, the setting objective function of the emulsion flow rate can be expressed as Formula (4).

$$\begin{cases} G(\mathrm{X}) = \psi g(\mathrm{X})_1 + (1\text{-}\psi) g(\mathrm{X})_2 \\ g(\mathrm{X})_1 = \mathrm{Max}\left\{F'(\mathrm{X})_{ij} - \bar{F}(\mathrm{X})_i\right\} - \mathrm{Min}\left\{F'(\mathrm{X})_{ij} - \bar{F}(\mathrm{X})_i\right\} \\ g(\mathrm{X})_2 = \dfrac{1}{nN}\sum_{i=1}^{n}\sum_{N=1}^{N}\sqrt{\left(F'(\mathrm{X})_{ij} - F(\mathrm{X})_i\right)^2} \\ \bar{F}(\mathrm{X})_i = \dfrac{1}{n}\sum_{i=1}^{n} F(\mathrm{X})_i \\ f(x) = \sqrt{\left(f_{x(i-1)} - f_{xi}\right)^2} + \sqrt{\left(f_{xi} - f_{x(i+1)}\right)^2} \le 2\xi \qquad i = 1,2,...,n-1 \\ w_i \le w_{\max} \qquad i = 1,2,...,n \\ \sum_{i=1}^{n} w_i = C \qquad i = 1,2,...,n \\ X = \{w_i,\ \ i = 1,2,...,n\} \end{cases} \tag{4}$$

where C is constant; w is total flow of emulsion; ψ is weighting factor; $g(X)_1$ is dynamic shape change function in space; $g(X)_2$ is the dynamic change function of the shape of the board over time; $F(X)_i$ is target board shape, I; $\bar{F}(\mathrm{X})_i$ is average target shape, I; $F'(\mathrm{X})_{ij}$ is the ith nozzle of the jth flatness monitoring section that corresponds to the actual strip shape, I; $f_{x(i-1)}, f_{xi}, f_{x(i+1)}$ is the plate shape value of the ith nozzle and three adjacent points corresponding to the plate shape on both sides; $w_{\max}$ is the maximum emulsion flow rate allowed by each nozzle of the equipment, L/min ; and ξ is two-point plate shape difference given accuracy.

The entire optimization process can be simply described as: finding a suitable $X = \{w_i, i = 1,2,...,n\}$ that minimizes the $G(X)$ under the constraint conditions. According to the optimization characteristics of this research, the Powell method is used for optimization calculation.

For the prediction device in the automatic prediction model of the emulsion point cooling system, there are mainly two situations. One is when one or more nozzles are blocked or insufficiently cooled along the lateral position of the work roll, causing serious plate shape defects at that position. The plate shape is more than 5–10 I larger than the plate shape of the adjacent area within a certain period. At this time, the computer predicts the blocked nozzle and manually adjusts it to clear the blocked nozzle. The other situation is that the emulsion flow at a certain point is too large, causing the work roll to be over-cooled. The computer predicts the over-cooled nozzle

and adjusts the flow through manual adjustment. For the above two cases, the cooling system starts to alarm when the difference in the shape values of the three adjacent sections exceeds the given accuracy value.

4. Development of local wave shape control model for tandem cold rolling mill

A domestic 1800 five-stand tandem cold rolling mill is optimized to improve the stability of strip rolling, reduce local wave shape caused by uneven distribution of emulsion, and finally obtain high-quality plate shape. Especially, this local wave shape control model is used to optimize the on-site flatness and application of this model to actual production. After the application of this model and six months of field data statistics, the proportion of strip steel with a wave height of ≦3 mm and a steepness of ≦1% in the local wave shape has increased from the previous 86.9% to no less than 95%. The quality of the plate shape is significantly improved.

The steel type DP0161D1 and the strip steel with the specification of 1.805 × 1118 mm are taken as example. Before optimization, the strip steel has a higher probability of local wave shape, and it is not easy to control. Therefore, the work roll temperature field and hot roll profile during the rolling of such a product are tracked, and the emulsion flow rate on site is redistributed. The fifth frame is taken as an example. Under the premise of ensuring the total flow of the emulsion is constant, it is redistributed along the axial direction of the work roll, and the thermal crown and thermal expansion of the work roll are obtained under different ratios, as shown in Figure 5.

Figure 5 illustrates that under the premise of a certain overall flow rate of the emulsion, as the on-site emulsion flow rate changes along the axis of the work roll, the sum of the thermal crown of the work roll changes accordingly. This is further reflected in the plate shape, causing local plate shape defects. Similarly, when the emulsion nozzle is blocked or the flow rate at a certain position has a sudden change, it easily causes uneven distribution of the emulsion flow rate and continues along the length direction. For this reason, the cloud diagrams with steel plates before and after the model is applied are compared, as shown in Figures 6 and 7.

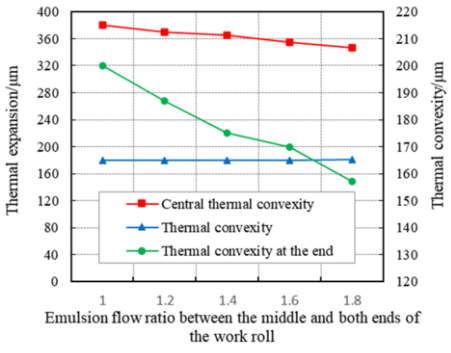


Figure 5. Variation curve of work roll thermal expansion and crown with emulsion flow ratio

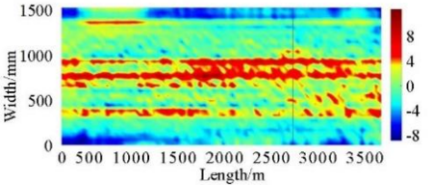


Figure 6. Shape before model application

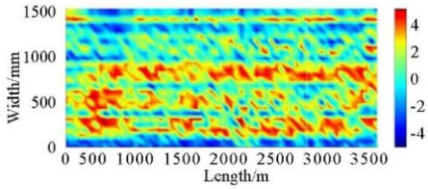


Figure 7. Shape after application of the model

Figure 6 shows that before the application of the model, during the rolling of the strip steel, there are relatively serious local flatness defects. The flatness value at certain positions even exceeds 10 I, causing the poor quality of the shape and affecting the user's product experience. Figure 7 depicts that after applying the emulsion point cooling automatic prediction model to field practice, the plate shape is significantly improved. The local plate shape defects are also effectively controlled. Most of the local flatness values are reduced to 5 I or less, which greatly improves the overall flatness quality.

5. Development of local wave shape control model for tandem cold rolling mill

(1) This article analyzes the generation of local waves. The formation of local waves is due to uneven distribution of emulsion along the width of the strip or blockage of the nozzles, resulting in uneven heating of the rolls. A partial wave shape is then formed, extending longitudinally.
(2) The influence mechanism of the lateral distribution of emulsion on the temperature field and hot roll shape of the work roll is studied. The hot roll shape can be affected by individually adjusting the concentration, temperature, and flow rate of the emulsion in each nozzle. In so doing, the treatment of the local wave shape of the strip steel can be realized.
(3) Based on the influence mechanism of emulsion on plate shape, an automatic prediction model of emulsion point cooling suitable for tandem cold rolling mills is established. This model can realize the prediction of local wave shape caused by nozzle clogging and uneven distribution of emulsion, thereby making manual intervention adjustments possible. The expected results are achieved through on-site application tests on products with typical specifications.

References

[1] Xiaogang L, Yiming F, Le L. Decoupling Predictive Control of Strip Flatness and Thickness of Tandem Cold Rolling Mills Based on Convolutional Neural Network. IEEE ACCESS,2020; 8: 3656-3667.
[2] Yu W, Changsheng L, Xin J, et al. Multi-objective optimization of rolling schedule for tandem cold strip rolling based on NSGA-II. Journal Of Manufacturing Processes,2020; 60: 257-267.
[3] Yanqing R, Wei Z. Application and optimization of shape model in automatic control of cold rolling mill[J]. Bengang Technology, 2018; (05):21-27.
[4] Xin J, Changsheng L, Yu W, et al. Investigation and Optimization of Load Distribution for Tandem Cold Steel Strip Rolling Process. METALS, 2020; 10(5):1-18.
[5] Xiaochen W, Quan Y, Hainan H, et al.Effect of work roll shifting control on edge drop for 6-hi tandem cold mills based on finite element method model. International Journal Of Advanced Manufacturing Technology,2020; 107(5): 2497-2511.
[6] Tao W, Qingxue H, Hong X, et al. Modification of roll flattening analytical model based on the plane assumption. Chinese Journal of Mechanical Engineering. 2018; 31(1):46-52.
[7] Lugt PM , Wemekamp AW , Napel W , et al. Lubrication in cold rolling: Elasto-plasto-hydrodynamic lubrication. Wear, 1993; 166(2):203-214.
[8] Smeulders B . Roll Cooling and Lubrication in Cold Rolling. 2014.
[9] Bohacek J, Raudensky Mi, Kotrbacek P. Remote Cooling of Rolls in Hot Rolling; Applicability to Other Processes. Metals,2021;11(7):
[10] Shao J, He AR , Yang Q, et al. Multi-parameter coupled subsection cooling regulation characteristics of work rolls in aluminum cold rolling. Chinese Journal of Engineering, 2015.

[11] Daron, Lloyd. Fundamentals of Roll Cooling and Control of Flatness at Primary Cold Reduction. Iron & Steel Technology, 2012.
[12] Wang P, Zhang D, Liu J, et al. Research and application of fuzzy selective roll cooling control for flatness control of cold-rolled strip. eucass proceedings, 2010.
[13] Xingtao Z, Gang T, Zhendong S. Fuzzy control of strip shape sectional cooling. Anshan Iron and steel technology, 2009; (01): 39-41.
[14] Liang XG. Research on shape control model optimization of six high tandem cold mill. Northeastern University, 2009.
[15] Zhenhua B. Mathematical model of high speed production process of tandem cold mill. China Machine Press, 2009.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 83
J. Xu et al. (Eds.)

doi:10.3233/ATDE220214

Numerical Study of the Effects of Tool Parameters on the Cutting Temperature Distribution and Ignition Risks of Magnesium Alloys

Linfeng LI[a], Jinyang XU[a,1], Guoqiang GUO[b] and Ming CHEN[a]
[a]*School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China*
[b]*Shanghai Spaceflight Precision Machinery Institute, Shanghai, China*

Abstract. Magnesium alloys have been widely applied in the advanced fields of aerospace, medical implants, automobile, *etc*. However, the ignition risks of magnesium alloys, especially at high cutting temperatures, have to be considered in the machining process. This article conducts numerical investigations on the effects of cutting tools on the cutting behaviors, especially the cutting temperature of magnesium alloy AZ31B. The impacts of the rake angle, the tool edge radius, and the friction coefficient are studied by simulations based on the orthogonal cutting models using the DEFORM software. The simulation results are studied and compared to analyze the correlations between the cutting parameters and the cutting temperature, as well as the underlying mechanisms. The conclusions of this numerical analysis can provide specific guidance to the design of cutting tools for the magnesium alloys.

Keywords. Magnesium alloys, numerical simulation, cutting temperature, tool geometry

1. Introduction

Among metallic materials, magnesium alloy has been extensively used in electronic products, high-speed railway, aerospace, and medical implant materials due to its extremely low density, high specific strength and stiffness, and good thermal and electrical conductivity [1].

Concerning the machining performance of magnesium alloys, various studies have been conducted. Tomac *et al.* [2] studied the impacts of cutting velocity and cooling conditions on the flank built-up formation in turning magnesium alloys. Tönshoff *et al.* [3] analyzed the contributions of tool coating to the cutting force and tool wear in magnesium cutting. Besides, the cutting force of magnesium-rare earth alloy in the plunge milling process was analyzed and modeled by Danis *et al.* [4]. Generally, magnesium alloys have good machinability in terms of relatively low power requirements, low cutting forces, and slow tool wear, and it is possible to achieve a high cutting efficiency in the machining of magnesium alloys.

[1] Corresponding Author. E-mail address: xujinyang@sjtu.edu.cn.

However, the machining of magnesium alloys is still challenging, especially under the use of large cutting parameters. The difficulties in the machining of magnesium are mainly problems caused by the heat generated in the metal machining process, *e.g.*, built-up edges and dimensional inaccuracy [5]. Besides, the high cutting temperature is prone to ignition risk due to the inherent low melting point and active chemical characteristics of magnesium. The thin chips with a large surface area are inflammable, and sparks indicating the possibility of fire are prevalent when the temperature exceeds 450 °C [6]. Therefore, the cutting temperature of magnesium alloys should be carefully analyzed to enhance the safety and efficiency of magnesium machining.

However, the conventional cooling method of applying cutting fluids during the machining process is not suitable for magnesium alloys due to the possibility of the explosion caused by the hydrogen gas generated in the reaction of magnesium alloy and water [7]. Hence, studies were conducted on the influencing factors of the dry cutting temperature of magnesium alloys. The methods for the mean flank temperature measurement of magnesium alloys were discussed by Fang *et al.* [8]. Mounting thermocouples in direct contact with the workpiece was suggested for a more accurate cutting temperature measurement. In their work, the cutting temperature was observed to rise with the increase of cutting velocity and the decrease of uncut chip thickness. An experimental study on the chip temperature of milling AZ31 and AZ91 magnesium alloys using thermography was conducted by Kuczmaszewski *et al.* [9]. The cutting depth was demonstrated to have the most significant contribution to the cutting temperature of magnesium, followed by the feed rate and cutting velocity. Besides, some of the advanced cooling methods, *e.g.*, the MQL and cryogenic cooling, are also adopted for the cutting of magnesium alloys. Viswanathan *et al.* [10] compared and optimized the cutting performance of AZ91D magnesium alloys under dry and MQL conditions by the Taguchi-based grey relational analysis. Experimental tests and simulations were conducted by Danish *et al.* [11] to compare the machining temperature of turning AZ31 magnesium alloys in dry and cryogenic environments, and the results showed a reduction of 60% on the maximum cutting temperature and reduced surface roughness for the cryogenic cooling.

The geometries and materials of cutting tools also have a significant impact on the temperatures in the metal cutting process. An investigation on the effects of cutting tool geometries on the cutting temperatures in the steel turning process was conducted by Saglam *et al.* [12]. Le Hieu Giang and Duc [13] concluded that the cutting temperature could be reduced by increasing of side and back rake angle, and the tool parameters were optimized by the Taguchi method. In the work conducted by M'Saoubi and Chandrasekaran [14], the effects of the microgeometries at the tool edge and the tool coatings on the cutting temperature were investigated. The lowest cutting temperature was obtained with sharp edges and tool coatings. Jagadesh and Samuel [15] studied the differences between the PCD tools and coated carbide tools in the micro-turning of Ti6Al4V. Temperature reductions up to 250 °C were observed for PCD tools in the simulation. The contributions of the tool geometries on the cutting temperatures were widely studied, particularly in the fields of processing hard-to-machine materials. However, there is still no quantitative analysis addressing the correlations between the tool geometry and the cutting temperature in machining magnesium alloys.

The finite element (FE) method has been widely adopted in the simulation of complex machining operations for various materials with proper material models. FE method has been proved as an effective and reliable tool that can partly replace the highly-costly and time-consuming experiments. Besides, direct observation for the

transient machining process as well as the distribution of underlying variables, which can be very difficult for the experimental methods, can be realized by the FE simulations. Regarding the magnesium cutting, serval FE investigations have been conducted. Salahshoor and Guo [16] studied the residual stresses of magnesium-calcium alloys in the high-speed milling process by FE simulation. Based on the Johnson-Cook (J-C) constitutive model, Pu *et al.* [17] developed a subroutine in the DEFORM to investigate the microstructural changes in AZ31B machining under two different cooling conditions: dry cutting and cryogenic cooling. There are also other constitutive models, *e.g*., the ZA model for magnesium alloys [18]; however, the J-C model is suitable for the conditions of coupled thermo-mechanical deformation with large strains and strain rates in the metal cutting processes and is available in most FE based software for its wide application on the numerical investigations of metal cutting processes [19].

Based on the literature review, it can be concluded that quantitative analysis on the relationship between the tool and the cutting temperature in the machining of magnesium alloys is significantly lacking. This work conducts numerical investigations on the effects of the cutting tools on the machining behavior, especially the cutting temperature of magnesium alloy AZ31B, and provides specific guidance to the design of cutting tools for magnesium alloys.

2. Simulation setup

To explore the impacts of the tool on the cutting performance of magnesium alloys, a 2-D orthogonal cutting model was developed in the DEFORM software. The cutting speed was set as 169.65 m/min, and the cutting depth was set as 0.15 mm. These cutting parameters are adopted from an actual machining process of magnesium alloys. The workpiece was assumed to be a rectangular plate with the size of 1.2 mm × 5 mm, and the tool size was approximately 1 mm × 1 mm. The impacts of the rake angle, the friction coefficient, and the radius of the tool edge were investigated by three sets of simulations. The detailed cutting parameters for the numerical simulations are listed in Table 1. The cutting responses of the AZ31B magnesium alloy were defined by the J-C constitutive model, and the adopted mechanical and thermal parameters are listed in Table 2 [20]. The carbide tool was set rigid, and only the thermal characteristics were defined by the parameters of carbide in the default material library, as shown in Table 3. The heat transfer coefficient between the tool and the workpiece is set as a default value of 45 N/s/mm/°C in DEFORM. Heat transfer with the environment is defined for non-contact edges of the tool and the workpiece. To increase the simulation accuracy with reasonable computational costs, the mesh density near the contact position is increased by a high mesh density window. Therefore, the tool is fixed in this simulation, and the X-direction velocity is applied at the bottom of the workpiece. To simulate the chip separation process, the ALE method was adopted in the DEFORM with the default parameters. The whole simulation model is shown in Figure 1.

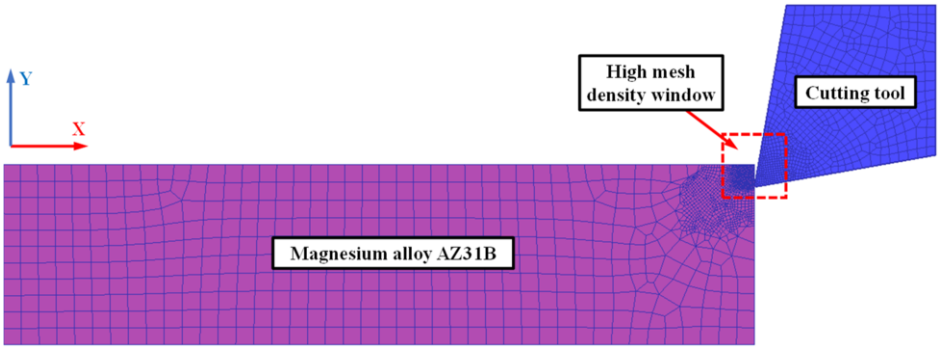


Figure 1. Schematic illustration of the simulation model.

Table 1. Tool parameters used in the simulations.

Set no.	Rake angle, α (°)	Clearance angle, γ (°)	Friction coefficient, μ	Tool edge radius, r (μm)
1	10/20/30		0.2	10
2	30	10	0.2	10/20/30
3	30		0.2/0.4/0.6	10

Table 2. Material parameters of magnesium alloy AZ31B [20].

Parameter	Value	Parameter	Value
Density	1780 kg/m^3	Elastic modulus	45 GPa
Poisson's ratio	0.34	Thermal expansion coefficient	2.7×10^{-5}
Thermal conductivity	78.5 W/(m·K)	Specific heat capacity	1.0 J/(kg·K)
A	172	B	360.73
C	0.012	m	0.95
n	0.45592	D1	-0.35
D2	0.6025	D3	-0.4537
D4	0.206	D5	7.2

Table 3. Thermal parameters of the carbide tool.

Density	Specific heat capacity	Thermal conductivity
15300 kg/m^3	0.95 J/(kg·K)	59 W/(m·K)

3. Results and discussion

3.1. Cutting performance under different rake angles

Figure 2 shows the variation of the maximum cutting temperatures and average cutting forces under different rake angles. The higher rake angle leads to the decreased maximum cutting temperature and the reduced X-directional cutting force, while the cutting force along the Y-axis rises. A significant reduction of the cutting temperature of 70.3 °C can be observed with the increase of the rake angle by 20°. This is mainly owing to the reduced heat generation in the first deformation zone during the chip

formation process. The shear angle in the cutting process increases with the rake angle, while the deformation of chips and the generated heat are reduced.

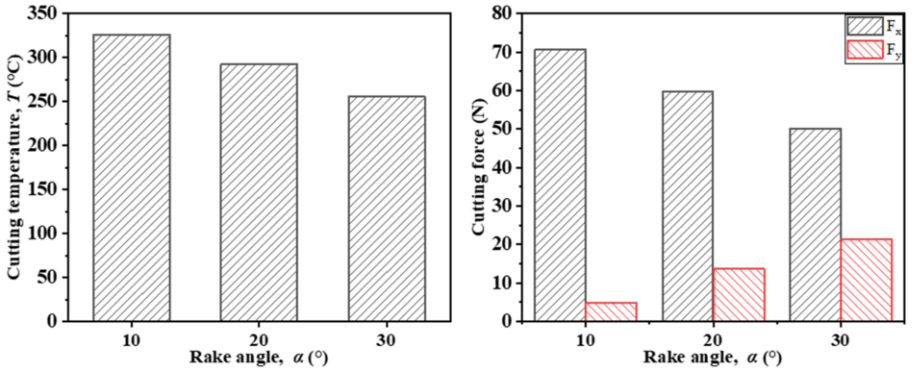


Figure 2. The cutting temperatures and cutting forces under different rake angles.

The temperature distributions at the end of the simulation under different rake angles are given in Figure 3. The thickness of the deformed chip is significantly reduced with the increase of the rake angle, which indicates less deformation in the first deformation zone. Besides, the temperature of the whole chip is affected by the reduced heat generation in the first deformation zone. Therefore, it can be concluded that the ignition risk in the machining of magnesium alloys could be significantly reduced by increasing the rake angle.

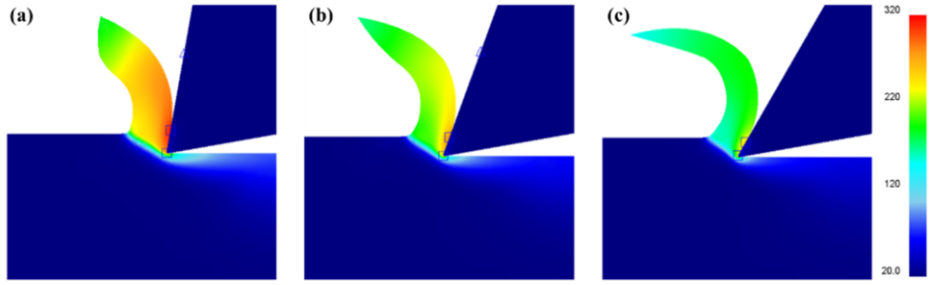


Figure 3. The temperature distributions under different rake angles: (a) 10°; (b) 20°; (c) 30°.

3.2. Cutting performance under different tool edge radii

The variations of the cutting temperatures and the cutting forces under different tool edge radii in the machining of magnesium alloys were plotted in Figure 4. With the increment of the tool edge radius, both the cutting temperatures and the cutting forces increase, while the cutting force on the Y-axis tends to decrease. The maximum difference in the cutting temperature is 62.5°C. In this situation, the additional heat generation in the third deformation zone caused by the blunt edge contributes greatly to the higher cutting temperature. The temperature distribution under different tool edge radii is given in Figure 5. It can be noticed that with the enlarged edge radius, the high-temperature zone transfers to the tool tip. This is mainly due to the large deformation and the heat accumulation caused by the rounded tool edge. However, the

high-temperature zone at the tool tip is isolated from air, which avoids the ignition risk. The edge radius has a minor contribution to the average chip temperature. The temperature increment at the separation point where the chip starts to contact air is slight because of the good thermal conductivity of magnesium. Although the temperature of the machined surface increases with a blunt tool edge, it maintains much lower than the average chip temperature. Therefore, an increased tool edge radius contributes slightly to the ignition risk in the magnesium cutting.

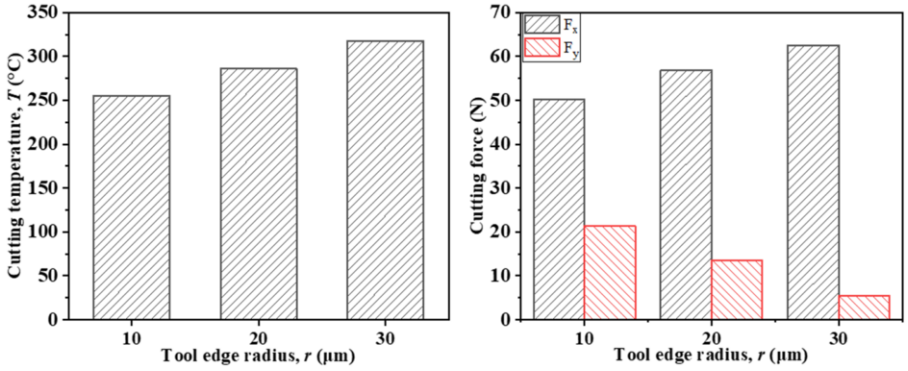


Figure 4. The cutting temperatures and cutting forces under different tool edge radii.

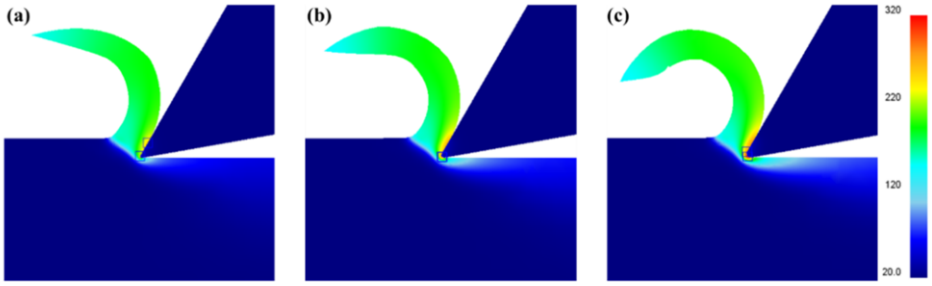


Figure 5. The temperature distributions under different tool edge radii: (a) 10 μm; (b) 20 μm; (c) 30 μm.

3.3. Cutting performance under different friction coefficients

Figure 6 shows the average cutting forces and the maximum cutting temperatures under different friction coefficients. The increase of the friction coefficient worsens the cutting performance of magnesium alloys. The cutting temperature and the main cutting force also increase with the friction coefficient, while the cutting force on the Y-axis drops down. A significant increment of the cutting temperature up to 85.9 °C can be observed when the friction coefficient increases by 0.4. This is mainly because of the increased frictional heat generation at the contact area. Figure 7 shows the temperature distribution under different friction coefficients. The intensification of the high-temperature zone near the tool tip can be noticed with an increased friction coefficient, which indicates the increased heat generation in the contact area. Besides, the deformation of the chip is also increased due to the increased frictional angle. The rise of the average chip temperature is slight; however, a high ignition risk may be

induced by the significant increment of the cutting temperature at the tool-chip separation point.

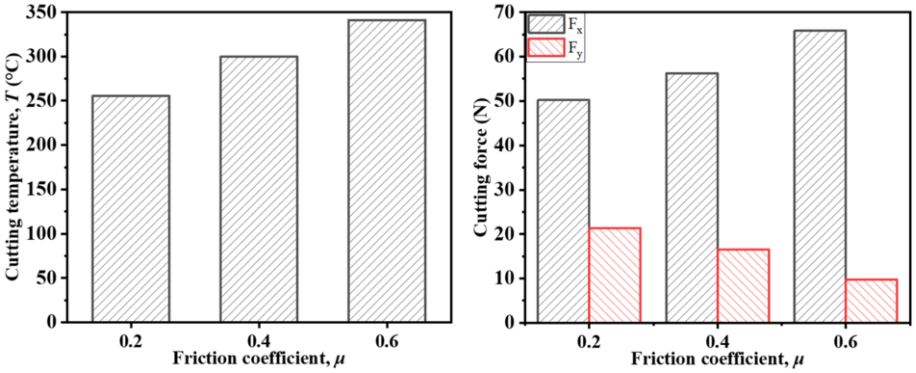


Figure 6. The cutting temperatures and cutting forces under different friction coefficients.

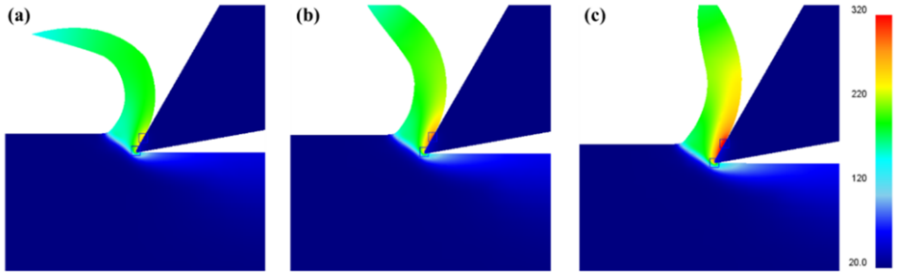


Figure 7. The temperature distributions under different friction coefficients: (a) 0.2; (b) 0.4; (c) 0.6.

4. Conclusions

The impacts of the tool geometries on the ignition risk and the cutting performances of magnesium alloy AZ31B were studied using the numerical method. The main conclusions obtained are summarized below.

(1) The tool parameters have significant effects on the cutting forces and cutting temperatures of magnesium alloys. Decrements in maximum cutting temperatures for about 60 °C can be observed in all the simulation sets. The cutting temperatures in all simulations are far below the ignition point of 450 °C.

(2) The chip morphologies vary significantly with the rake angle and the tool-workpiece friction coefficient, indicating their more significant impact on the chip deformation process in machining magnesium alloys.

(3) By analyzing the temperature distribution, it can be concluded that the ignition is most likely to occur on the chip, especially at the tool-chip separation point. The rake angle has the most significant impact on the ignition risk of magnesium alloys, followed by the friction coefficient and the tool edge radius.

Aknowledgment

This work is financially supported by the National Key Research and Development Program of China (2020YFB2010600).

References

[1] Lloyd JT, Jannotti PA, Jones TL. An overview of penetration behavior in magnesium alloys [J]. Mechanics of Materials 2021; 162: p. 104038.
[2] Tomac N, Tonnessen K, Rasch FO. Formation of flank build-up in cutting magnesium alloys [J]. CIRP Annals - Manufacturing Technology 1991; 40 (1): pp. 79-82.
[3] Tönshoff HK, Winkler J. The influence of tool coatings in the machining of magnesium [J]. Surface and Coatings Technology 1997; 94-95: pp. 610-616.
[4] Danis I, Monies F, Lagarrigue P, Wojtowicz N. Cutting forces and their modelling in plunge milling of magnesium-rare earth alloys [J]. International Journal of Advanced Manufacturing Technology 2015; 84 (9-12): pp. 1801-1820.
[5] Brito RF, Carvalho SR, Lima E Silva SMM. Experimental investigation of thermal aspects in a cutting tool using comsol and inverse problem [J]. Applied Thermal Engineering 2015; 86: pp. 60-68.
[6] Carou D, Rubio EM, Davim JP. Analysis of ignition risk in intermittent turning of UNS M11917 magnesium alloy at low cutting speeds based on the chip morphology [J]. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 2014; 229 (2): pp. 365-371.
[7] Shokrani A, Dhokia V, Newman ST. Environmentally conscious machining of difficult-to-machine materials with regard to cutting fluids [J]. International Journal of Machine Tools and Manufacture 2012; 57: pp. 83-101.
[8] Fang FZ, Lee LC, Liu XD. Mean flank temperature measurement in high speed dry cutting of magnesium alloy [J]. Journal of Materials Processing Technology 2005; 167 (1): pp. 119-123.
[9] Kuczmaszewski J, Zagórski I, Zgórniak P. Thermographic study of chip temperature in high-speed dry milling magnesium alloys [J]. Management and Production Engineering Review 2016; 7 (2): pp. 86-92.
[10] Viswanathan R, Ramesh S, Subburam V. Measurement and optimization of performance characteristics in turning of Mg alloy under dry and MQL conditions [J]. Measurement 2018; 120: pp. 107-113.
[11] Danish M, Ginta TL, Habib K, Carou D, Rani AMA, Saha BB. Thermal analysis during turning of AZ31 magnesium alloy under dry and cryogenic conditions [J]. International Journal of Advanced Manufacturing Technology 2017; 91 (5-8): pp. 2855-2868.
[12] Saglam H, Yaldiz S, Unsacar F. The effect of tool geometry and cutting speed on main cutting force and tool tip temperature [J]. Materials and Design 2007; 28 (1): pp. 101-111.
[13] Le Hieu Giang MDD, Duc PM. Investigation of effects of tool geometry parameters on cutting forces, temperature and tool wear in turning using finite element method and Taguchi's technique [J]. International Journal of Mechanical Engineering and Applications 2016; 4 (3): pp. 109-114.
[14] M'Saoubi R, Chandrasekaran H. Investigation of the effects of tool micro-geometry and coating on tool temperature during orthogonal turning of quenched and tempered steel [J]. International Journal of Machine Tools and Manufacture 2004; 44 (2-3): pp. 213-224.
[15] Jagadesh T, Samuel GL. Finite element simulations of micro turning of Ti-6Al-4V using PCD and coated carbide tools [J]. Journal of The Institution of Engineers (India): Series C 2016; 98 (1): pp. 5-15.
[16] Salahshoor M, Guo YB. Finite element simulation and experimental validation of residual stresses in high speed dry milling of biodegradable magnesium–calcium alloys [J]. International Journal of Mechanical Sciences 2014; 80: pp. 153-159.
[17] Pu Z, Umbrello D, Dillon OW, Lu T, Puleo DA, Jawahir IS. Finite element modeling of microstructural changes in dry and cryogenic machining of AZ31B magnesium alloy [J]. Journal of Manufacturing Processes 2014; 16 (2): pp. 335-343.
[18] Seif CY, Hage IS, Hamade RF. Extracting HCP Zerilli-Armstrong material parameters for magnesium alloy AZ31B from orthogonal cutting tests [J]. Journal of Materials Processing Technology 2021; 290: p. 116982.
[19] Korkmaz ME, Yaşar N, Günay M. Numerical and experimental investigation of cutting forces in turning of Nimonic 80A superalloy [J]. Engineering Science and Technology, an International Journal 2020; 23 (3): pp. 664-673.
[20] Feng F, Huang S, Meng Z, Hu J, Lei Y, Zhou M, et al. A constitutive and fracture model for AZ31B magnesium alloy in the tensile state [J]. Materials Science and Engineering: A 2014; 594: pp. 334-343.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 91
J. Xu et al. (Eds.)

doi:10.3233/ATDE220215

Dynamic Modeling and Simplified Analysis of EMA Pitching Mechanism

Ziping WAN[a], Rouyu TAN[a], Zhenyu ZHANG[a] and Dapeng FAN[a,1]
[a] *College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan, China*

Abstract. Aiming at the complexity of the dynamic model of the EMA pitching mechanism and the feasibility of simplification, this paper discusses the influence of rocker and pushrod on the dynamic characteristics of the EMA pitching mechanism and effectively simplifies the dynamic model on the basis of ensuring the accuracy. Firstly, a simplified dynamic model is established based on Lagrange dynamic equation, equivalent inertia, and equivalent force arm. Then, in the pitch range, the inertia and force arm of the rocker and pushrod are converted into the inertia and force arm of the load respectively, and the dynamic error compensation model is established. Finally, based on the position closed-loop control model after PID optimal parameter setting, the simplified characteristics of the EMA pitching mechanism in this prototype are clarified. In this paper, the simplified model error compensation value solution method improves the accuracy of the simplified model by 92%. The simplified feasibility judgment method can effectively judge the simplified feasibility of the EMA pitching mechanism.

Keywords. Equivalent inertia, Equivalent force arm, PID optimal parameter

1. Introduction

EMA(electric cylinder) pitching mechanism is mostly used in the servo system of dump trucks, radar, and artillery. These devices are non-high-precision servo devices. The main reason is that the kinematics and dynamics of the EMA pitching mechanism are nonlinear, it is difficult to establish an accurate model for control, and it includes nonlinear sources such as clearance, friction, and vibration. However, the EMA pitching mechanism has the characteristics of high rigidity and large load. When the load mass exceeds the fixed value, the advantages of the EMA pitching mechanism will be reflected. Therefore, it is necessary to model the EMA pitching mechanism with high precision.

The modeling of the EMA pitching mechanism is mostly based on various virtual prototype technologies[1][2][3], but this kind of method is difficult to realize online simulation and clarify the dynamic model equation. Based on analytical model of EMA is established, it is mostly analyzed based on the simplified model, but this simplified model does not consider the inertia and unbalanced torque of EMA and does not consider various nonlinear sources of the system. Instead, it uses a nonlinear control strategy for control[4-6], which is difficult to achieve a high-precision control effect.

[1] Corresponding Author, Dapeng Fan, College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan，China; E-mail: fdp@nudt.edu.cn

Therefore, based on high-precision modeling, this paper carries out high-precision modeling of EMA pitching mechanism by Lagrange, equivalent inertia, and equivalent force arm. On this basis, the simplified model, error compensation model[7], and complete model are established, and the accuracy of the three models is demonstrated.

2. Dynamic modeling of EMA pitching mechanism

In the EMA pitch servo system, the direct push EMA pitch mechanism is different from the single motion pair mechanism. The corresponding dynamic nonlinearity is introduced into different motion pairs, which makes the dynamic modeling of the mechanism very complex. The nonlinear source of the EMA pitching mechanism is shown in Figure 1.

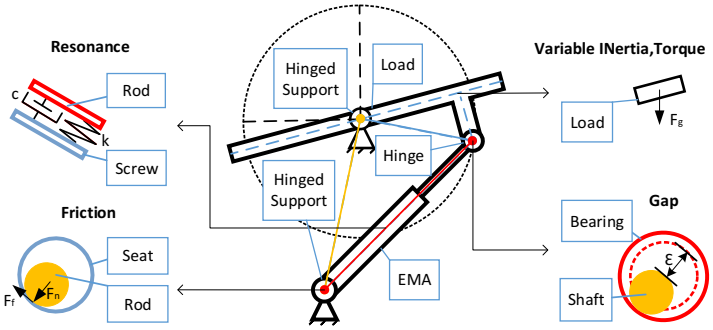


Figure 1. Nonlinear source of EMA pitch mechanism

In figure 1, the main nonlinear sources of the EMA pitching mechanism are variable inertia[10], variable torque[10], dynamic clearance[10], dynamic friction[10], and vibration[10], which will affect the dynamic characteristics of the system. Except for variable inertia and variable torque, other characteristics are related to operating conditions and need to be identified. Variable inertia and variable torque can be obtained through mechanical parameter calculation. Therefore, rigid body dynamics modeling of the EMA pitching mechanism can effectively separate the nonlinear sources caused by variable inertia and variable torque.

2.1. Complete modeling of EMA pitching mechanism

The simplified rigid body dynamics modeling of the EMA pitching mechanism is carried out without considering clearance, friction, and vibration. As the EMA pitching mechanism is a multi-part system, its disassembly can be described. The kinematic diagram of its load, rocker, and push rod within the pitching range is shown in figure 2.

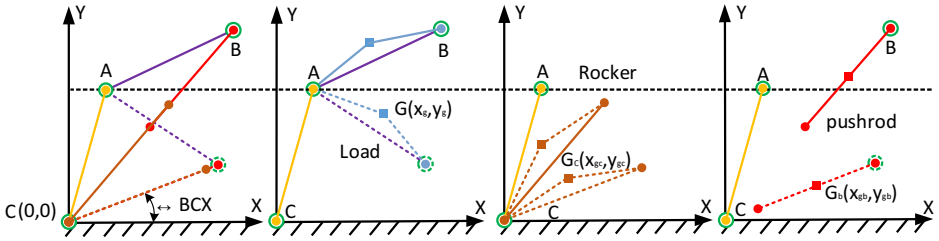


Figure 2. Decomposition of EMA pitching mechanism

In figure 2, G, Gc and Gb are the center of gravity of load, rocker and push rod respectively, and they need to be modeled in the same coordinate system, so the Lagrange method is the most suitable for modeling the above system. The kinetic energy equation (T) and potential energy equation(P) of the system are obtained.

$$\begin{cases} T = 0.5\times\left\{\left(J_c + m_c L_{CG_C}^2\right)\times\omega_{ACB}^2 + \left[J_b + m_b\left(L_{CB} - L_{BG_B}\right)^2\right]\times\omega_{ACB}^2 + J\times\omega_{BCA}^2\right\}; \\ P = F_g\times L_{AG}\cos\left(\angle BAX\right) + F_{gb}\times L_{CG_B}\cos\left(\angle BCX\right) + F_{gc}\times L_{CG_C}\cos\left(\angle BCX\right); \end{cases} \tag{1}$$

In equation (1), J, Jb, and Jc are the moment of inertia of load, pushrod and rocker respectively; mb is mass of push rod; ωACB and ωBAC are the angular velocity of ∠ACB and ∠BAC respectively; vb is the velocity of the length of LBC; Fg, Fgb and Fgc are the gravity of load, pushrod and rocker respectively; LAG, LCGB, and LCGC are the force arm of load, pushrod and rocker; ∠BAX is the angle between the load and horizontal line, ∠BCX is the angle between the load and horizontal line.

Based on the kinetic energy equation and potential energy equation, the system is modeled as a conservative system without considering the generalized output force. The dynamic model based on Lagrange is shown in equation (2).

$$\begin{cases} \dfrac{d}{dt}\dfrac{\partial T}{\partial\left(\omega_{BAC}\right)} - \dfrac{\partial T}{\partial\left(\angle BAC\right)} + \dfrac{\partial P}{\partial\left(\angle BAC\right)} = 0, \\ \dfrac{d}{dt}\dfrac{\partial T}{\partial\left(\omega_{BAC}\right)} = \left\{\left[\left(J_c + J_b + m_c L_{CG_C}^2\right) + m_b\left(L_{BC} - L_{BG_B}\right)^2\right]u + J\right\}\alpha_{BAC} \\ \qquad + \dfrac{d\left\{\left[\left(J_c + J_b + m_c L_{CG_C}^2\right) + m_b\left(L_{BC} - L_{BG_B}\right)^2\right]u + J\right\}}{dt}\omega_{BAC}, \\ \dfrac{\partial T}{\partial\left(\angle BAC\right)} = 0.5\times\left\{\dfrac{d\left[\left(J_c + J_b + m_c L_{CG_C}^2\right)u + m_b\left(L_{BC} - L_{BG_B}\right)^2 u\right]}{dt} + J\right\}\times\omega_{BAC}^2, \\ \dfrac{\partial P}{\partial\left(\angle BAC\right)} = F_g\times L_{AG}\cos\left(\angle BAX\right) + \dfrac{d\left[F_{gb}\times L_{CG_B}\cos\left(\angle BCX\right)\right]}{d\left(\angle BAC\right)} + \dfrac{d\left[F_{gc}\times L_{CG_C}\cos\left(\angle BCX\right)\right]}{d\left(\angle BAC\right)}, \\ L_{BC} = \sqrt{L_{AB}^2 + L_{AC}^2 - 2L_{AB}L_{AC}\cos\left(BAC\right)}, \\ u = \left\{\left(\dfrac{AB}{BC}\right)\left[\dfrac{\cos\left(BAC\right)BC - AC\sin\left(BAC\right)\sin\left(ACB\right)}{BC\cos\left(ACB\right)}\right]\right\}^2. \end{cases} \tag{2}$$

In equation (2), αBAC is the acceleration of ∠BAC, and ∠CAX is the angle between LAC and horizontal line. It can be seen from the above that the model is too complex and not suitable for dynamic numerical solutions, so it is difficult to carry out the online dynamic simulation.

2.2. Simplified modeling of EMA pitching mechanism

The lifting mechanism is a variation of the four-bar linkage. When doing the dynamic optimization and control of the four-bar linkage, because the mass of each link of the four-bar linkage is equal, it is necessary to carry out complete dynamic modeling to ensure the accuracy of the dynamic model. However, the application background of the EMA pitching mechanism is different, and its load mass is much larger than that of connecting rod and rocker in most cases, so it can be simplified. The kinetic energy equation and potential energy equation after simplified analysis are shown in equation (3).

$$\begin{cases} T = 0.5 \times J \times \omega_{BCA}^2; \\ P = F_g \times L_{AG} \cos\left(\angle BAX\right); \end{cases} \tag{3}$$

In equation (3), it can be seen that the dynamic model has been greatly simplified. Here, the driving force of EMA is introduced to change the conservative system into a non-conservative system. The dynamic model based on the Lagrange method is shown in equation (4).

$$\begin{cases} \dfrac{d}{dt}\dfrac{\partial T}{\partial\left(\omega_{BAC}\right)} - \dfrac{\partial T}{\partial\left(\angle BAC\right)} + \dfrac{\partial P}{\partial\left(\angle BAC\right)} = M_e, M_e = \left(F_b \times v_b\right)/\omega_{BAC}, \\ \dfrac{d}{dt}\dfrac{\partial T}{\partial\left(\omega_{BAC}\right)} = J \times \alpha_{BAC}, \dfrac{\partial T}{\partial\left(\angle BAC\right)} = 0, \dfrac{\partial P}{\partial\left(\angle BAC\right)} = F_g \times L_{AG} \cos\left(\angle BAX\right); \end{cases} \tag{4}$$

In equation (4), Fb is push force of EMA, Me is the equivalent torque. It can be set that the status variable is x1 = ∠BAC, x2=ωBAC, x3=αBAC, from which the state space equation of the pitching mechanism can be obtained as equation (5).

$$\begin{cases} x_1^{'} = x_2, x_2^{'} = \dfrac{1}{J} \times \left(M_e - M_g\right), \\ M_g = F_g \times L_{AG} \times \cos\left(\angle BAX\right); \end{cases} \tag{5}$$

In equation (5), the Mg is the gravity torque. The simplified dynamic model can simplify the dynamic simulation and controller design.

2.3. Error compensation modeling of EMA pitching mechanism

It can be seen from the simplified dynamic model that it can effectively simplify the dynamic model. However, due to ignoring the inertia and torque of EMA, has a slight

impact on the accuracy of the model, which will affect the amplitude response and phase response of the system under the frequency response. Therefore, it is important to compensate for the inertia and torque of EMA.

The inertia and moment of EMA are equivalent to the load by using the equivalent modeling method. The equivalent kinetic energy equation and potential energy equation are established, as shown in equation (6).

$$\begin{cases} T = 0.5\times\left\{\left(J_c + m_c L_{CG_C}^2\right)\times\dfrac{\omega_{ACB}^2}{\omega_{BCA}^2} + \left[J_b + m_b\left(L_{CB} - L_{BG_B}\right)^2\right]\times\dfrac{\omega_{ACB}^2}{\omega_{BCA}^2} + J\right\}\times\omega_{BCA}^2; \\ P = F_g\times\cos\left(\angle GAX\right)\left[L_{AG} + \dfrac{F_{gb}\times L_{CG_B}\cos\left(\angle BCX\right) + F_{gc}\times L_{CG_C}\cos\left(\angle BCX\right)}{F_g\times\cos\left(\angle GAX\right)}\right]; \end{cases} \tag{6}$$

In equation (6), if the changing inertia and the changing force arm are averaged within the motion range, the compensation values of inertia and torque can be obtained, which is equation (7).

$$\begin{cases} \Delta\left(J\right)_m = \dfrac{\int_{\min(\angle BAC)}^{\max(\angle BAC)}\left[\Delta\left(J\right)\right]d\left(\angle BAC\right)}{\max\left(\angle BAC\right) - \min\left(\angle BAC\right)}; \\ \Delta\left(L_{AG}\right)_m = \dfrac{\int_{\min(\angle BAC)}^{\max(\angle BAC)}\left[\Delta\left(L_{AG}\right)\right]d\left(\angle BAC\right)}{\max\left(\angle BAC\right) - \min\left(\angle BAC\right)}; \\ \Delta\left(J\right) = \left(J_c + m_c L_{CG_C}^2\right)\times\dfrac{\omega_{ACB}^2}{\omega_{BCA}^2} + \left[J_b + m_b\left(L_{CB} - L_{BG_B}\right)^2\right]\times\dfrac{\omega_{ACB}^2}{\omega_{BCA}^2}; \\ \Delta\left(L_{AG}\right) = \dfrac{F_{gb}\times L_{CG_B}\cos\left(\angle BCX\right) + F_{gc}\times L_{CG_C}\cos\left(\angle G_C CX\right)}{F_g\times\cos\left(\angle GAX\right)}; \end{cases} \tag{7}$$

Based on equation (7) and Lagrange method, the dynamic model after error compensation can be obtained, as shown in equation (8).

$$\begin{cases} \dfrac{d}{dt}\dfrac{\partial T}{\partial\left(\omega_{BAC}\right)} - \dfrac{\partial T}{\partial\left(\angle BAC\right)} + \dfrac{\partial P}{\partial\left(\angle BAC\right)} = M_e, M_e = \left(F_b\times v_b\right)/\omega_{BAC}, \\ \dfrac{d}{dt}\dfrac{\partial T}{\partial\left(\omega_{BAC}\right)} = \left[J + \Delta\left(J\right)\right]\alpha_{BAC}, \dfrac{\partial T}{\partial\left(\angle BAC\right)} = 0, \dfrac{\partial P}{\partial\left(\angle BAC\right)} = F_g\left[L_{AG} + \Delta\left(L_{AG}\right)\right]\cos\left(\angle BAX\right); \end{cases} \tag{8}$$

In equation (8), Δ (J) And Δ (LAG) the integral range can be changed, that is, the approximate linear model of the interval can be obtained.

3. Simplified analysis of EMA pitching mechanism

Based on the prototype, the mechanical parameters of the prototype are described. The parameter representation of the prototype is shown in Table 1.

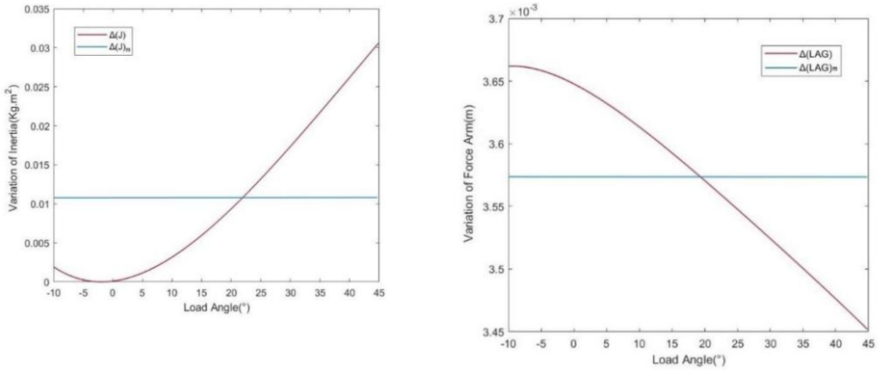
Table 1. Mechanical parameters of EMA pitching mechanism

J	J_b	$J_c+ m_c\cdot L_{CGC}$	m_b	m_c	m	K_I
21.61kg.m^2	0.01kg.m^2	0.34kg.m^2	1.55kg	9.57kg	170.91kg	1.9
$\angle XAC$	$\angle ACX$	$max(\angle BAC)$	$min(\angle BAC)$	$\angle GAB$	$\angle G_cCB$	K_M
97.14°	82.86°	17.77°	-37.23°	16.25°	16.09°	0.132
L_{AG}	L_{BGB}	L_{CGC}	F_c	F_{gb}	F_{ac}	d
0.05m	0.18m	0.11m	1674.92N	15.19N	93.79 N	0.004mm

In table 1, the inertia of the push rod and connecting rod is much less than that of the load, so it is possible to simplify the analysis.

3.1. Modeling method of EMA pitching mechanism

According to the system parameters and the method in equation (7), the error compensation value of the changed inertia and moment can be solved, and its function with ∠BAC is shown in figure. 3.



(a)Variation and error compensation of inertia (b)Variation and error compensation of force arm

Figure 3. Variation and error compensation of inertia and force arm

In the actual simulation, we use the Simulink model suitable for on-line simulation, including simplified model, error compensation model and complete model, as shown in figure 4.

Through the above model, the error compensation of inertia is 0.011, the error compensation of force arm is 0.003. The open-loop performance and closed-loop performance can be simulated. The following definitions can be used to simulate the accuracy and error influence degree of the model.

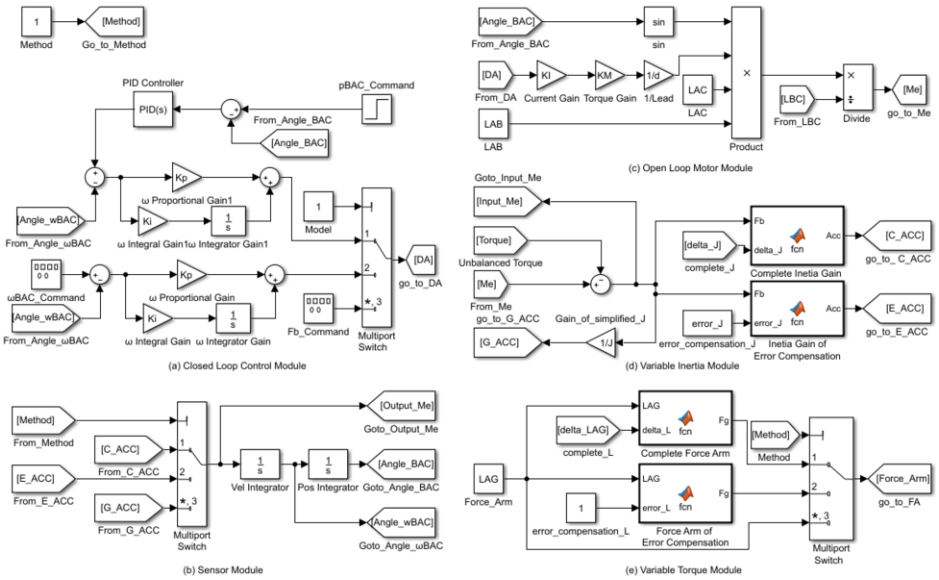


Figure 4. Simplified model, error compensation model and complete model

3.2. Model influence analysis of model simplification

To verify the effect of the error compensation model, the amplitude-frequency response signal under frequency excitation is considered, and the excitation function is set as the sinusoidal signal with the highest frequency in the pitch range. The signal parameters are shown in equation (9).

$$y = 0.48 * \sin(0.25 * t) \tag{9}$$

According to equation (9), the output torque signal and error signal of the complete model, simplified model and compensation model are simulated under the open-loop model. The simulation results are shown in figure 5.

Since the amplitude response and phase angle response under the open-loop model are related to the transfer function of the system, the error value can effectively represent the accuracy of the model before and after the error, and the improvement rate of the model accuracy can be characterized as the error integral value in one cycle, which is defined as equation (10).

$$\rho = \frac{\max\left[a_{J+\Delta(J)_m}\right] - \max\left[a_J\right]}{\max\left[a_{J+\Delta(J)}\right] - \max\left[a_J\right]} \tag{10}$$

In figure 5, it can be calculated that the improvement rate of model accuracy is 92%.

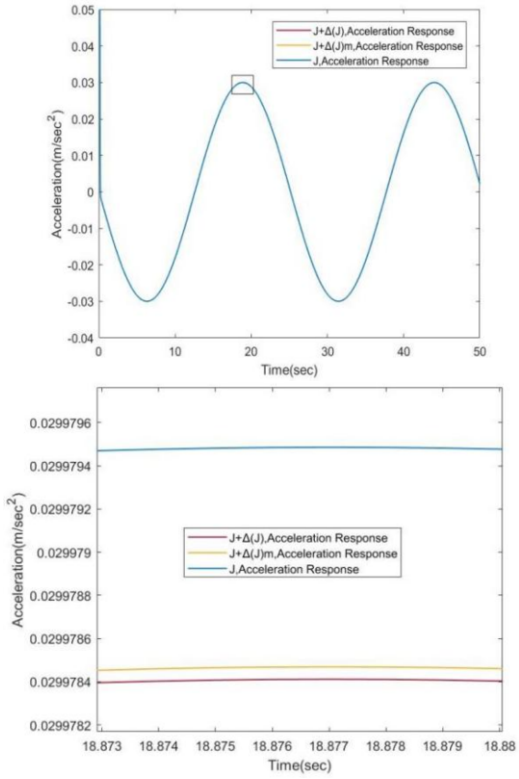


Figure 5. Acceleration response of complete model, simplified model and compensation model

3.3. Feasibility analysis of model simplification

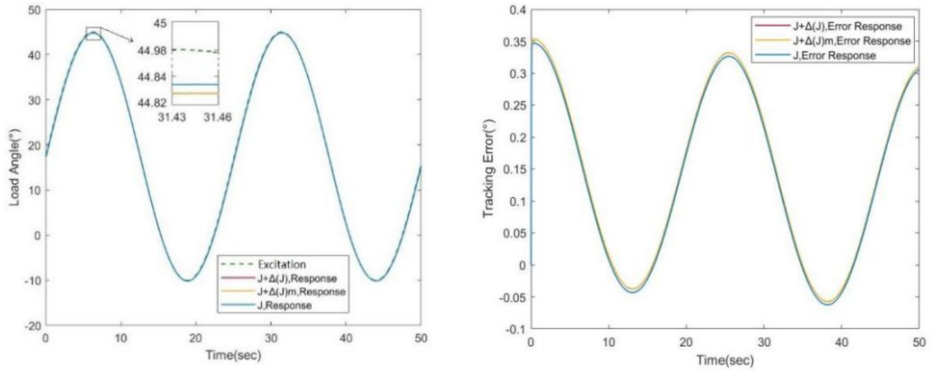
The feasibility analysis mainly considers the impact of model inaccuracy on the system in the case of closed loop. In the servo control of EMA pitch system, the most important index is the system target tracking accuracy, so the following methods are used for analysis:

- Taking the simplified model as the original model, an open-loop system is established.
- The optimal parameters (kp:99.99, ki:0.01) of PI controller of velocity loop and the optimal parameters (kp:36.45, ki:0.04, kd:0.047) of PID controller of position loop are adjusted based on the original model by 'Fmincon' function of Matlab.
- PI and PID controller is used to control the error compensation model and the complete model.
- Compare the control effects of the three models to verify the feasibility of the model.

Through the above method, the weighting function of error value, controller output, and rise time are used as the optimal objective function of the controller with step signal; The motor power constraint, motor speed constraint, motor torque constraint,

and Da output are used as the constraint functions of the controller, and the interior point method is used to optimize the proportional parameters K_p and K_i integral parameters of the controller. The set parameters are shown in table 1.

Using the signal of equation (9) as excitation, three response signals and error signals are obtained, which are shown in figure.6 respectively.



(a) Position loop command and response (b) Position loop error response

Figure 6. Tracking response and error of position loop

In figure 6, the tracking error caused by the push rod and connecting rod is 0.005 ° (0.087mrad), which is not the main factor affecting the tracking error. Therefore, it can be ignored, and the simplified model is used to replace the original model.

4. Conclusions

Based on the dynamic characteristics of EMA pitching mechanism, this paper discusses the sources of nonlinearity in the system, and establishes a simplified model, error compensation model and complete model. In this paper, the following conclusions can be summarized:

- The complete model of EMA pitching mechanism can be effectively modeled by Lagrange equation, equivalent inertia and equivalent force arm.
- The simplified model of EMA pitching mechanism can effectively represent the dynamic characteristics of EMA pitching mechanism.
- The error compensation model of EMA pitching mechanism can effectively improve the accuracy of the simplified model of EMA pitching mechanism.

Based on the above conclusions, the EMA pitching mechanism can be effectively simplified, which provides an accurate model for the model-based EMA pitching mechanism control method, and can effectively separate the other nonlinear error sources.

Acknowledgments

National key R & D plan support project (2019YFB2004700)

References

[1] Xin Z, Jianwu Z, Qingliang Z, Hanzheng D. Optimization Design of Four-Bar Linkage of Hydraulic Support Based on ADAMS. IEEE Computer Society, 2009; pp 338-341.
[2] Changlin M, Feng L, Yongbao F, Xiaoguang Y. The Study on Dynamic Modeling of Complex Mechanism Based on Multi Software Collaboration. Advanced Materials Research. 2014; 945-949: pp 87-90.
[3] Yinghong T, Zhongwei L, Bing H. Optimization Design for Crank-rocker Mechanism Based on Genetic Algorithm. IEEE Computer Society; 2012.
[4] Jiang XL, Qi C, Fan SX, et al. Disturbance observer based controller design of a speed servo system. CSAA/IET International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering.
[5] Lech MG, Tomasz T. PMSM serv-drive control system with a state feedback and a load torque feedforward compensation. COMPEL-The International Journal for Computation and Mathematics in Electrical and Electronic Engineering. 2013; 32(1): pp 364-382.
[6] Chao Q, Xianliang J, Xin X, Dapeng F. A SAKF-Based Composed Control Method for Improving Low-Speed Performance and Stability Accuracy of Opto-Electric Servomechanism. Applied Sciences, 2019, 9(21):4498. Applied Sciences, 2019, 9(21): p 4498.
[7] Xuemei N, Guoqin G, Xinjun L, et al. Dynamic modeling and simplified analysis of a new drive redundant parallel mechanism. Journal of mechanical engineering. 2014; 50 (19): p 9.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 101
J. Xu et al. (Eds.)

doi:10.3233/ATDE220216

Multi-Granularity Feature Fusion for Person Re-Identification

Yue YANG[a, 1] and Lizuo JIN[a]
[a]*School of Automation, Southeast University*

Abstract. Due to the limited ability of feature expression learned by one branch, designing a Multi-granularity network for feature extraction has become one of the important directions in the field of person re-identification. This paper designs a Multi-granularity feature fusion network (MFN) to enhance person feature extraction. The network is composed of global branches and local branches, and the former use the convolution pyramid to extract multiple scales features, through the channel attention module (Split Attention, SA) fusion of global branches and local branches, so that the global branch to obtain a strong ability to express persons features; the local branches is a feature map extracted from the backbone network based on the idea of Part-based convolutional baseline. Split the feature map horizontally into 4 branches, and the ID loss is calculated separately for each local branch. The classification loss in the total loss is consist of the ID loss of the 4 local branches. The two interact in parallel to improve the recognition ability. The experimental results on the Market1501 and DukeMTMC-ReID datasets, Rank-1/mAP reached 94.9%/86.4% and 87.5%/76.8% respectively.

Keywords. Person re-identification; Deep learning; Pyramid convolution; Attention mechanism

1. Introduction

Person re-identification focus on finding specific person target across surveillance devices, also known as pedestrian search. In recent years, person re-recognition technology has played an active role in video surveillance and public safety, and has become an important research significance. But various complex situations including background changes, illumination changes and changes in person posture can cause interference in recognition in the process of practical application. Therefore, extracting richer pedestrian identity features that can be distinguished from others is the focus of person re-recognition research.

Traditional Re-ID research directions are divided into representational learning and metric learning. The representational learning-based methods mainly extract pedestrian features such as color, SIFT [1] and LAP [2]. The pedestrian Re-ID methods based on metric learning achieves the purpose of distinguishing different persons by designing deep neural networks and distance functions to make the distance smaller for same persons and make distance larger for different persons. XUAN [3] used dynamic matching local information (NMLI) to dynamically match the minimum path to solve the problem of unreasonable distance of persons features due to unaligned feature

[1] Yue YANG: School of automation, Southeast University, 220191615@seu.edu.cn

chunks. SUN [4] horizontally cut the feature map extracted by the backbone network into six blocks, and the local network improves the model representation by learning local features. FAN [5] adopted the method of supervising channel features by spatial information based on the literature [3,4], and the SCP-Net made great progress in both half-body and full-body person re-identification by supervising global branches through local branches. Wang [6] argued that the PCB model ignored the influence of the whole on local learning, thus proposing the multi-grain size model MGN. ZHENG [7] proposed a progressive pyramid model Pyramid from coarse to fine grain size, which achieved good results. However, the image chunking approach still suffers from alignment problems between features, which can lead to information loss and greater computational effort, at the expense of some performance.

The feature extraction network in this paper is based on ResNet50 [8], modified by adding pyramid convolution [9] for multi-level feature extraction, and fusing the feature information of each level by means of split attention mechanism [10]. The network will learn the feature weights of each level and each channel, and share the shallow and deep information to extract a more comprehensive distance feature. The extracted global features are used to get triplet-loss and center-loss in the training process and as distance metric for retrieval. The local branches are extracted from the features of the backbone network ResNet50, and the features are horizontally chunked into 4 blocks. The global average pooling is used to obtain 4 local feature vectors, and the classification loss is calculated separately for each of these four vectors during network training. Testing discards local branches of the network and uses the distance metric of global features for pedestrian retrieval ranking.

2. Algorithmic Framework

The network framework used in this paper is shown in Figure 1 and consists mainly of a backbone network, global branches and local branches.

(a) Backbone：The backbone used in this paper is improved from the mainstream feature extraction network ResNet50. The residual structure used in this network can decrease the gradient degradation problem of deep learning, and the structure has certain cascading characteristics, which can better combine shallow feature information. The global feature branch is obtained by adding convolutional pyramids and split attention blocks. The pyramid convolution extracts feature information at different scales through convolutional kernels at different scales, resulting in a richer cascade of feature information, thus improving the Generalization and robustness of the network. The ResNet50 network can be used to fuse the feature information of each layer by stitching or adding them together to share the shallow and deep layer feature information.

(b) Global branches: The modified ResNet50 is used as input for the global branches. Adding the attention mechanism and GAP (Global Average Pooling) to get the global feature vectors for calculating triplet-loss and center-loss.

(c) Local branches: The local branch in this paper refers to the idea of the well-known paper PCB, which divides the feature level extracted by the backbone into 4 blocks, and then obtains 4 local feature vectors through global average pooling for calculate the classification loss, i.e. how many person IDs in the training set are divided into how many classes, and calculates the ID-Loss respectively.

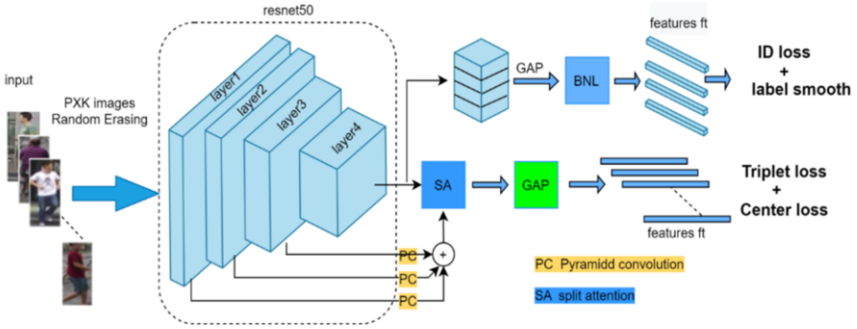


Figure 1. Network structure

2.1. Pyramid convolution

As in the case of multi-scale feature extraction, the pyramid convolution, extracts feature information at different scales through multi-scale and multi-depth convolution kernels, and controls the depth of the kernels through group convolution. Compared to the standard convolution, the convolution pyramid extracts richer semantic information without excessive computational and parametric effort thanks to the group convolution. Setting the size of the convolution kernel in pyramid to 9×9、7×7、5×5、3×3 with decreasing kernel size and increasing kernel depth. Pyramidal convolution is applied to different layers of the backbone, with each layer of the convolutional pyramid having different scales and depths of convolutional kernels, as shown in Table 1. The output of the three convolutional pyramids is superimposed on the channels as 2048 dimensions. The final feature map is fused with the layer4 output by way of channel attention.

Table 1. Pyramid structure corresponding to ResNet50

layer	kernel size	depth	group
1	3×3	64	1
	5×5	64	4
	7×7	64	8
	9×9	64	16
2	3×3	128	1
	5×5	128	4
	7×7	256	8
3	3×3	512	1
	5×5	512	4

2.2. Attention mechanisms

As some basic convolutional neural networks such as ResNet50 are designed for image classification, the limited perceptual field size and lack of cross-channel interactions make these networks less effective for niche areas such as person re-identification. This means that to improve the performance of a given computer vision task, ‘network surgery’ is required to modify ResNet50 to make it more effective for a given task. The idea of split attention (SA) used in this paper is derived from ResNeSt, and SA is borrowed from SE-Net [11] and SK-Net [12], where the innovation is to weight the

channels by channel attention, increasing the weight of useful channels and decreasing the weight of useless channels. Figure 2 is the structure of SA module.

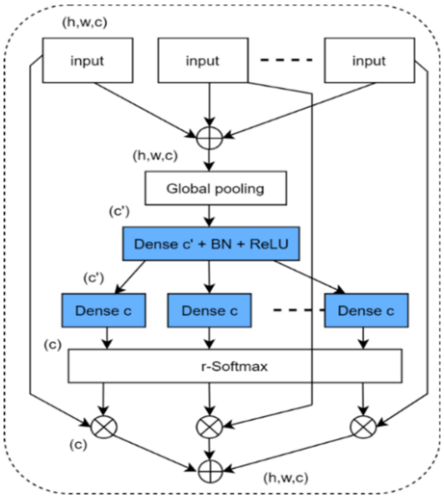


Figure 2. Split attention module structure diagram

2.3. Loss function

Many current person re-identification networks combine metric loss functions and classification loss functions trained together to jointly constrain features. The total loss in this paper are consist of Classification loss L_{id}, Center loss L_{center} , and Triplet-hard L_{tri_hard}. To prevent overfitting during training, a smoothing operation is performed on the identity labels (Label smoothing, LS), and LS is a common method to prevent overfitting in classification tasks. The classification loss function L_{id} with LS is as follows.

$$L_{id} = -\sum_{i=1}^{N} q_i \log p_i \tag{1}$$

$$q_i = \begin{cases} 1 - \frac{N-1}{N}, & i = y \\ \frac{\varepsilon}{N}, & otherwise \end{cases} \tag{2}$$

N is the number of people in training set., p_i is the predicted probability of specific person identity, and y denotes true label information of the person identity. formula 2 means the LS for identity label y. ε is a hyperparameter with a small value, set $\varepsilon = 0.1$ in this paper. The center loss L_{center} is defined as:

$$L_{center} = \frac{1}{2}\sum_{i=1}^{m} \left\| f_i - C_{y_i} \right\|_2^2 \tag{3}$$

Where m is the number of samples of a single person ID in the current batch, f_i denotes the feature tensor of sample I, and C_{y_i} denotes the feature center of the y_i class.

Triplet-hard loss L_{tri_hard} is improved from the triplet loss, which consists of three pictures: the anchor, the negative picture and the positive picture. Optimizing the triplet loss is to make the positive image and anchor's feature distance smaller, making the negative image and anchor's feature distance larger. The hard sample triplet loss is to make the most difficult positive image and anchor's distance smaller, making the most difficult negative image and anchor's feature distance larger in a batch. The Triplet-hard loss L_{tri_hard} is defined as:

$$L_{tri\_hard} = \frac{1}{P\times K}\sum_{an\epsilon batch}\left(max_{pos\epsilon A}D_{an,pos} - min_{neg\epsilon B}D_{an,neg} + M_{ar}\right)_{+} \quad (4)$$

Where $M_{ar}$is the set hyperparameter, $D_{an,pos}$ is the feature distance of the anchor and the positive image, $D_{an,neg}$ is the feature distance of the anchor and the negative image.

The overall loss function is as follows.

$$L_{total} = L_{id} + L_{tri\_hard} + L_{center} \quad (5)$$

3. Experimental results

3.1. Evaluation Metrics and Dataset

The effectiveness of this method was evaluated on the Market1501 and DukeMTMC-reID datasets. The Market1501 is the most commonly used dataset in the domain of pedestrian Re-ID, which was collected and released from Tsinghua University through 6 HD cameras. There are 12936 persons image of 751 different persons in training set, 19732 images of 750 different person in test set. The DukeMTMC-reID dataset was captured from 8 cameras, resulting in 34611 imag2es. There were 16522 persons image in training set and 2228 persons image in the gallery and 17661 persons image in test set. The evaluation metrics generally used Rank-k and Mean Average Precision (mAP) to assess the performance of the person re-recognition model. The evaluation metric used in this paper is Rank-1and mAP.

3.2. Experiment preparation

The experiments were conducted on Intel i7-8700K processor with 32GB of RAM and an NVIDIA GeForceRTX2060S GPU with 8GB of video memory, using the Pytorch deep learning framework to build a person re-identification network. The input person image size was set to 256×128, and each batch include 64 training images. During the training process, Adam was chosen as the optimizer. The experiment adopts L2 regularization. A learning rate decay strategy with Warm-Up was adopted. The initial learning rate has been set to 3.5×10^{-6}, which grew linearly to 3.5×10^{-4} within ten iterations. and then decays to 3.5×10^{-5} at the 50th epoch, decays to 3.5×10^{-6} at the 90th epoch. The whole training process lasted for 110 epochs.

3.3. Experimental result

Contrast experiments on Market1501 and DukeMTMC-reID datasets. An example graph of the experimental results is shown in Figure 3. The first cell of each row is the query image, and the next 10 images of each row are the top 10 query results of the test sample. The green labels indicate the correct query results, and the red labels indicate the wrong query results.

Figure 3. Example graph of person Re-ID results

The experiments result is shown in Table 2, which compare the accuracy of ResNet50_Baseline, ResNet50+Local Branches, MFN, and MFN+ (Label-Smooth, Random-Erasing, Center-Loss) on the above two datasets. To ensure a fair comparison, all experimental methods in this paper without using re-ranking [14].

Table 2. Performance metrics of our method (%)

Network	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
ResNet50	88.5	76.1	81.2	64.3
ResNet50+LB	90.7	77.4	82.2	70.4
MFN	91.4	80.3	82.4	71.2
+ Label-Smooth	92.3	82.1	83.9	72.5
+ Random-Erasing	93.8	85.8	86.2	75.9
+ Center-Loss	94.9	86.4	87.5	76.8

As shown in Table 2, the proposed multi-granularity feature fusion network MFN proposed in this paper together with some image enhancement means (horizontal flipping, random erasure, etc.) can achieve a better level of metrics on the two public datasets under appeal. Compared to ResNet50 as a feature extraction network, MFN improves the metrics without bringing too many parameters. The table 3 shows that the method has an advantage over the current mainstream methods in Rank-1 and mAP.

Table 3. Performance metrics of our method and the comparative method (%)

Network	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
SVDNet[15]	82.3	62.1	76.7	56.8
HA-CNN[16]	91.2	75.5	80.5	63.8
PCB+RPP[4]	93.8	81.6	83.3	69.2
SCPNet-a	94.1	81.8	84.8	68.5
IANet[17]	94.4	83.1	87.1	73.4
OSNet[18]	94.8	84.9	88.6	73.5
CtF[19]	94.2	84.9	86.9	75.6
MFN+(LS、RE、CL)	94.9	86.4	87.5	76.8

4. Conclusion

In order to improve person re-recognition backbone, this paper designed a method based on the ResNet50 benchmark network with a multi-granularity feature fusion network as the person feature extraction network, combining the advantages of local branching and global branching, extracting person feature information at multiple scales, paying more attention to the differences between person and the background while preserving person whole-body information and deepening person contours, so that the final feature representation is more discriminative. Experimental result on the two public datasets shows that the designed multi-granularity feature fusion method for pedestrian Re-ID has a significant advantage over other method. How to find better methods to extract more significant features and further promote the metric of pedestrian Re-ID on more datasets will be the next step to be carried out.

References

[1] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 2004; 60(2):91-110.
[2] Ojala T, Pietikainen M, Maenpaa T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence. IEEE, 2002; 971-987.
[3] Xuan Z, Hao L, Xing F, et al. AlignedReID: Surpassing Human-Level Performance in Person Re-Identification. 2017.
[4] Sun Y, Zheng L, Yang Y, et al. Beyond Part Models: Person Retrieval with Refined Part Pooling. 2017.
[5] Fan X, Luo H, Zhang X, et al. SCPNet: Spatial-Channel Parallelism Network for Joint Holistic and Partial Person Re-Identification. Springer, Cham, 2018.
[6] Wang G, Yuan Y, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification. Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea:ACM, 2018; 274 – 282.
[7] Zheng F, Deng C, X Sun, et al. Pyramidal Person Re-IDentification via Multi-Loss Dynamic Training. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.
[8] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. IEEE, 2016.
[9] Duta I C, Liu L, Zhu F, et al. Pyramidal Convolution: Rethinking Convolutional Neural Networks for Visual Recognition. 2020.

[10] Zhang H, Wu C, Zhang Z, et al. ResNeSt: Split-Attention Networks. 2020.
[11] Jie H, Li S, Gang S, et al. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; PP(99).
[12] Li X, Wang W, Hu X, et al. Selective Kernel Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
[13] Zhao C, Lv X, Zhang Z, et al. Deep Fusion Feature Representation Learning With Hard Mining Center-Triplet Loss for Person Re-Identificatio. IEEE Transactions on Multimedia, 2020; PP(99):1-1.
[14] Qi M, Y Wei, Gao K, et al. Re-Ranking Person Re-Identification withForward and Reverse SortingConstraints. Pacific Rim Conference on Multimedia. Springer, Cham, 2018.
[15] Sun Y, Zheng L, Deng W, et al. SVDNet for Pedestrian Retrieval. 2017 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2017.
[16] Li W, Zhu X, Gong S. Harmonious Attention Network for Person Re-Identification. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018.
[17] Hou R, Ma B, Chang H, et al. Interaction-And-Aggregation Network for Person Re-Identification. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
[18] Zhou K, Yang Y, Cavallaro A, et al. Omni-Scale Feature Learning for Person Re-Identification. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2020.
[19] Wang G, Gong S, Cheng J, et al. Faster Person Re-Identification. 2020.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 109
J. Xu et al. (Eds.)

doi:10.3233/ATDE220217

Research on Weak Vortex Signal Detection Based on Stochastic Resonance

Jianxiu LIU [a, 1], Zhaoxia SHI [a], Hui WANG [a], Yu ZHANG [a], Rong HUANG [a] and Zihao LIU [a]
[a] *School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou, China*

Abstract. Vortex flowmeters are easily affected by vibration interference of pipelines and various equipment when used in industrial field. Especially in the measurement of small flow, the noise signal will be superimposed on the output signal of the sensor, the vortex signal is easily submerged, and the measurement of small flow is limited. Aiming at the problem that the small flow vortex signal is very weak and difficult to detect under actual working conditions, an adaptive stochastic resonance(ASR) detection method based on ensemble empirical mode decomposition(EEMD) is proposed. Firstly, the vortex signal is denoised and preprocessed by EEMD, and the appropriate component is selected to reconstruct the signal by using the correlation coefficient as the screening criterion. Then, the system parameters are optimized in parallel by particle swarm optimization to achieve the best stochastic resonance result. The numerical simulation and experimental research results show that this method can improve the output power spectrum amplitude of the vortex signal with small flow, effectively obtain the vortex frequency, realize the measurement of small flow, and enhance the detection ability of weak signal.

Keywords. vortex flowmeter; signal processing; stochastic resonance; ensemble empirical mode decomposition

1. Introduction

Vortex flowmeter is a vibrating flowmeter, and its frequency is proportional to the flow rate. However, various vibrations in the industrial site make the vortex signal mixed with noise, especially in the measurement of small flow, the vortex signal is weak and easily overwhelmed by noise, resulting in limited measurement [1-2]. Scholars at home and abroad have done a lot of research on vortex signal processing methods, such as classical spectrum analysis method, wavelet transform method, adaptive filtering method, etc. [3-5]. These methods mostly weaken the noise extraction signal through the characteristic difference between the signal and the noise, which will inevitably damage the useful signal in the processing process, and have certain limitations.

Stochastic resonance(SR) is a nonlinear processing method. Under the combined action of nonlinear bistable system, signal and noise, noise energy is used to enhance weak signals,reflecting the influence of random motion of microscopic particles on macroscopic variables[6]. For the situation that the vortex signal is easily disturbed by

[1] Corresponding author,Jianxiu LIU, Zhengzhou University of Light Industry, Zhengzhou, China, E-mail: jinxiuliu@zzuli.edu.cn.

the noise in the field, the stochastic resonance method can improve the measurement accuracy and expand the lower limit of the measurement. The concept of SR was proposed by Benzi's team [7] to explain the periodic occurrence of ancient meteorological glaciers; in 1989, McNamara et al.[8] gave a bistable two-state stochastic resonance based on the assumption of adiabatic approximate small parameters mathematical representation of the model.According to the ability of SR noise to enhance the output signal and the characteristics of vortex signal, this paper proposes a classic bistable stochastic resonance (CBSR) method based on EEMD to extract weak vortex signal. This method makes full use of the signal decomposition performance of EEMD and the signal enhancement ability of stochastic resonance to achieve the purpose of enhancing the fault characteristics of the low frequency band of the FFT spectrum.First, The signal is decomposed by EEMD, and the normalized correlation coefficient is used as the evaluation standard, and the main IMF components are selected for signal reconstruction. Then, the system parameters are optimized and selected by particle swarm algorithm (PSO), and input the reconstructed signal after preliminary noise reduction into CBSR for signal feature enhancement. Finally, the output signal-to-noise ratio(SNR) is used to evaluate the system enhancement effect. To achieve the optimal output of nonlinear systems, enhance weak periodic signals, and improve the system's ability to extract vortex signals.

2. Adaptive Stochastic Resonance System Based on EEMD

2.1. Ensemble Empirical Mode Decomposition(EEMD)

Ensemble empirical mode decomposition (EEMD) is a multiple-time empirical mode decomposition of superimposed white Gaussian noise, and is a noise-assisted analysis method. The EEMD method is essentially an improvement to the EMD algorithm. It is mainly based on the characteristic that the mean value of white noise is zero. White noise is added to the signal, and EMD is still used to decompose it. The decomposed results are averaged. The higher the number of times, the smaller the influence of noise on the decomposition result. Using EEMD to decompose the signal, the steps to achieve preliminary filtering and noise reduction are as follows:

(1) Add normally distributed white noise to the vortex signal to form a series of new signals

(2) The signal added with white noise is taken as a whole, and then EMD is decomposed to obtain IMFs

(3) Average the IMFs of the corresponding modes to obtain the EEMD decomposition result

(4) Use the normalized correlation coefficient to filter the effective IMF corresponding summation, and obtain the signal after preliminary noise reduction

2.2. Classical Bistable Stochastic Resonance Model

SR can enhance weak signal features with appropriate noise through nonlinear dynamical systems. The classical bistable model is the most widely used nonlinear model. Its potential function expression is:

$$U(x) = -ax^2/2 + bx^4/4 \quad a>0, b>0 \tag{1}$$

where a and b are system parameters. Fig.1 shows the bistable potential well diagram with a=1, b=1. It can be seen from Fig.1 that the potential function has two steady-state points $x_{1,2} = \pm\sqrt{a/b}$ and one non-steady-state point x=0, and the potential barrier height is $\Delta U = a^2/4b$ .By changing the values of a and b, bistable potential well diagrams with different shapes are obtained as shown in Fig.2 The width of the potential well and the height of the potential barrier vary with the value of the system parameters.

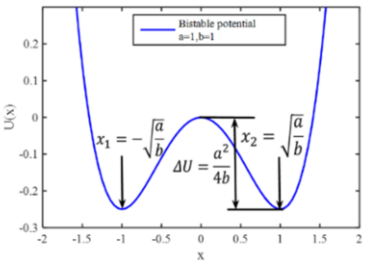


Figure 1. Bistable potential in SR

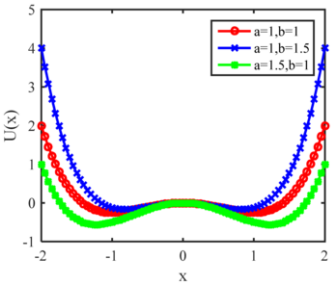


Figure 2. Bistable potential with different parameters

The occurrence process of stochastic resonance is usually expressed by the Langevin equation:

$$\frac{dx}{dt} = -U'(x) + A\cos(2\pi f_0 t + \varphi) + N(t) \tag{2}$$

where N(t) is additive white Gaussian noise.

SNR reflects the enhancement effect of the nonlinear system on the signal. The calculation formula of the output SNR based on the adiabatic approximation theory is:

$$SNR = \frac{\sqrt{2}a^2A^2}{4bD}\exp\left(-\frac{a^2}{4bD}\right) = \sqrt{2}\Delta U\left(\frac{A}{D}\right)^2 e^{-\Delta U/D} \tag{3}$$

With the increase of noise intensity, the output SNR firstly increases sharply, then peaks and then decreases continuously. When the output SNR reaches a peak value, the system is considered to have the best stochastic resonance.

2.3. Numerical Calculation Method

The output of the SR system needs to be solved by a numerical solution algorithm, because the approximate analysis or analysis is no longer applicable to the output of the nonlinear system, or the solution is extremely complicated, so the differential equation is usually converted into a difference equation to solve. In this paper, the Runge-Kutta algorithm is a numerical method for solving differential equations. The main idea of this method is to interpolate in the selected integral interval, optimize the slope continuously, and obtain updated results. The overall expression of the Runge-Kutta algorithm is shownin equation 4.

$$\begin{cases} x_{n+1} = x_n + \frac{1}{6}(k_1 + k_2 + k_3 + k_4) \\ k_1 = h(ax_n - bx_n^3 + u_n) \\ k_2 = h[a(x_n + \frac{k_1}{2}) - b(x_n + \frac{k_1}{2})^3 + u_n] \\ k_3 = h[a(x_n + \frac{k_2}{2}) - b(x_n + \frac{k_2}{2})^3 + u_{n+1}] \\ k_4 = h[a(x_n + \frac{k_3}{2}) - b(x_n + \frac{k_3}{2})^3 + u_{n+1}] \end{cases} \tag{4}$$

$x(n)$ represents the nth output, $u(n)$ represents the nth input item $s(t)$, and h is the iteration step size.

2.4. Selection of Control Parameters

Existing SR methods mainly focus on choosing appropriate optimization parameters and fitting functions. Since the noise level of a machine fault signal is difficult to estimate because the source signal is unknown and only the received signal can be measured. The vibration has to go through a complex path before reaching the sensor, and noise from different sources will be introduced into the collected signal. Therefore, noise is indeed contained in the acquired signal, but the details of the noise such as noise type, noise distribution and noise level are difficult to determine. Therefore, in adaptive SR methods designed for practical signals, system parameters are widely used as tuning objects.SR is a parameterized system, and the selection of optimal system parameters has a significant impact on the detection accuracy. The output SNR is used as the system evaluation index, The system parameters are optimized and selected by PSO. PSO is an iterative-based optimization algorithm. The system is initialized as a set of random solutions, and the optimal value is searched through iteration. The advantage of PSO is that it is simple and easy to implement, and there are not many parameters to adjust.

3. Simulation Results of Bistable Stochastic Resonance System

In order to verify the effectiveness and practicability of the method, MATLAB is used for signal simulation and analysis to verify the effectiveness of the method. When the input signal is a low-frequency signal, set the input signal amplitude A = 0.1, frequency f_0 = 0.01Hz, noise intensity D = 2, sampling frequency f_s =5 Hz, the number of sampling points N=10240, the time domain and frequency domain diagrams of the output signal are shown in the figure.

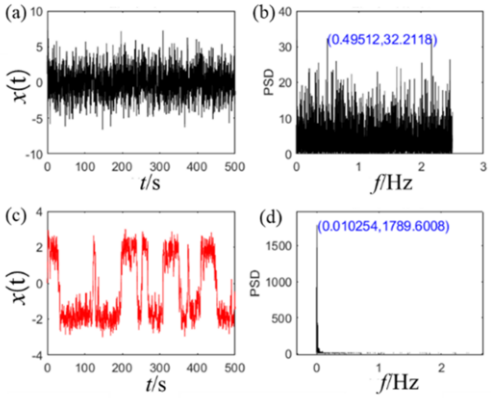


Figure 3. Time-domain waveform and power spectrum of the Simulation signal

Fig.3(a) shows the graph of noisy signal after adding Gaussian white noise to the low-frequency characteristic signal. Because the background noise is too large, the overall time-domain signal is cluttered, irregular, and loses periodicity; after FFT transformation, the graph The frequency of the target signal cannot be directly identified on the power spectrum shown in Fig.3(b); the time domain diagram of the signal output after being processed by the bistable system is shown in Fig.3(c). The power spectrum of the output signal is shown in Fig.3(d), the noise energy is significantly weakened in the frequency domain, and the useful input signal is enhanced. The target frequency f_0 = 0.01Hz has the highest amplitude, which is 1789.6. Simulation results confirm that the method can identify the vortex signal frequency submerged by noise, and is suitable for weak signal detection under strong noise background.

4. Vortex Signal Experiment Results

The vortex flow acquisition device adopts a vortex flowmeter with an accuracy of 0.5 and a diameter of 50mm, and collects a vortex signal with a flow rate of 6.48m^3/s for analysis. The corresponding frequency theoretical value under this flow rate is 16.66Hz. The smaller the flow, the smaller the lift generated by the vortex, and the weaker the periodic signal of the vortex street, so it is easy to be submerged in the site noise. Figures Fig.4(a) and Fig.4(b) are the time domain diagram and power spectrum of the vortex signal. The vortex signal is submerged in noise, and the characteristic frequency of the vortex signal cannot be directly extracted from the power spectrum of the signal. After the vortex signal is processed by the stochastic resonance system, there is a significantly larger power spectrum peak in Fig.4(d), and the corresponding frequency f=16.668Hz, which is the characteristic frequency, indicating that the bistable stochastic resonance control method based on EEMD can be effective to detect the characteristic frequencies of weak vortex signals.

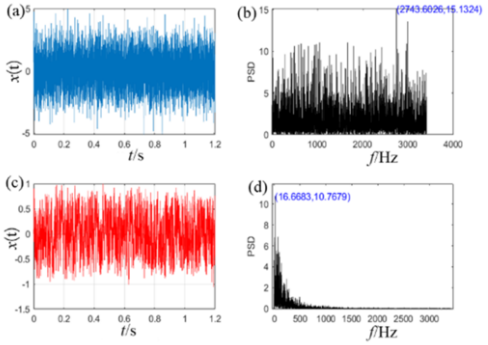


Figure 4. Time domain diagram and power spectrum diagram of vortex signal

5. Conclusion

Aiming at the characteristics that the small flow vortex signal is easily submerged by noise, CBSR system based on EEMD is proposed in this paper to improves the power spectrum at the characteristic frequency of the small flow vortex signal. The theoretical and experimental results show that this method can effectively extract the vortex features. This method is of great significance for expanding the application range of stochastic resonance to solve practical engineering problems.

References

[1] Xunjie T, Min L, Yongmei H. The enhancement of noise nonlinear effect and the detection method of weak vortex street signal [J]. Vibration and Shock, 2016, 35(15): p. 5.

[2] Junhong D, Yongmei H,Min L. Stochastic resonance characteristics and frequency detection method of small flow vortex signals [J]. Instrument Technology and Sensors, 2016(5): p. 4.

[3] Guiji T, Nannan L, Xiaolong W. A gear fault feature extraction method with comprehensive improvement of singular spectrum decomposition and singular value decomposition [J]. China Mechanical Engineering, 2020, 31(24): pp. 2988-2996.

[4] Zhixing L, Boqiang S. Extraction of Weak Fault Features by Stochastic Resonance Based on Adaptive Singular Value Decomposition [J]. Chinese Journal of Agricultural Engineering, 2017, 033(011): pp. 60-67.

[5] Lifang H, Yingying C, Tianqi Zh, et al. Fault Signal Detection Method Based on Power Function Bistable Stochastic Resonance [J]. Chinese Journal of Instrumentation, 2016, 37(7): p. 11.

[6] Shan W, Pingjuan N, Yongfeng G, et al. Bearing fault diagnosis based on adaptive segmented hybrid system [J]. Journal of Aerodynamics, 2021, 36(10): p. 11.

[7] Roberto‐B, Giorgio P, Alfonso‐S. A‐Theory of Stochastic‐Resonance‐in Climatic Change[J]. SIAMJournal_on_Applied Mathematics, 1983, 43(3).

[8] Mcnamara B, Wiesenfeld K .Theory of stochastic resonance[J]. Phys Rev A Gen Phys, 1989, 39(9): pp. 4854-4869.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 115
J. Xu et al. (Eds.)

doi:10.3233/ATDE220218

Optical Lens Anti-High Overload Design Technology

Yaxiong TAN[a,c,1], Xin ZHENG[b], Jun TANG[a,c], Yujiao JIA[a,c] and Ji DING[a,c]
[a] *Beijing Aerospace Automatic Control Institute*
[b] *Harbin Xinguang Optic-Electronics Technology Co., Ltd.*
[c] *National Key Laboratory of Science and Technology on Aerospace Intelligence Control*

Abstract. In view of the problem that the optical lens is difficult to withstand the impact of high overload, on the basis of the optical imaging function, measures in many aspects such as selecting the appropriate lens material, designing the appropriate diameter-to-thickness ratio and designing the colloidal vibration damping structure are adopted to improve the optical lens anti-high overload capability. The ANSYS simulation analysis and hammering test verify that the optical lens designed by this method can withstand overloads above 20,000g, and the optical imaging ability does not change before and after the hammering.

Keywords. High overload, optical lens, structural design

1. Introduction

With the development and application of information technology, a number of new types of ammunition such as TV reconnaissance shells, laser terminal guided shells, and TV terminal guided shells have emerged. The seeker contains various photoelectric sensors and microelectronic components. During the launch of the artillery, the shell will be subjected to the pressure in the chamber, and the projectile body will produce a large acceleration, making the seeker to withstand huge instantaneous, high-energy and strong impact load [1]. With the development of shells, the requirements for the range of the shells are getting higher and higher, so the firing overload that the shells need to bear is also increasing [2], This high overload environment will cause deformation of the internal support structure of the seeker and damage to various electronic components, and even cause the guidance system to fail[1]. In the face of high overloads of 10,000g or even more than 20,000g, how to withstand the optical lens is the most critical and difficult one [3]. This paper proposes a high-overload resistance design scheme for optical lens, which can meet the requirements for use in an overload environment of more than 20,000 magnitude after simulation analysis and experimental verification.

[1] Yaxiong TAN, Beijing Aerospace Automatic Control Institute, 50 Yongding Road, Haidian District, Beijing, China; E-mail: 1205145876@qq.com.

2. Optical lens anti-high overload design

2.1. Material selection

Since the system has certain requirements on the related parameters of mechanical mechanics of the lens, in addition to the optical properties of the lens such as refractive index, absorptivity, thermal deformation and other parameters, the mechanical related properties of the lens such as stability, impact resistance, hardness and weight, should also be considered when selecting the lens material.

In combination with the above material selection criteria, the lens materials to be used in this project are lanthanum crown glass, heavy flint glass and lanthanum flint glass, because these lens materials have good mechanical stability, suitable weight, and can prevent the lens from chipping due to its own material in the case of high overload or large impact.

The optical lens has a simple structure and small volume, but the mechanical and thermal environment is harsh. The optical lens needs to withstand an overload of more than ten thousand orders under working conditions. Therefore, considering the mass-strength ratio, the main structure material chooses low-density and high-stiffness super-hard aluminum alloy material 7A09 for aviation.

2.2. Anti-high overload structure design

The structure and fixation of the lens are the core of whether the lens can withstand high overload. In the optical design, the curvature of the lens should be as large as possible to improve its anti-overload ability; Comprehensive imaging design and lens strength simulation, the diameter-thickness ratio of the lens is not more than 3 to prevent the lens being unable to bear its own weight and broken under high impact.

The optical lens structure adopts optical centering and edge-taking process design. By adding a layer of mechanical frame to the lens, the outer diameter of each lens is consistent and the installation tolerance requirements are met through centering and edge removal processing. At the same time, because the lens is equipped with a mechanical frame, there is no direct contact between the lens and the lens, the lens depends on the positioning of the mechanical frame. Due to the high strength and processing accuracy of the mechanical metal material, the problem of positioning and force transmission between the lenses is solved. The boss for positioning the lens of each edge removal frame is on the side opposite to the acceleration direction, that is, the bearing side, which ensures that the lens does not fall out under the structural limit of the edge removal frame under working conditions. The appearance of the optical lens structure is shown in **Figure 1**.

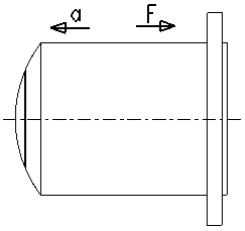


Figure 1. Schematic diagram of optical lens structure design.

The components of the optical system in this optical lens are installed in order from left to right, and the positioning of the first lens is ensured during installation, and then the rear lens is installed in sequence, and glue is applied to the gap between the lens and the frame, and finally the pressure ring is tightened.

3. Finite element analysis and test verification

The optical lens working normally in a high-overload environment must have sufficient strength and rigidity, so that the lens will not be broken and the relative positional relationship between the lenses will not change. From these two aspects, perform finite element analysis [4-6] and experimental verification on the optical lens.

3.1. Finite element analysis

In order to verify the strength of the optical lens in a high overload environment, a finite element analysis was performed on the optical lens[7].

3.1.1. Force model

The optical system accelerates from the stationary state and shoots out along the rifling of the barrel. At this time, the projectile is in a linear acceleration state, and the projectile is subjected to a force F opposite to the direction of acceleration a. The kinetic energy of the projectile gradually decreases after it exits the chamber. The force of the optical lens is shown in **Figure 2**.

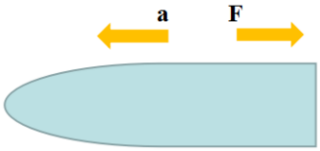


Figure 2. Optical lens force diagram.

3.1.2. Material parameters

The material parameters used in the structural finite element analysis are shown in **Table 1**.

Table 1. Material parameters in analysis

Material name	Density (kg/m^3)	Young's modulus (GPa)	Poisson's ratio	Thermal expansion coefficient 1/℃×10^{-6}
H-K9L	2520	79.2	0.211	7.6
H-ZF88	3520	108.81	0.243	6.0
H-FK61	3700	70.07	0.3	13.1
H-ZF4	4510	52.95	0.245	9.6
H-ZF4	4510	52.95	0.245	9.6
H-ZF88	3520	108.81	0.243	6.0
Aluminum alloy	2770	71	0.33	23
PTFE	2170	0.47	0.45	120

3.1.3. Finite element analysis results

The lens, lens holder, and threaded connection are set according to Frictional; The friction coefficient is set to 0.2; Restrict the flange connection of the shell. The overall stress distribution cloud diagram of the structure is shown in **Figure 3**.

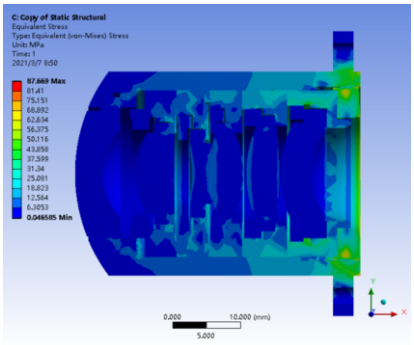


Figure 3. Lens stress cloud

The stress distribution cloud diagram of each lens is shown in **Figure 4.**

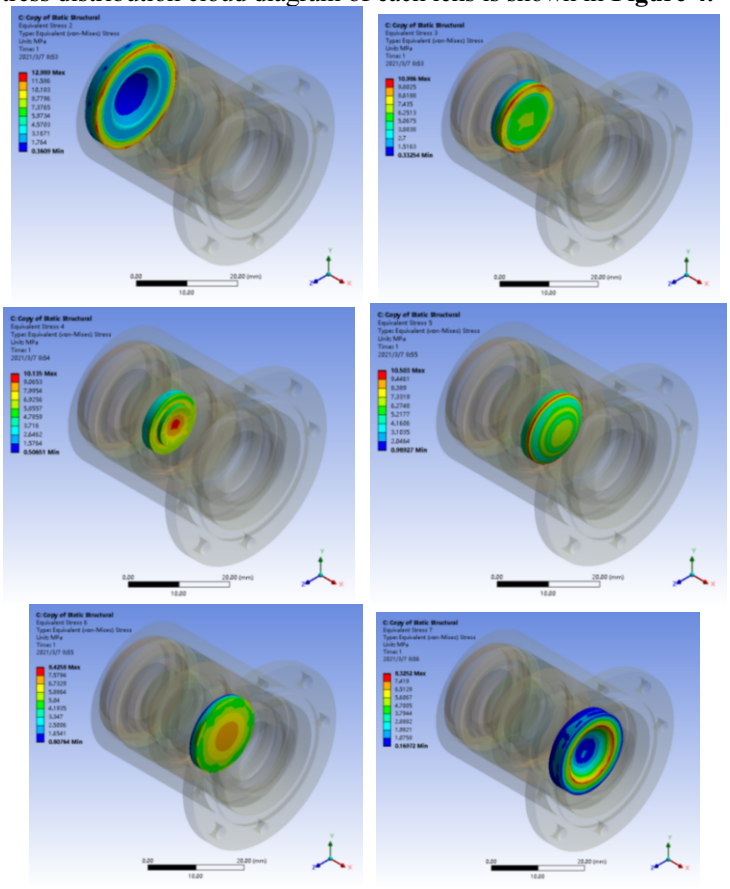


Figure 4. Stress cloud diagram of each lens

The stress cloud diagram of the threaded connection is shown in **Figure 5**.

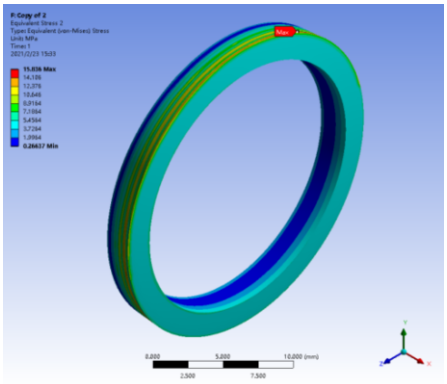


Figure 5. Stress cloud diagram at threaded connection

The analysis results show that under the target overload condition, the maximum stress position is at the connecting flange on the lens base, the size is 87.669Mpa, and the maximum stress at the threaded connection is 15.836Mpa.The maximum stress of each lens is shown in **Table 2**.

Table 2. Maximum lens stress table

	Lens 1	**Lens 2**	**Lens 3**	**Lens 4**	**Lens 5**	**Lens 6**
Stress (MPa)	12.989	10.986	10.135	10.503	8.4258	8.3252

In summary, the stress of each part of the optical lens is much smaller than the allowable stress of the material, and the analysis result meets the requirements of use.

3.2. Hammer test

A Marshall hammer test system[8] was used to conduct a high overload resistance test on the optical lens, and the test magnitude was close to and greater than the target overload magnitude.The same scene was imaged before and after the test, and the imaging comparison is shown in **Figure 6**.

Figure 6. Imaging comparison chart

It can be seen from the figure that after three hammering tests, the optical lens is intact, and the imaging effect is basically unchanged before and after the test.

4. Conclusion

This article adopts measures such as selecting the appropriate lens material, designing the appropriate diameter-to-thickness ratio, and designing the jelly damping structure to improve the high overload resistance of the optical lens.The ANSYS simulation analysis and hammering test verified that the optical lens designed by this method can withstand overloads above 20,000g, and the optical imaging ability does not change before and after hammering, which can meet the requirements of high overload environment.

References

[1] Mingdong SH, Junli Q, Mengdi Y, Xiaojing H, Taihui M, Yuhui CH. Anti-high overload technology and research progress of missile-borne seeker. Journal of Ordnance Equipmeng Engineering. 2019 July; 40(7): pp 85-89.
[2] Lizhi Q. Study of projectile-loaded equipment against high overload. Acta Armamentarii. 2007 Aug; 28(8): pp 1017-1020.
[3] Junwei L, Limin ZH, Xiaokai ZH. Study on laser-guided projectile seeker against high overload. Journal of Zheng Zhou University of Light Industry. 2014 Jun; 29(3): pp 65-67.
[4] Yuan L, Kai L, Xiaofei W, Yan H. Simulation analysis on vibration and high overload impact of missile guidance system. Journal of North University of China. 2014; 35(3): pp 293-297.
[5] Pan WT. A new fruit fly optimization algorithm: Taking the financial distress model as an example. Knowledge-Based Systems. 2012; 26: pp 69-74.
[6] An P, Zheng Y, Yan S. High-Q microsphere resonators for angular velocity sensing in gyroscopes. Applied Physics Letters. 2015; 106(6): p 327.
[7] Jian ZH, Xinyu W, Hong Y, Changcheng X, Jiangtao D. Research on Anti-High Overload Technology of Missile-Borne Attitude Measurement System. Piezoelectrics and Acoustooptics. 2021 Apr; 43(2): pp 270-273.
[8] Wen ZH. Protection Analysis and Technology Research of High Overload Receiver Potting Layer. Modern Manufacturing Technology and Equipment. 2021; 2021(3): pp 138-141.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 121
J. Xu et al. (Eds.)

doi:10.3233/ATDE220219

Vibration Control of Magnetorheological Semi-Active Seat Based on Reinforcement Learning

Yuxia LI[a,1] and Yuxuan LIANG[a]
[a] *State Key Laboratory of Mechanical Transmission, Chongqing University*

Abstract. The study aims to improve the vibration isolation effect of the semi-active seat suspension system using a reinforcement learning control scheme. Mechanical tests on the designed magnetorheological damper to establish the mechanical model of the damper. According to the parameters of the seat suspension, the modeling of the magnetorheological semi-active seat suspension system is derived, and a control scheme based on reinforcement learning is proposed. Finally, the performance of the proposed control strategy is compared with the passive system and the sky-hook control under random single. It is shown that the RMS value of acceleration of the seat suspension is reduced by 21.2% compared to the sky-hook control, and by 10.6% compared with the passive system.

Keywords. Seat suspension, reinforcement learning, magnetorheological damper

1. Introduction

Construction machinery vehicles often drive on non-structured roads. The long-term work of the driver in this high-intensity vibration environment will cause very serious damage to the driver's health. Therefore, the research on seat vibration reduction is of great significance. Helps improve driver comfort and operability of the vehicle[1-2].

The selection of the control algorithm can effectively improve the damping performance of the semi-active suspension. At present, the main control algorithms mainly include sky-hook control[3], PID control[4], fuzzy control[5], adaptive control[6] and so on. These algorithms usually require precise dynamic models to achieve control, while fuzzy control requires a lot of control experience. Compared with these algorithms, reinforcement learning can realize a self-learning process without model and experience samples.

2. Modeling of magnetorheological damper

Due to the magnetorheological damper has strong nonlinear characteristics such as hysteresis, yield and saturation, the error of the modeling using mathematical formula

[1] Corresponding Author, Yuxia LI, College of Mechanical and Vehicle Engineering, Chongqing University, Shapingba District, Chongqing 400044, China; E-mail: liyuxia1997@gmail.com.

is large and the adaptability of the modeling is weak. Secondly, the parameter of the model is particularly complicated, so it is difficulty to solve the model and the error of the model is large. The shortcomings that occur when using parametric modeling can be avoided by using non-parametric modeling. Neural network can approximate any nonlinear function after training with a large amount of data, this study uses the neural network to establish the dynamic model of the magnetorheological damper.

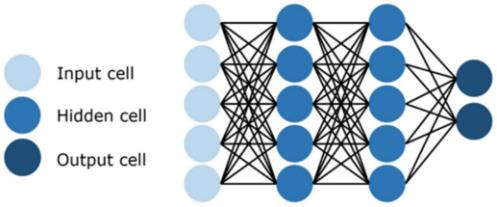
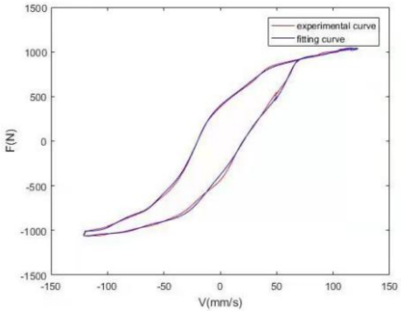
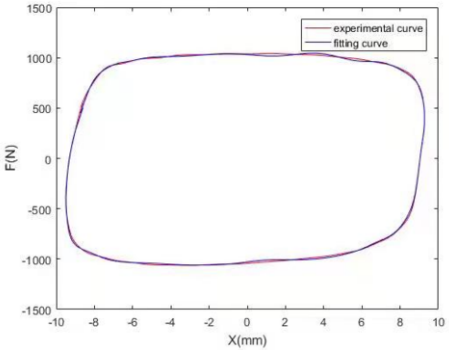


Figure 1. Neural network structure.

A common neural network structure is shown in Figure 1. The input layer has 3 nodes, namely displacement, velocity, and current; the output layer has a node, which is the damping force; the number of hidden layers is 20, the training algorithm adopts Bayesian-Regularization, and the training samples are collected in the experiment, the number of current training data samples in each group is 5000 groups, and the number of training iterations is 1000. The modeling result of the model of the damper is shown in Figure 2. The predicted value of the model is basically consistent with the test data, the damper model has good results.



a. F-V curve with current of 2.0A



b. F-X curve with current of 2.0A

Figure 2. Fitting results of the damper modeling.

The modeling result of the damper model is shown in Figure 2. The neural network can fit the experimental curve well, and the curve fitting accuracy is high.

3. Modeling of semi-active seat suspension

The seat suspension model shown in Figure 3 is established, and the dynamic modeling of the seat suspension system can be described as:

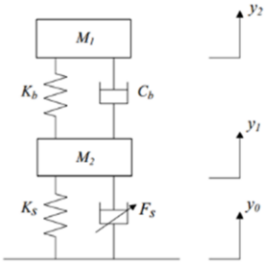


Figure 3. The modeling of the seat suspension system.

$$\begin{cases} M_2\ddot{y}_1 = -F_s - K_s(y_1 - y_0) - C_b(\dot{y}_1 - \dot{y}_2) - K_b(y_1 - y_2) \\ M_1\ddot{y}_2 = C_b(\dot{y}_1 - \dot{y}_2) - K_b(y_2 - y_1) \end{cases} \tag{1}$$

where M_1 is the seat suspension bearing mass, M_2 is the seat bracket mass, K_s, K_b,and $C_b$represent suspension spring stiffness, stiffness and damping of cushion, F_s is damping force, magnetorheological damper controller and other parts. Among them, for the sake of simplicity, the seat body mass and the driver mass are regarded as a whole, that is, the seat suspension bearing mass M_1 is the sum of the seat body and the driver. The above-mentioned seat suspension model parameters are shown in table 1.

Table 1. Parameters of seat suspension model

Parameters	M_1	M_2	K_s	K_b	C_b
Value	70kg	8kg	7.3N/mm	31.5N/mm	0.95Ns/mm

4. Control algorithm

As a semi-supervised learning algorithm, reinforcement learning is different from the supervised learning algorithm. this kind of algorithm does not need a training set in advance, but obtains sample data through multiple interactions between the agent and the environment, and completes the self-learning process. Initially, the agent don't have any information on which best measure to take in each environmental state, but the agent repeat the process that the agent take actions in the environment and obtain rewards. At the same time, the environment updates the state and transmits the new state to the agent. By repeating this process, the agent will continue to learn based on the experience, and finally obtain the optimal strategy.

Markov Decision Process (MDP) is the memoryless process that decision makers obtain states from dynamic random systems with Markov properties and make decisions

continuously or periodically. Tuples (S, A, P, R, γ) are usually used to describe MDP, where:

- S represents the state collection of agents in the environment;
- A represents the action set of the agent, all the actions that the agent can take;
- P indicates the probability of state transition, and the action is performed under the state s_t to the next state s_{t+1}. For each state and action, P_{sa} is a distribution on the state space;
- R represents the reward function, the reward and punishment that the agent receives after performing the action at time t is $r_{t+1} = R(s_t, a_t, s_{t+1})$;
- γ Represents the discount factor, used to calculate the cumulative return, $\gamma \epsilon [0,1]$

The dynamics of the Markov decision process is as follows:

Start at a certain initial state s_0, then select a certain action $a_0 \epsilon A$ to execute the MDP process. According to the selected action, there will be a corresponding result, and the state of the MDP will transition to a subsequent state s_1 which is obtained according to $s_1 \sim P_{s_0 a_0}$. Then select another action a_1, and there will be a state transition corresponding to this action, the state is $s_2 \sim P_{s_1 a_1}$, and then another action a_2 is selected, and so on. The following process can be used as a representation:

$$s_0 \overset{a_0}{\rightarrow} s_1 \overset{a_1}{\rightarrow} s_2 \overset{a_2}{\rightarrow} s_3 \overset{a_3}{\rightarrow} \cdots \tag{2}$$

Through the state and action in the process, the reward value in the whole process can be obtained, the overall benefit function is

$$R(s_0, a_0) + \gamma R(s_1, a_1) + \gamma^2 R(s_2, a_2) + \cdots \tag{3}$$

If the reward function is only a function related to the state, then this value can be described as:

$$R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + \cdots \tag{4}$$

The goal of reinforcement learning is to find a set of actions that maximize the expected value of the overall return function that can be written as:

$$maxE[R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + \cdots] \tag{5}$$

5. Controller design

In order to achieve the optimization of vehicle ride comfort, the vehicle body vertical acceleration and vertical displacement are the optimization objects. In the DDPG algorithm, the output action is directly related to the state of the system, so the input of the state of the network must have a strong correlation with the action. This study chooses sprung displacement, sprung speed, unsprung displacement and unsprung velocity to be used as the state input, $S = \{y_1, \dot{y}_1, y_2, \dot{y}_2\}$.

Choosing control current of the damper as the control action of the algorithm. The DDPG algorithm directly outputs the control current from the Actor network, and the output action range covers all feasible ranges of the control current, $I\epsilon[0,2]$.

In order to optimize the ride comfort of the vehicle, the study selects the vertical acceleration and vertical displacement of the vehicle as part of the reward and punishment function. Similar to the objective function of the LQR algorithm based on the optimal control theory, this study also converts the vertical acceleration and vertical displacement into quadratic forms as the reward and punishment function of the DDPG algorithm, which can be written as:

$$R(\ddot{z}_s, z_s, u) = \lambda_1 {\ddot{z}_s}^2 + \lambda_2 {z_s}^2 \tag{6}$$

where λ_1, λ_2 and λ_3 are weights coefficients of the three optimization goals.

The DDPG algorithm has an independent Actor strategy network dedicated to output actions and adopts a deterministic strategy. It adds Gaussian noise with time attenuation properties to the action, as:

$$a = \mu(s \mid \theta^{\mu}) + N \tag{7}$$

where μ represents the current Actor policy network, θ^{μ} represents the weight of the policy network, and N represents Gaussian noise.

6. Results and discussions

In order to verify the validity of the algorithm, the performance of the proposed control strategy contrasts with the passive system and the sky-hook control under random single.

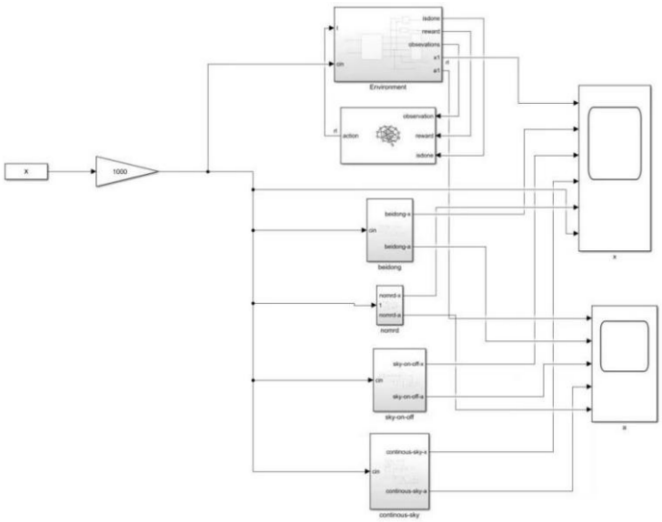


Figure 4. Algorithm simulation block diagram.

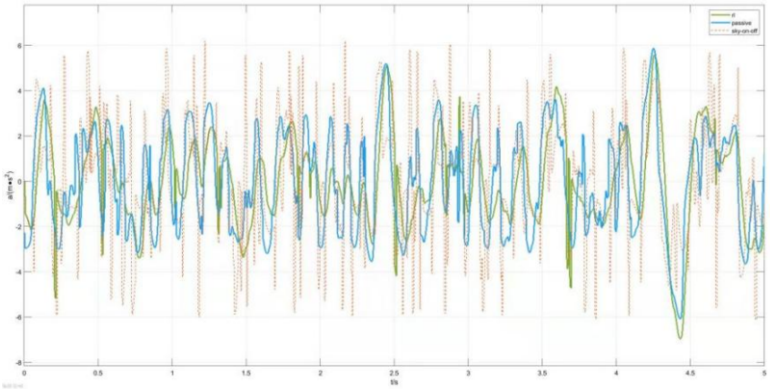


Figure 5. Human body acceleration under random signal.

It can be seen from the figure 5 that DDPG can greatly improve the ride comfort of the vehicle under random road conditions. The simulation results show that the semi-active seat suspension system under the reinforcement learning control can adapt to different conditions.

Table 2. RMS value of three control modes under random signal.

Control type	RMS of Acceleration/(m/s^{-2})
passive system	2.3088
sky-hook control	2.6164
reinforcement learning control	2.0628

In this study, firstly, the designed magnetorheological damper is tested mechanically, then the mechanical model of the damper was established. The modeling of the magnetorheological semi-active seat suspension system is derived, and a control scheme based on reinforcement learning is proposed. Finally, the effectiveness of the proposed control strategy contrasts with the passive system and sky-hook control. The results show that the proposed reinforcement learning control strategy has relatively good results. The RMS value of acceleration of the seat suspension is reduced by 21.2% compared to the sky-hook control, and by 10.6% compared with the passive system.

Acknowledgement

This work was financially supported by graduate scientific research and innovation foundation of Chongqing, China (Grant No. CYB19009, No. CYB21012). These supports are gratefully acknowledged.

References

[1] Ramalingam M, Raju P, Arul Thirumurugan M, et al. Control policies used for semi-active for automotive seating system: a review. International Journal of Dynamics and Control. 2018; 24(6): pp 268-282.

[2] Haiping D, Weihua L, Zhang N. Semi-active control of an integrated full-car suspension with seat suspension and driver body model using ER dampers. International Journal of Vehicle Design. 2013; 63(2/3): pp 159-171.

[3] Karnopp D. Vibration control using semi-active force generators. Transactions of the Asme Journal of Engineering for Industry. 1974; 96(2): pp 619-626.
[4] Jie M, Haipeng Y, Qingzhang C, et al. Simulation research of the PID controller of vehicle suspension based on the genetic algorithm. Modern Manufacturing Engineering. 2013(6): pp 92-96.
[5] Anand Raj R, Shrivastava S, Trikande M.W. Modelling and analysis of skyhook and Fuzzy logic controls in semi-active suspension system. International conference on Industrial Instrumentation and Control. Pune: ICIC; 2015.
[6] Yechen Q, Mingming D, Feng Z, et al. Suspension semi-active control of vehicles based on road profile classification. Journal of Northeastern University(Natural Science). 2016; 37(8): pp 1138-1143.

doi:10.3233/ATDE220220

3D Printing of Smart Materials and Actuators

Yixian WANG[a], Bingsen JIA[a], Sen Liu [a], Xinle YAO [a] and Chufeng SUN [a,b,1]
[a] *Key Laboratory of Environment-Friendly Composite Materials of the State Ethnic Affairs Commission, School of Chemical Engineering, Northwest Minzu University, Lanzhou, 730030, China*
[b] *Gansu Provincial Biomass Function Composites Engineering Research Center, Lanzhou, 730030, China*

Abstract. Smart actuators can sense external stimuli and produce controllable mechanical responses, and convert these energies into mechanical energy. They have great applications in the aerospace, electronic circuits, medical and other fields. As a new manufacturing method, the combination of 3D printing and smart actuators had developed rapidly in recent years. In this paper, we summarize the research progress of 3D printing smart actuators and its materials. The smart driver includes water responsive driver, pH responsive driver, temperature responsive driver, light responsive driver and magnetic field responsive driver. The smart driver materials can be divided into shape memory materials, piezoelectric materials, responsive smart hydrogels and electroactive polymers. In addition, their stimulative effect and driving mechanism have been studied emphatically.

Keywords. 3D printing; Smart actuators; Drive materials

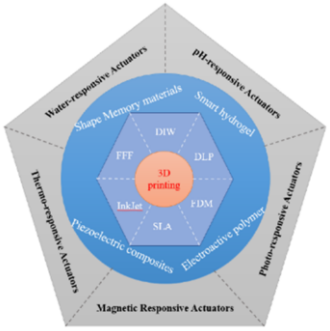


Figure 1. Schematic overview of 3D printing

1. Introduction

Smart actuators can be stimulated by a certain field to convert all kinds of energy into mechanical energy, and then the material undergoes reversible deformation and

[1] Corresponding Author, Chufeng Sun, School of Chemical Engineering, Northwest Minzu University, Lanzhou, 730030, China; E-mail: cfsun@licp.cas.cn.

movement [1-6]. Over recent years, the applications of smart actuators in biomimetic robots, bionic robots and sensors. etc. had attracted much attention [7]. However, most smart actuators were processed by traditional methods, such as molding and soft lithography[8]. Although these machining methods have high machining accuracy, but the two-dimensional plane processing ability and machining efficiency are low, so that the materials are only suitable for simple and small-sized structure. With the increasing complexity of smart driving mechanisms, the above-mentioned traditional processing method can no longer meet the current requirements. According to the material morphology and molding principle, it is mainly divided into three categories: extrusion molding, granular material molding and photopolymerization synthesis. Based on this idea that 3D printing can quickly prototype of the target structures, it can avoid the tedious manufacturing process of traditional processing method.

The smart actuators with 3D printing complex structure can realize directional movement through stimulation in special environment. Recently 3D printed smart actuators have great potential value in drug transportation, medical equipment, micro devices, soft robots and so on.

The smart actuator is an actuator, which is an essential part of an automatic control system, and can respond to changes in external stimuli. Thus, it has received wide concern of researchers aiming to enhance their efficiency and agility. Smart actuators include water-responsive actuators, pH-responsive actuators, Thermo-responsive actuators, photo-responsive actuators, magnetic responsive actuators and stimuli responsive actuators.

Smart materials include carbon-based materials, liquid crystalline polymers, biological materials, smart hydrogels, dielectric elastomers, ionic polymer-metal composite. These smart materials respond to external stimuli including water or humidity, pH, temperature, magnetic and electrical fields.

In this study, we present the latest progress in the field of 3D printing smart actuators and their materials. We provide a toolbox for interdisciplinary audiences who are interested in the latest 3D printed smart actuator technology. In particular, we mainly explored the driving methods of 3D printing smart actuators, focusing on the form of motion and its driving effects. The first section discusses the actuating methods of 3D printing smart actuators and explore their functions and the realization of researchers. The materials for 3D printing smart actuators are then discussed.

2. Smart actuators for 3D printing

As one of the main components of robot system, smart actuator has been widely paid attentions by researchers in order to improve its efficiency and flexibility. Recently, researchers have focused on design with smart functions and advanced manufacturing techniques.

Plentiful examples of changes in biological systems can make changes according to their environment in nature. The wide range of stimulus response and different functions provide a good foundation for biological driving system to become smart actuators. In addition, researchers have invested many work into the development of smart actuators. Research in driving mechanism, material selection, 3D printing and creative design realization confirmed the potential of this field[9]. In view of people's interest in 3D printing smart drives, this paper reviews the latest research on smart drives and their biological driving applications, and classifies them according to their

driving mechanisms. Here provided the stimulus is water, pH, thermal, optical, magnetic, electric, and the combination of these. These stimulate diagram is shown in figure 2. This paper reviews the latest research on intelligent drives and their biological drive applications, and classifies them according to their driving mechanisms in view of people's interest in 3D printing intelligent drives.

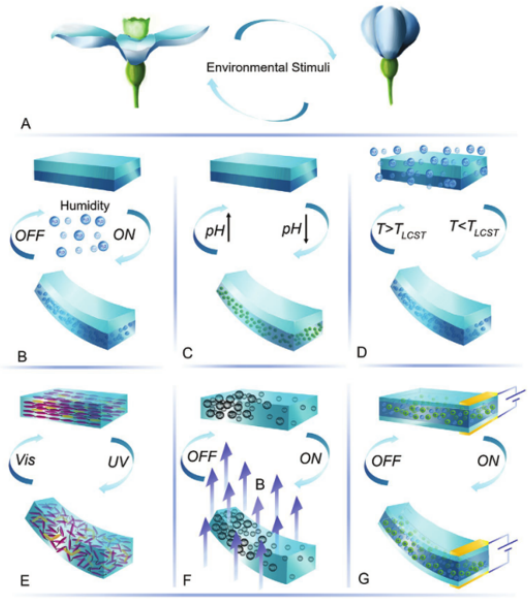


Figure 2. 3D printing smart actuators made of soft smart materials (reproduced from ref. 31 with permission of the John Wiley and Sons)

2.1. Water-responsive Actuators

In nature, many plants with two-layer structure can drive deformation through the change of h content in external water. The anisotropic deformation of the two-layer structure is caused by the different responses of each layer to water stimulation. Therefore, the researchers were prepared a variety of water drive response inspired in nature.

Water-responsive actuators are widely used in human body with the increasing demand of biomedical and sensor applications. The drive behavior is usually manifested by a changing in shape. Over recent years, bilayer composites have been widely used in the design of water responsive actuators. Consists of two layers, with different water absorbing capacity of double layer composite materials can realize response to water drive. Ren et al.[10] constructed a water responsive actuator based on composite membrane shown in Fig. 3. The composite membrane can be converted between two different states, thus showing bidirectional shape memory effect in response to water.

Han[11] et al. prepared skin collagen fiber/polyurethane (SCF/PU) composites with double network structure by paper making method with water as solvent for the first time shown in Fig. 3. SCF/PU composite has completely non-warm water responsive shape memory ability, and has high shape fixation rate (> 95%) and shape recovery rate (> 90%) in repeated memory cycle. In this study that exposure to water

can destroy the hydrogen bond in collagen fibers, which makes the flexible elastomer matrix change into the desired shape. In addition, the existence of collagen fiber network makes SCF/PU composite biodegradable and with low cytotoxicity.

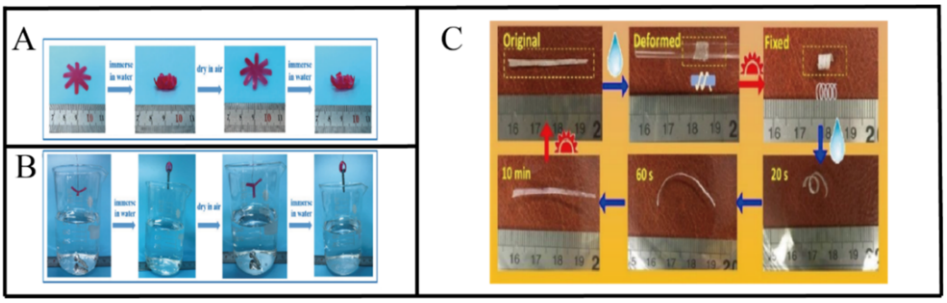


Figure 3. 3D printing water-responsive actuators. A-B) flower-like and claw-like actuators (reproduced from ref. 39 with permission of the Royal Society of Chemistry); C) water responsive shape memory behavior (reproduced from ref. 40 with permission of the Royal Society of Chemistry)

2.2. pH-responsive Actuators

Although the water responsive smart actuators is powerful and adaptable, it is not practical in fluid environment, such as precision mechanical structure and microfluidic application. The materials used in these drives will deform with the change of pH. Researchers found that this feature can be applied to the development of smart actuators. The development of different synthetic polymers has enabled researchers to make pH-responsive drives. Therefore, researchers have developed a number of pH-response smart actuators based on groups of different polymers. The polymers used to prepare pH-responsive actuators can be divided with basic group and acid group.

2.2.1. Ionization of Acidic Group

Dutta et al[12]. reported a pH-responsive smart hydrogel actuator. The responsive polymer displayed reversible cross-links with fluctuations in pH between pH 2.0 and 7.4 shown in Fig. 4.[12] In another study, Okwuosa et al[13] .recently used 3D dual extrusion FDM printing for the fabrication of the drug release capsules shown in Fig. 4. This 3D printed polymer shell structure has pH response characteristics. The core and shape memory capsule structure are designed using CAD modeling technology, and the best shell thickness is >0.52 mm. Due to the presence of the pH-response characteristic shell structure, its limited solubility under acidic pH protects the effect of the drug in an acidic gastric environment. The control of drug release is achieved by 3D printing with accurate shell thickness. This method can be widely used in the medical field in the future.

2.2.2. Ionization of Basic Groups

A number of basic groups have been used to develop basic polymers, including tertiary amine such as poly(N-(dimethylamino) (PDMA), morpholino such as poly[(2-N-morpholino)ethyl methacrylate] (PMEMA), and pyridine groups such as poly(4vinylpyridine) (P4VP), to name a few. Nadgorny et al[14].. prepared a 3D

printing of an actuator using pH-responsive P2VP. When P2VP is blended with ABS, the mechanical strength can be significantly improved, because P2VP has excellent plasticizing properties in the extrusion manufacturing process.

The 3D printing of pH-responsive actuators shows great potential for their applicability in the medical field due to their biodegradation and biocompatibility[13]. However, the applicability of pH-responsive protein smart hydrogels in 3D printing actuators and materials are limited due to the non-physiological pH gel effect. Therefore, this is a problem that many researchers need to solve at present.

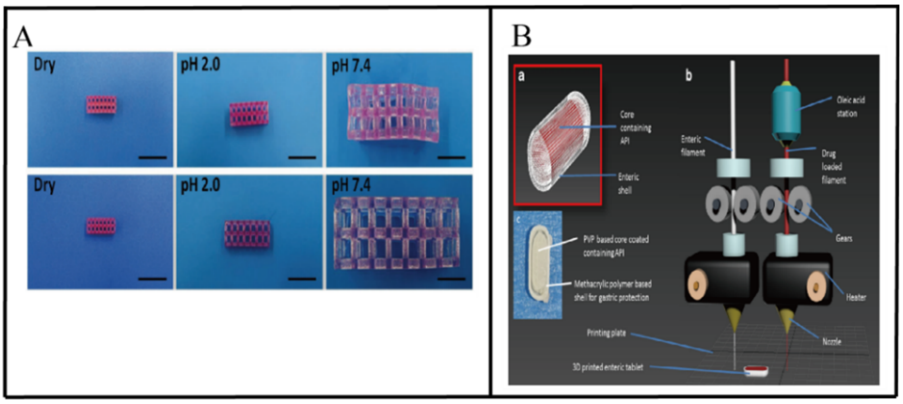


Figure 4. 3D printing water-responsive actuators. A) flower-like and claw-like actuators (reproduced from ref. 44 with permission of the Royal Society of Chemistry); B) water responsive shape memory (reproduced from ref. 45 with permission of the Springer Nature)

2.3. Thermo-responsive Actuators

Although a number of water-responsive and pH-responsive actuators have been developed with excellent features based on 3D printing technology. This has prompted researchers to invest in the development of drives suitable for dry environments. Among these actuators, the advantage of temperature responsive actuators is that they can realize various driving methods, because they can respond to various heating methods. In addition, these actuators are capable of showing large volume and shape changes. Therefore, researchers have developed a number of temperature-sensitive response drivers for 3D printing technology.

Shin[15] and others successfully manufactured a temperature response soft actuator by 3D printing technology, demonstrating its switching function in electrical applications. In another report, thermo-responsive PNIPAm along with polyether-based PU into 3D printed poly (2-hydroxyethylmethacrylate) (PHEMA) has been reported for shape morphing using 3D printing shown in Fig. 5[16] .The thermo-responsive actuator has fast response performance and is manufactured by this method. The thermo-responsive smart hydrogel smart actuator with LCST is limited by the phase change when heated, it has huge application potential.

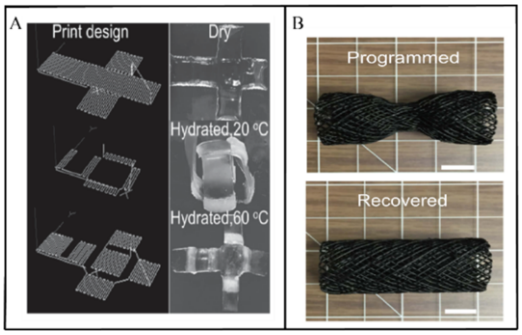


Figure 5. 3D printing Thermo-responsive actuators. (A) Smart hydrogel actuator (reproduced from ref. 54 with permission of the John Wiley&Sons, Inc). (B) Cylindrical support. (reproduced from ref. 55 with permission of the Springer Nature).

2.4. Photo-responsive Actuators

Photo-responsive actuators are considered to be one of the most promising smart actuators, because it can realize non-contact driving, selective and precise driving controlled by wavelength and intensity, and high-resolution temporal and spatial control. Different materials and technologies have led to photo-responsive actuators responding to different wavelengths, from ultraviolet to near infrared (NIR), direct (photochemical) or indirect (photothermal).

Rob et al[17]. developed a novel, light responsive thermoplastic actuator with versatility in shape design, reprogramming and multi-mode driving. In addition, this simple preparation method can also be used for the functionalization of other thermoplastic polymers, such as polyimide and polyamide. Due to its outstanding versatility and ease of manufacture, the photo-responsive thermoplastic actuator has established a new toolbox for future smart soft robot equipment, which requires fast and reversible driving, mechanical robustness, as well as activate any geometric shape that is controllable.

2.5. Magnetic Responsive Actuators

Magnetic responsive actuators have great applications including micro-optics systems, electronics, flexible robotics and biomedicine. Magnetic stimulation drives have unique advantages over other drives based on 3D printing technology. These advantages include: 1) non-contact remote control, widely used in various complex environments; 2) Precisely control the actuator; 3) the ability to generate high-frequency AC magnetic fields; 4) tunable ability of the actuators from nano scale to macro scale. However, the disadvantage of magnetic field responsive actuator is that it needs a complex and heavy external device to control the magnetic field and its gradient accurately.

Lee et al.[18] showed 3D printed micro-magnetic actuators with controllable dimensions which can respond to a variety of stimulus. The miniature magnetic actuator can expand and contract reversibly in multiple cycles with temperature, pH and divalent cations. It can expand and contract through narrow passages at designated locations. Their potential to control the occlusion of small capillaries. Furthermore, once the micro-actuator expands, it can only be removed by enzymatic degradation. This kind of micro drive can be widely used in the biomedical field.

3. 3D printing Smart driving materials

3D printing smart driving materials are mainly divided into: 3D printing shape memory materials, 3D printing piezoelectric materials, 3D printing responsive smart hydrogels and 3D printing electroactive polymers. At present, researchers have carried out a lot of work in 3D printing smart materials and demonstrated their applications in the preparation of various kinds of drive devices such as mechanical power generation. Over the past 5 years, researchers have used smart materials, such as shape memory materials, piezoelectric composites, smart smart hydrogel and electroactive polymer to produce smart actuator. These materials have been widely used in medical, health care, aerospace, bio-robot, underwater exploration and auxiliary equipment

3.1. 3D printing shape memory materials

The combination of shape memory material (SMM) and 3D printing technology not only optimizes the material preparation process, but also has the advantages of small mass, small impact and simple structure as smart driving materials compared with mechanical structure driving. At present, great progress has been made in the field of 3D printing SMM, and it has been applied in aerospace, electronic communication, smart robot and other fields. Shen Xinxin et al.[19] used 3D printing method to prepare shape memory carbon fiber composite pod rod, which broke through the limitation of mold in traditional process and effectively reduced the cost. Matt et al. [20] used the general and simple method of 3D printing methacrylate macromonomer to manufacture SMM. The responsive object is used to make soft robots, minimally invasive medical devices, and electronic devices. Sampada et al.[21]developed a composite material composed of SMM PLA and pea, as well as piezoelectric barium titanate nanoparticles by using 3D printing technology, to prepare a robust sensor. The sensor can withstand temperatures from 23 ℃ to 100 ℃ and more than 5000 cycles. However, the performance of 3d printed shape memory materials can not be used in special environment. Therefore, can be widely used in aerospace field of 3D printing shape memory materials received extensive attention of the researchers.

3.2. 3D printed piezoelectric composites

Piezoelectric materials are important parts of a series of devices such as transducers, sensors and actuators. Their functions are usually realized by a single brittle ceramic chip. In order to control the size and shape of piezoelectric materials, mechanical chips or saws are usually used. With the current cutting technology, it is almost impossible to mold brittle piezoelectric materials into more precise structures, which cause many hazards on precise sensor design, efficient transducers and diagnostic equipment. Kanguk et al.[22] used 3D printing technology to optically print efficient piezoelectric nano particle polymer composites into 3D microstructure. By doping barium titanate nanoparticles into light curable polymer solutions and exposing them to a dynamically changing digital optical mask, a user-defined three-dimensional microstructure was generated, and a piezoelectric polymer was prepared. Cui et al.[23]proposed a method to design electromechanical coupling anisotropy and orientation effect, and produced high response piezoelectric materials by 3D printing technology. This creates the freedom to design arbitrary piezoelectric tensors in reverse design, including symmetry

and failure characteristics, and goes beyond the common coupling modes observed in piezoelectric monolithic and foam. With this material, the user can design, amplify or suppress any operating mode (DNM) for the target application.

3.3. 3D printing smart hydrogel

Smart hydrogel is a kind of hydrophilic polymer material with three-dimensional network structure, and its three-dimensional polymer network is filled with a large amount of bound water, interface water and free water, showing dual characteristics of quasi-solid and quasi-fluid. Due to its unique physical and chemical properties such as softness, smoothness, stimulus response, biological compatibility and material transport and exchange, smart hydrogel has attracted extensive attention in the construction of functional devices and mechanical equipment in the fields of sensing, flexible electronics, drive, coating, optics and water collection. In addition, smart hydrogels are widely used in biomedical research, including tissue engineering, medical regeneration, drug delivery and biological mechanisms, due to their physiological characteristics similar to those of tissues and organs. However, in order to ensure that smart hydrogel and its derived devices have good controllability and practicability, smart hydrogel materials are required to have intrinsically excellent mechanical properties and rich functional characteristics in practical application exploration. Ji et al.[24] introduced secondary microstructures on the sides of the smart hydrogel strips, which resulted in bending or distortion due to asymmetric expansion. With the advantages of free-form design and manufacturing, various smart hydrogel structures are constructed through 3D printing based on stereolithographic printing, which achieve complex and controllable shape deformation through programmed microstructures on characteristic surfaces. It is worth mentioning that various response smart hydrogels are compatible with this method, which can stimulate reversible shape deformation. At the same time, a set of electric response smart hydrogel flexible driving grippers was manufactured by using 3D printing technology. Zhu et al.[25] prepared a complex structure of PIC smart hydrogel using 3D printing technology. The PIC smart hydrogel has different ionic bonding strength in a variety of parameters of 3D printing. In concentrated brine solution, PIC formed a viscous solution, which can be directly extruded from the nozzle into the water, and the salts and counterions are dialyzed, resulting in the sol gel transition, forming a PIC gel with complex structure and tough physical properties.

In recent years, with the rapid development of high strength and toughness smart hydrogel and adhesive smart hydrogel material system and the continuous innovation of smart hydrogel processing and manufacturing technology, functional smart hydrogel has become one of the important materials of the next generation of intelligent machinery, especially soft machinery. Naturally, smart hydrogel machinery was established and developed rapidly as a new concept. Core purpose is to explore in the field of the smart hydrogel as an important part of intelligent devices and machinery, according to actual needs and conditions of application of the traditional hard materials (metals, ceramics, plastics, etc.) of parts to replace or supplement, expand the smart hydrogel machinery in biomedicine, software robots, flexible electronics, energy and environment and other important practical applications.

3.4. 3D printing electroactive polymer

Electroactive polymer are materials that can adjust its volume mechanically in response to electrical stimulation. However, the manufacturing process of electroactive polymer is very complex. The size or shape of electroactive polymer changes under the stimulation of electric field. David et al.[26] designed a novel 3D printing flexible dielectric material and characterized its application as an electroactive polymer (deap) actuators. The expansion area of the actuator is 5.48% under 4.3kv applied voltage, and the initial prestrain applied to the dielectric material is 63.21%.

4. Application of 3D printing smart driving materials

4.1. Application of 3D printing smart driving materials in soft robot field

The smart response of natural creatures have always been the source for researchers to design soft robots. With the research of natural molluscs, such as elephant nose, earthworm body, octopus tentacle, a variety of biomimetic soft robots have been widely applied. The contact part of traditional rigid robot is composed of rigid materials, which is easy to damage soft objects. In terms of human-computer interaction, medical operations, item grabbing, grab sorting, underwater exploration etc., the rigid robots rely on sensor control to achieve safe interaction, which not only brings manufacturing difficulties, but also makes the complete set of equipment more expensive and is not conducive to the popularization and application. Therefore, the rapid manufacture of mechanical grippers with a variety of functions has been widely concerned by researcher. The actuators use the gas produced by chemical reaction to drive by air pressure, which realizes the movement of crawling and swimming.

4.2. Application of 3D printing smart driving materials in medical field

Custom implants can be made for a fraction of the initial cost because of the specificity of 3D printing. Although the automotive, art and aerospace sectors have benefited a lot of the medical devices sector accounts for a huge market in the 3D printing industry. Therefore, customized devices and implants need to be manufactured. In the biomedical field, 3D printing is often used to manufacture biocompatible implants, artificial tissues and organs. In recent years, researchers have successfully designed and manufactured heart valves and knee meniscus through 3D printing.3D printing rib cage is designed for patients who have lost their ribs due to accidents and other circumstances. Oral cavity or oral cavity includes maxilla (maxilla), mandible (mandible), 32 teeth (16 teeth per mandible), muscles, nerves and blood vessels. 3D printing plays an important role in creating complex biocompatible devices to replace these damaged tissues. Therefore, whether in industrial production or daily life, the medical industry related to 3D printed intelligent devices is in a good stage of development and has become a hot industry for research and development. In this context, it is of great significance and value to develop 3D printing intelligent devices with excellent performance. In the future, the research and application of 3D printed intelligent drivers will use the cutting-edge achievements of many disciplines to achieve high flexibility, versatility, high affinity and other performance characteristics,

so as to be widely used in various medical, aerospace, military and other fields.

5. Conclusions

3D printing smart actuators is one of the important directions of the development of modern high-tech materials. As a new manufacturing method that intelligent actuators using 3D printing technology has developed rapidly in recent years. The smart driving materials include shape memory materials, piezoelectric composites, smart smart hydrogels, electroactive polymers, etc. The fields involved include aerospace, soft robots, etc. The research and application of 3D printing smart actuators is far beyond these. In the future, the research and application of 3D printing smart actuators would make use of the cutting-edge achievements of many disciplines to achieve high flexibility, versatility, high affinity and other performance characteristics, so as to be widely used in various medical, aerospace, military and other fields.

The key factors to for the applications of 3D printing smart actuator include the usability of external stimuli, drive efficiency and mechanical performance requirements. Therefore, it can create a tradeoff between performance and material printability due to limitations.

The latest development of 3D printing technology has achieved multi-smart material printing, embedded smart material printing, and the use of magnetic fields and particles to calibrate filling materials. The combination with smart actuator materials means that this technology may become a viable to the next generation of intelligent technology. Many reports discuss the proof that the entire review focused on concepts and potential applications. However, with the continuous advancement of 3D printing smart drive technology and materials, these concepts may soon be realized in the near future.

References

[1] Deng J, Li J, Chen P, et al. Tunable Photothermal Actuators Based on a Pre-programmed Aligned Nanostructure. Journal of the American Chemical Society. 2016; 138(1): pp 225-230.
[2] Liu L, Jiang S, Sun Y, et al. Giving Direction to Motion and Surface with Ultra-Fast Speed Using Oriented Smart hydrogel Fibers. Advanced Functional Materials, 2016; 26(7): pp 1021-1027.
[3] Shahsavan H, Salili SM, Jakli A, et al. Thermally Active Liquid Crystal Network Gripper Mimicking the Self-Peeling of Gecko Toe Pads. Advanced Materials, 2017; 29(3).
[4] Sotiriou GA, Blattmann CO, Pratsinis SE. Flexible, Multifunctional, Magnetically Actuated Nanocomposite Films. Advanced Functional Materials. 2013; 23(1): pp 34-41.
[5] Terasawa N, Asaka K. High-Performance PEDOT:PSS/Single-Walled Carbon Nanotube/Ionic Liquid Actuators Combining Electrostatic Double-Layer and Faradaic Capacitors. Langmuir. 2016; 32(28): pp 7210-7218.
[6] Wang F, Jeon J-H, Kim S-J, et al. An eco-friendly ultra-high performance ionic artificial muscle based on poly(2-acrylamido-2-methyl-1-propanesulfonic acid) and carboxylated bacterial cellulose. Journal of Materials Chemistry B. 2016; 4(29): pp 5015-5024.
[7] Ban J, Mu L, Yang J, et al. New stimulus-responsive shape-memory polyurethanes capable of UV light-triggered deformation, hydrogen bond-mediated fixation, and thermal-induced recovery. Journal of Materials Chemistry A. 2017; 5(28): pp 14514-14518.
[8] Zhang YF, Ng C JX, Chen Z, et al. Miniature Pneumatic Actuators for Soft Robots by High-Resolution Multimaterial 3D Printing. Advanced Materials Technologies. 2019; p 4(10).
[9] Ilami M, Bagheri H, Ahmed R, et al. Materials, Actuators, and Sensors for Soft Bioinspired Robots. Advanced Materials, 2020.

[10] Ren S, Feng J. Water-responsive actuators based on the solution casted PVA/epoxidized-SBS two-way shape memory bilayer composite film. Journal of Materials Chemistry C, 2020; 8(42): pp 14826-14833.
[11] Han Y, Hu J, Chen X. A skin inspired bio-smart composite with water responsive shape memory ability. Materials Chemistry Frontiers. 2019; 3(6): pp 1128-1138.
[12] Dutta S, Cohn D. Temperature and pH responsive 3D printed scaffolds. Journal of Materials Chemistry B, 2017, 5(48): pp 9514-9521.
[13] Okwuosa T C, Pereira B C, Arafat B, et al. Fabricating a Shell-Core Delayed Release Tablet Using Dual FDM 3D Printing for Patient-Centred Therapy. Pharmaceutical Research. 2017; 34(2): pp 427-437.
[14] Symes MD, Kitson PJ, Yan J, et al. Integrated 3D-printed reactionware for chemical synthesis and analysis. Nature Chemistry. 2012; 4(5): pp 349-354.
[15] Shin S, So H. Time-dependent motion of 3D-printed soft thermal actuators for switch application in electric circuits. Additive Manufacturing, 2021, 39.
[16] Naficy S, Gately R, Gorkin R, Iii, et al. 4D Printing of Reversible Shape Morphing Smart hydrogel Structures. Macromolecular Materials and Engineering. 2017; 302(1).
[17] Verpaalen RCP, Da Cunha MP, Engels TP, et al. Liquid Crystal Networks on Thermoplastics: Reprogrammable Photo-Responsive Actuators. Angewandte Chemie-International Edition. 2020; 59(11): pp 4532-4536.
[18] Lee YW, Ceylan H, Yasa IC, et al. 3D-Printed Multi-Stimuli-Responsive Mobile Micromachines. Acs Applied Materials & Interfaces. 2021; 13(11): pp 12759-12766.
[19] Shen X, Jia B, Zhao H, et al. Study on 3D printing process of continuous carbon fiber reinforced shape memory polymer composites, 2019 International Conference on Advanced Electronic Materials, Computers and Materials Engineering. 2019.
[20] Zarek M, Layani M, Cooperstein I, et al. 3D Printing: 3D Printing of Shape Memory Polymers for Flexible Electronic Devices (Adv. Mater. 22/2016). Advanced materials (Deerfield Beach, Fla.), 2016; 28(22): 4166-4166.
[21] Bodkhe S, Ermanni P. 3D printing of multifunctional materials for sensing and actuation: Merging piezoelectricity with shape memory. European Polymer Journal, 2020, 132.
[22] Kim K, Zhu W, Qu X, et al. 3D Optical Printing of Piezoelectric Nanoparticle - Polymer Composite Materials. Acs Nano. 2014; 8(10): pp 9799-9806.
[23] Cui H, Hensleigh R, Yao D, et al. Three-dimensional printing of piezoelectric materials with designed anisotropy and directional response. Nature Materials. 2019; 18(3): p 234.
[24] Ji Z, Yan C, Yu B, et al. 3D Printing of Smart hydrogel Architectures with Complex and Controllable Shape Deformation. Advanced Materials Technologies. 2019; 4(4).
[25] Zhu F, Cheng L, Yin J, et al. 3D Printing of Ultratough Polyion Complex Smart hydrogels[J]. Acs Applied Materials & Interfaces. 2016; 8(45): pp 31304-31310.
[26] Gonzalez D, Garcia J, Newell B. Electromechanical characterization of a 3D printed dielectric material for dielectric electroactive polymer actuators. Sensors and Actuators a-Physical. 2019; 297.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 139
J. Xu et al. (Eds.)

doi:10.3233/ATDE220221

Operating Efficiency of Software Platform for Machine-Vision-Oriented CNC System

Chuhao QIU [a], Kunzi WANG [a], Wenyou DONG [a] and Liming XU [a,1]
[a] *School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

Abstract. The combination of machine vision and open CNC system contributes to develop various functions of CNC system such as visual inspection, condition diagnosis and machining error compensation. Focusing on the two commonly used software development platforms (Visual Studio, Qt) and three machine vision libraries (OpenCV, Halcon and EmguCV), this paper studies the operating efficiency of software platform for development of open CNC system with machine vision function. The operating efficiency of various schemes is experimentally compared. Finally, the machine-vision-based error compensation efficiency is studied by taking the curve grinding CNC system as an example. The optimization scheme based on operating efficiency is presented, which provides a basis for the selection of software platform for machine vision oriented CNC software.

Keywords. Image processing, open CNC system, machine vision, software development, operating efficiency

1. Introduction

The requirements for portability and flexibility have contributed to the birth and development of open Computerized Numerical Control (CNC) system. The common architecture of open CNC system is a combination of industrial PC and real-time control unit. As the upper computer, PC is responsible for the man-machine interface, complex or massive calculation, while the real-time control unit is responsible for high real-time functions such as servo drive of the lower computer. With the continuous development of machine vision technology, the introduction of digital image processing technology into CNC system has become a developing direction. Many functions such as online visual inspection, diagnosis, machining error compensation can be developed in CNC control system with machine vision. But at the same time, the host computer not only needs to have the function of traditional human-computer interface, but also have the ability of fast real-time image processing and display. It puts forward a high demand on how to choose the right software development language, integrated development environment and vision development tools, in order to improve the operating efficiency of software development platform.

[1] Corresponding Author, Liming XU, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, Shanghai, China; E-mail: limingxu@sjtu.edu.cn.

In the aspect of introducing machine vision into open CNC system, Tian [1] et al. developed a classification system for vegetable grafting machines, and developed image processing and corresponding algorithms using the open source computer vision library (Open CV) on the Visual C++ platform. Zou [2] et al. designed a six-axis robotic arm weld tracking experimental platform based on the Halcon machine vision library. Its control system includes the upper computer software module of Visual Studio Microsoft. The upper computer software module mainly performs image acquisition, image processing and coordinate calculation. Schumacher [3] et al. developed model-based computer vision applications based on Visual C++, which can be used in systems such as industrial robots, but there are still limitations in visual recognition. Hazrat Ali [4] et al. developed an industrial sorting robot manipulator based on image processing by using VB. Yan [5] et al. established an image processing algorithm library through C# program, and added machine vision function to the industrial sorting system. Hou [6] et al. developed an online tool wear monitoring system based on machine vision. With the help of Matlab software, a self-matching algorithm is proposed for monitoring tool wear. The corresponding user-friendly graphical user interface (GUI) of the algorithm is developed, improving the accuracy and reliability of tool wear detection. Rui [7] et al. developed a machine vision system for inspecting the dimensions of flat coil springs. The detection algorithm of the system is written in C++ language. The machine vision system improves the stability of detecting the dimensions of flat coil springs.

In terms of the operating efficiency of software platform, Alnaser [8] et al. compared the efficiency of horizontal algorithms based on C++ and Java, but did not consider the efficiency of the development platform. Elsayed [9] et al. analyzed the speed and efficiency of the algorithm run on Matlab and OpenCV, concluding that Matlab is more convenient in development and data display, and OpenCV is more conducive to improving execution efficiency. For the unmanned aerial vehicles control problem, Parfiryev [10] et al. proposed a correlation filtering target detection algorithm based on two-dimensional discrete Fourier transform, using OpenCV and C++ to realize the software implementation of the detection algorithm. The efficiency of the object detection algorithm is improved.

Commonly used software platforms include Visual Studio (VS) and Qt, etc. As a Microsoft IDE, VS contains almost all the tools in the software development life cycle and the code supports most of the products of the Windows. At present, Visual Studio supports mainstream programming languages such as Visual Basic (VB), C, C++, C#. The programming languages are based on .NET architecture, which greatly facilitate the development of various API (application programming interface); Qt is a C++ based cross-platform graphical user interface application framework. It supports Linux, Windows, iOS platforms, and the rich GUI library can easily be used to develop various kinds of human-computer interaction.

Typical machine vision development libraries include OpenCV, Halcon, EmguCV, etc. OpenCV is an open source machine vision library which provides interfaces with C++, Java, Python, etc. OpenCV focuses on improving the computing efficiency of image processing, that is, efficient processing in real-time application. Halcon is a commercial machine vision library. It has been applied in various fields of medicine, measurement and control, and industrial automation. Halcon provides a convenient hybrid programming mode compared to other visual libraries. EmguCV is a product developed by OpenCV developers to package OpenCV under the .Net framework for the situation that OpenCV cannot be used in Visual Studio's C# and VB.

This study selects the commonly used software platforms and machine vision development libraries, conducts a time-consuming test of typical image processing algorithms, studies the operating efficiency of different software platforms, and provides data support for the selection of open CNC system development platforms.

2. Operating efficiency of software development platform

In order to explore the impact of various development platforms, languages and visual library on the processing efficiency, two algorithms of image processing, gray-scale and thresholding are selected to test the processing efficiency. The Lena image (resolution 512×512, RGB image) is chosen as original image and the processing results are shown in Figure 1.

To ensure the accuracy of the test time, the QueryPerformanceCounter function of WindowsAPI is selected as the timer, which can provide 64-bit timing accuracy. The average running time is taken as the result after 1000 times operation of each algorithm on each platform.

The test computer configuration is: The CPU is i5-6300HQ, GPU is NVIDIA GeForce GTX 960M, the memory is 4G, and the system is Windows 10 Home Edition (64-bit), the software is Qt 5.7.0, Visual Studio 2013, Halcon 10.0, OpenCV 2.4.13 and EmguCV 3.0. The test results are shown in Table 1 and Figure 2.

(a) Original drawing (b) Grayscale (c) Thresholding

Figure 1. Test picture and result after algorithm processing.

Table 1. Execution efficiency test using different vision libraries and development platforms (unit: ms)

Time / Platform / Algorithm	Halcon + Qt	Halcon + (C++)	Halcon + (C#)	OpenCV +Qt	OpenCV + (C++)	EmguCV + (C#)
Grayscale	0.357	2.231	8.147	0.626	2.193	3.375
Thresholding	0.194	1.561	5.776	0.124	2.530	2.857

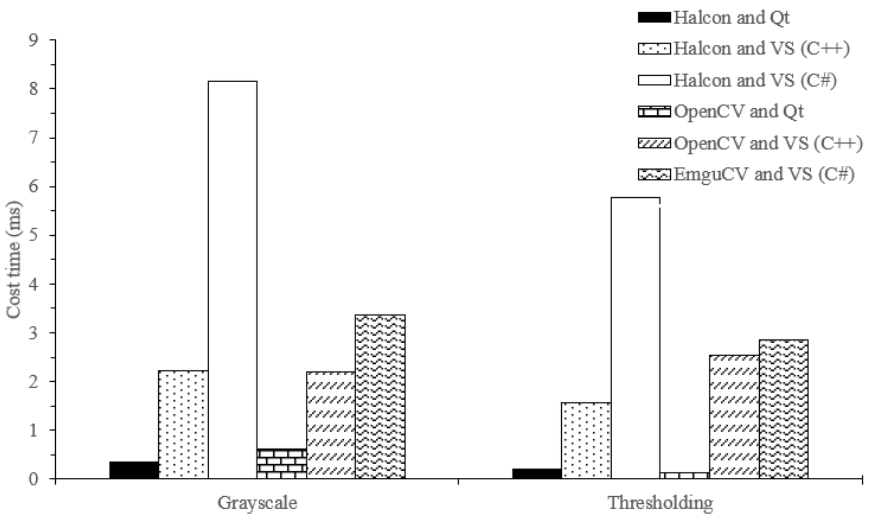


Figure 2. Comparison of execution speed of image algorithms based on different software platforms.

It can be seen from Figure 2 that the efficiency of C++ related platforms is generally better than that of C# related platforms, especially in the comparison of C++, Halcon with C#, Halcon. However, Qt has the best effect on the C++ platform. OpenCV has better performance than Halcon on Qt platform, but is close to Halcon on VS (C++) platform.

3. Experiment on error compensation efficiency based on machine vision

In the machining process, open CNC system is responsible for image shooting, image transmission and capture, image processing, compensation value calculation, etc., as shown in Figure 3. Therefore, higher requirements are put forward for real-time performance of CNC system. In addition, in order to avoid the running stuck of man-machine interaction, the CNC system with visual compensation function must be a multi-threaded system. The main thread is only responsible for man-machine interaction and some low-demand processing processes. The complex and massive computing real-time processing tasks should be placed in the secondary thread. In order to evaluate the efficiency of machining error compensation based on visual images, the complete online compensation process of CNC system was analyzed.

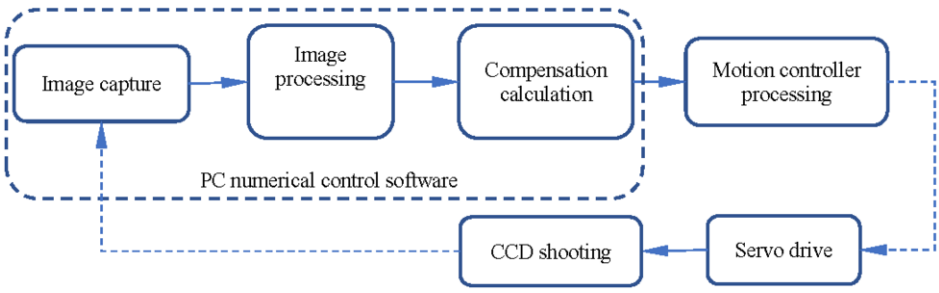
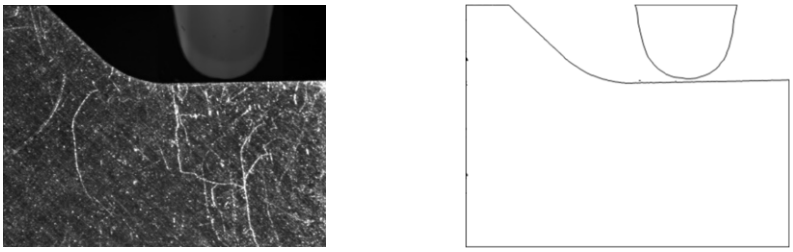


Figure 3. Online CNC machining error compensation process based on visual image.

The experimental platform is the machine vision oriented CNC curve grinding testbed. The PC configuration is the same as the previous experiment, the software platform is Qt 5.7.0, the visual library is OpenCV 2.4.13, and a black and white camera is model GC2441M with 5 mega-pixel global shutter CCD. The Motion controller is MC664-X from Trio Motion company, and the servo motion system is PD series linear motor from Parker and A6N series servo motor from Panasonic.

The experiment mainly tests the time-consuming of each function module in Figure 3 except the servo motion. Figure 4 (a) is the actual CCD image, and the image in Figure 4 (b) can be obtained after operations such as Blur filtering and findContours in OpenCV library. The edge of the black area of the image is the contour of the tool edge and the workpiece edge. The contour is compared with the ideal contour, and the tool motion compensation amount can be obtained by algorithm processing. The amount of compensation will be transferred to the motion controller to complete the compensation process. Table 2 shows the time cost of each process item.



(a) Original image (b) Image after edge extraction

Figure 4. The original image and the image after edge extraction in curve grinding.

Table 2. Consuming time analysis of error compensation in CNC curve grinding

Process item	Consuming time (ms)	Time proportion
CCD shooting and image capture	101.013	83.4%
image processing	16.284	13.4%
compensation calculation	3.617	2.99%
motion controller processing	0.250	0.02%
total	121.164	100%

As can be seen from Table 2, CCD image shooting and image capture occupy most of the time in the online compensation process, while image processing occupies the highest proportion of the remaining time. The running time of first two items should be reduced in order to improve the response speed of the online compensation of CNC system.

4. Conclusions

The operating efficiency of software platform for machine-vision-oriented CNC system is studied, and the conclusions are as follows:

(1) The platform operating efficiency is directly affected by the combination selection of development software and vision library. The efficiency of C++ related software platforms is higher than C# related software platforms with the same vision library. Qt is slightly more efficient than VS under the same vision library. Halcon

library and OpenCV library are basically close in operating efficiency under the same development environment.

(2) CCD shooting and image capture account for more than half of the whole operation time. Image processing occupies the largest proportion in the remaining time and its efficiency directly affects the operation efficiency of the CNC system.

(3) Software development platform for vision oriented CNC system should be selected considering both the operation efficiency and development cost. For example, in order to improve operating efficiency, Qt or VS (C++) can be selected to use with Halcon library or OpenCV library.

Acknowledgment

Funding for this study was provided by the National Natural Science Foundation of China (No. 52075331).

References

[1] Subo T, et al. Development of an automatic visual grading system for grafting seedlings [J]. Advances in Mechanical Engineering, 2017, 9(1): pp. 119-125.
[2] Yanbiao Z and Tao C. Laser vision seam tracking system based on image processing and continuous convolution operator tracker [J]. Optics and Lasers in Engineering, 2018, 105: pp. 141-149.
[3] Paul S and Musa J. A system for automated disassembly of snap-fit covers [J]. The International Journal of Advanced Manufacturing Technology, 2013, 69(9-12): pp. 2055-2069.
[4] Md. Hazrat A et al. Vision-based Robot Manipulator for Industrial Applications [J]. Procedia Computer Science, 2018, 133: pp. 205-212.
[5] Juan Y and Huibin Y. Research on Workpiece Sorting System Based on Machine Vision Mechanism [J]. Intelligent Control and Automation, 2015, 6(1): pp. 1-9.
[6] Qiulin H, Jie S and Panling H. A novel algorithm for tool wear online inspection based on machine vision [J]. The International Journal of Advanced Manufacturing Technology, 2019, 101(9): pp. 2415-2423.
[7] Rui Z, et al. Flat spiral spring dimension inspection based on machine vision [J]. Journal of Physics: Conference Series, 2018, 1074(1): pp. 012179-012179.
[8] Alnaser A M, Alheyasat O, Abuein A K, et al. Time Comparing between Java and C++ Software [J]. Journal of Software Engineering & Applications, 2012, 5(8): pp. 630-633.
[9] Ahmed A. Elsayed and Waleed A. Yousef. Matlab vs. OpenCV: A Comparative Study of Different Machine Learning Algorithms [J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2019, abs/1905.01213.
[10] Parfiryev A V, et al. Control of unmanned aerial vehicles based on the detection algorithm [J]. Journal of Physics: Conference Series, 2019, 1202(1): p. 012014.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 145
J. Xu et al. (Eds.)

doi:10.3233/ATDE220222

Design of Target Tracking System Based on Apriltag

Xin HUANG[a], Zuoshi LIU[a,1], Pengsheng CHENG[a]
[a] *School of Mechanical and Electrical Engineering, Jiang Xi University of Science and Technology, Ganzhou341000, Jiangxi, China*

Abstract. In the research and development of target tracking, vision-based target tracking technology still has the problems of low accuracy and high cost. This paper designs a tracking system that uses a four-wheel differential mobile chassis as a carrier and uses the 36H11 series tags in Apriltag as a moving target. STM32F103 single-chip microcomputer is used as the core of motion control, and the classic PID control algorithm is used to adjust the wheel speed to achieve target tracking. STM32F765 single-chip microcomputer is used as the image processor of the OV7725 camera to solve the label information. The experimental results show that the Apriltag tag target can be better tracked when the PID parameters are adjusted properly. It can achieve near real-time tracking effect when the tag moving speed is slow, and it can move to the specified position quickly when it is far away from the target.

Keywords. Apriltag; Target tracking; PID algorithm;STM32 microcontroller

1. Introduction

The continuous development of target tracking technology has made it widely used in various fields. In the research of target tracking technology, vision-based target tracking technology has always been a popular research topics. Shafique used the point target tracking method and proposed a multi-frame method to maintain the current consistency of speed and position[1]. Omaniciu uses a target tracking method based on the idea of image segmentation, and proposes a mean-shift method to find clusters in the space where space intersects with color[2]. In order to improve the robustness of target detection and reduce the consumption of computing resources in 2011, the University of Michigan proposed a new visual benchmark system for 2D bar graphs, which combined with the fast edge detection system to form a more superior digital encoding system[3]. Based on previous research, this paper combines vision and mobile chassis to realize a target tracking system based on Apriltag. Through PID control algorithm, adjust the parameters to make it achieve the best tracking effect.

2. Experimental system construction

The experimental system adopts a four-wheel differential mobile chassis as the carrier,

[1] Corresponding Author, Liu Zuoshi, School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China; Email: liuzuoshi@163.com.

STM32F103ZET6 as the chassis control core, OV7725 camera as the vision sensor, and STM32F765VIT6 as the visual information processing chip. The TB6612FNG motor drive chip adjusts the speed of the motor. Powers the system with a 1500MAH, Li-Po battery. The block diagram of the target tracking system is shown in Figure 1. The physical map of the experimental system is shown in Figure 2. In Figure 2, a: visual image processing module; b: STM32F103 motion control core; c: motor drive module; d: power supply; e: wireless transmission module.

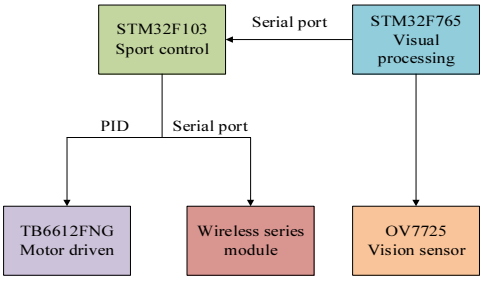


Figure 1. Block diagram of target tracking system.

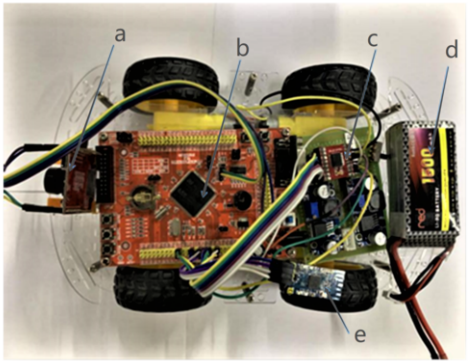


Figure 2. Physical map of the experimental system.

2.1. The working principle of the experimental system

The label is randomly placed on the ground, the camera on the car is used to find the label, and then the car is controlled to reach the specified position, and the relationship between the data obtained and the actual value is observed. Then move the tag in real time and experiment with the tracking function to see if it can track in real time. After the camera recognizes the label, the visual processing chip calculates the position of the label relative to the camera. At this time, the origin of the coordinate system is centered on the camera. Since the experimental platform uses a mobile chassis, it is only necessary to obtain the position information in Z and X directions, and the position information of the tag is sent to the motion control chip through the vision processing chip. The motion control algorithm of the mobile chassis uses the classic PID control algorithm. According to the actual debugging situation of the system, it is found that only PI control is needed to achieve a better control effect. The flow chart of the working principle of the experimental system is shown in Figure 3.

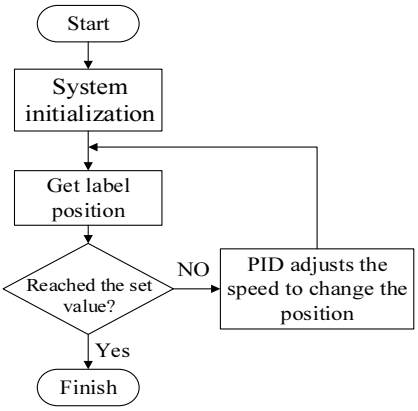


Figure 3. Working principle diagram of the experimental system.

3. Apriltag tag recognition

Label recognition is mainly divided into two steps of detection and decoding. The detection process is mainly divided into three parts (1) identifying line segments, (2) identifying quadrilaterals, (3) homography and external estimation. The decoding of tags is by calculating the relative coordinates of the tags of each bit field, using homography to convert them into image coordinates, and then thresholding the resulting pixels. In order to reduce the influence of light on the recognition effect, the intensity "black" pixel of the spatial variation model and the intensity "white" model of the second model were established in Apriltag [4]. The expression is shown in formula (1).

$$I(x, y) = Ax + Bxy + Cy + D \tag{1}$$

The parameters of the model can be solved through least square regression analysis, and the black and white models are calculated separately and the average value is used as the predicted intensity value. There are currently multiple families of Apriltag tags. The 36H11 family was selected for the experimental system, which has 586 family members, that is, 586 different tags. The larger the number after H, the more family members and the greater the density of the color block of the label. The larger the color block of the same size, the longer the detection distance, and the easier it is to lose the label information. The No. 1 label of the 36H11 family is shown in Figure 4.

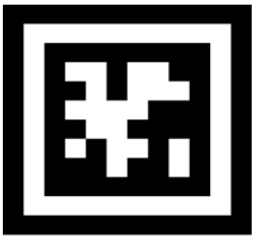


Figure 4. 36H11 family label No. 1.

In this article, Apriltag's recognition method is to import the library file in micopython into the STM32 microcontroller and directly call it to improve the real-time performance of recognition. The Apriltag image taken by the camera is shown in Figure 5.

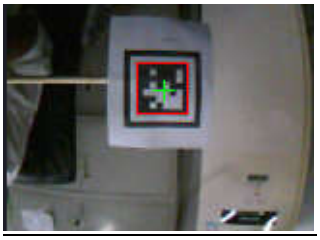


Figure 5. Apriltag image taken by the camera.

3.1. Camera coordinate system transformation

The coordinate origin of the identified position information is the camera, if the position information of the image coordinate system needs to be multiplied by the rotation matrix R. Optionally set a point t in the image to correspond to the camera down vector and the image coordinate system down vector (t1,t2,t3), $(t_1^{'}, t_2^{'}, t_3^{'})$T respectively.

The transformation relationship:

$$\begin{pmatrix} t_1^{'} \\ t_2^{'} \\ t_3^{'} \end{pmatrix} = R \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} + P \tag{2}$$

Where P is the translation amount of coordinate transformation.

4. Application of PID Control Algorithm in Target Tracking System

The PID control algorithm is a classic control algorithm. Although there are many new optimized control algorithms, its essence is the adjustment of proportional or integral derivative. There are two types of PID control algorithms: positional and incremental. Position control will produce a large cumulative error in the control of the motor. In this experimental system, an incremental PI controller is used, making it only related to

the errors of the last three times. The PI control model block diagram is shown in Figure 5.

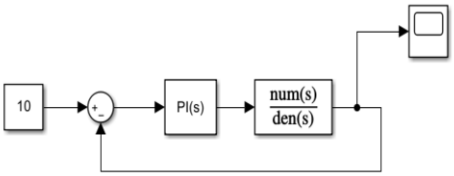


Figure 6. Block diagram of PI control model.

4.1. Mathematical Model of PID Control Algorithm for DC Motor

According to the principle of automatic control and the parameters of the motor, the transfer function of the DC motor can be calculated to establish the mathematical model of PID control through the matlab simulink module. 130 DC motor parameters are shown in Table 1. The speed transfer function of the DC motor is shown in equation (3).

Table 1. 130 DC motor parameters

The main parameters	Value
Rated torque	6.31g·cm
Rated power	0.75W
Rated voltage	3V
Rated speed	8530r/min
Reduction ratio	48:1
Armature inductance	4.9mH
Armature resistance	1.1Ω

Find the transfer function when the motor is running as:

$$G(s)=\frac{K_T}{L_\alpha Js^2+(L_\alpha B+R_e J)s+K_e K_T+R_\alpha B} \tag{3}$$

Where *KT*: Torque constant; *Lα*: Armature inductance; *J*: Moment of inertia; *Rα*: Armature resistance: *Ke*: Back EMF coefficient; *B*: Viscous damping coefficient.

By calculation:

$$G(s)=\frac{0.05}{3.16\times10^{-9}s^2+7.25\times10^{-6}s+2.8\times10^{-3}} \tag{4}$$

Substitute the obtained transfer function results into the Simulink simulation model, and adjust the PI parameters to obtain the control effect as shown in Figure 6.

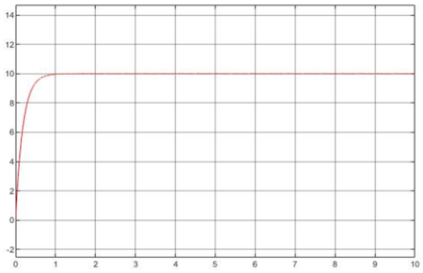


Figure 7. PI control effect diagram.

5. Analysis of results

5.1. Distance calibration error analysis

The calculated position information data is not the real distance. If the real position information is obtained, it needs to be calibrated. Place the labels at different distances to obtain multiple sets of measured values at different distances, average the measured values at different distances, and finally perform curve fitting to obtain the fitting curve between the true value and the measured value as shown in Figure 7. It can be seen from the curve that there is a linear relationship between the real value and the measured value and the fitting effect is better. The fitting equation is shown in equation (5).

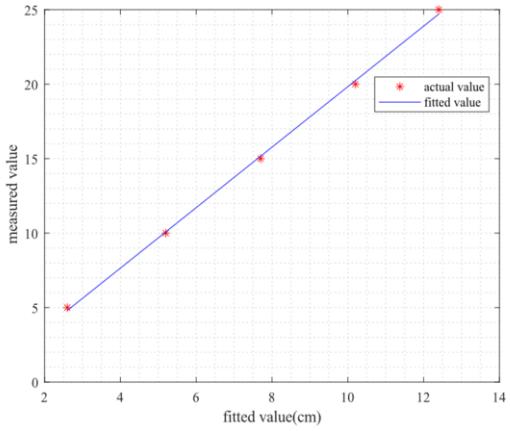


Figure 8. Fitting curve of z-axis true value and measured value.

$$y = 2.031x - 0.475 \tag{5}$$

5.2. Dynamic tracking experiment analysis

The experiment process is divided into three stages. The label is quickly moved to a position about 10 cm away from the lens, then slowly moved for 5 seconds, and finally

moved quickly for a certain distance and then stopped. Figure 8 shows the tracking curve of the mobile platform dynamically tracking Apriltag tags. The time period of 0-3 seconds is the fast moving stage, and 3-8 seconds is the uniform moving stage. The label stops moving at 11 seconds. The last point and the first point in the figure are basically in a straight line, that is, the distance between the label and the camera. The distance is the same. The fitting curve expression is shown in formula (6).

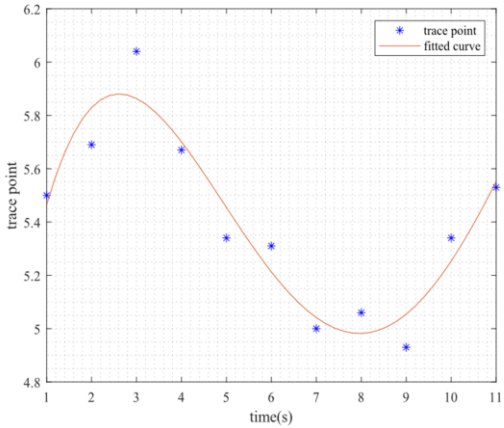


Figure 9. Dynamic tracking curve.

$$y=-0.001x^4+0.032x^3-0.335x^2+1.160x+4.608 \tag{6}$$

The camera distance tag measurement value set in the experiment program is 5.5, and the measurement value in the dynamic tracking process is subtracted from 5.5 to get the dynamic tracking error. The error value recorded in Table 2 at 10 seconds from the initial state of 0 to the final state error back to 0.03, it can be seen that the dynamic response of moving the site in the process of fast moving the tag is better. The tracking stability is better at a constant speed.

Table 2. Dynamic error value

Time	1	2	3	4	5	6	7	8	9	10
Error	0	0.19	0.54	0.17	-0.16	-0.19	-0.5	-0.44	-0.16	0.03

6. Summary and outlook

This article mainly uses Apriltag visual positioning and tracking algorithm to realize the tracking of tags, and uses the mobile chassis as the carrier to realize the function of target tracking. The mathematical model of the PID transfer function of the experimental platform is established, which is convenient to quickly adjust the PID parameters. The experimental results show that the use of Apriltag algorithm for target tracking has the advantages of strong robustness and good dynamic response, and has wide applicability in the future of moving target tracking. There are many tags in the Apriltag family, which can be used in robot formations, which not only reduces the

hardware pressure of mutual communication, but also greatly improves the anti-interference ability. Of course, if you are in an environment with poor visibility or foggy weather, it will affect the accuracy of tag identification.

Reference

[1] SHAFIQUE K, SHAH M. A non-iterative greedy algorithm for multi-frame point correspondence[J]. IEEE Trans. Patt. Analy. Mach. Intell , 2005, 27(1) : 110-115.
[2] COMANICIUD，MEERP．Mean shift: a robust approach toward feature space analysis[J].IEEE Trans. Patt. Analy. Mach. Intell, 2002, 24(5) : 603-619.
[3] Olson E. AprilTag: A robust and flexible visual fiducial system[C]. international conference on robotics and automation, 2011: 3400-3407.
[4] Abbas S M, Aslam S, Berns K, et al. Analysis and Improvements in AprilTag Based State Estimation.[J]. Sensors, 2019, 19(24).
[5] Mark Petersen, Chad Samuelson, Randal W. Beard. Target Tracking and Following from a Multirotor UAV[J]. Current Robotics Reports,2021.
[6] Xiu Kan, Jia He, Zhenghao Xi. Visual Tracking based on Moving Monocular Camera[J]. International Journal of Performability Engineering,2020,16(8).
[7] Abbas Syed Muhammad, Aslam Salman, Berns Karsten, Muhammad Abubakr. Analysis and Improvements in AprilTag Based State Estimation.[J]. Sensors (Basel, Switzerland),2019, 19(24).
[8] Yu J, Jiang W, Luo Z, et al. Application of a Vision-Based Single Target on Robot Positioning System[J]. SENSORS. 2021, 21(5).
[9] Kyristsis S, Antonopoulos A, Chanialakis T, et al. Towards Autonomous Modular UAV Missions: The Detection, Geo-Location and Landing Paradigm[J]. SENSORS. 2016, 16(11).

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 153
J. Xu et al. (Eds.)

doi:10.3233/ATDE220223

Lightweight Detection Model Based on Attention Mechanism

Yun SHEN [a, 1] and Lizuo JIN [a]
[a] *School of automation, Southeast University, Nanjing, China*

Abstract. Object detection algorithms have been widely used in important fields such as national defense, military industry and transportation. However, due to the limitation of computing power and power consumption of hardware devices, most of the object detection models deployed in mobile devices are lightweight models, and their detection accuracy can not fully meet the needs of complicated scenes. This paper proposes a new model adding path aggregation and attention mechanism to nanodet, an anchor-free detection model, and improves the detection accuracy with litte cost. On our dataset, tested with RTX 3080TI the AP50 reaches 62.09% and only cost little additional inference time. It has a good reference value for the deployment of lightweight model in engineering practice.

Keywords. deep learning, attention mechanism, convolutional neural network, object detection

1. Introduction

The manufacturing level of computing devices has advanced rapidly these years and neural network has become a popular research area. Among them, the fastest growing and most widely used is the object detection network. Object detection refers to finding out the object of interest in an image and determining its specific location and classification.

The earliest target detection network is two-stage, which detection step includes two steps. First find some regions of interest on the feature map based on the original image, and then infer the exact location and size of the target through a series of convolutional layers. Its typical representatives are Faster RCNN [1], Cascade RCNN and so on. However, the network structure of the two-stage network is more complex and the detection speed is very slow, which is difficult to be applied to mobile devices.

Later, one-stage network came into being, which removed the time-consuming RPN module in two-stage network and directly enumerated anchors on the pixels on the feature map, significantly improving the inference speed of the network. Its typical representatives are the YOLO series, which have made it possible to deploy object detection model to low computing power devices.

Although YOLO is an one-stage network, it is still anchor-based, which has many drawbacks because the shape and size of the anchors are pre-customized. In the past several years, many anchor-free networks have appeared. They do not need to use sliding windows to enumerate anchors, but directly generate corresponding prediction

[1] Yun SHEN: School of automation, Southeast University, Nanjing, China, 220191611@seu.edu.cn

bounding box and confidence for each point. The anchor-free network further simplifies the network structure and improves the network speed, and its accuracy is comparable to that of the anchor-based network, which is an important development direction for future lightweight models.

Attention mechanism is originally a technique in the field of NLP. Later, some teams tried to apply it to computer vision. The essence of attention mechanism is weight assignment, which improves the detection performance of the network by increasing the weights of the effective part of the network. In the field of computer vision, the attention mechanism mainly consists of spatial attention and channel attention. The former applies different computational weights to different pixels of the image, and the latter applies different computational weights to different channels of the convolutional feature layer.

The application of attention mechanism in the field of computer vision has become a popular research direction, and there are many related works that have achieved success. External Attention [2] is an improvement on Self Attention [3]. ECA [4] implements a lightweight attention mechanism module. Coordinate Attention [5] is an attention mechanism module for mobile-based networks.

In this paper, we use the anchor-free network nanodet as the basis, and incorporate the spatial attention and channel attention mechanisms into the network backbone. This architecture improves the detection accuracy and still maintains fast detection speed.

2. Problem Description

The dataset in this paper is produced by intercepting images from video recordings of actual scenes, and the main detection objects are pedestrians and vehicles.

The cameras are mounted high up and shoot diagonally downwards, and the pictures of the dataset are mainly taken from a top-down perspective. The shooting locations including campuse, markets, intersections, etc. The shooting time period includes day and night. Therefore, the target size in the dataset varies, the scene is complex and the lighting conditions are variable. This dataset can more completely simulate the needs of real surveillance scenarios. Some of the dataset images are shown in Figure 1.

Figure 1. Example pictures of dataset

3. Network Structure

3.1. Backbone

After a series of comparisons, we finally chose ShuffleNetV2 as backbone for the network. ShuffleNetV2 is a fast and accurate lightweight network. It consists of multiple shuffle blocks, and in each block a series of techniques are applied to ensure accuracy.

During the training process, a large network with too many layers may cause negative effects such as vanished gradient and reduced accuracy. In order to solve the problem, this paper applies a residual connection branch similar to ResNet in each block. It transfers the shallow features directly to the deep layer and uses concatenate operations to minimize the information lost during convolutional layer transfer.

In addition, this paper uses depth separable convolution instead of traditional convolution, which reduces the model complexity. The inference time consumption of a neural network mainly stems from the number of parameters of the convolutional layers. The number of input channels, the number of output channels and the kernel size of the convolutional layer all have an impact on the number of parameters. Assuming the size of convolution kernel is K, the number of input channels of feature map is M and the number of output channels is N, the parameter of the traditional convolution kernel is $P_n = K \times K \times M \times N$.

Depth separable convolution has two parts, depthwise convolution and pointwise convolution. Depthwise convolution enables each convolution kernel to convolve only a single channel of the input feature map, which equals group convolution with the same number of groups as the number of input channels. The number of parameters is $P_d = K \times K \times 1 \times M$. Pointwise convolution is an ordinary convolutional layer which kernel size is 1. Its number of parameters is $P_p = 1 \times 1 \times M \times N$. Compared with the traditional convolutional layer, the parameter quantity of the depth separable convolutional layer is only $(P_d + P_p)/P_n = 1/N + 1/K^2$. Under normal circumstances, the value of N is very large and can be ignored. When K is set to 3, the number of parameter can be reduced to approximately 1/9 of the traditional layer.

After using residual join and concatenate measures, the feature maps of the two parts are directly superimposed on each other. In order to make a connection between them, the channel shuffle technique is used at the end of each block, which breaks up the channels of the two parts and then pairs them one by one and integrates them back. The structures are shown in Figure 2 and Figure 3.

By using these methods, the network greatly reduces the number of parameters and increases information transfer, thus improving network performance.

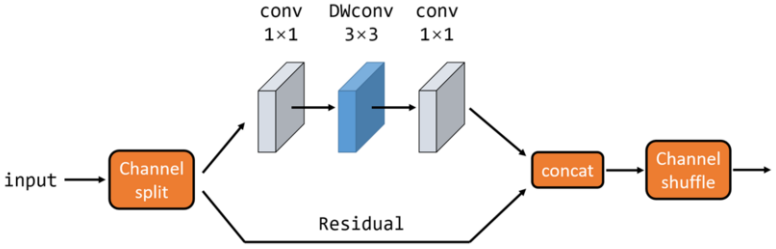


Figure 2. Block with stride=1

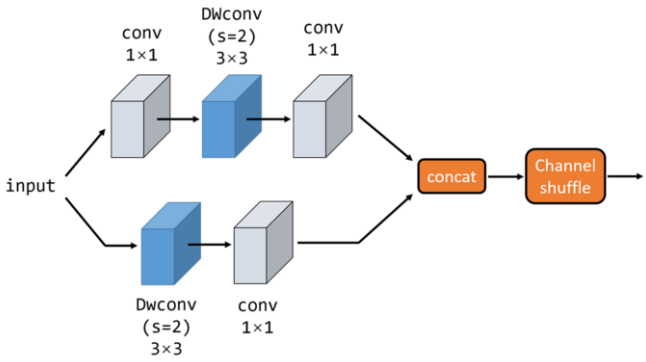


Figure 3. Block with stride=2

3.2. Path Aggregation

Image detection network downsamples every few layers. The traditional network model is based on the last layer of feature map for object classification and bounding box regression. Due to multiple downsampling, the traditional network model performs object classification and border regression based on the last layer of feature maps, which have lost a lot of information of the original image, especially information of some small targets. This will lead to the limitation of the detection capability of the network..

In order to solve this problem, a structure named Feature Pyramid Network has been proposed. FPN upsamples the smaller feature map of lower layer and enlarges it to the same size as the upper layer, then adds them together. At the same time, FPN no longer only detects the last layer of feature maps, but detects last several layers of feature maps.

The Path Aggregation Net used in this article uses an additional feature transmission on the basis of FPN. FPN uses a feature transmission from bottom to top, while PAN adds a transmission from top to bottom. The architecture is shown in Figure 4.

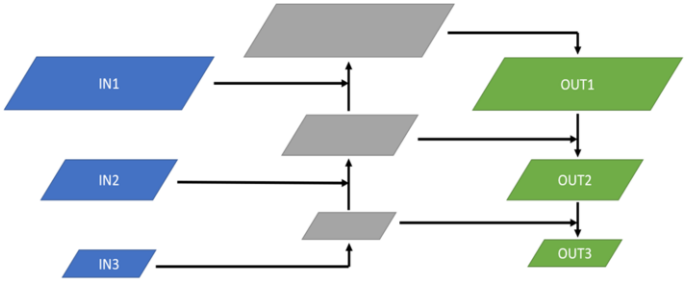


Figure 4. Path Aggregation Network

After adopting this structure, the network can fully take into account the targets of multiple sizes. For large targets, the deeper feature maps contain richer semantic information, which ensures the detection accuracy. For small targets, the shallower feature maps have not yet lost too much detail information and can also have better detection results.

3.3. Attention Blocks

The attention mechanism refers to focusing more on the parts we want to detect and less on the irrelevant parts. The essence of its implementation is to learn a set of weighted distribution data through the self-learning ability of the network. The attention part is weighted more and the irrelevant part is weighted less, thus simulating the behavior of focusing attention on the focused information.

We use the CBAM [6] attention module in this article. It consists of a channel attention module and a spatial attention module. The former one uses MaxPool and AvgPool operations. If the input shape is C×H×W, the output will be C×1×1. Each channel has a corresponding weight value, thus implementing the mechanism of channel attention. The spatial attention module also uses MaxPool and AvgPool operations, but effect dimensions are different. For the input shape of C×H×W, the output shape will be 1×H×W, which means each position of the input feature map has a corresponding weight value, so as to realize the spatial attention mechanism.

This article inserts the CBAM module into each block of the backbone. The architecture is shown in Figure 5 and Figure 6. Using both attention mechanism at the same time can help improve the inference accuracy.

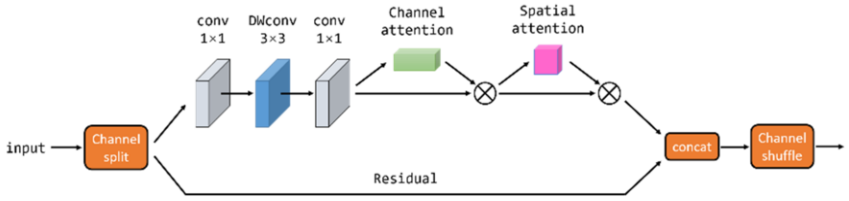


Figure 5. CBAM-Block with stride=1

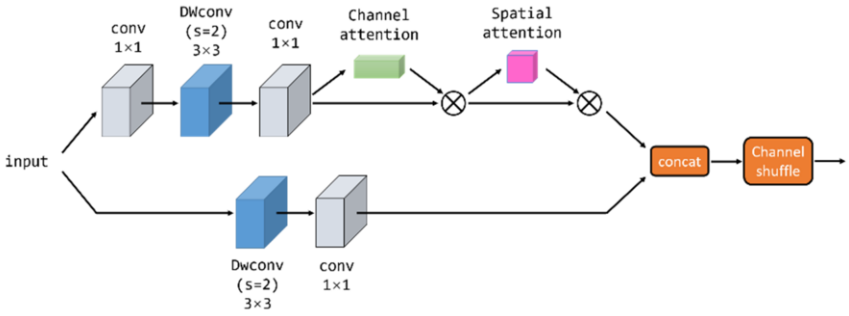


Figure 6. CBAM-Block with stride=2

3.4. Architecture

In the backbone part, this article uses 3 stages, each containing [4, 8, 4] CBAM-Block. Each stage applies stride=2 downsampling at the first block.

In the neck part, this paper uses PAN to process the output of these three stages to increase the multi-scale perception ability of the network. After PANet, 3 feature maps of different sizes are output, and each feature map connected with a detection head.

In the final part of classification and regression, this paper uses a shared convolutional layer of classification regression to combine classification of objects and regression of bounding box into one for training and prediction, which improves the accuracy of target detection.

The final network structure is shown in Figure 7.

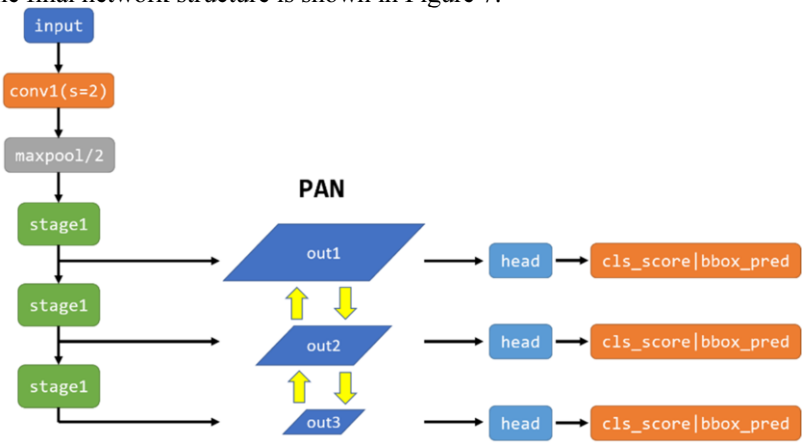


Figure 7. Network structure

4. Experiments and Comparisons

The CPU we used to do experiments is I7-11700, the GPU is NVIDIA GeForce RTX 3080TI and the memory size is 32G. The operating system for training and testing is Ubuntu 18.04.

The image resolution of the dataset is 1920×1080, the training set has 5000 images, and the verification set has 1000 images.

Since the dataset targets are mostly small and medium-sized, if the image is scaled too small, the small target will only occupy a few pixels, which is too difficult to distinguish and is not friendly to the lightweight detection network. For improving the detection performance, this article scales the picture into 640×640 and 800×800 size to input to the network. We compare and test the native network, network only with channel attention module, network only with spatial attention module, and network with both attention modules.

The accuracy and speed test under 640*640 input are shown in Table 1.

Table 1. Modle performance with 640*640 input(+C denotes add channel attention blocks, +S denotes add spatial blocks)

Model	mAP	AP50	Speed(GPU)	Speed(CPU)
Nanodet	26.64	55.28	4.82ms	17.31ms
Nanodet+C	26.83	55.82	5.01ms	18.43ms
Nanodet+S	26.70	55.51	5.04ms	18.56ms
Nanodet+C+S	27.52	57.12	5.24ms	19.12ms

The accuracy and speed test under 800*800 input are shown in Table 2.

Table 2. Modle performance with 800*800 input(+C denotes add channel attention blocks, +S denotes add spatial blocks)

Model	mAP	AP50	Speed(GPU)	Speed(CPU)
Nanodet	29.80	60.68	4.98ms	25.19ms
Nanodet+C	29.97	61.01	5.12ms	26.11ms
Nanodet+S	29.89	60.92	5.16ms	26.28ms
Nanodet+C+S	30.78	62.09	5.40ms	27.25ms

As can be seen from the above two tables, the increase in input resolution helps a lot in network detection, and the additional inference time is very little when using GPU to inference, but greater when using CPU inference. When the channel attention or space attention module is used alone, the improvement is small. When they are used together, the mAP can be increased by about one point with little additional time cost.

We use the trained model (800*800 input) to test some pictures, and its performance is shown in Figure 8.

Figure 8. Example results

5. Conclusion

This article attempts to add an attention module to a lightweight detection network. Experiments show that the attention module can improve the detection performance of the network to a certain extent and the additional time cost is negligible, which is an effective way to improve the network performance.

In our experiments, we found that the accuracy of the network is not always improved after adding the attention mechanism to the network. Corresponding adjustments to the training parameters need to be made according to the training curve. Also, the convergence speed of training becomes slower, and the number of training epochs needs to be increased.

Since the dataset used in the experiment is intercepted from the surveillance video, the number of samples is small, and some pictures are highly repetitive, which affects the result of training. In the future, we can try to pre-train on a large dataset first and then use migrate training on this dataset.

References

[1] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28: pp. 91-99.

[2] Guo M H, Liu Z N, Mu T J, et al. Beyond self-attention: External attention using two linear layers for visual tasks[J]. arXiv preprint arXiv:2105.02358, 2021.
[3] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017: pp. 5998-6008.
[4] Qilong W et al. "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks." 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: pp. 11531-11539.
[5] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: pp. 13713-13722.
[6] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: pp. 3-19.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 161
J. Xu et al. (Eds.)

doi:10.3233/ATDE220224

Self-Powered Angle Sensor for Drill Pipe Based on Triboelectric Nanogenerator

Delong ZHANG [a,b], Chuan WU [c,d,1], Qiang GUO [a] and Qing ZHOU [c]
[a] *Beijing Institute of Exploration Engineering, Beijing, 100083, China*
[b] *School of Engineering and Technology, China University of Geosciences (Beijing), Beijing, 100083, China*
[c] *Faculty of Mechanical and Electronic Information, China University of Geosciences (Wuhan), Wuhan 430074, China*
[d] *Hubei Intelligent Geological Equipment Engineering Technology Research Center, Wuhan 430074, China*

Abstract. Drill pipe angle is one of the key parameters for adjusting drilling technology, which needs real-time measurement. In this paper, an angle sensor based on triboelectric nanogenerator is proposed. The sensor can be set to different resolutions by designing different EVA sheets. Taking 12 EVA sensors as an example, the test results indicated the sensitivity is 0.08°/V, the linearity is 5.2%, and the maximum relative error is 6%. In terms of self-supply capability, the self-supply sensor can output maximum voltage are 29 V, a maximum current are 50 nA, and a maximum load power are 110 nW. In the case of 50000 working cycles, its test performance has not significantly decreased, showing high stability.

Keywords. Drill pipe, Angle sensor, Triboelectric nanogenerator, Self-powered

1. Introduction

Drilling is an engineering technique that uses the drilling tools to drill, and is one of the important means of mining the ground floor or the natural resources of the seafloor , and for drilling tools, the rotation angle is one of the important parameters to control the drilling process of the drilling tools [1]. At present, Angle sensors of the drill pipe are widely used in the market. Common Angle sensors are photoelectric[2], electromagnetic[3], eddy current[4], optical grating[5] and inertial measurement unit. However, these Angle sensors generally have high power consumption and cannot be self-powered. Therefore, it is necessary to develop an angle sensor with self powering function to adapt the well working conditions.

Triboelectric nanogenerator have been widely used in sensor and energy collection fields since they were proposed. In the field of sensors, such as speed sensing[6], flow sensors acceleration sensing, vibration sensing[7] and human movement monitoring. In the field of energy collection, it is applied to the collection of wind energy[8], sound wave energy, rain energy[9], mechanical energy[10] and wave energy. It can be seen that triboelectric generator has great advantages in Angle sensor application. Therefore,

[1] Chuan Wu, Corresponding author, Faculty of Mechanical and Electronic Information, China University of Geosciences (Wuhan), Wuhan 430074, China; E-mail: wuchuan@cug.edu.cn.

this paper presents a drilling rod angle sensor based on a triboelectric Nanogenerator. Compared with traditional Angle sensors, the drill pipe Angle sensor based on triboelectric nanogenerators can be self-powered, which provides a solution to the problem of difficult power supply for downhole equipment.

2. Structural design and working principle

2.1. Structural design

The structural model of the sensor is shown in Figure 1. The cylindrical blade angle sensor consists a rotor and a stator. The stator is a large cylinder, and the rotor is composed of a shaft, and a small cylinder and a connecting mechanism. Copper foil is pasted on the outer wall of a large cylinder as a negative electrode, and PTFE (polytetrafluoroethylene) is pasted on the copper foil as a negative friction material. The PTFE area on each EVA sheet is not the same, to produce different output when contacting with the positive electrode to realize the judgment of rotation angle. A rectangular EVA sheet is pasted on a large cylinder tangentially, and copper foil is pasted on the EVA sheet as a positive friction material and a positive electrode. The outer wall of the large cylinder is in close contact with the small cylinder, and the shaft is connected with the large cylinder by bearings. The small cylinder is connected to the middle rotating shaft by a connecting mechanism on the outside. When the rotating shaft rotates, the small cylinder can be driven to rotate. The resolution angle is determined by the number of rectangular EVA sheet. For example, in Figure 1a, where pasted 12 EVA sheets, it can distinguish 360°/12, that is, 30° angle. If the number is larger, the smaller the angle that the sensor can distinguish, that is, the higher the resolution.

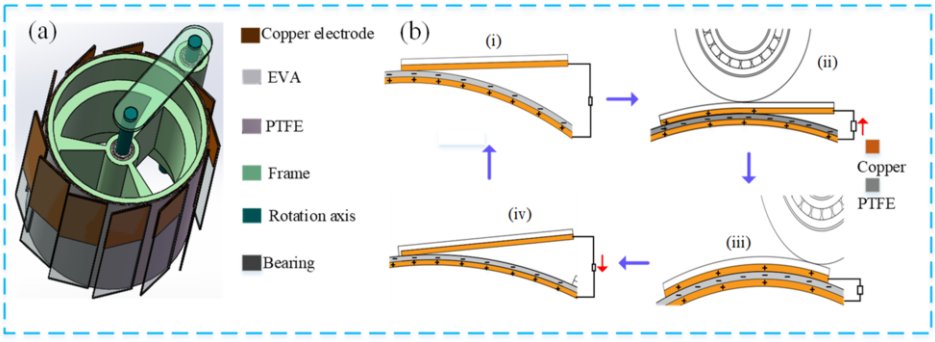


Figure 1. (a) Schematic diagram of the sensor structure; (b) Sensor working principle

2.2. Working Principle

The generating principle of the structure can be attributed to the electrostatic induction and the triboelectric effect. Take a small cylinder touching an EVA sheet once. As shown in Figure 1b (i), when the small cylinder does not contact the EVA sheet, the electronegativity of the two friction layer materials (PTFE and copper foil) on the large cylinder is different. The surface of copper foil has a positive charge, while the surface

of PTFE will generate an equal amount of negative charge. When the small cylinder rotates to contact the EVA sheet, the copper foil on the EVA sheet gradually contacts the PTFE, as shown in Figure 1b (ii). At this time, the positive charge is gradually transferred from the negative to the positive. The rotation axis continues to rotate, and the small cylinder presses down the EVA sheet to fully contact the PTFE, as shown in Figure 1b (iii). The negative charge on the PTFE forms a charge balance with the upper and lower copper electrodes again. There is no charge transfer. As the cylinder moves away from the EVA sheet, the EVA sheet gradually separates from the PTFE, and the positive charge is gradually transferred from the positive to the negative. At this point, the small cylinder gradually contacts another EVA sheet, as shown in Figure 1b (i) again, the next working principle is consistent with the above, but the maximum output current is different due to different areas of the friction layer. As the cylinder rotates, it produces several different peaks, depending on how many pieces of EVA sheet.

3. Test results

3.1. Sensor Power Generation Characteristics

As shown in figure 2a shows that the cylinder vane from the experimental process of the power supply Angle sensor, mainly is to use a rotary motor drive motion sensor, the output is connected to the KEITHLEY 6514 types of the electrometer to measure and get the voltage and current of the structure and transfer charge, then by connecting the resistance of the different sizes, and measuring the load output characteristic; Finally, the stability of the material is measured. As shown in Figure 2c, TENG has a maximum

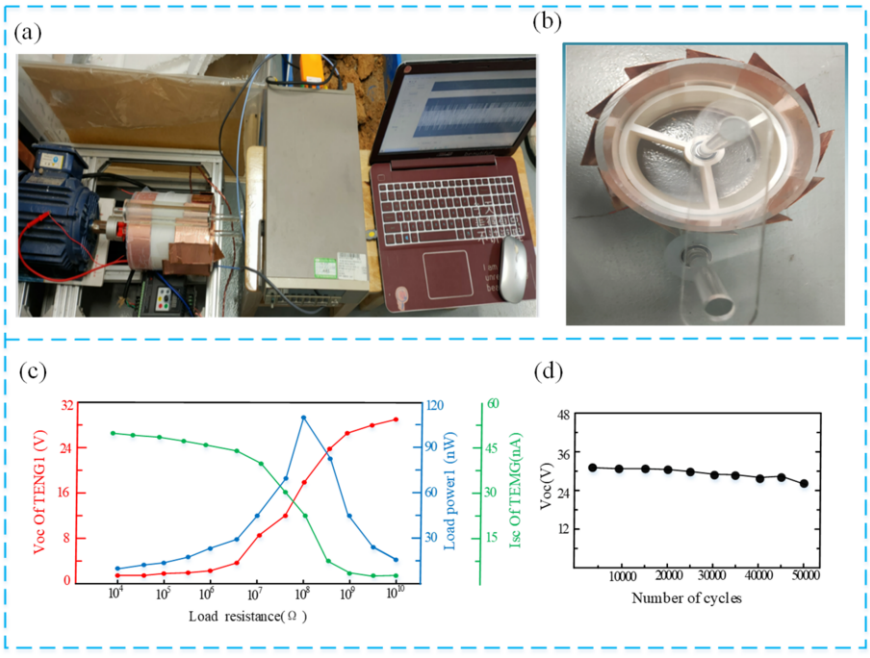


Figure 2. (a) Sensor test device and power generation characteristics. (b) Sensor photographs; (c) Voltage, current and output power of the sensor at different resistances; (d) Sensor stability test

output voltage of 29 V, a maximum current of 50 nA and a maximum output power of 110 nW. Figure 2d shows that the sensor has stable output performance even after 50000 rotations.

3.2. Sensor Characteristics

As shown in Figure 3a, when the rotor drives the small cylinder to rotate, the small cylinder presses the EVA sheet, making the inner copper foil contact with the PTFE outside the large cylinder. The contact area increases from small to large, and the output voltage increases gradually. There will be a process of voltage increase to wave peak every time an EVA sheet is pressed, and because of the different areas of PTFE from different angles, the wave peak shows an increasing trend. As can be seen from Figure 3b, as the rotation angle increases, its current will also increase. Because the large voltage and small current of the triboelectric nanogenerator, the sensor's output voltage and rotation angle are calibrated to obtain a large SNR (signal to noise ratio). As shown in Figure 3c and Figure 3d, the sensitivity of the sensor is 0.08°/V, the linearity is 5.2%, and the maximum relative error is less than 6%.

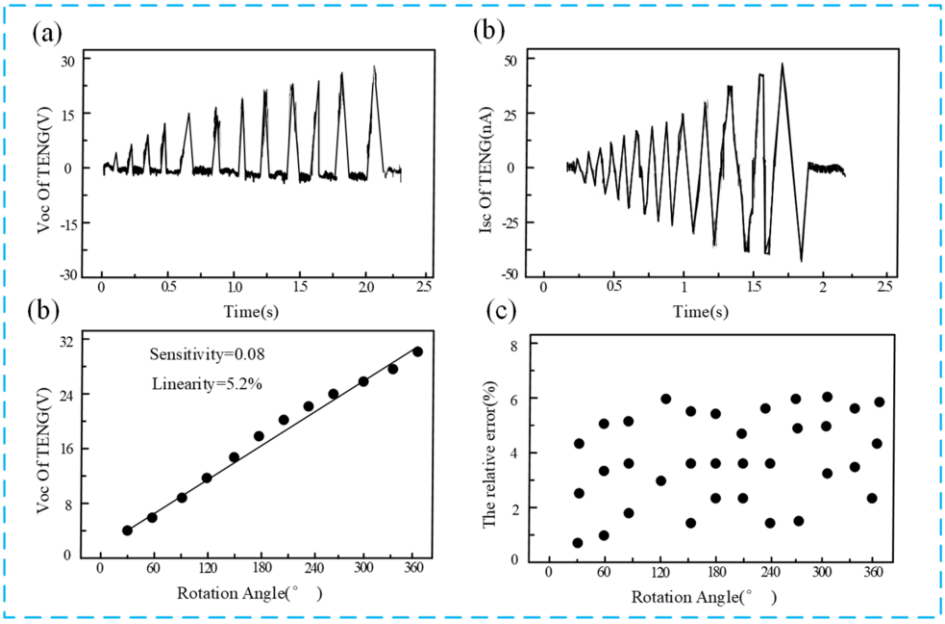


Figure 3. (a) Output voltage waveform of the TENG; (b) Output current diagram of the TENG; (c) Calibration curve of TENG output voltage and rotation angle; (d) Sensor error diagram

4. Conclusions

In this paper, a self-powered angle drill pipe sensor based on TENG is proposed. The sensor can output voltage with different amplitude at different rotation angles. Through the experimental results, the resolution of the sensor was 360°/12, i.e. 30°, with the sensitivity is 0.08°/V, the linearity is 5.2%, and maximum relative error is 6%. In terms of power generation capacity, the sensor can output voltage of 29 V, current of 50 nA,

and load power of 110 nW. In the case of 50000 working cycles, its test performance has not significantly decreased, showing a high stability. However, the output performance of this sensor is still low, especially the output current is low, so how to improve the performance of the friction layer materials to improve the output performance of the sensor is the focus of the next research.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 41802197), the Project of China Geological Survey (No. DD20201102) and the Research Foundation of Key Laboratory of Deep Geodrilling Technology, Ministry of Natural Resources (No. KF202103).

References

[1] Feito N, Díaz-Álvarez J, Díaz-Álvarez A, et al. Experimental analysis of the influence of drill point angle and wear on the drilling of woven CFRPs. Materials. 2014; 7(6): pp 4258-4271.
[2] Kolosov MP, Gebgart AY. Variants of nonmisadjustable optical systems of turning-angle sensors based on a BR-180° prism and a photoelectric autocollimator. Journal of Optical Technology. 2019; 86(9): pp 539-543.
[3] Tho D, Beaulieu L. Identification of an optimal electromagnetic sensor for in vivo electromagnetic - tracked scintillation dosimeter for HDR brachytherapy. Medical Physics. 2019; 46(5): pp 2031-2036.
[4] Allegro MicroSystems LLC; Researchers Submit Patent Application, "Angle Sensor Using Eddy Currents", for Approval (USPTO 20190265018). Politics & Government Week. 2019.
[5] García-Valenzuela A, Sandoval-Romero G E, Sánchez-Pérez C. High-resolution optical angle sensors: approaching the diffraction limit to the sensitivity. Applied optics. 2004; 43(22): pp 4311-4321.
[6] Huang HH, Wang RC, Chen YJ. Fluorinated graphite paper used for self-powered water speed sensors by immersion-type tribovoltaic effect-dominated triboelectric nanogenerators. Nano Energy. 2022; 93: p 106887.
[7] Chen J, Wang Z L. Reviving vibration energy harvesting and self-powered sensing by a triboelectric nanogenerator. Joule. 2017; 1(3): pp 480-521.
[8] Liu S, Li X, Wang Y, et al. Magnetic switch structured triboelectric nanogenerator for continuous and regular harvesting of wind energy. Nano Energy. 2021; 83: p 105851.
[9] Zhao L, Duan J, Liu L, et al. Boosting power conversion efficiency by hybrid triboelectric nanogenerator/silicon tandem solar cell toward rain energy harvesting. Nano Energy. 2021; 82: p 105773.
[10] Liu Y, Zheng Y, Wu Z, et al. Conductive elastic sponge-based triboelectric nanogenerator (TENG) for effective random mechanical energy harvesting and ammonia sensing. Nano Energy. 2021; 79: p 105422.

doi:10.3233/ATDE220225

Construction and Application of Digital Twin for Propulsion System in New Energy Ships

Jin WANG[a], Zheng XIAO[b,1], Teng WU[b]
[a]*Jiangsu Shipping College, P.R. China*
[b]*School of Mechanical and Electronic Engineering, Wuhan University of Technology, P.R. China*

Abstract. Digital Twin (DT) is considered to be the general purpose technology of the 4th industrial revolution, and realizing its engineering application universality is the common endeavor objective of both universities and companies. Based on the theoretical framework and technical route of DT, this paper focuses on its application exploration for new energy ships. Via blending technologies of cloud computing, open source software, the marine control system and characteristics of new energy, authors carry out a feasibility analysis from the shipboard and cloud architecture and implementation plan respectively. A DT of propulsion system is developed, deployed and operated online on a new energy ship (NES). Meanwhile, the DT data is used to correct the calculation deviation of the battery State of Charge (SOC) by the ship's physical system. The research in this paper will provide a decision-making platform for situations such as safe operation, fault diagnosis, and condition-based maintenance, and also provide an effective solution for future DT system design of new energy ships.

Keywords. Digital Twin, New Energy Ships, Propulsion System, Internet of things

1. Introduction

In recent years, new technologies, such as 4G/5G wireless communications, Internet of Things, Cloud Services and Artificial Intelligence, are undergoing rapid development. The physical entity world and the digital virtual world have formed two systems, which develop and interact in parallel. Digital virtual technology serves physical entities, which are more efficient and organized with the support of digital technology. In this context, the digital twin is born.

The earliest concept of the digital twin can be traced back to the "a virtual, digital equivalent to a physical product" proposed by Prof. Michael Grieves of the University of Michigan in 2003. He emphasized that digital representations of a physical product should be able to abstract itself, and the physical product can be tested under real or simulated conditions based on digital representations, which is to construct an equivalent virtual entity of the physical entity [1]. In 2009, the US Air Force Research

[1]Zheng Xiao, Jiangsu Shipping College, email:reallylaugh@whut.edu.cn

Laboratory (AFRL) also proposed the concept of "Airframe Digital Twin", in which Digital Twin first appeared [2-3]. In 2011, Prof. Michael Grieves officially cited Digital Twin (DT) in his new book "Virtually Perfect: Driving Innovative and Lean Products through Product Lifecycle Management", which quoted the suggestion of John Vickers of NASA [4]. Then, Dr. Fei Tao of Beihang University extended the three-dimensional model of Digital Twin "physical entity-virtual mirror-connection" to five dimensions. The five-dimensional model uses the manufacturing workshop as a carrier, which is combined with the emerging hardware and communication technologies, increases the dimensions of data and services. That caused a sensation in academia [5-6]. Based on DT five-dimensional model, technologies of application and exploration of DT have been widely carried out subsequently. Academician Jianrong Tan of Zhejiang University presented an assembly precision analysis method based on a general part DT model, which would integrate multi-source heterogeneous geometric models and maps assembly information from assembly semantics to geometry elements, allowing automatic assembly positioning of parts and improving the efficiency of assembly simulation [7]. Academician Deren Li of Wuhan University discussed the relationship between DTs and smart cities, analyzed the characteristics of smart cities based on DTs, and focused on the five main applications of smart cities based on DTs [8]. Those scholars provide method reference for DT's landing application.

Because DT is being an effective way to realize the intelligent interconnection and interaction of the manufacturing physical world and the information world, it is also highly concerned by technology companies. Global industrial and software giants such as Siemens, Ansys, Microsoft, and GE have made great achievements in the fields of DT platform construction and engineering applications after years of intensive cultivation. DT is also considered to be the general purpose technology of the fourth industrial revolution [3], and it is being applied in the fields of manufacturing, service industry, urban development and military industry. However, the current application research in academia still focuses on the theory, methodology, feasibility of the application direction, and experiments on a single-point problem, which is difficult to be adopted by engineering projects for reference and implementation. Most engineering applications in enterprises build a full-featured service platform with huge economic and labor costs, while application users only use the platform in a black box way. It is difficult to accurately implement the functional modules of the platform on the application object, and it is impossible to focus on the object mechanism to construct an accurate description model, which causes a waste of resources, and the application effect is also relatively poor.

As a general purpose technology of the fourth industrial revolution, DT's engineering application should be more universal. The achievements of the previous industrial revolutions such as steam engines, electricity/lights, and computers have proven this. Only when DT is conveniently, economically and accurately applied can it create real value for industrial progress and social development. Thus, the paper is based on the DT theoretical framework and the technical route, focusing on the analysis of its application technology in the systematic design and engineering implementation. We take a new energy ship as the object of implementation, which is with many monitoring points, large data volume, high stability requirements, lack of industry standards and implementation plans. Through the study of cloud computing, open source software, marine control system and characteristics of new energy, we carry out feasibility analysis from the shipboard and cloud system architecture and implementation plan respectively, and completed the development, deployment and

online operation on the new energy ship. Finally, the DT data is used to correct the calculation deviation of the battery SOC (State of Charge) by the ship's physical system.

2. An overall architecture of DT engineering application

There are two main development directions for enterprise-led DT engineering applications. One is the development of infrastructure and software platforms, in which companies that own physical assets develop twins on their infrastructure and platforms, such as Siemens and Microsoft. The other is a physical asset-oriented design model, in which DT of specific asset objects is realized through data collection and model construction on the basis of existing software functions, such as Ansys. Also, some choose to develop in both directions, such as GE.

With the continuous upgrading of technologies, the theoretical and application requirements are constantly expanding and upgrading. The development and application of DT presents new requirements, including: (i) application field expansion, (ii) deep integration with New IT technology, (iii) data fusion with CPS, (iv) intelligent services, (v) universal industrial interconnection, (vi) model of dynamic multidimensional and multispacetime scale [6]. Combining these development requirements and the status of enterprise applications, an overall architecture of DT engineering application is shown in Fig 1.

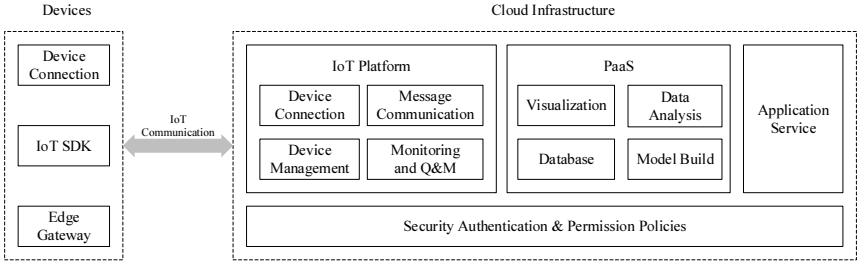


Figure 1. An overall architecture of DT engineering application

3. DT construction for propulsion system in new energy ships

So far, DT application technologies and engineering implementation for ships are still in their infancy, and there are very few cases and schemes that can be used for reference. China has vigorously promoted the development of new energy ships (NESs) from the perspective of energy security and ecological environmental protection, such as the "Guiding Opinions on Promoting the Development of Green Shipping in the Yangtze River Economic Belt" issued by the Ministry of Transport of the P.R.China [9], Ministry of Ecology and Environment of the P.R.China issued "Limits and Measurement Methods of Ship Engine Exhaust Pollutant Emissions" [10], "Inland Green Ship Regulations" issued by China Classification Society (CCS) [11], etc. The International Maritime Organization (IMO) also mandates a ship energy efficiency design index. The essence of inland green shipping is to change the form of propulsion from traditional diesel engine drive to electric propulsion or pure battery power,

especially the pure battery power, it is the main development direction of the industry, due to its advantages of economy, safety and convenience. Since the propulsion, control and sensing of pure battery-powered ships are driven by current and voltage, while the collection, transmission and storage of current and voltage data have higher accuracy, precision and real-time, it is natural to build a DT system. The DT systematic design for new energy ships is mainly for a pure electric system. The system consists of four major parts: (1) battery pack, (2) DCDC power supply, (3) DC power distribution system, (4) propulsion motor, which is shown in Fig 2.

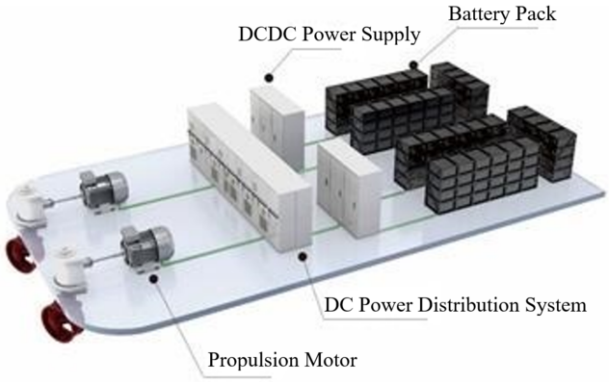


Figure 2. Major parts for pure electric system

A NES integrates with a high-safety lithium battery system, a high-efficiency variable frequency drive system, a high-mobility pod propeller, an intelligent ship handling and operation and maintenance system, and the mute effect reaches the luxury car level standard. The main design parameters are shown in Table 1 [12-13].

Table 1. Main design parameters of a NES

Design Parameters	Value
Total Length	53.20 Meters
Waterline Length	≈ 49.96 Meters
Total Width	13.40 Meters
Type Width	10.80 Meters
Type Depth	2.7 Meters
Design Draft	1.55 Meters
Displacement	410 Tons
Top Speed	10±0.2kn(6m Water Depth)
Endurance	8 Hours(10 Knots Speed)
Passenger	300P
Propulsion Power	2×200kW
Battery Capacity	2.24MWh

The NES is equipped with a 2240kWh LiFePO4 battery. After DC conversion, it outputs a stable DC voltage, and then the DC power distribution system outputs 380V power AC power and 220V daily AC power, which are used for propulsion system and shipboard equipment respectively. The ship's control system and data acquisition system are implemented by a variety of combined PLCs and supporting communication modules. The communication protocol mainly uses CANopen and Modbus TCP. The collected data is aggregated to the bridge industrial computer for local monitoring management. The monitoring system is shown in Fig 3.

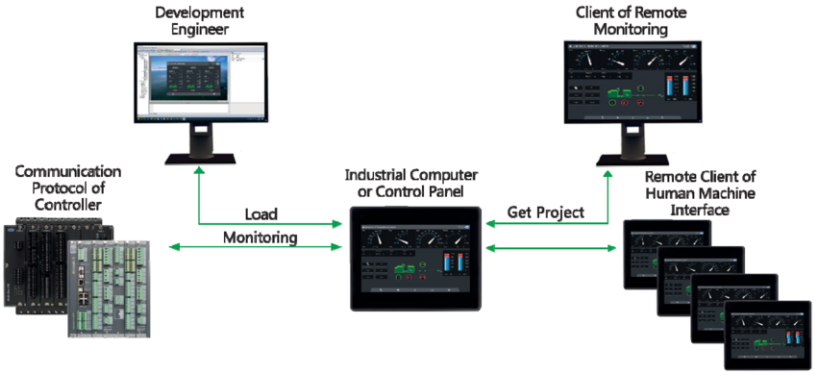


Figure 3. The monitoring system for a NES

Due to the fact that the capability of computing power, storage capacity, and application service of the shipboard industrial computer is not able to directly build a DT system, the cloud services must be used for core computing and storage. For the shipboard terminal, the client application developed must achieve both local monitoring and data upload to the cloud. Therefore, we believe that the most appropriate solution is based on Node.js development. Node.js is an open source and cross-platform JavaScript runtime environment, the software type is open source with cross-platform capabilities, the license agreement is a friendly MIT agreement. It runs the V8 JavaScript engine of the Google Chrome kernel outside the browser, and its application runs in a single process without the need to create a new thread for each program request. Node.js standard library provides a set of asynchronous I/O native functions to prevent JavaScript code from being blocked, and is written in a non-blocking paradigm, so that blocking behavior becomes an exception rather than a norm. When Node.js performs I/O operations, such as: reading data from the network, accessing the database or file system, the program will resume the operation when the response returns, instead of blocking the thread and wasting the CPU loop waiting, which makes Node.js handle thousands of concurrent connections on one computer without the burden of managing thread concurrency [14], this feature is very good at solving the blocking problem of shipboard industrial computer when processing with high concurrent connections.

The NES has about 2,000 collection points on the whole ship, and the relevant data of ship communication and navigation equipment such as depth sounder, ship automatic identification system (AIS), GPS is also connected to the industrial computer through the serial port (DB9), so that its concurrent connections present a larger magnitude. In addition, as the requirements for battery safety monitoring are further improved, battery monitoring point data will increase exponentially, and the stability and performance requirements for concurrent connections will also be higher, while Node.js has solved this problem in the framework design. The development tool uses Node-RED, which is an application developed on the basis of Node.js, it is an open source, dataflow-based visual programming tool and has been widely used in IIoT and control systems. It connects hardware devices, application interfaces, and online services together, provides a browser-based programming environment and rich node types, makes it very easy to create a process, and also provides a running environment for the operation of the process, has the one-click deployment capability. The shipboard terminal program developed under the Node.js framework using the Node-RED

development tool contains two main modules, of which the module pushing to the cloud is connected to the IoT platform through the MQTT protocol.

Table 2. Program module function on shipboard

Module Function	Function Description
■ Collect Data ■ Storage Data	To collect system equipment point data such as EMS, BMS, propulsion system, monitoring alarm system, pod device steering gear side control box, GPS, AIS and other system equipment points and store local database
■ Update Local Library ■ Push to Cloud	To collect system equipment point data such as EMS, BMS, propulsion system, monitoring alarm system, pod device steering gear side control box, GPS, AIS, etc. and push it to the cloud database

When data is connected to the Alibaba Cloud IoT platform, the most important indicator for product and parameter selection is system throughput, which is: TPS (Transactions per Second). The calculation formula is as follows:

$$tps = n_d \times q_u + n_d \times q_d \tag{1}$$

Where n_d is the number of access devices, q_u is the amount of messages reported by a single device per second, and q_d is the amount of messages sent by a single device per second. Consider that the NES is encapsulated and reported according to different message types, there are at least 4 types, and 1 report every 2 seconds, the uplink for 30 ships to connect at the same time is: $tps = 30 \times 4 \div 2 = 60$; consider the difference between ships, the initial configuration of the IoT platform is at least: the message downlink TPS is 100 messages/sec; the rule engine TPS is 100 messages/sec; the time sequence data written into the TPS is 5,000 messages/sec.

The choice of the database should fully consider the performance and stability under high throughput when the query and analysis business is paralleled with the exponential growth of the data volume. The storage space consumed by the average daily data volume of a single ship is as follows:

$$C = n \times b \times f \times t \div p \tag{2}$$

Where C is the storage space capacity, unit is byte (Byte); n is the number of acquisition points; b is the storage space required for a single point, and the ship data is mainly switching value and power electronic parameters; f is the sampling frequency, it is usually 1 time per second; t is the sampling time, unit is second (s); p is the reporting cycle, that is, the data uploaded to the cloud has undergone a certain down sampled. A high-performance database must be selected to support the normal throughput of twin data. We chose the Alibaba Cloud Hologres database, which belongs to the enterprise analytical data warehouse and has the ability to analyze and query massive big data in seconds. Hologres integrates real-time services and big data analysis scenarios, it is fully compatible with the PostgreSQL protocol and seamlessly

connects with the big data ecology, it can use the same data architecture to support real-time write real-time query and real-time offline federated analysis. Its execution engine is a general architecture that supports complex queries and high-performance real-time service queries, it has the following characteristics: (1) distributed execution model, (2) fully asynchronous execution, (3) vectorization and column processing, (4) adaptive incremental processing, (5) deep optimization of specific queries [15]. According to the design and implementation of this plan, DT data is stably collected, transmitted and stored. Also, a visual service is provided to remotely monitor real-time operating status as shown in Fig 4.

Figure 4. A visualization service from DT

4. DT applicaton exploration for new energy ships

In the NES, the most critical changes compared to diesel main engine-driven and diesel-electric ships are power batteries and BMS (Battery Management System). BMS is responsible for detecting battery working conditions, estimating battery SOC and battery health, completing the functions such as thermal management, charge and discharge control, data communication, balance detection, fault diagnosis and liquid crystal display. Among them, SOC is the most important parameter, it is the basis of everything else, if there is no accurate SOC, the battery will always be in a protected state, and no amount of protection functions can make the BMS work normally, nor can it prolong the service life of the battery. In addition, high-precision SOC estimation can effectively reduce the required battery cost.

The SOC estimation battery is based on the ampere-hour integration method, the algorithm and model are solidified in the BMS controller and calibrated in the factory state. Algorithms and models are related to many physical characteristics, such as: temperature, polarization effects, battery life, etc., and have non-linear characteristics, especially SOH (State of Health) will accumulate with the use of time, and eventually lead to the SOC error in the controller becomes bigger and bigger. Re-modeling calculations using twin data in the cloud, and iterative optimization using DT is an effective strategy to solve this problem. If the initial state of charge and discharge is recorded as SOC_0, then the SOC of the current state is calculated as:

$$SOC = SOC_0 - \frac{1}{Q_N}\int_0^t KId\tau \tag{3}$$

$$K = K_t \times \eta \tag{4}$$

Where Q_N is the rated capacity of the battery at standard temperature; I is the charge and discharge current; η is the charge and discharge efficiency, also known as the Coulomb efficiency; K is a temperature-related constant, and K_t is the correction coefficient, the calculation formula is:

$$K_t = 1 + 0.008(\mathrm{T}_a - \mathrm{T}) \tag{5}$$

Ta is the standard temperature at the rated capacity of the battery, and T is the real-time ambient temperature.

The output charge is calculated from 2021-07-24 19:27:36 to 2021-07-24 19:39:31, when the controller reports that SOC is 97.6%:

$$\Delta C = \int_0^t Id\tau \div 0.95 = 34800\,\mathrm{C} \tag{6}$$

The value of I used in the calculation here is the bus current of the battery system in the left compartment, that is, the real load current. Therefore, it needs to be divided by $\eta = 0.95$ to be equal to the discharge current of the battery system itself, and the total power of the battery system in the left compartment is:

$$Q_n \approx 1120kWh \times 97\% \times 60 \times 60 \div 640V \times 1000 \approx 6111000C \tag{7}$$

Where 97% is the DCDC conversion efficiency from the battery cluster to the left compartment battery system, and 640V is the bus voltage output by the left compartment battery system, the SOC change during this time period is:

$$\Delta SOC = 34800C \div 6111000C \approx 0.57\% \tag{8}$$

We use the DT system to calculate the SOC, where the current and voltage are the output data of the left compartment battery system, which represents the real load value, and the charging and discharging efficiency uses laboratory experience values, the results are shown in Fig 5.

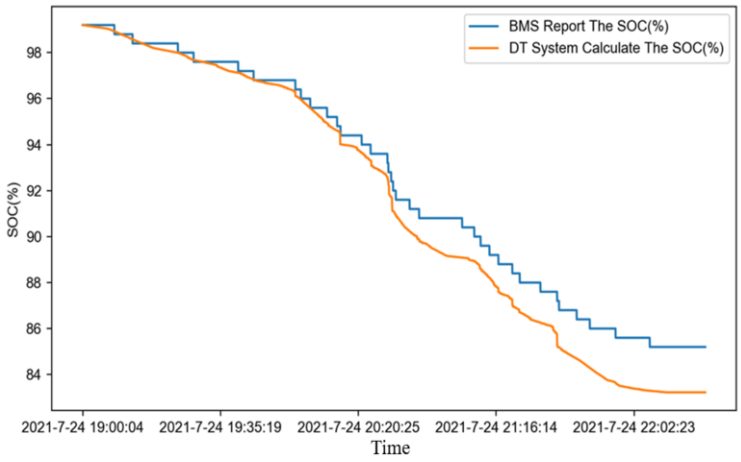


Figure 5. Comparison of SOC reported by BMS of the left compartment battery system and SOC of the DT system calculation

From the comparison of the calculation results, it can be seen that the calculated value of the DT system is continuously changing, which is basically consistent with the change of the battery's continuous output current, and conforms to the true physical characteristics of the battery. Meanwhile, it can be seen that as time increases and power consumption increases, the cumulative error of the SOC value reported by the BMS is gradually increasing.

5. Conclusion

In this paper, taking a new energy ship as the object, we summarize the definition, development status and engineering application of the DT technology, and present its overall architecture and key technical scheme. Based on this, an application exploration has been carried out in SOC estimation for BMS, and it is expected to perform a certain reference role in the design and development of new energy ships and inland green shipping.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (GrantNo.51905397).

References

[1] Grieves M. Digital twin: manufacturing excellence through virtual factory replication. J. White paper. 2014; 1: 1-7.
[2] Tuegel EJ, Ingraffea AR, Eason TG, et al. Reengineering aircraft structural life prediction using a digital twin. J. International Journal of Aerospace Engineering. 2011.
[3] Pera Corporation Ltd Digital Twin Lab 2019 Digital Twin Technology White Paper http://www.peraglobal.com/upload/contents/2019/12/20191230095610_31637

[4] Grieves M. Virtually perfect: driving innovative and lean products through product lifecycle management Space Coast Press. 2011.
[5] Tao F, Cheng Y, Cheng J, et al. Theories and technologies for cyber-physical fusion in digital twin shop-floor. J. Computer integrated manufacturing systems. 2017; 23(8): pp 1603-1611.
[6] Tao F, Liu W, Zhang M, et al. Five-dimension digital twin model and its ten applications. J. Computer integrated manufacturing systems. 2019; 25(1): pp 1-18.
[7] Wang K, Liu D, Liu Z, et al. An assembly precision analysis method based on a general part digital twin model. J. Robotics and Computer-Integrated Manufacturing. 2021; 68: p 102089.
[8] Deren L, Wenbo Y, Zhenfeng S. Smart city based on digital twins. J. Computational Urban Science. 2021; 1(1): pp 1-11.
[9] Shang H, Jiang S, et al. Guiding Opinions of the Ministry of Transport on Promoting the Development of Green Shipping in the Yangtze River Economic Zone. J. Gazette of the State Council of the People's Republic of China. 2018; 1: pp 74-78.
[10] Implementation Guidelines for Limits and Measurement Methods of Ship Engine Exhaust Pollutant Emissions (China Phase I and Phase II) (Rev.1). J. Ship Standardization Engineer. 2018; 51(06): p 12.
[11] China Classification Society. "Inland Green Ship Regulations (2020)". J. China Ship Survey. 2020; 01: p 104.
[12] Liu X, Wu Y. New technology for battery-powered passenger ships. J. Urban Public Transport.2020; 02: p 29
[13] Eworldship. The first large-scale all-electric Yangtze River cruise ship "Jun Lv Hao" made its maiden voyage http://www.eworldship.com/html/2020/OperatingShip_0624/160958.html. 2020.
[14] OpenJS Foundation Node.js https://nodejs.org/
[15] Jiang X, Hu Y, Xiang Y, et al. Alibaba hologres: a cloud-native service for hybrid serving/analytical processing. J. Proceedings of the VLDB Endowment. 2020; 13(12): pp 3272-3284.

doi:10.3233/ATDE220226

Structural Parameter Sensitivity Study of a Quadrupedal Bionic Horse Robot Driven by a Cam-Linkage Mechanism

Liangwen WANG [a,1], Yalei SHI [a], Liwei LI [a], Tianyun HE [b], Hongwei HAO [b], Jihao ZHANG [a] and Guizhong XIE [a]
[a] *School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Henan Provincial Key Laboratory of Intelligent Manufacturing of Mechanical Equipment, Zhengzhou 450002, China*
[b] *Henan Huanghe Whirlwind Co. LTD, Changge 461500, Henan, China*

Abstract. In order to determine the main parameters affecting the kinematic performance of the quadrupedal bionic horse robot, this paper investigates the sensitivity analysis of structural parameters on performance changes. Based on the introduction of the structure of the bionic horse robot used for equine-assisted therapy, the establishment of a kinematic model, and the kinematic analysis by the closed-loop vector method, the kinematic relationship between the trajectory of the foot end of the bionic horse and the structural parameters is calculated by full differential calculations, and the influence laws of the changes of the structural parameters on the change of the trajectory of the foot end of the bionic horse are analyzed, and the indexes for evaluating the sensitivity of the structural parameters are introduced, so that the key structural parameters with high sensitivity to the robot performance are obtained. The results are verified by simulation. This research lays the foundation for further optimization of the structural parameters of the robot.

Keywords. Quadruped robot, bionic horse, kinematic model, parameter sensitivity, key parameters

1. Introduction

The quadruped walking robots have been intensively studied by scientists [1,2]. As a nontraditional therapy, equine-assisted therapy offers horseback riding [3] that allows the patient to feel the movement of the pelvis and trunk, improve balance control, promote trunk extension, and increase the patient's endurance and cardiorespiratory fitness. [4]. Because of the high cost of horse-riding-assisted therapy, bionic equine therapy has been used as a relatively low-cost solution. In order to mimic the motion of a horse, the authors designed a bionic horse robot driven by a cam-link mechanism [5].

[1] Corresponding Author, Liangwen WANG, School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Henan Provincial Key Laboratory of Intelligent Manufacturing of Mechanical Equipment, Zhengzhou 450002, China; E-mail: w_liangwen@sina.com.

In order to optimize the performance of the bionic horse robot, we need to determine the main structural parameters that affect the performance of the robot. And in order to establish the relationship between the influence of structural parameter changes on the robot motion trajectory changes, we used the method of building a robot error model. Many scholars have conducted more in-depth research on mechanism errors. Fugui X et al. [6] established a new error mapping model of parallel mechanism based on the modeling method of spatial error vector chain; Ruiqin L et al. [7] carried out error simulation and analysis of 3-RSR parallel mechanism based of D-H matrix method; Yu R et al. [8] used the closed-loop vector loop method with error sources for error study of parallel mechanism. In the above literature, most of the scholars have conducted error modeling and influence factor analysis studies on parallel mechanisms, but the bionic horse robot studied in this paper involves a cam-linkage mechanism, and it is rare to conduct error modeling studies on robots with such combination mechanisms.

In this paper, the structural parameter sensitivity of the robot is investigated by modeling the error relationship between the trajectory of the foot end of the bionic horse and the structural parameters in order to have the key structural parameters with high sensitivity to the robot performance.

2. Kinematic modeling and analysis of a bionic horse robot driven by a cam-linkage mechanism

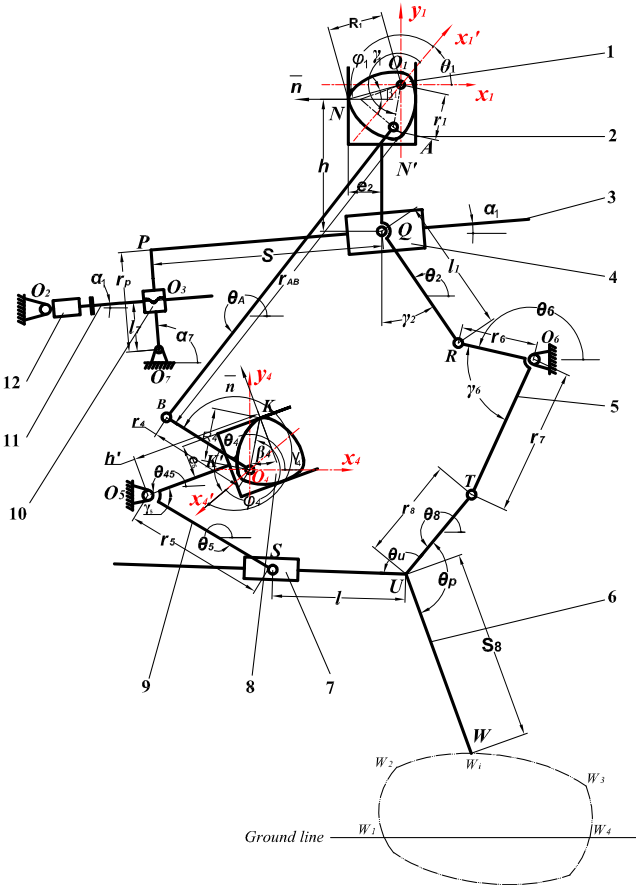
The bionic horse robot was designed and developed with a single leg motion driven by a cam-linkage mechanism [5]. The schematic diagram of the mechanism is shown in Figure 1.

The structure includes: (1) The stride length adjusting mechanism $O_1O_7O_3O_2O_1$; (2) The stride length mechanism $O_1NN'QRO_6O_1$; (3) The four-bar linkage mechanism $O_1ABO_4O_1$; (4) The lifting mechanism $O_1O_4KK'O_5O_1$; (5) The leg mechanism $O_1O_5SUTO_6O_1$. (6) The constant-breadth three-center cam is shown in Figure 2, N is the contact point between the cam and the follower, φ is the input angle, β is the output angle, and R is the radius vector, we can determine the correspondence between the output angle β and the radius vector R and the input angle φ.

From the closed-loop $O_1O_7O_3O_2O_1$, the kinematic relationship is:

$$\left(X_{O_7} - X_{O_2}\right)\cos\alpha_1 + \left(Y_{O_7} - Y_{O_2}\right)\sin\alpha_1 = l_0 \pm s\frac{\theta_3}{2\pi} \tag{1}$$

where, l_0 is the initial length of the lead screw, θ_3 is the rotation angle of the motor for stride length adjusting, and α_1 is the angle between the lead screw and the horizontal direction.



1 the stride length cam, 2 the stride length fork, 3 the rectangular swing rod, 4 the short slider, 5 the connecting rod, 6 the walking leg, 7 the long slider, 8 the stride height cam, 9 the stride height fork, 10 the nut slider, 11 the lead screw, 12 the step length adjustment motor.

Figure 1. The schematic of the single leg walking mechanism.

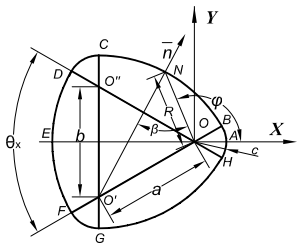


Figure 2. The constant-breadth three-center cam.

From the closed-loop $O_1O_7PQO_6O_1$ and the closed-loop $O_1NN'QPO_7O_1$, we have:

$$\begin{cases} X_{O_7} + r_p \cos(\alpha_1 + \pi/2) + S\cos\alpha_1 + l_1\cos\theta_2 = X_{O_6} + r_6\cos\theta_6 \\ Y_{O_7} + r_p \sin(\alpha_1 + \pi/2) + S\sin\alpha_1 + l_1\sin\theta_2 = Y_{O_6} + r_6\sin\theta_6 \end{cases} \tag{2}$$

$$\begin{cases} R_1\cos(\theta_1+\varphi_1) + h\cos(\theta_1+\varphi_1+\beta_{10}+\pi/2) + e_2\cos(\theta_1+\varphi_1+\beta_{10}+\pi) \\ = X_{O_7} + r_p\cos(\alpha_1+\pi/2) + S\cos\alpha_1 \\ R_1\sin(\theta_1+\varphi_1) + h\sin(\theta_1+\varphi_1+\beta_{10}+\pi/2) + e_2\sin(\theta_1+\varphi_1+\beta_{10}+\pi) \\ = Y_{O_7} + r_p\sin(\alpha_1+\pi/2) + S\sin\alpha_1 \end{cases} \tag{3}$$

where $\beta_{10} = -\beta_1 SIGN(\varphi_1)$ and $\theta_2 = \pi/2+\theta_1+\beta_{10}+\varphi_1+\gamma_2$, and θ_1 is the motor input angle.

From the closed-loop $O_1ABO_4O_1$, we have:

$$\begin{cases} r_1\cos(\theta_1+\gamma_1) + r_{AB}\cos\theta_A = X_{O_4} + r_4\cos(\theta_4+\gamma_4) \\ r_1\sin(\theta_1+\gamma_1) + r_{AB}\sin\theta_A = Y_{O_4} + r_4\sin(\theta_4+\gamma_4) \end{cases} \tag{4}$$

From the closed-loop $O_1O_4KK'O_5O_1$, we have:

$$\begin{cases} X_{O_4} + R_4\cos(\theta_4+\varphi_4) + h'\cos(\theta_4+\varphi_4+\beta_{40}+\pi/2) \\ +e_5\cos(\theta_4+\varphi_4+\beta_{40}+\pi) = X_{O_5} \\ Y_{O_4} + R_4\sin(\theta_4+\varphi_4) + h'\sin(\theta_4+\varphi_4+\beta_{40}+\pi/2) \\ +e_5\sin(\theta_4+\varphi_4+\beta_{40}+\pi) = Y_{O_5} \end{cases} \tag{5}$$

where $\beta_{40} = -\beta_4 SIGN(\varphi_4)$.

From the closed-loop $O_1O_5SUTO_6O_1$, we have:

$$\begin{cases} X_{O_5} + r_5\cos\theta_5 = X_{O_6} + r_7\cos(\theta_6+\gamma_6) + r_8\cos\theta_8 + l\cos(\theta_U+\theta_8-\pi) \\ Y_{O_5} + r_5\sin\theta_5 = Y_{O_6} + r_7\sin(\theta_6+\gamma_6) + r_8\sin\theta_8 + l\sin(\theta_U+\theta_8-\pi) \end{cases} \tag{6}$$

where $\theta_5 = 3\pi/2+\theta_4+\beta_{40}+\varphi_4-\gamma_5$.

See Figure 1, the coordinates of the leg end-point W of the legs are obtained:

$$\begin{cases} X_W = X_{O_6} + r_7\cos(\theta_6+\gamma_6) + r_8\cos\theta_8 + S_8\cos(\pi+\theta_8-\theta_P) \\ Y_W = Y_{O_6} + r_7\sin(\theta_6+\gamma_6) + r_8\sin\theta_8 + S_8\sin(\pi+\theta_8-\theta_P) \end{cases} \tag{7}$$

3. Sensitivity analysis of parameters affecting the output performance of the mechanism

In a mechanism, the trajectory can be expressed as a function related to input variables and structural parameters. Suppose the trajectory output of the mechanism is expressed

as Φ, then we have:

$$\Phi = F(\theta, P_1, P_2, ..., P_n) \tag{8}$$

where θ is the driving displacement; $P_i(i = 1, 2, ..., n)$ is the structural parameter.

In a mechanism, there may be errors for θ and $P_i(i = 1, 2, ..., n)$. Expanding Eq. (8) and omitting the higher-order terms, we have:

$$\Delta\phi = C_0 \Delta\theta + \sum_{i=1}^{n} C_i \Delta P_i \tag{9}$$

where $C_0 = \partial F / \partial \theta$, $C_i = \partial F / \partial P_i (i = 1, 2, ..., n)$.

In most cases, when the active part moves as required, take $\Delta\theta = 0$. The influence coefficients (or sensitivity coefficients) C_0 、 $C_i(i = 1, 2, ..., n)$ of the input of the mechanism and the structural dimensional error of the mechanism on the output result error indicate the relative importance and influence of each component scale of the mechanism on the output scale. The key is to find the intrinsic connection between the variation of the influence coefficient and the influence of each structural scale on the output requirements.

As the input angle changes, we express the change in the sensitivity coefficient over a cycle as a curve, called the sensitivity coefficient curve. To assess the sensitivity of the variation of each structural parameter to the output error, we introduce the efficacy coefficient. The so-called efficacy coefficient is the degree of sensitivity of all parameters to an output indicator, and scores each parameter. i.e. $d = d(C_i)$ (i=1,2,…, ρ), where C: the sensitivity value of a parameter to the j requirement of the output, it can be obtained by analyzing the sensitivity coefficient curve and the output requirements of the mechanism, the calculation is performed by taking the absolute value of the sensitivity value; d: score of the sensitivity of the evaluation parameter to the jth requirement of the output; ρ: the number of items (i.e., scoring items) for the output indicators. The relationship between C and d is as follows:

$$d = (C - C_{\min}) / (C_{j\max} - C_{\min}) \tag{10}$$

In order to obtain the sensitivity of each parameter to the overall performance of the output, we weight the scores of the parameters in the sensitivity score to the individual performance according to the relative importance of each index. That is, the efficacy D of the parameter in affecting the overall performance of the output is obtained, expressed by Eq:

$$D = \sum_{j=1}^{\rho} \rho_i \alpha_i \tag{11}$$

where ρ: the weighting factor of the corresponding item indicator introduced according to the relative importance of each individual requirement; the size of ρ is determined

according to the specific situation, combined with the experience of production practice and the requirements of the specific mechanism on the output.

Based on the above analysis method, we investigate the single leg mechanism. Through a series of derivations and synthesizing the relevant expressions, we have:

$$\begin{aligned}\Delta X_W = {} & C_{xxo7}\Delta X_{O_7} + C_{xx\theta 6}\Delta X_{o6} + C_{xxo5}\Delta X_{O_5} + C_{xxo4}\Delta X_{O_4} + C_{xxo7}\Delta Y_{O_7} \\ & + C_{xyo6}\Delta Y_{O_6} + C_{xyo5}\Delta Y_{O_5} + C_{xyo4}\Delta Y_{O_4} + C_{xr1}\Delta r_1 + C_{xr4}\Delta r_4 + C_{xr5}\Delta r_5 \\ & + C_{xr6}\Delta r_6 + C_{xr7}\Delta r_7 + C_{xr8}\Delta r_8 + C_{xrAB}\Delta r_{AB} + C_{xs8}\Delta S_8 + C_{xl1}\Delta l_1 + C_{xrp}\Delta r_p \\ & + C_{xe2}\Delta e_2 + C_{xe5}\Delta e_5 + C_{x\theta 1}\Delta\theta_1 + C_{x\theta 6}\Delta\theta_6 + C_{x\alpha 1}\Delta\alpha_1 + C_{x\theta U}\Delta\theta_U \\ & + C_{x\theta P}\Delta\theta_P + C_{x\gamma 1}\Delta\gamma_1 + C_{x\gamma 2}\Delta\gamma_2 + C_{x\gamma 4}\Delta\gamma_4 + C_{x\gamma 5}\Delta\gamma_5 + C_{x\gamma 6}\Delta\gamma_6 \\ & + C_{xa1}\Delta A_1 + C_{xb1}\Delta B_1 + C_{xc1}\Delta C_1 + C_{xa4}\Delta A_4 + C_{xb4}\Delta B_4 + C_{xc4}\Delta C_4\end{aligned} \tag{12}$$

$$\begin{aligned}\Delta Y_W = {} & C_{yxo7}\Delta X_{O_7} + C_{yx\theta 6}\Delta X_{o6} + C_{yxo5}\Delta X_{O_5} + C_{yxo4}\Delta X_{O_4} + C_{yxo7}\Delta Y_{O_7} \\ & + C_{yyo6}\Delta Y_{O_6} + C_{yyo5}\Delta Y_{O_5} + C_{yyo4}\Delta Y_{O_4} + C_{yr1}\Delta r_1 + C_{yr4}\Delta r_4 + C_{yr5}\Delta r_5 \\ & + C_{yr6}\Delta r_6 + C_{yr7}\Delta r_7 + C_{yr8}\Delta r_8 + C_{yrAB}\Delta r_{AB} + C_{ys8}\Delta S_8 + C_{yl1}\Delta l_1 + C_{yrp}\Delta r_p \\ & + C_{ye2}\Delta e_2 + C_{ye5}\Delta e_5 + C_{y\theta 1}\Delta\theta_1 + C_{y\theta 6}\Delta\theta_6 + C_{y\alpha 1}\Delta\alpha_1 + C_{y\theta U}\Delta\theta_U \\ & + C_{y\theta P}\Delta\theta_P + C_{y\gamma 1}\Delta\gamma_1 + C_{y\gamma 2}\Delta\gamma_2 + C_{y\gamma 4}\Delta\gamma_4 + C_{y\gamma 5}\Delta\gamma_5 + C_{y\gamma 6}\Delta\gamma_6 \\ & + C_{ya1}\Delta A_1 + C_{yb1}\Delta B_1 + C_{yc1}\Delta C_1 + C_{ya4}\Delta A_4 + C_{yb4}\Delta B_4 + C_{yc4}\Delta C_4\end{aligned} \tag{13}$$

The sensitivity coefficient curves for each parameter at position α_1=14.5° are plotted according to Eqs. (12) and (13). Taking the angular parameters of θ_U, θ_P, γ_1, γ_5, and γ_6 as an example, its sensitivity coefficient curves for ΔX_W and ΔY_W are shown in Figures. 3(a) and 3(b).

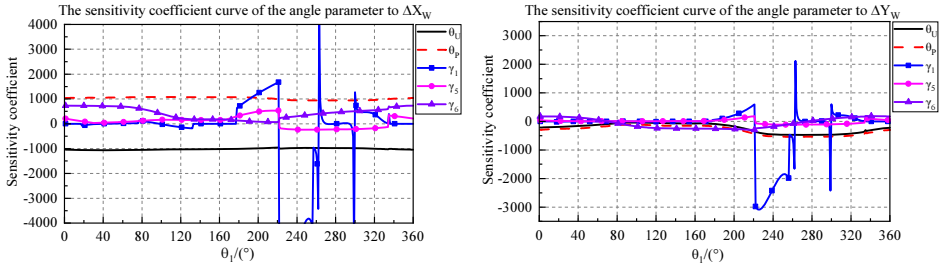


Figure 3. The sensitivity coefficient curve.

The typical trajectory of the foot end of the bionic horse is shown in Figure 4.

The basic requirements are as follows: (1) reduce leg lifting and increase body movement; (2) in the section CF, require flat and straight; (3) in the section BAG, require change smoothly, have a certain height, and avoid the sharp point; (4) in the section BC and section FG, require a certain height to meet the patient's rehabilitation training requirements; (5) maintain a certain distance between section BG.

The magnitude of the area between the sensitivity coefficient curve and the x-axis is defined as the absolute sensitivity quantity, and the difference between the highest and lowest values of the sensitivity coefficient curve in a certain segment is defined as the relative sensitivity quantity. Eleven sensitivity quantities are selected from the sensitivity coefficient curves for each parameter.

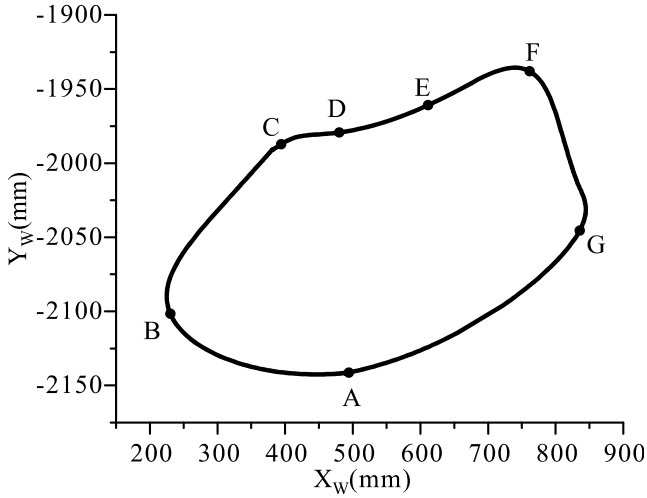


Figure 4. The typical leg-end trajectory.

A_1: the sensitivity of the parameter to the y-axis coordinates of the point A of the trajectory, determining the starting point of the trajectory; A_2: the sensitivity of the parameter to the straightness of the trajectory within the segment CF; A_3: the sensitivity of the parameter to the average height within the segment CF; A_4: the sensitivity of the parameter to the displacement in the y-axis direction of the trajectory in general, with C_{1ymax} indicating the highest value of the sensitivity coefficient within the 150°-170° segment and C_{2ymin} indicating the lowest value of the sensitivity coefficient within the 170°-190° segment, and $C_{3ymax}=C_{1ymax}-C_{2ymin}$ reflecting the sensitivity of the parameter to the amount of change in the displacement in the y-axis direction of the trajectory in general; B_1: the sensitivity of the parameter to the x-axis coordinates of the trajectory point A; B_2: the sensitivity of the parameter to the leg span $|X_B-X_G|$; B_3: the sensitivity of the parameter to the positions of the elevated segments BC and FG; $B_4(B_5)$: the sensitivity of the parameter to the rate of change of the trajectory along the x-axis (y-axis) direction, the average value of the sensitivity coefficients in θ_1=150°-170°, θ_1=170°-190° and θ_1=190°-210° is taken; C_1: the sensitivity of the parameter to the span of BC section $|X_B-X_C|$; C_2: the sensitivity of the parameter to the span of DA section $|X_F-X_G|$.

In order to reflect the degree of influence of each structural parameter on the overall performance of the leg end-point trajectory, the above parameters are weighted according to the degree of influence of each index requirement on the overall performance of the mechanism, and this score is used as an important index to evaluate the sensitivity of each parameter. That is to take P_{A_1}, P_{A_2}, P_{A_3}, P_{A_4}, P_{B_1}, P_{B_2}, P_{B_3}, P_{B_4}, P_{B_5}, P_{C_1}, P_{C_2} eleven weighting factors. Then the efficacy of each parameter on the overall performance of the leg end-point trajectory is:

$$\begin{aligned} C = {} & P_{A_1}C_{A_1} + P_{A_2}C_{A_2} + P_{A_3}C_{A_3} + P_{A_4}C_{A_4} + P_{B_1}C_{B_1} + P_{B_2}C_{B_2} \\ & + P_{B_3}C_{B_3} + P_{B_4}C_{B_4} + P_{B_5}C_{B_5} + P_{C_1}C_{C_1} + P_{C_2}C_{C_2} \end{aligned} \tag{14}$$

4. Example of calculation

We chose α_1=14.5° to analyze the effect of each structural parameter on the output of the single legged mechanism of the quadruped bionic horse robot. According to the specific requirements of the actual bionic horse, we think that: A_2, A_4, B_1, B_2, B_3 and C_2 can reflect the sensitive amount of close relationship with the trajectory performance, A_1, A_3, B_4 (B_5) and C_1 can take the second place. After a comprehensive evaluation, the weighting factor was taken as $P_{A_2}=P_{A_4}=P_{B_1}=P_{B_2}=P_{B_3}=P_{C_2}$ =0.8, $P_{A_1}=P_{A_3}=P_{B_4}=P_{C_1}$ = 0.6. According to Eq (14), we calculate the sensitive amount as well as the efficacy value of each parameter, as shown in Table 1.

Table 1. The sensitivity of each parameter to the trajectory change of the mechanism at α_1=14.5°.

Parameters	Y_{O_6}	X_{O_6}	r_5	r_7	r_8	r_P	Y_{O_7}	l_1	r_1
Sensitive amount	1	0.3486	0.2644	0.2637	0.2517	0.8392	0.8193	0.7094	0.4016
Efficacy value	1	0.9429	0.9342	0.9316	0.8891	0.8410	0.8211	0.6907	0.4197
Parameters	r_4	r_6	r_{AB}	X_{O_4}	α_1	θ_P	θ_U	γ_5	γ_1
Sensitive amount	0.3297	0.3397	0.2478	0.1317	1	0.3106	0.2857	0.2103	0.1548
Efficacy value	0.3584	0.2683	0.2669	0.2565	1	0.3104	0.2829	0.2240	0.2100
Parameters	Y_{O_4}	Y_{O_5}	A_1	C_1	A_4	X_{O_7}	e_2	X_{O_5}	S_8
Sensitive amount	0.2329	0.2314	0.1860	0.2111	0.1539	0.1828	0.1575	0.1343	0.1470
Efficacy value	0.2519	0.2507	0.2402	0.2218	0.1890	0.1820	0.1715	0.1563	0.1439
Parameters	B_1	B_4	e_5	C_4	γ_6	γ_4	θ_1	θ_6	γ_2
Sensitive amount	0.1173	0.0775	0.0775	0.0779	0.0898	0.0964	0.1514	0.0700	0.0518
Efficacy value	0.1287	0.0946	0.0944	0.0944	0.1555	0.1460	0.1376	0.0745	0.0639

From the table 1, we have that the changes of parameters Y_{O6}, $r5$, $r7$, $r8$, γ_1, γ_5, γ_6, θ_p and θ_U have a greater effect on the trajectory. The rest of the parameters have less influence on the trajectory. We can combine this result to select the optimization variables. Some leg end-point trajectories for changing different parameters are shown in Figure 5. (Here, only changing different parameters Y_{O6}, $r5$, $r7$ are given.)

(1) Y_{O6} from -511mm to -556mm, the result is as in Figure 5(a); (2) $r5$ from 331mm to 376mm, the result is as in Figure 5(b); (3) $r7$ from 449mm to 494mm, the result is as in Figure 5(c). Obviously, the nine parameters mentioned above, have a greater impact on the leg end-point trajectory.

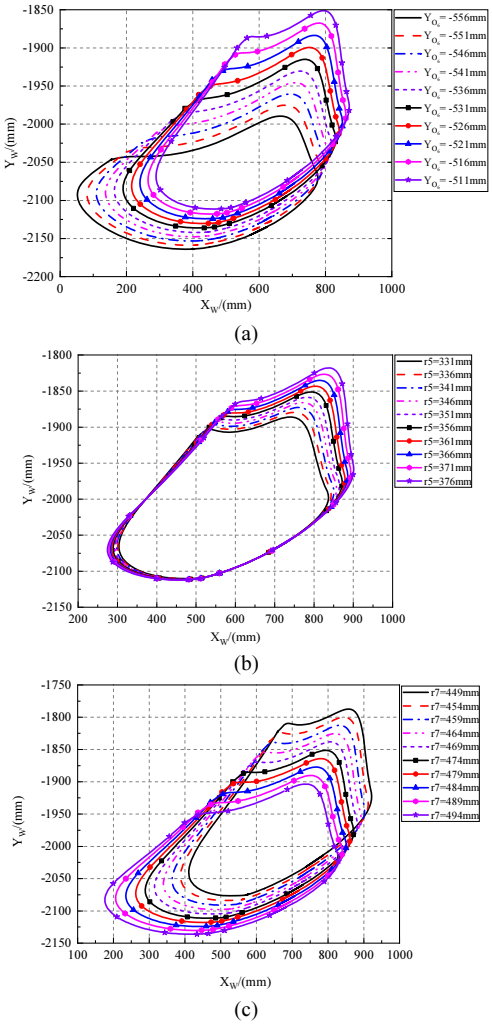


Figure 5. The leg end-point trajectory for changing different parameters.

5. Conclusion

Based on the kinematic model of the bionic horse, researching the influence law of the structural parameters changes on the foot end trajectory changes is carried, and key structural parameters with high sensitivity to the robot performance are obtained by introducing the indexes for evaluating the sensitivity of the structural parameters, and the calculation results are verified through simulation.

Funding

This project was supported by National Natural Science Foundation of China (Grant No.52075500, 52005453), the Key Science and Technology Research Project of the

Colleges and Universities of Henan Province, China (21A460034), and Key Scientific Project of Henan Province (211110220200), China.

References

[1] Wooden D, Malchano M, Blankespoor K, Howardy A, Raibert M. Autonomous Navigation for BigDog. Robotics and Automation (ICRA).2010 IEEE International Conference on Robotics and Automation; 2010 May 3-8; Anchorage, AK2010. pp 4736-4741.

[2] Guobiao W, Diansheng C, Kewei C, Ziqiang Z. The current research status and development strategy on biomimetic robot. Journal of Mechanical Engineering. 2015; 51(13): pp 27-44.

[3] Faber M, Schamhardt H, Weeren René Van, Johnstont C, Roepstorff L, Barneveld A. Basic three-dimensional kinematics of the vertebral column of horses walking on a treadmill. American Journal of Veterinary Research. 2001; 62(5): pp 757-64.

[4] Thais, Borges, de, Araújo, Ricardo, Jacó. Effects of hippotherapy on mobility, strength and balance in elderly. Archives of Gerontology & Geriatrics. 2013; 56(3): pp 478-481.

[5] Liangwen W, Weiwei Z, Caidong W, Fannian M, Wenliao D, Tuanhui W. Conceptual Design and Computational Modeling Analysis of a Single-Leg System of a Quadruped Bionic Horse Robot Driven by a Cam-Linkage Mechanism.Applied Bionics and Biomechanics. 2019; 2019: pp 1-13.

[6] Fugui X, Xinjun L, Yuzhen C. Error sensitivity analysis of novel virtual center mechanism with parallel kinematics. Journal of Mechanical Engineering. 2013; 49(17): pp 85-91.

[7] Ruiqin L, Bin Y. Error modeling of 3-RSR parallel robot based on D-H transformation matrix. Journal of Measurement Science and Instrumentation. 2014; (3): pp 53-59.

[8] Yu R, Shuangyong L, Yong H. Error model and verification of three degrees of freedom parallel mechanical leg on hexapod drilling robot. Transactions of the Chinese Society of Agricultural Engineering. 2016; 32(15): pp 18-25.

doi:10.3233/ATDE220227

Registration of Visible Image and Three-Dimensional Point Cloud Based on Point Features

Qianyi LU [a] and Lizuo JIN [a,1]
[a] *School of Automation, Southeast University, Nanjing, China*

Abstract. The accurate registration of visible image and three-dimensional point cloud is the prerequisite of image fusion. How to improve the registration accuracy is the focus of current research. In this paper, a registration algorithm of visible image and three-dimensional point cloud based on point features is proposed. Firstly, the visible image is preprocessed to correct the distortion of the camera. Then, the appropriate self-build calibration field is set, and the homonymous feature points in visible image and point cloud are obtained by using the blob detection algorithm. Finally, the corresponding registration model parameters of visible image and point cloud are calculated based on multiple pairs of homonymous feature points. Experimental results show that this method can achieve centimeter level registration accuracy.

Keywords. visible image, three-dimensional point cloud, distortion correction, image registration

1. Introduction

With the advent of the big data era, image sensor technology continues to develop, generating a large number of image data for target detection, positioning, recognition and other work [1-3]. However, due to the limitations of single sensor imaging mechanism, it is often necessary to fuse multi-channel sensor data to obtain more valuable information. The accurate registration among different sensors is the prerequisite for image fusion operation.

Three-dimensional laser scanning technology is an emerging technology with rapid development in recent years. Based on the principle of laser ranging, it realizes the rapid acquisition of three-dimensional position information of objects and establishes the point cloud model of objects [4]. Laser scanning technology has the characteristics of strong robustness, non-contact and large field of vision. Compared with visible imaging technology, it is not easy to be disturbed by environmental factors such as temperature and illumination. But at the same time, compared with the visible image, the three-dimensional point cloud data has lower resolution and lacks the color information of the objects [5]. Three-dimensional laser scanning technology and visible imaging technology have their own advantages and limitations. The registration and

[1] Lizuo JIN, School of Automation, Southeast University, Nanjing, China; E-mail: jinlizuo@qq.com.

fusion of visible image and three-dimensional point cloud can realize the complementary advantages of the two technologies.

Removing camera distortion is the basis of all visual positioning technologies. At present, Zhang Zhengyou calibration method is often used to obtain the distortion parameters of the camera. But the premise of this method is that multiple groups of template images must be taken from different angles for subsequent parameter calculation, which may be difficult to meet in engineering practice. This paper adopts a more general distortion correction algorithm, which only needs a single image. Firstly, an appropriate camera distortion model is defined. Then, according to the distortion of multiple lines in the world coordinate system in the image, the camera distortion model parameters can be obtained, which makes the distortion correction of the camera realized.

The key of registration between visible image and three-dimensional point cloud is to clarify the relative position relationship between laser scanner coordinate system and camera coordinate system. Then the corresponding transformation matrix can be found according to this relationship. Several pairs of homonymous feature points are needed to calculate the corresponding transformation matrix. At present, the commonly used feature points matching algorithms are mainly based on SIFT feature [6] and SURF feature [7]. However, these algorithms are highly dependent on the environment and are easily affected by illumination and scene. These problems directly affect the registration accuracy. Aiming at the problem of low quality of homonymous feature points, this paper artificially builds an appropriate calibration field. In this self-build calibration field, easily detected feature points are provided, which help the algorithm to realize the automatic calculation of registration parameters and ensure the registration accuracy.

2. Image preprocessing

Fish-eye lens is a kind of ultra-wide angle lens [8]. Compared with ordinary lenses, fish-eye lens has a larger field of view. But at the same time, the camera distortion of fish-eye camera will be more serious. According to the imaging principle of fish-eye lens [9], the straight line in the world coordinate system will show the shape of curve in the camera imaging picture due to the existence of barrel distortion. Therefore, it is necessary to obtain appropriate distortion parameters to make the curve show the shape of straight line on the corrected image.

In this paper, the camera distortion model in [10] is adopted, and its specific manifestation is

$$\begin{pmatrix} x_u - x_0 \\ y_u - y_0 \end{pmatrix} = \frac{1}{1+\lambda[(x_d - x_0)^2 + (y_d - y_0)^2]} \begin{pmatrix} x_d - x_0 \\ y_d - y_0 \end{pmatrix} \tag{1}$$

where λ is the distortion coefficient, (x_0, y_0) is the distortion center, (x_d, y_d) is the point coordinates on the original image and (x_u, y_u) is the point coordinates on the undistorted image. To calculate the distortion coefficient and distortion center, the corrected linear equation can be defined as

$$Ax_u + By_u + C = 0 \tag{2}$$

where A, B, C are the equation coefficients. The curve equation before correction is

$$x_d^2 + y_d^2 + Dx_d + Ey_d + F = 0 \tag{3}$$

where D, E, F are the equation coefficients. The algebraic relationship between distortion coefficient and distortion center is obtained by simultaneous Eqs. (1), (2) and (3):

$$x_0^2 + y_0^2 + Dx_0 + Ey_0 + F = \frac{1}{\lambda} \tag{4}$$

It can be seen that in order to obtain the camera distortion coefficient and distortion center, it is necessary to find out three groups of fitting circle parameters of straight lines in camera imaging picture. In this paper, marker straight lines are set artificially, and the circle parameters is fitted by the least square method. The specific steps are as follows:

1. Set the marker straight lines. Select three long straight lines and place them at the position with serious distortion in the camera imaging picture;
2. Extract the coordinates of multiple points on each straight line, and calculate the center and radius of its fitting circle respectively;
3. Simultaneous equations and substituting parameters into Eq. (4) to obtain the distortion center and distortion coefficient;
4. According to the transformation relationship in Eq. (1), the distortion corrected image can be calculated.

3. Feature extraction

The key to realize the registration of visible image and three-dimensional point cloud is to obtain high-quality homonymous feature points in laser scanner coordinate system and camera coordinate system. Aiming at the problem that the traditional feature matching algorithm cannot find high-quality homonymous feature points, this paper uses the form of self-built calibration field to assist in feature matching.

In the self-built calibration field, it is necessary to place multiple discs with different colors and heights in the observation area. The number of discs should be set to more than six, but the maximum number is unlimited. Each disc number corresponds to a unique RGB value and height value. The disc attribute table is established by using disc number, color attribute and height attribute. The disc centroid coordinates on the visible image and point cloud are extracted respectively, and can be matched to the corresponding disc number through color attribute and height attribute.

The centroid detection in visible image can be realized by blob detection algorithm. A blob is an area with a significant gray difference from the surrounding area. Compared with a single corner point, blob has stronger stability and anti-interference. Commonly used blob detection algorithms include derivative based differential method and watershed algorithm based on local extremum. The principle of differential method

is to convolute the image with the filter kernel. The closer the blob in the image is to the blob model, the greater the convolution response value is. Thus, the regions with the same characteristics as the filter core are found. Such algorithms usually combine scale space to realize multi-scale detection. At present, LOG algorithm, DOH algorithm, DOG algorithm and so on are commonly used. The principle of watershed algorithm is to first binarize the original image with multiple sets of thresholds, so as to obtain a series of binary images. Then, the center coordinates of the contour can be calculated according to the connected region on each binary image. After all the center coordinates are counted, the center point within the threshold distance is considered to be the same blob feature. And then the spot center coordinate set of the image can be constructed. Finally, these blobs are filtered according to the attributes of the target to be detected. The common filtering attributes mainly include color, area, roundness, eccentricity and convexity.

Watershed algorithm is used in this paper. Thus, the characteristic parameters can be set in advance according to the attributes of the calibration object to reduce the influence of interfering objects in the scene and ensure the reliability of the calculated value of the centroid. After obtaining the blob center coordinates, the RGB value at the coordinate position is corresponding to the disc RGB value, and the disc attribute table is searched to obtain the corresponding relationship between the disc number and the two-dimensional coordinates of the feature points.

The centroid detection in three-dimensional point cloud can be realized by using the projection of three-dimensional point cloud. Firstly, the ground and other interfering objects in the point cloud are filtered according to the height value. And then project the three-dimensional point cloud to the z axis direction to obtain the two-dimensional orthophoto image in the x axis and y axis direction. Similarly, the disc centroid (X', Y') on the projection map is solved by blob detection, and then back project (X', Y') onto the point cloud to obtain the corresponding three-dimensional point cloud coordinates (X, Y, Z). The Z value corresponds to the disc height value, and the circle attribute table is searched to obtain the corresponding relationship between the disc number and the three-dimensional coordinates of the feature points.

4. Registration algorithm

The world coordinate system is established on the laser scanner. The camera coordinate system has transformation relationships such as displacement and rotation for the world coordinate system. The transformation relationship is related to the internal and external parameters of the camera. The transformation relationship from the world coordinate system to the camera coordinate system can be expressed as

$$\begin{bmatrix} \lambda x \\ \lambda y \\ \lambda \end{bmatrix} = K \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{5}$$

where K is the camera internal parameter matrix, $\begin{bmatrix} R & t \end{bmatrix}$ is the optical system external parameter matrix, $P = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix}$ is the registration matrix with the size of 3*4, (X, Y, Z) is the three-dimensional point cloud coordinates, and (x, y) is the two-dimensional visible image coordinates. Let $\mathrm{x}_i = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$, $\mathrm{X}_i = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$, $\mathrm{x}_i \times P\mathrm{X}_i = 0$ can be obtained from $\lambda \mathrm{x}_i = P\mathrm{X}_i$, which as same as $\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \times \begin{bmatrix} P_1^T \mathrm{X}_i \\ P_2^T \mathrm{X}_i \\ P_3^T \mathrm{X}_i \end{bmatrix} = 0$, expand it to obtain

$$\begin{bmatrix} 0 & -\mathrm{X}_i^T & y_i \mathrm{X}_i^T \\ \mathrm{X}_i^T & 0 & -x_i \mathrm{X}_i^T \\ -y_i \mathrm{X}_i^T & x_i \mathrm{X}_i^T & 0 \end{bmatrix} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} = 0 \tag{6}$$

Any two terms in Eq. (6) are independent of each other. So, the first two lines can be selected to construct the equation. Multiple pairs of characteristic points can be used to establish Eq. (7).

$$\begin{bmatrix} 0^T & \mathrm{X}_1^T & -y_1 \mathrm{X}_1^T \\ \mathrm{X}_1^T & 0^T & -x_1 \mathrm{X}_1^T \\ \cdots & \cdots & \cdots \\ 0^T & \mathrm{X}_n^T & -y_n \mathrm{X}_n^T \\ \mathrm{X}_n^T & 0^T & -x_n \mathrm{X}_n^T \end{bmatrix} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} = 0 \tag{7}$$

Let $M = \begin{bmatrix} 0^T & \mathrm{X}_1^T & -y_1 \mathrm{X}_1^T \\ \mathrm{X}_1^T & 0^T & -x_1 \mathrm{X}_1^T \\ \cdots & \cdots & \cdots \\ 0^T & \mathrm{X}_n^T & -y_n \mathrm{X}_n^T \\ \mathrm{X}_n^T & 0^T & -x_n \mathrm{X}_n^T \end{bmatrix}$, the problem of finding the registration parameters is transformed into the problem of finding the optimal solution of $MP = 0$. Because matrix P has 11 degrees of freedom and a set of corresponding points can provide two linearly independent equations, it only needs six sets of corresponding points to calculate matrix P. According to the principle of least square method [11], the element values in matrix P are the same as those in the eigenvector corresponding to the minimum eigenvalue of matrix $M^T M$.

In conclusion, to calculate the registration matrix, homonymous feature points should be replaced into the expression of matrix M. And then record the feature vector corresponding to the minimum eigenvalue of the matrix $M^T M$ as the registration matrix P. Substitute matrix P into Eq. (5) to obtain the corresponding position of point cloud coordinates on the visible image. It should be noted that when calculating the corresponding three-dimensional point cloud position corresponding to the two-dimensional visible image coordinates, the projected Z value should be provided.

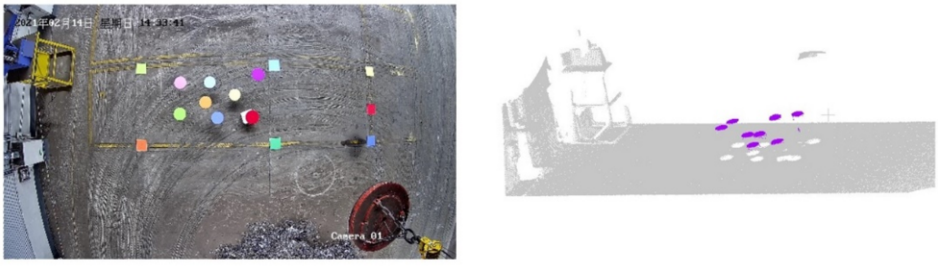
5. Experimental results and analysis

The fish-eye camera used in this paper has a resolution of 1920 * 1080, and the point cloud coverage is 24 meters long and 8 meters wide. Firstly, the camera is corrected for distortion. According to the ground straight line, the calculated distortion center of the camera is (1004.75,534.138) and the calculated distortion coefficient is -3.36243e-07. Figure 1 shows the comparison of visible image before and after distortion correction.

(a) Original image (b) Undistorted image

Figure 1. Comparison of visible image before and after distortion correction

Establish a suitable calibration field, as shown in Figure 2, where (a) is a two-dimensional visible image picture and (b) is a three-dimensional point cloud picture.



(a) Two dimensional image (b) Three dimensional image

Figure 2. Schematic diagram of calibration field

Build an image attribute table and calculate the coordinates of multiple groups of homonymous feature points in the visible image and point cloud respectively. The disc attributes table and the calculated coordinate values of disc centers are shown in Table 1. The registration matrix between the visible image and the three-dimensional point cloud is calculated according to the coordinate values of disc centers, and the calculation result is

$$P = \begin{bmatrix} -0.00303082 & -0.0337023 & -0.027803 & 0.962485 \\ -0.0328093 & -0.000811075 & -0.013961 & 0.265392 \\ -2.80975e\text{-}06 & -2.59073e\text{-}06 & -3.05421e\text{-}05 & 0.000383686 \end{bmatrix}.$$

In addition, six groups of feature points are selected for error calculation, and the registration error is defined as the Euclidean distance between the calculated value of three-dimensional coordinates and the actual value. The specific error results are shown in Table 2.

It can be seen that the registration error of the algorithm in this paper is controlled within 5 centimeters, so this method can be considered as an effective registration method.

Table 1. Disc attributes and coordinate values of disc centers

Disc number	Disc RGB value	Disc height (m)	Visible image coordinates	Point cloud coordinates (m)
1	(255,210,255)	1.16	(739.85, 324.00)	(4.292,21.036,1.165)
2	(220,255,255)	0.52	(871.704, 325.88)	(4.353,19.862,0.523)
3	(255,255,230)	0.40	(980.26, 381.72)	(3.824,18.810,0.407)
4	(210,60,255)	1.40	(1080.98, 291.55)	(4.545,17.964,1.407)
5	(200,255,150)	1.38	(1080.98, 291.55)	(3.052,21.015,1.384)
6	(255,210,130)	0.83	(854.31, 413.73)	(3.508,20.046,0.854)
7	(150,190,255)	0.65	(906.32, 480.44)	(2.857,19.563,0.652)
8	(255,0,60)	1.95	(1055.70, 478.51)	(2.889,18.272,1.949)

Table 2. Test points registration error statistics

Test point number	Real value of visible image coordinates	Calculated value of point cloud coordinates (m)	True value of point cloud coordinates (m)	Accuracy error (cm)
1	(527.72,264.72)	(4.763,22.769,1.400)	(4.780, 22.784, 1.400)	2.229
2	(772.58,323.31)	(4.273,20.713,1.362)	(4.280, 20.727, 1.362)	1.486
3	(995.89,250.42)	(5.049,18.665,0.657)	(5.016, 18.681, 0.657)	3.735
4	(930.13,370.54)	(3.875,19.323,1.126)	(3.865, 19.343, 1.126)	2.266
5	(999.37,443.98)	(3.205,18.627,0.455)	(3.214, 18.646, 0.455)	2.090
6	(1140.25,295.11)	(4.516,17.418,1.397)	(4.503, 17.449, 1.397)	3.372
Average value				2.530

6. Conclusion

In this paper, a registration algorithm of visible image and three-dimensional point cloud is proposed based on point features. The distortion correction of the camera is completed by using a single visible image, which simplified the steps in engineering practice. A brand-new method of building calibration field was proposed, which provided high quality homonymous feature points in visible image and point cloud, so as to realize the automatic calculation of registration parameters. Experiments show that this method can achieve centimeter level registration accuracy.

References

[1] Mingge X, You H, Xiaoming T, Shichang X. Overview of multisensor image fusion. Electro Optic and Control. 2002; (04): 1-7.

[2] Sriram KV, Havaldar RH. Analytical review and study on object detection techniques in the image. International Journal of Modeling, Simulation, and Scientific Computing. 2021; 12(05).

[3] Jiefu L, Mingtao L, Lijia C. Overview of automatic target recognition technology for aerial remote sensing. Scientific Journal of Intelligent Systems Research. 2021; 3(1).
[4] Yonghui C. Research on three-dimensional point cloud data processing technology based on laser scanning. University of Science and Technology of China. 2017.
[5] Emmanuel PB. A comparison between photogrammetry and laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing. 1999; 54(2).
[6] Qiu P, Liang Y. The improved algorithm of remote sensing image registration based on SIFT and CONTOURLET transform. IEEE; 2013; pp 906-909.
[7] Sheng Z, Peihua L, Yuli L, Mingsi Q, Meng Z. Image registration method based on optimized SURF algorithm. American Journal of Optics and Photonics. 2019; 7(4).
[8] Peishuang S. Research on fish eye image distortion correction algorithm. Tianjin University of Technology. 2016.
[9] Hughes C, Denny P, Jones E, et al. Accuracy of fish-eye lens models. Optical Technology and Biomedical Optics. 2010; 49(17): p 3338.
[10] Zhang L, Shang H, Wu F, et al. Robust line-based radial distortion estimation from a single image. IEEE Access. 2019; 7: pp 180373-180382.
[11] Nikolai C. Circular and linear regression: Fitting circles and lines by least squares. CRC Press. 2010.

doi:10.3233/ATDE220228

Forecast-Oriented Feature Extraction of Cutting Chatter Based on Mean Square Frequency

Liming XU [a], Chao ZHOU [a, 1], Chuhao QIU [a], Lun SHI [a]
[a] *Shanghai Jiao Tong University, Shanghai, 200240, China*

Abstract. Cutting chatter has a tremendous impact on the machining steadiness of machine tools and seriously restricts the machining efficiency. In order to realize the on-line monitoring and especially forecasting of cutting chatter, a method for chatter feature extraction based on mean square frequency (MSF) is proposed. By analyzing the cutting force signal in frequency domain, the ratio of MSF is calculated as a feature based on chatter sensitive frequency band extracted by wavelet packet transform. The results display that the presented method is able to identify the cutting chatter in advance effectively and realize chatter early warning, which provides conditions for further chatter suppression.

Keywords. Cutting chatter, mean square frequency, wavelet packet decomposition, feature extraction, forecast.

1. Introduction

In metal processing, cutting chatter is one kind of spontaneous vibration phenomenon due to the self-oscillation effect between workpiece and tool. Chatter greatly affects the processing quality of the workpiece, the service life of the tool and even the service life of the machine tool. Due to the complexity of machine tool dynamics, the dynamic characteristics of cutting system will change instantaneously with the change of cutting parameters. There is often a certain difference between the actual model and the theoretical model, which affects the extensive use of chatter models in engineering. Therefore, on-line monitoring and prediction of chatter by using sensor signals plays an important role in practical applications.

The rapid development of computer and sensor technology provides good conditions for on-line chatter monitoring. The common sensing signals for chatter monitoring include cutting force signal, vibration signal, acoustic pressure signal and so on. Different sensing signals have different sensitivity to different cutting chatter. Generally, the original signals can not effectively identify the chatter. The classification technology is applied in on-line monitoring of cutting chatter by extracting the characteristic quantities that can sensitively respond to chatter from the original signal. Liu [1] et al. proposed an on-line monitoring method for extracting chatter character based on wavelet packet transform of current and acceleration signals. The energy ratio

[1] Chao ZHOU: School of Mechanical Engineering, Shanghai Jiao Tong University, joeciao96@sjtu.edu.cn

and the standard deviation of characteristic band signals are used as chatter monitoring features. Somkia [2] et al. solved the mean variance of dynamic cutting force signal and extracted the corresponding features to realize the real-time monitoring of chatter in the machining process. Gao [3] et al. constructed the vibration datasets of different chatter states and established a machine learning model for chatter identification to achieve a higher mean identification accuracy. Xu [4] et al. proposed a short-time difference spectrum analysis for identification of early chatter, which is effective for early stage recognition of chatter frequency and amplitude track from low signal-to-noise vibration signal. Yung[5] et al. used a relative wavelet packet energy entropy in the high-frequency band as a chatter feature for early detecting and classifying chatters, respectively. Liu [6] et al. combined wavelet packets decomposition (WPD) and EMD as the chatter identification method to avoid the effect of modal aliasing, and the results verify the effectiveness for identification of the chatter features in the milling process. Furthermore, Tran M Q[7] et al. proposed a kind of multi-sensor data fusion for milling chatter detection which is low-priced and easy to use.

However, there are few studies on feature extraction for early warning of cutting chatter at present. Early warning can avoid huge economic losses and improve productivity and efficiency. Therefore, a novel method for feature extraction of cutting chatter is proposed focusing on the rapidity of early warning. The feature extraction algorithm is analyzed and experiments are conducted. The results lay a foundation for further chatter suppression.

2. Feature extraction

A feature extraction method of cutting chatter is presented based on mean square frequency (MSF) and WPD. To determine the frequency domain range of cutting chatter, firstly use fast Fourier transform to transform the cutting force signal obtained from milling process. The signal is decomposed and reconstructed by discrete wavelet packet. Afterwards, the mean square frequency of the characteristic frequency band is extracted as the characteristic quantity.

Wavelet packet decomposition makes up for the deficiency that wavelet decomposition only deals with the low-frequency signals and do not process the high-frequency signals. WPD adaptively select the frequency band and spectrum based on the signal features. The fast iterative algorithm of wavelet packet decomposition is as follows:

$$c_{2p}^{j+1}(k)=\frac{1}{2}\sum_{k\in Z}h_{l-2k}c_{p}^{j}(l) \tag{1}$$

$$c_{2p+1}^{j+1}(k)=\frac{1}{2}\sum_{k\in Z}g_{l-2k}c_{p}^{j}(l) \tag{2}$$

The reconstruction algorithm is expressed as

$$c_{p}(k)=\sum_{l\in Z}h_{k-2l}c_{2p}(l)+\sum_{l\in Z}g_{k-2l}c_{2p+1}(l) \tag{3}$$

Where, $c_p(l)$ represents the original signal, $c_{2p}^{j+1}(k)$ represents the decomposed low-frequency part, $c_{2p+1}^{j+1}(k)$ represents the decomposed high-frequency part, and

h_{l-2k}、g_{l-2k} represents the filter decomposition coefficient. $j = 1,2,3,\ldots,J$ (J is the number of decomposition layers).

The mean square frequency reflects the signal change in the frequency domain, which is the weighted average of the signal frequency square, and the weight is the amplitude of the power spectrum. Due to the chatter occurrence, the main frequency band shifts down, and the value of mean square frequency also drops sharply. The MSF is calculated as follows:

$$MSF = \sum_{i=1}^{H} f_i^{\,2} S(f_i) / \sum_{i=1}^{H} S(f_i) \tag{4}$$

Where, f represents frequency and $S(f)$ represents power spectrum amplitude.

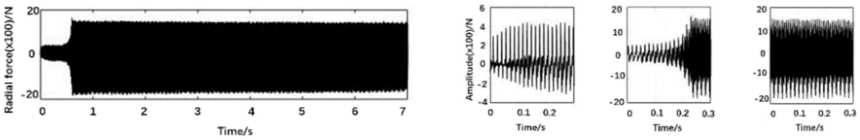
3. Experiments

The milling chatter test setup and method are as follows: the milling force acquisition system is composed of Kistler 9272 dynamometer, data acquisition card, Kistler 5070a charge amplifier. The workpiece material is 400 mm×100 mm×100 mm aluminum alloy. The tool is four tooth carbide ball end milling cutter, with 4 mm ball radius and 40 mm overhanging length. The sampling frequency of milling force is 20k Hz. By changing the cutting parameters, such as feed speed, spindle speed, axial cutting depth, five groups of cutting force signal are obtained, of which the first three groups are in stable state, and chatter occurs in the last two groups. The test conditions are listed in Table 1.

Table 1. Experimental conditions

Group number	State	n/(r·min^{-1})	f/(mm·min^{-1})	a_p/mm
1	stable	4000	200	2
2	stable	4000	200	3
3	stable	4000	200	4
4	chatter	4000	580	6
5	chatter	4000	640	6

Taking 5th group as an example, figure 1(a) displays a 7 seconds force signal from stable processing to chatter occurrence in time domain. Figure 1(b), (c) and (d) are the three stages of stability, transition and chatter.



(a) (b) (c) (d)

Figure 1. Waveform of force signal in processing: (a) entire, (b) stability, (c) transition, (d) chatter.

The signals of the three states are transformed by fast Fourier transform to obtain the corresponding spectrum. Based on the analysis of signal in both time and frequency domain, a conclusion can be drawn that when chatter occurs, the signal amplitude in time domain increases and the energy shifts with frequency.

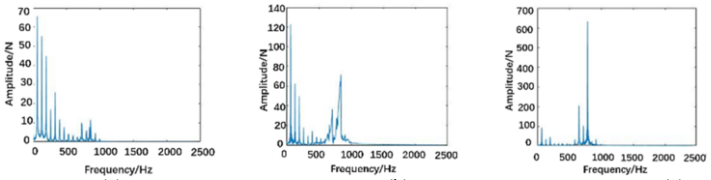


Figure 2. Signal spectrum diagram: (a) stable, (b) transition, (c) chatter.

4. Results and analysis

It is found that the generation of cutting chatter is accompanied by the energy shift with frequency, which reaches the peak at 788Hz. A signal segment containing three states of stability, transition and chatter is intercepted from the entire time series in Figure 1(a), and is displayed in Figure 3. By 4-layer wavelet packet decomposition of any time slice in the signal segment, 16 frequency bands from d1 to d16 are obtained. The frequency range of d2 is from 625Hz to1250Hz and contains the most abundant chatter information. The wavelet packet coefficients of d2 are reconstructed and the mean square frequency of the reconstructed signal is calculated.

The chatter characteristic quantity T is defined as the ratio of the mean square frequency r of the cutting force signal to the maximum mean square frequency r_{max} in the chatter state. Thus the MSF ratio can be expressed as

$$T = \frac{r}{r_{\max}} \tag{5}$$

In order to verify whether the chatter character can sensitively respond to the generation of chatter, calculate the T of the time slice signal from the stable state to the chatter state as the criterion. The variation trend of mean square frequency ratio is displayed in Figure 3.

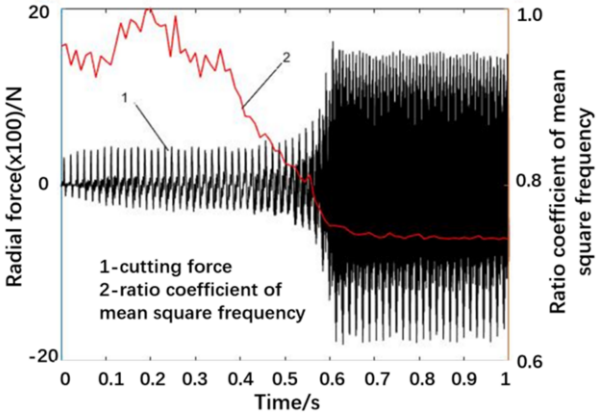


Figure 3. Variation trend of mean square frequency ratio

As displayed in Figure 3, the overall change of MSF ratio fluctuates in a narrow range from 0.9 to 1. When the time is about 0.4s, the cutting enters the chatter

transition stage. At 0.7s, the character value *T* decreases to 0.75, and the cutting enters the stable chatter stage. The chatter transition time reaches about 0.3s. Generally, the chatter transition time identified by the character value extracted in time domain or frequency domain is less than 0.1s. Thus, the feature of MSF ratio can significantly advance the time of chatter prediction, which provides a good condition for chatter suppression.

5. Conclusions

In this study, a method for early identification of cutting chatter is presented. The results show that:

- In time domain, the signal amplitude increases when chatter occurs, and the energy shifts with frequency as well.
- The frequency band sensitive to cutting chatter can be located through wavelet packet decomposition and signal spectrum analysis, and can be reconstructed into the state signal most suitable for feature recognition.
- Three states of stability, transition and chatter of the cutting process can be effectively identified and what is more, the forecast time of cutting chatter is significantly advanced by using the feature quantity of MSF ratio, which provides a foundation for chatter suppression.

Acknowledgment

National Natural Science Foundation of China (No. 52075331) provided the funding of this study.

References

[1] Liu Y, Wang X, Lin J. Early Chatter Detection in Gear Grinding Process Using Servo Feed Motor Current [J]. International Journal of Advanced Manufacturing Technology. 2016; 83(9): pp. 1801-1810.
[2] Somkiat T, Narongsak P. Development of Chatter Detection in Milling Processes [J]. International Journal of Advanced Manufacturing Technology. 2013; 65(5): pp. 919-927.
[3] Gao H N, Shen D H, Yu L, et al. Identification of cutting chatter through deep learning and classification[J]. International Journal of Simulation Modelling. 2020; 19(4): pp. 667-677.
[4] Xu X, Zhou T, Hu H, et al. Chatter Frequency Identification and Amplitude Tracking Using Short-Time Difference Spectrum Analysis. IEEE Transactions on Instrumentation and Measurement. 2020; 69(12): pp. 9844-9852.
[5] Yungchen Y, Chen YH, Chienhao L, et al. Real-time chatter detection and automatic suppression for intelligent spindles based on wavelet packet energy entropy and local outlier factor algorithm. The International Journal of Advanced Manufacturing Technology. 2019; 103(1-4): pp. 297-309.
[6] Liu C, Zhu L, Ni C. The chatter identification in end milling based on combining EMD and WPD. The International Journal of Advanced Manufacturing Technology, 2017, 91(9): pp. 3339-3348.
[7] Tran M Q, Liu M K, Elsisi M. Effective multi-sensor data fusion for chatter detection in milling process. ISA transactions; 2021.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 199
J. Xu et al. (Eds.)

doi:10.3233/ATDE220229

Research Progress of Lubricant Oxide Films in a Wide Temperature Range

Bingsen JIA[a] ,Yixian WANG[a], Tingting GUO [a], Xuanyu LI [a] and Chufeng SUN [a,b,1]
[a] *Key Laboratory of Environment-Friendly Composite Materials of the State Ethnic Affairs Commission, School of Chemical Engineering, Northwest Minzu University, Lanzhou, 730030, China*
[b] *Gansu Provincial Biomass Function Composites Engineering Research Center, Lanzhou, 730030, China*

Abstract. High-temperature lubrication is widely used in modern industrial fields such as aviation, aerospace, nuclear, metallurgy and metal processing. The service temperature of corresponding mechanical equipment is getting higher and higher, and the problems of high-temperature lubrication and wear failure are becoming more and more prominent.This article reviews the latest research progress of binary and ternary oxide coatings with high- temperature solid lubricant application potential, discusses the structure, chemical and electrical properties of these oxides and their tribology, and focuses on the main types of solid lubricating materials. The correlation of performance and its lubrication mechanism are analyzed.This review also discusses the industrial applications of these coatings and predicts that lubricating coatings might develop in the direction of self-lubricating and self-healing smart lubricating materials.

Keywords. Solid lubrication; Binary oxide; Ternary oxide; Self-lubricating; Self-repairing

1. Introduction

Extremely harsh working conditions such as high/low temperature (wide temperature range), special media and high speed are common in modern aviation, aerospace, nuclear technology and electric power and other high-tech industries. The lubrication and wear resistance of related sliding parts have an impact on the reliability of the entire system, key technology of longevity and life.And therefore, there is an increasingly urgent need for new lubricating materials with continuous lubrication function in a wide temperature range[1,3].

However, conventional fluid lubricating materials are very "sensitive" to temperature. For example, the maximum use temperature of grease-based lubricating materials usually does not exceed 200 ℃, and the limit use temperature of polymer-based lubricating materials is 400 ℃. Too high temperature would cause the lubricant to fail due to oxidative deterioration, causing direct contact with the surface of the moving pair, resulting in serious scratches and even seizures on the moving parts, and thus,the failure of the entire mechanical moving parts[4].The traditional solid lubricants

[1] Corresponding Author, Chufeng Sun, School of Chemical Engineering, Northwest Minzu University, Lanzhou, 730030, China; E-mail: cfsun@licpcas.cn.

(such as graphite and molybdenum disulfide, etc.) would be oxidized and become invalid at around 350°C. Therefore, it is of great significance to develop a wide temperature range solid lubricant that adapts to room temperature to high temperature. Continuous lubrication from low temperature to high temperature, that is, wide temperature range lubrication.In the friction process, the solid lubricant, solid lubricating material and surrounding media react physically and chemically with the friction surface to form a solid lubricating film, which can reduce friction and wear. But to achieve this goal, the solid film must meet the following requirements [5]: good adhesion to the metal substrate; low shear strength; good stability; strong load-bearing capacity.

Therefore, scholars at home and abroad have conducted a lot of research on solid lubrication technology and tried different types of materials. It has been proven that they still have good friction properties at high temperatures. The solid lubricants that can be used normally can be divided into three categories [6-7]:(1) Soft metals (such as Ag, Cu, Au, Pb and In); (2) Fluorides (such as CaF2, BaF2 and CeF3); (3) Metal oxides (such as V_2O_5, Cr_2O_3). For solid lubricating materials with continuous lubrication in a wide temperature range, the most leading and representative work in solving the friction, wear and lubrication problems of related moving parts under harsh working conditions is NASA Green The PS series thermal spraying wide- temperature self-lubricating coating developed by the research center, and the temperature-adaptive lubricating coating ("chameleon" coating) developed by the U.S. Air Force Material Research Laboratory.

This article reviews the solid lubrication. Through a lot of research, we found that under the harsh conditions of room temperature-high temperature (1000°C)-room temperature alternating temperature, oxides are usually stable in structure and chemical/thermodynamics, such as $(Al ,Cr)_2O_3$ and other solid solution oxides form corundum phase in the relatively low temperature PVD process. This is a promising low friction and wear coating. The metastable structure at medium temperature is conducive to reducing friction and wear, so oxidation Material may be the best choice for solid lubrication. This article discusses the application of high-temperature solid lubricants, introduces the latest experimental studies of oxide coatings and applications under wide temperature range conditions, and emphasizes future development trends.

1.1 Solid lubrication mechanism

1.1.1 Soft metal lubrication mechanism

The lattice structure of soft metals is anisotropic, easy to slip between the crystals, has low shear strength and good plastic deformation ability, and it is easy to produce shear slip during the sliding process and has a lubricating effect.Common soft metals include Ag, Au, Pb, Cu, etc. Among them, the metal Ag has a high melting point (961.93°C) and stable physical and chemical properties, and can maintain good self-lubricating properties in the middle and low temperature range. Wang et al. [8] designed a Pb-Sn-Ag-RE composite solid lubricant that can be used for lubrication in a wide temperature range of 300°C to 800°C. H. Torres et al. [9] found that MoS2 coated Ag and uniformly distributed in the self-lubricating composite coating, which effectively prevented Ag from floating during the laser cladding process. The increase in Ag content makes the coating have better friction at room temperature. Learn performance.

Table 1. Phsical properties and tribological properties of common soft metal materials

metal	Crystal structure	Density ($g \cdot cm^{-3}$)	Melting point (℃)	Moh's hardness	Coefficient friction
					Start-End
Au	FCC	19.32	1063	2.5-3.0	0.40-0.57（788℃）
Ag	FCC	10.53	961	2.5-4.0	0.70-0.40（788℃）
Pb	FCC	11.34	327	1.5	0.33-0.17（788℃）

1.1.2 Metal oxide lubrication mechanism

Because of its good thermal and chemical stability, metal oxides are not prone to oxidation and decomposition, and are often used as an effective method to solve the problem of lubrication in high-temperature oxidizing atmospheres. In a high-temperature environment, oxides are softened. Some metal oxides have lubricating properties due to plastic deformation or shearing during friction to form an oxide lubricating layer, and the transition from brittleness to plasticity has greatly improved their friction and wear. performance.Peterson et al. [11] systematically investigated the high-temperature tribological properties of a large number of metal oxides (PbO, B_2O_3, Cu_2O, CuO, CoO, MoO_3, V_2O_5), and their results showed that the lubricity of the oxide is not directly related to the hardness, shear strength and crystal structure of the oxide. The high temperature lubrication mechanism of most oxides is related to their thermal softening at high temperatures.Erdemir [12] introduced the ion potential to explain the lubrication characteristics of different oxides, where the ion potential refers to the ratio of the number of ion charges (Z) to the ion radius (r, pm) (represented by Φ), that is, $\Phi = Z/r$, The size of this parameter affects the thermal stability of the compound. Research in recent years has further shown that binary metal oxides such as molybdate and vanadate (MexTyOz) have a relatively low friction coefficient between 350 and 700℃[13], which has further broadened the research scope of oxide lubricating coatings.

Table 2. Melting point, friction test conditions and friction coefficient of some oxide

Oxide	Melting point（K）	Coefficient friction (temperature range, K)
ReO_3	433	—
B_2O_3	723	0.3-0.15（823-1000）
V_2O_3	945	0.32-0.3（873-1273）
MoO_3	1068	0.27-0.2（870-1073）
WO_3	1743	0.3-0.25（873-1073）
TiO_2	2123	0.55-0.35（1073-1273）
Al_2O_3	2313	0.5-0.3（1073-1273）
ZrO_2	3073	0.5（1073）
MgO	3173	0.5-0.35（773-793）
FeO	—	0.6（573）

1.1.3 Metal fluoride lubrication mechanism

Fluoride lubricants are mainly fluorides of alkaline earth metals such as BaF_2, CaF_2, LiF, NaF, CeF_3, LaF_3 and rare earth metals, all of which have high melting points and

good chemical stability.At the same time, they show good lubricating properties by softening at high temperatures [14], among which BaF_2 and CaF_2 with cubic crystal structure have been well used in high-temperature moving parts in aviation and aerospace.Such as alkaline earth metal fluoride BaF_2 and CaF_2, when the temperature exceeds at 500℃, from brittle material to plastic material transformation and lubricity, can be used to 900℃ without oxidation failure.Ma et al [15] found that nanoscale LaFR has better self-lubricating wear resistance properties than LaF. Besides, rare earth fluorides such as CeF also have certain lubricating properties. The Ni-based MoS2 /LaF3-CeF composite coating prepared by Zhang[16] also has better wear and friction reduction properties.

Table 3. Physical properties and friction coefficient of some fluoride solid lubricats

Oxide	density ($g \cdot cm^{-1}$)	Melting point (℃)	hardness (GPa)	Coefficient friction
CaF_2	3.18	1418	4.0	0.2-0.4（25℃-900℃）
BaF_2	4.78	1353	3.0	0.2-0.4（25℃-900℃）
62wt.%BaF_2 38wt.%CaF_2	4.01	1022	1.5	0.15-0.2（25℃-900℃）
CeF_3	4.50	1437	4.5	0.2-0.5（25℃-1000℃）
LaF_3	4.50	1490	4.5	0.2-0.4（25℃-1000℃）

1.2 Preparation technology of solid lubricating coating materials in a wide temperature range

There are various techniques for the preparation of solid self-lubricating coating/ cladding in wide temperature range, such as electroplating, thermal spraying, wet spraying, laser cladding, powder metallurgy, vacuum vapor deposition, etc. Currently, the main applications are thermal spraying, laser cladding, physical vapor deposition, etc.

1.2.1 Thermal spray technology

Thermal spraying technology is a new technology for material surface strengthening and protection, and occupies an important position in surface modification technology. It is the use of some kind of heat source, such as electric arc, plasma arc, combustion flame, etc. to heat powdered or filamentary metallic or non-metallic coating materials to the molten or semi-molten state, and then atomized to form molten droplets with the power of the flame flow itself or the external high-speed airflow, and sprayed to the surface of the pretreated substrate material at a certain speed, the molten droplets are flattened, rapidly cooled and solidified, and formed on the surface of the substrate with a certain thickness and function of the coating on the substrate surface [17,18]. Hou et al [19] prepared WC-based coatings on stainless steel surfaces using a spraying method, and as the friction temperature increased, WO3 and CoWO4 phases were generated in the coatings, which exhibited excellent tribological properties in the

600°C range. Yuan et al [20] prepared a WC-Co-Cu-BaF2/CaF2 self-lubricating wear-resistant coating by plasma spraying technology, and found that although the friction coefficient and wear rate were slightly higher at 200°C due to the presence of WC hard particles, the coating had a lower friction coefficient and wear rate at 400°C-600°C. The reason for this is that At 400°C, Cu and BaF2/CaF2 form a smooth and dense friction product layer on the surface of the coating, while at 600°C, the decomposition of the material in the coating and the generation of oxides form a dynamic equilibrium, and the wear of the coating at this time is mainly manifested as oxidation wear. O.J. Gerald et al [21] deposited Cr2O3 coating on Q235 steel substrate using plasma spraying technique, which resulted in better denseness and lower porosity with increasing spraying current, producing the least coating wear and the best frictional resistance performance at higher spraying currents.

1.2.2 Laser cladding technology

Laser cladding technology refers to irradiating the selected coating material through different filling methods and using high-energy density laser light to melt and rapidly solidify the coating material and the thin layer on the surface of the substrate at the same time, forming a good metallurgical bond with the substrate[22]. This technology significantly improves the wear resistance, corrosion resistance, high temperature oxidation resistance and other properties of the base surface, so as to meet the specific performance requirements of the material surface. The NiCrAlY-Ag-MoO composite coating prepared by Wang et al[23] uses Ag as an effective lubricant at 200°C. As the temperature increases, Ag loses its lubricating effect. When the temperature continues to rise to 600°C, the friction In the process, the AgzMoO4 lubricating phase is formed by reaction, which can achieve better lubrication at 800°C. Ke et al[24] used the laser cladding technology to prepare the composite powder on the surface of Ti6Al4V alloy with Ni-48%Mo-32%Si (wt.%, in which the mass of Mo and Si is 1:2 by atomic ratio) mixed powder. Experiments show that the composite coating has no cracks and only a few pores, and achieves a good metallurgical bond with the substrate. The hard phases Ti_5Si_3, $MoSi_2$ and Mo_5Si_3 are uniformly distributed in the matrix α-Ti and NiTi and after being oxidized at a constant temperature of 800 ℃ for 100 h, the oxide film of the composite coating is mainly composed of TiO_2, SiO_2 and NiO, and the structure is continuous and dense, showing good performance. The high temperature oxidation resistance of Ti_6Al_4V alloy is mainly loose TiO_2, and the surface oxidation is serious, showing poor oxidation resistance. Laser cladding NiMoSi composite coating is one of the effective ways to improve the high temperature oxidation resistance of Ti_6Al_4V alloy.

1.2.3 Physical Vapor Deposition Technology

Vapor-phase deposition is one of the common techniques for preparing thin films, where film-forming substances are vaporized and transported onto a substrate to be deposited as a solid-phase film.In recent years, the technology of preparing wide temperature domain solid self-lubricating coatings by physical vapor deposition has been rapidly developed, which refers to the reaction of the evaporated material with gas ionization to form compounds by evaporation, ionization and sputtering under vacuum, and then deposited on the workpiece surface under the acceleration of electric

field. Physical vapor phase technology is mainly divided into three kinds of vacuum evaporation coating, sputtering coating, ion coating, etc[25]. The prepared coatings mainly include multi-metal coating, bimetallic oxide coating, nitride based temperature adaptive coating, etc.

In the face of extremely harsh use environments, multi-metal composite coatings show greater advantages. Du et al [26] prepared NiCrN hydrophobic coatings using multi-arc ion plating technique at different deposition temperatures. Nitride was chosen as the substrate based on its high strength, hardness and wear resistance. As the deposition temperature was increased from 150 to 450 °C, the surface morphology of the coatings changed from simple microstructures to micro-nano structures. During storage in air, the wettability changed from hydrophilic to hydrophobic, and the contact angle reached 143.7°. The samples at 450°C were so strong that the contact angle remained at 134.4° even after 100 sandpaper wear tests at 10 kpa pressure. And the wettability of the coating after abrasion was exposed to air showed self-healing ability. The experiments showed that the NiCrN coating prepared at 450°C has good corrosion resistance compared with stainless steel.

1.2.4 Others

Rajnesh Tyagi et al [27] prepared Ni-Ag-hBN coatings using powder metallurgy and found that the coatings have certain friction and wear reduction properties from room temperature to 600°C. It was also found that the hardness of the coatings decreased with the increase of Ag and BN content, which was mainly due to the increase of soft metal silver content and the poor sintering properties of BN.

In summary, although laser cladding technology has gained some application in the preparation of wear-resistant, corrosion-resistant, anti-oxidation, thermal barrier, and biological coatings, the cladding material melts under the action of laser to form a liquid melt pool and then solidifies, forming a micro-metallurgical casting process. The chemical compatibility and physical matching between the components of the cladding layer will affect the crystallization, growth, and solidification of the components in the laser cladding process, thus affecting the organization, mechanical, and physicochemical properties of the cladding layer. Thermal spraying technology is widely used to prepare various coating layer materials because of its advantages such as stable gauge, simple operation, high spraying efficiency, low thermal influence on the substrate during spraying, and applicability to almost all refractory materials. However, the disadvantages of this technology are obvious, namely, the prepared coating layer is a layered isometric crystal, the coating layer contains a large number of pores and porosity, the bond with the substrate material is mainly physical bonding, micro-metallurgical bonding as a supplement to the weak strength, under the action of hot and cold cycles are easy to flake, and it is not suitable for the spraying of complex structural shapes and precision workpieces. Compared with thermal spray technology, physical vapor deposition (PVD) technology has the following obvious advantages: denser coating, improved thermal cycle life, significantly increased bonding force, better resistance to oxidation and thermal corrosion, smooth surface without affecting the accuracy of the substrate, especially suitable for the preparation of coating/ cladding materials on precision workpieces, this series of advantages also makes PVD technology widely used in the surface modification of wear-resistant materials, lubricating materials, biomedical materials, thermoelectric materials and some mold materials, etc. Its disadvantages are mainly slow deposition speed, high equipment

costs, more complex process, etc.

2. Current status of research on lubricious oxide coatings

2.1 Binary oxide coating

Metal oxides have good thermal and chemical stability, are not easy to oxidize and decompose, and the oxides soften at medium and high temperatures, and the transition from brittleness to plasticity also greatly improves their friction and wear properties[28]. The lubricity of oxides is not directly related to the hardness, shear strength and crystal structure of oxides. The high-temperature lubrication mechanism of most oxides is related to their thermal softening at high temperatures.Erdemir et al [12,28] developed a crystal chemistry model to obtain the relationship between the high temperature lubrication properties of metal oxides and the ionic potential of metal oxides. Figure 1 shows the curves of ionic potential versus average friction factor for different oxides, reflecting the close correlation between ionic potential and their tribological properties. Swedish scientist Magnéli discovered for the first time substoichiometric compounds similar to Me_nO_{2n-1}, Me_nO_{3n-1}, and Me_nO_{3n-2} in oxides of molybdenum and tungsten, which are prone to shear planes in the structure due to the presence of oxygen vacancies in the crystal. As a result, the bond strength is reduced, and the resulting structural defects are called Magnéli phases, and vanadium oxides have a similar structure.

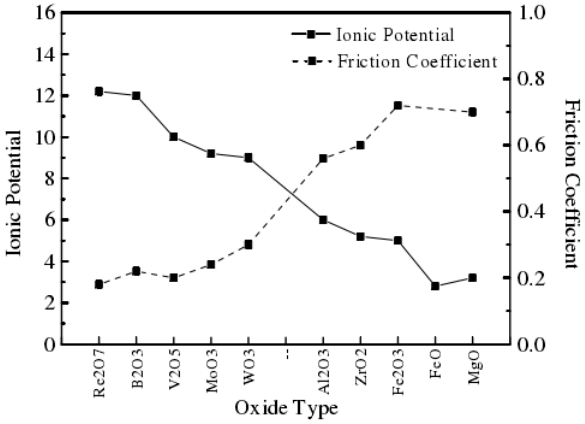


Figure 1. Relationship between ionic potentials and averaged friction coefficients of various oxides[26]

Gassner et al[30] investigated the tribological properties of MoN and Mo_2N hard coatings at high temperatures using non-equilibrium reactive magnetron sputtering deposition and showed that oxidation occurs at high temperatures to generate the Magneli phase $MonO_{3n-1}$, but at temperatures above 500°C, volatile MoO_3 is formed, limiting its use at higher temperatures.

In order to further study the role played by the Magnéli phase, Gardos [31] studied the relationship between the Magnéli phase and the redstone polymorphic TiO2 as a high-temperature lubricant. The experiment showed that the existence of oxygen vacancies is the significant shear strength of the rutile crystalline TiO2 surface area.

main reason for the decrease. When the oxygen content of stoichiometric TiO2 is slightly reduced to sub-stoichiometric TiOx (x is in the range of 1.93 to 1.98), a (132) crystal easy shear plane with an ordered structure is formed, and the shear strength decreases significantly from . And when x=1.98, the friction factor value is reduced to 0.08 due to the large reduction of shear strength, showing good lubricity.

In addition to the Magnéli phase, studies have found that binary oxides with specific growth structures can also reduce the friction and wear of materials at high temperatures. Such as ZnO with subcrystalline structure, this type of lubricating oxide has a highly defective nano columnar grain structure and contains a high density of (0002) stacking fault planes. The ZnO coating with this structure exhibits excellent tribological properties at room temperature, with a friction coefficient of 0.1～0.2 and a wear rate as low as 10-7mm3/(N·m).Zabinski [28] compared the wear life of nano-sized ZnO coatings and hot-pressed ZnO coatings by preparing them, and in fact the friction factor of nano-grain coatings was about 0.18, compared to hot-pressed coatings that wore rapidly during the test with a friction factor as high as 0.6. Therefore, controlling the defective form of the material has also become a way to achieve material lubrication.

2.2 Ternary oxide coating

Ternary oxides were found to be a promising class of new oxide lubricant materials for wide temperature range environments and to maintain good performance of substrates, the concept of which was first proposed by Zabinski et al. It was found that the ternary oxide MexTMyOz. has a rather low friction factor value of 0.1-0.3 at 200°C-800°C, where Me is a noble metal (Ag) TM is a transition metal [32]. Ternary oxides are classified as silver-based and non-silver-based ternary oxide lubricating materials.

2.2.1 Silver-based ternary oxide lubricating materials

Ag is a soft metal, very friendly to the environment, has lubricity at higher temperatures (T > 300°C), and Ag is not easy to oxidize, the addition of Ag elements can reduce the melting point of the original binary metal oxide in the ternary oxide structure, can enhance the film toughness to improve the performance of the matrix, so in the past decade or so is a hot research in high temperature tribology. Research on silver-based ternary oxides has mostly focused on silver molybdate.

Aouadi et al[33] prepared a MO_2N / MOS_2 / Ag high-temperature self-lubricating coating by a plasma spray process, which achieves lubrication effect from room temperature to 600 ° C to maintain a relatively low friction coefficient and wear rate, mainly because of friction Silme-bor bone is formed in the wear test, and the coating coefficients can be reduced to 0.1 when testing at 600 ° C. During the wear, when the temperature is higher than 500 ℃, the oxygen atom is easily interchanged, and the chemical reaction of the Ag and Mo oxide forms a molybdate compound. Li et al[34] The nickel-based self-lubricating composite material prepared by powder metallurgy technology. During the sintering process, V_2O_5 has a reduction reaction, transition to VO_2. As the Ag content (mass fraction) increases, the porosity of the composite is lowered, the hardness is lowered; with the increase in the V_2O_5 content (mass fraction), the porosity of the composite increases, the hardness increases. And in the temperature range of 25 to 600 ℃, the friction coefficient of the composite is generally increased

first. At room temperature, with the increase of the Ag content (mass fraction), the coefficient of friction decreases, the coefficient of friction is 0.32; at 600 ℃, with the addition of V_2O_5, the composite material exhibits better tribological properties, friction The coefficient is at least 0.22. Under medium and low temperature conditions, MOS_2 and AG have mainly lubricated; under high temperature conditions, the composite materials have a friction chemical reaction, and the new molybdenum silver-lubrication phase is formed, and the lubrication mechanism is molybdiblybdate and V_2O_5 lubrication of friction chemical reaction. The same synergistic effect. To further explore, Xue[35] uses high energy ball mills to prepare NiAl matrix composite materials with vacuum hot press sintering techniques, and the addition of silver-borned silver lubrication phase makes NiAl-based composite density and hardness increase. The addition of molybdate ($Ag_2Mo_2O_7$) lubrication phase of different nanostructures significantly improved high temperature tribological properties of NiAl compound composite. Further, the prepared particulate molybdate can reduce the width temperature domain friction coefficient and the wear rate of the NiAl group composite. Compared to linear and rod-shaped borid silver, the particulate molybdate nickel aluminum matrix composite has more excellent tribological properties at 800 ℃, the friction coefficient is 0.27, and the wear rate is only 2.90×10^{-5} mm^3 / nm. Improvement of friction and wear performance at high temperatures in composite is due to the main high temperature lubrication of the main high temperature lubrication in the surface of the wear surface, which is mainly reduced, lubricating.

Stone et al [36] reduced friction by preparing nanocomposite films of niobium nitride intercalated with nanosilver, where silver migrated to the surface to form Nb_2O_5 in the medium and low temperature range. At higher temperatures, oxygen, silver and transition metals reacted to form $AgNbO_6$, a metal oxide phase with lubricity, which kept the friction factor between 0.15 and 0.3 above 700°C. The researchers then added MoS_2 to the composite film. The addition of MoS_2 did not improve the tribological behavior of the original film at medium temperatures, but the reaction at higher temperatures produced silver niobate and silver molybdate phases to further improve its tribological properties.

2.2.2 Non-silver based ternary oxides

The study of non-silver based ternary oxides is gradually increasing due to the relatively high migration rate of silver and the possibility of condensation on the surface to destroy the surrounding components.Zabinski [37] prepared PbO-MoS_2 composites to form a $PbMoO_4$ lubricant layer at high temperatures, and although this material has a small friction factor at high temperatures (about 0.3 at 700°C), at low temperatures aggravates the wear. In order to achieve wide temperature domain lubrication, the material can maintain low frictional wear at low, medium and high temperatures.

He et al [38,39] deposited Cr_2O_3 films on the surface of Incone1718 high-temperature alloy by multi-arc ion plating technique, which has poor hardness, film-base bonding, and tribological properties, and after high-temperature annealing treatment of the original films, due to the diffusion of Cr and Ti elements in the matrix along the grain boundaries, the metal atoms will react with oxygen after diffusion to the surface, which greatly improves the The hardness and toughness of the film are greatly improved, and a composite phase of $Cr_2Ti_{17}O_{17}$ and Cr_2O_3 with a reticulation-like

raised structure is generated. This special morphology is very similar to the surface of the woven sample, and secondly, it has been shown that for $Ti_{n-2}Cr_2O_{2n-1}$ phase, when $6 \leqq n \leqq 9$, its structure is similar to that of Magneli phase, and the presence of shear-prone surface can reduce the friction factor during the friction process, so there is also a shear-prone surface for the $Cr_2Ti_{17}O_{17}$ phase [40].That is, Cr_2O_3 in the film is the main lubricating phase in the low and medium temperature environment, and when the parts experience high temperature a composite phase with excellent lubricating properties is generated, and this kind of mesh-like raised structure is also found to have a self-healing function after multiple thermal cycle tests. Therefore, the material has good performance under multiple thermal cycles from room temperature to 1000°C.

3. Application of oxide lubricating materials

3.1 Air foil bearing

Gas foil bearings are an innovative technology in the rotating machinery industry and have broad application prospects at high speeds. No need to use other lubricating materials in the process of use, only the solid lubricating coating is used as the friction lubricating medium of the contact surface during the start-stop phase, which simplifies the structure of the bearing, makes the whole system compact and reliable, and avoids the lubricating oil under high speed. It can also solve the potential pollution of lubricating oil to the environment[41]. It has great advantages for mechanical structures that are not easy to replace and add lubricants and require long-term stable operation, and have the advantages of low maintenance or maintenance-free. Some researchers used arc ion plating method to deposit chromium oxide coating on nickel-based superalloy, and conducted five continuous thermal cycle friction experiments in the temperature range of 25 to 1000 ℃, and found that the friction coefficient was less than 0.3, The wear rate is maintained at $10^{-7}mm^3/Nm$, indicating that the coating has reusable lubricating properties over a wide temperature range, making it ideal for high temperature lubrication of sealing foils.

3.2 High-speed machine tools

High-speed cutting technology is an efficient new technology for cutting with a cutting speed much higher than conventional cutting. High-speed cutting can be used for processing non-ferrous metals, cast iron, steel, fiber reinforced composite materials, etc., and can also be used for cutting processing Various difficult-to-process materials. Now, high-speed cutting technology has gradually matured and has begun to show its talents in the manufacturing field. The unit technology and overall level of high-speed machine tools are gradually improving. Controlling the interface temperature between the tool and the workpiece so that the coating material has stable mechanical properties at high temperatures is an important way to successfully increase productivity. The self-hardening mechanism of TiAlN coatings may be further prolonged at high temperatures due to the decomposition of spinodal lines, thus further improving wear resistance [42].

4. Summary and Outlook

Although great progress has been made in the research of lubrication mechanism and the development of high-temperature solid lubricating materials, with the development of modern industry, high-temperature lubrication is facing many practical challenges. Further explore the structure of self-lubricating materials and their mechanical and tribological properties It is of great significance to study the materials used in the ultra-high temperature range and develop a single lubricating material to achieve continuous lubrication in a wide temperature range from room temperature to high temperature.

References

[1] Xue Q J， Lv J J. Research status and developing trend ofsolid lubrication at high temperatures. Tribology. 1999; 19(1): pp 91-96 (in Chinese).
[2] Sliney H E.Solid lubricant materials for high temperatures a review. Tribology International. 1982; 15(5): pp 303-315.
[3] Xue Q , Dang H . Developing situation and tendency ofthe tribological research. Tribology. 1993; 13(1): pp 73-81 (in Chinese).
[4] Eryong L, Yaping B, Yimin G, Gewen Y, Junhong J,Tribological properties of NiAl-based composites containing Ag3VO4 nanoparticles at elevated temperatures, Tribology International. 2014; 80: pp 25-33
[5] Linqing Z . Principles of Tribology. Beijing: Higher Education Press. 1994 (in Chinese).
[6] K. Miyoshi, Solid Lubricants and coatings for extreme environments. NASA Technical Memorandum. 214668; 2007.
[7] J.S. Zabinski, J.H. Sanders, J. Nainaparampil, S.V. Prasad, Lubrication using a microstructurally engineered oxide: performance and mechanisms. Tribol. Lett. 8 (2000) p 103.
[8] Jingjing W, Yanjun W, Zhen W, etc. Research on the design and tribological properties of infiltrating wide temperature range composite solid lubricant . Lubrication and Sealing. 2018; 43(1): pp 50-54 (in Chinese)
[9] H. Torres, S. Slawik, C. Gachot, et al.Microstructural design of self-lubricating laser claddings for use in high temperature sliding applications. Surface and Coatings Technology. 2018(337) pp 24-34
[10] Peterson M B, Murray S F, Florek J J. Consideration of lubricants for temperatures above 1000 F. ASLE Transactions. 1959; 2(2): pp 225-234.
[11] Peterson, Murray, Florek. Consideration of lubricants for temperatures above 1000°C. ASLE Trans, 1959. 2: pp 225-234.
[12] Erdemir A. A crystal-chemical approach to lubrication solid oxides. Tribology Letters, 2000, 8(2): pp 97-102.
[13] Rui W, Dongqiang G, Nairu H , etc. Research progress of oxide lubricating materials. Surface Technology. 2017; 46(9): pp 127-133 (in Chinese).
[14] Deadmore D L, Sliney H E.Hardness of CaF2 and BaF2 solid lubricants at 25 to 670°C [R].Cleveland:NationalAeronautics and Space Administration. Glenn Research Center;1987.
[15] Yanjun M, Yinping Y, et al. Tribological behaviors of the UV curing polyurethane acrylate resin-polytetrafluoroethylene bonded solid lubricating coat-ings filled with LaF3. Progress in Organic Coatings. 2018(121): pp 218-225.
[16] Jie Z . Laser cladding molybdenum disulfide/fluoride composite self- lubricating coating and its friction and wear characteristics. Shanghai: Shanghai University of Engineering Science. 2015(in Chinese).
[17] Jianhui Y,Yingchun Z, et al.Microstructures and tribological properties of plasma sprayed wC-Co-Cu-BaF2/CaF2 self-lubricating wear resistant coatings .Applied Surface Science. 2010.256(16); 4938-4944.
[18] Ouyang J H, Sasaki S. Unlubricated friction and wear behavior of low-pressure plasma-sprayed ZrO coating atelevated temperatures . Ceramics International. 2001; 27(3): pp 251-260.
[19] Guoliang H, Yulong A, Jie C, et al. Tribological properties of two WC- based coatings at 600°C. China Surface Engineering. 2012; 25(6): pp 61-67(in Chinese)
[20] Jianhui Y, Yingchun Z, Qiang L. WC-CO-Cu-BaF2/CaF2 self-lubricating and wear-resistant coating prepared by plasma spraying and its high-temperature friction properties. China Surface Engineering. 2012; 25 (2): pp 31-36(in Chinese).

[21] Gerald O J, Wenge L, Tao Z Y, et al. Influence of plasma spraying current on the microstructural characteristics and tribological behaviour of plasma sprayed Cr2O3 coating. Bol. Soc. Esp Cerám. Vidr. (2020)
[22] Zhenzhong G. Manual of Laser Processing Technology. Beijing: China Metrology Press, 2005: p 265 (in Chinese).
[23] Lingqian W, Jiansong Z, et al. Phase transformation and tribological properties of Ag-MoO3 contained NiCrAlY based composite coatings fabricated by laser cladding. Optics&Laser Technology, 2017, 93: pp 79-86.
[24] Jin K, Xiubo L, Jiangning X, Jue L, Yingshe L. Research on high temperature oxidation resistance and mechanical properties of TC4 alloy laser cladding NiMoSi composite coating. Chinese Chemical Society. 2018; p 86.
[25] Zhaohe G, Zhenbo Z, et al. A conformable high temperature nitride coating for Ti alloys, Acta Materialia. 2020; 189: pp 274-283.
[26] Xiaoye D, Bo G, Yanhuai L, Zhongxiao S.Super-robust and anticorrosive NiCrN hydrophobic coating fabricated by multi-arc ion plating, Applied Surface Science. 2020; 511: p 145653
[27] Tyagi, Xiong, L. Effect of load and sliding speed on friction and wear behavior of silver/h-BN containing Ni-base P/M composites. Wear, 2011; 270(7–8): pp 423-430.
[28] Erdemir A. A crystal chemical approach to the formulation of self-lubricating nanocomposite coatings. Surface & Coatings Technology. 2005; 200(5-6): pp 1792-1796.
[29] Magnéli A.Structures of the ReO3-type with recurrent dislocations of atoms: homologous series of molybdenum and tungsten oxides, Acta Crystallogr. (1953) p 495.
[30] Gassner G, Mayrhofer P H, Kutschej K, et al. Magnéli phase formation of PVD Mo–N and W–N coatings. Surface&Coatings Technology. 2006; 201(6): pp 3335-3341.
[31] Gardos M N, Magnéli phases of anion-deficient rutile as lubricious oxides. Part II. Tribological behavior of Cu-doped polycrystalline rutile (TinO2n-1). Tribology Letters. 2000; 8(2-3): pp 79-96.
[32] Aouadi S M, Kohli P, et al. Progress in coatings for high temperature tribological the development of adaptive applications . Surface&Coatings Technology. 2009; 204(6-7): pp 962-968.
[33] Aouadi S M, Paudel Y, Simonson W J, et al. Tribological investigation of adaptive Mo2N/MoS2/Ag coatings with high sulfur content. Surface&Coatings Technology. 2009; 203(10-11): pp 1304-1309.
[34] Jian L, Changsheng L, Zhaoyu D. Effect of MoS2-Ag-V2O5 on friction and wear properties of nickel-based materials . Powder Metallurgy Technology. 2021; 39(02): pp 141-146.
[35] Jiali X. Design preparation and friction mechanism of NiAl-Ag2Mo2O7 composites. Shaanxi University of Science and Technology. 2019.
[36] Stone D S, Migas J, Martini A, et al. Adaptive NbN/Ag coatings for high temperature tribological applications . Surface&Coatings Technology. 2012; 206(19-20): pp 4316-4321.
[37] Zabinski J S, Donley M S, Dyhouse V J, et al. Chemical and tribological characterization of Pb0-MoS2 films grown by pulsed laser deposition. Thin Solid Films. 1992; 214(2): pp 56-163.
[38] Nairu H, Feifei W, Li J, et al. Diffusion of Ti and Cr elements in Cr2O3 film and their influence on tribological properties. China Surface Engineering. 2016; 29 (6): 67-74(in Chinese).
[39] Feifei W, Nairu H, Li J, et al. Study on the preparation, mechanical and tribological properties of Cr/Cr2O3 multilayer films. Chinese Journal of Tribology. 2016; 36 (2): pp 226-232(in Chinese).
[40] Zabinski J S, Day A E, Donley M S, et al. Synthesis and characterization of a high-temperature oxide lubricant. Journal of Materials Science. 1994; 29(22): pp 5875-5879.
[41] Dellacorte C, Fellenstein J A, Benoy P A. Evaluation of Advanced Solid Lubricant Coatings for Foil Air Bearings Operating at 25° and 500°C. A S L E Transactions. 1999; 42(2): pp 338-342.
[42] Mayrhofer P H, Hörling A, Karlsson L, et al. Self-organized nanostructures in the Ti–Al–N system. Applied Physics Letters. 2003; 83(10): pp 2049-2051.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 211
J. Xu et al. (Eds.)

doi:10.3233/ATDE220230

Volute Optimization Based on NSGA-Ⅱ Algorithm

Fannian MENG[a,1] , Ziqi ZHANG[a], Liangwen WANG[a] and Yiyang LIU[b]
[a]*School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China*
[b]*Zhengzhou College of Finance and Economics, Zhengzhou 450002, China*

Abstract. Optimizing the volute performance can effetely improve the efficiency of a centrifugal fan by changing the volute geometric parameter, so the Kriging model is used to optimize the volute geometric parameter. Firstly, Volute radius Rd, the radius of tongue r and outlet angle of the volute θ are selected as the optimization parameters of the volute, and Latin hypercube sampling is used to configure the initial sample points, the corresponding three-dimensional aerodynamic model under each sample point configuration is constructed. CFD software is used to simulate the efficiency value and total pressure of the centrifugal blower under each initial sample point configuration; Secondly, the Kriging surrogate model of initial sample point configuration parameters, efficiency value and total pressure of volute is constructed, the high-precision Kriging surrogate model is used as the fitness function of NSGA-II algorithm to get the optimal solution. The rationality of the above method is verified by optimizing the 9-19.4A type centrifugal fan volute. The efficiency of the optimized fan under working conditions is increased by 1%, and the total pressure under working conditions is not reduced.

Keywords. Volute, Optimization, Centrifugal fan, Kriging surrogate model, NSGA-II algorithm

1. Introduction

The function of the volute is to collect and guide the gas leaving the impeller, gradually reduce the airflow speed, and turn part of the kinetic energy into static pressure. At present, Shuiqing Zhou et al.[1] proposed a Latin hypercube sampling method combined with the surrogate model to parameterize the blade profile of centrifugal fan, and multi-objective optimization is carried out under NSGA-II algorithm. Selvaraj T et al.[2] treated the blade inlet angle, impeller diameter and width of centrifugal fan as optimized parameters using orthogonal experimental design method. Shuiqing Zhou et al.[3] studied a CST function to parameterize the blade, the surrogate model combined with the intelligent method are used in the research, which provides a new idea for fan energy saving. Honggang Fan et al. [4] researched a centrifugal fan optimization method. Abolfazl Khalkhali et al. [5] proposed neural network and CFD technology in the optimization processing. Xinfeng Li et al. [6] proposed a genetic algorithm to optimize the blade outlet installation angle and the number of blades. Finally, the

[1] Fannian Meng, School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450002, China, Email: mengfannian123@163.com

compression ratio of the impeller at the design working point is increased by 3.57%, the isentropic efficiency is increased by 0.79%, and the overall condition of the impeller is improved. Konrad Bamberger et al. [7] used an evolutionary algorithm combined with surrogate model to quickly optimize the geometric parameters of centrifugal fan.

In this paper, the Latin hypercube sampling design method combined with the self-adaption Kriging surrogate model is used to optimize the centrifugal fan volute radius Rd, radius of tongue r and outlet angle of the volute θ. The response value is calculated by aerodynamics software CFX, and the NSGA-II algorithm is to find the Pareto optimal volute parameter; after optimization, the airflow distribution in the volute is smoother..

2. Volute Optimization problem description

The general mathematical model of the volute optimization problem can be described as:

$$\begin{aligned} &\text{find } \boldsymbol{x} = (x_1, x_2, \cdots, x_n) \\ &\min\ f(\boldsymbol{x}) \\ &\text{s.t. } g_j(\boldsymbol{x}) \le 0 \\ &\quad \boldsymbol{x}^{LB} \le \boldsymbol{x} \le \boldsymbol{x}^{UB} \\ &j = 1, 2, \cdots, m \end{aligned} \tag{1}$$

Where $\boldsymbol{x}$ is the n dimensions vector of design variable vector, $\boldsymbol{x}^{LB}$ and $\boldsymbol{x}^{UB}$ are upper and lower bounds value for design variable respectively, f is the objective function, g_j is the constraint condition, the analysis model for volute optimization can be described as a black-box model.

To cut down calculation cost to improve optimization efficiency, the surrogate model usually takes the place of the mathematical model. In this case, the optimization model is shown as follows:

$$\begin{aligned} &\text{find } \boldsymbol{x} = (x_1, x_2, \cdots, x_n) \\ &\min\ \widehat{f}(\boldsymbol{x}) \\ &\text{s.t. } \widehat{g}_j(\boldsymbol{x}) \le 0 \\ &\quad \boldsymbol{x}^{LB} \le \boldsymbol{x} \le \boldsymbol{x}^{UB} \\ &j = 1, 2, \cdots, m \end{aligned} \tag{2}$$

Where $\widehat{f}$ and $\widehat{g}_j$ are the surrogate model approximate response value, Kriging model, support vector machine (SVM) model, deep learning machine (DLM) model and so on usually used as a surrogate model.

In the volute optimization process, the mathematical relation equation between the volute shape and aerodynamic response value can't be directly constructed, fortunately; computational fluid dynamics (CFD) soft can be used to calculate the aerodynamic

response value with geometric volute shape, but CFD calculation is time-consuming, so surrogate mode can be used to respect CFD calculation to save time, high precision surrogate model relies on adding sufficiency sample point, so the self-adaption sample point adding method is used to enhance the surrogate model accuracy. In the article, the self-adaption Kriging model is used to set as an objection function to optimize.

3. Volute Optimization problem description

3.1. Centrifugal fan structure

Collector, impeller, and volute are the main parts of the ventilation facility; the task of the volute is to guide the gas leaving the impeller to the outlet of the volute. The overall structure of the centrifugal fan and volute structure are drawn in Figure 1.

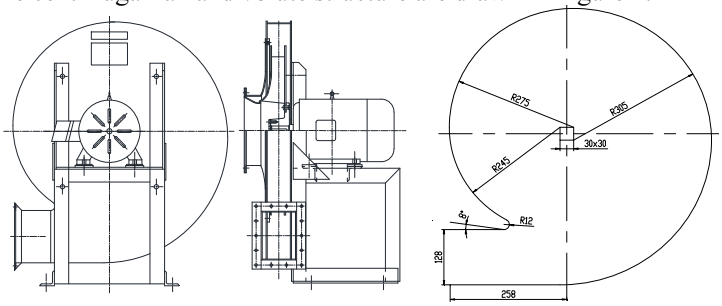


Figure 1. Centrifugal fan structure: (a) overall structure and (b) volute structure

3.2. CFD simulation analysis

In order to calculate the aerodynamic response parameters under a specific centrifugal fan model, CFD software needs to be used for numerical simulation calculation. Here, *. X_t model parameters are imported into ICEM software to divide the grid. The divided grid is shown in Figure 2.

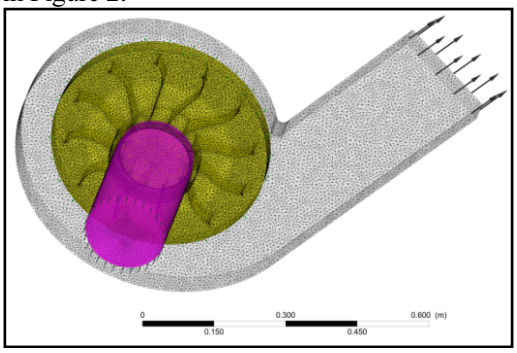


Figure 2. Grid mesh

3.3. Sample Point Design

First, volute radius R_d, radius of tongue r and outlet angle of the volute θ are selected are selected as optimization input parameters, Efficiency (*Eff*) and total pressure (*Tp*) are elected as optimization target parameters. The factors and levels are indicated in Table 1.

Table 1. Volute parameters and their levels

Factors	Parameters	Range
A	Volute radius R_d	235~255
B	Radius of tongue r	8~16
C	Outlet angle of the volute θ	0~10

The concrete optimization can be written in the below:

$$\begin{cases} \max\ Eff(R_d, r, \theta) \\ \max\ Tp(R_d, r, \theta) \end{cases} \quad \begin{cases} 235 \le Rd \le 255 \\ 8 \le r \le 16 \\ 0^\circ \le \theta \le 10^\circ \end{cases} \tag{3}$$

In the article, the LHS technology is adapted for parameter design; it is obvious that the distribution of sample points in each region of subspace is relatively uniform. The sample points also show a uniformly distributed full coverage state in the subspace formed by other factors.

3.4. NSGA- Ⅱ Algorithm optimization

NSGA-II Algorithm is adapted in the article, and the parameters are set as; Maximum evolutionary algebra 300; Population size is equal to 100; Setting the crossover parameter value as 0.9; Variation probability is 0.1; Scaling factor 0.5; The distribution index of crossover and mutation algorithm is 20. The optimal combination of optimization parameters is calculated. After the combination of the parameters is rechecked by CFD, the total pressure evaluation parameter error is 0.5% and the efficiency fitting error of the fan is 0.1%, indicating that the response surface accuracy is high and the result is reliable. The fan efficiency is increased by 0.9%. The Pareto optimal front distribution is shown in Figure 3.

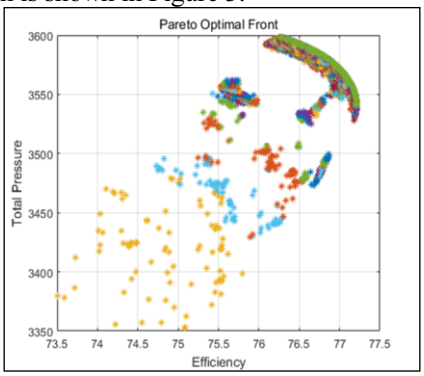


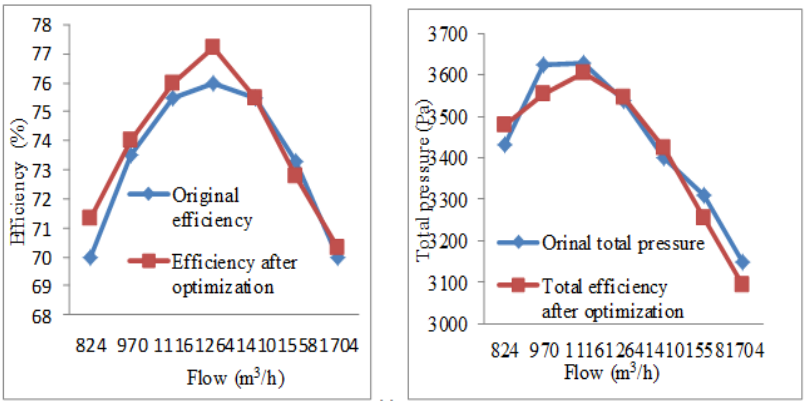
Figure 3. The Pareto optimal front distribution

Table 2 shows the Performance comparison of the centrifugal fan before and after optimization. It is known that at the design points, the total pressure value of the optimized centrifugal fan is raised from 3538Pa to 3545Pa, and the efficiency is increased from 76.3% to 77.2%. The efficiency of the optimized centrifugal fan is improved under the condition that the total pressure value is not lower than that of the original fan. The optimized centrifugal fan is more energy-saving and has better performance.

Table 2. Performance comparison of the centrifugal fan before and after optimization

Centrifugal fan	Total pressure at design point	The efficiency at design point
Before optimization	3538Pa	76.3%
After optimization	3545Pa	77.2%

Figure 4 are the centrifugal fan result curves before and after optimal design ((a) corresponds to the efficiency curve and (b) corresponds to the total pressure curve). It is known from this that under the design conditions after optimization, the total pressure of the optimized centrifugal fan is not less than of the original fan and its efficiency is increased by about 1%.

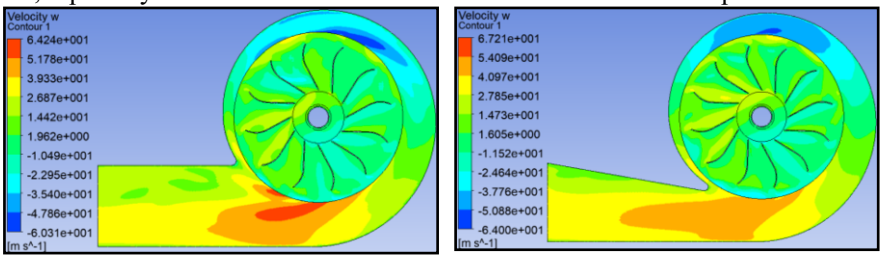


(a) **(b)**

Figure 4. Performance curve before and after optimization: (a) efficiency and (b) total pressure

3.5. Volute internal flow performance comparison before and after optimization

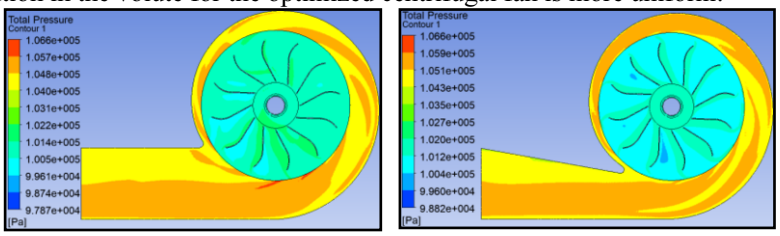
Figure 5 show the velocity distribution for the original fan and the optimized fan. It can be concluded that the velocity distribution in the volute after optimization is more uniform, especially in the area where the volute is in contact with the impeller.



(a) **(b)**

Figure 5. Velocity distribution: (a) before optimizaiton and (b) after optimization

Figure 6 show the total pressure distribution for the original centrifugal fan and the optimized centrifugal fan. It can be concluded from the figure 6 that the total pressure distribution in the volute for the optimized centrifugal fan is more uniform.



(a) **(b)**

Figure 6. Total pressure distribution: (a) before optimizaiton and (b) after optimization

4. Conclusion

In the article, the Kriging surrogate model is established, the EI adding points technology is used to raise the precision of the model, and the genetic method is used for optimization. The volute optimization design process considering the aerodynamic characteristic is proposed, and the centrifugal fan with better comprehensive performance is optimized.

- CFD method is complex and time-consuming to solve aerodynamic forces. The point-adding optimization method combining the Kriging model with EI point-adding optimization method can obtain a higher precision model with fewer sample points, which significantly improves the optimization efficiency.
- After optimizing the volute, the aerodynamic characteristic is improved, and the efficiency of the integral centrifugal fan is increased by 1%.

Funding

This paper is supported by the National Natural Science Foundation of China (No. 52075500), by the Key Science and Technology Research Project of the Henan Province (No. 212102210071).

References

[1] Shuiqing Z, Huaxin Z, Ke Y, Haobing D, Zengliang G. Research on blade design method of multi-blade centrifugal fan for building efficient ventilation based on Hicks-Henne function. Sustainable Energy Technologies and Assessments. 2021; 43: pp 1-13.

[2] Selvaraj T, Hariharasakthisudhan P, Pandiaraj S, Sathickbasha K, Mohamed Aslam Noorani AB. Optimizing the Design Parameters of Radial Tip Centrifugal Blower for Dust Test Chamber Application Through Numerical and Statistical Analysis. FME Transactions. 2020; 48: pp 236-245.

[3] Shuiqing Z, Ke Y, Weitao Z, Kai Z, Chihu W, Weiya J. Optimization of Multi-Blade Centrifugal Fan Blade Design for Ventilation and Air-Conditioning System Based on Disturbance CST Function. applied sciences. 2021; 11: pp 1-20.

[4] Honggang F, Jinsong Z, Wei Z, Bing L. Multiparameter and Multiobjective Optimization Design Based on Orthogonal Method for Mixed Flow Fan. Energies. 2020; 13: pp 1-15.

[5] Abolfazl K, Mehdi F, Hamed S. Modeling and multi-objective optimization of forward-curved blade centrifugal fans using CFD and neural networks. Transactions of the Canadian Society for Mechanical Engineering. 2011; 35(1): pp 63-79.
[6] Xingfeng L, Xiangyun Y, Guofu Y. Parameter Optimization of Centrifugal Compressor Impeller Based on CFD and Multi-objective Algorithm. Fluid machinery. 2019; 47(3): pp 31-36.
[7] Konrad B, Thomas C, Julian B, Oliver N. Development, Validation, and Application of an Optimization Scheme for Impellers of Centrifugal Fans Using Computational Fluid Dynamics-Trained Metamodels. Journal of Turbomachinery. 2020; 142: pp 1-7.

doi:10.3233/ATDE220231

Machine Vision-Based Defect Detection Method for Flexible Circuit Boards

Zuoshi LIU [a], Liang ZHONG [a,1] and Hanbin CHEN [a]
[a] *School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China*

Abstract. A machine vision-based FPC defect detection method is designed to address the problems of low efficiency of FPC defect detection. The Otsu algorithm is applied for image segmentation to decompose the image into two parts: foreground and background; the SURF algorithm is applied to achieve image alignment, and then an experimental platform is built for detecting FPC defects. For the foreign colors, foreign objects, and pressure/scratch/scratch defects in the appearance of FPC boards, detection algorithms are designed to achieve automatic recognition and classification of FPC board defects according to their defect characteristics.Finally, based on the above, by calculating the detection of 10 groups of FPC boards containing defects, the average over-inspection rate is 5.66% and the average leakage rate is 8.74%, which are better than the values of manual visual inspection in enterprises.

Keywords. Defect detection; FPC; Image registration; Image segmentation.

1. Introduction

With the massive use of modern society electronic products, FPC circuit board demand is greatly increased at the same time will be more and more inclined to thin and light, then the degree of precision and manufacturing complexity of FPC will gradually increase, at this time for the quality control of FPC circuit boards will be more stringent.

At present, the appearance of FPC defects are still mainly dependent on manual visual inspection, which is a great consumption of manpower, and the more high-density lines will make manual visual inspection extremely difficult, and in the same day under long-term operation, the spirit of workers will be greatly reduced, then it will be easy to lead to defects on the FPC board[1] Be over-checked or missed, so the board on the subsequent manufacturing process will bring unpredictable impact[2-4].

With the advent of the Internet of Things era and the rapid development of machine vision, the manufacturing industry will also win a huge innovation, the realization of machine vision-based FPC circuit board defect detection system has very great significance and value.

Therefore, for the characteristics of FPC, such as a wide range of defects, this paper focuses on the automatic identification and classification methods of foreign colors, foreign objects, and scratch defects in FPC board defects, and combines image

[1] Liang ZHONG: School of Mechanical and Electrical Engineering，Jiangxi University of Science and Technology, zhongliang202201@163.com

processing and motion control theory to achieve the purpose of replacing manual visual inspection. The Otus algorithm is used for image segmentation, the SURF algorithm is selected to achieve image alignment, and the camera is connected to the computer to build a monitoring platform, and the research of machine vision-based FPC defect detection method is carried out, and the detection efficiency is significantly improved.

2. Image processing

2.1. Image segmentation

The Otsu algorithm is the maximum interclass variance method, also known as the Otsu algorithm, and is regarded as the best algorithm among the threshold-based image segmentation algorithms, advantages of simplicity, high adaptive capacity, suitability for many scenes, and high efficiency[5].

The main steps of the OTSU algorithm[5] are as follows.

1) Calculate the normalized histogram of the input image. Using p_i , the $i = 0，1，2，3，\ ...L-1$ denotes each component of this histogram.

2) Using equation $P_1(k) = \sum_{i=0}^{k} p_i$ (1), for $k = 0，1，2，3，\ ...L-1$, calculate $P_1(k)$.

3) Using equation $m_1(k) = \sum_{i=0}^{k} i\,P(i|C_1) = \sum_{i=0}^{k} \frac{IP(i|C_1)P(i)}{P(C1)} = \frac{1}{P_1(k)} \sum_{i=0}^{k} ip_i$ (2), for $k = 0，1，2，3，\ ...L-1$, calculate $m(k)$.

4) Using equation $P_2(k) = \sum_{i=k+1}^{L-1} p_i = 1 - P_1(k)$ (3) to calculate the global grayscale mean value mG.

5) Using equation $\sigma_G^2(k^*) = \max_{0 \le \mathrm{Re}\, L-1} \sigma_B^2(k)$ (4), for $k = 0，1，2，3，\ ...L-1$ the inter-class variance is calculated $\sigma_B^2(k)$.

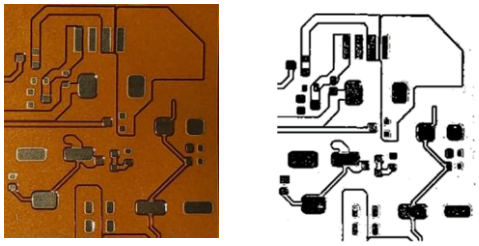
6) Obtain the Otsu threshold k^* to obtain $\sigma_B^2(k)$ the maximum k value. If the maximum value is not unique, the individual maximum k value is averaged to obtain the k^* value.

7) Calculate the separability measure at $k = k^*$ the place to calculate the divisibility metric η.

To verify the effectiveness of Otsu's algorithm, the principle is reproduced in this paper, and the core part of the program is shown in Figure 1.The rendering is shown in Figure 2 below.

```
//计算类间方差
double Var = 0;
for (int k = 0; k < Grayscale; ++k) {
    if ((MK[Grayscale - 1] * PK[k] - MK[k])*(MK[Grayscale - 1] * PK[k] - MK[k]) / (PK[k] * (1 - PK[k])) > Var)
    {
        Var = (MK[Grayscale - 1] * PK[k] - MK[k])*(MK[Grayscale - 1] * PK[k] - MK[k]) / (PK[k] * (1 - PK[k]));
        thresh = k;
    }
}
```

Figure 1. Core part of Otsu algorithm program.



A.Original image. **B**.Results.

Figure 2. Otsu algorithm effect diagram.

From the results, the segmentation effect is better, and the running time is 0.6808ms, which has good real-time performance and is the most applicable image segmentation method in this paper.

2.2. Image Alignment

Using the SURF (Speed Up Robust Features) algorithm for image alignment, the efficiency of the algorithm is improved by using integral maps and reduced dimensional feature descriptors, which improves the noise immunity of the algorithm, reduces the computational effort, and greatly improves the computational speed.[6] The algorithm has improved the noise immunity and reduced the computational effort.

The specific implementation process of the SURF algorithm is 1. Construct Hessian matrix to generate interest points for feature extraction; 2. Construct scale space[6]. The main directions of the feature points are assigned; 5. The feature point descriptors are generated; 6. The feature points are matched.

The following graph shows the experimental results of the image alignment SURF algorithm.

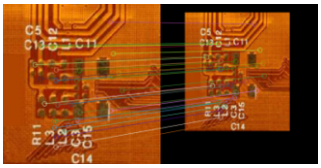


Figure 3. Processing results of image scaling.

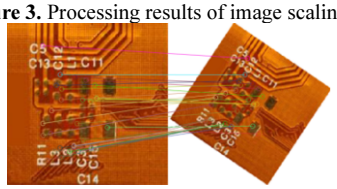


Figure 4. Processing results of image mirroring.

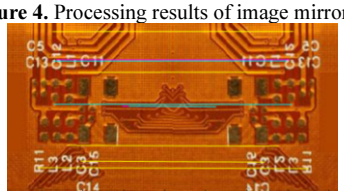


Figure 5. Processing result of image rotation.

The SURF algorithm takes very little time to perform the image alignment process, statistically 1.7997s to process the reduced image, 1.6867s to process the mirrored image, and 1.374s to process the rotated image.

3. Research on FPC defect detection and classification

3.1 Construction of testing platform

This paper builds a platform around FPC defect detection, the experimental platform is mainly composed of a computer, CCD camera, camera fixed with camera fixed linkage, lighting source, etc. Build the inspection platform shown in Figure 6, the camera is connected to the computer, in the upper computer to run the program to start the inspection.

Figure 6. FPC defect detection platform.

3.2 Defect identification

The determination of each defect is different, and the benchmarks for determining the FPC surface defects involved in this paper are shown in Table 1 below.

Table 1. FPC surface defects determination benchmark.

Serial number	Inspection items	Judgment benchmark
1	Indentation/scratch/scratch	1.There should be no pressure/scratch/scratch marks that are sharp and make the cover film broken. 2.If the pressure/scratch/scratch causes the exposed copper is considered bad and is not allowed to exist.
2	Foreign objects	Foreign objects must not make the FPC protrude significantly or crack the cover film.
3	different color	There can be no discoloration that affects the performance of the product and the service life of the customer, and for those that can be wiped need to be wiped clean before judging.

In this paper, the following two main identification methods are used.

1) Defect recognition based on color features.

Color and grayscale features are the most intuitive perceptual features. In this paper, the identification of heterochromatic and foreign defects is distinguished by color features, and in this work, the color features must be extracted accurately, so the grayscale histogram is used to extract the color features. Since the obtained image is a three-color RGB map, three-color components can be extracted separately. The specific implementation is as follows.

The gray level of the image is divided equally, and the number of pixels in each gray level interval is counted separately. Compare with the grayscale of the image without defects to observe the presence of defects.

2) Defect recognition based on geometric features.

Geometric features are also one of the main features of FPC boards, mainly describing the geometric nature of the defects, which can be achieved by a combination of the following two approaches.

Calculate the rectangularity The length of the long side and the length of the short side of the smallest external rectangle of the detected defect area are calculated using and are expressed, then there is the following equation.

$$R_t = \frac{L_l}{L_s} \tag{1}$$

Calculate the regional duty cycle: The ratio of the minimum outer rectangle of the defective region to the area of the defective target has the following equation.

$$R_q = \frac{\mathrm{S}}{L_l \times L_s} \tag{2}$$

According to equation (1) as well as equation (2), combining the two approaches, discriminative classification can be performed.

3.2.1 Indentation/scratch/scratch defect identification

Indentation/scratch/scratch type defects[7] The maximum width to length ratio of the defect can be calculated to identify the defects that arise from abrasion or scratching during the production process. That denotes the threshold value and denotes the height of the smallest outer rectangle of the defect, and denotes the length of the smallest outer rectangle of the defect, and has the following equation.

$$T = \frac{H}{L} \tag{3}$$

A pre-set maximum threshold is set and when this maximum threshold is exceeded, it can be considered as having a crush/scratch/scratch type defect, part of the core procedure is shown in Figure 7 below.

```
var defect_Reg = p.defectRegion.Connection();
for (int i = 1; i < defect_Reg.CountObj() + 1; i++)
{
    var reg = defect_Reg.SelectObj(i);
    if (reg.Area < AreaLimit) return;
    double row, col, phi, r1, r2;
    reg.SmallestRectangle2(out row, out col, out phi, out r1, out r2);
    if (r1 <= 0) return;
    if (r2 / r1 > AspectRatio) return;
    AddDefectRegion(reg, this.TargetLayer.ToString() + ScratchDetectTypeAttribute.Scratch);
}
```

Figure 7. core code of the core code for the recognition of defects of the pressure/scratch/scratch type.

3.2.2 Foreign body defect identification

Foreign matter defects are often due to the generation process, raw materials or products mixed with substances other than the required materials, there are many reasons for this phenomenon: impure raw materials, reaction with by-products, process control is not standardized, or immature process formulations, etc. Identification of foreign matter defects, followed by the analysis of the foreign matter specifically, can quickly determine the composition of foreign matter or impurities, the analysis of its causes can further improve the product yield, eliminate the production of products hidden problems, to ensure the stability of the production process.

The defects are identified based on the difference between the color of the foreign object and the color of other locations. By calculating the gray value of R, G, and B

channels of the image respectively and defining the acceptable area of the foreign object, finally, if there is a difference between the color of the foreign object and the color of other locations is greater than the foreign object judging tolerance and the area of the foreign object is also greater than the previously defined acceptable value, then it can be judged to be unqualified and belongs to the foreign object defect. Some of the core codes are shown in Figure 8 below.

```
var defectReg = p.defectRegion.Connection();
for (int i = 1; i < defectReg.CountObj() + 1; i++)
{
    var selReg = defectReg.SelectObj(i);
    if (selReg.Area < AreaSelect) return;
    double dev;
    double r = rImage.Intensity(selReg, out dev);
    double g = gImage.Intensity(selReg, out dev);
    double b = bImage.Intensity(selReg, out dev);
    double dis = Math.Sqrt(Math.Pow(r - ForeignMatterRCenterValue, 2) +
                 Math.Pow(g - ForeignMatterGCenterValue, 2) +
                 Math.Pow(b - ForeignMatterBCenterValue, 2));
    if (dis < DisThresholdValue)
        AddDefectRegion(selReg, TargetLayer.ToString() +
                        ForeignMatterDetectTypeAttribute.ForeignMatter);
}
```

Figure 8. Foreign body defect recognition part of the core code.

3.2.3 Identification of heterochromatic defects

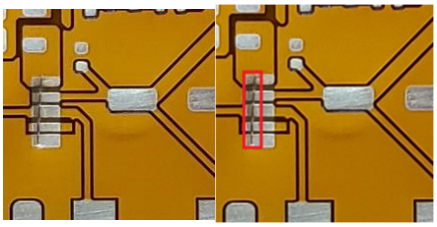
Wire discoloration, also known as heterochromatic defects, is commonly found in the production process of gold surface brushing and surface gold different and cause differences in color, similar to foreign object detection, by calculating the image of the R, G, B channel gray value, and the definition of the acceptable area of the heterochromatic, if the area is greater than the value, regardless of the color of the heterochromatic is dark brown or lighter color is as a substandard product. Some of the core codes are shown in Figure 9 below.

```
var defectReg = p.defectRegion.Connection();
Parallel.For(1, defectReg.CountObj() + 1, (i) =>
{
    var selReg = defectReg.SelectObj(i);
    if (selReg.Area < AreaSelect) return;
    double dev;
    double r = rImage.Intensity(selReg, out dev);
    double g = gImage.Intensity(selReg, out dev);
    double b = bImage.Intensity(selReg, out dev);
    double dis = Math.Sqrt(Math.Pow(r - ColorDiffRValue, 2) +
                 Math.Pow(g - ColorDiffGValue, 2) +
                 Math.Pow(b - ColorDiffBValue, 2));
    if (dis < DisThresholdValue)
        AddDefectRegion(selReg, this.TargetLayer.ToString() +
                        ColorDifferenceDetectTypeAttribute.ColorDifference);
});
```

Figure 9. Heterochromatic defect recognition part of the core code.

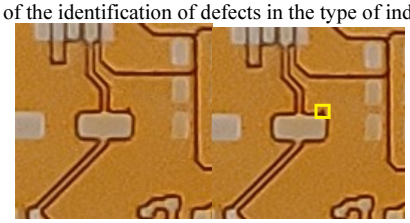
4. Analysis of experimental results

Based on the identification algorithm described above, some of the defects detected are shown in Figures 10, 11 and 12 below.



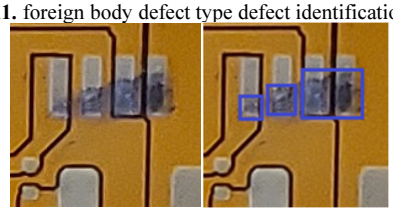
A.Original image. **B**. Results graph.

Figure. 10. The results of the identification of defects in the type of indentation/scratch/scratch.



A. Original image. **B.** Results graph.

Figure 11. foreign body defect type defect identification results.



A. Original image. **B.** Results graph.

Figure 12. heterochromatic defect type defect identification results.

From the experimental results as above, the identification algorithm used in this paper is effective and obvious. Afterward, the leakage and overdetection rates are calculated, and the number of manually detected defects is used as a criterion, which leads to the following equation.

$$\text{Missed detection rate} = \frac{\text{Number of undetected defects}}{\text{Actual number of defects}} \times 100\% \tag{4}$$

$$\text{Pass rate} = \frac{\text{Number of erroneously detected defects}}{\text{Number of defects detected by the algorithm}} \times 100\% \tag{5}$$

According to equation (4) as well as equation (5), the following experimental results can be obtained Table 2.

Leakage rate and overdetection rate in the production process is extremely important indicator, in the actual production process to achieve 0 leak detection is a very difficult thing, because in the actual detection process, whether the light source remains stable, whether the camera is blurred, in the image acquisition process is too much noise and other irresistible factors, these factors can cause the detection of leakage and overdetection phenomenon.

In the enterprise, the leakage rate of manual visual inspection is about 30%,and the over-inspection rate is about 10%, while the average leakage rate and over-inspection rate of this subject can be less than 10%, which shows that in terms of stability, this system is significantly better than manual visual inspection.

In summary, the method has some practical value and can be used in industrial production lines.

Table 2. Test data results

FPC Board number	Physical presence Number of Defects	Algorithm detection Number of Defects	Error detection Number of Defects	Not detected Number of Defects	Inspection rate	Missing detection rate
1	6	5	0	1	0.00%	16.67%
2	7	8	1	0	12.50%	0.00%
3	5	5	0	0	0.00%	0.00%
4	8	8	1	1	12.50%	12.50%
5	5	4	0	1	0.00%	20.00%
6	11	10	0	1	0.00%	9.09%
7	12	11	1	2	9.09%	16.67%
8	9	10	1	0	10.00%	0.00%
9	8	8	1	1	12.50%	12.50%
10	6	6	0	0	0.00%	0.00%
Maximum value	12	11	1	2	12.50%	20.00%
Minimum value	5	4	0	0	0.00%	0.00%
Average value	7.7	7.5	0.5	0.7	5.66%	8.74%

5. Conclusion

We design and implement automatic recognition and classification of FPC board defects, such as foreign matter, foreign color, and pressure/scratch/scratch defects in FPC board appearance. The Otsu threshold segmentation algorithm, which is capable of fast and accurate threshold segmentation of FPC images, is used for image segmentation, and the SURF algorithm is used for image alignment. We design and implement the automatic recognition and classification of FPC board defects, such as foreign matter, foreign color, and pressure/scratch/scratch defects. Firstly, the three different defects are divided into two categories: scratch defects and color defects, and then a recognition algorithm is designed for each type of defects, color defects by calculating the gray value of R, G, and B channels, and scratch defects by calculating the maximum aspect ratio. Finally, by calculating the over-and under-detection rates, the algorithm proved to be of practical value.

References

[1] Tian T, Chunlei X. Method of mirroring to generate FPC single panel: China, 105188267[P]. 2015-12-23.
[2] Qinwei Z, Min Z, Qiang G, Zhehao W. PSO-SVM-based surface defect detection of FPC pads[J]. Combined Machine Tools and Automatic Machining Technology, 2020(05): pp. 78-81. p. 85.
[3] Chegini S N, Bagheri A, Najafi F, et al. Application of a new EWT - based denoising technique in bearing fault diagnosis [J]. Measurement, 2019: pp. 275-297.
[4] Huikai F. Simulation of intelligent detection of FPC patch defects based on machine vision[J].

Computer Simulation, 2020, 37(02): pp. 385-389.
[5] Gong J. Fast recursive algorithm for two-dimensional thresholding [J]. Pattern Recognition, 1998, 31(3): pp. 295-300.
[6] Shusheng C. Research on key image processing techniques in AOI-based FPC defect detection system [D]. Guangdong: South China University of Technology, 2014.
[7] Tian T, Chunlei X. Method of mirroring to generate FPC single panel: China, 105188267[P]. 2015-12-23.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 227
J. Xu et al. (Eds.)

doi:10.3233/ATDE220232

Improve Object Detection with Knowledge Distillation

Lixiang WANG[a] , Lizuo JIN[a,1] and Jun Yan[b]
[a] *School of Automation, Southeast University*
[b]*Department of Geriatric Neurology, Affiliated Brain Hospital of Nanjing Medical University*

Abstract. Knowledge distillation is an effective method for model lightweighting. However, the previous distillation methods in object detection, in order to solve the problem of extreme imbalance between positive and negative examples, rely on manually adjusted hyper-parameters and lack generalization. And ignore the relationship information between different detection instances. Therefore, we propose a new distillation algorithm for the task. This algorithm automatically selects the Top k detection instances that need distillation most based on the network output, and fully considers the feature distillation based on the attention mechanism and the instance relationship distillation based on the Euclidean distance. Our results show that under various object detection frameworks, the student model has achieved a significant effect improvement with a lighter structure.

Keywords. Object detection, instance selection, attention mechanism, relational distillation

1. Introduction

With the vigorous development of deep learning, the accuracy of object detection models has been greatly improved. However, considering the applications of object detection models, the advanced models usually have large model parameters that consume a large number of resources, which limits their application on edge devices, such as self-driving cars or cell phones.

To solve the deployment problem of large models, a variety of model compression methods have been proposed and practiced in academia and industry, including model pruning [1], parameter quantification [2], compact structure design [3,4], and knowledge distillation [5]. Knowledge distillation is a model compression method based on knowledge transfer. It expresses the output of a teacher network with a complex structure and stronger learning ability as knowledge and guides lightweight students to learn knowledge online by designing a loss function. Knowledge distillation has an excellent performance in classification tasks, but it has many difficulties in more difficult visual tasks, such as object detection. The specific performance is as follows:

(i): Unbalanced number of foreground and background instances

[1] Corresponding Author, Lizuo JIN. School of Automation, Southeast University. E-mail: jinlizuo@qq.com.

In the object detection task, after the image has been extracted with features by the backbone network, there will be several instances to be detected. Since the image itself has far fewer foreground pixels than background pixels, this leads to an imbalance in the number of foreground and background instances. Since foreground instances are more vital in detection, the imbalance of instances seriously hurts the performance of the knowledge distillation model.

Chen[6] first addressed this question by reducing the weight of background distillation loss in the classification head and simulating the full feature map in the backbone network. Li et al [7] designed a distillation framework with a two-level detector to apply L2 distillation loss to the RPN sampling feature of the student network, which consists of randomly sampled positive and negative examples with a certain proportion of truth labels differentiation.

(ii): Lack of learning of relational information between instances

Research believes that the relationship between different objects contains valuable information in object detection. Many researchers have successfully improved the performance of detectors by enabling them to capture and exploit these relationships, such as nonlocal modules (Wang et al. [8]) and relational networks. The current mainstream knowledge distillation methods based on object detection only extract information about individual pixels or detection instances, while ignoring the relationships between different pixels (instances).

Considering the previous knowledge distillation methods applied to object detection are manually set to learn the ratio between positive and negative instances to cope with the imbalance between foreground and background regions in the detection task. And the relationship between different detection instances is not considered. Based on this, the method in this paper is shown in Figure 1 and includes the Instance Selection Module (ISM) and the Additional Distillation Module (ADM).

- **ISM**: Instead of relying on truth labels and manually setting the ratio between pos && neg instances, our approach automatically selects the instances to be distilled based on the inference results of the teacher and student network.
- **ADM**: Guiding the model to learn background and foreground knowledge to generate different feature responses through feature distillation based on attention mechanism; learning inter-instance relationship information through a measure of instance feature distance.

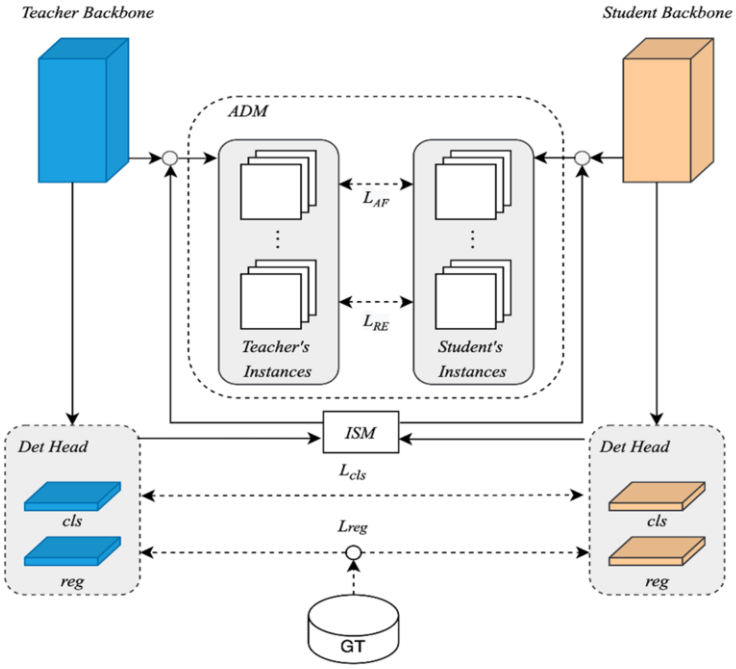


Figure 1. Schematic diagram of the model.

2. Instance Selection Module

In the object detection model, each instance predicted with a detection box is usually a meaningful region, and the difference in accuracy between the teacher and the student model is reflected in the difference in instances. ISM quantifies the difference between each instance.

We consider two aspects of quantifying instance differences: score and image area, defined as follows:

$$P_{ISM}^{m} = \max_{0 \le c \le CLS} |P_T^{mc} - P_S^{mc}| \tag{1}$$

$$B_{ISM}^{m} = \begin{cases} B_T^m, & \max\limits_{0 \le c \le CLS} P_T^{mc} > \max\limits_{0 \le c \le CLS} P_S^{mc} \\ B_S^m, & \max\limits_{0 \le c \le CLS} P_T^{mc} \le \max\limits_{0 \le c \le CLS} P_S^{mc} \end{cases} \tag{2}$$

$$ISM = NMS(P_{ISM}, B_{ISM}) \tag{3}$$

Where P_{ISM}^{m} and B_{ISM}^{m} respectively represent the screening scores and regions of interest for each instance.P_{ISM} and B_{ISM} is the set of all instances.m is the index of the set of instances. CLS and c are the total number of categories and the category indexes, respectively.P_T^{mc} and P_S^{mc} are the instance classification scores of the teacher and student model under the current instance in the one-stage algorithm, and RPN prediction scores for the two-stage algorithm. Similarly, the B_T^m and B_S^m are the

border regions of the output of the teacher and the student model under the current instance.

ISM takes the L_1 distance of the classification score as the instance score. And selects the box with the higher score as the instance box filtered by ISM. Considering that after ISM screening, the instances with higher P_{ISM}^{m} may have highly overlapping boxes with redundant information thereby substantially increasing the computational effort during distillation training. Therefore, we use NMS to filter the redundant boxes.

3. Additional Distillation Module

3.1. Feature Distillation Based on Attention Mechanism

To focus on the response differences between foreground and background during distillation training, we use an attention mechanism to generate attention masks for the feature maps of the detection head.

Denote $F \in \mathbb{R}^{C,H,W}$ as the feature of the detection head in the network, where C,H,W denote its number of channels, height and width, respectively. Considering both spatial attention and channel attention, we obtain the respective attention representations by GeM pooling as follows.

$$\boldsymbol{A^c}(F) = \left(\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left|F_{.,i,j}\right|^{p_c}\right)^{\frac{1}{p_c}} \tag{4}$$

$$\boldsymbol{A^s}(F) = \left(\frac{1}{C}\sum_{k=1}^{C}\left|F_{k,.,.}\right|^{p_s}\right)^{\frac{1}{p_s}} \tag{5}$$

Where i,j,k are the indexes of F in (height, width, channel,) respectively. p_c and p_s are the parameters of GeM pooling in channel attention and GeM pooling in spatial attention, respectively, which are obtained by model training. Based on the attention representation, the attention masks used in distillation can be generated as follows:

$$M^s = HW \cdot softmax\left(linear\left(\boldsymbol{A^s}(F^{\mathcal{S}}), \boldsymbol{A^s}(F^{\mathcal{T}})\right)/T\right) \tag{6}$$

$$M^c = C \cdot softmax\left(linear\left(\boldsymbol{A^c}(F^{\mathcal{S}}), \boldsymbol{A^c}(F^{\mathcal{T}})\right)/T\right) \tag{7}$$

Where $F^{\mathcal{S}}$ and $F^{\mathcal{T}}$ denote the features corresponding to the student and the teacher model, respectively, T is emperature parameter in knowledge distillation. We encourage the student model to mimic the features of the teacher model through L_2 parametric loss masked by M^s and M^c, which is formulated as:

$$L_{AF} = \left(\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{C}\left(F_{i,j,k}^{\mathcal{T}} - F_{i,j,k}^{\mathcal{S}}\right)^2 \cdot M_{i,j}^{s} \cdot M_{k}^{c}\right)^{\frac{1}{2}} \tag{8}$$

3.2. Instance Relationship Distillation Based on Distance Metric

In the object detection task, instances in the same example are correlated, no matter foreground or background, and this point can promote student model converge more quickly. Here we take Euclidean distance to measure the relationship between different instance features and represent the difference information between the relationships by L_1 loss. For the instances output by ISM, to calculate the Euclidean distance of the instance and to avoid the bias of the object size on the calculation of the Euclidean distance of the features, we resize the feature corresponding to the instances to the same size by ROIAlign.

$$L_{RE} = \sum_{(i,j)\in K^2} \left| \frac{1}{Norm(t)} \left\| t_i - t_j \right\|_2, \frac{1}{Norm(s)} \left\| s_i - s_j \right\|_2 \right| \tag{9}$$

$$Norm(x) = \frac{1}{|K^2|} \sum_{(i,j)\in K^2} \left\| x_i - x_j \right\|_2 \tag{10}$$

For the instances filtered by ISM, $K^2 = \{(i,j) | i \neq j, 1 \leq i,j \leq K\}$, $Norm(x)$ is the normalization function. t_i and s_i are the ROIAlign-adjusted features of the i^{th} instance of the teacher and student model, respectively.

4. Overall Loss Function

Denoting the classification outputs of the student and teacher networks as Z_s and Z_t. Considering the temperature T in knowledge distillation, the distributions of the respective outputs are defined as $P_t = softmax(Z_t/T)$ and $P_s = softmax(Z_s/T)$, and the classification loss is defined as:

$$L_{cls} = -\sum P_t \log P_s \tag{11}$$

In addition to considering the classification loss, we also calculate the regression loss to adjust the position and size of the boxes. Chen et al [8] analyzed the distillation loss of regression and classification differently, the teacher's regression output may provide the very wrong guidance to the student model because the regression output of real values is unbounded and the teacher's model may provide a regression direction that contradicts the true direction.

$$L_{reg} = \begin{cases} \|R_s - y\|_2^2, & if\ \|R_s - y\|_2^2 + margin > \|R_t - y\|_2^2 \\ 0, & otherwise \end{cases} \tag{12}$$

Where y is the true label, and R_s and R_t are the regression outputs of the student model and the teacher model, respectively. The overall loss function is as follows.

$$L_{total} = L_{cls} + L_{reg} + \alpha_1 L_{AF} + \alpha_2 L_{RE} \tag{13}$$

Where α_1 and α_2 are the hyper-parameters used to balance the L_{total}.

5. Experiments

5.1. Validity and Robustness Experiments

To validate the validity and robustness of the method, we conduct experiments on different detection frameworks using different backbone networks. The MS COCO2017 dataset is used for the evaluation of our knowledge distillation method, and we consider mainstream one-stage and two-stage object detection frameworks, including Faster RCNN, Cascade RCNN and RetinaNet , while ResNet-50 and ResNet-101 are used as the backbone networks for each object detection framework in our evaluation. We used RetinaNet with a backbone network of ResNeXt-101 as the teacher network for the one-stage model distillation training and Cascade Mask RCNN with a backbone network of ResNeXt-101 as the teacher network for the two-stage model distillation training. The experimental results are shown in Table 1.

Table 1. Experimental results of validity and robustness.

Methods	Backbone	mAP	AP_{50}	AP_{75}
Faster RCNN	ResNet50	37.3	58.8	42.6
+ Ours	ResNet50	39.2	60.1	44.9
Faster RCNN	ResNet101	38.2	59.7	43.1
+ Ours	ResNet101	41.6	62.9	46.3
Cascade RCNN	ResNet50	41.2	59.2	44.1
+ Ours	ResNet50	44.1	62.5	47.9
Cascade RCNN	ResNet101	42.5	60.5	46.7
+ Ours	ResNet101	45.4	63.3	49.1
RetinaNet	ResNet50	36.1	56.7	39.2
+ Ours	ResNet50	38.2	58.4	41.7
RetinaNet	ResNet101	37.4	58.2	41.3
+ Ours	ResNet101	39.8	60.9	44.1

The experimental show that our algorithm obtains effectual improvement on several object detectors with strong robustness. The average mAP improvement is 2.7 on Faster RCNN, 2.9 on Cascade RCNN, and 2.3 on RetinaNet. mAP of the same object detection algorithm exceeds that of ResNet101 when the backbone network is ResNet50 using our distillation training method.

5.2. Ablation Experiments

We perform ablation experiments on the ISM in the method as well as the two-loss terms added in the ADM. The one-stage method RetinaNet with ResNet-101 as the backbone network and the two-stage method Faster RCNN with ResNet-50 as the backbone network are considered respectively, and the results are shown in Tables 2 and 3, where 0 means that this module is not in the experiment and 1 means that this module is added in the experiment.

Table 2. Ablation experiments on RetinaNet (backbone=ResNet101).

Modules	ISM	0	1	0	1	1	1
	$ADM-L_{AF}$	0	0	1	0	1	1
	$ADM-L_{RE}$	0	0	1	1	0	1
Results	mAP	36.1	36.5	37.3	37.7	38.0	38.2
	AP_{50}	56.7	57.1	57.6	57.9	58.1	58.4
	AP_{75}	39.2	39.8	40.2	41.0	41.3	41.7

Table 3. Ablation experiments on Faster RCNN (backbone=ResNet50)

Modules	ISM	0	1	0	1	1	1
	$ADM-L_{AF}$	0	0	1	0	1	1
	$ADM-L_{RE}$	0	0	1	1	0	1
Results	mAP	37.3	38.6	38.5	38.8	39.0	39.2
	AP_{50}	58.8	59.4	59.2	59.7	59.8	60.1
	AP_{75}	42.6	43.4	43.9	44.5	44.8	44.9

The experimental results show that ISM improves the mAP by 0.4 and 1.3 in both experimental groups, respectively. ADM improves the mAP by 1.2 in both experimental groups, and $ADM-L_{AF}$ is more effective, improving the mAP by 1.9 and 1.7 in combination with ISM, respectively. these experimental data show that each module of our method can improve the distillation effect, and the combination of them can achieve the best distillation effect.

5.3. Application Experiments

We transfer the knowledge distillation method to the accuracy recovery training after model pruning, using the unpruned network as the teacher network and the pruned network as the student network. We used yolov3 in our experiments and performed channel pruning on the backbone network Darknet-53 following the method of Liu et al [2]. The experiments were based on the Visdrone 2019 dataset, trained and tested on Nvidia GeForce RTX 3090, and different channel pruning ratios were tried:

Table 4. Accuracy recovery training based on yolov3.

Ratio	0%	20%	50%	70%	80%	90%
Pruned map	45.6	43.1	37.3	30.8	24.3	19.2
(+ fine-tuned) map	-	44.5	41.6	41.3	37.5	35.2
(+ fine-tuned + distilled) map	-	45.7	45.3	45.9	45.4	45.7
Model size / (M)	248	226	158	79	38	13
Inference time / (s)	0.0316	0.0293	0.0236	0.0173	0.0152	0.0121

From the experimental results, we can see that when the pruning ratio is high, it is not enough to do accuracy recovery by subsequent fine-tuned training only. By adding our distillation method to the accuracy recovery training, the accuracy recovery effect can be well achieved, thus keeping the model size greatly reduced without dropping points in accuracy and shortening the model inference time by 62%.

6. Conclusion

For the characteristics of the object detection task, we adopted the ISM method of automatically filtering distillation instances based on the model results. At the same time, we enhanced the learning of feature map responses in distillation training based on an attention mechanism and considered learning of instance relationships based on distance metrics. From the results of the ablation experiments, both ISM and ADM can bring the improvement of model effects, and the best distillation effect can be achieved when used together. Meanwhile, our method combined with the model pruning method can greatly reduce the model inference time without losing model accuracy, which is a strong reference for model deployment on edge or mobile devices.

References

[1] Zhuang L, Jianguo L, Zhiqiang S. et al. Learning efficient convolutional networks through network slimming.IEEE international conference on computer vision. 2017; pp 2736-2744.
[2] Aojun Z, Anbang Y, Yiwen G, et al. Incremental network quantization: Towards lossless cnns with low-precision weights. arXiv preprint.2017; 1702.03044.
[3] Mark S, Andrew H, Menglong Z, et al. Mobilenetv2: Inverted residuals and linear bottlenecks.IEEE conference on computer vision and pattern recognition. 2018; pp 4510-4520.
[4] Ningning M, Xiangyu Z, Hai-Tao Z, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design. European conference on computer vision (ECCV). 2018; pp 116-131.
[5] Geoffrey H, Oriol V, Jeff D. Distilling the knowledge in a neural network (2015). arXiv preprint arXiv:1503.02531. 2015; p 2.
[6] Guobin C, Wongun C, Xiang Y, et al. Learning efficient object detection models with knowledge distillation. Advances in neural information processing systems. 2017; p 30.
[7] Quanquan L, Shengying J, Junjie Y. Mimicking very efficient network for object detection.IEEE conference on computer vision and pattern recognition. 2017; pp 6356-6364.
[8] Tao W, Li Y, Xiaopeng Z, et al. Distilling object detectors with fine-grained feature imitation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019; pp 4933-4942.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 235
J. Xu et al. (Eds.)

doi:10.3233/ATDE220233

Operation and Maintenance and Development Trend of 1.5Mw Wind Turbine in Xinjiang

Yulong CHEN[a],Xue HU[a,b,1],Lixin ZHANG[a,b] and Xiongfei ZHENG[a]
[a]*Shihezi University,Xinjiang Shihezi,832003,China*
[b] *Industrial Technology Research Institute,Xinjiang Shihezi,832003,China*

Abstract.With the rapid development of wind power industry in Xinjiang, the 1.5MW wind power generating unit has a large use base. After running for a period of time, the problems of maintenance and repair after its failure appear on paper.This article in view of the xinjiang region of the 1.5 MW wind turbine variable flow system, electric control system, generator system, variable propeller system, yaw system, hydraulic system and other aspects of the fault, has carried on the detailed statistics and analysis, combining with the actual production situation puts forward the solutions to these faults daily operations, and put forward combined with preventive maintenance operations strategy.Correct operation and maintenance is the guarantee of improving productivity. Finally, the paper points out that intelligent operation and maintenance is the future development trend of wind power operation and maintenance, which can effectively promote the improvement of wind farm operation and maintenance management level.

Keywords.Wind Turbines; Operations; Fault diagnosis; Maintenance

1. Introduction

Xinjiang has nine major wind regions and is extremely rich in wind energy resources. The proven reserves account for about one-fifth of the country's total, reaching 890 million kilowatts. In recent years, Xinjiang has accelerated the construction of "three bases and one channel", made every effort to develop the wind power industry, and strived to promote the transformation of the energy structure. After more than 30 years of development, Xinjiang's wind power industry has achieved remarkable results. As of October 27, 2020, the installed capacity of Xinjiang's wind power equipment has jumped over the 20 million kilowatt mark, reaching 20.099 million kilowatts, and has gradually developed into a Xinjiang power grid. The second largest type of power supply [1-3].

The 1.5MW wind turbines accounted for 65.4% of the total installed capacity of all wind turbines in Xinjiang in 2013. The rapid development has made the 1.5MW wind turbines have a very large base and is currently one of the main models of wind power generation in Xinjiang. After a period of production and operation of 1.5MW wind turbine, it faces problems such as daily operation and maintenance and maintenance after failure [4].

[1] Xue HU: Shihezi University, Email：huxue@foxmail.com

Xinjiang wind farms, especially some wind farms located in southern Xinjiang, are usually located in deserts, mountains and other inaccessible places, with inconvenient transportation and long distances from cities, coupled with poor communication in some areas, which greatly affects the operation and maintenance management of wind farms, and the failure of wind farms. Post-overhaul brings a lot of challenges [1,5]. The operation and maintenance of wind power equipment must be analyzed through practical and specific analysis. Based on this research, the development of feasible solutions and operation and maintenance management methods can significantly reduce operating costs and increase operating efficiency.

2. Fault characteristics of 1.5MW wind turbine in Xinjiang

Xinjiang wind turbines operate in a poor natural environment, and each system is extremely prone to failures under the influence of alternating loads and large temperature differences between morning and evening. In order to accurately grasp the fault characteristics of 1.5MW wind turbines, the total failures of all 1.5MW wind turbines in a large wind farm in Xinjiang for a total of 8 months from January 2018 to August 2018 were counted, and a total of 318 failures occurred, and the failures were mainly concentrated. For IGBT (Insulated Gate Bipolar Transistor) fan feedback loss, speed comparison failure, generator circuit breaker failure, pitch capacitor voltage unbalance, pitch substation bus failure, pitch position comparison failure, motor side capacitor fuse feedback Loss of feedback, loss of capacitor fan feedback, main control cabinet UPS battery failure, etc., the total number of these failures accounted for more than 80% of the total number of failures.

By classifying all the faults, it can be found that the faults of Xinjiang 1.5MW wind turbine mainly occur in the converter system, electronic control system, generator system, pitch system, yaw system and hydraulic system. The proportion of faults in each system is shown in Figure 1. It can be seen from Figure 1 that the faults are mainly concentrated in the converter system, electronic control system, generator system, pitch system and yaw system.

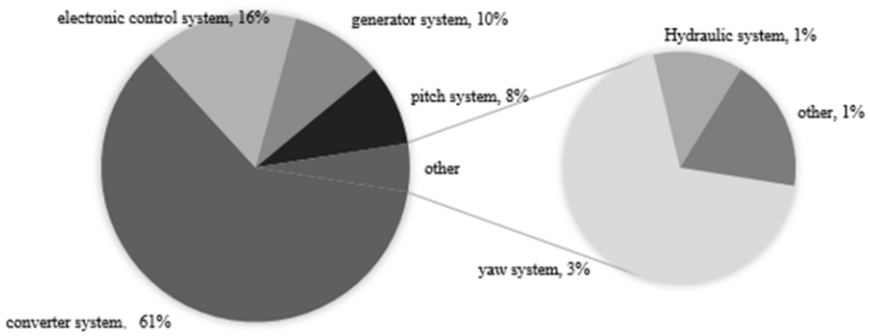


Figure 1. The proportion of failures in each system

3. Common fault analysis and operation and maintenance suggestions

3.1.converter system

The main function of the converter system in the entire wind power generation system is to change the wind energy into electrical energy suitable for the grid, and feed the electrical energy back to the grid. The electricity generated by the generator is alternating current. At this time, the frequency of the electricity, including the voltage, changes. In addition, it also changes with the change of the impeller speed. The electricity needs to be passed through the rectifier unit and then through the chopper boost system. It is sent to the DC busbar, and then inverted by the inverter unit, so that it can be matched with the grid, and finally fed back to the grid [6-7].

The faults of the converter system are mainly concentrated in the loss of IGBT fan feedback, the failure of the generator circuit breaker, the loss of the capacitor fan feedback, the current imbalance and the overcurrent of the converter braking system. Among them, IGBT_ok is lost (IGBT feedback is lost), IGBT fan feedback is lost, current is unbalanced, grid voltage is low, and the longest fault duration of generator circuit breaker fault is more than one hour. Poor on-site environment, heavy pollution and some harsh conditions such as electromagnetic interference will affect the performance of the converter system, and it is also very easy to cause failures.

Table 1.Converter system fault statistics

Classification	Fault description	Occurrence	Proportion	Maximum failure time
converter system	IGBT_ok lost	4	2%	3:18:41
	IGBT fan feedback lost	136	70%	5:43:24
	Inverter braking system overcurrent	6	3%	0:14:54
	Charge contactor feedback lost	2	1%	0:00:53
	current imbalance	8	4%	5:17:57
	Capacitor Fan Feedback Lost	11	6%	0:31:43
	low grid voltage	1	1%	9:49:31
	Generator breaker failure	20	10%	1:55:14
	Bottom fan feedback lost	5	3%	0:25:17

During the operation and maintenance of the converter system, in the face of serious faults, necessary inspections are required, and remote reset operations are strictly prohibited. In order to ensure that the tower door is tightly sealed, check and update all the sealing strips on the cabinet and the tower, and ensure that all cabinet doors on the cabinet are locked. For units with ultra-low inlet (outlet) valve pressure, carefully check the pipes of the water cooling system. If there is a shortage of water, immediately carry out water replenishment operations. If there is water leakage, repair it immediately. Finally, it must be noted that if the unit experiences a long-term power failure, dehumidification must be performed before powering on again [8-9].

3.2.Electronic control system

The control system of the wind turbine is composed of some modules that can realize information transmission. It plays an extremely important role in the entire wind power

system. Its main function is to sense all the devices in the entire wind power system, so as to achieve monitoring In order to achieve the smooth operation of the system [10].

The faults of the electronic control system mainly include the bus fault of the pitch substation, the OK fault of the safety chain (the feedback of the safety chain is lost), the feedback loss of the capacitor fuse on the motor side, and the fault of the UPS battery of the main control cabinet. shown, where the safety chain OK fault and motor side capacitor fuse feedback loss, the longest fault time is more than four hours. The working environment of wind turbines is harsh, and the work is accompanied by high temperature and vibration, resulting in a high failure rate of the electronic control system.

Table 2. Electronic control system fault statistics

Classification	Fault description	Occurrence	Proportion	Maximum failure time
Electronic control system	Pitch substation bus fault	12	24%	0:19:56
	LVD control cabinet temperature is high	2	4%	0:11:41
	Safety chain OK failure	7	14%	4:54:57
	Motor side capacitor fuse feedback lost	12	24%	7:41:07
	high cabin temperature	2	4%	0:01:40
	Impeller lock not fully released	2	4%	0:02:15
	The main circuit breaker fails to close	4	8%	0:02:26
	Main control cabinet UPS battery failure	10	20%	0:34:09

In the operation and maintenance process of the electronic control system, the first step is to check the working status of the PLC, whether a crash condition occurs, whether the PLC display and operation interface are normal, and whether the PLC shell is overheated. If the PLC check is normal, carry out the second-step communication check, including the working conditions of the module power supply, the wiring between the communication modules and other main contents. If the communication check is normal, proceed to the third step of the safety chain inspection to carefully check all nodes in the safety chain loop to ensure that all problems in the safety chain loop can be found [11-12].

3.3.Pitch system

The pitch system installed in the fan hub can control the power of the unit by changing the pitch angle of the wind rotor, and can also be used as an air brake to make the wind rotor feather brake, which is an important protection for the wind turbine. and control devices [13-14].

The faults of the pitch system include the pitch position comparison fault, the unbalanced pitch capacitor voltage, the pitch 5° proximity switch fault and the high temperature of the pitch motor. Balance, pitch position comparison failure and pitch 5° proximity switch failure all have a maximum failure time of more than ten hours. During the operation of the fan, the hub of the fan is in a state of continuous rotation, which causes all the components inside the pitch system to bear the pulsating load caused by the centrifugal force and the continuous change of gravity in the direction. Poor conditions are prone to failures [15].

Table 3. Pitch system fault statistics

Classification	Fault description	Occurrence	Proportion	Maximum failure time
Pitch system	Pitch capacitor voltage unbalance	12	44%	10:52:40
	Pitch motor temperature is high	1	4%	0:02:49
	Pitch 5° proximity switch failure	2	7%	12:24:01
	Pitch position comparison failure	12	44%	17:19:46

During the operation and maintenance of the pitch system, in order to ensure that the braking torque of the pitch motor meets the requirements for use, it is necessary to check the working conditions of the pitch motor brake pads on time. If the wear is serious, it should be replaced in time. In addition, in the process of operation, if the limit switch is found to be faulty, it should be replaced immediately. At the same time, pay attention to the aging of the battery and the tightness of the mechanical and electrical connections. Finally, make sure to clean the slip ring regularly. The pitch system can significantly reduce the occurrence of failures by performing component inspections and regular maintenance in daily operation and maintenance [16-17].

3.4.Generator system

The generator in the wind turbine is used to convert mechanical energy into electrical energy and achieve the purpose of supplying power to the power grid. The generator is prone to failures when it works continuously under various variable conditions including electromagnetic interference [18].

The faults of the generator system are mainly speed comparison faults. As shown in the fault statistics of the generator system in Table 4, the fault occurred 31 times in eight months. The fault may be caused by loose wiring of the generator system, failure of the impeller speed proximity switch, failure of the speed measurement circuit fuse, damage to the system module, deformation of the speed detection disc, etc.

Table 4. Generator system fault statistics

Classification	Fault description	Occurrence	Proportion	Maximum failure time
Generator system	Speed comparison failure	31	100%	0:39:09

When a speed comparison fault occurs in the generator system, the following methods can be used to check during the operation and maintenance process: The first step is to observe the signal detected when the impeller is in a free rotation state, and determine whether the proximity switch is faulty. disposal and replacement. The second step is to check the Overspeed, Gspeed and Gpulse modules to see if the wiring between these modules is correct. At the same time, check whether the insurance of the Gpulse circuit works. If the above problems exist, the wiring needs to be corrected. The third step is to check the speed proximity switch. On the premise that the impeller is locked, determine whether the distance from the generator speed detection plate meets the actual needs, and at the same time, check whether the grounding of the shielding layer can work normally, and observe by placing a metal object on the top. If the fault still exists after completing the first three steps of troubleshooting, you need to use the troubleshooting method to find the fault, replace the same module between the units, and observe whether the fault is transferred to the replacement unit [19].

3.5.Yaw system

The unique servo system of the wind turbine is the yaw system, also known as the wind system, which is a part of the wind turbine cabin. Quasi-wind direction [20].

The faults of the yaw system are mainly left yaw feedback loss, yaw position overrun and yaw speed fault, as shown in Table 5 yaw system fault statistics, among which the left yaw feedback loss not only occurs more frequently, but also has the longest fault duration. more than three hours.

Table 5. Fault statistics of yaw system

Classification	Fault description	Occurrence	Proportion	Maximum failure time
Yaw system	Yaw speed failure	1	9%	0:03:17
	Yaw position overrun	2	18%	0:34:19
	Left yaw feedback lost	8	73%	3:11:15

The left (right) yaw feedback is lost in the wind turbine. During the operation and maintenance process, check whether the wiring in the system is correct, whether the contactor, yaw solenoid valve, yaw motor resistance, fan yaw and air switch are correct. Normal operation, whether there is peculiar smell in the cabin, etc., such faults should be analyzed in detail, and targeted analysis and processing should be carried out for the fault phenomenon [21-23].

3.6.Hydraulic system and others

The 1.5MW wind turbines that have been put into operation in Xinjiang all use variable pitch technology [24]. In the variable pitch wind turbine, the main function of the hydraulic system is to control the power of the wind turbine by controlling the pitch mechanism. As well as the rotational speed, at the same time, it can also control the drive yaw reducer and the mechanical brake mechanism [25].

The failure of the hydraulic system is mainly due to the low oil level of the hydraulic system, as shown in the statistics of the hydraulic system and other failures in Table 6. The possible cause of this failure is that there is an oil leakage in the hydraulic system. Due to the existence of the oil leakage point, the hydraulic oil inside the system leaks out, resulting in a low oil level. Other failures mainly include anemometer failure and abnormal operation of the cabin acceleration sensor.

Table 6. Statistics of hydraulic system and other faults

Classification	Fault description	Occurrence	Proportion	Maximum failure time
Hydraulic system	Hydraulic system oil level is low	2	40%	0:06:18
others	Anemometer failure	1	20%	0:01:13
	The cabin accelerometer works abnormally	2	40%	1:27:57

If the oil level of the hydraulic system is low, check the following procedures in sequence: whether each oil pipe in the hydraulic system is damaged, whether the sealing gasket of the body is damaged, whether the sealing gasket at the interface of the high-pressure filter is damaged, and check the oil level of the oil level gauge. Check the oil level of the oil level gauge again after the oil filling is completed. If it is not normal, the

oil level gauge is damaged and needs to be replaced. According to the detected oil leakage point, select the appropriate maintenance method [26-27].

The causes of the anemometer failure may be the damage of the anemometer, the loose connection of the anemometer loop, or the damage to the measurement module corresponding to the anemometer. In the operation and maintenance process, first observe whether the wind speed data has obvious jumps. If there is, it may be caused by a problem with the resistance. At this time, the resistance needs to be replaced. Next, check the measurement module of the anemometer. If it is abnormal, the module needs to be replaced. At the same time, check whether the wiring of the anemometer is loose. If it is loose, strengthen the wiring. Finally, if the fault still exists, the anemometer needs to be replaced[28].

The abnormal operation of the cabin acceleration sensor may be caused by loose or disconnected feedback signal loop wiring, damage to the acceleration module, damage to the measurement module, extreme severe wind conditions, and sensor signal interference. During the operation and maintenance process, first judge whether the fault is caused by extreme wind conditions according to the wind condition information at that time, and then check whether the corresponding wiring is loose according to the drawings, and tighten the wiring at the same time, and then check whether the internal wiring is burnt. If not, you can try to replace the acceleration sensor, and finally check whether the indicator light of the module is abnormal, and replace it if it is abnormal.

4. Preventive Maintenance

During the daily operation and maintenance of wind farms, regular inspection and post-fault maintenance should be combined with preventive maintenance, including technical means such as spectrograph analysis, vibration analysis and temperature recorder analysis. By adding preventive maintenance, potential problems can be predicted before equipment failures, and related components can be dispatched and replaced in time, reducing equipment failure rates, reducing downtime, increasing power generation, and better managing spare parts inventory, providing flexible and flexible adjustment maintenance cycle [29-32].

4.1.Spectrograph analysis

Every six months is a cycle, take part of the oil samples and send them to the laboratory for analysis to detect the viscosity, moisture, suspended particles and pH in the oil samples. If the increase of the above indicators exceeds the limit value, it means that new oil needs to be replaced.

The change in viscosity indicates that the oil is in extreme temperature conditions, and the additives in it have lost their function; if the moisture exceeds the upper limit, the lubricating performance will be reduced, and the maximum allowable upper limit for the fan is 2%; the increase of suspended particles indicates that the components inside the gearbox are seriously damaged Wear; an increase in total acidity indicates an increase in the acidity of the oil, which will cause the internal components of the gearbox to be susceptible to corrosion.

4.2 Vibration Analysis

The vibration of key parts is regularly detected with a vibration analyzer, and the analysis of the vibration signal can reflect the internal operation of the machine. Displacement detection can be used to detect low frequency vibration, velocity detection can be used to detect high frequency vibration, which can reflect unbalance, misalignment and loose foundation, and acceleration detection can be used to analyze bearing operation and high frequency vibration.

For example, in the fan impeller, by monitoring the vibration of the blade, the state characteristics of the blade during operation are extracted, and then the self-associative artificial neural network is used to analyze the fatigue of the blade to identify and identify the faults in the blade. In the gearbox, the frequency spectrum of the high-speed horizontal axis and the time-domain waveform can be observed by the vibration analysis method. If there are some peaks exceeding the threshold in the frequency spectrum, combined with multiple frequency spectra, it means that the deviation between the generator and the gearbox is serious. middle.

4.3 Analysis of temperature recorder

The temperature is measured at various points, and an abnormal rise in temperature will indicate a problem. Temperature recorders (infrared cameras) are used to identify hot spots on electrical equipment, or other hard-to-reach areas of machinery.

Whether the temperature is normal or not is an important observation method for judging whether the bearing is abnormal. If the temperature rises abnormally, it means that the bearing is in an abnormal state. Excessive temperature will cause the working clearance of the bearing to decrease rapidly and the thermal stress to increase rapidly. The problem of frictional heat generation will be aggravated by this. At the same time, the bearing will generate a large amount of heat in a short period of time, which will cause large thermal stress deformation, thereby reducing the working efficiency and reliability of the bearing.

5. Development Trend

With the development of computer technology and data acquisition technology, a unified online monitoring system platform has been established. Some of the existing advanced monitoring methods include unit online condition monitoring system, expert system, oil online monitoring system and big data, etc. By adopting the online condition monitoring system, predictive maintenance can be carried out, excess maintenance can be eliminated, and equipment availability can be improved, increasing the power generation [33]. The use of expert systems can provide a basis for preventive maintenance of wind farms and improve fault diagnosis capabilities [34]. Using the oil online monitoring system, the wear condition of the gearbox can be judged by monitoring the oil quality used, which helps to find faults in advance [35-36]. Through the use of big data, and the fusion and deep mining of big data, the efficiency of fault diagnosis can be effectively improved [37].

Through the construction of a unified online monitoring operation and maintenance system, effective results have been shown in improving operation and maintenance efficiency, reducing operation and maintenance costs, and intensive management. With

the continuous development of "Internet +" technology, the operation and maintenance of wind turbines will become more intelligent, and the intelligent operation and maintenance of the whole life cycle will definitely be the key and core technology of wind power operation and maintenance. Intelligent operation and maintenance is the comprehensive use of Internet technology and big data processing technology, combined with control technology and sensor technology, by matching the big data platform and the Internet of Things platform built by wind turbines, to coordinate the operation, management, monitoring, maintenance and other work, so as to realize wind power generation. An intelligent operation and maintenance model that maximizes the economic benefits of the farm.

Application Scenario 1: Analyzing Cross-Application Systems

At present, the operation and maintenance monitoring system of most wind farms is composed of multiple platforms, and different platforms are usually completed by different development teams. The development environment and code language between platforms are inconsistent, which leads to various internal monitoring systems. The platforms work independently and lack integrity and unified management capabilities. Based on the existing IT monitoring system and log data of the enterprise through the cloud smart business operation and maintenance platform, the intelligent data collector is used to obtain performance-related indicator data. The business process is completely sorted out and displayed.

Application Scenario 2: Fault Location and Analysis

The unified online monitoring operation and maintenance system will be set to a fault warning mode, the faults will be sorted by level, the severity of the fault will be distinguished, the priority of the fault will be determined, and the fault will be sent to the operation and maintenance management department of the wind farm as soon as possible. ” function, the site photos, fault location and root cause analysis are sent in time, so that the site operation and maintenance personnel can diagnose the fault in time and determine the maintenance strategy before arriving at the fault site. The use of intelligent operation and maintenance makes the operation and maintenance work prioritized, the on-site work is more orderly, and the fault location and troubleshooting efficiency are significantly improved.

6. Conclusion

1) In view of the frequent failures of the converter system, electronic control system, generator system, pitch system, yaw system, hydraulic system and other aspects of the 1.5MW wind turbine in Xinjiang, wind farms should pay attention to these problems in the daily equipment management of wind farms. Scientific management and allocation of common equipment and spare parts ensure that maintenance time is not wasted due to shortage of equipment and spare parts.

2) As the wind power generation industry has entered a stage of rapid growth, the importance of maintenance and operation and maintenance after wind farm failures has also increased significantly. Fault handling methods, and continue to explore and summarize, sum up experience and methods, establish a set of detailed operation and maintenance procedures suitable for their own wind farms, and formulate scientific and reasonable wind farm management regulations.

3) The advantages of intelligent operation and maintenance are fully demonstrated, and the ability of professional and technical personnel in the daily operation and

maintenance of wind farms has been continuously improved. It is not only necessary to deal with daily faults and safe operation and maintenance of wind farms, but also to combine the characteristics of enterprises to make intelligent High-quality and efficient operation and maintenance, continuous innovation in the process of intelligentization, and improvement of the operation and maintenance management level of wind farms.

Acknowledgements

National Natural Science Foundation of China (51675354), Research on Fatigue Failure Assessment and Operation and Maintenance Management of Onshore Direct Drive Wind Turbine Main Spindle (ZZZC201837B), Design and Research on Wind Turbine Converter Cooling System (CXPY202012)

References

[1] LI Xiu, WU Zhengping, LAI Yang, HAN Ximeng, SHEN Gaoyan, LAI Weijun.Research on the development status and relationship of wind power industry and policy in Xinjiang: Based on the perspective of innovation system function [J].Science and Technology Management Research,2020,40(19):214-222.
[2] Wang Junna, Zhang Ningning.Analysis on the development status and prospect of wind power in Xinjiang [J]. Neijiang Science and Technology,2020,41(08):124+127.
[3] Lin hong, zhu wei. Statistical analysis and application of xinjiang wind power output characteristics [J]. Electrical measurement & instrumentation,2017,54(20):116-121.
[4] Hu Yongqiang. Application Research on Localization Transformation of Control System of GE1.5MW Double-fed Generator [D].North China Electric Power University,2016.
[5] Chen Songli, Li Ming, Wan Daqian.Design of 1.5MW Wind Turbine Blade Model Based on Similarity Theory [J]. Science and Technology Innovation,2020(36):47-48.
[6] Zhang Yusen, Zhao Chunyi, Jiang Jing, Tian Tao.Optimization Design and Simulation of Dual-fed Wind Turbine Converter [J]. Large Electric Machine Technology,2021(02):75-81.
[7] Zhao Zihang, WANG Haiyun, TANG Xinan, Wang Jia. Case study on current harmonic level of wind turbine converter [J]. Electrical measurement & instrumentation,2018,55(24):114-120.
[8] Zhang Hongyi, Shi Min, Wang Bing.Failure Analysis and Treatment of Power Unit of Wind Power Converter [J]. Inner Mongolia Electric Power Technology,2015,33(02):53-56.
[9] Zhang Zhicheng, Zhang Zhidan.Brief introduction to the failure reason and improvement of variable current power module of wind turbine [J]. Urban Construction Theory Research (Electronic Edition), 2011, 000(035):1-3.
[10] ZHANG Qian. Maintenance Analysis of Electrical Control System of Wind Turbine [J]. Times Agricultural Machinery,2019,46(12):60-62.
[11] ZHANG Xuejian. Analysis on Maintenance of Electrical Control System of Wind Turbine [J]. Electronic Testing,2019(09):110-111.
[12] Jin Min, Du Rina. Analysis and Solution of Common Faults in Electrical Control Part of Wind Turbine Operation [J].Journal of Inner Mongolia Agricultural University (Natural Science Edition),2012,33(Z1):309-312.
[13] JIANG Zhiwei. Daily Maintenance and Fault Treatment of Wind Turbine [J]. Science and Technology Innovation and Application,2019(28):127-128.
[14] Huang Zhangjian, LI Hui, LIU Xingzhong, WANG Kun, Xie Xiangjie, Wang Jie. Model simplification and operation characteristics comparison of wind turbine paddle system [J]. Electrical measurement & instrumentation,2018,55(01):65-71+77.
[15] Chen Qian, Li Luping, Liu Rui, Yang Bo, Li Zhonggui, Deng Zihao.Research progress on fault diagnosis method and technology of variable rotor system of high power wind turbine [J]. Power System Engineering,2020,36(01):1-7.
[16] Gan Huaizhang, Zhou Xinsheng.Fault Analysis of Wind Turbine's Variable Oar System [J]. Hunan Electric Power,2012,32(06):35-37.

[17] YAN Baibing. Fault Analysis and Measures of Wind Farm Fan Oar Change System [J]. Hubei Agricultural Mechanization,2020(12):43-44.
[18] Li Gang, Qi Ying, Li Yinqiang, Zhang Jianfu, Zhang Lihui.Research progress on fault diagnosis and state prediction of wind turbine [J]. Automation of Electric Power Systems, 201,45(04):180-191.
[19] XU Yingjian. Fault Analysis and Diagnosis of Wind Turbine Generator [D]. North China Electric Power University,2013.
[20] Wu Chun, Li Zhi, Yang Dong, Deng Niuwa.Research on Abnormal Yaw Performance of Wind Turbine and Its Maintenance Method [J]. Electric Age,2021(02):37-40.
[21] Ning Wengang, Jiang Hongwei, Wang Yuefeng.Common Fault Analysis of Yaw System of Wind Turbine [J]. Machine Management Development,2018,33(11):67-68+116.
[22] Zhang Junyan. Fault Analysis and Solutions of Yaw System of 1.5MW Wind Turbine [J].Forum of Association for Science and Technology (second half month),2010(06):33.
[23] LIU Jiahui. Fault Analysis of Yaw System of Wind Turbine [J]. Science and Technology Innovation Heralds,2019,16(12):106-107.
[24] Cong Zhihui. Principle and Maintenance of Wind Turbine Pitch System [J]. Inner Mongolia Science, Technology and Economy,2016(18):77-79.
[25] Ju Bin, Wang Binghui. Mechanical Braking Analysis of Megawatt Wind Turbine Based on AMESim [J].Wind power, 2017 (01) : 78-80.
[26] Ma Binrui. Fault Analysis and Treatment of Hydraulic Station of Wind Turbine [J]. Electrical Manufacturing,2012(03):36-38.
[27] Li Yufeng. Fault Analysis and Treatment of Hydraulic System of a Direct Drive Wind Turbine [J].Energy and Environment,2015(04):90-92.
[28] YU Yun. Application and Operation and Maintenance Experience Analysis of Wind Energy Monitoring Equipment [J]. Northeast Electric Power Technology,2020,41(03):44-47.
[29] YE Wei. Exploration on Reliability Management of High Altitude Fan [J]. Internal Combustion Engine & Accessories,2018(16):186-188.
[30] Zhuang Su. Analysis of Problems Existing in Operation and Maintenance of Wind Power Generation Equipment and Improvement Measures [J]. China Equipment Engineering,2021(03):38-40.
[31] Zhao Hongshan, Zhang Lupeng.Reliability Based Preventive Chance Maintenance Strategy for Wind Turbine [J]. Proceedings of the Csee,2014,34(22):3777-3783.
[32] Liu Hongyang, Deng Chun, Song Peng, Yang Weixin, Zhao Hongshan, Dong Wenqi, Liu Ximei.Preventive Opportunity Maintenance Strategy of Wind Turbine Considering Wind Energy Loss [J]. North China Electric Power Technology,2017(03):65-70.
[33] LI Zheng. Predictive Maintenance of Equipment Based on On-line Condition Monitoring System [J]. User of Instrument and Meter,2020,27(09):41-44.
[34] WANG Bin, DONG Xing-hui, LIU Hao, YANG Zhi-ling.Application of Fault Tree Based Expert System in Wind Power Gearbox [J]. Equipment Manufacturing Technology,2011(10):104-107.
[35] Jin Xin, Wang Dongya, Yin Wei, Cao Lianshan.Design and Application of On-line Oil Product Monitoring System for Wind Turbine [J]. Instrument Technology and Sensor,2011(09):62-63+82.
[36] Ding Xian, Xu Jin, Teng Wei, Liu Yibing.Research status and development trend of wind turbine state detection technology [J]. Renewable Energy,2017,35(10):1551-1557.
[37] Qin Zichuan, Su Hongsheng. Reliability Evaluation of Key Components of Wind Turbine Based on Improved Weibull Distribution [J]. Electrical Measurement & Instrumentation, 201,58(03):68-73.

doi:10.3233/ATDE220234

Research on the Method for the Reconstruction of Complex Surface Based on Deformable Template

Rong YU [a,1] and Chang YANG [a]
[a] *Inner Mongolia University School of Transportation, Huhhot, 010030, China*

Abstract. Due to the long hours running under the harsh environment such as high temperature, high speed, high pressure, complex components such as gas turbine blades are often deformed and the shape of complex components have been changed from their nominal CAD models. To solve this problem, a reconstruction method of complex surface based on ICP registration and surface deformation is proposed for the measured points and given design template surface, and the registration and reconstruction model of complex surface is established under the condition of minimizing the deflection of design template surface. The alternative iteration optimization strategy of the registration and template surface deformation establishes the analytic relationship between the design template surface and measured points. Meanwhile, the reconstruction accuracy and efficiency including the minimization of design template surface deflection. The proposed method effectively avoids complex process in conventional surface reconstruction such as regularization and parameterization of sample points, calculation of control points. Finally, the feasibility and effectiveness of the proposed method is validated by the given typical example.

Keywords. Reconstruction, Registration, Deformable template, B-spline surface

1. Introduction

Non-rigid registration based on deformation has always been a research hotspot in medical images, computer vision and astronomical observation [1-3]. In manufacturing, surface rigidity registration technology is very mature, the main application in machining orientation and quality assessment [4-8], but for the long running under high temperature, high temperature, high speed and bad environment prone to deformation of complex curved surface parts such as aircraft engine blades, rigid registration application technology cannot make workpiece original CAD design model and the actual measurement point fully fit, The design model cannot continue to be used in design, machining, etc. Therefore, surface reconstruction of this kind of products has economic benefits and practical significance. In this case, non-rigid registration is

[1] Corresponding author: Yu Rong (corresponding author), male, born in 1984, ph.D. candidate, whose main research directions are CNC machining Surface reconstruction Reverse engineering. E-mail: yurong.bao@163.com.The paper is supported by the Inner Mongolia Natural Science Foundation (2020MS05058)

proposed, which is a new research focus in the field of manufacturing. It is just starting, and there is still a lot of room for improvement. Yong[9] completed the reconstruction of blade section profile through affine transformation combined with free deformation to form non-rigid registration of template curve of blade section profile and measured data of sectional profile. A. Sahar[10] proposed an improved ICP registration algorithm based on the difference between the workpiece and CAD design model after cold forming, such as deep stamping and spinning, and realized the non-rigid matching between triangular mesh surface and discrete surface, so as to achieve the purpose of analyzing and compensating the workpiece shape error. The application of non-rigid registration technique in surface reconstruction has not been seen.

Surface reconstruction is the core content of CAD/CAM. The traditional surface reconstruction method can be divided into interpolation of measuring points and approximation of measuring points. The surface reconstruction by interpolation and approximation of scattered measurement points requires segmentation, sequencing, parameterization, determination of node vectors, inverse control points and other key steps. Due to the multiple factors, sometimes manual operation, increased uncertainty and variability, prompting the development of new refactoring techniques. In this paper, a new method of complex surface reconstruction based on ICP registration algorithm and surface deformation is proposed. For complex curved surface measurement point and the design template surface, the first rough registration actual measurement point and the surface template, and then USES the alternating iterative registration and deformation of non-rigid registration method, template surface and the measurement point within the maximum limit, at the same time of reconstruction error control within the scope of the permit minimize template surface deformation. The proposed method can realize surface reconstruction of complex parts quickly and effectively.

2. Mathematical model of surface reconstruction

Through the non-rigid registration of the actual measuring points set of the template surface $\{\mathbf{p}_{i'}\}\ 1 \le i' \le \mathrm{h}$ and complex surface $\mathbf{S}(u, v)$, the surface reconstruction is completed under the condition of the minimum deformation of the template surface. The reconstruction process mainly includes the determination of template surface and measuring point, rough registration of template surface and measuring point, alternate iterative ICP precision registration and non-rigid registration of surface deformation.

The template surface is generally NURBS surface:

$$\mathbf{S}(u,v) = \sum_{i=0}^{m}\sum_{j=0}^{n} N_{i,t}(u)N_{j,t}(v)\mathbf{d}_{\mathrm{i,j}} \tag{1}$$

In the formula, $\mathbf{d}_{\mathrm{i,j}}$ is the surface control point. $N_{i,t}(u)$ is is the basis function in the u direction. $N_{j,t}(v)$ is the basis function in the v direction. t is the number of spline bases [11].

The expression for the deformed template surface $\overline{\mathbf{S}}(u,v)$ is:

$$\overline{\mathbf{S}}(u,v)=\sum_{i=0}^{m}\sum_{j=0}^{n}N_{i,t}(u)N_{j,t}(v)(\mathbf{d}_{i,j}+\boldsymbol{\delta}_{i,j}) \tag{2}$$

$\boldsymbol{\delta}_{i,j}$ is the control point disturbance amount.

The key step of non-rigid registration is the determination of the measurement point at the nearest point of the surface and the calculation of the transformation matrix.

The mathematical model of the non-rigid registration surface reconstruction of the actual measurement point $\{\mathbf{p}_{i'}\}$ with the deformed template surface $\overline{\mathbf{S}}(u,v)$ is:

$$e(\mathbf{f}^k)=\sum_{i'=0}^{h}\left\|\mathbf{p}_{i'}-\mathbf{f}^k(\overline{\mathbf{S}}(u_{i'},v_{i'}))\right\|^2 \tag{3}$$

$\overline{\mathbf{S}}(u_{i'}, v_{i'})$ is the corresponding point $\{\mathbf{p}_{i'}\}$ on the deformed template surface $\overline{\mathbf{S}}(u,v)$. $\mathbf{f}^k:\{\mathbf{T}^k,\boldsymbol{\delta}^k\}$ is the transformation matrix. k is the number of iterations. $\boldsymbol{\delta}^k$ is the template surface deformation transformation. $\mathbf{T}^k$ is the rigid transformation including the rotation transformation and the translation transformation. Equation (2) substitution formula (3). The mathematical model of deformation surface reconstruction after expansion is:

$$e(\mathbf{f}^k)=\sum_{i'}^{h}\left\|\mathbf{p}_{i'}-\sum_{i=0}^{m}\sum_{j=0}^{n}N_{i,t}(u_{i'})N_{j,t}(v_{i'})(\mathbf{d}_{\mathrm{i,j}}\mathbf{R}^k+t^k+\boldsymbol{\delta}^k{}_{i,j})\right\|^2 \tag{4}$$

3. Non-rigid registration algorithm for template surfaces and measurement points

Through the non-rigid registration of the template surface and the measurement point, the specific steps for completing the complex surface reconstruction are as follows:

(1) Coarse registration of template surface $\mathbf{S}(u, v)$ and measurement point $\{\mathbf{p}_i\}$.

(2) Precise registration is achieved through ICP algorithms. First, the rapid iterative method calculates the corresponding point of measurement point $\{\mathbf{p}_{i'}\}$ on the template surface $\mathbf{S}(u,v)$. The transformation matrix $\mathbf{T}^k:\{\mathbf{R}^k,\mathbf{t}^k\}$ is then computed using the singular value decomposition method.

(3) The amount of control point movement $\boldsymbol{\delta}^k$ of the deformation template surface $\overline{\mathbf{S}}(u,v)$ is calculated from the free deformation of the surface control point.

3.1 Coarse registration of the measurement point with the template surface

The initial transformation $\mathbf{T}^0 : \{\mathbf{R}^0, \mathbf{t}^0\}$ is accurately determined by the coarse matching of the measurement point set $\{\mathbf{p}_{i'}\}$ and the template surface $\mathbf{S}(u,v)$.

Take three points $\mathbf{p}_{i'}$ ($i' = 0,1,2$) on the measurement point set $\{\mathbf{p}_{i'}\}$ that are not collinear. The corresponding three points on the template surface are $\mathbf{q}_{i'}$ ($i' = 0,1,2$). Construct a local coordinate system $\tau_p = (\mathbf{x}(p), \mathbf{y}(p), \mathbf{z}(p))$ with a as the origin $\mathbf{p}_0$. This is shown in Figure 1.

Similarly, a local coordinate system $\tau_{\mathbf{q}} = (\mathbf{x}(q), \mathbf{y}(q), \mathbf{z}(q))$ can be constructed as the origin $\mathbf{q}_0$.

$$\begin{cases} \mathbf{x}(p) = \dfrac{\mathbf{p}_1 - \mathbf{p}_0}{|\mathbf{p}_1 - \mathbf{p}_0|} \\ \mathbf{y}(p) = \mathbf{x}(p) \times \dfrac{\mathbf{p}_2 - \mathbf{p}_0}{|\mathbf{p}_2 - \mathbf{p}_0|} \\ \mathbf{z}(p) = \mathbf{x}(p) \times \mathbf{y}(p) \end{cases} \tag{5}$$

The method given by Chua[12] was used to determine the initial transformation array $\mathbf{T}^0 : \{\mathbf{R}^0, \mathbf{t}^0\}$ for rotation and translation:

$$\begin{cases} \mathbf{R}^0 = [\mathbf{x}(q), \mathbf{y}(q), \mathbf{z}(q)]^T \bullet [\mathbf{x}(p), \mathbf{y}(p), \mathbf{z}(p)] \\ \mathbf{t}^0 = \mu_{\mathbf{p}} - \mu_{\mathbf{q}} \mathbf{R}^0 \end{cases} \tag{6}$$

In the formula $\mu_{\mathbf{p}} = (\mathbf{p}_0 + \mathbf{p}_1 + \mathbf{p}_2)/3$ and $\mu_{\mathbf{q}} = (\mathbf{q}_0 + \mathbf{q}_1 + \mathbf{q}_2)/3$.

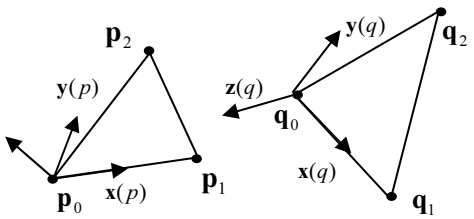


Figure 1. Coarse registration of measuring point sets and template surface.

3.2 Accurate registration of measurement points and template surfaces

3.2.1 Calculates the corresponding point of the measurement point on the template surface

The problem of pointing to the closest point of a B-spline surface can be translated into calculating the projection point of the point in the direction of the surface normal vector. This article uses the fast iterative method [13] to calculate the nearest point of a point on a surface.

The measurement point set $\{\mathbf{p}_{i'}\}$ corresponds to the template surface $\{\mathrm{q}^{\mathrm{k}}{}_{i'}\}(1 \le i' \le \mathrm{h})$. Assuming that the initial point of the measurement point $\mathbf{p}_0$ is $\mathbf{q}'_0$. The projection point in the direction of the normal vector $\overrightarrow{\mathbf{p}_0\mathbf{q}_0}$ is $\mathbf{q}_0$. So:

$$\overrightarrow{\mathbf{p}_0\mathbf{q}'_0} = \mathbf{S}_u \Delta u + \mathbf{S}_v \Delta v + \overrightarrow{\mathbf{p}_0\mathbf{q}_0} \tag{7}$$

$\mathbf{S}_u$, $\mathbf{S}_v$ are the partial derivatives of the template surface: $\mathbf{S}_u = \dfrac{\partial S(u,v)}{\partial u}$, $\mathbf{S}_v = \dfrac{\partial S(u,v)}{\partial v}$. $\mathbf{q}'_0$ moves Δu, Δv in the u and V directions respectively and the two sides of equation (7) are multiplied by $\mathbf{S}_u$ and $\mathbf{S}_v$ respectively:

$$\begin{cases} \mathbf{S}_u \bullet \overrightarrow{\mathbf{p}_0\mathbf{q}'_0} = \mathbf{S}_u \bullet \mathbf{S}_u \Delta u + \mathbf{S}_u \bullet \mathbf{S}_v \Delta v + \mathbf{S}_u \bullet \overrightarrow{\mathbf{p}_0\mathbf{q}_0} \\ \mathbf{S}_v \bullet \overrightarrow{\mathbf{p}_0\mathbf{q}'_0} = \mathbf{S}_v \bullet \mathbf{S}_u \Delta u + \mathbf{S}_v \bullet \mathbf{S}_v \Delta v + \mathbf{S}_v \bullet \overrightarrow{\mathbf{p}_0\mathbf{q}_0} \end{cases} \tag{8}$$

Solve the above system of equations to solve Δu and ΔV ,make $u + \Delta u \to u$, $v + \Delta v \to v$. Repeat the above process after moving several times until satisfied $|\Delta u| \le \xi$, $|\Delta v| \le \xi$, $\mathbf{q}_0$ is the nearest point corresponding to $\mathbf{p}_0$. The fast iterative method is fast and efficient, and can generally converge.

3.2.2 Singular value decomposition calculates the transformation matrix

After determining the corresponding point $\{\mathrm{q}^{\mathrm{k}}{}_{i'}\}$ between the measurement point set $\{\mathbf{p}_{i'}\}$ and the template surface during the fine registration process, the rigid transformation $\mathbf{T}^k : \{\mathbf{R}^k, \mathbf{t}^k\}$ is calculated.

Construct the least squares function as follows:

$$e(\mathbf{T}^k) = \sum_{i'=0}^{h} \left\| \mathbf{p}_{i'} - \mathbf{T}^k(\mathbf{q}^k{}_{i'}) \right\|^2 \tag{9}$$

Formula (9) can also be written as:

$$e(\mathbf{R}^k, \mathbf{t}^k) = \sum_{i'=0}^{h} \left\| \mathbf{p}_{i'} - (\mathbf{q}_{i'}\mathbf{R}^k + \mathbf{t}^k) \right\|^2 \tag{10}$$

The transformation matrix $\mathbf{T}^k : \{\mathbf{R}^k, \mathbf{t}^k\}$ is calculated in two steps. The rotation matrix $\mathbf{R}^k$ is calculated using the singular value decomposition method (SVD)[14] and then the translation transformation matrix $\mathbf{t}^k$ is calculated. $\mathbf{H}^k$ is the covariance matrix between $\{\mathbf{p}_{i'}\}$ and $\{\mathrm{q}^k{}_{i'}\}$:

$$\mathbf{H}^k = \sum_{\mathrm{i}'=0}^{\mathrm{h}} (q^k{}_{i'} - \mathbf{c}^k{}_q)(\mathbf{p}_{i'} - \mathbf{c}_p)^T \tag{11}$$

In the formula $\mathbf{c}^k{}_q = \frac{1}{h}\sum_{i'=0}^{h} \mathbf{q}^k{}_{i'}$, $\mathbf{c}_p = \frac{1}{h}\sum_{i'=0}^{h} \mathbf{p}_{i'}$. Singular value decomposition of matrix $\mathbf{H}^k$ are $\mathbf{H}^k = \mathbf{U} \wedge \mathbf{V}^T$, $\mathbf{X} = \mathbf{U}\mathbf{V}^T$. $\det(\mathbf{X})$ is calculated. If $\det(\mathbf{X}) = +1$, $\mathbf{R}_k = \mathbf{X}$, and if $\det(\mathbf{X}) = -1$, the calculation fails, but this rarely happens. After the rotation matrix is obtained, the translation matrix is calculated:

$$\mathbf{t}^k = \mathbf{c}_\mathbf{p} - \mathbf{c}^k{}_q \mathbf{R}^k \tag{12}$$

3.3 Template surface deformation

For complex surfaces, deformation is generally achieved through multipoint constraints. B-spline multipoint constraints calculate the amount of surface control point movement by least squares and reference control points to achieve the required deformation[15].The specific calculation process is as follows.

The matrix form of equation (1) is:

$$\mathbf{S}(u,v) = \mathbf{N}(u,v)\mathbf{d} \tag{13}$$

The basis function in the form of a matrix in the above equation is:

$$\mathbf{N}(u,v)=[\mathrm{N}_{0,\mathrm{t}}(u)\mathrm{N}_{0,\mathrm{t}}(v),\mathrm{N}_{0,\mathrm{t}}(u)\mathrm{N}_{1,\mathrm{t}}(v),\cdots,\mathrm{N}_{0,\mathrm{t}}(u)\mathrm{N}_{\mathrm{n},\mathrm{t}}(v),\ \mathrm{N}_{1,\mathrm{t}}(u)\mathrm{N}_{0,\mathrm{t}}(v),\mathrm{N}_{1,\mathrm{t}}(u)\mathrm{N}_{1,\mathrm{t}}(v),\cdots,\mathrm{N}_{1,\mathrm{t}}(u)\mathrm{N}_{\mathrm{n},\mathrm{t}}(v),\cdots,\ \mathrm{N}_{\mathrm{m},\mathrm{t}}(u)\mathrm{N}_{0,\mathrm{t}}(v),\mathrm{N}_{\mathrm{m},\mathrm{t}}(u)\mathrm{N}_{1,\mathrm{t}}(v),\cdots,\mathrm{N}_{\mathrm{m},\mathrm{t}}(u)\mathrm{N}_{\mathrm{n},\mathrm{t}}(v)]$$

,The control point matrix form is:

$\mathbf{d}=\left[d_{0,0},d_{0,1},\cdots,d_{0,n},d_{1,0},d_{1,1},\cdots,d_{1,n},\cdots,d_{m,0},d_{m,1},\cdots,d_{m,n}\right]$,The matrix form of the control point perturbation is:

$\boldsymbol{\delta}=\left[\delta_{0,0},\delta_{0,1},\cdots,\delta_{0,n},\delta_{1,0},\delta_{1,1},\cdots,\delta_{1,n},\cdots,\delta_{m,0},\delta_{m,1},\cdots,\delta_{m,n}\right]$ .The matrix form of the deformed template surface formula (2) is:

$$\overline{\mathbf{S}}(u,v)=\mathbf{N}(u,v)(\mathbf{d}+\boldsymbol{\delta}) \tag{14}$$

For b-spline surfaces that implement deformations with multipoint constraints, assume that there are l initial points on the surface. The corresponding parameters are (u_r,v_r) ($r=0,1,\cdots,l$). $\Delta\mathbf{S}(u_r,v_r)$ is the displacement from the initial point of the surface to the target point, then the template surface multipoint constraint equation system can be written as:

$$\begin{cases}\Delta\mathbf{S}(u_r,v_r)=N(u_r,v_r)\boldsymbol{\delta}_r\\ 0\le r\le l\end{cases} \tag{15}$$

Making $\Delta\mathbf{S}=[\Delta\mathbf{S}(u_0,v_0),\Delta\mathbf{S}(u_1,v_1),\cdots,\Delta\mathbf{S}(u_r,v_r)]$,equation (15) is converted to matrix forma as follow:

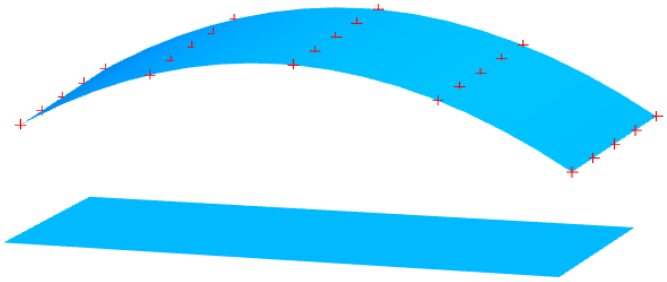
$$\Delta\mathbf{S}=\mathbf{N}(u,v)\boldsymbol{\delta} \tag{16}$$

Given $\Delta\mathbf{S}$ and $\mathbf{N}(u,v)$, the control point of motion $\boldsymbol{\delta}$ is derived , and the least square solution to equation (6) can be solved by computing the generalized inverse matrix $\mathbf{N}^{+}(u,v)$ of the basis function $\mathbf{N}(u,v)$:

$$\boldsymbol{\delta}=\mathbf{N}^{+}(u,v)\Delta\mathbf{S} \tag{17}$$

Figure 2 shows the deformation of the surface during multipoint constraints.

(a) The initial position of the point set and surface template surface



(b) Template surface deformation

Figure 2. Multipoint constrained surface deformation

4. Experimental validation

Blades are the main components of the aircraft engine. After long-term operation in high temperature, high speed, high pressure and other harsh working environment, it is easy to deform or damage, so its actual measurement point set and CAD design model is different. Using the complex surface reconstruction method based on ICP registration algorithm and deformation proposed in this paper, the surface of the aero engine blade is reconstructed for the given measurement point set and template surface to verify the effectiveness of the method. The proposed algorithm is implemented in VC++. The set of measurement points is 16×64 truncated data, in order to display clearly, select 16×21 in the figure. The size of the template surface is 45.3722mm in length, 22.2998mm in width, and the control point is a array. The maximum distance from the initial position measurement point set to the template surface is 67.1312mm. After the coarse registration, the maximum distance is 4.011mm, the non-rigid registration of the measurement point set and the template surface are formed by alternate iteration ICP

registration and template surface deformation. The minimum deformation of the template surface is ensured during the calculation process. The template surface and the measurement point set are fully matched to the greatest extent. The maximum distance between the template surface and the measurement after coarse matching, and the position at different iterations. Point set after 123 iterations is 0.035mm. Table 1 shows the algorithm errors for different iterations. Figure 3 shows the initial position of the measured point set and template surface, the position.

Table 1. Errors at different iterations

Iterations times	1	15	35	55	85	123
error (mm)	0.9704	0.6528	0.0973	0.0956	0.069	0.00457

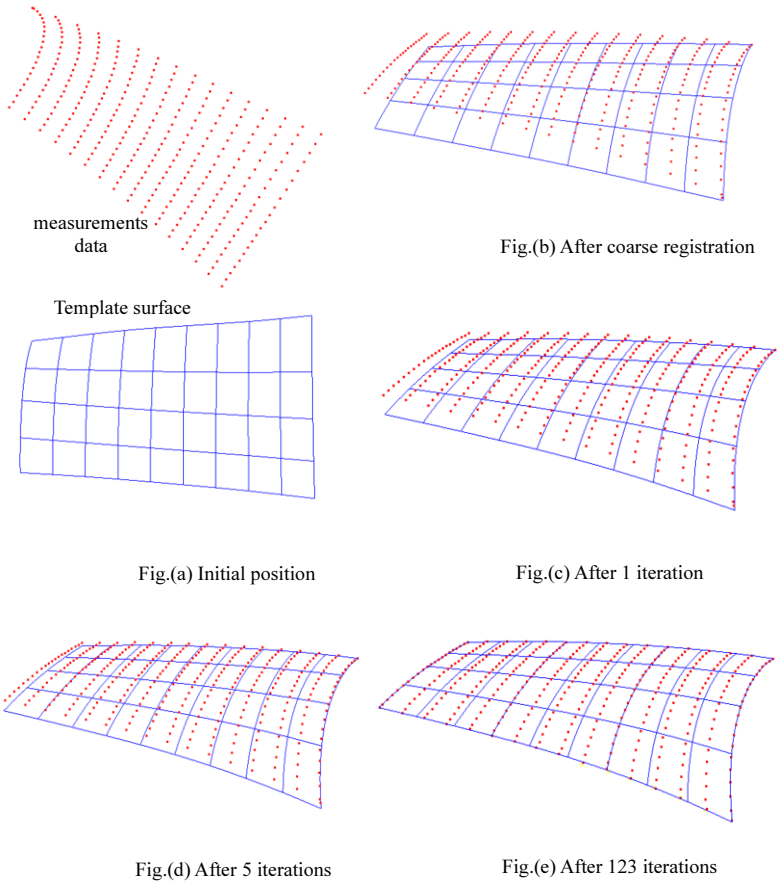


Fig.(a) Initial position

Fig.(b) After coarse registration

Fig.(c) After 1 iteration

Fig.(d) After 5 iterations

Fig.(e) After 123 iterations

Figure 3. Non-rigid registration process for measuring point sets and template surface

5. Conclusion

(1) The complex surface reconstruction based on ICP registration and template surface deformation proposed in this paper effectively avoids the steps of sequence, segmentation and parameterization of the actual measurement point set in the conventional B-spline surface reconstruction. In the process of surface reconstruction, the degree of automation is high, and no manual intervention is required, which effectively avoids the lack of manual operation of traditional B-spline surface reconstruction.

(2) The strategy of alternate iterative ICP registration and template surface deformation constitutes non-rigid registration, which ensures that the deformation of template surface is minimized and the actual measurement point and template surface are coincident to the greatest extent. The singular value decomposition method of determining the corresponding point and finding the transformation array is used to ensure the accuracy and reliability of the registration process.

(3) For the long-term operation of the workpiece that is deformed in the harsh working environment or deformed during the manufacturing process, the difference between the actual measurement point and the design model causes the design model to be used for design, processing and other steps, and the complex surface reconstruction based on ICP registration and template surface deformation proposed in this paper solves the problem of surface reconstruction of such products, and the example verification shows the effectiveness and practicality of the proposed method.

References

[1] Rueckert D, Sonoda L, Hayes C, et al. Nonrigid Registration Using Free-form Deformations: Application to Breast MR Image[J]. IEEE Trans on Medical Imaging,1999,18(8): pp. 712-721.
[2] Huang Q, Adams B, Wicke M, Guibas L.J. Non-rigid registration under isometric deformations[J]. Computer Graphics Forum, 2008, 27 (5): pp. 1449–1457.
[3] Xu L, Chen Y, Qiu J, etal. Nephogram Nonrigid Registration Method Based on B-spline[J]. Acta Geodaetica et Cartographica Sinica, 2011, 40(3): pp. 326-331.
[4] Jingting X.Shape-adaptive CNC milling for complex contours on deformed thin-walled revolution surface parts[J].Journal of Manufacturing Processes,2020,59:pp.760-771.
[5] Jingting X, Jinbo Niu, Mansen C etal. Research Progress of Multi-Axis NC Machining Technology for Precision Complex Surface Parts [J]. Acta AERONAUtica ET Astronautica Sinica, 2021, 42(10): pp.31-54.
[6] Jingting X etal.A Method of Generating Spiral Tool Path for Direct Three-Axis Computer Numerical Control Machining of Measured Cloud of Point[J].Journal of Computing and Information Science in Engineering,2019,19(4):041015.
[7] Xu J, Sun Y, Liu W. Optimal localization of free-form shaped parts in precision inspection [J]. Chinese Journal of mechanical engineering, 2007, 43 (6): pp. 175-179.
[8] Menq C, Yau H, Lai G. Automated precision measurement of surface profile in CAD-directed inspection. IEEE Transactions on Robotics and Automation, 1992, 8(2): pp. 268- 278.
[9] Yongqing L, Jun N. Constraints based nonrigid registration for 2D blade profile reconstruction in reverse engineering[J]. Journal of computing and information science in engineering. 2009, 9(4): pp. 57-62.
[10] Sacharow A, Balzer J, Biermann D, Surmann T. Non-rigid isometric ICP: A practical registration method for the analysis and compensation of form errors in production engineering[J]. Computer-Aided Design, 2011, 12(43): pp. 1758–1768.
[11] Pieg L, Tiller W, The NURBS Book. Springer-Verlag. 1997
[12] Chua CS, Jarvis R. 3D free-form surface registration and object recognition[J]. International Journal of Computer Vision, (1996), 17, (1): pp. 77-99.

[13] Peng QS. "An Algorithm for Finding the Intersection Lines Between Two B-Spline Surfaces[J].Computer Aided Design 16, 4 (1984): pp. 191-196.
[14] Zhu XX. Free-form curves/surface modeling technology [M]. Beijing: Science Press, 2000(in Chinese)
[15] Hus WM, Hughes JF, Kaufman H. Direct manipulation of free form deformation. Computer Graphics, 1992, 26(2): pp. 177-184.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 257
J. Xu et al. (Eds.)

doi:10.3233/ATDE220235

Research Progress on Optimization Methods of Powder Injection Molding Process Parameters

Tong ZHAO[a], Baozhen EI[b,1], Xunwei WANG[b], Guanghao QI[a] and Harald LOWE[c]
[a]*Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing, China*
[b]*Beijing Intelligent Machinery Innovation Design Service Engineering Technology Research Center, Beijing Union University, Beijing, China*
[c]*Institute for Partial Differential Equations, Technical University of Braunschweig, Braunschweig, Germany*

Abstract. The powder injection molding (PIM) process provides a possibility for high-volume, low-cost manufacturing of workpiece. During powder injection molding, besides mold design and material characteristics, injection parameters are critical in affecting the capability of molded commodities. Traditional injection molding parameter design relies on the experience of technicians, and it is difficult to ensure the consistency in the quality of injection parts. In this work, we analyze the optimization of technical parameters of current PIM technology and provide some perspectives on the future development of this field.

Keywords. Powder injection molding, parameter optimization, design of experiment

1. Introduction

Powder injection molding (PIM) is a technique that involves metal and non-metallic powders for manufacturing a large number of small and medium sized parts with complex shapes [1]. Due to its net or near-net shape process, high efficiency and low cost, it has had a wide of applications in the fields of aerospace, automotive medical, military, electronic, industrial, personalized products and others [2]. Being a branch of PIM, metal injection molding (MIM) is ranked in the top two advanced manufacturing technologies for the future by McKinsey & Company [3]. At the same time, parts with high density, good mechanical properties, and small surface roughness can be obtained, and parts with complex structures can be produced in large quantities and with high efficiency.

During the production process, warpage and shrinkage are the main defects in the products [4]. These problems are related to several factors, and it is difficult to avoid these problems by relying solely on the experience of the engineers. Especially for complex precision parts where there are dimensional matching requirements, the shape

[1] Baozhen Lei, Beijing Union University, No.97 Beisihuan East Road, Chao Yang District, Beijing, China, E-mail:leibaozhen@126.com.

and dimensional accuracy of the parts are often very high. Therefore, the development of an appropriate method for the optimization of the forming process of the parts is highly desired by the industry. In the following, we review some current research methods for PIM technical parameter optimization and provide some perspective on future development trend of this field.

2. The Status Quo of Optimization of PIM Process Parameters

2.1. Injection Molding Technology

The parts produced by MIM technology have better mechanical properties than plastic injection molding parts. The process consists of 4 stages: feedstock preparation, injection molding, debinding and sintering [5], as shown in Figure 1. In the first stage we need to mix metal powder and thermoplastic binders [6]. Since the metal powders' flow ability and permeability are mainly influenced by the binder, a suitable, precise recipe is important for MIM [7]. Then the feedstock is processed by the MIM machine into the green part. In the following debinding stage, removing the binder from the green part yields the so-called brown part [8]. During the final sintering stage, high temperature heating to densify. The parts has good mechanical properties [9]. Each stage must be strictly controlled, because it will directly affect the quality of MIM products. Therefore, the optimization in each stage will directly affect the quality of the final formed product [10].

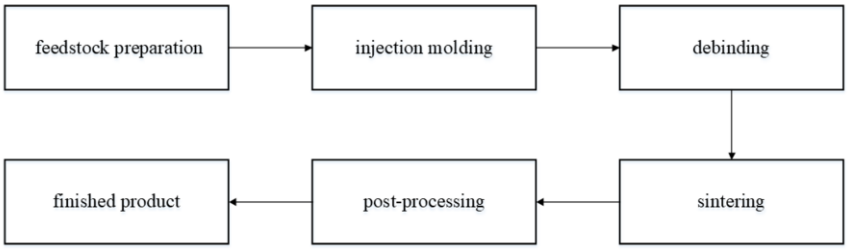


Figure 1. MIM process

The choice of suitable injection molding process parameters is critical for a good molding quality. One approach to a reliable parameter design is to use computer aided engineering (CAE). For example, Mahmud *et al*. [11] studied the main technical factors which influence the shrinkage of plastic injection molding product. Li *et al*. [12] used the Kriging model and design of experiment (DOE) to analyze the dependency of the shrinkage on the technical parameters; the results were applied to an effectively reduced warpage of PIM products. Saunders [13] applied the functional Gaussian process surrogate models to link microstructure morphology to mechanical properties. Nayak *et al*. [14] utilized Taguchi analysis, PCA and GA hybrid methods to research, analyze and optimize MIM technical parameters. Song *et al*. used SVM and BPNN to establish a function for predicting the quality of PIM products. Kun *et al*. [15] applied DOE approach, RSM and NSRGA-II for optimization. Qiao *et al*. [16] proposed an ant colony algorithm model to optimize MIM technical parameters.

Islam [17] found that the shrinkage of MIM products on the micro scale is uneven and that the thinner area of the part shows a higher relative shrinkage than the thicker

area. Lin [18] used the Taguchi method and GRA minimize volume shrinkage. Maulana [19] adopted the orthogonal matrix method and found that the temperature had the greatest influence on the shrinkage rate.

There has been a major focus on the warpage of plastic products and the shrinkage of metal products. To produce high-precision parts, especially for gears, the characteristics of anisotropic shrinkage are worthy of deep studying.

2.2. Design of Experiment

Before establishing a model describing the relationship between technical parameters and shrinkage, experiments should be designed. Researchers have adopted a variety of experimental design methods, such as orthogonal experiments, central composite design (CCD) experiments, Box-Behnken design (BBD) experiments and sampling methods. Orthogonal experiment is a common experimental design method for studying multiple factors and multiple levels. It can be characterized as "evenly dispersed, neat and comparable". Sateesh *et al*. [20] used the orthogonal experiment method to design 9 sets of experiments with the L9 orthogonal table, and minimized warpage, volume shrinkage, and cycle time. Moayyedian *et al*. [21] applied a mixture orthogonal experiment design method and designed 18 sets of experiments with L18 orthogonal table. Bensingh *et al*. [22] used the orthogonal experiment method to design 44 sets of experiments on 7 process parameters. The count of tests will increase geometrically with the count of test factors, but for the accuracy of the test, certain factors cannot be ignored because of the increase in the number of tests.

Kastner *et al*. [23] utilized the CCD test method to design 25 sets of tests. Heidari used CCD approach to design 45 sets of tests. Kim employed the BBD design approach to design 15 sets of tests for 3 process parameters, and adjusted the optimized conditions by changing the variables. Guo *et al*. [24] applied the Latin Hypercube Sampling (LHS) approach to design 488 groups of experiments for network training, and performed finite element simulations and physical experiments.

Tsai *et al*. [25] made use of the Taguchi parameter design to analysis injection molding parameters and found the important factors affecting the shrink. Range analysis is used to research various technical parameter on the warpage deformation; the design of the test was done by an LHS method. Tian applied the DOE approach to experiment and process data, and then used statistical approach to condition the greatest association technical parameters. Amin *et al*. utilized the Taguchi method and took 7 process parameters into consideration in the design of experiments for MIM. Prathabrao analyzed MIM in detail and applied the L9 orthogonal table to design the experiment for MIM.

The CCD and BBD methods have higher experimental accuracy than the normal process experiment, but the corresponding number of experiments are also relatively large. Selection of experimental method depends on the specific production needs.

2.3. Mathematical Model

Traditional injection process optimization methods are mostly based on experimental design, analysis of variance and average response, and simple processing of data. This method can optimize the product quality of the injection process to a certain extent, but the quality prediction for the continuous unknown parameter combination product is

Unable to do it. According to mathematical statistical logic to find the best level of each control factor from the distribution of discrete points, only the existing local best combination can be obtained. Therefore, considering the non-linear relationship between control factors and optimization goals, many models of mathematical professional fields are used in the injection molding process.

Being a applicable multivariate regression model, RSM is widely used due to its simple application and mature technology. Rizvi *et al*. [26] used experimental design and RSM methods, as well as set the CCD and the technical parameters. Sudsawat *et al*. utilized orthogonal experiment, RSM, firefly algorithm (FA) and annealing process, and carried out experiments based on Moldex3D and injection molding machine and obtained the best solution of optimized warpage.

On the other side, neural networks have strong performance in processing nonlinear relationships between multi-inputs and multi-outputs, and can cope especially with the highly complex nonlinear relationships in manufacturing processes. Therefore, artificial neural networks (ANN) are extremely widely applied in process optimization parameter scenarios.

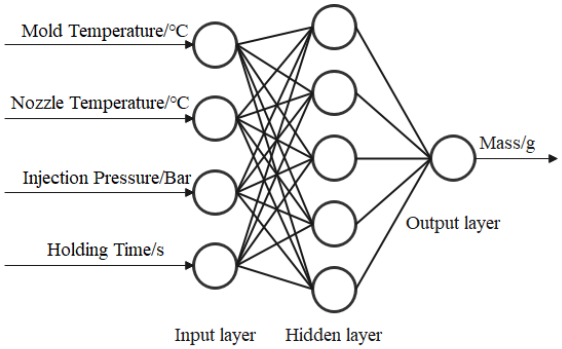


Figure 2. Structure of Artificial Neural Network.

Kitayama [27] used a RBF network and a few simulations to determine the optimal injection speed and pressure distribution. Shi *et al*. [28] deployed BPNN models and a parameter sampling evaluation (PSE) strategy to optimize product warpage.

Table 1. traditional method and the neural network method

traditional method	neural network method
RSM	BPNN
FA	RBF-NN
GA	BPNN
PCA	SVM

In summarize, neither the single traditional method nor the neural network method can solve complex engineering problems. For obtaining higher performance, suitable methods should be adopted.

3. Development Direction

Much research has been done to optimize MIM technical parameters. According to different needs, they have commonly optimized certain quality goals of a part, respectively conducted different experimental design schemes to obtain data. The

model is analyzed and calculated to finally achieve the goal of quality optimization. Due to their high complexity, these optimization methods are rarely used in actual industrial manufacturing. Therefore, it is necessary to commercialize and modularize the technologies in these studies, integrate these functions into the injection molding machine, and operate the experiment through a visual interface, as well as to save the data in the process of an injection experiment. Moreover, the model should have the ability of self-learning, i.e. it should be capable to learn from data and to optimize itself.

It is also important to classify and compare existing models and establish a special model library. For example, one could collect typical materials, 3D models of parts and molds. It can provide the technicians with an initial range of processing parameters, and on this basis, the subsequent optimization is carried out. The above two strategies benefit reducing the experiment costs.

4. Conclusions and Prospects

In summary, techniques such as experimental design methods, genetic algorithms, and neural network models have been widely used for injection molding process optimization, but most of them optimize a single objective. However, the performance of both plastic and metal products does not only depend on a few indicators such as mechanical properties or shrinkage. Therefore, it is important to multiple indicators of products at the same time and to achieve comprehensive improvement of product quality through multi-objective optimization method, to guide injection molding more effectively.

Acknowledgments

This research is financially supported by National Natural Science Foundation of China(No.5177050098), Collaborative innovation project of Chaoyang District, Beijing (CYXC2102、CYXC1905), Academic Research Projects of Beijing Union University (No. JZ10202003). The authors gratefully acknowledge above financial support.

References

[1] Prathabrao M, et al. Review on sintering process of WC-Co cemented carbide in metal injection molding technology. IOP Conference Series: Materials Science and Engineering. 2017 Dec;165.
[2] Moon AP, Dwarapudi S, Sista KS, Kumar D, Sinha GR. Opportunity and Challenges of Iron Powders for Metal Injection Molding. ISIJ International. 2021 Feb; 61(7): pp 2015-2033.
[3] Ben S, Colin S, Colin S. Factory of the Future Issue One. 2018 Mar; pp 5-18.
[4] Engström S. Metal Injection Molding: A review of the MIM process and its optimization. 2017 July.
[5] Hossain, et al. Experimental and theoretical investigation of powder‐binder mixing mechanism for metal injection molding. Materials and Manufacturing Processes. 2015 Nov; 30(1): pp 41-46.
[6] Heaney DF. Powders for metal injection molding (MIM). 2019 Sep; pp 45-56.
[7] Banerjee, S, et al. Debinding and sintering of metal injection molding (MIM) components. 2019 Dec; pp 129-171.
[8] Soyama J, et al. Sintering behavior and microstructure formation of titanium aluminide alloys processed by metal injection molding. 2017 Oct; 69(4): pp 676-682.

[9] Shahbudin, S, et al. A Review of Metal Injection Molding-Process, Optimization, Defects and Microwave Sintering on WC-Co Cemented Carbide. In IOP Conference Series: Materials Science and Engineering. 2017 Aug.
[10] Ali. Metal injection moulding of titanium and titanium alloys: Challenges and recent development. Powder Technology. 2017 Dec; 319: pp 289-301.
[11] Mahmud, Nurul N, et al. Rheological properties of irregular-shaped titanium-hydroxyapatite bimodal powder composite moulded by powder injection moulding. Journal of Materials Research and Technology. 2021 Nov;11: pp 2255-2264.
[12] Li S, et al. Optimization of Injection Molding Process of Transparent Complex Multi-Cavity Parts. Arabian Journal for Science and Engineering. 2021 Jun; pp 1-11.
[13] Saunders R, et al. Mechanical behavior predictions of additively manufactured microstructures using functional Gaussian process surrogates. npj Computational Materials. 2021 Mar;7(1): pp 1-11.
[14] Nayak C, et al. Analysis and optimization of metal injection moulding process. Materials Forming, Machining and Post Processing. 2020 July; pp 41-74.
[15] Li K, et al. Multi-objective optimization of the fiber-reinforced composite injection molding process using Taguchi method. 2019 Nov; 91:pp 69-82.
[16] Qiao B. Optimization of copper powder injection molding process parameters based on ant colony algorithm. Fenmo Yejin Jishu(Powder Metallurgy Technology. 2011 Feb;29(3): pp 200-205.
[17] Islam A., et al. The shrinkage behavior and surface topographical investigation for micro metal injection molding. AIP Conference Proceedings. 2015 May; p 1664.
[18] Lin CM., et al. Hybrid Taguchi‐Gray Relation Analysis Method for Design of Metal Powder Injection-Molded Artificial Knee Joints. Polymers. 2021 Jan;13(6): p 865.
[19] MAULANA, A. R. Pengaruh Temperature, Holding Time, dan Penambahan SN Terhadap Cacat Shrinkage Pada Proses Sintering Metal Injection Molding Al-Pp. Doctoral dissertation. 2020 Nov.
[20] Sateesh N, Reddy SD, Kumar GP, Subbiah R. Optimization of injection moulding process in manufacturing the flip cover of water meter. Materials Today: Proceedings. 2020 Nov;26:240-246.
[21] Moayyedian M, et al. Optimization of injection molding process based on fuzzy quality evaluation. CIRP journal of manufacturing Science and technology. 2018 May;21:pp 150-160.
[22] Bensingh RJ, et al. Injection molding process optimization of a bi-aspheric lens using hybrid artificial neural networks (ANNs) and particle swarm optimization (PSO). Measurement. 2019 Feb;134: pp 359-374.
[23] Kastner C, et al. Influence of process parameters on mechanical properties of physically foamed, fiber reinforced polypropylene parts. Journal of Applied Polymer Science. 2019 Dec;136(14): p 47275.
[24] Guo W, et al. A hybrid back-propagation neural network and intelligent algorithm combined algorithm for optimizing microcellular foaming injection molding process parameters. Journal of Manufacturing Processes. 2020 Feb;50: pp 528-538.
[25] Tsai KM. An inverse model for injection molding of optical lens using artificial neural network coupled with genetic algorithm. Journal of Intelligent Manufacturing. 2017 Nov;28(2): pp 473-487.
[26] Amin, et al. Green density optimization of stainless steel powder via metal injection molding by Taguchi method. In MATEC Web of Conferences. 2017 Feb;135(39).
[27] Prathabrao, et al. Parameter optimization of metal injection moulding: A review. Journal of Mechanical Engineering (JMechE). 2018 Oct; pp 100-114.
[28] Kitayama S, et al. Multi-objective optimization for minimizing weldline and cycle time using variable injection velocity and variable pressure profile in plastic injection molding. The International Journal of Advanced Manufacturing Technology. 2020 Feb;107(7): pp 3351-3361

New Materials

This page intentionally left blank

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 265
J. Xu et al. (Eds.)

doi:10.3233/ATDE220237

Experimental Study on Shear Mechanical Characteristics of Jujube Branches in Winter Pruning Period

Ning LI[a,b], Longpeng DING[a,b], Xianfei WANG[a,b], Gaokun SHI[a,b] and Jingbin LI[a,b,1]
[a] *College of Mechanical and Electrical Engineering, Shihezi University, Shihezi,* 832003, *China;*
[b] *Xinjiang Production and Construction Corps Key Laboratory of Modern Agricultural Machinery, Shihezi,* 832003*, China*

Abstract. The objective of this study was to design and optimize the pruning and crushing machinery of jujube branches, and to provide the corresponding data reference, so as to improve the cutting and crushing efficiency of jujube branches. Take fresh jujube branches (31.7±5 % w.b) as the research object. The morphology and microstructure of jujube branches were detected by scanning electron microscope. Four kinds of cutting blades were selected and manufactured, and the cutting characteristics of four kinds of blades against branches were evaluated. The shear fracture process of jujube branches was analyzed according to the microstructure of branches. Covariance analysis was used to separate the influence of water content on shear strength, and the significance level of the influencing factors is evaluated. The analysis of the test results showed that: the significant effects on shearing strength of the branches are shearing angle, cutter shape and shear speed from large to small. The force-displacement curves of cutting tool branches with different shapes of edge lines are different, and the peak of the shear force generally appears in theII&IVstages (xylem) shearing process. Under the same shear mode and the same shear velocity, the shear strength of the cambered cutter is the smallest.

Keywords. Jujube branch, Mechanical characteristics, Shear, Shearing strength

1. Introduction

Jujube, belong to the genus Jujube in the family Rhamnus, are a unique fruit tree native of China (Fu et al., 1993), and are widely planted with Xinjiang, Shandong, Shanxi, and other places in China (Qu et al., 2013). In the process of jujube growth management, jujube pruning is an important link, which can improve the light condition of jujube trees, improve the yield and quality of jujube (Chen et al., 2021). As forestry waste, the pruned jujube branches are mainly incinerated (Hu et al., 2017), which will cause pollution and waste of resources (Liu et al., 2011; Zhang et al.,2019). The sheer mechanical properties of jujube branches have a direct impact on the service life of the cutting tool and the energy consumption of jujube branches crushing machinery (Phillip et al., 2012).

[1]Corresponding author: Jingbin LI, College of Mechanical Electrical Engineering, Shihezi University, Shihezi, China; Email: ljb8095@163.com.

In recent years, different methods have been used to study the mechanical properties of crop straw in various aspects. The shear characteristics of crop straw are the basis for the development of agricultural harvesting, cutting and crushing machinery (Shi et al., 2017). At the macro level, Chen et al. (2012) passed the quadratic regression orthogonal rotation test, and found that the loading speed and water content had significant effects on the shear strength of corn stalk husk. Through the shear test, Li et al. (2011) found that the shear strength of cotton stalk was closely related to the water content. Kamandar et al. (2018) through the shear test of Ligustrum lucidum stems, and found that the loading speed and height had a significant impact on the shear strength and energy consumption of stems. Igathinathane et al. (2010) established the shear energy consumption model of maize straw, and found that the shear direction, internode and internode materials of maize straw had a significant impact on the shear characteristics of maize. At the micro level, Gibson et al. (2018) pointed out that plant cell walls were mainly composed of lignin, cellulose, hemicellulose and pectin, which provided a basis for the evaluation of mechanical properties of porous structures. Du et al. (2019) analyzed the structure of the cutting incision through X-ray and micro-CT, and found that the maximum cross-sectional area ratio and volume ratio of the cutting groove had a significant impact on the cutting force, cutting depth and blade angle. Wang et al. (2015) studied by scanning electron microscopy that the tissue composition of stems was the material basis for the mechanical properties of stems.

In summary, although there are many studies on the shear properties of crop straw in China and abroad (Zhang et al., 2019), there are few studies on the shear mechanical properties of jujube branches. Moreover, the research direction mainly focuses on the analysis of factors affecting the shear characteristics (Wang et al., 2020), and no analysis of the shear blade shape is carried out. Existing research mainly focuses on the analysis of shear mechanical properties, and does not combine with the microstructure analysis. In this study, the microstructure of jujube branches was observed by SEM, and the interaction of shearing angle, cutter type and other factors on the shear characteristics of jujube branches was studied. Combined with the microstructure of branches, the shear process of branches was analyzed. This study can provide a basic reference for the selection of cutter type and parameter optimization of jujube branch cutting machinery and branch crushing machinery.

2. Materials and methods

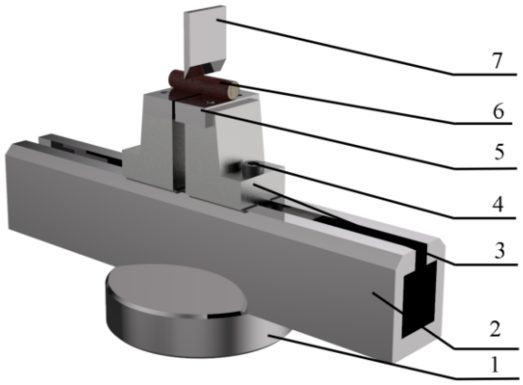
2.1. Testing material

The jujube branches(Variety: Junyou; Growth cycle: 1~2 year) used in the experiment were taken from the 2–1 section of the main jujube standardized test garden (79.29°E, 37.24°N) of 224 regiment of Xinjiang Province. The jujube branch samples were collected on November 28, 2020. The collection method of wood physical and mechanical materials in GB/T1927–2009 was referred to in the shear test sample collection (Wu et al., 2012). The branch samples were required to be straight and few bifurcations, no pests and diseases, no damage to the epidermis, no less than 50 mm in branch length, and 8~20 mm in branch diameter (Wang et al., 2020). The food preservation film was used for wrapping at the shear end of the branch. After sampling,

the branches were grouped and numbered, and put into a sealed bag for constant temperature preservation (Zong et al., 2012). In the shear test, the branches were trimmed into branches 50 mm in length and flat ends. Some branches were taken, sliced from the radial and transverse sections, and stored in the fixative for the observation of stem microstructure.

2.2. Experimental setup

The main equipment used in the shear test are:INSTRON-8801 hydraulic servo universal material testing machine (range: ±100KN), the accuracy of the test force indication: 0.5% of the indicated load value, INSTRON in the United States, self-made shear test bench as shown in Fig. 1, digital display vernier caliper (0.001mm), electric shear and cutting saw, jujube branch transverse cutting and radial slice samples were characterized by German Zeiss SIGMA 300 field emission scanning electron microscope. The universal material testing machine is connected to the PC computer through the controller. The data acquisition, processing and control are carried out by INSTRON 4.11 version software.



1.Supporting Platform; 2. Slide rail support; 3. Adjustment bracket; 4. Set bolt; 5. Stage; 6.Jujube branches; 7. Shear blade

Figure 1. Self-made shear test rig

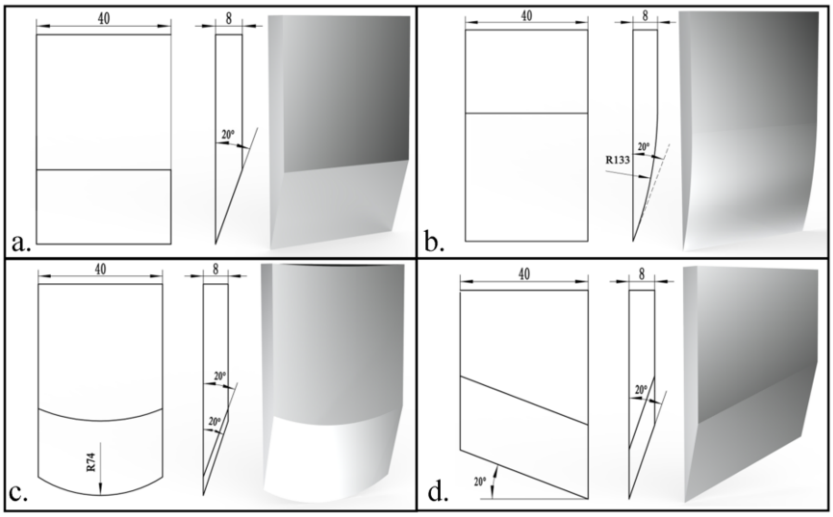
The water content was measured by 101-1B electric heating constant temperature blower dryer (temperature range: +10°C~300°C, temperature fluctuation: ±1°C, Shaoxing Luda Machinery Equipment Manufacturing Co., Ltd.) and JM-B100 electronic balance (precision: 0.001g, range: 200g).

2.3. Test method

According to Reference (Phillip et al., 2020), the cutting efficiency of the machine can be significantly improved by optimizing the cutting speed and blade shape. Therefore, four kinds of cutting tools are selected in this study and named flat cutter, sliding cutter, arc surface cutter and arc blade cutter. The four kinds of shear cutter type as shown in Fig.2. And cutter type parameters as shown in Table 1.

Table 1. The parameters of shear cutter type

Cutting type	Thickness (mm)	Edge angle (°)	Edge type	Others
Flat cutter	8	20	Straight edge	——
Sliding cutter	8	20	Oblique blade	Sliding angle: 20°
Arc surface cutter	8	20	Straight edge	Radius of arc surface: 133mm
Arc blade cutter	8	20	Arc edge	Radius of the blade line: 74mm



a. Flat cutter, b. arc surface cutter, c. arc blade cutter, d. sliding cutter

Figure 2. Four kinds of cutting blades

Before the test, adjust the cutting blade closer to the side of the branch platform. Adjust the blade to face side gap⩽1mm, adjust the gap between the left and right platform to 5mm, and lock the position. It doesn't make cutting happen continuously. The no-load inspection was carried out on the universal material testing machine to ensure that the platform and blade did not interfere. Check the blade force sensor work to verify the force-displacement diagram is in line with expectations.

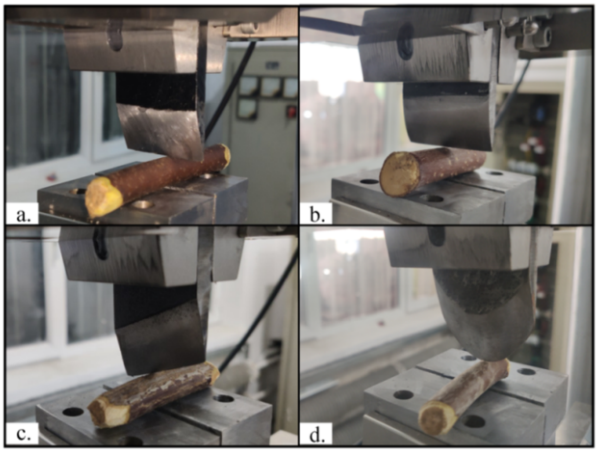


Figure 3. Shear test of jujube branches with different cutter types

Edit the shear test process using INSTRON 4.11 software. In the shear test, the shear cutter is fixed, and the bearing support and branch do uniform linear motion upward. When the cutting force is greater than 5N, the computer data acquisition system begins to record the instantaneous shear force and displacement and generate the force-displacement curve. When the shear force decreases by 40% instantaneously (the branches can be completely cut off), the test ends. Then the tool and the shear platform recover the initial position of the test and generate the force-displacement curve. Each shear test was repeated three times to take three averages for the final results. In order to obtain the shear performance of jujube branches and stems, the shear stress of branches and stems was obtained through experiments, and shearing strength can be calculated in Eq. 1 (Chen et al.,2016; Lien et al.,2015).

$$\tau = \frac{4F_{\max}}{ab\pi} \tag{1}$$

Where τ is the shear strength (The maximum shearing stress) (MPa), $F_{\max}$ is the maximum shear forces (N), a is the short diameter (mm) of jujube branch section, and b is the long diameter (mm) of jujube branch section.

After the shear test, the number and weight of the cut branches were weighed and oven-dried at 103 ℃ for 24 h and then weighed again to determine the water content(Wen et al., 2020). The water content ($w(w.b)$) was calculated by:

$$w(w.b) = \frac{w_W - w_d}{w_w} \times 100\% \tag{2}$$

The formula w_W is the branch weight (g) before drying. w_d is the quality of branches after drying (g).

2.4. Experimental design

Comprehensive literature found that many factors are affecting the shear performance of branches (Liang et al.,2020). It mainly includes branch cross-sectional area, blade type, shear mode (oblique angle), shear velocity, branch diameter, branch water content and so on (Wang et al., 2020; Moiceanu et al.,2013). The water content of branches is related to local precipitation and temperature, which cannot be adjusted in the production environment. Therefore, the water content of branches cannot be used as an experimental factor, but only as an external condition factor. In this experiment, flat cutter, sliding cutter, arc surface cutter and arc blade cutter were selected as four levels of shear cutter.

According to literature (Wang et al. 2014) and production experience, 5, 75 and 100 mm min^{-1} were selected as three levels of landing speed.

In summary, the shearing angle, cutter shape and shear speed were selected as experimental factors in this study. Due to the internal factors of the branch itself, the water content and shear section also affect the change of the curve. Water contented as the test external conditions discuss the significance level of three test factors, and through the principle of fixing the correlation to selects the appropriate cutting edge. In order to eliminate the effect of branch cross-sectional area on the shear performance of jujube branches, shearing strength of branch stem during shearing was selected as the test index.

In order to study the comprehensive influence of the above three factors on the shear performance of jujube branches and stems, the three-factor multi-level full-factor experimental design was adopted, and the instantaneous shear force and displacement curves of the cutter were obtained. A three-factor test and multi-level full-factor test were used in the experiment. Test factors and level table were shown in Table 2.

Table 2. Factors and evaluation indexes of the shear test of jujube branches

Level	Factor			targets of test
	Cutting type	Landing speed（mm min^{-1}）	Shearing angle（° ）	
1	Flat cutter	50	0	
2	Sliding cutter	75	30	Shearing strength (MPa)
3	Arc surface cutter	100	60	
4	Arc blade cutter			

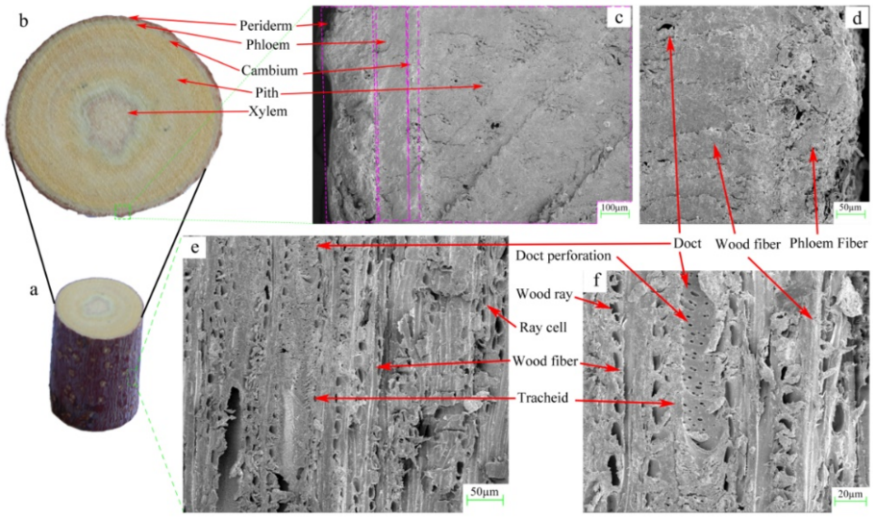
3. Results and analysis

3.1. Microstructure of jujube branches

The shearing force of jujube branches is closely related to the microstructure and the internal structure of materials (Zhao et al. 2011). To thoroughly understand and analyze the mechanical properties of branches, we must first understand the microstructure of stems. Fig.4(b) is the cross-section of jujube branches, and the color of each part is clear through observation in Fig.4(b). From outside to inside, it is divided into five parts: cortex phloem, cambium, xylem and bone marrow (Hu et al. 2017). As shown in Fig.4(c), the four purple regions are shown in turn from left to right, Periderm, Phloem, Cambium and Pith, corresponding to Fig.4(b). Fig.4(c) shows the scanning images of jujube branches under SEM. From the images, we can compare the residual branches to study the cambium, xylem and phloem, and have been marked in the images. Fig.4(b) and Fig.4(c) are the scanning images of bark-xylem on the cross-section under the conditions of amplification of 100 and 200 times, respectively. The structures observed in the images are bark, phloem, cambium and xylem (catheter, wood ray and wood fiber). It can be seen from the images that the regional organizational structure of the cambium and xylem is similar, and the area of a single phase is less than that of the xylem. Fig.4(e)and Fig.4(f) showed the longitudinal scan of jujube branches. From Fig.4(f), it can be seen that the catheter scattered, and there were often two catheters gathered together in the stem direction. There were no inclusions in the lumen of the catheter, and the wood ray was obvious. The catheter and the xylem were segmented. The tissue structure was supported by wood fiber, catheter and ray cells. Dense ducts are thinner in the cambium than in the xylem.

3.2. Shear process analysis

The cutting of jujube branches is a complex mechanical process of multi-factor coupling. When jujube branches are cut, they are subjected to the shearing force of the blade on the branch and the compression force of the blade on the branch (Zhao et al.2011).



a. Jujube branches samples, b. Cross-sectional view of jujube branches, c. Cross section of jujube branches 100 times lower, d.200 times the cross section of jujube branches, e. Vertical section of jujube branches at 200 times, f. Vertical section of jujube

Figure 4. Microstructure and SEM results of jujube branches . branches at 500 times

Because the stem tissue of jujube from outside to inside is composed of bark, xylem and pulp three parts (Fei al.2014). because the mechanical properties of each part are inconsistent with the mechanical superposition of branches. Four different regions (in Fig.4) can be identified in the observed force-displacement curve corresponding to the four stages of the cutting branch, and these regions of each blade are different, which can also illustrate the nature of the cutting process. Since the waveform changes little when the shear velocity changes, the force-displacement curves of branches under different cutter shapes and shear modes are shown in Fig.5.

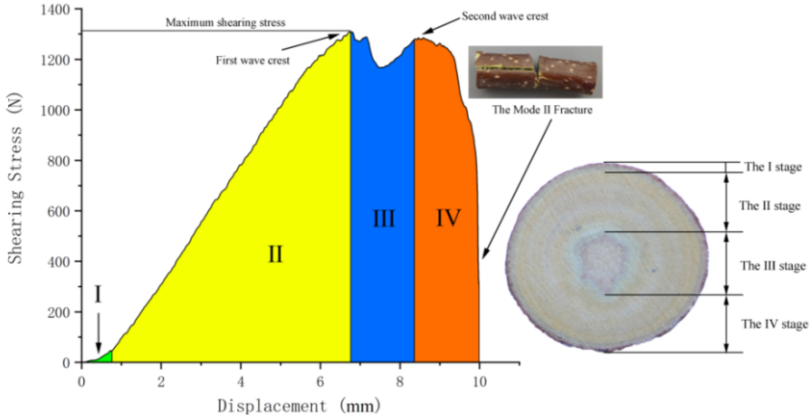
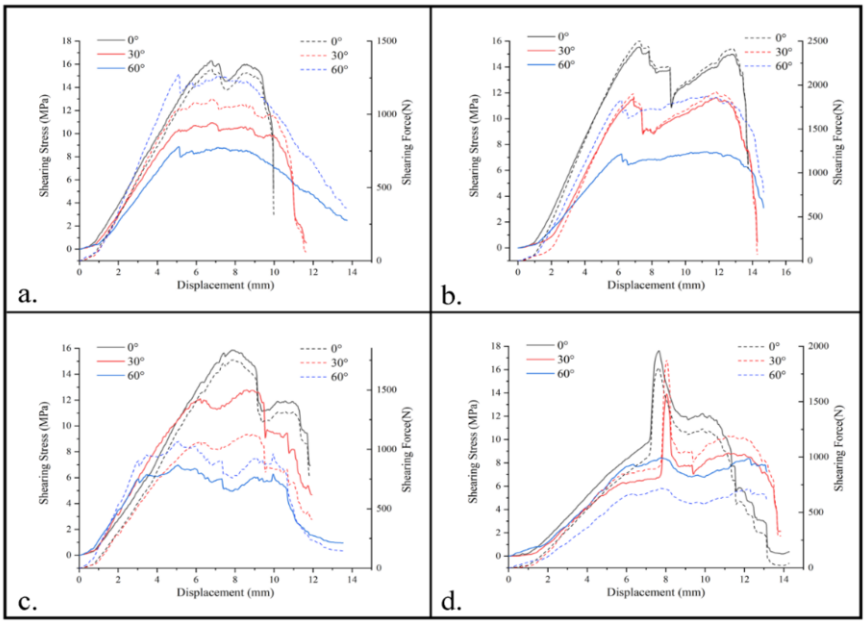


Figure 5. Different regions of cutting force-displacement characteristics of jujube branches explain the nature of cutting process

Due to the large difference in the instantaneous shear force between the structural anisotropy of jujube branches and its internal structure in the test. The cutting process of jujube branches and stems can be divided into four stages. In the I stage, the blade is close to and in contact with the branches. It compresses the branches to the near level

of shear failure. And the shear force gradually increases with the shear displacement. The blade cuts the first stage by compressing the branch and penetrating the bark (epidermis and phloem) of the branch. In the II stage, the blade passes through bark tissue. Then the cell wall in the xylem gradually thickened, the wood degree was higher, and the hardness was higher. It makes the shear force increase in a short time. When the sheer force of the xylem reaches the critical value, the shear force becomes larger and larger, and reaches the first peak of shear force, so that the upper xylem of the branch is cut. In the III stage, the cutter cuts into the middle tissue of the stem. Because the pulp is softer than the xylem, the shear force decreases rapidly and fluctuates slightly. The cutting position includes xylem, pulp and bark tissue, and the xylem still bears the main cutting force. Cutter cuts to the middle of the pulp, and then the cutting resistance will gradually increase until the pulp is completely cut. In the IV stage, the blade cuts into the xylem again after penetrating the pulp. The hard texture of the xylem requires a large load, so there is a second peak. Then the contact line between the blade edge and the xylem tissue gradually shortened. Its cutting organization gradually reduces, so the load decreases until the end of the test. But the anisotropy of the transverse and longitudinal arrangement of branches show in Fig 3.(a). In the shear process, the type II fracture along the axial direction of the blade surface will occur (in Fig. 4), resulting at the end of the test.



a. Flat cutter, b. Arc surface cutter, c. Arc blade cutter, d. Sliding cutter

Figure 6. Four Cutter Types of Shear Stress and Shear Stress Curves at Different Angles

Different regions of cutting force-displacement characteristics of jujube branches explain the nature of cutting process. Fig.5(a, b) and Fig.5(c, d) are compared. Fig.5(a, b) are straight edges. Fig.5(c) is arc edge. Fig.5 (d) is an oblique edge. By comparing the curves of four kinds of cutters, it can be found that four kinds of curves can be regarded as bimodal graphs. Because Fig.5(a, b) are straight edges, their edge lines are similar. Fig.5(c, d) and Fig.5(a, b) form curves are not similar, so the waveform generated by the inconsistent edge lines is also different, so the type of tool edge lines

will change. By changing the branch shear form (shearing angle), the formation of its curve is affected. Because the cutting angle changes, the length of the blade line and xylem will also change. In shearing, according to the end of stage II, stage III and the beginning of stage II, and the length of the edge has little changed. So the shearing load changes a little.

3.3. Results of branch shear test

Table 3. Results of branch shear test

Cutting type	Test. no	Shearing angle (°)	Landing speed (mm min^{-1})	sectional area (mm^2)	maximum shear force (N)	Shearing strength (N mm^{-2})	Water content (%)
Flat cutter	1	0	50	89.39	1543.2	17.26	32.58%
	2	0	75	124.36	1536.6	12.36	33.00%
	3	0	100	156.16	2302.4	14.74	32.75%
	4	30	50	100.91	1103	10.93	32.53%
	5	30	75	147.95	1298.1	8.77	35.48%
	6	30	100	131.05	1195	9.12	32.76%
	7	60	50	143.72	1276	8.88	32.05%
	8	60	75	76.60	702.3	9.17	34.07%
	9	60	100	131.01	1042.9	7.96	38.43%
Arc surface cutter	1	0	50	163.02	2309.9	14.17	35.89%
	2	0	75	116.38	1400.9	12.04	33.31%
	3	0	100	131.30	1460	11.12	30.44%
	4	30	50	146.61	1079.1	7.36	30.82%
	5	30	75	165.24	1222.1	7.40	25.73%
	6	30	100	149.32	1443.6	9.67	26.84%
	7	60	50	144.24	1130.2	7.84	28.26%
	8	60	75	179.82	1274.5	7.09	32.60%
	9	60	100	170.85	1054.8	6.17	29.11%
Sliding cutter	1	0	50	123.67	1908.8	15.43	33.60%
	2	0	75	100.74	1671.5	16.59	33.38%
	3	0	100	143.62	2133.6	14.86	31.50%
	4	30	50	100.74	1674.3	16.62	33.38%
	5	30	75	133.01	1284.1	9.65	34.41%
	6	30	100	156.89	1977.2	12.60	33.54%
	7	60	50	80.21	1031.1	12.85	34.23%
	8	60	75	161.05	1240	7.70	33.54%
	9	60	100	88.67	682.4	7.70	33.37%
Arc blade cutter	1	0	50	86.31	1416.2	16.41	30.64%
	2	0	75	110.57	1755.1	15.87	39.50%
	3	0	100	132.78	2031.5	15.30	31.31%
	4	30	50	100.64	1640	16.30	28.19%
	5	30	75	88.21	1128.3	12.79	32.16%
	6	30	100	81.98	974.4	11.89	36.76%
	7	60	50	112.78	134.6	1.19	27.38%
	8	60	75	153.34	1532.3	9.99	34.54%
	9	60	100	123.99	1068.8	8.62	33.18%

In the experiment, the water content is an independent variable and is not manipulated by the experimenter, but according to References (Wen et al. 2020), water content has an impact on shear force. Therefore, the branch water content was used as a covariant, the cutter type, shearing angle and shear speed were used as independent variables, and shearing strength was used as an evaluation index. SPSS Statistics 25

software was used to analyze the data in table 2 by the covariance analysis method. Under the condition of excluding the influence of covariant water content, the correlation between the three factors and shearing strength was analyzed. The covariance analysis results are as follows in Table 4.

Table 4. Principal Effect Analysis of Covariance of Cut Factors of Jujube Branches

Source	SS	d*f*	F-value	Eta-squared (η^2)	P-value
Model	380.86	9	9.247144	0.761958	P<0.001**
Cutting type	89.07	4	4.866	0.428121	P<0.005*
Shearing angle	151.35	2	16.536	0.559859	P<0.001**
Landing speed	14.36	2	1.568	0.107657	0.227464
Water content	4.43	1	0.967	0.035874	0.334390
Pure error	123.41	26			
Cor total	5042.82	36			

*Significant factor ($0.01 < p \leq 0.05$), ** extremely significant factor ($p \leq 0.01$), $p > 0.05$ non-significant factor. SS is sum of squares.

According to Table 4 the variance analysis results after the coverable (water content) are included. Model (P<0.001) indicates that the regression model is very significant and the equation fits well. Table 4 hows that water content is a covariate. Water content (P>0.05) had correlation with maximum shear stress but had no significant effect. Under the condition of deducting the influence of coverable, the cutter shape (P<0.005) factor has a significant influence on shearing strength. The shearing angle (P<0.001) factor has a significant effect on shearing strength. The shear velocity (P>0.005) factor has no significant influence on shearing strength, because the mechanical properties of individual branches differ greatly and the change of shear velocity is small. Therefore, the experimental shear velocity factor is not enough to have a significant effect on shearing strength. According to the partial Eta square value, the significance of each variable on shearing strength of the branch is the shearing angle, cutter shape and shear speed from large to small.

3.4 Effect of Cutter Type and Inclined Angle on Maximum Shear Stress of Branches

Inclined cutting refers to the oblique cutting surface and stem axis. But the cutting direction is perpendicular to the stem axis. At each angle (0 °, 30 °and 60 °) oblique cutting, loading speed of 50mm min^{-1}, compared four kinds of the blade on jujube branch maximum shear stress. A total of 12 samples were set up with 5 replicates per sample. The standard deviation and maximum shearing stress were shown in Fig.7. The maximum shear force decreases with the increase of oblique cutting angle. The reason is that the essence of shear is the tensile and compression of the internal organizational structure of the branch by the tool. According to the literature, the longitudinal tensile compressive strength of branches is higher than the transverse tensile compressive strength (Wang et al. 2020). When the oblique shearing angle is 0°, namely, transverse shear, shearing strength is also the largest because the longitudinal tensile strength and compressive strength along the cross-section of the stem need to be overcome. With the increase of oblique cutting angle (the angle between the cutting surface and the normal direction formed by the branch pipe), the tensile strength gradually shifts from longitudinal direction to transverse direction. Therefore, their shearing strength also decreases.

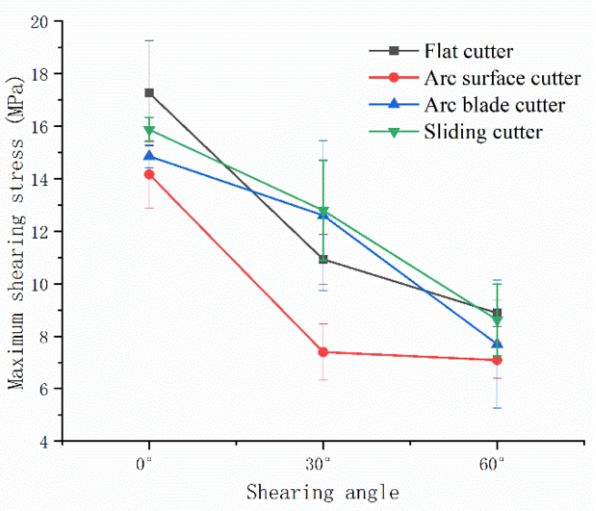


Figure 7. Shearing strength of four types of cutters at different angles, each half representing a standard deviation

Cutter type and shearing angle have a significant effect on shearing strength of branches. Fig. 7 shows the comparative analysis results of four types of tools under different oblique cutting angles. It can be seen from the above figure that the arc blade cutter and the arc surface cutter are significantly better than the other two cutter types at the oblique angles of 0° and 60°. And the arc cutter is better than the above three cutters at different shearing angles. Both the smooth cutter and the arc cutter belong to the straight-edged cutter (Teng et al.2009). The way they cut the branches is cutting. The oblique cutter and the arc cutter have the sliding cutting effect, which makes shearing strength of the two tools lower than that of the straight cutter. Both the smooth cutter and the arc cutter belong to the straight cutter with a blade angle of 20°. However, due to the change of the blade edge of the arc cutter, the normal force required by the tool cutting into the branch along the edge deviation direction is gradually reduced.

4. Conclusions

Four kinds of blades were selected and made, and the shear test bench was made to explore the mechanical properties of jujube branches and the influence of the shear process and cutter type on the shear characteristics of branches. In this study, the microstructure of the cross-section of branches was analyzed by SEM. Under quasi-static conditions, the shear test of jujube branches with the moisture content of 31.7±5 % was carried out and analyzed by covariance method. The conclusions are as follows :

The influence degree of single factor of branch shear from large to small is: shearing angle ($p<0.01$), cutter type ($p<0.05$) and loading speed. When sheared at different shearing angles, the mechanical properties of jujube branches in different directions and different shear planes showed significant anisotropic differences. As the shearing angle increases, the shear strength increases.

The force-displacement curves of the cutting process of the branch with different cutting tools are different. The SEM observation shows that the structure of jujube

branches is a great factor affecting the shearing process. The peak value of the maximum shear force generally occurs in the II&IV stages of the shear process.

As the same shearing angle and shearing speed, the shear strength of the smooth cutter is the largest, and that of the arc cutter is the smallest. The comprehensive performance of arc cutter is significantly better than that of the other three types. Through the comparative analysis of the smooth cutter and the curved cutter, it is found that the important reason for the smaller shearing strength of the curved blade is that the slope of the curved blade surface decreases gradually. This study could provide basic parameters for selecting and optimizing tool selection, cutting mode and cutting angle of pruning crusher.

Acknowledgement

This work was funded by the National Key research and development program of China (2016YFD0701504), The project to upgrade agricultural science and technology of the Xinjiang Production and Construction Corps (2130106-13), and Shihezi University Achievement Transformation and Technology Promotion Project (CGZH201909).

References

[1] Chen L., Yan G., Li Y.D., 2016. Experiment on mechanical properties of sorghum straw. Chinese Journal of Agricultural Machinery Chemistry, 37 (05), 130–135.
[2] Chen Y. Y., Zhang B., and Fu Y.X., 2021. Path Planning of Jujube Tree Pruning Robot (in Chinese). China Agricultural Mechanization Research, 43(12):37-41.
[3] Chen Z.G., Wang D.F., and Li L.Q., 2012. Experiment on tensile and shearing characteristics of rind of corn stalk. Chinese Journal of Agricultural Engineering, 28(21).
[4] Du Z., Hu Y.G., and Wu W.Y., 2020. Lu Yongzong,Noman Ali Buttar. Structural analysis on cutting notch of tea stalk by X-ray micro-computed tomography. Information Processing in Agriculture, 7(2). https://doi.org/10.1016/j.inpa.2019.09.003
[5] Fang Y., Ge Y., and Wang P., 2012. Experimental study on shear stress and tissue damage stress of jujube branches.Agricultural mechanization research, 34 (11): 182-185.
[6] Fei B.H., 2014. Characterization technology and application of mechanical properties of wood cell wall. Science Press.
[7] Feng Z.L., Du Y.Z., and Zhang J.G., 2008. Development of 4QZ-15 self-propelled green forage harvester. Research on agricultural mechanization in China, 000(3): 110-112.
[8] Fu W., He R., and Kan Z., 2013. Experimental Research on the Mechanical Properties of Jujube.Journal of Shihezi University (Natural Science Edition), 31(04):518-522.
[9] Glukhikh V., and Okhlopkova A., 2018. Predicting wood strength upon compression along fibers. Architecture and Engineering, 3(1). https://doi.org/ 10.23968/2500-0055-2018-3
[10] Hu Y. S. and Wang W., 2017. Discussion on the utilization of red jujube pruning residual branches. Xinjiang Agricultural Mechanization, (01):9-12.
[11] Hu Y.H., 2017. Experimental study on mechanical properties of jujube pruning stumps in Xinjiang. Tarim University.
[12] Igathinathane C., Pordesimo L.O., and Schilling M.W., 2011. Fast and simple measurement of cutting energy requirement of plant stalk and prediction model development. Industrial Crops and Products, 33(2): 518－523. https://doi.org/ 10.1016/j.indcrop.2010.10.015
[13] Igathinathane C., Womac A.R., and Sokhansanj S., 2010. Corn stalk orientation effect on mechanical cutting. Biosystems Engineering, 107(2): 97 － 106. https://doi.org/ 10.1016/j.biosystems-eng.2010.07.005
[14] Kamandar M.R., Massah J., and Khanali M., 2018. Quasi-static and mpact cutting behavior definition of privet stem. Agricultural Engineering International: CIGR Journal, 20(1):70−80.
[15] Li Y.D., Du X.J., and Song Z.H., 2011. Test on shear mechanical properties of cotton straw. Journal of Agricultural Engineering, 27 (2), 124-128.

[16] Liang R., Chen X., and Zhang B., 2020. Tests and analyses on mechanical characteristics of dwarf-dense-early major cotton variety stalks. International Agrophysics, 34(3), 333-342. https://doi.org/ 10.31545/intagr/122575
[17] Lien C.C.and Liu H. W., 2020. Shear Characteristics of Napier Grass Stems. American Society of Agricultural and Biological Engineers, In Agriculture, 31(1): 5-13. https://doi.org/10.13031/aea.31.10497
[18] Liu H. J., 2011. Comprehensive utilization technology of orchard pruning branches. Agricultural mechanization research in China, 33 (2): 218-220.
[19] Moiceanu G., Voicu G., and Paraschiv G., 2013. Cutting resistance of miscanthus plants using v shaped blades with opening of 50 degrees and different sharpening angles. Engineering for Rural Development - International Scientific Con. Jelgava, 05,509-515.
[20] Phillip C.J.,Clairmont L.C., and Sunil K.M., 2012. Cutting energy characteristics of Miscanthus X giganteus stems with varying oblique angle and cutting speed. Biosystems Engineering, 112(1). https://doi.org/ 10.1016/j.biosystemseng.2012.02.003
[21] Qu Z.Z. and Wang Y.C.,1993. Chinese Fruit Jujube Roll. Beijing. China Forestry Publishing House.
[22] Shi N., Guo K.Q., and Fan Y.J., 2017. Test on shear mechanical properties of cotton stalk extrusion peeling. China Journal of Agricultural Engineering, 33(18):51-5.
[23] Teng S.M., Wang Z.Q., and Li Y., 2009. Theoretical research on cutting mode and cutting resistance.Agricultural mechanization research, 031 (005), 89-90, 96.
[24] Wang B., He M., and Li C., 2020. Microstructure and biomechanical characterisation of jujube branches. Biosystems Engineering, 194, 165-176. https://doi.org/ 10.1016/j.biosystemseng.2020.04.004
[25] Wang H.L., Zhang W., and Dong X.W., 2014. Experiment and analysis of shear mechanical properties of corn stubble. Experimental Mechanics, 29(01): 73-82
[26] Wang T.Z., Zhang L., and Han Q., 2015. Effect of cell wall and tissue construction of maize stem on compressive strength. Journal of Plant Science, 33 (01): 109-115.
[27] Wang Y., Yang Y., and Zhao H., 2020. Effects of cutting parameters on cutting of citrus fruit stems. Biosystems Engineering, 193, 1-11. https://doi.org/10.1016/j.biosystemseng.2020.02.009
[28] Wen B.Q., Li Y., and Kan Z., 2020. Experimental research on the bending characteristics of glycyrrhiza glabra stems. American Society of Agricultural and Biological Engineers. Vol. 63(4):1-8. https://doi.org/10.13031/trans.13802
[29] Wen B.Q., Li Y., and Kan.Z., 2020. Experimental study on microstructure and mechanical properties of stalk for glycyrrhiza glabra. Journal of Biomechanics, 118(11), 110198. https://doi.org/10.1016/j.jbiomech.2020.110198
[30] Wu L.J., Yang Z., and Hong T.T., 2012. Experimental study on mechanical properties of litchi. Agricultural Engineering Journal, 28 (16): 68-73
[31] Zhang C., Chen L., and Xia, J., 2019. Effects of blade sliding cutting angle and stem level on cutting energy of rice stems. International Journal of Agricultural and Biological Engineering, 12(6), 75-81. https://doi.org/10.25165/j.ijabe.20191206.4604
[32] Zhang J., Ding L. P, and Li J.B., 2019. Design and test of branch crushing and returning machine in dense planting jujube orchard Research on agricultural mechanization in China, 41 (02) : 128-133.
[33] Zhang L.L., 2015. Bionic design and experimental analysis of cutting blade of branch crusher. Jilin University.
[34] Zhao C.H., Han Z.H., and Shi S.L., 2011. Mechanical properties and microstructure of newly bred forage stem during harvest. Journal of Agricultural Engineering, 27 (08): 179–183. https://doi.org/10.3969/j.issn.1002-6819.2011.08.030
[35] Zhao Y.B. and Chi S.Y., 2005. Design of drum-type crushing device for corn silage. Feed industry, 26 (15): 10-13.
[36] Zong W., Huang X., and Xu A., 2012. Bending and shearing characteristics of cotton seedling stem. Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, 28, 118-124. https://doi.org/10.3969/j.issn.1002-6819.2012.z2.021

doi:10.3233/ATDE220238

Design of the Penetrating-Cabin Assembly System for Slender Engine

Shuanli JIA [a], Long HE [b], Hanpeng WEI [b], Yunfei YANG [a], Na PENG [c], Jiayu SHE [a], Naiming QI [a] and Yanfang LIU [a,1]
[a] *School of Astronautics, Harbin Institute of Technology, Harbin, China*
[b] *Beijing Xinghang Electro-mechanical Equipment Co., Ltd, Beijing, China*
[c] *Shanghai Aerospace Electronic Technology Institute, Shanghai, China*

Abstract. In the process of assembling a slender engine into a cabin, it is inevitable to face the constraints of the narrow space. This paper proposes an assembly system that assembles a slender-diameter ratio engine and other compartments into a cabin. The assembly system utilizes visual position measurement technology and robotics and improves assembly efficiency and assembly quality.

Keywords. The slender-diameter ratio engine, Penetrating-cabin assembly, Visual measurement technology, Robotics

1. Introduction

The engine's assembly quality directly affects the performance of the entire product. At present, the domestic engine assembly mainly focuses on fixed stations and adopts the method of visual inspection-adjustment-visual inspection for docking installation. The position and posture of each unit body is adjusted manually, which is likely to cause bumps and jams. The assembly quality is unstable and the assembly efficiency is low [1]. Statistics shows that assembly process consumes about 45% of production work load and nearly 54% of cost [2-3]. Robots are widely used to improve assembly efficiency [4-5]. Thus, it is of great significance to study the engine assembly method with advanced technology.

This article mainly focuses on the system design and analysis of the assembly docking of a slender engine. Compared with the original assembly system, the assembly system utilizes visual position measurement technology and robotics and improves assembly efficiency and assembly quality. The requirement and composition of the assembly system is presented in Section 2. In Section 3, the trajectory of the cabin is planned and demonstrate by using RobotStudio. The visual position measurement system is presented in Section 4 and the assembly efficiency is given in Section 5. Results show that the assembly efficiency increases by 73% and the number of workers reduce by about 60%.

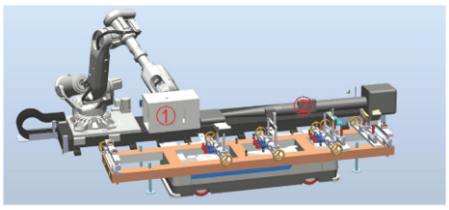
[1] Yanfang Liu, School of Astronautics, Harbin Institute of Technology, Harbin 150001, China; E-mail: lyf04025121@126.com.

2. Design of assembly system

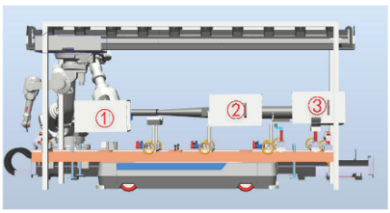
2.1. Analysis of assembly requirements and difficulties

The assembly process is shown in Figure 1: Move Cabin 1 from position ① to ③; Install component into Cabin 1 at position ②; Monitor small gaps at positions ① and ③.

The assembly stroke of Cabin 1 is about 3.5m. The minimum gap between engine and Cabin 1 is 4.5mm. It is difficult to manually adjust the position and attitude of the cabin with high accuracy. What's worse, it is hard to observe directly as the assembly environment is poor visibility.



a) Engine and Cabin 1

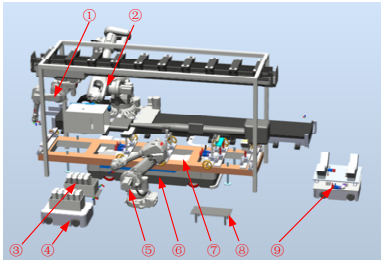


b) The process of assembly

Figure 1. Product and assembly process.

2.2. System composition and hardware structure design

As shown in Figure 2, the penetrating-cabin assembly system includes engine adjustment mechanism, robots, AGVs, storage platform, and work platform.



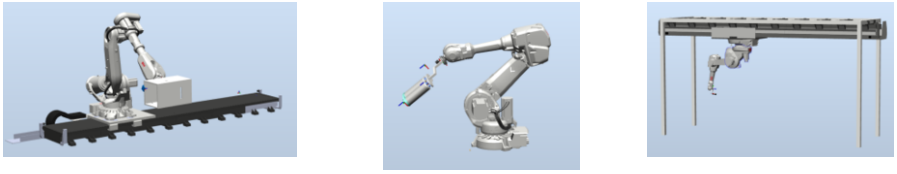
① robot3, ② robot1, ③parts storage platform, ④ AGV-1, ⑤ robot2, ⑥ AGV-2, ⑦ engine adjustment mechanism, ⑧ work platform, ⑨AGV-1

Figure 2. System composition.

Robot 1 equips with a camera and a grabbing tools, as shown in Figure 3 (a). It grabs Cabin 1 from AGV-1 and moves it along the planned trajectory.

Robot 2 also equips with a camera and a parts grasping tools, as shown in Figure 3 (b). It grabs parts from the storage platform and moves them into Cabin 1 following the planned trajectory.

Robot 3 consists of the seventh axis and equips with a vision system, as shown in Figure 3 (c). It monitors the key positions during the passage of the penetrating-cabin.



a) Robot1 b) Robot2 c) Robot3

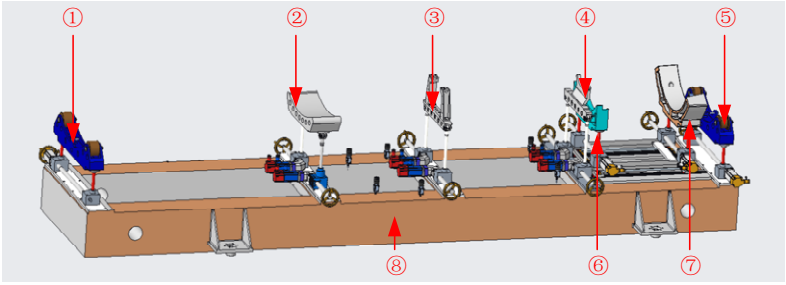
Figure 3. Robot 1, 2 and 3

Engine adjustment mechanism consists of 7 adjustable supports and a chassis, as shown in Figure 4. It supports and adjusts the engine, Cabin 1.

The assembly process is shown in Figure 5.

3. The trajectory planning of robot

The trajectory of the penetrating-cabin planned, simulated in RobotStudio and shown in Figure 6. The trajectory of the penetrating-cabin is not complicated but difficulty to control accuracy. Thus, the method of teaching is directly used for trajectory planning.



①⑤-cabin2 support, ②③④-engine support, ⑥⑦-cabin1 support, ⑧- chassis

Figure 4. The engine adjustment mechanism.

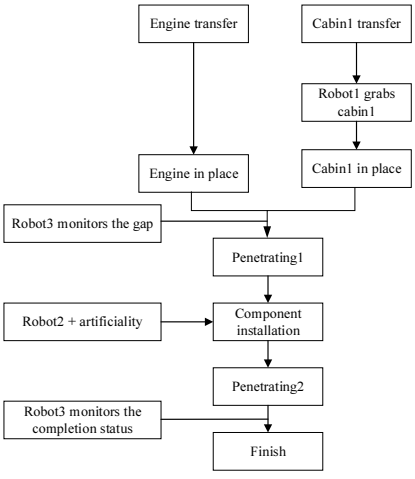
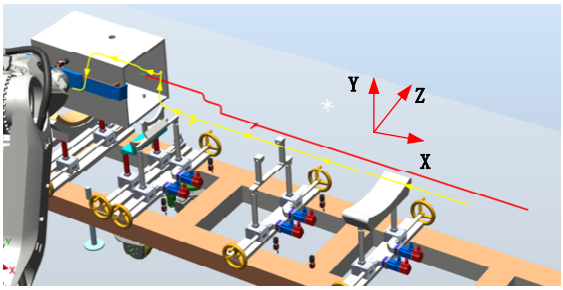
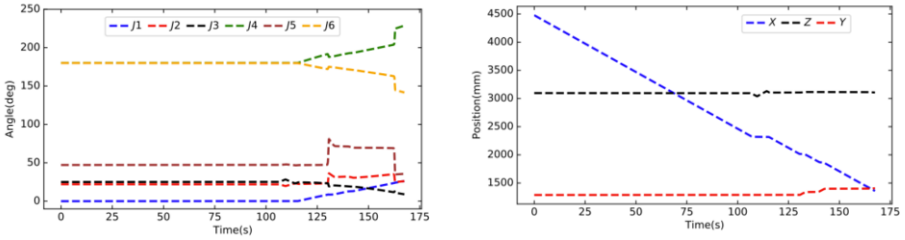


Figure 5. The process of assembly.



red line: the track of the center of the rear end face of cabin1, yellow line: the track of the end of the robot

Figure 6. The trajectory planning of robot.



a) Joints Angles of Robot 1 b) Position of the end of Robot 1

Figure 7. Simulation results.

The angle curve of 6 joints of Robot 1 and the position of the end of Robot 1 are shown in Figure 7. At the first step, the seventh axis of Robot 1 moves. At about 100s, Cabin 1 moves to the position where components are installed. Then, there is an offset in the z direction to leave space for Robot 2 to install components. During 100s~160s, the seventh axis keeps in the same position, and other 6 axis moves. The angles of Robot 1 changes smoothly, which indicates that the planned trajectory is effective.

4. Image processing and analysis

During the assembly process, it is necessary to monitor the narrow gap to ensure that there is no collision and the safety of the product. The visual measurement technology is utilized to monitor the narrow gap.

Bilateral filter is used to preserve edges and reduce noise. The characteristic of edge preservation is mainly realized by the combined operation of spatial function and range function. The output pixel value depends on the weighted combination of adjacent pixel values:

$$g(i,j) = \sum_{k,l} f(k,l)\omega(i,j,k,l) \Big/ \sum_{k,l} \omega(i,j,k,l) \tag{1}$$

The weight function is expressed as:

$$\omega(i,j,k,l)=\exp(-(i-k)^2+(j-l)^2/2\sigma_d^2-\|f(i-j)+f(k-l)\|/2\sigma_r^2) \tag{2}$$

In order to achieve better edge extraction in the later stage, the image needs to be binarized. According to the actual characteristics of the image after the first two preprocessing, the OTSU threshold for binarization is used.

In order to find the strength and direction of the edge at the $f(x,y)$ position of an image, the gradient is selected as the main tool. The gradient is represented by ∇f , and is defined by a vector:

$$\nabla f \equiv grad(f) \equiv \begin{bmatrix} g_x & g_y \end{bmatrix}^T = \begin{bmatrix} \partial f/\partial x & \partial f/\partial y \end{bmatrix}^T \tag{3}$$

The gradient direction is shown in the Figure 8, and α is the angle reflected on the x axis. The overall gradient of a picture is derived from $\partial f/\partial x$ and $\partial f/\partial y$ of each pixel in the picture. The simplest gradient approximation expression is:

$$\begin{cases} g_x = \partial f(x,y)/\partial x = f(x+1,y)-f(x,y) \\ g_y = \partial f(x,y)/\partial y = f(x,y+1)-f(x,y) \end{cases} \tag{4}$$

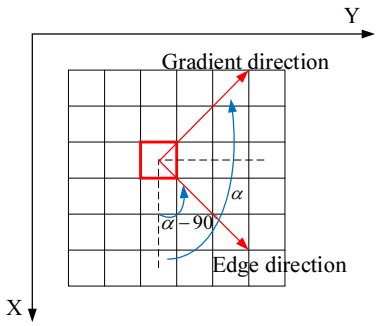
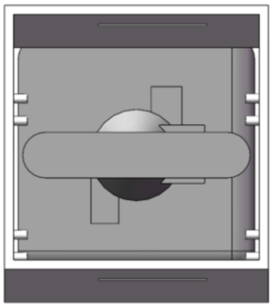


Figure 8. Schematic diagram of gradient direction.

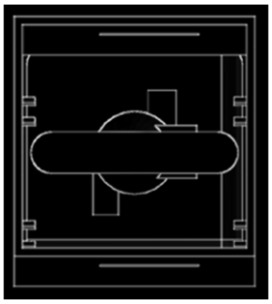
The picture effect processed by the Sobel operator is regarded as one of the best operators because of its excellent noise reduction and clear outline[6-8]. The Sobel operator expression is as follows:

$$s=(dx^2+dy^2)^{1/2} \tag{5}$$

The templates of the two directions of the Sobel operator are applied to the detection, and the detection result is shown in Figure 9. Analyze the edges of the engine and cabin1 to get the size of the gap.



a) Original image



b) Processed image

Figure 9. Edge detection.

5. Analysis of assembly efficiency

The original assembly system is assembled by manual + special tooling. The number of workers in the original assembly system is 8-10. The number of workers in the new assembly system is 2-3. A specific assembly time is given in Table 1.

Table 1. Comparison of assembly time between new and old assembly systems

Process	Time /min	Process	Time/ min
Engine transfer (parallel)	15.0	Engine transfer (parallel)	5.0
Transshipment of cabins to be crossed (parallel)	10.0	Transshipment of cabins to be crossed (parallel)	5.0
Tooling is ready to be in place	10.0	Robot1 grabs cabin1	2.0
Product and tooling connection	30.0	Robot3 monitors the gap	2.0
Penetrating2	20.0	Penetrating1	1.7
Component installation	30.0	Component installation(Robot2 + artificiality)	20.0
Penetrating2	20.0	Penetrating2	0.6
-	-	Robot3 monitors the completion status	2.0
Total (out parallel time)	125.0	Total (out parallel time)	33.3

According to the data, the proposed assembly system increases the assembly efficiency by 73% and reduces the number of workers by about 60%.

6. Conclusion

An assembly docking system for slender engines is designed. First, the hardware structure was designed, and the model was built in RobotStudio software. Then the trajectory of cabin1 is planned. Through simulation analysis, it is known that the running time of the entire trajectory is about 160s without considering the time of manual installation of components. The angle of each joint changes smoothly. Then the image processing method is analysed, Bilateral filtering is performed on the image to reduce the influence of noise, and then threshold processing is performed to reduce unnecessary information in the image. Choosing the Sobel operator in the edge detection, the resulting image contour has fewer fractures, better continuity, and can better monitor the narrow gap value. Finally, the assembly efficiency of the new and

old systems was analysed, and results showed that the presented assembly system increases the assembly efficiency by 73%, and reduces the number of workers by about 60%.

References

[1] Ling W. Study on automatic docking technology for aero-engine low pressure turbine installation. Gas Turbine Experiment and Research. 2018 October. p. 1-5.
[2] Wenlei Z, Chengen W, Jiapeng Y. A Design for Assembly System for Aero Engine. International Conference on Computer and Automation Engineering. Bali, Indonesia.2015. p. 328-331.
[3] Abhilasha S, Takashi N, Jun K, Kazuhiro K. Automatic path planning for 3-D assembly system using configuration space and CAD model. International Conference on Mechatronics and Automation. Beijing, China. 2015 August 2-5. p.1355-1361.
[4] Yaotong J, Gang Z, Wei W, Xian C. A Practical Path Planning Method in Aircraft Assembly Process Simulation. International Conference on Intelligent Computing, Automation and Systems (ICICAS). Chongqing, China. 2019 April 2. p. 411-415.
[5] Xue W, Jianfeng L, Ruxiao C, Mengying X, Luyao X. Research on Design and Planning of Pulsating Aero-engine Assembly Line based on Plant Simulation. Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). 2020 December. p. 591-595.
[6] Junjie D. Research on object recognition and surface defect detection based on machine vision. M.E dissertation, Dali University, Yun Nan, China.
[7] Saeedeh A, Mohammad A. Tehran, Mark D. Fairchild. Colour metrics for image edge detection. Colour research and application. 2020 August. p. 632-643.
[8] Xueqin L, Zhen L, Guofu Y, et al. Ferrite Magnetic Tile Defects Detection Based on Nonsubsampled Contourlet Transform and Texture Feature Measurement. Russian journal of nondestructive testing. 2020 April. p. 386-395.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 285
J. Xu et al. (Eds.)

doi:10.3233/ATDE220239

Size Optimization of a Carbide Anvil Based on Thermodynamic Coupling Analysis

Guizhong XIE[a,1], Tao WANG[a], Liangwen WANG[a], Xiaoyun GONG[a], Shixin ZHANG[a], Zeheng ZHI[a], Ziyong ZHAO[b], and Xiaojun YANG[b]
[a]*Mechanical and Electrical Engineering Institute, Zhengzhou University of Light Industry, Zhengzhou 450002, China*
[b]*Henan Huanghe whirlwind Co.. LTD, Changge, 461500*

Abstract. This paper presents size optimization of a carbide anvil based on thermodynamic coupling analysis. In our method, the established carbide anvil assembly through SolidWorks is firstly imported into the finite element software. Then, the temperature field and thermal-mechanical coupling field of the carbide anvil assembly of the diamond press were analyzed. From the simulation results, it can be found that the contact stress of steel ring under temperature load increases by 17.9% compared with that without temperature load. Thus the service life of carbide anvil under temperature load is lower than that without temperature load. It is consistent with the actual crack position of the carbide anvil, which proves the accuracy and effectiveness of thermal-mechanical coupling simulation. The optimization results show that the optimum effect is best when the thickness of carbide anvil is 1.895 cm, the face of carbide anvil width is 16.452 cm, and the height of anvil is 21.03 cm.

Keywords. Thermodynamic coupling, carbide anvil, size optimization, cemented carbide layer

1. Introduction

As the largest producer of synthetic diamond in the world, China's synthetic diamond output accounted for about 90% of the world's synthetic diamond production in 2009, with a total annual output of 10-12 billion carats, occupying an important position in the industry. In particular, for cutting and polishing tools, it is estimated that about 98% of industrial-grade diamonds use artificial diamonds. Because the diamond has high hardness, good thermal conductivity, wear-resistance and excellent light transmittance and corrosion resistance, etc., it is widely used in the electronic industry. To improve the quality of diamond, the development of high-performance cubic synthetic diamond press and carbide anvil system has become the primary task of artificial diamond industry [1-5].

[1] Corresponding Author, Guizhong XIE, School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China; E-mail: xieguizhong@126.com

2. Thermal analysis of carbide anvil

The internal environment of the carbide anvil system of the six-sided anvil press is high temperature and high pressure. Under such working conditions, the failure of the carbide anvil may be accelerated, resulting in cracking of the carbide anvil in the synthetic cavity, which affects the quality of the diamond [6-7]. Therefore, it is necessary to analyze the carbide anvil and observe the temperature distribution.

2.1. Finite element modeling of thermal analysis of carbide anvil system

Import the carbide anvil system model into finite element software (each part are demonstrate in Figure 1). The carbide anvil system assembly comprises a carbide anvil, a steel ring and a cushion block. The top material of the carbide anvil is YG8 carbide and others are made of low alloy ultra-high strength steel 45CrNiMoNA [8]. Stratified carbide anvil is shown in Figure 2. See Table 1 for relevant parameters of two different materials. Hexahedral element is used to divide the mesh for cemented carbide layer and cushion block, and tetrahedral element with good adaptability is used for other parts. Element size is 0.001 m. The result is shown in Figure 3. The carbide anvil top surface of the carbide anvil system is subjected to a temperature load of 200℃ [9]. The outer surface temperature of the steel ring is 85℃, and the flow heat transfer coefficient is 1000 ($W/m^2 \cdot ℃$). The inclined surface of the steel ring and carbide anvil defines the convective coefficient as 20 ($W/m^2 \cdot ℃$) and which surface temperature is 20℃.

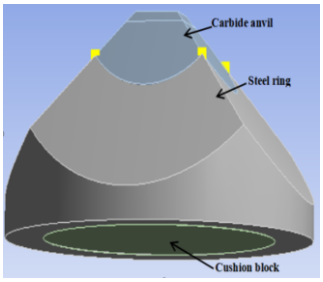


Figure 1. Drawing of all parts of carbide anvil system.

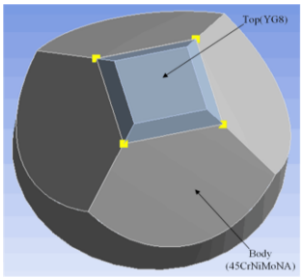


Figure 2. carbide anvil stratification.

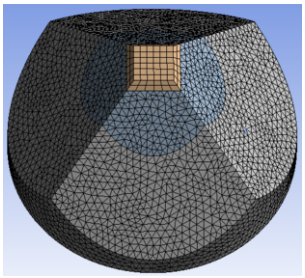


Figure 3. Assembly grid division of carbide anvil.

Table 1. material parameter of YG8 and 45CrNiMoNA.

material	density (kg/m^3)	thermal conductivity $(W/m \cdot K)$	thermal expansivity (m/K)	elastic modulus (pa)	Poisson ratio
YG8	15000	75.4	4.5×10^{-6}	6×10^{11}	0.22
45CrNiMoNA	7900	20	14.5×10^{-6}	2.01×10^{11}	0.3

2.3. Temperature field of the carbide anvil system

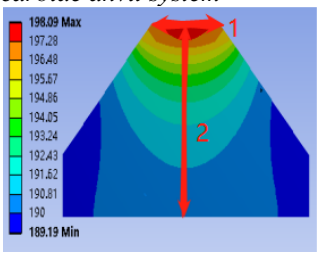


Figure 4. The temperature distribution of carbide anvil.

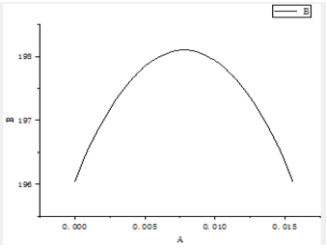


Figure 5. Path 1 temperature distribution.

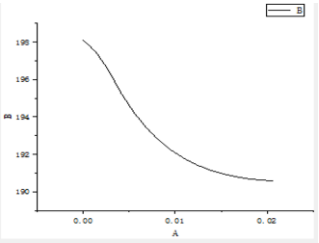


Figure 6. Path 2 temperature distribution.

It can be seen from the simulated temperature trend diagram, and the maximum temperature is 198.09℃, which is not much different from the actual measured temperature of 200℃ [10]. From the path defined in Figure 5, the results can be seen that the exit temperature gradually adds from the diagonal point to the center point, the temperature gradually decreases from the center point to the oblique point on the other side. From path 2, the result is shown in Curve 6, the temperature gradually decreases with the increase of the distance from the carbide anvil top surface to the bottom surface. This is due to the primary heat source applied to the carbide anvil top surface, heat transfer along the axis, the temperature is decreasing.

2.4 Temperature field of the steel ring and cushion block

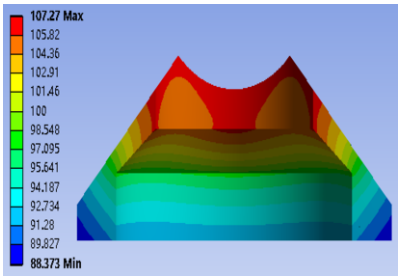


Figure 7. Temperature distribution of steel ring.

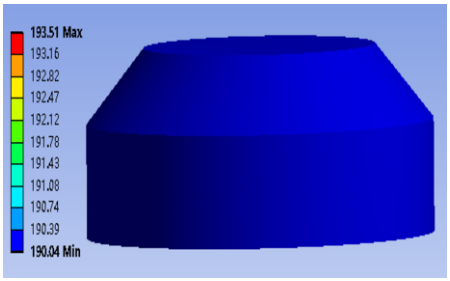


Figure 8. Temperature distribution of cushion block.

The maximum temperature of the steel ring is 107.27℃. Air contact and water flow can remove some of the heat, resulting in the uneven temperature distribution of the steel ring. Because the cushion block is in contact with the carbide anvil, the temperature loss is less, and the temperature is stable at about 190℃.

2.5 Analysis of pre-tightening force of steel ring

As shown in Figure 9(a) below, the pre-tightening force of steel ring without load and temperature influence. The contact stress of the steel ring is very uneven due to the thinning of the steel ring containing the carbide anvil after the material is removed and the cutting of the steel ring incline 46° down the cylinder [11]. The maximum pre-tightening force is 852.6Mpa. Then, the analysis results of the temperature field were imported into the analysis of the stress field, and the results are shown in Figure. 9(b). The maximum stress is 1005Mpa. Compared with the preload under the condition of no load, the preload increased by about 17.9%, so it can be seen that the temperature will lead to the thermal deformation of the material and produce thermal expansion force, which affects the stress of the carbide anvil system.

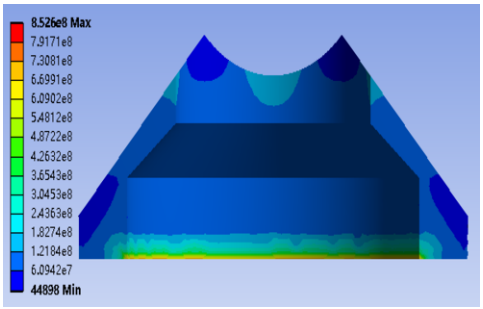


Figure 9(a). Preload of steel ring without load.

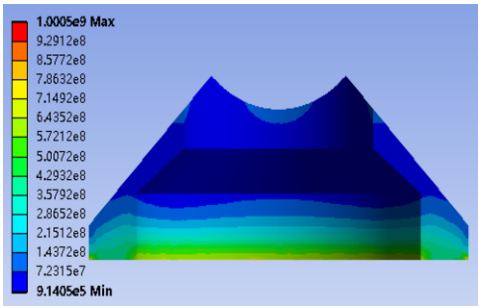


Figure 9(b). Preload of steel ring under temperature load.

3. Carbide anvil stress analysis under thermodynamic coupling

Based on the above temperature field analysis results, the simulation results are imported into the stress field, and the corresponding boundary conditions are applied to analyze the coupling field. In the actual process of diamond production, the carbide anvil is not only affected by the high-pressure $P_1 = 1$Gpa acting on the carbide anvil top surface by the synthetic cavity, but also by the positive pressure P_2 acting on the small inclined plane of the carbide anvil by the sealing gasket and the friction force f acting on the small inclined plane by the composite block. A fixed constraint is applied to the bottom of the anvil assembly. The load distribution of the carbide anvil is shown in Figure 10. The stress of coupled field carbide anvil system after applying boundary conditions is shown in the figure in Figure 11 to Figure 14.

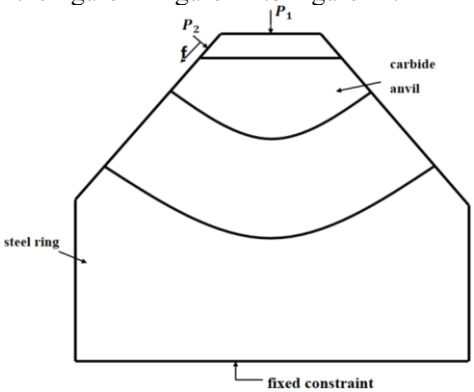


Figure 10. Boundary conditions of anvil system.

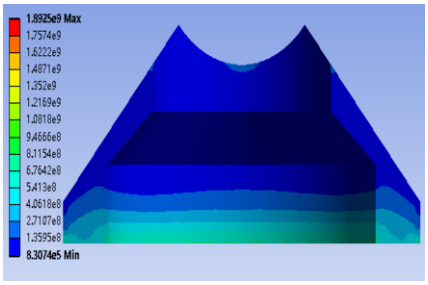


Figure 11. Stress diagram of steel ring under thermodynamic coupling.

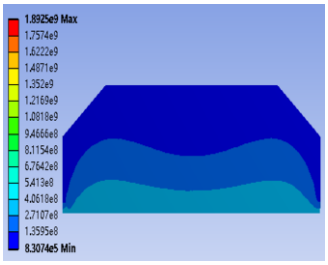


Figure 12. cushion block strain diagram under thermodynamic coupling.

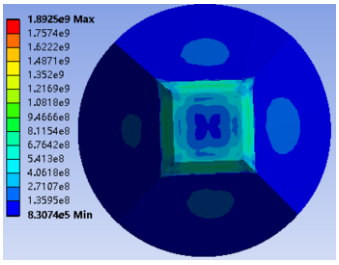


Figure 13. carbide anvil contact stress diagram under thermodynamic coupling.

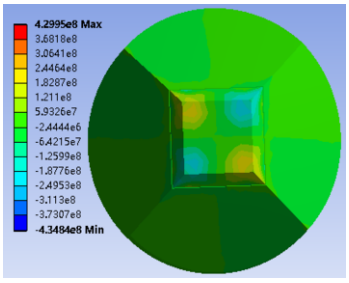


Figure 14. Shear stress of carbide anvil under thermodynamic coupling.

It can be seen that the carbide anvil system is more affected by temperature than before, and the stress after thermodynamic coupling generally increases a lot due to the generation of thermal expansion force. Due to the material of the carbide anvil of the cubic synthetic diamond press being brittle material, and the fatigue failure of brittle material is mainly shear failure caused by shear stress [12]. The comparison between Figure 13 and Figure 14 shows that the shearing force of the carbide anvil increases a lot under the influence of temperature. The simulation results also explain the fact that the service life of the carbide anvil under the temperature load is lower than that without the temperature load.

4. Optimize the size of cemented carbide layer of carbide anvil

Because the cemented carbide layer thickness can affect the carbide anvil stress distribution, however, the stable stress can better simulate the diamond forming conditions in natural environment for graphite [13]. Therefore, it is very important to determine the optimal thickness of cemented carbide layer for producing high-quality diamonds. In the 3D software SolidWorks, a parametric model is established according to the dimensions of carbide anvil. The basic dimensions are imported into the finite element software, and the associated dimensions are established with the software. The multi-objective optimization mathematical model is:

$$\min\{f_1, f_2, \cdots, f_n\} \tag{1}$$

$$g_i(x) < 0 \qquad i = 1, 2, \cdots l \tag{2}$$

$$h_j(x) = 0 \qquad j = 1, 2, \cdots m \tag{3}$$

$$x_L \le x \le x_U \qquad x = (x_1, x_2, \cdots x_r) \in X \tag{4}$$

4.1 Sensitivity analysis of carbide anvil size

Taking the dimensions of carbide anvil as input variables, the mass, average temperature, maximum contact and shear stress were defined as target variables. Then the sensitivity analysis of carbide anvil size parameters is carried out to find out the parameters that have a greater impact on the target variables. As shown in figure 15. Parameter $P_1 - P_7$ successively represents height and width of the cushion block, carbide anvil width, the face of carbide anvil width, carbide anvil height, edge length of carbide anvil top surface, and thickness of cemented carbide layer. Parameter $P_8 - P_{12}$ successively represents the carbide anvil mass, average temperature, maximum contact and shear stress of carbide anvil [14]. It can be seen that the influential parameters of the anvil are carbide anvil width, carbide anvil height and thickness of cemented carbide layer.

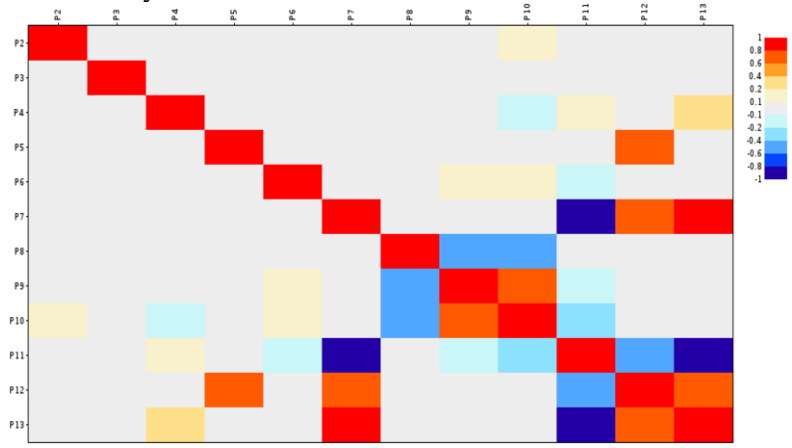


Figure 15. Correlation matrix diagram.

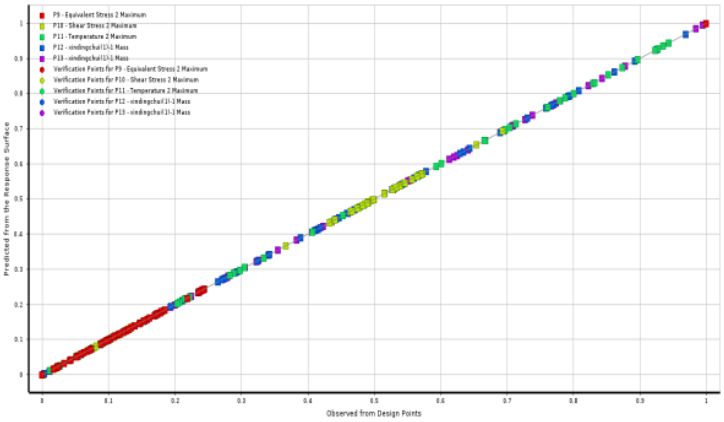


Figure 16. Parameter fitting curve.

In this paper, the upper and lower 20% variation values of each size are assigned as the parameter variation range, and a total of 70 sample points are selected for optimization calculation. Kriging model was selected for optimization, as shown in figure 16 This model can well fit the sample points near a straight line, and the constructed response variation model is suitable for the model in this paper.

4.2 Analysis of optimization results

The optimization simulation uses the previous calculation results of thermodynamic coupling to bring different sample points into the constructed coupling field for calculation again. Until the calculation is complete, choose the best result. The final optimization results show that the optimization effect is the best when the layer thickness is 1.895 cm, the carbide anvil height is 21.759 cm, and the face of carbide anvil width is 16.452 cm; The maximum contact stress decreased by 249.5Mpa and the maximum shear stress decreased by 166.25Mpa.

5. Conclusion

In this paper, the finite element method is used to simulate the carbide anvil under the joint action of temperature field and stress field. The analysis results show that the center temperature of the carbide anvil is the highest in the temperature field. The temperature distribution of the steel ring is not uniform, and the temperature of the cushion block is approximately close to 190 ℃, which is almost the same as the temperature of the carbide anvil. Due to the influence of thermal expansion, the contact stress of the steel ring increases by 17.9% under the temperature field. In the analysis of the coupling field, the shear force is the largest at the four corners of the carbide anvil. Because the carbide anvil system is fixed at the bottom, the stress is greatest at the bottom. In the optimization analysis, in the 20% change above and below each size, when the thickness of cemented carbide layer is 1.895 cm, the face of carbide anvil width is 16.452 cm, and the carbide anvil height is 21.759 cm, the optimum stress results are obtained. The research of carbide anvil system in this paper has important practical significance for improving the service life of carbide anvil and designing the size of carbide anvil [15].

Acknowledgments

This work was supported by Key Scientific Project of Henan Province (211110220200).

References

[1] Ge Y, Qigang H, Mingzhe L, et al. Finite element analysis of the high-pressure tungsten carbide radius-anvil. Acta Physica Sinica. 2012; 61(4): pp 273-335.
[2] Xianjun Z, Zhongwen X, Renquan C, et al. Coupled Thermal Analysis on Carbide Anvil of Cubic Press. Advanced Materials Research. 2014; 852: pp 629-633.
[3] Singh AK, Divakar C, Mohan M. Step-loading technique for tungsten carbide opposed anvil high-pressure setup. Review of Scientific Instruments. 1983; 54(10): pp 1407-1409.
[4] Qigang H, Qiang Z, Mingzhe L, et al. An Effective Solution for the Best Set of Beveling Parameters of the Cubic High-Pressure Tungsten Carbide Anvil. Chin.phys.lett, 2012; 29(11): pp 116201.
[5] Jun Y, Bin C, Yanan W, et al. Crack detection in carbide anvil using acoustic signal and deep learning with particle swarm optimisation. Measurement. 2020; 173: p 108668.
[6] Singh AK, Devi SU . Use of Energy Dispersive Method with Tungsten Carbide Opposed Anvil High Pressure Set Up. 1986.
[7] Li H, Bin C, Baocheng G, et al. Fault Detection of Carbide Anvil Based on Hurst Exponent and BP Neural Network. Advanced Materials Research. 2013; 805-806: pp 1881-1886.
[8] Kumar P, Sinha A N. Studies of temperature distribution for laser welding of dissimilar thin sheets through finite element method. Journal of the Brazilian Society of Mechanical Sciences and Engineering 2018; 40(9): p 455.
[9] Panda P C, Ruoff A L. The relation of the yield stress of high-pressure anvils to the pressure attained at yielding and the ultimate attainable pressure. Journal of Applied Physics. 1979; 50(2): pp 582-588.
[10] Osakabe T, Kakurai K, D Kawana, et al. Development of a hybrid-anvil type high-pressure device and its application to magnetic neutron scattering studies. Journal of Magnetism & Magnetic Materials, 2007; 310(2-part-P3): pp 2725-2727.
[11] Singh, A. K. Uniaxial stress component in tungsten carbide anvil high-pressure x-ray cameras. Journal of Applied Physics. 2003; 45(11): p 4686.
[12] Lin W. Nondestructive Testing Method of Cemented Carbide Anvil Based on Stress Analysis. Cemented Carbide. 2010; 27(2): pp 107-109.
[13] Qiang W, Dongmei C, Dong Z. Structure optimization on carbide anvil of BELT apparatus. Heavy Machinery. 2004; 53(4): pp 53-57.
[14] Wen P. Research on preparation method of fine grained cemented carbide anvil with WC/Co nano composite powder. Superhard Material Engineering. 2017; 29(5): pp 17-24.
[15] Wen P, Tao D. Application of advanced technology of cemented carbide in manufacturing large size anvil. Superhard Material Engineering. 2013; 25(1): pp 25-29.

doi:10.3233/ATDE220240

Strain Engineering on Thermoelectric Properties in 2,7-Dioctyl[1]benzothieno-[3,2-b][1]benzo-thiophene

Ziman WANG [a, b], Ming YANG [a, b] and Hang ZHANG [a, b, 1]
[a] *Institute of Engineering Thermophysics, Chinese Academy of Sciences, Beijing 100190, China.*
[b] *University of Chinese Academy of Sciences, Beijing 100049, China.*

Abstract. 2,7-dioctyl[1]benzothieno[3,2-b][1]benzothiophene (C_8-BTBT) is a kind of organic semiconductor materials with high carrier concentrations. In this work, we reported the thermoelectric performances of C_8-BTBT under isotropic compressive strain of 5%. The results indicate that the figure of merit (zT) for N-type C_8-BTBT could be improved by 17.6% and strain engineering plays an important role in the regulation strategy for thermoelectric materials.

Keywords. Thermoelectric materials; Organic semiconductors; Strain engineering; Thermoelectrical transport coefficients

1. Introduction

Thermoelectric materials, as a kind of important functional materials, can transform the thermal energy in various scenes into electric energy without pollutant emission. And they have broad prospects in thermoelectric power generation and thermoelectric refrigeration, such as recovery and utilization of industrial waste heat, infrared sensing, and space special power supply [1-3]. The energy conversion efficiency for thermoelectric materials is usually measured using zT, which comprising the Seebeck coefficient S, the electrical conductivity σ, the thermal conductivity κ, and the absolute temperature T,

$$zT = \frac{S^2\sigma}{\kappa}T \tag{1}$$

Strain engineering is a usual semiconductor processing technology, which can change the crystal structure of the material by mechanical stretching or compression. Thus, the electron structure and the macroscopic characteristics of the system will also

[1] Hang Zhang, Institute of Engineering Thermophysics, Chinese Academy of Sciences, No.11, North Sihuan West Road, Beijing, China; E-mail: zhanghang@iet.cn.

be influenced. Therefore, strain engineering is expected to be a feasible way to optimize the performance of thermoelectric materials.

Organic semiconductors have significant advantages in the preparation of highly flexible and lightweight electronic devices due to its higher carrier mobility and better transparency [4, 5]. It is not only widely used in field effect transistors, but also considered as a class of thermoelectric materials. High carrier mobilities can improve the conductivities of thermoelectric materials without introducing a high carrier concentration, which is beneficial to its applications in thermoelectricity. According to previous studies, C_8-BTBT ($C_{30}H_{40}S_2$) has great carrier mobility and excellent stability at room temperature, because of its appropriate band structure and excellent hole transport properties [6-8]. Improving the carrier transparent properties of C_8-BTBT from the molecular structure selection and optimization is of great significance to improving its thermoelectric performance [9, 10]. This work analyzes the regulating effect of strain on the electrical and thermal transport characteristics of C_8-BTBT through theoretical calculation, and discusses the strain regulation of its thermoelectric properties. The influence of compression on the thermoelectric characteristic for C_8-BTBT can be understood through analyzing the band structure.

2. Methods

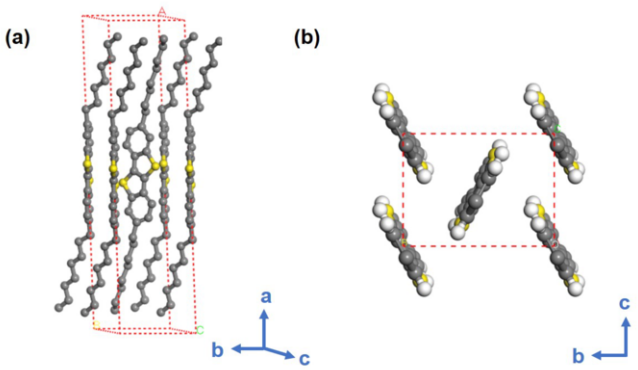


Figure 1. Crystal structure of C_8-BTBT: (a) side view and (b) top view

The structural relaxation and electronic band structure were calculated through the Vienna Ab-initio Simulation Package (VASP) [11]. We set the cutoff energy to 600 eV and chose a 1 × 4 × 4 grid as the k-mesh for structural relaxation. The energy convergence criterion of electronic iteration was set to 10^{-4} eV. The thermoelectric characteristics were obtained through Boltzmann transport theory which was performed by the BoltzTraP code [12].

C_8-BTBT has a monoclinic crystal structure and shows a lamella-like structure (Figure 1(a)). The herringbone positioning for the cores of the structure (BTBT) in the bc plane (Figure 1(b)) is beneficial for two-dimensional charge transport.

3. Results and Discussion

The transport properties of thermoelectric materials are closely related to the band structure and the partial density of states (PDOS), which are shown in Figure 2 and Figure 3. The band structure shows the energies of the lowest and second-lowest conduction bands (CB) decrease from 0% to 5% compressive strain, which results in bandgap narrowing under compression. Furthermore, the Fermi level increases as the compressive strain increases, which means the compression tend to cause the overall electron level to increase, and the highest energy level occupied by electrons will tend to increase with the overall electron level.

By comparing the PDOS under strain (Figure 3), compressive strain leads to a decline of the CB minimum. The effect of strain on the bottom of the CB is greater than that on the top of the valence band (VB). The major contribution to the VB is from C 2p state and S 3p state, and that to the CB is from C 2p state and a little contribution from S 3p state is appeared.

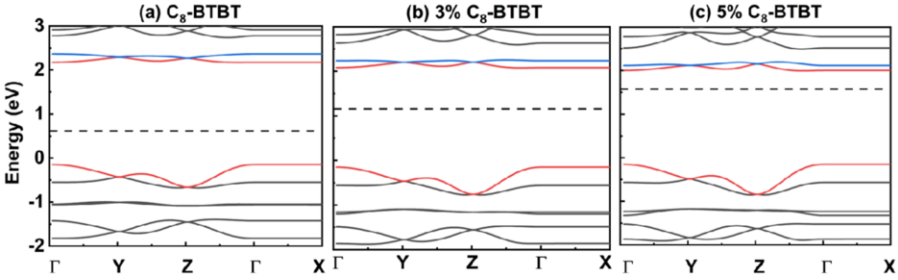


Figure 2. Band structures of C_8-BTBT under (a) 0%, (b) 3% and (c) 5% compressive strains. The red lines represent the lowest CB and highest VB and blue lines represent the second-lowest CB. The dashed lines represent the Fermi levels

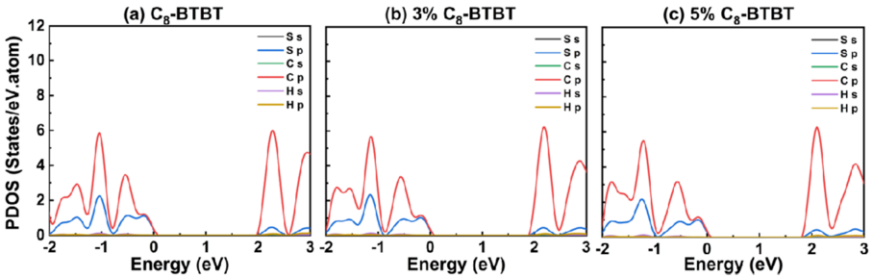


Figure 3. PDOS of C_8-BTBT under (a) 0%, (b) 3% and (c) 5% compressive strains

The analysis of transport coefficients of C_8-BTBT demonstrates an optimization strategy based on strains for practical application. Figure 4 provides the relationships among the thermoelectric parameters and chemical potential under compressive strain (5%) at room temperature (300 K). The positive or negative chemical potentials represent the major carriers are the hole (P-type) or the electron (N-type). And the following calculations were obtained based on the experimental relaxation time (19.9 fs) according to the literature [13].

The chemical potential dependences of the thermoelectric parameters, including S, σ, κ and zT, at 300 K are illustrated in Figure 4(a)–(d). Seebeck coefficient is the internal voltage response to the temperature gradient ($S = \Delta U/\Delta T$) in thermoelectric materials, which is an important parameter related to the electronic structure. The

heated end carriers migrate from the hot end to the cold end, and the formation of electric current is due to the directional movement of the carriers. The S exhibits two distinct peaks on both sides of the Fermi level for P-/N-type semiconductors. The variation of the peak values of S under 1% to 5% compressive strains is small. The increases of the highest S are 0.84% and 0.47% in positive and negative chemical potentials. It should be noted that the critical points of S at around $\pm$1.25 eV. C_8-BTBT presents a large S within the critical points, and beyond these points S basically disappears.

The electrical conductivity ($\sigma = ne\eta$) depends on the charge carrier concentration (n) and its mobility (η). Therefore, a material with a high carrier concentration and mobility is required to obtain an excellent σ. Furthermore, the κ is the thermal conductivity which including the electric thermal conductivity κ_e and the lattice thermal conductivity κ_l. The calculations of BoltzTraP are only for the electric thermal conductivity. Here, we use the κ_l = 0.2 W/m·K, which is cited from Shi et al. [1]. For σ and κ, it is clear that the values both increase with the increasing of strains. The greatest improvements for maximum σ values under 5% compressive strain are 52.9% (from 12.1 × 10^2/Ω·cm to 18.5 × 10^2/Ω·cm) and 54.1% (from 3.7 × 10^2/Ω·cm to 5.7 × 10^2/Ω·cm) for P- and N-type semiconductors. As for κ, these values are 43.0 % (from 10.0 × 10^2/Ω·cm to 14.3 × 10^2/Ω·cm) and 23.5% (from 3.4 × 10^{-3} W/cm·K to 4.2 × 10^{-3} W/cm·K). It can be inferred that C_8-BTBT exhibits higher σ under compressive strain, that is due to the fact that C_8-BTBT possesses a smaller bandgap in that condition.

In Figure 4(d), we illustrated the chemical potential dependences of zT for C_8-BTBT under 1% to 5% compressive strains at 300 K. The highest peak of zT for P-type C_8-BTBT is at around -1.25 eV and for N-type C_8-BTBT is at around 1.0 eV. In the inset of Figure 4(d), we can clearly observe that the maximum of zT value slightly decreases (from 0.33 to 0.32) as the compressive strain increases for P-type C_8-BTBT. While for N-type C_8-BTBT, this value increases significantly as the compressive strain increases. The largest increment and increasing percentage of zT appear in C_8-BTBT under a 5% compressive strain, which is 0.06 (17.6%) for N-type, and the maximum of zT value reaches 0.4, which means that N-type C_8-BTBT could be the candidate organic thermoelectric material under compressive strains.

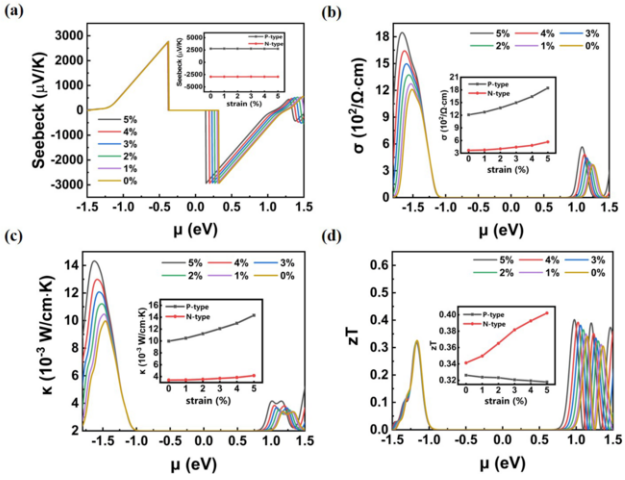


Figure 4. The (a) S, (b) σ, (c) κ and (d) zT under compressive strains for C_8-BTBT. The insets show the strain dependence of the thermoelectric properties.

4. Conclusion

We have investigated the effect of chemical potential and compressive strain on the thermoelectric properties of C_8-BTBT at room temperature. The results reveal that the Fermi level rises and the bandgap narrows under compressive strain. According to the results, the maximum value of zT for N-type C_8-BTBT reaches 0.4 under 5% compressive strains, which is improved by 17.6% compared with the original structure. This study demonstrates that strain regulation is a feasible pathway to enhance the transport properties for C_8-BTBT organic semiconductors, which is important for realizing improved thermoelectric materials.

Acknowledgments

This work was supported by the Basic Science Center Program for Ordered Energy Conversion of the National Natural Science Foundation of China (No. 51888103, 51606192) and the CAS Pioneer Hundred Talents Program.

References

[1] Ssennoga T, Jie Z, Yuying Y, Bo L. A comprehensive review of thermoelectric technology: Materials, applications, modelling and performance improvement. Renewable & Sustainable Energy Reviews. 2016 Jul; 65: pp. 698-726.

[2] Francis J. DS. Thermoelectric Cooling and Power Generation. Science. 1999 Jul; 285(5248): pp. 703-706.

[3] Chunxiao Z, Chao S, Shen W, Yuan W, Guoquan L, Cheng S. A Review on Recent Development of Cooling Technologies for Photovoltaic Modules. Journal of Thermal Science. 2020 Nov; 29(6): pp. 1410-1430.

[4] Daowei H, Jingsi Q, Linglong Z, Junya W, Tu L, Jun Q, Yun L, Yi S, Yang C, Wei L, Luis K. Ono, Yabing Q, Jianbin X, Wei J, Xinran W. Ultrahigh mobility and efficient charge injection in monolayer organic thin-film transistors on boron nitride. SCIENCE ADVANCES. 2017 Sept;3:e1701186.

[5] Michael JF, John GL, Ming W, Hengbin W, Thuc QN, Guillermo CB. Carrier‐Selective Traps: A New Approach for Fabricating Circuit Elements with Ambipolar Organic Semiconductors. Advanced Electronic Materials. 2017; 3(3).

[6] Takafumi, Eigo M, Kazuo T. Molecular Ordering of High-Performance Soluble Molecular Semiconductors and Re-evaluation of Their Field-Effect Transistor Characteristics. Advanced Materials. 2008; 20(18): pp. 3388-3392.

[7] Shitan W, Dongmei N, Lu L, Yingbao H, Xuhui W, Can W, Haipeng X, Yongli G. Interface electronic structure and morphology of 2, 7-dioctyl[1]benzothieno[3,2-b]benzothiophene (C8-BTBT) on Au film. Applied Surface Science. 2017 Apr; 416: pp. 696-703.

[8] Yongbo Y, Gaurav G, Alexander L. A, Arjan P. Z, Stefan C. B. M, Jihua C, Dennis N, Michael F. T, Jinsong H, Zhenan B. Ultra-high mobility transparent organic thin film transistors grown by an off-centre spin-coating method. Nat Commun. 2014; 5: p. 3005.

[9] Guillaume S, Vincent L, Claude N, Christian R, Ying D, Osamu G, Wen YL, Yeongin K, Jean BA, Jolanta K, Alan RK, Sean RP, Yoann O, Stefan C. B. M, Jérôme C, Yves H. G, Zhenan B. Bulky End‐Capped [1]Benzothieno[3,2‐b]benzothiophenes: Reaching High‐Mobility Organic Semiconductors by Fine Tuning of the Crystalline Solid‐State Order. Advanced Materials. 2015; 27(19): pp. 3066-3072.

[10] Yingfeng W, Sufen Z, Jianhua G, Huarong Z, Guoqiao L, Chengdong Y, Hui X, Renren F, Hongxiang L, Wenping H. High-performance organic field-effect transistors based on single-crystalline microribbons of a two-dimensional fused heteroarene semiconductor. Chemical Communications. 2015; 51(60): pp. 11961-11963.

[11] G. Kresse, J. Furthmuller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Physical Review B. 1996 Oct; 54(16): pp. 11169-11186.

[12] Georg K. H. M, David J. S. BoltraP. A code for calculating band-structure dependent quantities. Computer Physics Communications. 2006 Jul 1; 175(1): pp. 67-71.
[13] Wen S, Jianming C, Jinyang X, Dong W, Zhigang S. Search for Organic Thermoelectric Materials with High Mobility: The Case of 2, 7-Dialkyl[1]benzothieno[3,2-b][1]benzothiophene Derivatives. Chemistry of Materials. 2014 Mar; 26(8): pp. 2669-2677.

doi:10.3233/ATDE220241

The Effect of Deposition Current on Structure and Properties of Ag Coatings Prepared by Middle Frequency Magnetron Sputtering

Junwei CHEN [a], Bo LI [a], Yan LIU [d], Zhuoyi LIU[b], Jian YAO [a], Quan HU [a], Wei LIU[a], Tianjing SHAO[c], Ruijing YANG[c], Bin WANG[c], Bing YANG[d,1], Zhengang LI[d]
[a] *Electric Power Research Institute of Guizhou Power Grid Co., Ltd., Guiyang 550002, Guizhou*
[b] *Guizhou Power Grid Co., Ltd., Guiyang 550001, Guizhou*
[c] *Xingyi Power Supply Bureau of Guizhou Power Grid Co., Ltd., Guiyang, Guizhou 562400)1. Electric Power Research Institute of Guizhou Power Grid Co., Ltd, Guiyang, 550000*
[d] *Department of Materials, School of power and mechanical engineering, Wuhan University, Wuhan, 430072*

Abstract. Silver coatings were prepared by middle frequency (MF) magnetron sputtering with different deposition current. The effects of currents on the microstructure were investigated by using scanning electron microscopy (SEM), atomic force microscopy (AFM) and X-ray diffraction (XRD). The mechanical properties and corrosion resistance were studied by using nanoindentation and electrochemical workstation. The results showed that the roughness and thickness of coatings rise obviously with increasing current. The XRD patterns indicated that the silver coating exhibit a strong Ag (111) preferred orientation and the grain size increased from 61.6 to 68.4 nm with incremental current from 1 A to 3 A. The deposition current also exhibited obvious influence on the mechanical properties and chemical corrosion resistance of the silver coatings, which decrease the hardness and elastic modulus of the coating. At deposition current with 1 A, the coating shows highest elastic modulus and hardness and of 218 GPa and 2.1GPa, respectively, as well as the lowest corrosion current density of 0.26 A•cm2 in the 3.5wt% NaCl solution.

Keywords. Silver coating, magnetron sputtering, Deposition Current, Microstructure, Properties, Corrosion resistance

1. Introduction

High temperature alloys such as super-alloys are usually used as the mechanical components like high strength bolts in aerospace. However, these materials usually exhibit high friction coefficient and adhesion behavior during their service process in

[1] Corresponding Author, Bing Yang, Department of Materials, School of power and mechanical engineering, Wuhan University, Wuhan, 430072; E-mail: toyangbing@whu.edu.cn.

high temperature and vibration environment. To avoid such phenomena, the components are often coated with a soft coating like Ag or gold to act as a lubricant materials[1]. Electroless plating is the wildest used method to deposit silver or gold film. However, the solution used in such process contains cyanide, which was harmful to the environment and limits the applications of the soft metal coating. Magnetron sputtering technology is an environment-friendly technique to produce industrial scale coatings and has been used to fabricate silver lubricant coatings recent years[2]. Wang et al. investigated the influence of sputtering power on the resistance of Ag coating by magnetron sputtering technology. The results exhibited that the resistance of coating decreased with increasing sputtering power[3]. P. Asanithi et al investigated the effect of the distance between the substrate and the target on the grain size of silver coatings, and obtained the optimum nanostructured silver coating, which shows an average crystal size of 5.9 nm at the fixed distance of target-substrate at 10 cm[4]. P. Dutheil et al probed the influence of the substrates on the wettability of the silver coating and found that the coating deposited on Si wafer was hydrophobic, while on the W plate was hydrophilic[5]. Although lots of studies of the silver coatings deposited using magnetron sputtering have been reported, there are seldom literatures focused on the effects of the target current on the structure as well as the performance of silver coatings.

In this paper, silver coatings were deposited using MF magnetron sputtering technology under different current conditions, and the microstructure, mechanical properties and chemical corrosion resistance were investigated, systematically.

2. Experimental details

The silver coatings were deposited onto silicon, carbides and stainless steels substrates, using a homemade magnetron sputtering system, which was equipped with an Ag target (Φ70mm). Before deposition, all substrates were cleaned automatically in alcohol and acetone for 10 min, respectively. The samples were fixed on the sample holder and the distance of the substrate to the target was fixed at 15 mm. The base pressure was pumped down to 7×10-3 Pa, and then the temperature of the substrates were heated up to 300 ℃. Before deposition, the argon bombardment was applied to clean the substrates under a bias voltage of 800 V for 30 minutes. Subsequently, the silver coatings were deposited onto substrates at a bias voltage of 50 V and 0.5 Pa in Ar ambient for 60 min. The current was set to 1A, 1.5A, 2A, 2.5A and 3A, respectively.

3. Characterization

The morphologies of the coatings were analyzed by a field emission scanning electron microscope (MIRA 3 LMH). An atomic force microscope (Shimadzu spm-9500j3) was used to investigate the surface morphologies and roughness of the coatings, and the analyze area was 10×10 μm2; The structures of the coatings were measured by an X-ray diffractometer (XPert Pro) with Cu Kα radiation. The nanohardness and the elastic modulus of the Ag coatings were tested using a MTS G200 nano indenter. The depth of indentation was controlled less than 10% of the coating thickness to prevent the effects of the substrates. The chemical corrosion resistance of the coatings was investigated by

testing the Tafel curves of the silver coatings. The experiment were conducted in the 3.5 wt.% NaCl solutions by the three-electrode method. The Pt and saturated calomel acted as the counter electrode and indication electrode, respectively. The silver coating was the working electrode and the tested area was designed as 1×1cm2.

4. Results and discussions

4.1. Surface and cross-section morphologies

Figure 1(a)-(e) shows the surface morphologies of silver coatings deposited at varied deposition current. It was observed that the current has great influence on the surface morphologies. At 1A, the coating exhibited a smooth surface with small Ag droplets contaminations. With increasing current, the number and size of the droplets increased. Which is believed to relate to the high target temperature of the target caused by high power density. Fig.1 (f) showed the size of the particles on the coating surface rise from 290 to 1461 nm with the current increased from 1A to 3A. Fig.2 is the top-view morphologies and the roughness of the silver coatings. As shown in surface morphologies of the coatings, it exhibits a hill-like shape, which was consistent with the SEM observation. As the current increase from 1 A to 2.5 A, the average surface roughness (Ra) of the coating increases from 13.7 nm and17.4 nm and the root mean square roughness (RMS) and increase from 36.3 nm to 45.7 nm, respectively. When the current increased to 3 A, the RMS and Ra of the coating significantly increased to 168.8 nm and 121.3 nm respectively. The augment in roughness can be attributed to that the particles contamination at higher target power density with increasing current, which enhancing the bombardment and etching effect of the particles on the coating surface.

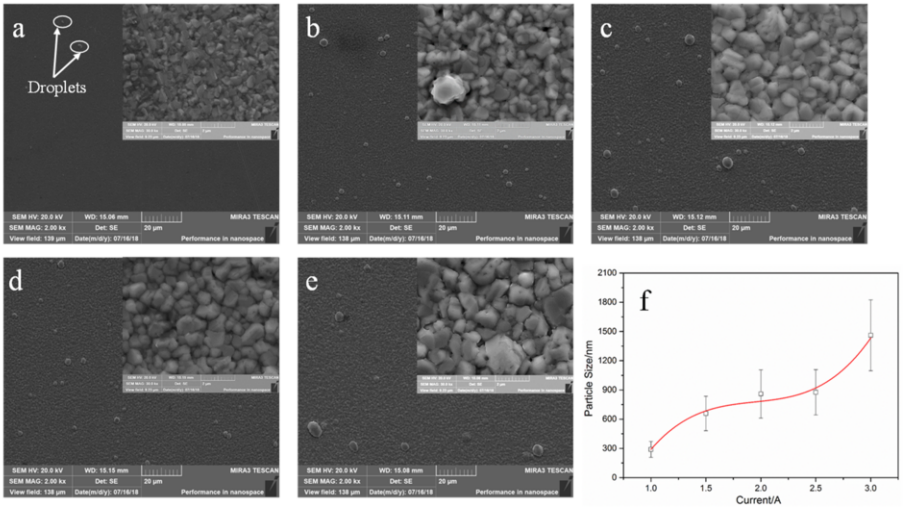


Figure 1. The surface morphologies ((a) 1A, (b) 1.5A, (c) 2A, (d) 2.5A, (e) 3A) and the particle size (f) of the silver coatings deposited at varied current

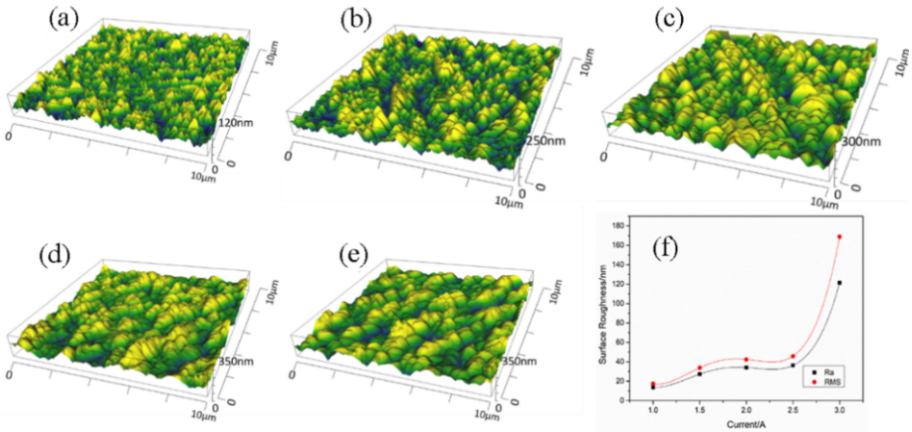


Figure 2. The top surface morphologies ((a) to (d) is 1A, 1.5A, 2A, 2.5A and 3A, respectivelly) and the roughness (f) of the Ag coatings deposited at varied current

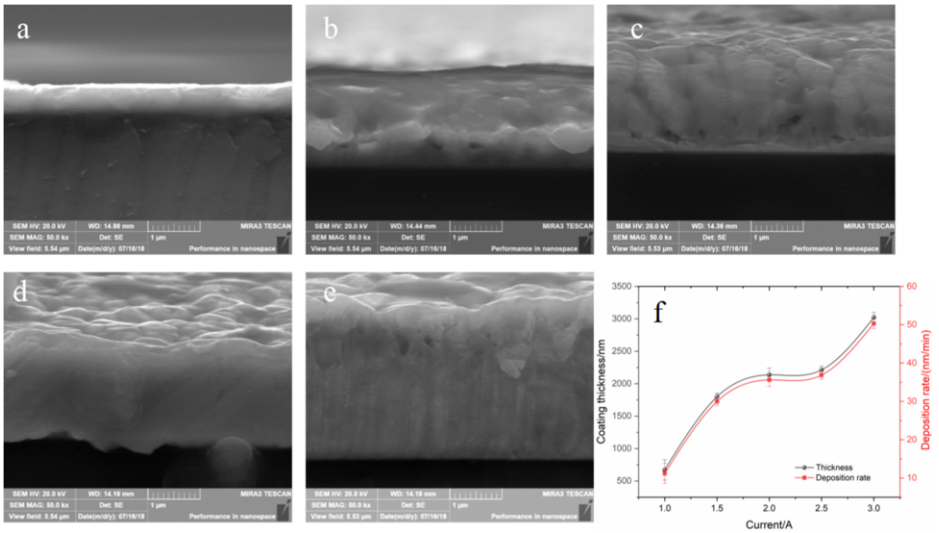


Figure 3. The cross-sectional morphologies of the coating ((a) 1 A, (b) 1.5 A, (c) 2 A, (d) 2.5 A, (e) 3 A) and the thickness and the deposition rate (f) of the silver coatings

The cross-section morphologies were displayed in Fig.3, indicating that the current exhibited a significant influence on the thickness of the silver coating. Calculated from the Fig.3 (a) to (e), the deposition rate and the thickness of the coatings was presented in Fig.3 (f). It was observed that the thickness and deposited rate of the coating increased from 673 to 3019 nm and 11.2 to 50.3 nm•min^{-1}, respectively. During the deposition process, the argon ions with high energy bombarded the Ag target, which caused the silver particles to escape from the target and deposited onto the substrates. The increase in current ease to generate more silver particles, resulting in an improved deposition rate[6].

4.2. Structures

The XRD patterns of the silver coatings were showed in Figure.4 (a). It was observed that Ag coating shows stronger (111) preferred orientation at deposition current of 1 A and relatively more diffraction peaks of (200), (220), (311) and (400) orientation with increasing target current. The preferred orientation was determined by the overall energy of the as-deposited film consisting of surface energy and strain energy. The strain energy is reported to be portioned to the thickness of the film. At 1 A, the Ag film exhibits a small thickness, therefore the overall energy of the film was controlled by the surface energy. For Ag coating, the (111) plane has the lowest surface energy [6], which will be favorable to form the Ag (111) preferred orientation. As the current increased, the energy obtained by the substrate will be augmented, and the coating system has enough energy to support the diffusion of Ag atoms to the high-energy plane. The grain size of the (111) plane was calculated from the XRD patterns, and the result was showed in Fig.4 (b). It was shown that the grain size of the Ag coating increased from 61.6 to 68.4 nm with the increasing current from 1 A to 3 A. The increase in grain size could be attributed to the enhanced current, which can provided more energy for the growth of the coatings.

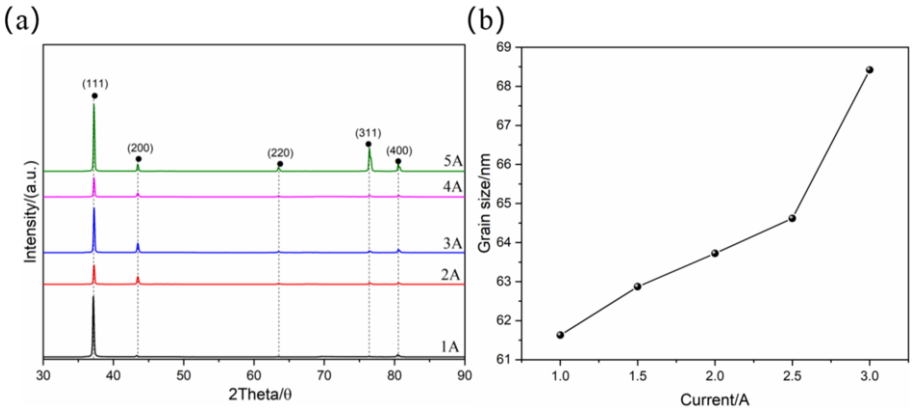


Figure 4. The XRD patterns (a) and the grain size (b) of the Ag coatings sputtered at varied current

4.3. Hardness and toughness of coating

Figure.5 (a) shows the hardness and modulus of the Ag coatings under different current conditions. It was observed that the hardness and modulus of the coating decreased with rising current. As the current rises from 1 A to 3 A, the hardness and elastic modulus of the coatings decreased from 2.1GPa and 218 GPa to 1.1 GPa and 96 GPa, respectively. The reducing hardness could be caused by the Hall-Petch effect. According to the Hall-Petch relationship, the grain refinement could strengthen the hardness of the materials [7]. Therefore, the coating deposited at high current will result in reducing grain size and exhibits higher hardness and modulus. The value of the ratio for H^3/E^2 could be used to estimate the toughness of the coatings. The high the ratio for H^3/E^2, the tough the coating will be obtained. The values of H^3/E^2 of the Ag coatings were exhibited in Fig.5 (b). It was observed that the H^3/E^2 of the coatings ranged from 1.75×10^{-4} to 2.6×10^{-4} GPa, and the coating deposited at 1A exhibited the highest H^3/E^2.

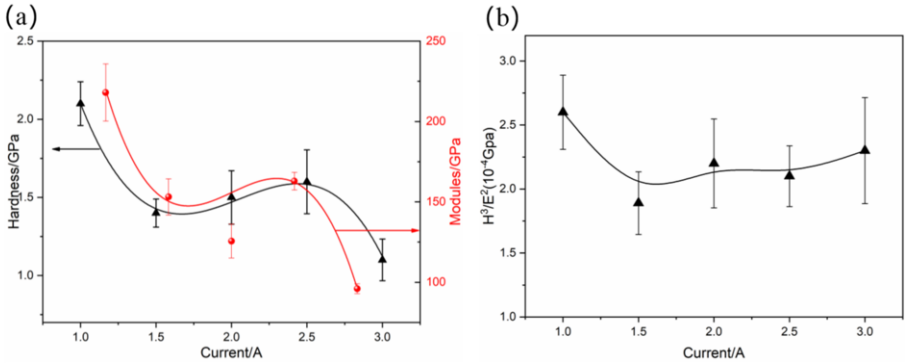


Figure 5. The elastic modulus and nanohardness (a) and H^3/E^2 (b) of the Ag coatings deposited at different current

4.4. Corrosion performance of the coatings

Figure 6 (a) shows the Tafel plots of the silver coatings in 3.5 wt% NaCl solution. Passivation and pitting behavior were observed in the anode zone for the coatings deposited at 1 A, 1.5 A, 2 A and 2.5 A. While for the coating deposited at 3 A had no passivation process, and it directly dissolved in the NaCl solution. The corrosion current density was calculated from the Tafel plots using the epitaxy method and the results were revealed in Fig.6 (b). The results revealed that the current change from 1 A to 2.5 A, the corrosion current density raised slightly from 0.26 to 0.33 A•cm^2. The corrosion current density increased sharply to 0.99 A•cm^2 at 3A, indicating that when the current was lower than 3A, the effect of the current on the corrosion performance was not significant, while further increase in current, the corrosion resistance of the silver coating would be obviously reduced. This may lead to the increasing grain size with the growing current and reduce the number of active atoms on the coating surface, which will weaken the ability of the Ag coating to form a passivation film. As a result, the corrosion resistance of the coating drops. Moreover, the rougher surface with increasing current also degraded the corrosion resistance. It was reported that the higher surface roughness will lead to the non-uniform Volta potential, which could cause micro-galvanic corrosion. The rougher surface provided more localized corrosion sites, and consequently accelerated the corrosion rate of the coating.

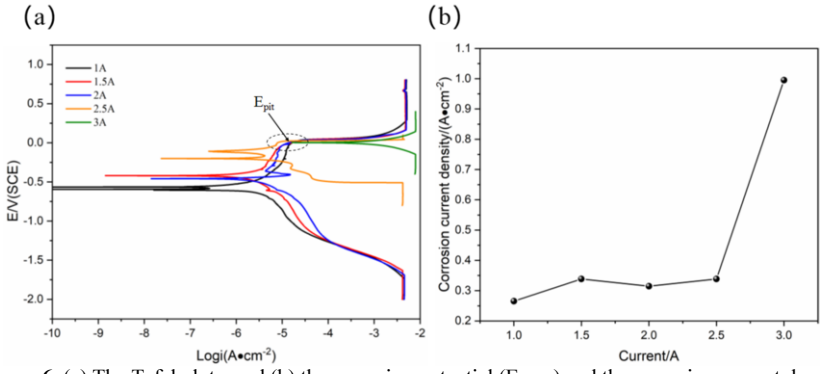


Figure 6. (a) The Tafel plots, and (b) the corrosion potential (Ecorr) and the corrosion current density of the silver coatings deposited at different current

5. Conclusion

Silver coatings were deposited at different current using MF magnetron sputtering, and the influence of the target current on structure and properties of the coatings were investigated systematically. The results showed that the current could significantly affect the surface morphologies and deposition rate of the coatings. The increasing current caused a rougher surface morphologies and a high deposition rate. As the current increased from 1 A to 3 A, the Ra, RMS and deposition rate of the coating increased from13.7 to 121.3 nm, 17.4 to168.8 nm and 11.2 to 50.3 $nm \cdot min^{-1}$. The silver coating deposited at 1 A exhibited a strong Ag(111) preferred orientation, and the diffraction peak intensities of (200), (220), (311) and (400) crystal planes were enhanced with the increasing current. The grain size of coating increased from 61.6 to 68.4 nm with the current rising from 1A to 3A. The current also exhibited obvious effect on the mechanical properties and chemical corrosion resistance of the silver coatings. The hardness and elastic modulus of the coatings decreased from 2.1GPa and 218 GPa to 1.1 GPa and 96 GPa, respectively, from the current increasing from 1A to 3A. The H^3/ E^2 of the coatings ranged from 1.75×10^{-4} to 2.6×10^{-4} GPa, and the coating at 1 A exhibited the highest H^3/ E^2 value. The corrosion resistance of the coatings degraded as the raising current, and the corrosion current density increased from 0.26 to 0.99 $A \cdot cm^2$ when the current changed from 1 A to 3 A. These results indicated that 1A was the best deposition parameter to obtain the coating with the best mechanical properties and corrosion resistance.

References

[1] Tronci G, Marshall MB. Understanding the behavior of silver as a low friction coating in aerospace fasteners, Tribol. Int. 100(2016): pp 162–170.
[2] Safi I. Recent aspects concerning DC reactive magnetron sputtering of thin films: A review, Surf. Coatings Technol. 127(2000): pp 203–218.
[3] Linwen W, Ling L, Weidong C. INVESTIGATION of the PROPERTIES of SILVER THIN FILMS DEPOSITED by DC MAGNETRON SPUTTERING, Surf. Rev. Lett. 24 (2017).
[4] Asanithi, PS. Chaiyakun, Limsuwan P. Growth of silver nanoparticles by DC magnetron sputtering. J. Nanomater. 20122(2012): pp 1-8.
[5] Dutheil P, Thomann AL, Lecas T, Brault P, Vayer M. Sputtered Ag thin films with modified morphologies: Influence on wetting property, Appl. Surf. Sci. 347 (2015): pp 101–108.
[6] Dushman S. Thermionic Emission. Rev. Mod. Phys. 2 (1930): pp 381–476.
[7] Hall BEO. The Deformation and Ageing of Mild Steel: III Discussion of Results, Proc. Phys. Soc. Sect. B. 64 (1951): p 747.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 307
J. Xu et al. (Eds.)

doi:10.3233/ATDE220242

Research Status of Wear Resistance Enhancement of Fan Impeller Surface

Xue HU[a], Yulong CHEN [a, 1], Lixin ZHANG[a], Shengli ZHANG[b] and Feng DONG[b]
[a] *School of Mechanical and Electrical Engineering, Shihezi University,832003,China*
[b] *Shihezi Shengli Hard surface Engineering Technology Co., LTD,832000,China*

Abstract. Centrifugal fans have poor operating conditions, and the problem of wear failure has been around for a long time, and the blades are the most severely affected parts of wear. This article discussed the wear mechanism of fan blades from the perspectives of erosion wear and gas-solid two-phase flow field, and then summarizes the blades including wear-resistance strengthening measures with active anti-wear and passive anti-wear.It is necessary to combine the microstructure of the material to form a complete erosion and wear mechanism; in the gas-solid two-phase flow field analysis, it is necessary to strengthen the numerical simulation and experimental analysis. Further research will be conducted on the viscous flow, abrasive particles of the gas and the direction of secondary flow, etc., to further improve the prediction accuracy. When selecting the method of strengthening the wear resistance of the fan blades, the structure of the centrifugal fan, the gas-solid two-phase flow field and the actual operating conditions should be considered comprehensively. Analyzed reasonably to achieve appropriate economic benefits

Keywords. Centrifugal fan; Impeller blade; Wear mechanism; Wear-resistant strengthen

1. Introduction

In electricity, mines, tunnels, metallurgy and other fields, there is a high concentration of dust in the working environment. Currently the most popular use for the demand to reduce the concentration through air discharge is centrifugal fans. The main structure of the centrifugal fan is shown in Figure 1, including the volute and the impeller, of which the impeller is the most critical component, consisting of a curved front disk, a blade and a flat rear dsk[1].

[1] Yulong CHEN: School of Mechanical and Electrical Engineering, Shihezi University, 314991798@qq.com

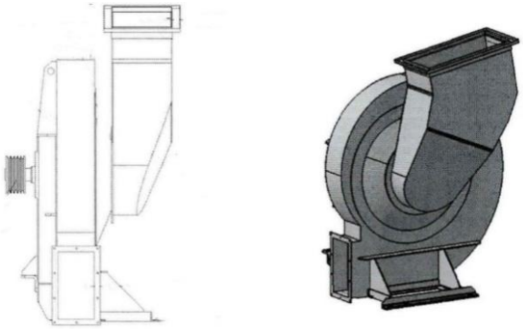


Figure 1. Centrifugal fan

Although there is a protective wear layer on the outer surface of the fans, due to the poor operating conditions of the centrifugal fans and the gas-solid two-phase flow, the impact wear of solid particles in the gas on the fan is still very serious, and even causes structural imbalance and other phenomena. Among them, the impeller is most affected. The part of the impeller which is most affected by wear is the blade working surface at the entrance and exit of the impeller. At the exit, the blade becomes thin due to the wear, and even appears jagged. Due to abrasion in the inlet vane will produce gap, the resulting centrifugal fan down the increase of the frequency and the maintenance cost, and reduce use frequency and transmission efficiency[2,3]. Therefore, the wear mechanism of the impeller was clarified through the study of wear resistance enhancement on the surface of the fan impeller, and appropriate wear resistance enhancement measures are selected to effectively improve the service life of the centrifugal fan.

2. Erosion wear mechanism of blade

The phenomenon of wear is usually the result of several wear mechanisms. The process is very complicated and there are many forms of wear. At present, many researchers generally believe that erosion wear is the main wear mechanism of fan blades. Erosion wear is caused by abrasive particles with a diameter of less than 1mm hitting the target surface at a speed of less than 550m/s and at a certain angle. In-depth study on the mechanism of erosion wear is helpful for solving the surface failure phenomenon caused by wear of fan blades [2,4].

Finnie [5] first put forward the micro cutting theory of plastic materials, which pointed out that the material on the surface of the object would be cut off by abrasive particles like a sharp knife, resulting in wear failure phenomenon. Assuming that the surface of the impact target is a polygonal abrasive particle with a mass of m, incident angle α and velocity v, the erosion wear quantity V changes with the change of incident angle α through theoretical analysis, and the expression is as follows:

$$V = K\frac{mv^2}{p}f(\alpha) \tag{1}$$

Where, p is the flow stress of the target material and K is a constant.

This model can better interpret the wear law of plastic materials with the impact incident angle of polygonal abrasive less than 16.84°. However, the error of erosion

wear quantity V can be large when the impact incident Angle of non-polygonal abrasive is large, brittle materials and impact incident Angle are large.

Tilly[6] proposed that the occurrence process of Erosion wear was not only related to Primary Erosion wear (Primary Erosion wear), but also caused Secondary Erosion wear (Secondary Erosion wear) due to the surface wear of target material brought by Primary Erosion wear, especially the fragmentation of abrasive particles during Primary Erosion. The model proposed for the first time that wear did not occur when the abrasive particle diameter was smaller than the critical value.

Hutchings[7] proposed that when rigid body abrasive particles collided with target surface, the volume of indentation was related to the compressive constant of target surface and the kinetic energy of rigid body abrasive particles. At the same time, Hutchings proposed that the temperature of the target material rised rapidly under high strain rate. In this theory, the critical value of deformation was used as the measurement index of material properties for the first time, which was determined by the microstructure of the material, providing a new idea and theoretical basis for the study of wear resistance mechanism.

3. Wear mechanism of gas-solid two-phase flow field

The gas-solid two-phase flow field is a macroscopic analysis of the wear mechanism of the fan impeller, and the flow track and motion characteristics of the gas-solid two-phase flow field in the fan are studied based on fluid dynamics. At the same time, numerical simulation and experimental analysis are used to obtain the impact Angle of the abrasive particles and the distribution of the velocity in time and space.

Lin et al. [8] proposed for the first time the wear mechanism of centrifugal fan impeller, and the effect of abrasion was the same as erosion wear. The model uses the finite element method to solve the invistic model as the airflow field in the inlet area of the fan impeller, and uses the quasi-three-dimensional streamline curvature method to solve the airflow field in the fan impeller passage. Then, the concept of "computational particle". Taba-Koff and some erosion models are used to calculate the motion trajectory and wear amount of abrasive particles. Fig. 2 shows the movement path of abrasive particles with different diameters after entering the impeller, and Fig. 3 shows the erosion wear of blades.

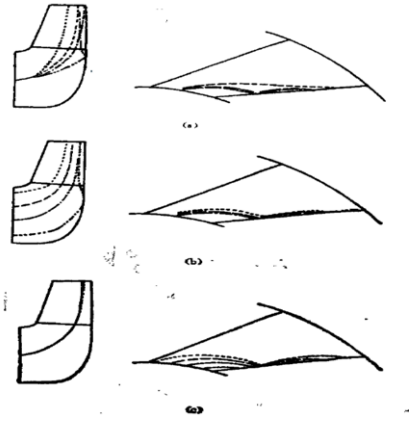


Figure 2. Movement path of abrasive particles

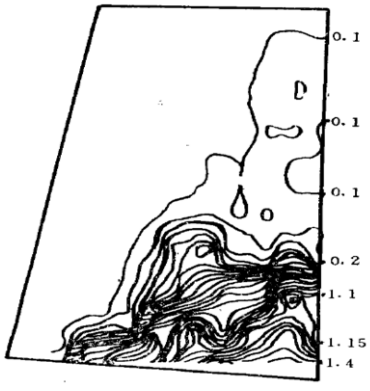


Figure 3. Pressure erosion wear pattern

The model explained well that the wear of fan impeller was mainly composed of abrasion and erosion wear. The abrasion is concentrated in the tail of blade, and erosion wear is concentrated in the head of blade. In the preparation of fan impeller, in order to achieve wear-resisting effect, the blade can be rationally and scientifically arranged rib to reduce abrasive wear.

Meidan et al. [9] realized the visualization of the wear degree of the impeller. Fig. 4 shows the visible view of the wear rate of the back cover of the impeller by abrasive particles, and Fig. 5 shows the visible view of the wear rate of the front disk of the impeller by abrasive particles. By using numerical simulation, Reynolds stress model, particle track model and Taba-koff & Grant wear model, the different wear phenomena of different diameters on different parts of impeller were explained and verified by experiments.

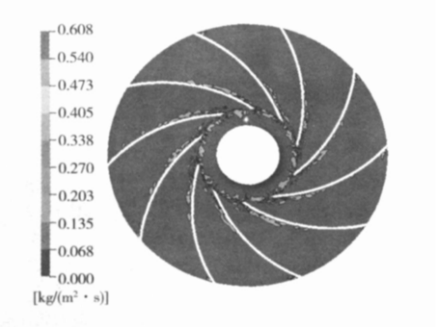


Figure 4. Wear rate of rear cover of impeller

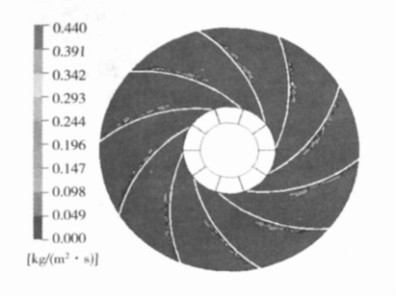


Figure 5. Wear rate of impeller front disk

The model reveald that the pressure surface was the most seriously worn part of the fan impeller, and the pressure surface near the outlet end was the most seriously affected by the wear. The blade wear at the inlet was mainly caused by impact wear, while the blade wear at the outlet is mainly caused by abrasion wear, and was mainly caused by small abrasive particles.

Luo et al. [10] used the experimental bench as shown in Fig. 6 to prepare the impeller from plexiglass and used tobacco seeds as abrasive particles to study the gas-solid two-phase flow field inside the centrifugal fan impeller through experiments. The research conclusioned that the main reason for the impact wear of the fan impeller was the radial velocity of the particles at the impeller entrance.

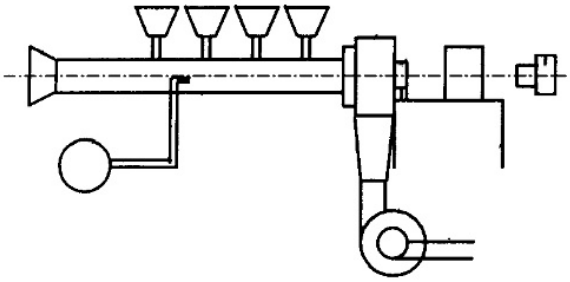


Figure 6. Schematic diagram of the experimental platform

The experiment showed that by changing the flow path profile of the fan impeller, the inlet impact Angle of abrasive particles can be reduced, the gas-solid two-phase flow field can be improved and the radial velocity of the solid phase can be increased, which can benefit for reducing the abrasion wear and erosion wear at the impeller

entrance, and improving the wear resistance of the fan impeller.

4. Wear resistance strengthening measures

While studying the wear mechanism of centrifugal fan impeller, researchers have also done a lot of effort on the wear resistance enhancement of fan impeller, and achieved a series of achievements. At present, passive anti-wear technologies are the most widely used, such as modification of blade materials and adding wear-resistant materials on blade surfaces. The other part is combining the characteristics of fluid mechanics, adopt active anti-wear technology, such as changing the blade structure, front anti-wear blade and improving the aerodynamic design. The wear resistance of centrifugal fan impeller can be effectively improved by adopting the above wear-resisting strengthening technology, and the service life of centrifugal fan can be extended.

4.1. Passive anti-wear technology

(1) Wear resistant material should be added to the surface of the blade

① Surfacing of wear resistant layer

Liang [11] found that the part of centrifugal fan impeller, which was most seriously affected by wear ,was the stagnation point of airflow at the blade head, especially the part from the inlet to the wheel, mainly caused by the cutting effect of abrasive particles and the impact on the target surface. The treatment method was to first use the welding rod in the blade head at the stationary point of the air flow for surfacing treatment, then use 16Mn liner from the surfacing to the outlet for welding treatment, and finally use the same welding rod on all the lining plate for surfacing again, at this time, the surfacing of the corrugated single arc center Angle of 60°, the height of 3-4mm corrugated. Through the above operation, the impact Angle of abrasive particles almost appeared around the wave valley of corrugated surfacing welding. At the same time, the airflow whirling in the middle of the two piles of corrugated surfacing welding, played the role of air cushion for abrasive particles and effectively reduces the wear degree of abrasive particles.

② Wear resistant ceramics

Zhang [12] inlaid a layer of wear-resistant ceramics on the surface of the centrifugal fan blade. Through practical observation, it was found that under normal operating conditions, the wear-resistant ceramics and the base material of the blade were firmly combined, the annual wear was less than 0.2mm, and the wear resistance was 30 times more than that of the base material of the blade 16Mn, which could improve the working efficiency of the centrifugal fan by about 20%. At the same time, it can also meet the needs of the fan in the high temperature environment of not more than 500℃, increasing the use of wear-resistant ceramics in cement plants and other fields increasingly popular.

③ Wear-resistant ribs

Lin et al. [13] pointed out that the turbulence of the flow field around the fan blade was conducive to reducing the influence of abrasive particles on blade wear. The effective way to promote the turbulence of the flow field was to add wear-resistant ribs on the blade surface.

④ The coating

Sun et al. [14] used laser cladding technology to cladd a layer of TiC/Co coating on the impeller surface, and proved through experiments that when the TiC content in the TiC/Co coating was 20%, the friction factor was low with little fluctuation, and the impeller surface had the best wear resistance and surface condition.

Shi et al. [15] applied a high-speed arc spraying system to spray a layer of SR-MI coating on the surface of fan impeller. The coating thickness was 0.7±0.1mm, the bonding strength was greater than 35MPa, and the hardness was greater than HRC50.

(2) Modification of leaf materials

Jia [16] adopted b-V-RE multiple co-osmotic technology to solve the problem of blade wear. After 8h co-osmotic treatment at 950℃, the self-developed multiple co-osmotic agents can obtain blades with surface hardness up to Hv1800, which is 1.3 times higher than that of abrasive particles. At the same time, the b-V-RE multiple co-osmotic technology of 16Mn steel can effectively eliminate the FeB phase in the permeable layer and reduce the brittleness of the material. The experimental study showed that the wear resistance of the blade after co-osmotic treatment was significantly improved. After 13 months of work, there was no obvious wear phenomenon of the blade.

Boriding and carbonization technology can make the surface of fan blade with good wear resistance and keep the toughness of blade. For fan blade, the thicker wear-resisting layer, the stronger the ability to resist wear and tear, but leaves the brittleness of increase, fracture problem, are more likely to happen in the actual boride, carbonization technology processing fan blades, the process is difficult and requires accurate mastery of boride, the position of the carbide layer, and thickness, in order to obtain ideal wear-resisting carbonitriding layer.

4.2. Active anti-wear

(1) Optimization of aerodynamic design

The shape of the inlet of centrifugal fan affects the intake conditions, which can be effectively improved by the design of the appropriate shape. First of all, it is necessary to ensure the minimum inlet relative velocity of the fan impeller, and reduce the speed as far as possible under various conditions to meet the actual operating conditions. Secondly, a reasonable impeller runner shape is designed. The design principle is to make the radian curvature radius of the blade gradually increase from the inlet to the outlet, so as to reduce the probability of abrasive particles hitting the blade and slow down the blade wear rate.

(2) Changing the blade structure

①Solid straight blade

At present, many fan blades are hollow. When the friction between the abrasive particles and the blade occurs, the abrasive particles will penetrate the blade, seriously destroying the dynamic balance of the blade and reducing the utilization rate of the fan. Therefore, solid straight blade can be used to replace the existing blade through the design.

② Narrower blades

The reason why the blade pressure surface is the most severely worn part of centrifugal fan is that the airflow incidence Angle and installation Angle of blade have a great influence on the wear situation, and the arc-shaped pressure surface always

exists in a position conducive to the Angle of abrasive particles hitting the blade [17]. Therefore, the design of narrower blades, with reasonable installation, can reduce blade wear.

③Serrated pressure surface

The pressure surface of the fan blade is designed to be pointed in the direction of the tooth groove, which changes the Angle of the grinding blade of the grinding grain, reducing the abrasion of the grinding grain on the surface of the blade, and the wear-resisting metal on the surface of the fan blade, and the ability to improve the blade of the fan blade.

Jackson et al. [18] designed the pressure surface of fan blade into a zigzag shape, so that the airflow direction was perpendicular to the direction of the tooth groove. By designing the clamping groove, the Angle of abrasive particles hitting the blade was changed to reduce the erosion wear caused by abrasive particles on the surface of the blade, and wear-resistant metal was coated on the surface of fan blade, effectively improving the wear resistance of fan blade.

(3) Increasing the front anti-wear cascade

Centrifugal fan parts affected by wear and tear is not the same, this is because the grinding particles in the air flow distribution caused by uneven, grits of the wear rate is big, high concentrations of small particle concentration, wear rate, increase the front wear such as impeller blade method, heat causes air currents to grind grain distribution in uniform, can significantly reduce the wear of impeller, prolong the service life of the centrifugal fan.

Yi etc. [19] introduced a front-facing wear blade impeller. The front anti-wear cascade will rotate with the work of the fan impeller, producing in the process of rotating airflow can dispel grits, forcing the operation of the grinding grain radius, abrasive particle trajectory will not reach after the blade root and plate, which in blade inlet particle uniformly, At this time, the wear condition changes from centralized wear to uniform wear, which effectively slows down the wear rate of the impeller.

5. Conclusion

1) At present, based on the erosion wear mechanism of fan blades, most research scholars to research and analysis from the Angle of mechanics, it is to be from a macro mechanics analysis of abrasive and material into micro- and macro-organization interaction to clarify material influence on the wear and the organization design and development high wear-resisting of new materials, comprehensive analysis and the ability to impact the fan blade wear, Forming a complete erosion wear mechanism.

In the gas-solid two-phase flow field analysis of numerical simulation and experimental analysis needs to be improved. More research will be gathered in the solid grain of secondary flow, the two-way coupling model and gas such as viscous flow direction, to further improve forecasting precision, optimization design, including the flow field, the whole shape of the fan size, reducing and gradually achieving the effect of the uniform wear.

3) When choosing the way of wear-resisting enhancement of fan blades, factors such as the structure of centrifugal fan, gas-solid two-phase flow field and actual operating conditions should be taken into comprehensive consideration to reasonably analyze and achieve appropriate use value and economic benefits.

Acknowledgements

The authors acknowledged the financial support by Shihezi Science and Technology Project of the eighth Division（2020PT02）

References

[1] Mingzhu S. Research on Working Characteristics and Optimization of Wear resistance of Sweeper Fan [D]. Yangzhou University, 2021.
[2] Zhigang Y. Wear Analysis and Reconstruction of Impeller of Environmental Protection Dust Removal Fan [D]. Northeastern University, 2015.
[3] Zengqiang Z. Wear Analysis and surface modification of Circulating Fan Impeller [D]. Northeastern University, 2015.
[4] Ying M, Jun R, Yuandong L, Tijun C, Bing L. Progress of Erosion Wear Research [J]. Journal of Lanzhou University of Technology, 2005(01): pp. 21-25.
[5] Finnie I . Erosion of surfaces by solid particles[J]. Wear, 1960, 3(2): pp. 87-103.
[6] Tilly G P . A two stage mechanism of ductile erosion[J]. Wear, 1973, 23(1): pp. 87-96.
[7] Hutchings I M. A model for the erosion of metals by spherical particles at normal incidence[J]. Wear, 1981, 70 (3) : pp. 269-281.
[8] Jianzhong L, Xinnan L, Bolong Z. Fluid Machinery,1994(01): pp. 12-16, p. 64. (in Chinese)
[9] Mei D, Happiness Hall. Numerical Prediction of Wear Rate of Gas-solid Two-phase Flow Fan [J]. Fluid Machinery, 2007(10): pp. 25-28.
[10] Guangjie L, Ren D, Kangmin C. Journal of University of Shanghai for Science and Technology, 2001(04): pp. 305-308.
[11] Jijun L. Wear Analysis and Treatment of Centrifugal Induced Draft Fan Impeller [J]. General Machinery, 2006(04): pp. 41-42, p. 94.
[12] Kaishen Z. Application of Wear-resisting Ceramics on High Temperature Fan Impeller [J]. Sichuan Cement, 2011(05): p. 110.
[13] Jianzhong L, Jiang L, Yulin L, Liangcheng C. Theoretical Research and Development of Wear Resistant Gas-solid Two-phase Flow Centrifugal Fan [J]. China Mechanical Engineering, 2003(01): pp. 12-15, p. 92.
[14] Mingzhu S, Hui S, Meng Z, Chao Y, Xi L, Mengyuan W. Tribological properties of TiC/Co coating on impeller surface by laser cladding [J]. Special casting and non-ferrous alloys, 2020, 40(06): pp. 658-663.
[15] Jie S, Xiaoliang H, Yan Y, Yuan F. Anti-corrosion and wear resistance treatment of impeller based on thermal spraying process [J]. Chemical equipment technology, 2020, 41(06): pp. 66-68.
[16] Ming J. Application of Multi-permeability Anti-Wear Technology [J]. Journal of Shenyang University of Technology, 2004(03): pp. 271-273.
[17] Feng P. Three-dimensional Numerical Calculation and Research of Centrifugal Fan [D]. Wuhan University of Technology, 2004.
[18] Wang J, Wang J, Wang J, et al. Corrosion resistance of fan blade by corrosion resistance [J].
[19] Ming F, Asheng W. Design of Anti-wear Cascade of Centrifugal Ventilator [J]. Fluid Machinery, 1994(07): pp. 16-18, p. 64.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 315
J. Xu et al. (Eds.)

doi:10.3233/ATDE220243

Study on Structure and Ultrasonic Performance of ZnO Piezoelectric Coatings on Smart Bolts After 300 °C Annealing

Chuan WANG[a], Yanghui JIANG[b], Jingyu LI[b], Xiaomei ZENG[b], Guangming JIAO[a], Wenseng LI[a], Yan LIU[b], Binhong XIA[a], Binhua WAN[a], Jun ZHANG[b], Bing YANG[b,1], and Vasiliy PELENOVICH[c]

[a] *Aerospace Precision Products Inc., Ltd Tianjin, 300300, China*
[b] *School of Power and Mechanical Engineering, Wuhan University, Wuhan, 430072, China*
[c] *The Institute of Technological Sciences, Wuhan University, Wuhan, 430072, China*

Abstract. The accurate measurement of the preload on bolt has been extensively investigated ecent years using longitudinal ultrasonic waves. For ultrasonic measurement of bolt preload ZnO coatings have exhibited excellent piezoelectric properties and stable structure. In this paper, ZnO piezoelectric coatings were prepared onto Si(100) substrates by radio frequency (RF) magnetron sputtering, and the effect of annealing on the structure and ultrasonic performance of the coatings was investigated. The morphology, structure and ultrasonic signal characteristics were characterized by SEM, AFM, XRD and ultrasonic measurement instruments. The experimental results show that ZnO coatings have good stability at 300 °C for 200 h. There are not obvious effects of annealing on the surface morphology and crystalline orientation of the coating. The ultrasonic echo signals of the coatings deposited on the bolts demonstrate good stability after longtime heat treatment.

Keywords. Preload of the bolt, ZnO piezoelectric coating, Annealing, Ultrasonic

1. Introduction

Bolts are indispensable fasteners in machines and mechanisms such as wind-power generators, aeroplane engines, trains, playing the role of connection, reinforcement and sealing, etc. The key factor affecting lifetime and safety of the components is the accurate preload control of the bolts [1]. However, the current measuring methods in industry show poor accuracy or complex analysis process of the obtained data [2-3]. The current methods for controlling and measuring the bolt preload are mainly the corner method, torque method, strain measurement method, elongation method and ultrasonic method [4]. Among them, the torque method can control the preload with an accuracy of ± 30% since the variation of the friction coefficient between the interface of the bolt and the nut or other mating component; the rotation corner method is

[1] Corresponding author, Bing YANG, School of Power and Mechanical Engineering, Wuhan University, Wuhan, China; E-mail: toyangbing@whu.edu.cn.

slightly more accurate compared with torque, but the error is still higher than 15%, which is not enough to ensure the safety of the critical components. The basic requirement of the elongation method is that the two ends of the bolt should be contacted, which limits its application. The strain measurement method has relatively higher accuracy, but the operation of installing strain gauges is complicated and costly, which is not suitable for large-scale applications. Therefore, using ultrasonic to measure preload is a promising and non-destructive technique with high accuracy and stability [5], which has received wide attention over the world. The measurement of preload by the ultrasonic wave uses the value of flight time of a sound wave transmitting through the bolt [6]. Since the ultrasound technology has been used to measure the bolt preload, many groups have paid much attention on how to improve the accuracy of ultrasound measurement using electromagnetic excitation [7] or piezoelectric excitation of ultrasound [8]. Furthermore, methods for measuring stresses by transducer with longitudinal and longitudinal-transverse combined ultrasonic waves were used to measure the preload with higher accuracy [9]. However, a transducer needs a coupling agent to reduce the reflection of the ultrasonic wave at the interface with an object under investigation. Therefore, it is necessary to develop a new ultrasonic transducer, which can measure the stress without coupling agent and with error less than 3%.

Deposited piezoelectric coatings onto the surface of the bolts to form a transducer is an effective technique to perform an accurate preload measurement, which can avoid the use of coupling agent or adhesive. Currently, the main piezoelectric coatings which can generate ultrasonic waves are ZnO [10] and AlN [11], both of them have columnar hexagonal wurzite structure without the requirement for polarization. Piezoelectric ZnO has high piezoelectric coefficients, can be grown on a wide variety of substrates, its epitaxial growth temperature requirements are lower and its stoichiometry is easier to control. Therefore, ZnO coating with strong piezoelectric effect can be chosen as an acoustoelectric transducing layer [12]. There have been numerous studies on the generation of pure longitudinal waves by ZnO piezoelectric films [13-14], but studies on ZnO coatings for use at higher temperatures have not been reported. In this paper, ZnO piezoelectric coatings were prepared by RF magnetron sputtering. The prepared coatings were heat-treated to investigate the effect of annealing on morphology, crystal structure, and efficiency of ultrasonic generation.

2. Materials and methods

2.1. Coating preparation

ZnO piezoelectric coatings were prepared on Si(100) substrates and titanium alloy (Ti6Al4V) bolts using RF magnetron sputtering techniques, see Fig.1. The sputtering target was a 99.99% pure ZnO ceramic target with a diameter of 150 mm and 5 mm thickness. The distance from target to substrate was 110 mm. Before the deposition, silicon substrates were cleaned in acetone, alcohol and deionized water for 10 min, consequently, and then dried in nitrogen flow. The base vacuum in the deposition chamber was 7×10^{-3} Pa. To clean the surface of samples and to improve the adhesive force between the film and the substrate, before the deposition the plasma etching of the samples was performed at bias voltage of -150 V, 80% duty cycle, pressure of 0.5 Pa and the arc current of 70 A. After etching, a mixture of argon (99.99%) and oxygen

(99.99%) was introduced to deposit the ZnO piezoelectric coating. The as-deposited ZnO coatings were subjected to annealing at 300 ℃ for different annealing time(5, 10, 50, 100 and 200 h).

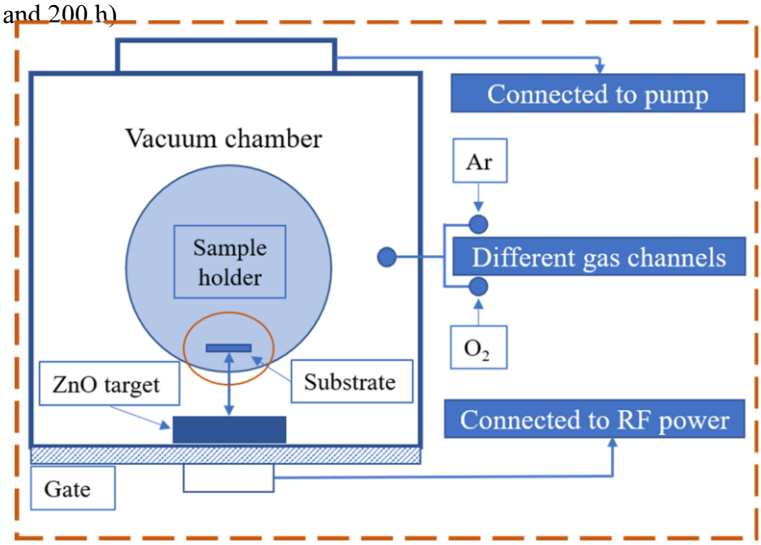


Figure 1. Schematic diagram of the RF magnetron sputtering system.

2.2. Characterization

The morphology of the coatings was observed by a field emission scanning electron microscope (SEM) MIRA3. The surface roughness of the samples was characterized by an atomic force microscope (AFM) Shimadzu SPM-9500 J3. The crystalline phases of the coating were tested by X-ray diffraction technique (XRD) (Tongda TDM-10) using the Cu Kα source, λ=0.15405Å. The electrodes onto the ZnO coatings were prepared by using silver paste of 99.9% purity and dried in oven at a fixed temperature of 100 °C for 1 h. For the ultrasound echo response, -200 V electric pulses (spikes) generated by DPR300 ultrasonic pulser-receiver (JSR Ultrasonics) were applied between the Ag electrode on the ZnO film and grounded bolt. The feedback signal was collected using the same pulser-receiver connected with PC, see Fig.2.

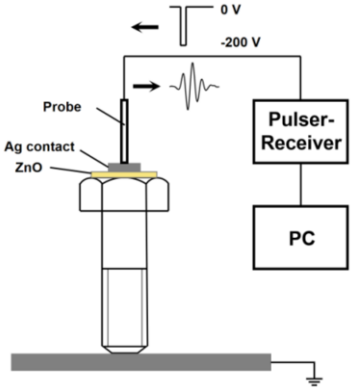


Figure 2. Schematic diagram of the experimental setup for ultrasound echo response measurement.

3. Results and discussion

3.1. Morphologies of coatings after annealing treatment

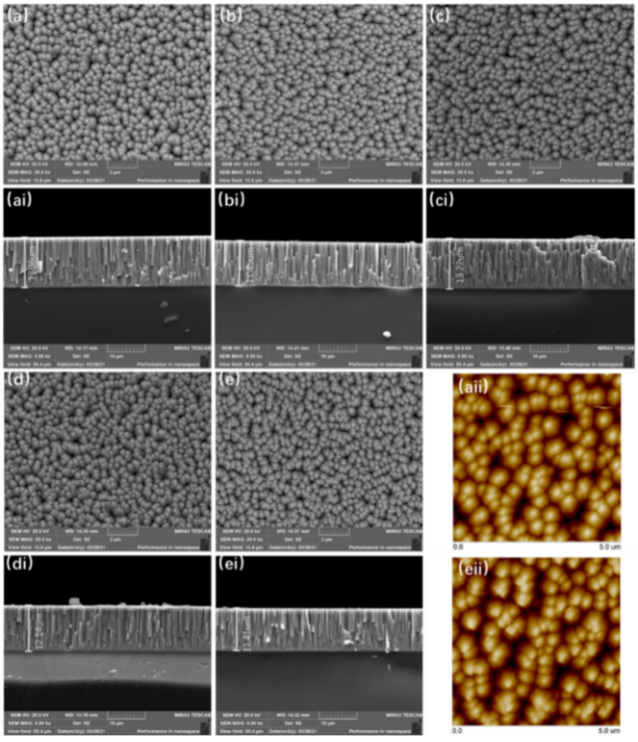


Figure 3. Morphology of ZnO coatings after annealing at 300 °C for different time (a-aii) 5 h; (b-bi) 10 h; (c-ci) 50 h; (d-di) 100 h; and (e-eii) 200 h.

The coated samples were subjected to heat treatment at 300 °C for 5, 10, 50, 100 and 200 h. As shown in Fig. 3(a)-(e), the coating surface is composed of hemispheric-like features, which are top ends of the columnar crystals. One can see that with increasing the annealing time, the morphology of the coating does not demonstrate big change. In Fig. 3(ai)-(ei) the coatings still have a clear columnar structure after being kept at 300 °C for 200 h, and the columnar grain boundaries remain clear. Since the surface morphology does not change much, samples with processing duration of 5 and 200 h were selected for surface roughness test. The AFM images are shown in Fig. 3(aii) and (eii): the roughness values are Rq (5 h) = 104 nm and Rq (200 h) = 105 nm. Generally, the surface changes of heat-treated samples are caused by factors such as increase of oxidized particles, merging of crystal grains, changes in micro-nano structure, etc., but in our experiment such changes are not observed on the treated ZnO coating samples, indicating that the annealing at 300 °C has no critical effect on the coating morphology.

3.2. Crystal structure of coatings after annealing

Fig. 4(a) shows the XRD patterns of the ZnO coating after annealing for different time. The results show that the coatings after heat treatment have only one (002) diffraction peak, indicating that all the coatings have highly c-axis preferred orientation. The (002) diffraction peaks have good symmetry, indicating that the coatings have a good crystalline quality after heat treatment for different time. In Fig. 4(b) the diffraction peak intensity as a function of annealing time is exhibited, indicating the increase of annealing duration will rise the diffraction peak intensity of the coatings. Since the thickness of the film is relatively uniform, the increase of the (002) diffraction peak indicates that the crystal quality of the coating is improved.

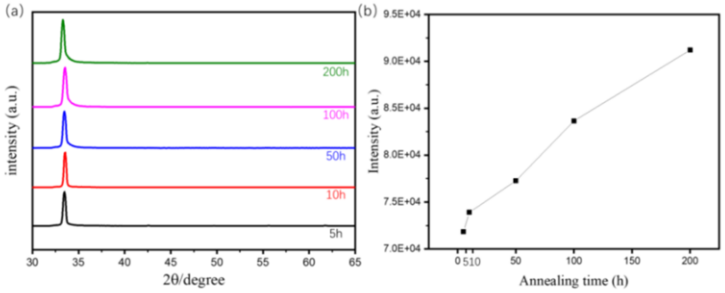


Figure 4. (a) XRD patterns of ZnO coatings at different annealing treatment; (b) Intensity of (002) diffraction peak as a function of annealing time.

The long-term heat treatment provides energy for the atoms inside the coating to overcome the potential barriers and migrate. The atoms can have enough time to migrate from the non-equilibrium positions to the (002) surface with the lowest energy. During the annealing the defect density inside the coating is reduced and the crystalline quality is improved. Therefore, the annealing at 300 °C for 200 h can improve the quality of the original coating.

3.3. Ultrasonic properties of smart bolts after annealing

The bolt samples after annealing are shown in Fig. 5. The ZnO piezoelectric layer on the surface of all bolts remains bright and intact, indicating that the long-term treatment at 300 °C does not cause the peeling damage of the coatings. The ultrasonic echo signal tests are performed on the processed bolts after annealing, and the result is shown in Fig. 6(a). The first four equidistant echoes can be clearly observed in all the charts. The echo signal was identified as a pure longitudinal wave.

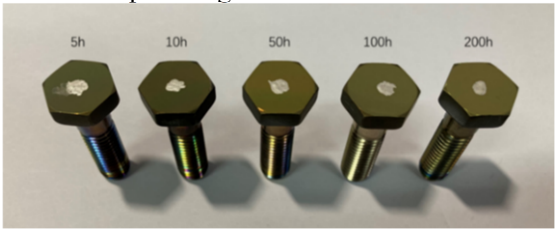


Figure 5. Bolt samples after different annealing time at 300 °C.

As shown in Fig. 6(b), the amplitude attenuation rates were calculated according to the bolt echo signal values. The decay rate first decreases and then increases as the annealing time increases, the lowest and highest decay rates in this group of samples are 7% and 50%, respectively. Next, we calculate the initial (not attenuated) ultrasonic signals using the attenuation rate, the result is shown in Fig. 6(c). It can be found from the figure that after 5 h annealing, the original ultrasonic signal amplitude is 13.22 V. With the increasing annealing time, the amplitude of the initial ultrasonic signal remains relatively stable at about 7.5 V. As the annealing time increases, the crystalline quality of the coatings improves, see Section 3.2. Therefore, the ultrasonic performance of the coating would be expected to be improved, that is, the signal amplitude will increase or remain unchanged, however, in our experiment we observe that the signal amplitude first decreases and then remains stable, see Fig. 6(c). The specific reasons of this effect need to be further studied. The above results show that the deposited on the bolt coatings after heat treatment for different time do not show any change of efficiency to generate the ultrasonic longitudinal waves, and the minimum amplitude of the echo signal is 2.72 V. Therefore, the coating deposited on the bolt has good stability after a long time annealing at 300 °C.

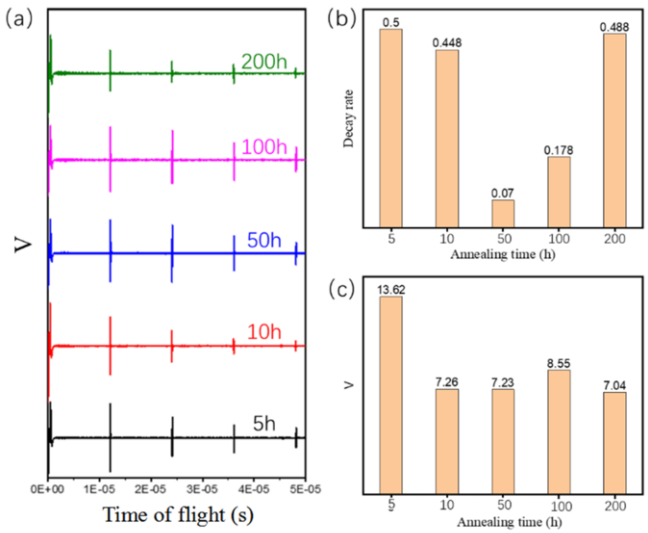


Figure 6. Ultrasonic echo signals of smart bolts annealed for different time (a) Ultrasonic echo signal; (b) Decay rate of the ultrasonic signal; (c) Initial value of ultrasonic signal.

4. Conclusions

The effects of different annealing time at 300 °C on the coating structure and ultrasonic wave generation were investigated for RF magnetron deposited ZnO piezoelectric coatings. The results show that annealing for different time does not only cause any damage to the coating, but also enhances the coating properties. Therefore, the annealing experiments prove that the coatings are morphologically and structurally stable up to 300 °C and can meet the minimal requirements for applications for bolts in non-room temperature service.

References

[1] Rusong M, Ruilin S, Songhan Z *et al*. A review of bolt tightening force measurement and loosening detection. Sensors (Switzerland). 2020; 20(11).
[2] Salim C, Gilles C, Ivan L et al. Combination of longitudinal and transverse ultrasonic waves for in situ control of the tightening of bolts. Journal of Pressure Vessel Technology, Transactions of the ASME. 2007; 129(3): pp 383–390.
[3] Wentao S, Chunguang X, Qinxue P et al. Nondestructive testing and characterization of residual stress field using an ultrasonic method. Chinese Journal of Mechanical Engineering (English Edition). 2016; 29(2): pp 365–371.
[4] Seyed M Y N, Masoud G. A review paper on looseness detection methods in bolted structures. Latin American Journal of Solids and Structures. 2017; 14(12): pp 2153–2176.
[5] Xinxin Z, Yiming L, Xiaoguang L *et al*. Ultrasonic measurement on axial force of high-strength bolt in service. Journal of Low Frequency Noise Vibration and Active Control. 2020; 39(3): pp 596–603.
[6] Haibo L, Tianran L, Yapeng L *et al*. Uniaxial stress in-situ measurement using EMAT shear and longitudinal waves: Transducer design and experiments. Applied Acoustics. 2021; p 175.
[7] Y. P. S, X. X. Z, J. H. Z. Asia Pacific Conference for Non-Destructive Testing (APCNDT2017), Singapore. Bolt stress inspection by Electromagnetic Acoustic Transducer(EMAT) and PZT. 2017; pp 1–8.
[8] Junhua S, Tao W, Heyue Y *et al*. Bolt looseness detection based on piezoelectric impedance frequency shift. Applied Sciences (Switzerland). 2016; 6(10).
[9] Henri W, Patrick B. Application of ultrasonic measurements to stress assesment on already tightened bolts 1 . Theoretical basis of ultrasonic stress measurement 2 . Application of the ultrasonic method to ultrasonic control of bolt tightening. Ecndt. 2014; 2014(1).
[10] Hanada T. Basic Properties of ZnO, GaN, and Related Materials. 2009: pp 1–19.
[11] Marta C, Enrique I, Jimena O *et al*. On the effectiveness of lateral excitation of shear modes in AlN layered resonators. Ultrasonics, 2014, 54(6): 1504–1508. DOI:10.1016/j.ultras.2014.04.014.
[12] W. L. D, Y. Q. F, J. K. L *et al*. Deposition and characterization of sputtered ZnO films. Superlattices and Microstructures, 2007, 42(1–6): 89–93. DOI:10.1016/j.spmi.2007.04.081.
[13] S. B, Laurent L B, Omar E *et al*. Deposition of ZnO inclined c-axis on silicon and diamond by r.f. magnetron sputtering. Physica Status Solidi (A) Applications and Materials Science. 2007; 204(9): pp 3091–3095.
[14] S. B, Laurent L B, Omar E et al. SAW devices based on ZnO inclined c-axis on diamond. Diamond and Related Materials. 2008; 17(7–10): pp 1420–1423.

doi:10.3233/ATDE220244

Statistics and Analysis of Wind Turbine Failure Data for a Wind Farm in Xinjiang for One Year

Xiongfei ZHENG [a], Xue HU [a,1], Lixin ZHANG [a,b], Yulong CHENG [a] and Chunliang MAI[a]
[a] *School of Mechanical and Electrical Engineering, Shihezi University, Shihezi 832003, Xinjiang, China*
[b] *Corps Industrial Technology Research Institute, Shihezi 832003, Xinjiang, China*

Abstract. This article conducts statistical analysis on the wind turbine failure data of a wind farm in Xinjiang in 2019. The main statistical methods include common failure types, total duration of each type of failure, average duration of each type of failure, etc. Based on the results of the previous data statistics, each type of the severity levels of the faults are divided, and finally the number of faults varies with month. The conclusion of the statistical analysis can provide certain guidance to relevant wind turbine operation and maintenance personnel.

Keywords. Wind turbine operation and maintenance, Failure analysis, Fault type

1. Introduction

Wind energy is a clean renewable energy. With the increasing global resource shortage problem, the development of the wind energy industry has received more and more attention. Xinjiang is rich in wind resources. The theoretical reserves of wind energy (at a height of 10m) are 872 million kilowatts, which accounts for about 20.8% of the country's total wind energy resources. Wind power is a clean energy source that is most likely to achieve large-scale and economical alternatives to coal power in the short term, and it is also an important channel for farmers and herdsmen in remote areas to use electricity. However, Xinjiang's environment is complex and changeable (large temperature difference, strong radiation, sudden wind speed, dust, snow, etc.), which adversely affects the safe operation of wind turbines. The multi-unit parallel coupling of wind farms is complicated, and various impacts such as low-voltage ride-through (LVRT) and high-voltage ride-through (HVRT) are often caused by grid voltage drops and swells, making them endure various tests and frequent failures[1-3].

The significance of wind turbine fault analysis and fault diagnosis is mainly manifested in reducing the accident rate of the generating unit, reducing maintenance costs, reducing maintenance time, and improving the operating efficiency and reliability of the unit, providing guidance for designers. According to statistics, after adopting maintenance management and fault diagnosis centered on reliability analysis technology, the accident rate is reduced by 75%, and the maintenance cost is reduced by 25-50%. No matter from the perspective of safety, from the perspective of economic

benefits, the fault analysis of wind turbines and the establishment of a complete fault diagnosis system are extremely important[4-7].

Based on the theory of mechanical fault diagnosis, early scholars analyzed the types of faults and the corresponding causes of faults in the actual operation of the main drive chain of a kilowatt-level wind turbine generator[8-11]. With the advancement of technology, intelligent diagnosis methods are more and more widely used in the operation and maintenance of wind turbines. Subsequent scholars are based on neural networks, genetic algorithms, etc. Mainly analyze the faults of the main shaft gearbox of the wind turbine, and propose a wind turbine state monitoring and fault diagnosis system[12-16]. In summary, the main research object of the above-mentioned fault diagnosis method is wind turbines with kilowatt-level power in the early stage, which cannot provide guidance for the operation and maintenance of current megawatt-level wind turbines. On the other hand, the object of studying the fault is a part of the wind turbine, and the wind turbine is a huge and complex system. Therefore, the proposed operation and maintenance plan is not comprehensive enough.

This article is based on the statistical analysis of unit operation failure data of a wind farm in Xinjiang in 2019. Combining with the climate conditions of the power plant in the current year, the corresponding failure causes of the failure types of each subsystem of all units are analyzed in different statistical methods. It also proposes feasible handling methods for each type of failure of the subsystem. The results of the analysis and research can provide certain guidance for the operation and maintenance of wind power plants.

2. Data source Wind farm introduction

The total installed capacity of the wind farm is 47.5MW, and the wind turbine power is 2.5MW. The average temperature of the region this year is 16℃, July is the hottest month of the year, with an average maximum temperature of 21.5℃ and an extreme maximum temperature of 34℃. January is the lowest temperature month of the year. Its average lowest temperature is -15.5℃, and in extreme cases the lowest temperature can even reach -28℃. The number of days when the lowest temperature is below 0℃ in the year is 175; the number of days when the temperature is below -10℃ is 85 days; and the number of days when the temperature is below -20℃ is 10 days. The average rainfall in 2019 is about 70mm, and the average annual wind speed is 8.8m/s. Table 1 is a detailed introduction to the climatic conditions throughout the year.

Table 1. Statistics of main climatic factors of wind farms throughout the year

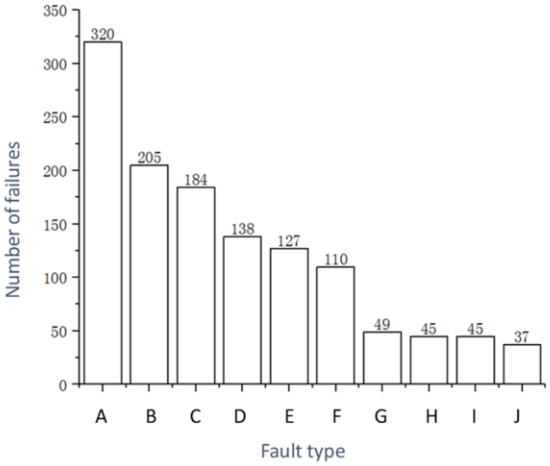
month	Minimum temperature (℃)	Maximum temperature(℃)	Average temperature(℃)	Temperature difference (℃)	Average wind speed (m/s)
1	-28	-2	-15.5	26	9.17
2	-19	11	-6.5	30	7.80
3	-12	24	5.5	36	8.92
4	-6	29	9	35	9.16
5	0	30	13	30	9.23
6	6	33	20.5	26	9.1
7	11	34	21.5	23	9.26
8	7	33	20.5	26	9.27
9	-1	25	11.5	26	7.19
10	-12	23	5	35	7.16
11	-18	8	-4	26	9.76
12	-1	-23	-10.5	22	9.86

3. Statistical analysis of fault data from different angles

In order to operate and maintain the wind turbine better, the most ideal way is to fully understand the causes of common failure types, and then solve the problems from the source through theoretical and technological innovation to eliminate or reduce the occurrence of failures. Secondly, after the failure occurs, you need to be familiar with the handling methods of different failures and solve related failure problems quickly and timely. At the same time, in different seasons and different climatic conditions, it is necessary to have a certain ability to predict possible failures.

3.1. The composition and common fault types of wind turbines

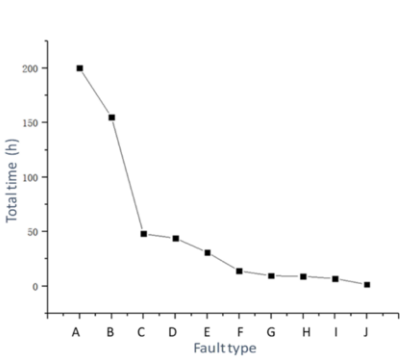
The composition of the wind turbine is very complex, which is a huge system, mainly including the hub, wind wheel, tower, brake system, transmission system, yaw system, pitch system, main control system, frequency conversion system, generator, transformer and parallel Network system and other subsystems.And each subsystem is composed of several parts. Wind turbines often work under complex and changeable working conditions, so its internal systems are prone to failures, which will affect the normal operation of the generators and affect the power generation. In order to ensure the output and benefits of wind farms, relevant operation and maintenance personnel are required to clarify the faults that frequently occur during the operation of the wind turbine and the countermeasures for different faults. Through the classification and statistics of the fault data of a wind farm in one year, the top ten fault types with the number of faults in a year are sorted out, as shown in Figure 1.

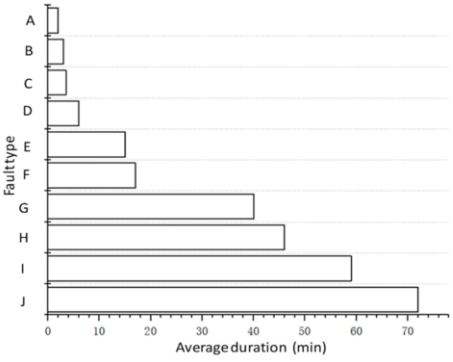


A: Emergency shutdown of converter; B: Converter warning; C: Gear lubrication pump operation failure; D: Low inlet valve temperature; E: Gear oil pressure failure; F: 2#Pitch motor temperature is high; G: PCH substation bus failure; H: Pitch safety chain failure; I:1#The bus of the pitch substation is abnormal; J: Yaw direction failure

Figure 1. The top ten types of faults in a wind farm

It can be seen from Figure 1 that the most faulty component is the converter, mainly because the converter is often in a complex and changeable environment (large temperature difference, strong radiation, sudden wind speed, sand, dust, snow, etc.). This will adversely affect the safe operation of wind turbines. The multi-unit parallel coupling of wind farms is complicated. The voltage drop and swell of the grid often cause various impacts such as low voltage ride-through (LVRT) and high voltage ride-through (HVRT), which saturate the converter. Subject to various tests, failures frequently occur. However, the economic loss caused by the converter failure is relatively small. Generally, the corresponding parts can be replaced, and more spare parts should be added. Secondly, the gear lubrication system and the pitch system have many faults, which can be focused on during maintenance and repair, and spare parts can be appropriately increased. PCH substation bus and yaw system failures are relatively rare, but as the operating time of the unit increases, failures will increase, and regular maintenance and maintenance of these systems should still be strengthened. The number of failures is an important indicator to measure the probability of a certain type of failure.But it is not complete to rely on this indicator to measure it. Therefore, this article also analyzes the total duration and average duration of each type of failure. Statistics, the statistical results are shown in Figure 2 and Figure 3.





A: Emergency shutdown of converter; B: Converter warning; C: Gear lubrication pump operation failure; D: Low inlet valve temperature; E: Gear oil pressure failure; F: 2#Pitch motor temperature is high; G: PCH substation bus failure; H: Pitch safety chain failure; I:1#The bus of the pitch substation is abnormal; J: Yaw direction failure

Figure 2. Total duration of each type of failure **Figure 3**. Average duration of each type of failure

The top three failure types in each evaluation index are taken as an example for analysis. According to the statistical results, the top three types of failures, from most to least are: converter emergency shutdown, converter warning, and gear lubrication pump operation failure. The top three types of failures in the order of total failure duration are: converter emergency shutdown, converter warning, and pitch safety chain failure. The top three types of failures, from longest to shortest, for the average duration of failures are: pitch safety chain failure, 1# pitch substation bus abnormality, and converter warning. From these statistical results, it can be seen that converters and gear lubrication pumps are prone to frequent failures, and the total duration of converter emergency shutdown and converter warning are also ranked in the top two places. The converter warning average failure time is in the third place, so the converter needs more maintenance times per year, and the maintenance time for each failure is longer. Although the number of failures of gear lubrication pumps is ranked third, the total duration of failures and the average duration of failures are ranked lower,

indicating that such failures are easy to occur and easy to repair. Pitch safety chain failure and 1# pitch substation bus abnormality have relatively low failures. But the average failure time is the first two types of failures, indicating that the probability of these two types of failures is low, but the failures are not easy to repair.

3.2. Statistics of the number of failures of the main subsystems of the wind turbine

In the previous section, statistical analysis of all types of failures was carried out in different ways. On this basis, it is also necessary to analyze the occurrence of failures of different subsystems of the wind turbine.

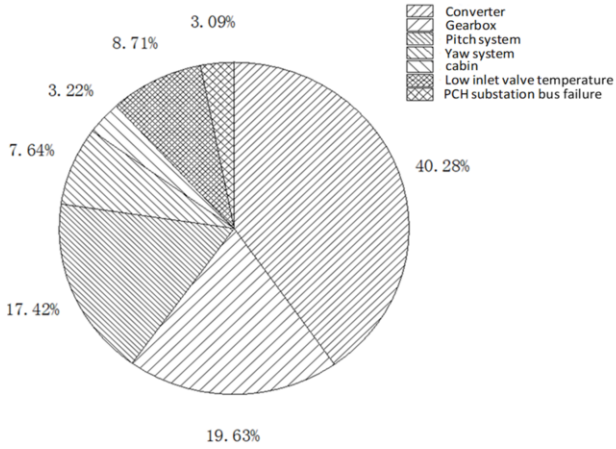


Figure 4. The top eight types of faults in a wind farm

From the statistical results in Figure 4, The most failure-prone subsystem in a wind turbine is the converter, which accounts for 40.28% of the total number of failures of the wind turbine system, followed by the gearbox and the pitch system, and the number of failures accounts for the total number of failures. The proportions are 19.63% and 17.42% respectively. The total number of failures of these three subsystems accounted for 77.33% of the wind turbine system. In the operation and maintenance management of wind turbine systems, converters and gearboxes have always been the two main subsystems that have been widely concerned. In addition to objective natural conditions, the main factors affecting the annual power generation of wind farms are the faults of converters and gearboxes. These problems are also a test for relevant operation and maintenance personnel.

3.3. Sorting of the number of failures in different months and analysis of the causes

In order to explore the relationship between the occurrence of failures and months, the number of occurrences of failures in different months was counted. The statistical results are shown in Figure 5:

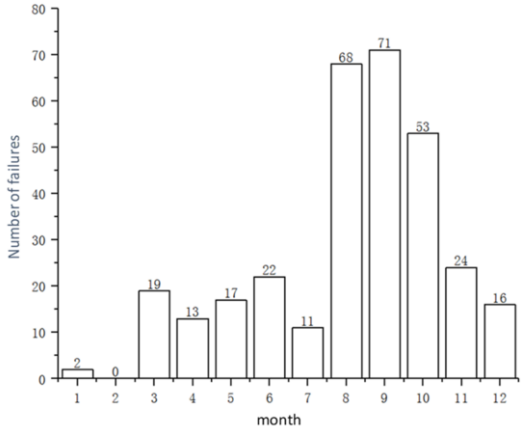


Figure 5. Number of failures in different months

It can be seen from the figure that the failures of the units in a year are uneven. Among them, there are relatively more failures in winter and summer, and fewer failures in spring and autumn. Mainly because there are more extreme weather in winter and summer, and January and July are the two coldest and hottest months in Xinjiang when the faults occur the most. The wind opportunities are due to the harsh climate and the temperature is too high or too low. The wind speed is too small or too high and the machine stops or malfunctions. Therefore, in the winter and summer seasons, it is necessary to invest more time and cost in the operation and maintenance of the unit, and to do the overhaul and maintenance of the unit before the arrival of the harsh working environment.

3.4. Statistics of the number of occurrences of major faults in different months

In order to explore the relationship between the number of major failures and months, the top three types of failures: converter emergency shutdown, converter warning, gear lubrication pump operation failure, and the relationship between the number of failures in different months perform statistics, and the statistical results are shown in Figure 6:

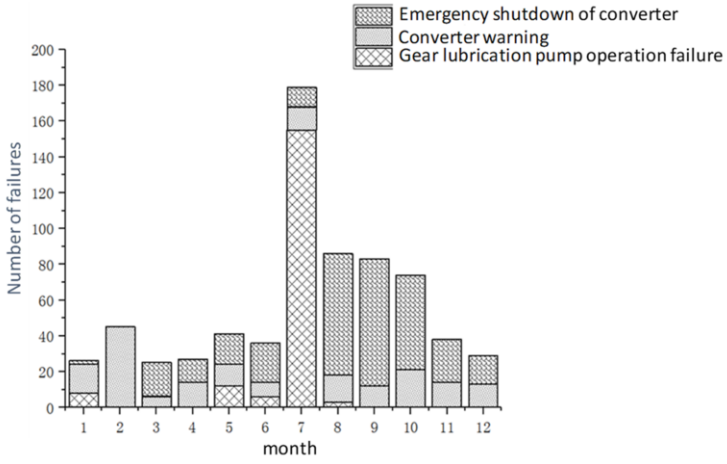


Figure 6. The number of main fault types in different months

It can be seen from Figure 6 that the emergency shutdown of the converter mainly occurred in August, September, and October, because the converter is often in a fluctuating working condition, which will generate a large amount of heat instantaneously, and the main reasons for the failure are also the same. Temperature is related. Therefore, the failure rate is higher in the months of higher temperature in Xinjiang, such as August, September, and October, and the maintenance and repair work of the converter needs to be done. Judging from the statistical results in Figure 6, this fault warning of the converter has many times in each month of the year, showing a relatively average characteristic, except for the larger number of faults in February. The characteristics of the statistical results in Figure 6 are very clear. The vast majority of gear lubrication pump operation failures occurred in July, and only a few failures occurred in the remaining months. Therefore, the prevention of such failures and the corresponding maintenance work were mainly concentrated in July.

4. Apply FMECA method to classify the severity of the fault

Based on the conclusion of the previous statistical analysis of the data, the FMECA method is introduced in this section to analyze the severity of the fault. The severity of the failure is the main technical indicator of the severity analysis. It is equal to the ratio of the number of occurrences of a certain failure mode during the working period of the product to the number of all failures that occurred during the period. Class A (occurs frequently): Class B (occurs sometimes): Class C (occurs occasionally): Class D (occurs rarely): Class E (occurs very rarely).

Table 2. Severity analysis of different fault types

	Fault type	Failure severity	Probability level	Severity level
1	Emergency shutdown of converter	2.28%	Class C	Level 2
2	Converter warning	1.77%	Class C	Level 2
3	Pitch safety chain failure	0.55%	Class D	Level 3
4	1# Bus abnormality of pitch substation	0.50%	Class D	Level 3
5	2# Pitch motor temperature is high	0.35%	Class D	Level 3
6	Low inlet valve temperature	0.16%	Class D	Level 3
7	Gear lubrication pump operation failure	0.12%	Class D	Level 3
8	Yaw direction failure	0.10%	Class D	Level 3
9	Gear oil pressure failure	0.08%	Class E	Level 4
10	PCH substation bus failure	0.02%	Class E	Level 4

The following four levels are used to classify the severity of the fault: the first level is catastrophic, which will damage the entire wind turbine or cause casualties; the second level is serious, which will seriously damage the wind turbine and cause the unit to fail; and the third level is general. It will cause general damage and affect the operating capability of the unit; Level 4 is secondary and usually does not damage the unit, but may require unplanned maintenance.

5. Principles and guidance of wind turbine operation and maintenance

After passing the previous statistics and analysis, we have a more in-depth understanding of the laws of different types of failures in different seasons and months of the year, and then analyzed the causes and mechanisms of these failures, and obtained some failures. The conclusions of prevention and treatment are summarized in this section, providing certain guidance to relevant operation and maintenance personnel.

(1) The component with the most failures is the converter, but the economic loss caused by the converter failure is relatively small, and it is generally sufficient to replace the corresponding parts. Secondly, the gear lubrication system and the pitch system have many faults, which can be focused on during maintenance and repair, and spare parts can be appropriately increased.

(2) The total fault duration of the converter emergency shutdown and converter warning are also ranked in the top two places. The converter warning average failure time is in the third place, so the converter needs more maintenance times per year, and the maintenance time for each failure is longer. Although the number of failures of gear lubrication pumps is ranked third, the total duration of failures and the average duration of failures are ranked lower, indicating that such failures are easy to occur and easy to repair.

(3) The failures of the units during the year are uneven, with relatively more failures in winter and summer, and fewer failures in spring and autumn. Therefore, in the winter and summer seasons, it is necessary to invest more time and cost in the operation and maintenance of the unit, and to do the overhaul and maintenance of the unit before the arrival of the harsh working environment.

(4) The emergency shutdown of the converter mainly occurred in August, September, and October, and maintenance and repair of the converter were required. Most of the gear lubrication pump operation failures occurred in July, and there were only a few failures in the remaining months. So the prevention of this type of failure and the corresponding maintenance work are mainly concentrated in July.

6. In conclusion

Based on the statistical analysis of the annual failure data of a wind farm in Xinjiang, this paper analyzes the failure causes of the top ten failure types in the number of failures, and also uses different rating indicators, such as total failure time, average failure time, etc. The cause of the main fault types in each evaluation index and the operation and maintenance strategy is discussed. The most failure-prone device in a wind turbine is the converter, and its failures mainly occur in August, September, and October, and most of the gear lubrication pump operation failures occur in July.

References

[1] Yu W, Shenchao Z, Fangbin C, Sheng Z. A preliminary study on the coordinated development of nuclear power and renewable energy power generation in China. Renewable Energy. 2021; 39(08): pp. 1069-1077.
[2] Peng L, Tingling W. Research on Optimal Power Flow Considering the Uncertainty of Renewable Energy. Renewable Energy. 2021; 39(08): pp. 1117-1123.
[3] Xieben W, Xujin L, Peiming S, Tongbin L. Analysis of vibration status monitoring and fault diagnosis of wind turbines. China Equipment Engineering; 2021(15): pp. 142-143.
[4] XinanT . Fault diagnosis of 600KW wind turbine generator. Xinjiang University. 2006.
[5] Weichen L, Xianglian X, Ke P, Mingming Y. Discussion on the realization method of fault diagnosis for wind turbines. High Voltage Apparatus. 2011; 47(08): pp. 57-62 p. 67.
[6] Mingming Y. Statistical analysis and fault diagnosis of failure modes of large-scale wind turbines. North China Electric Power University (Beijing); 2009.
[7] Shengkui Z. System Reliability Design Analysis Course. Beijing University of Aeronautics and Astronautics Press; 2001.
[8] Baibing Y. Fault analysis and countermeasures of wind turbine pitch system in wind farm. Hubei Agricultural Mechanization. 2020(12): pp 43-44.
[9] Jiaan Z, Dong L, Zhijun L, Xu H, Hui L, Cun D, Junyan W, Chenyu L, Yunpeng X. Power prediction of a wind farm cluster based on spatiotemporal correlations. Applied Energy. 2021; p. 302.
[10] Yingchao D, Hongli Z, Cong W, Xiaojun Z. Wind power forecasting based on stacking ensemble model, decomposition and intelligent optimization algorithm. Neurocomputing. 2021; p. 462.
[11] Bian C, Liu S, Xing H, et al. Research on fault-tolerant operation strategy of rectifier of square wave motor in wind power system[J]. China Electrotechnical Society Transactions on Electrical Machines and Systems. 2021; 5(1): pp. 62-69.
[12] Han H. Study on the Method of Fault Set for AC/DC System Operation Considering Wind Power. The Journal of Engineering; 2018.
[13] Hui Z. Fault prediction of key components of wind turbine based on BP neural network. Power Station System Engineering 2021; 37(02): pp. 21-22.
[14] Wenbo C, Xi P, Qian L, Bosen Z, Jin X, Lingling H. Fault Analysis of Catalytic Main Fan Unit Based on Condition Monitoring and Fault Diagnosis Technology. China Equipment Cheng; 2021(05): pp. 152-154.
[15] Jinling L, Xiangguo Z, Wei Z, Luyu G, Ruotong W. Fault Diagnosis of Fan Main Bearing Based on Improved Auxiliary Classification Generative Adversarial Network. Electric Power Automation, 2021; 45(07): pp. 148-154.
[16] He S, Wang W Q, Wang HY, et al. Research on Electrical Power System with Analysis of Fault Operation State Magnetic Field about DFIG in Wind Power. Advanced Materials Research. 2013; 676: pp. 181-185.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 331
J. Xu et al. (Eds.)

doi:10.3233/ATDE220245

Pd (CH_3COO)$_2$ and MIL-101 (Cr) Composites: A Novel In-situ Oxidation Catalyst for the Removal of BT from Fuel Oil Under Neutral Condition

Xiao ZHANG [a,b], Yueyun ZHU [a,b], Qinqin HAN[c], Xiaoli MA [a,b], Ping CHEN [a,b], Zhixi ZHAO [a,b] and Qing WANG [a,b,1]

[a]*School of Chemistry &Chemical Engineering, Xinjiang Normal University, Urumqi 830054, China*

[b]*Xinjiang Key Laboratory of Energy Storage and Photoelectrocatalytic Materials, Urumqi 830054, China*

[c]*Urumqi Environmental Monitoring Central Station, Xinjiang, Urumqi 830054, China*

Abstract. In the field of environmental science research, an important research direction is to find a suitable, effective, material and method for desulphurization of fossil fuels and prevent the release of sulfide in fuels into the environment.In this study, a novel catalyst consisting of Pd(CH_3COO)$_2$ and MIL-101(Cr) was used to study the sulfur oxidation of organic sulfides, which are difficult to treat, In simulated fuel oil.The optimum reaction conditions were discussed, including the loading of The optimum reaction conditions were discussed, including reaction temperature, the content of Pd in MIL-101(Cr), catalyst dosage, initial BT concentration and cycle times.The BT conversion efficiency of the catalyst in simulated fuel is 88.77%,It shows that the catalyst has good desulfurization performance.

Keywords. Neutral condition, in situ oxidation desulfurization, Pd(CH_3COO)$_2$, MIL-101(Cr), benzothiophene

1. Introduction

In the field of environmental science, the desulfurization of petroleum distillate is still an urgent problem to be solved.The burning of petroleum and other chemical dyes has irreparably affected the global environment[1]. In industrialized countries, there is strict regulation of allowable levels of sulphur in commercial fuels, but there is a trend to tighten standards. Therefore, refining desulphurization will be imperative[2]. In the current commonly used desulfurization system, although hydrodesulfurization technology is mature and widely used, there are still problems such as operating conditions are demanding, with high requirements for temperature, pressure and cost. Thiophene is difficult to reduce its sulfur content below 10mg /kg due to its stable conjugated structure[3]. However, oxidation desulfurization technology does not

[1] Corresponding Author, Qing Wang, Xinjiang Normal University 102 Xinyi Road, Urumqi, Xinjiang, China; Emall: wangqing2002108@163.com.

require expensive hydrogen at room temperature (below 100℃) and has high reactivity to transform stable sulfur compounds, Therefore, it is widely used in desulfurization research[4].

Most of the desulfurization processes by in-situ oxidized oxidants are required to be carried out under the reaction conditions of high temperature and high pressure at present. Therefore, in-situ oxidized desulfurization by in-situ oxidized oxidants under neutral conditions at low temperatures has become a new research direction. In this study, MOFs material was prepared by hydrothermal method, dibenzothiophene was selected as the sulfide of simulated fuel, and other factors such as different time, temperature, catalyst quality and different Pd salt loads and loads were adjusted to achieve the effect of deep desulfurization, and the reaction mechanism was explored by characterization means.

2. Experiment

2.1 Reagents

Terephthalic acid (AR, 97%), Beijing Balinwei Technology ;Chromium nitrate (AR, 99%) Tianjin Jiangtian Chemical;Acetic acid (AR, 36%), Tianjin Komil Chemical Reagent ;Palladium acetate(AR, 46-48.0%), Shanghai Aladdin Biochemical Technology.

Isooctane (AR, 99%),Acetonitrile(AR, 99%) and Isopropanol(AR, 99%) supplied by Tianjin Institute of Guangfu Fine Chemicals ;

Dibenzothiophene(AR, 99%) Beijing Balinwei Technology ;Sodium Borohydride, (AR, 99%).Tianjin Chemical Reagent Supply and Marketing .

2.2 Synthesis of catalyst

The terephthalic acid, Cr(NO3) 3.9H20, acetic acid solution and deionized water were poured into a 100mL special container and stored in an oven at 220℃ for 8 h. The obtained materials were separated by centrifugation with DMF and anhydrous ethanol at 70℃, and the samples were dried in vacuum at 150℃. Therefore, according to chemical analysis, the obtained green crystalline material is MIL-101(Cr), with high crystallinity.

The synthesized MIL-101(Cr) material was added with acetone and Pd(CH3COO)2. Then transfer the mixture to a beaker and add 5 mL DMF. The solids obtained after drying were placed in a vacuum drying oven at 90℃ for 6 h. The catalyst was obtained by reducing KBH4 ethanol solution in ice bath for 2h.
The desulfurization device used is shown in the figure1 below:

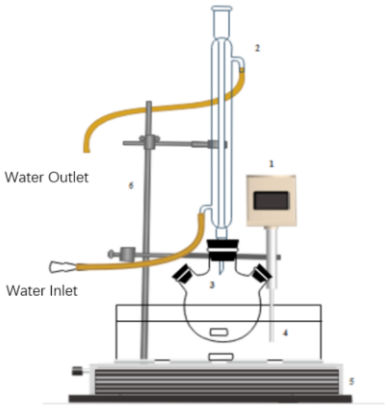


Figure 1. Catalytic testing reactor set-up.

(1) Temperature controller, (2) water condenser, (3) round bottom flask, (4) Oil Bath and (5) Heating plate with stirrer controller (6) Iron stage [5]

2.3Characterization of XRD of synthetic materials

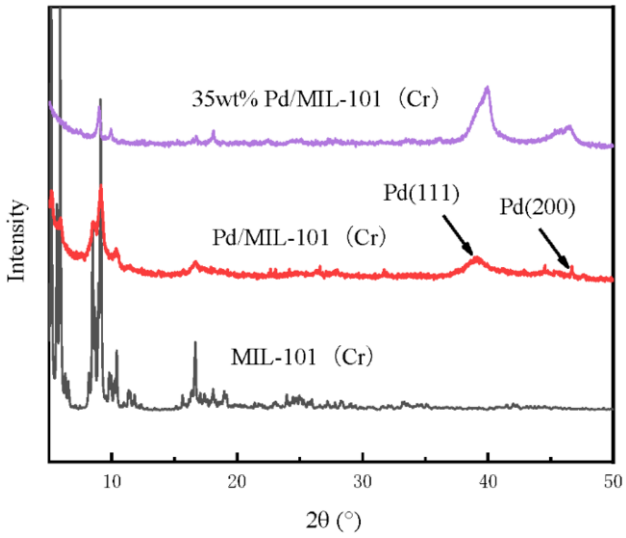


Figure 2. XRD patterns of MIL-101(Cr) and different loading percentages of Pd/ MIL-101(Cr).

In order to analyze the composition of the catalyst composites, XRD tests were carried out for different materials. The diffraction peak of mesoporous MIL-101(Cr) is located in the small Angle region (2θ<10°), which is in good agreement with the published XRD patterns simulated by the same MIL-101 (Cr) single crystal diffraction data. It can be seen from the figure that the peak value of the catalyst composite is Pd(111) crystal plane at 2θ=40.2°. At 2θ=47.0°, the new peak of the catalyst represents the crystal face of Pd(200), which is consistent with previous reports, indicating that the

loading of the catalyst material is successful. However, when the mass load rate reaches 35wt%, it can be seen from the figure that its diffraction peak has been sharply reduced at a small Angle, indicating that 35wt% the catalyst structure has collapsed.

2.4 XPS characterization of synthetic materials

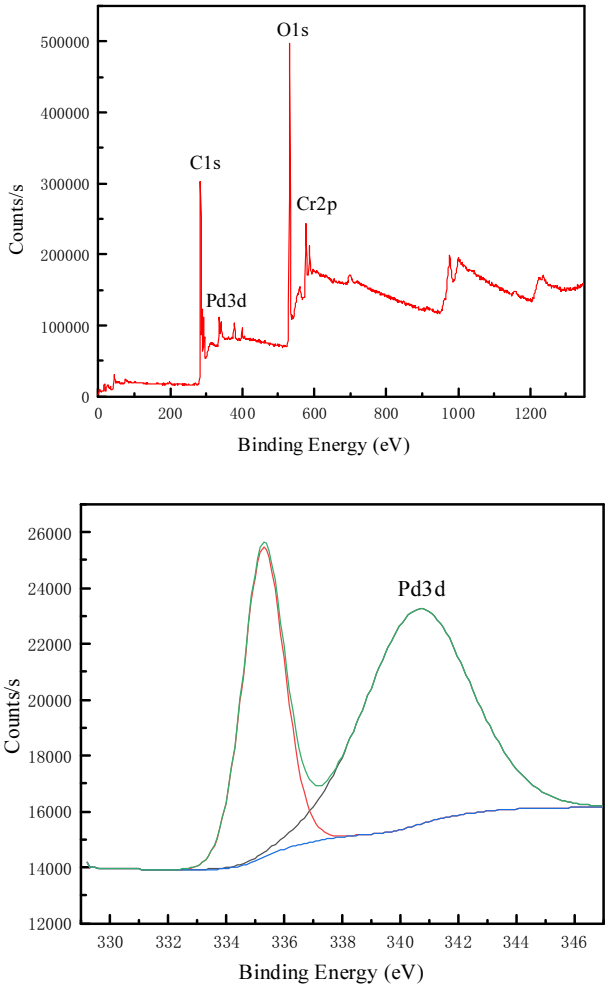


Figure 3. XPS pattern of Pd / MIL-101(Cr).

The chemical composition and electronic structure of the catalyst were analyzed by XPS.The resolution of Pd,C, Cr and O XPS spectra are given respectively.The peak of Pd3d 5/2 binding energy of 335.1eV can be identified as the peak of zero-valent palladium, and the slightly shorter peak on the left is the spin splitting peak.This is consistent with the expected results of this experiment [6].After the reduction, the zero-valent palladium loaded on MIL-101(Cr) can better perform the oxidation desulfurization reaction, and prepare for the next better desulfurization.

3. Results and discussion

3.1 Catalytic performance test

(1)The influence of reaction time and temperature on desulfurization effect

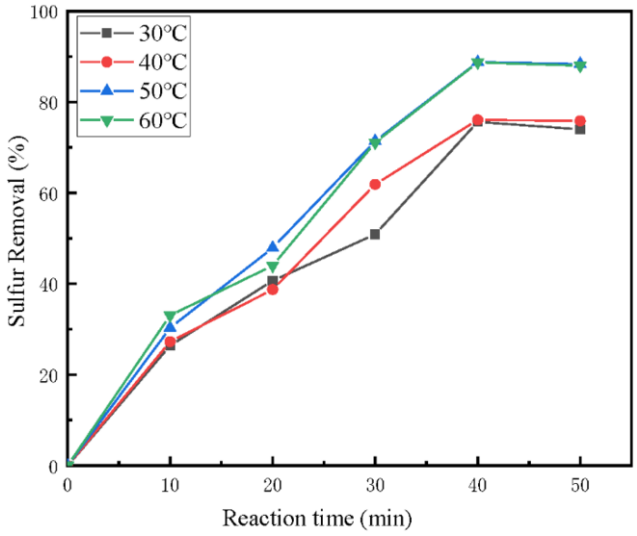


Figure 4. Influence of reaction time and temperature on desulfurization effect.

In summary, under the conditions of simulated fuel 20 mL, isopropanol 10 mL, acetonitrile 10 mL, the catalyst 0.1g, the effect of temperature on the desulfurization effect of simulated fuel containing BT generally shows that the desulfurization efficiency increases with the increase of temperature.The specific performance is 50℃ > 60℃ > 40℃ > 30℃, indicating that temperature is a decisive factor to determine the reaction.The desulfurization efficiency at 30℃ and 40℃ is significantly lower than that at 50℃, indicating that the higher the reaction temperature, the higher the oxidation desulfurization rate, which is mainly because the temperature increases, the speed of molecular movement is accelerated, and the effective collision frequency between catalyst and thiophene sulfide molecules is increased.According to Vant Hoff's law, the reaction rate increases 2-4 times with every 10 K increase of reaction temperature [7]. The increase of temperature will accelerate the mass transfer between phases, increase the activity of catalyst, and increase the content of in-situ peroxide. Therefore, the desulfurization rate shows an upward trend with the increase of temperature.When the temperature is 60℃, the desulfurization efficiency of the first 30 min is higher than that of the temperature 50℃, but after 30 min, the desulfurization efficiency is lower than that of the temperature 50℃.This reason is the oxidation reaction is exothermic and exothermic, and the rise of temperature will inhibit the reaction forward, reduce the reactive center, and partially decompose the peroxide generated in situ [18], thus reducing the desulfurization rate, that is, when the time is 40 min and the temperature is 50℃, the desulfurization effect is optimal.

(2) The effect of different catalyst content

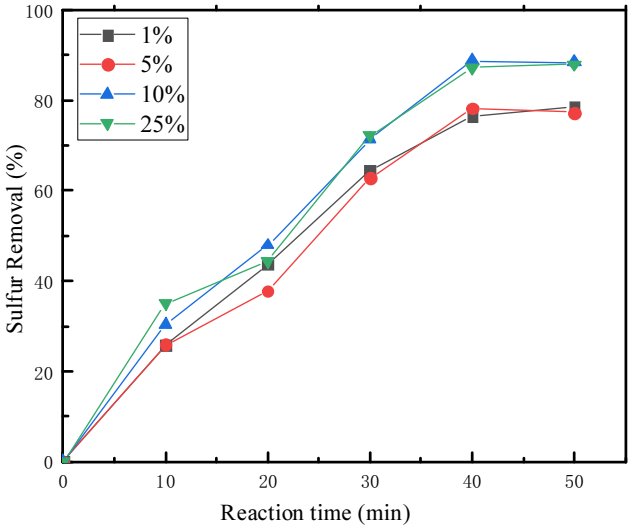


Figure 5. The effect of different catalyst content.

It can be seen from the figure that 25% catalyst and 10% catalyst have the best desulfurization effect, reaching 88.77% and 88.70%. Desulfurization efficiency load 10% catalyst >25% catalyst >5% catalyst >1% catalyst.The desulfurization efficiency of 10% catalyst and catalyst is more than 88%,almost achieve deep desulfurization effect.

(3) Influence of Pd/MIL-101 (Cr) with different mass on desulfurization effect

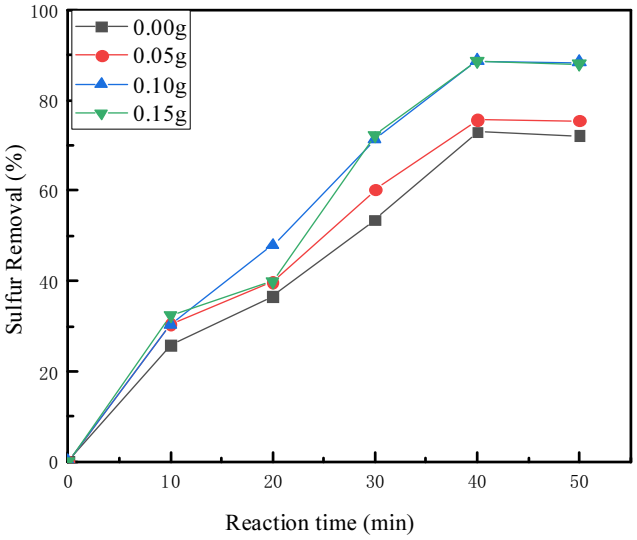


Figure 6. Influence of Pd/MIL-101 (Cr) with different mass on desulfurization effect.

In the above experiments, the catalyst was selected as the desulfurizer to carry out the experiment. The results show that with the increase of load percentage, the desulfurization rate of catalyst desulfurizer is higher and higher.When the dosage of 10% catalyst was 0.10g, the desulfurization efficiency reached the highest of 88.77%.

This may be because when Pd/MIL-101 (Cr) is used to a certain amount, the intermediate pore structure of the new catalyst, there are many active sites for oxidizing isopropanol in the pore channels.The size and number of crystal particles, as well as their dispersion on the inner surface of the catalyst all have influences on desulfurization efficiency to a certain extent.Increasing the amount of the catalyst can increase its active site and speed up the reaction efficiency.

(4) The influence of initial sulfur concentration on desulfurization effect

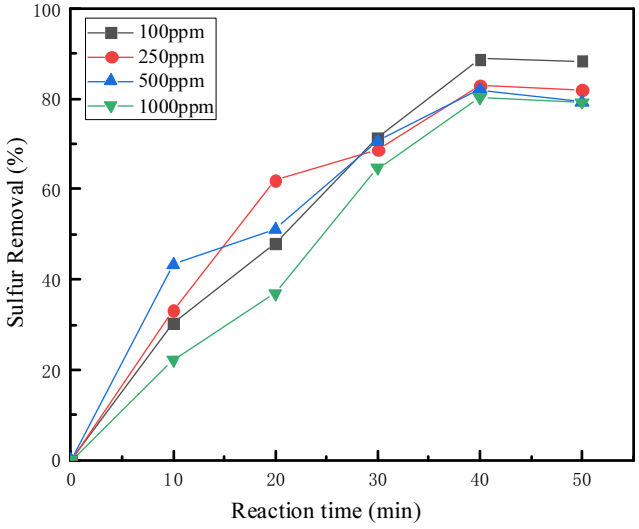


Figure 7. Influence of initial sulfur concentration on desulfurization effect.

As can be seen from the figure, the desulfurization efficiency of different initial concentrations of sulfur content is also different. With the increase of the initial sulfur concentration, the desulfurization rate decreased. This is due to the presence of sulfones on the surface of the catalyst. It should be noted that BT sulfone is insoluble in the model mixture, so the above experimental phenomenon occurred.

(5) The influence of catalyst recirculation on BT conversion rate

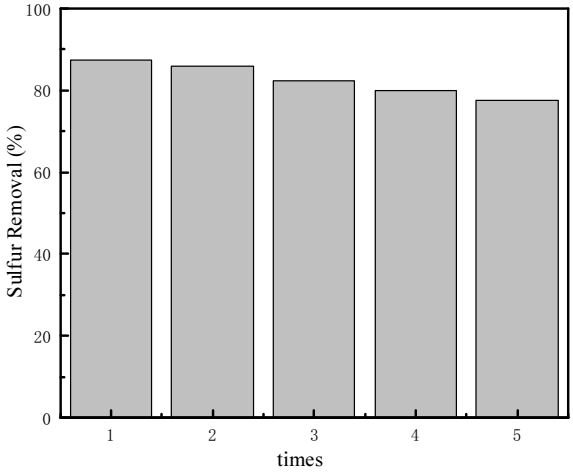


Figure 8. Influence of catalyst recirculation on BT conversion rate.

Recycling performance is an important factor that needs to be studied.In the desulfurization system, the volume ratio of acetonitrile and isopropanol is 1:1.After each reaction, the upper layer is the oil phase and the lower layer is the acetonitrile phase.The model oil can be separated from the acetonitrile phase system by catalytic oxidation.The results show that the desulfurization rate of BT simulation fuel decreases by 9.90% in the first five cycles of experiments.

The degradation of desulfurization performance after the catalyst recycling is due to the catalytic oxidation of thiophene sulfides into corresponding sulfoxides or sulfoxides, which coat the surface of the catalyst grains and block the active point of contact with thiophene sulfides. In addition,In the recycling process of the catalyst, there is a loss of its active point. At the same time, in the process of recycling and regeneration experiment, there will be trace loss of the catalyst, so the desulfurization efficiency of the catalyst decreases obviously after 5 times of recycling.

4. Conclusion

After the synthesis of MIL-101 (Cr) by hydrothermal method, the catalyst was synthesized by impregnation method, which can achieve the effect of deep desulfurization at low temperature and atmospheric pressure. Reaction time is 40 minutes, This material had excellent desulfurization performance.MIL-101 (Cr) loaded with Pd salt was used as the catalyst with high catalytic activity. The effect of deep desulfurization was studied by adjusting other factors such as different time, temperature, catalyst quality, different catalyst content. The results showed that when dibenzothiophene was used as the model fuel, the temperature was 50 ℃ and the time was 40 minutes, The salt type was Pd (CH3COO) 2 and the loading is 10 wt%.Taked 0.1g of the catalyst can reach the fuel oil from 108.12 mg / kg to 12.14 mg / kg,it had good desulfurization performance.

References

[1] Hossain M, Park H, and Choi H. A Comprehensive Review on Catalytic Oxidative Desulfurization of Liquid Fuel Oil. Catalysts. 2019; 9(3).
[2] Oliver K, Widmer M, Elsener M, et al. Adsorption and Desorption of SOx on Diesel Oxidation Catalysts. Industrial & Engineering Chemistry Research. 2017; 48(22): pp 9847-9857.
[3] Piccinino D, Abdalghani I, Botta G, et al. Preparation of wrapped carbon nanotubes poly(4-vinylpyridine)/MTO based heterogeneous catalysts for the oxidative desulfurization (ODS) of model and synthetic diesel fuel. Applied Catalysis B Environmental. 2017; 200: pp 392-401.
[4] Marcatti G E, Resende R T, Resende M, et al. GIS-based approach applied to optimizing recommendations of Eucalyptus genotypes.Forest Ecology and Management. 2017; 392: pp 144-153.
[5] Moulder J F, Chastain J, King R C et al. Size and morphological control of a metal–organic framework Cu-BTC by variation of solvent and modulator. Journal of Porous Materials. 2014.
[6] MOULDER J F, STICKLEWF, SOBOLP E, et al.Handbookof X-ray photoelectron spectroscopy. Minnesota of USA:Perking Elmer Corporation. 1990.
[7] Tian Y, Wang G, Long J, et al. Ultra-deep oxidative desulfurization of fuel with H2O2 catalyzed by phosphomolybdic acid supported on silica. Chinese Journal of Catalysis. 2016; 37(12): pp 2098-2105.

doi:10.3233/ATDE220246

Microstructure and Properties of Ni-Based Alloy Coatings by High-Efficiency Intelligent Cladding

Yingchun WANG[a], Shiquan JIANG [a], Xiaomin MA[b,1], Changsheng ZHAI[c], Guodong ZHENG [a], Ping WANG[a], Zan WANG [a], Fang XIE [c] and Jian ZHANG[a]
[a]*Joint R&D Center for Wear-resistant and Corrosion-resistant Petroleum Parts,Henan University of Technology,Zhengzhou 457000,China*
[b]*National Equipment New Material & Technology (Jiangsu)Co.,Ltd*
[c]*Nanyang Institute of Technology, Nanyang 473004, China*

Abstract. In this paper, an versatile and potential surface remelting strengthening technology, high-efficiency intelligent cladding (HEI-cladding) is presented. Ni-based alloy coatings were prepared by HEI-cladding. The microstructures of the cladding coatings were investigated by scanning electron microscopy, and the compositions were analyzed by electron diffraction spectroscopy. The phases of the cladding coating were clarified by X-ray diffraction technology. The microhardness of the coating was tested by microhardness tester. It shows that cladding coatings at 1298K are denser and have fewer defects, while the number of the hard phases such as bulk structure, carbide and borideis are obviously increased. It shows that the grain size is much smaller than the overmelting temperature, the interfacial defects decreased significantly. The diffusion transfer band (DTB) width is 10-15um at the interface of the coating prepared at 1298K, which results in a good metallurgical reaction at the interface. The microhardness results show that the cladding coating at 1298K has a smaller grain size, higher hardness in DTB near the interface and better microhardness distribution.

Keywords. High-efficiency intelligent cladding, ni-based alloy coating, reduction remelting, microhardness

1. Introduction

In recent years, high-frequency induction remelting cladding has become one of the most concerned methods of cladding treatment, whose essence is to heat the surface coatings of metal substrates by using electromagnetic eddy current, and then obtains good-performance cladding coatings[1-3]. Because of its high power density, fast heating speed and easy-control of temperature, time and heating depth, it has great advantages over traditional flame spray welding, plasma cladding and laser cladding in improving coating quality, production efficiency and working environment and reducing damage to work-piece[4-9]. However, the existing pure induction cladding equipment still relies on manual operation. Over-melting and under-melting are prone to occur in the cladding process, resulting in poor stability and consistency of the

[1] Corresponding Author, Xiaomin MA, National Equipment New Material & Technology (Jiangsu)Co., Ltd; Email: 1901794848@qq.com.

coating quality, which makes induction cladding not widely applied. Through the cross-disciplinary integration of thermal spraying technology, induction heating technology, intelligent control technology, and human-computer interaction technology, a patented high-efficiency intelligent cladding technology (HEI-Cladding Tech.) has been successfully developed by our resarch teams[10-12]. The human computer interaction function, intelligent tracking control, high-efficiency, low-cost operation in the production process is realized, and special metals on the surface is fully melted. Thus it has been developed into an ideal coating refining mode, which strongly promotes the deep development of induction cladding technology. In this paper, the effect of different cladding temperature on the microstructure of HEI-Cladding ni-based alloy coating was studied to obtain the optimized coating structure and mechanical properties.

2. Experimental procedures

In this experiment, GCr15 steel was used as the matrix, and NF201 nickel-based alloy powders are been chosen as coating material, which are produced by Wuhan HANGAUN New Material Sci. & Tech. Co., Ltd. The chemical composition of NF201 powders is shown in Table 1. NF201 nickel-based alloy Coatings has the surface of GCr15 steel by HEI-Cladding. The specifications of the samples are φ80mm*100mm. The process mainly includes: Sandblasting treatment to matrix surface, high-energy flame spraying pre-preparing coating, HEI-Cladding treatment. HG-FMS-II high-energy flame spraying system and HG3000 HEI-Cladding system were used, which are independently developed by Henan HANGAUN Mechanical Reman Tech. Co., Ltd. The process parameters of high-energy flame spraying are shown in Table 2. The preparation process of HEI-

Table 1. Chemical components of NF201 ni-based alloy powers (mass fraction/%)

C	Cr	Si	B	Fe	Ni
0.5-1.0	14-19	3.5-5.0	3.0-4.5	< 8.0	allowance

Table 2. Processing parameters of high-energy flame spraying NF201 ni-based alloy coating

Oxygen pressure /MPa	Acetyle pressure /MPa	Oxygen flow Rate /$m^3 \cdot h^{-1}$	Acetylene flow Rate /$m^3 \cdot h^{-1}$	Air flow Rate /$m^3 \cdot h^{-1}$	Powder delivery /$Kg \cdot h^{-1}$	Spraying distance /mm
0.45	0.09	1.6	0.8	1.5	6	185

Cladding coating is a temperature feedback control process. According to the cladding temperature of the coating material, through the adaptive PID closed-loop control of induction heating power and set cladding temperature, the power output of induction heating can be automatically adjust through PLC, so as to achieve the consistency of the cladding layer quality. The HEI-Cladding temperature of NF201 material is set at 1233K, 1298K and 1313K, respectively, with an accuracy of ±3K.

The phase of coating was analyzed by D-MAXLL X-ray diffractometer. The microhardness of the coating was tested by HXD-1000 microhardness tester. The indenter load was 500 grams and the holding time was 15 seconds. Quanta 250 FEG field emission scanning electron microscopy (SEM) produced by Dutch FEI company

was used to observe the microstructure and X-ray energy dispersive spectroscopy (EDS).

3. Results and discussion

3.1. Microstructure detection

Fig.1 shows the cross-section microstructure and EDS line scanning results of NF201 coatings prepared by high-energy flaming spraying. It shows that closed, crossed and continuous cracks were found in the sprayed coatings and interfaces. Fe and Ni showed steep drop characteristics at the interface, indicating the multi-defect characteristics of the sprayed coatings and the mechanical bonding characteristics of the interface. XRD diffraction pattern shows that hard phases such as Fe_7C3, $Cr7C_3$, $Fe_3(C,B)_6$ carbides and borides appeared in the fabricated high-energy flame sprayed coating

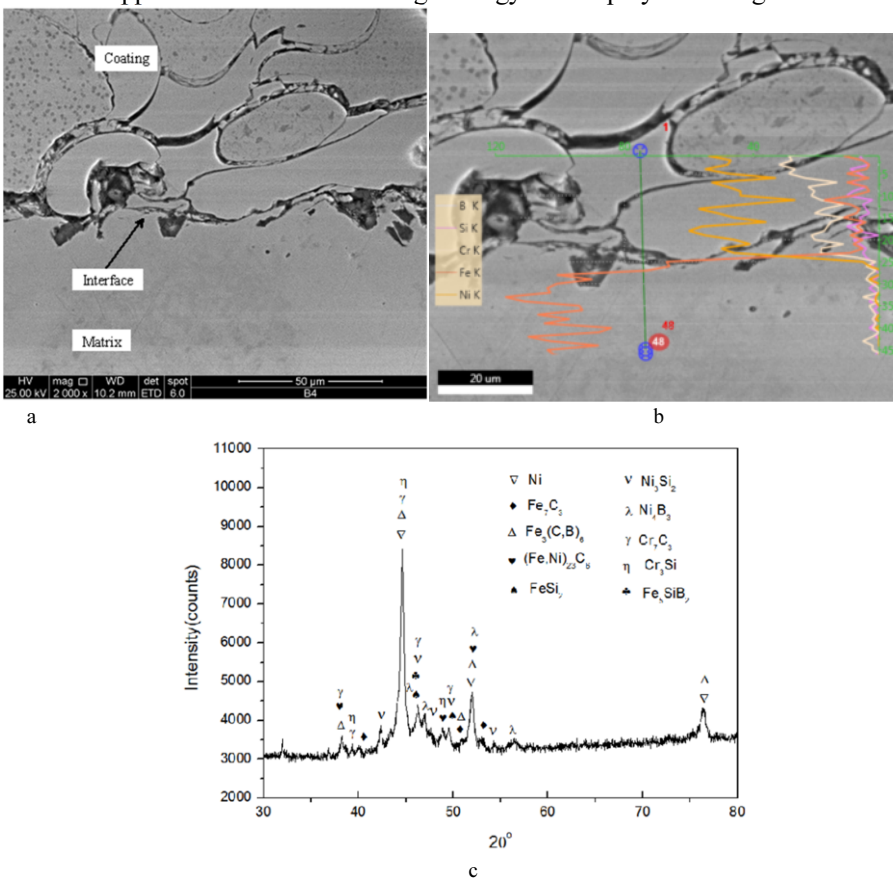


Figure 1. (a) SEM and (b) EDS images (c) XRD diffraction pattern of theNF201coating by high-energy flame spraying.

3.2. Microstructure of the cladding coating surface

Microstructure of the cladding coatings at different temperatures are shown in Fig. 2. The three coatings are composed of Ni solid solution (white), eutectic structure (grey) of Ni and other alloys, and dispersed hard phases such as bulk structure, carbide, boride(black particles)[13-16] . However, there are essential differences among the three kinds of cladding coatings. The amount of Solid solution and eutectic of the coating obtained at the cladding temperature of 1233K is much larger than that of 1298K, while the hard phase is relatively small. The carbide grains of the coatings prepared at the cladding temperature of 1313K grew obviously, and a large number of holes were distributed in the coating. The size of the holes was much larger than that of the other two kinds of cladding coatings. Compared with the other two temperatures, the number of black particles (hard phases such as bulk structure, carbide and boride) increased significantly at the cladding temperature of 1298K, and the grain size was much smaller than that of the overmelting temperature, with fewer defects in the coating, the surface microstructures were obviously improved.

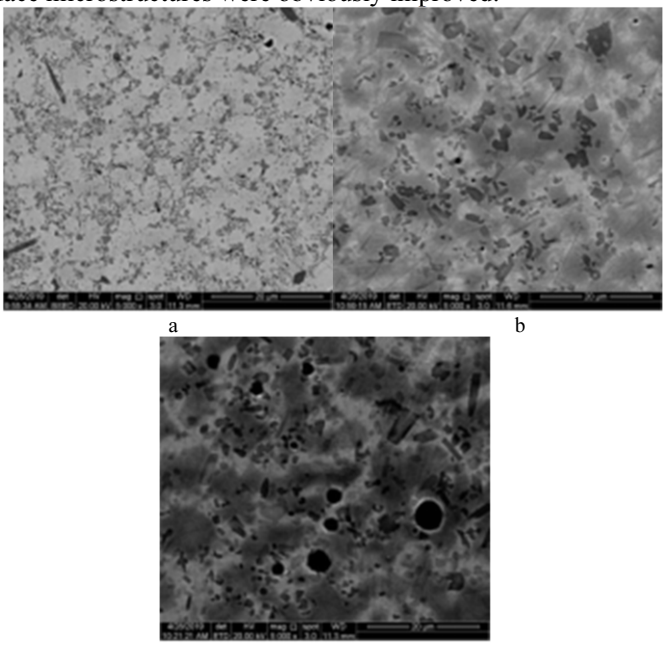


Figure 2. The surface morphology of the coatings at different cladding temperatures:(a) 1233K, (b) 1298K and (c) 1313K.

Figure 3 shows the XRD spectra of the coatings obtained at different cladding temperatures. These cladding coatings have different phase systems from those of high-energy flame spraying coatings. The phase of the cladding coating prepared at 1233K is composed of Ni solid solution, various soft phases ($B(Si,Fe)_3$, $FeSi_2$, $Fe_{4.5}Ni_{18.5}B_6$, $Fe_2Si_{0.4}B_{0.6}$, Ni_3Si_2, $Cr_3Ni_5Si_2$, Cr_3Si_2, Ni_3C, Ni_3Si), and hard phases such as Cr_7BSi_4、Fe_3C、Fe_7C_3、Cr_7C_3、$(Cr,Fe)_7C_3$.The non-hard phase of the cladding coating obtained at 1313K increases obviously. The cladding coatings at 1298K are obviously different from the former two. The main phases of the coatings are Ni solid solution and hard phases of Fe_7C_3,$Fe_{23}(C,B)_6$, $(Cr,Fe)_7C_3$,Cr_7C_3, and CrB4, and there are a few

non-hard phases of Ni_3Si_2、Ni_4Si_3、$Cr_3Ni_5Si_2$.The difference of microstructures and phases of the three coatings depends on the induction heating electromagnetic field formed at different cladding temperatures [17,18]. Only appropriate cladding temperature can promote the full reversal of Nickel-based Self-fluxing alloys.

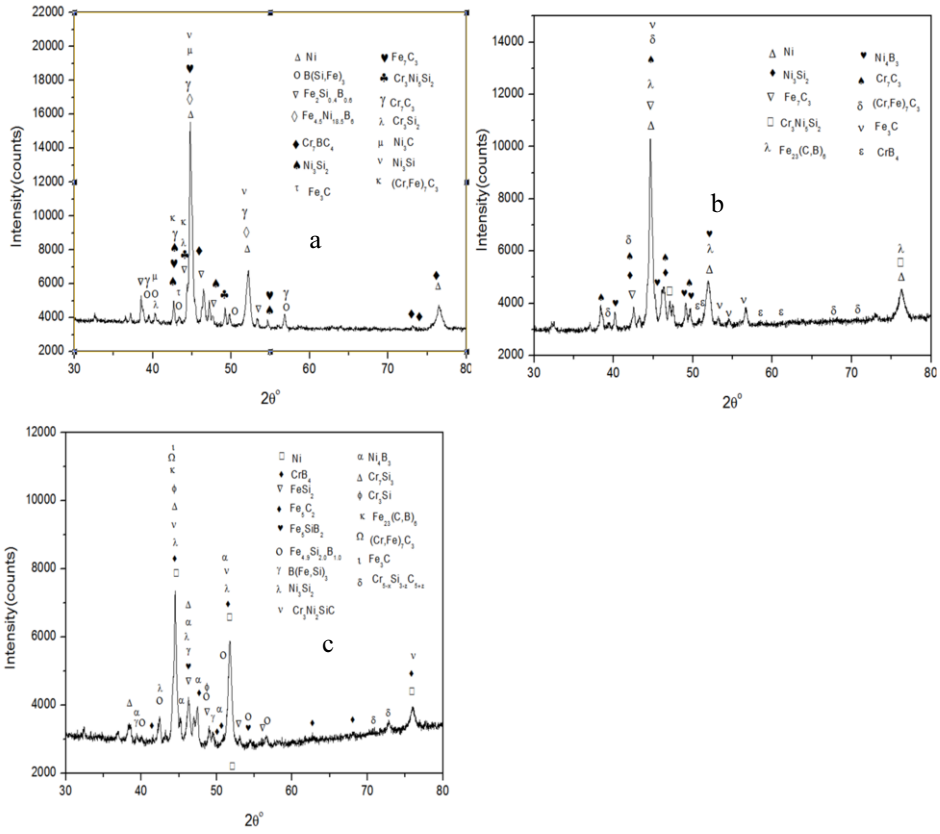


Figure. 3 The XRD patterns of the coatings at different cladding temperatures: 1233K, (b) 1298K and (c) 1313K.

3.3. Micro-area components and cross-section microstructure of the cladding coating

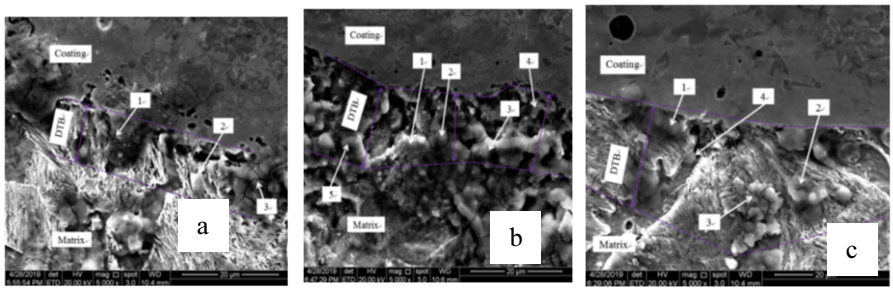


Figure.4. Cross-section microstructure of the HEI-Cladding coatings at: 1233K, (b) 1298K and (c) 1313

The SEM photo of the cross-section of the coating at three cladding temperatures are given in Fig.4. There are many holes and defects in the interface of the coating at 1233K, as indicated by arrows(1-3) in Fig. 4(a). It shows that the width of the diffusion transfer band (DTB) formed between the matrix and the coating is less than 8um, and there are a large number of granular structures in DTB. DTB width is relatively small.

As shown by arrows(1-4) in Figure 4(b), no obvious defects such as interface cracks and voids were found at the interface of the cladding coating at 1298K. The results show that DTB with a width of 10-15um is formed by the cross-distribution of Ni and Fe at the interface caused at 1298K, which results in a good metallurgical reaction at the interface. There are interfacial cracks and densely distributed holes at the interface of the coating prepared at 1313K, as indicated by arrows(1-3) of Figure4 (c). There are undulating structures in DTB. The width of DTB is not less than 20um. The DTB at the interface between the matrix and the cladding coating is the result of the diffusion of elements in the cladding coating to the matrix at high temperature and the retention of alloying elements in the matrix after rapid cooling [19]. The results show that the width, element diffusion and distribution of DTB are closely related to the cladding temperature, at a certain frequency[20].

3.4. Microstructure of the cladding coating surface

The structure of the coating prepared at the optimized cladding temperature is mostly non-equilibrium γ-tough solid solution and γ-solid solution+carbide eutectic, which is dispersed among dendrites, and the content of intermetallic compounds such as alloy cementite is small. The combination of carbides and chromium carbides results in high microhardness and small hardness extreme value of the coatings under the optimized process conditions.

Fig. 5 shows the cross-section hardness distribution of the coatings at different cladding temperatures. From Fig. 5, it can be seen that the microhardness of the cross-section of the sample shows a stepped distribution at different melting temperatures. After re-melting at 1233K, 1298K and 1313K, the average cross-sectional

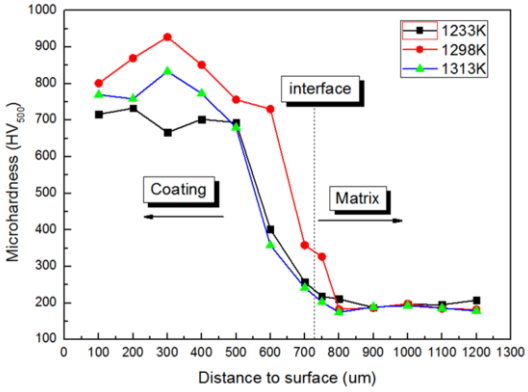


Figure 5. Hardness curves of the HEI-Cladding layers at different temperature.

hardness of the nickel-based coatings were 723.36HV_{500}, 850.32HV_{500} and 754.68HV_{500}, respectively. The average cross-sectional microhardness of the coatings prepared at 1298K was significantly higher than that of the other two coatings. Near the interface, the microhardness of the materials decreased significantly, and the matrix far

from the interface had similar microhardness. It can be concluded that the cladding coating at 1298K has better microhardness distribution.

4. Results and discussion

(1) The high-energy flame sprayed coatings were flattened obviously, with a lot of cracks and holes. There occurred the multi-defect characteristics of the sprayed coatings and the mechanical bonding characteristics of the interface. The three coatings are composed of Ni solid solution, eutectic structure of Ni and other alloys, and dispersed hard phases such as bulk structure, carbide and boride. The cladding coatings at 1298K are denser and have fewer defects, while the number of hard phases is significantly increased, the grain size is much smaller than the overmelting temperature, and the surface microstructures are obviously improved.

(2) Compared with the other two cladding temperatures, the interfacial defects of the coating prepared at 1298K decreased significantly. The diffusion transfer band (DTB) is formed at the interface. DTB width is 10-15um at the interface of the coating prepared at 1298K, which results in a good metallurgical reaction at the interface.

(3) The average microhardness of the cladding coating prepared at 1298K is 822.1HV_{500}, which is higher than that of the coating at 1233K and 1313K, respectively. The cross-section microhardness distribution of coating is ladder-like. Compared with 1233K and 1313K, the cladding coating at 1298K has a smaller decrease, higher hardness in DTB near the interface and better cross-section microhardness distribution.

References

[1] Ohmori A, Takasaki N, Tomiguchi A. Fusing of Sprayed Ni-base Coatings by Induction Heating(Physics, Process, Instruments & Measurements. Transaction of JWRI. 1992; 21: pp 195-200
[2] Yancong L. Simulation on the high frequency induction cladding and experimental validation. Journal of Information and Computational Science. 2015; 12(2): pp 641-56.
[3] Bensheng H, Yuxiao G, Peng C. Microstructural and mechanical properties of Co-based coating by high-frequency induction cladding. Materials Review. 2018; 32(13): pp 2272-7.
[4] Cheng-kai L, Yan-cong L, Yong-jun S. Modeling of high-frequency induction heating surface cladding process: numerical simulation, experimental measurement and validation. In: Proceedings of the 6th International Asia Conference on Industrial Engineering and Management Innovation. 2015; pp 747-59.
[5] Afanasieva L E, Ratkevich G V. Laser cladding of NiCrBSiFe-WC coating with laser. Letters on Materials. 2018; 8(3): pp 268-73.
[6] Chinmaya K S., Manoj M., Effect of pulse laser parameters on TiC reinforced AISI 304 stainless steel composite coating by laser surface engineering process. Optics and Lasers in Engineering. 2015; 67: pp 36-48.
[7] Hyung-Jun K,Byoung-Hyun Yoon,Chang-Hee Lee. Wear performance of the Fe-based alloy coatings produced by plasma transferred arc weld-surfacing process. Wear. 2001; 249: pp 846-852.
[8] Bergant Z, Trdan U, Grum J. Effect of high-temperature furnace treatment on the microstructure and corrosion behavior of Ni-Cr-B-Si flame-spraying coatings. Corrosion Science. 2014; 88: pp 372-386.
[9] Shibang M, Zhihao D, Zhenwei X. Parameter Optimization and Microstructure Evolution of In-Situ TiC Particle Reinforced Ni-based Composite Coating by Laser Cladding. Journal of Engineering Science and Technology Review. 11 (2), 2018; pp 88- 95.
[10] Chang-sheng Z, Yingchun W, Fang X. Effect of Remelting Temperature on Tribological Properties of Nickel-based Coatings Prepared by High Efficiency Intelligent Induction Cladding. Transactions of Materials and Heat Treatment. 2019; 40(11): pp 166-76.
[11] Chang-sheng Z. A man-machine interactive system of induction preheating cladding integrated device. China: ZL 201521102964.3. 2018.

[12] Chang-sheng Z. An intelligent operation-control system of an integrated induction preheating cladding device [P]. China: ZL 201521102962.4. 2018.
[13] Hong-xie Z, Hong-li Z, Cheng L. Effect of induction cladding technology on quality of Ni60 cladding layer. Special Casting & Nonferrous Alloys. 2008; 8: pp 129-131.
[14] Helong Y, Min W, Mengqing Z. Microstructure Characteristics and Nano Mechanical Properties of In-situ TiC/Ti Composite Coating by Induction Cladding. China Surface Engineering. 2018; 31(5): pp 150-8.
[15] [15] Xin W, Gui Qin W, Yong Feng C. Study on WC-Ni60 Complex Coating by High Frequency Induction Cladding. Materials Science Forum. 2011; 675-677: pp 1299-302.
[16] Zeng-zhi Z, Gui-quan H, Yue-wen F. Controlling means for the distribution of eddy current in coating by induction cladding. Heat Treatment of Metals. 2003; 28(7) : pp 58-61.
[17] Ying M, Xiao-peng Z , Peng-cheng D. High Frequency Induction Remelting of Ni60-40WC Coating Prepared by High Velocity Oxy-Fuel Spraying and Its Water-Lubricated Wear Resistance. Materials Protection' 2018; 1(12): pp 375-379.
[18] Chang JH, Chou JM, Hsieh RI. Influence of fusing temperature on microstructure, wear and corrosion resistance of induction melted bimetal of Co-based alloy and AISI 4140 steel. Materials Chemistry and Physics , 2009, 118: 314-21.
[19] Rui S, Qi L, Rui-Hai W. Microstructure and Mechanical Properties of High-Frequency Induction Cladding Ni-Based Alloy Coating with La2O3 Addition. Materials Science Forum. 2018; 934: pp 111-6.
[20] Zhou SF, Dai XQ. Microstructure evolution of Fe-based WC composite coating prepared by laser induction hybrid rapid cladding. Applied Surface Science. 2010; 256: pp 7395-99

doi:10.3233/ATDE220247

Lightweight Design of Hinged Beam Structure Based on Agent Model

Guizhong XIE[a], Shixin ZHANG[a,1], Liangwen WANG[a], Xiaoyun GONG[a], Tao WANG[a], Shuguang WANG[b], Zhiqiang CHEN[b], Chongmao ZHAO[a] and Hangqi JIA[a]

[a]*Mechanical and Electrical Engineering Institute, Zhengzhou University of Light Industry, Zhengzhou, 450002, Henan, China*
[b]*Henan Huanghe whirlwind Co.. LTD, Changge, 461500*

Abstract. This paper presents an optimization method based on the Kriging and multi-objective genetic algorithm. Firstly, taking the hinged beam structure of the cubic diamond press as the design object, the optimization design mathematical model is established with the quality as the objective function, the stress peak and displacement peak as the constraint conditions. Secondly, in combination with SolidWorks and ANSYS Workbench, parametric modeling analysis was conducted to obtain a large number of sample points sparing less time, and the agent model constructed by Kriging was trained and verified. Finally, taking advantage of global search of the multi-objective genetic algorithm, the lightweight design is realized and the quality of the hinge beam structure is effectively reduced, which would be a guiding significance for the lightweight design of other mechanical parts.

Keywords. Hinge beam structure, lightweight, finite element analysis, sensitivity analysis, the agent model

1. Introduction

With the rapid development of industry and the promotion of national policies in recent years, the market demand for diamond products is increasing, and the production process and application equipment of diamond manufacturing are constantly optimized and improved. In order to obtain higher yield and synthesize diamond of higher grade and larger grain, the large-scale step of the cubic diamond press is still advancing[1]. The large-scale mechanical structure also means the improvement of structural design rationality, structural safety and reliability. so the lightweight design of the hinge beam structure of the cubic diamond press is of great significance [2].

Agent model technology has been widely used in the optimization of engineering problems, and many scholars have carried out research on the optimization of agent model. Qingping Zhang[3] used multi-objective genetic algorithm to optimize the bed structure. Kang Di[4] obtained the optimal solution by using non-dominated sorting genetic algorithm on the basis of Kriging agent model. Di Trapani Fabio[5] proposed a new specific optimization framework for minimization problems and defined a new

[1] Corresponding Author, Shixin ZHANG, School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, E-mail: 1262747467@qq.com.

genetic algorithm, which can solve reconstruction optimization problems. Zhang Jian [6] adopted ant colony algorithm to optimize rail size and proposed a lightweight optimization idea based on TC4 titanium alloy frame structure design. All these studies provide ideas for the optimization of hinged beam structure.

In this method, firstly, the correlation analysis is carried out, and the structural parameters that have a greater impact on the output response are selected as the input variables, and the mass, volume stress peak and displacement peak of the hinge beam are taken as the output variables to establish a mathematical optimization model. Secondly, the optimal space design method is used to obtain the initial sample points, and parametric modeling was carried out with the SolidWorks and ANSYS Workbench, to obtain multiple groups of sample points. Finally, sample points were used to fit the Kriging agent model and verify its accuracy. After the fitting accuracy of the agent model reached the standard, multi-objective genetic algorithm was used to conduct global search and iterate to pursue the optimal design. Finally, the lightweight design of the hinged beam structure was completed.

2. Correlation analysis

Thirty groups of sample combination points were obtained through eight design variables. Correlation analysis was conducted on each design variable according to the input and output of the thirty groups of sample points to obtain the degree of influence of each design variable on the output. The correlation histogram is shown in Figure 1. Correlation analysis is based on Spearman's Rank Correlation statistical method. It calculates correlation coefficients by sorting the values of sample variables. It is applicable to the correlation between variables with nonlinear monotone change function relationship. The closer the correlation coefficient is to 1, the stronger the quadratic correlation is.

Figure 2. shows the correlation matrix between input and output. The darker the color in the matrix, the higher the correlation between the two parameters. According to the calculated correlation coefficient, the lug thickness (x_1, x_2), inner diameter(x_3), inner wall depth (x_4)and the lug length（x5）of the hinge beam are selected as the design variables of the lightweight of the hinge beam after filtering.

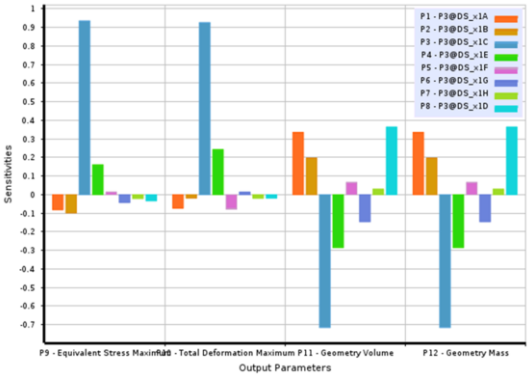


Figure 1. Sensitivity histogram.

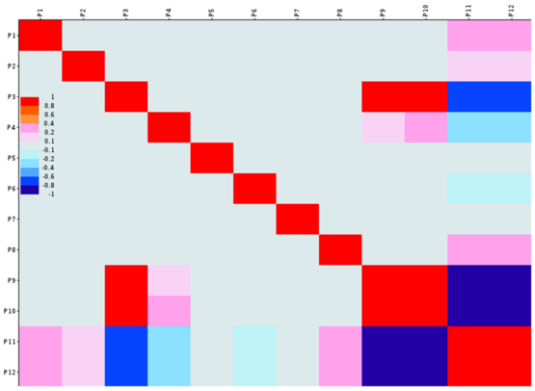
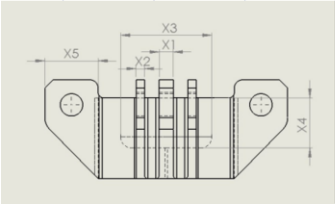


Figure 2. Correlation matrix diagram.

3. Establishment of mathematical model for structural optimization of hinged beam

Hinge beam structure is one of the key components of the cubic diamond press, which is of great significance to its lightweight. The optimization model of hinged beam structure is established as follows[7].

Firstly, according to the correlation analysis results, the final design variables selected are shown in Figure 3.



X1	Hinge beam lug thickness 1 (large)
X2	Hinge beam lug thickness 2 (small)
X3	Inner diameter of hinge beam
X4	Inner wall depth of hinge beam
X5	Lug length of hinge beam

Figure 3. Hinge beam structure and its design variables.

Secondly, the objective function of hinge beam lightweight design is to minimize the mass of hinge beam on the premise of satisfying its performance. Therefore, the quality of the hinged beam is taken as the objective function:

$$\min f(x) = f(x) = m(x_1,\ x_2,\ x_3,\ x_4,\ x_5) \tag{1}$$

Finally, the strength and stiffness of hinged beam structure will change with the size of each part, so the strength and stiffness should be checked, that is, the strength condition should meet the stress constraint, and the stiffness condition should meet the displacement constraint. Most domestic manufacturers use 35CrMo as the material for casting hinge beams. At room temperature, its elastic modulus $E = 200000MPa$, Poisson's ratio $\mu = 0.3$, yield strength $\sigma_s = 730MPa$ and allowable displacementδ=1.5mm, and its safety factor is 1.5, then the stress constraint function and displacement constraint function of the material are respectively:

$$g_1(x) = \sigma_x - \frac{[\sigma_S]}{1.5} \leq 0 \tag{2}$$

$$g_2(x) = \delta_x - [\delta] \leq 0 \tag{3}$$

Boundary constraints are the value range of design variables. According to the size standard, the value range of each design variable is:

$$\begin{cases} 80 \leq x_1 \leq 120 \\ 40 \leq x_2 \leq 80 \\ 500 \leq x_3 \leq 800 \\ 300 \leq x_4 \leq 420 \\ 320 \leq x_5 \leq 440 \end{cases} \tag{4}$$

4. FEA of hinge beam

Through the static analysis of the hinge beam structure by ANSYS finite element analysis software, the stress distribution of the hinge beam structure in general state can be obtained[8~11]. The linear tetrahedral quadratic isoparametric element is used to divide the finite element mesh because it has a good effect on boundary fitting. The finite element model of the hinge beam is divided into 54588 elements and 95367 nodes. A full constraint was applied to the lug through hole of the hinged beam, and a working oil pressure of 100MPa was applied to the beam wall.

The hinged beam model is solved and the displacement and stress cloud maps of the hinged beam are obtained, as shown in Figure 4. The maximum displacement of the hinged beam is distributed at the center of the inner bottom of the beam, which is 0.67059mm, and the maximum stress is located at the lug through hole of the hinged beam, which is 457.62MPa. The stress at the transition fillet at the bottom of the hinge beam is about 391.98MPa. In practical production, the fracture of hinge beam often occurs at the bottom transition fillet and lug through hole. The finite element analysis results are consistent with the actual situation, which verifies the rationality of the finite analysis results.

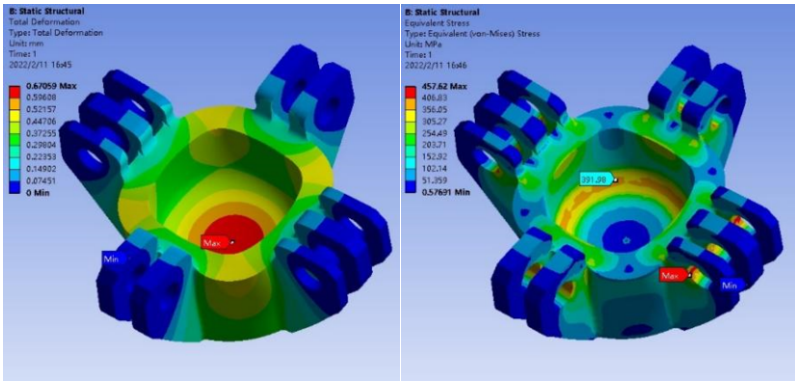


Figure 4. Displacement and stress cloud of hinge beam.

5. Agent model optimization

5.1. Design of experiment

In this paper, the optimal space filling design sampling method is adopted to carry out reasonable sampling in the feasible region composed of the value range of design variables[12]. optimal space filling sampling method(OSF), with its maximum interpoint distance and more uniform point distribution, solves the extremum problem more effectively, provides better coverage for design space, and ensures the filling of design space.

The SolidWorks parametric design combined with ANSYS Workbench finite element analysis, finally obtained 130 groups of design variables and quality, stress peak and displacement peak output sample points.

Table 1. Sample points obtained in DOE experiment.

Design point	X1	X2	X3	X4	X5	P9(MPa)	P10(mm)	P11(mm^3)	P12(KG)
1	90.2	74.6	564.5	407.4	411.8	388.78	0.59951	4.9934E+08	3919.8
2	103.4	47.8	789.5	324.6	353	775.72	1.2642	4.138E+08	3248.3
3	85.8	78.6	558.5	349.8	357.8	340.2	0.47432	4.9003E+08	3846.7
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
130	119.76	76.6	793.01	303.38	418.2	595.59	0.9241	5.0039E+08	3928

5.2. Establishment of agent model

Kriging agent model is a kind of interpolation-free estimation model, which can not only accurately and quickly predict the response value of the unknown point, but also measure the model's inaccuracy performance[13]. It has good approximation ability to nonlinear problems. Therefore, the Kriging agent model was used to construct the response surface, with the five design variables of the hinge beam model as inputs, and the stress peak (P9), displacement peak (P10), volume (P11) and hinge beam mass (P12) as outputs, to establish the relationship between inputs and outputs. 100 groups of design points were selected for the training of the agent model, and 30 groups of design points were used to verify the accuracy of the response surface.

	A	B	C	D	E
1		P9 - Equivalent Stress Maximum	P10 - Total Deformation Maximum	P11 - Geometry Volume	P12 - Geometry Mass
2	Coefficient of Determination (Best Value = 1)				
3	Learning Points	1	1	1	1
4	Root Mean Square Error (Best Value = 0)				
5	Learning Points	2.4515E-06	1.3435E-07	0.45788	3.5943E-09
6	Verification Points	2.1558E-06	1.4351E-07	0.22039	1.73E-09
7	Relative Maximum Absolute Error (Best Value = 0%)				
8	Learning Points	0	0	0	0
9	Verification Points	0	0	0	0
10	Relative Average Absolute Error (Best Value = 0%)				
11	Learning Points	0	0	0	0
12	Verification Points	0	0	0	0

Figure 5. Precision evaluation of agent model.

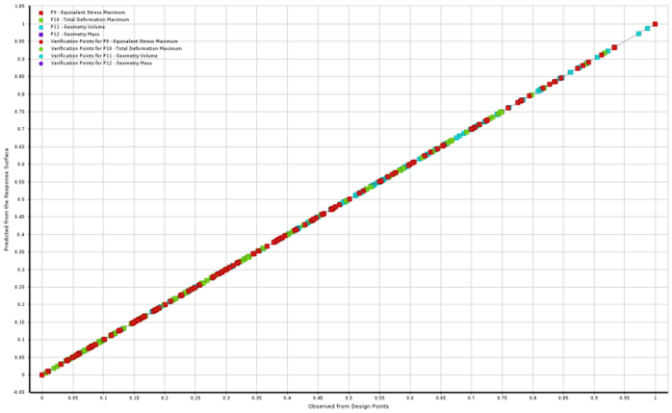


Figure 6. Distribution of sample points and verification points.

Figure 5 shows the goodness of fit of the agent model. As can be seen from the table, the root mean square error of Kriging agent model fitting, the relative maximum absolute error and the relative mean absolute error are both within the allowed error range, and the fitting accuracy of the agent model is generally good. As shown in Figure 6., is the coordinate distribution diagram of output predicted value and calculated value of sample points and verification points. The predicted value and calculated value are equal on the contour line, indicating that the fitting error is 0. As can be seen from the figure, their output points are basically scattered near the contour lines, which indicates that the fitting accuracy of the response surface meets the requirements and verifies the feasibility of the established agent model. It can be used for lightweight design.

5.3. optimization

On the basis of establishing the mathematical model of optimal design, the multi-objective genetic algorithm (MOGA) was combined to conduct optimization calculation[14~16]. Table 2 shows the candidate points finally obtained by the agent model combined with the optimization algorithm. As can be seen from the table, the error between the calculated value and the predicted value of the two candidate points is less than 2.1%, which verifies the fitting accuracy of the agent model. After

comprehensive analysis of the structure and analysis results of the hinge beam, candidate point 1 is selected as the optimal scheme.

Table 2. Calculated and predicted values of the two candidate points after optimization.

Design point	X1	X2	X3	X4	X5	P9(MPa)	P10(mm)	P11(mm^3)	P12(KG)
Sample points 1	80.388	52.6	612.02	385.19	322.5	473.75	0.69234	4.2448E+8	3332.1
Predictive value 1						483.67	0.69412	4.2446E+8	3332
Sample points 2	80.388	52.6	611.47	385.33	322.5	473.39	0.69124	4.2463E+8	3333.4
Predictive value 2						482.97	0.69312	4.2462E+8	3333.3

The results before and after optimization are shown in Table 3. The strength and stiffness of the optimized hinged beam structure meet the requirements. The optimal results are mass 3332.1Kg, stress peak 473.75MPa (satisfying the allowable stress), and maximum displacement 0.69234mm. The original mass was 3684.2kg. Compared with the initial scheme, the mass and volume of the hinged beam are reduced by 9.55%, the maximum stress is increased by 3.52%, and the maximum displacement is increased by 3.24%, but they are all within the allowable range of the material. The weight of the optimized hinge structure is significantly reduced, and the purpose of lightweight is effectively realized.

Table 3. Comparison of results before and after optimization.

Design point	X1	X2	X3	X4	X5	P9(MPa)	P10(mm)	P11(mm^3)	P12(KG)
Before optimization	100	60	650	360	380	457.62	0.67059	4.6932E+8	3684.2
The optimized	80.388	52.6	612.02	385.19	322.5	473.75	0.69234	4.2448E+8	3332.1
decrease	-19.61%	-12.34%	-5.84%	6.99%	-15.13%	3.52%	3.24%	-9.55%	-9.55%

6. Conclusion

(1) SolidWorks parametric modeling combined with ANSYS Workbench finite element analysis provides a large number of sample points for agent model fitting, saving a lot of time and calculation cost.

(2) The hinged beam structure was analyzed by finite element method, and the stress peak value and displacement peak value in the analysis results were used as constraint conditions to check the strength and stiffness of the optimized design scheme, which ensures the accuracy and reliability of the hinge beam structure design.

(3) Kriging was used to build the agent model to fit the relationship between input and output, and multi-objective genetic algorithm was used to complete the lightweight design of hinged beam structure. Compared with the initial scheme, the mass of the final optimization scheme is reduced by 9.55%. Under the condition of satisfying the strength and stiffness, the quality of the hinge beam structure is improved obviously, which also provides a reference for the lightweight design of other mechanical parts.

Acknowledgments

This work was supported by Key Scientific Project of Henan Province (211110220200).

References

[1] Xiaohu F, Jianjie W, Ye Y. Rapid development of large-scale cubic hinge press and relevant issues in China. Superhard Material Engineering. 2011; 23(01): pp. 42-45.
[2] Ye Y, Jianjie W, Ying L, Xiaohu F. The new understanding of increasing the quality of diamond depend on large-size of cubic hinge apparatus. Superhard Material Engineering. 2013; 25(06): pp. 15-19.
[3] Qingping Z, Jianhua Z. Lightweight Design of Vertical Lathe Bed for Repairing Wheel based on ANSYS Workbench. Journal of Physics: Conference Series. 2020; 1676(1): p. 012031.
[4] Di K, Zejun C, Youhua F, Li Cheng, Chengji M, Yinghong T. Optimization on kinematic characteristics and lightweight of a camellia fruit picking machine based on the Kriging surrogate model. Mechanics & Industry; 2021; 22: p. 16.
[5] Trapani FD , Sberna AP, Marano G C . A new genetic algorithm-based framework for optimized design of steel-jacketing retrofitting in shear-critical and ductility-critical RC frame structures. Engineering Structures, 2021,243:112684.
[6] Jian Z, Wei R. Lightweight Optimization Design of a Light Electric Commercial Vehicle Frame. Journal of Physics: Conference Series. 2021; 1939(1): p. 012038.
[7] Dongchen Q, Ying L, Liping C, Yifang Z. Research for hinge sleeve' s structural optimal design based on fatigue intensity. Journal of Mechanical Strength. 2006; 28(2): pp. 306-310.
[8] Bruno MS, Dunn KJ. Stress analysis of a beveled diamond anvil. Review of Scientific Instruments. 1984; 55(6): pp. 940-943.
[9] Liang M, Beihua L, Yuan Y, et al. The Stress Analysis and Tests on the Hinge Beam of the Diamond Synthesis Cubic Press. Matec Web of Conferences. 2016; 77: p. 01036.
[10] Rui L, Bojing X, Qingchao Z, Xue G, Guoliang Z, Hongan M, Xiaopeng J. Finite-element analysis on pressure transfer mechanism in large-volume cubic press. High Pressure Research. 2016; 36(4): pp. 575-584.
[11] Wenzheng Q, Xuan S. Analysis on Fatigue of Hinge Sleeve in Cubic Press Based on ANSYS. Applied Mechanics and Materials. 2014; 3082(543-547).
[12] Zhang Y, Li F, Jia D. Lightweight design and static analysis of lattice compressor impeller. Scientific Reports. 2020; 10(1): p. 18394.
[13] Xiang P, Chan Q, Jiquan L,Huaping W,Zhenyu L,Shaofei J. Multiple-scale uncertainty optimization design of hybrid composite structures based on neural network and genetic algorithm. Composite Structures. 2020(prepublish).
[14] Ponnambalam SG, Aravindan P, Naidu GM. A Multi-Objective Genetic Algorithm for Solving Assembly Line Balancing Problem. International Journal of Advanced Manufacturing Technology. 2000; 16(5): pp. 341-352.
[15] Jia J, Jian C, Chang G, et al. Energy efficient coverage control in wireless sensor networks based on multi-objective genetic algorithm. Computers & Mathematics with Applications. 2009; 57(11-12): pp. 1756-1766.

[16] Yu Z, Ping Z, and Guanlong C, et al. Study on Structural Lightweight Design of Automotive Front Side Rail Based on Response Surface Method . Journal of Mechanical Design. 2007; pp 553-557.

Vehicle Engineering

This page intentionally left blank

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 359
J. Xu et al. (Eds.)

doi:10.3233/ATDE220249

A Three-Stages Control Strategy for Semi-Active Seat Suspension

Yuxuan LIANG[a,1] , Yuxia LI[a] and Jinchao RAN[a]
[a] *State key laboratory of mechanical transmissions, Chongqing University*

Abstract. Commercial vehicle drivers are exposed to low-frequency vibration that causes many health problems, and the semi-active seat suspension can isolate the vibration well. In this paper, a three-stages control is proposed based on frequency-range selector from Single-Sensor control. The jerks appearing in original algorithm and the End-Stop impact coming from different sitting posture are discussed and reduced greatly by designing the control logic. At last, different excitation is used to simulation and the three-stages control is proved that can improve the comfort and safety of passengers.

Keywords. Semi-active seat, three-stages control, jerks, end-stop impact

1. Introduction

Commercial vehicle drivers are typically exposed to extended driving hours and severe working environment, which is mainly composed of low-frequency vibration (2-10hz), and that is a major factor in health disorders[1].

Suspension seats are widely used in vehicles to isolate the vibration transmitted to drivers or passengers[2]. The damper is the important part of seat suspension, which can reduce the power of vibration. Compared with traditional passive dampers, the advantages of active and semi-active dampers have been proposed in many studies[3-4]. Among them, semi-active suspension has a wide application prospect due to its low cost, good reliability and close control effect to active suspension[5].

Since the different sitting postures, the seats always can not stay at the middle height as the balance state, and it can be very dangerous that the seat suspension will hit the limits easily because there is not enough space for vibration. To solve this problem, some investigations discussed the adaptive algorithms[6], however, it is difficult to apply the complex control logic, and it will increase the cost of semi-active devices.

In this study, three-stages control based on frequency-range selector is proposed, it improves the logic to reduce the jerks during controlling and avoid the End-Stop impact caused by different seat height during driving. As the result, the comfort and safety of semi-active seat are improved.

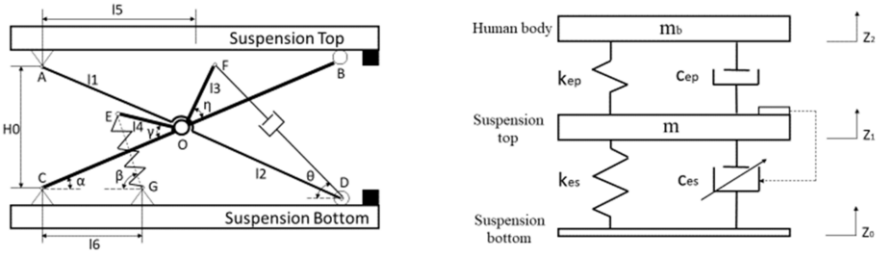
[1] Yuxuan Liang, State key laboratory of mechanical transmissions, Chongqing University; E-mail: 20152197@cqu.edu.cn

2. Dynamics Analysis of Seat Suspension

A commercial seat is chosen in this study including the damper and air spring. The damper is replaced with the magnetorheological damper so that the damping can be controlled with the current changing.

2.1. Physical Model Building

The seat suspension system is shown in the Fig.1(a) including damper and air spring. It can be simplified to a two degree of freedom system including human body and seat, which is shown in Fig.1(b):



(a) Seat suspension structure (b)Kinetic model of seat suspension

Figure 1. Schematic diagram of seat suspension .

The parameters required are measured as Table.1, and the simulation model is built with MATLAB Simulink.

Table 1. The seat suspension parameters

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
l_1/m	0.17	l_5/m	0.21	F	0.05	γ/ °	31
l_2/m	0.17	l_6/m	0.12	m_b/kg	70	m/kg	18
l_3/m	0.004	η/ °	40.5	H_0/m	0.11	k_{ep}/Ns·mm^{-1}	29.982
l_4/m	0.007	V_0/m^3	2.77×10^{-4}	n	1.38	c_{ep}/N·mm^{-1}	0.938

2.2. Simulation Model Building

According to the kinetic analysis and measurement of seat suspension, a simulation system including physical model, control block and excitation is built with Simulink program, which is shown as Fig.2:

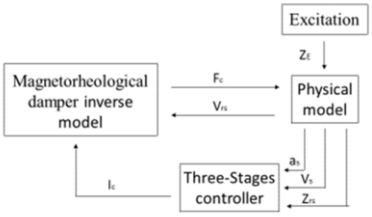


Figure 2. Block diagram of the control structure of seat suspension system

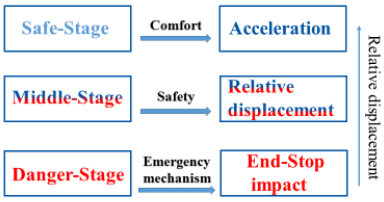


Figure 3. Diagram of three-stages controller logic

where the Z_E is the displacement of excitation, the a_s, V_s, and Z_{rs} are the acceleration, velocity of seat surface and the relative displacement of seat suspension, the I_c is the current from controller and the F_c and V_{rs} are the desired force of damping and the relative displacement of seat suspension. Sinusoidal signals, shock signals and random road according to the ISO/ TC108/ SCN67 and the GB7031 are used as the excitation.

3. Three-Stages Controller Design

A three-stages control algorithm based on frequency-range selector in Single-Sensor control from Savaresi.S and Spelta[7] is proposed in this study, which is divided into 3 control stages according to the relative displacement of seat suspension and shown as Fig.3.

3.1. Comfort Control Design in Safe-Stage

In safe-stage, the control objective is comfort, and the algorithm for this stage always trying to attenuate the acceleration of seat.

The seat suspension can be simplified to a single degree of freedom system, the transfer function of seat suspension is defined by:

$$G(s) = \frac{k+Cs}{ms^2+k+Cs} \tag{1}$$

and the bode diagram of seat suspension can be drawn as follows:

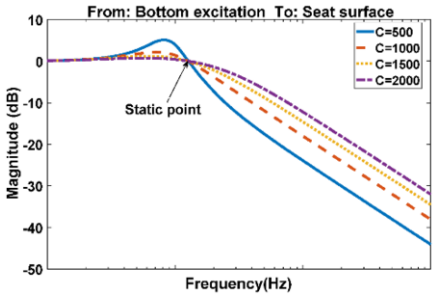


Figure 4. Magnitude of seat vibration in bode diagram

It is easy to find in Fig.4 that there is a static point in frequency-domain of seat suspension vibration. And the amount of damping required on both sides of this point is different to get the lower magnitude. Therefore, a selector about small and large damping according to the frequency is needed.

In order to regulate the vibration frequency, a static rule according to acceleration $\ddot{z}$ and velocity $\dot{z}$ is used in single-sensor control as follows:

$$\begin{cases} c_{in}(t) = c_{max}, if(\ddot{z}^2 - \alpha^2\dot{z}^2) \le 0 \\ c_{in}(t) = c_{min}, if(\ddot{z}^2 - \alpha^2\dot{z}^2) > 0 \end{cases} \tag{2}$$

where the $(\ddot{z}^2 - \alpha^2\dot{z}^2)$ is a simple "frequency-range selector" that can regulate the vibration frequency[7], which meets the requirement of damping control exactly.

In this study, $(\ddot{z}^2 - \alpha^2\dot{z}^2)$ is used to judge the vibration frequency range of seat surface to control the damping of seat suspension. It can be defined as follows:

$$\begin{cases} c_d = c_{max}, if\ \varepsilon \le 0 \\ c_d = c_{min}, if\ \varepsilon > 0 \\ \varepsilon = (\ddot{z}^2 - \alpha^2\dot{z}^2) \end{cases} \tag{3}$$

The parameter a is chosen according to the resonance frequency q(rad/s) of seat system and can be defined that:

$$a=\sqrt{2}q \tag{4}$$

where the q can be measured from vibration test and q≈8.8 rad/s (1.4Hz), as the result a≈12.5 rad/s.

When ε> 0, it means that the vibration frequency of seat surface on the right side of static point, which needs the small damping to reduce the vibration of seat surface, and it is the opposite when ε≤0.

However, some problems can be found during simulation. One is the harmonics components that are higher than seat resonance in frequency domain appearing in seat vibration as Fig.5(a):

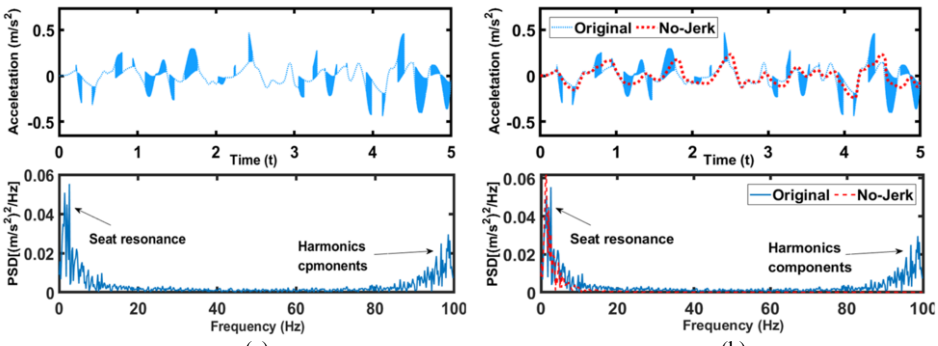


Figure 5. Acceleration characteristics of seat surface with the random rode excitation

It is easy to find that serious jerks appeared when the result of ε change cross 0 because of the binary logic in Eq.(3).Although in practical application, people may not feel shock at high frequency like this, the binary logic switching will also cause impact in lower frequency range and make people uncomfortable[8].

To solve this problem, an analytic continuous functions is employed to replace the original logic, which is defined as follows:

$$\begin{cases} c_d = c_{min}, if\ \varepsilon > 0 \\ c_d = c_{min} + K\varepsilon, if\ \varepsilon \leq 0 \end{cases} \tag{5}$$

where K is a constant gain. In this formulation, the damping c_d is related to the ε to make the damping force transit smoothly between two stages of algorithm and the testing results are shown in Fig.5(b). Noticed that the no-jerk algorithm function is very similar to the original algorithm, it means that the jerk reducing improving has almost no effect to original control function with appropriate parameter K. And it is clear that the zero-crossing jerk is avoided efficiently.

3.2. Emergency-Switch Function in Middle-Stage and Danger-Stage

In middle-stage, the algorithm will tend to reduce the relative displacement back to Safe-stage, and in danger-stage, the emergency mechanism will act and try to avoid the End-Stop impact.

Naturally, passengers tend to adjust the height of seat to make them feel comfortable and that may make there is not enough space for seat movement. This condition needs large damping to reduce the vibration no matter what the result of ε is. It may cause the sprung part of the suspension mechanism to hit the buffers on suspension top and bottom, which is named “End-Stop impact” in this study, and make passengers feel uncomfortable or even injure both the seat and people.

The End-Stop happening means the seat height crossed the limit of suspension. So the relative displacement of seat suspension is the most intuitive indicator that shows the danger degree of seat suspension, which is used to be the control basis of emergency mechanisms that is named “emergency-switch” in this study.

In three-stages control, emergency-switch is used to reduce the relative displacement H_1 of seat suspension to avoid End-Stop impact, which is defined as follows:

$$\begin{cases} c_d = c_{min} + K\varepsilon', if\ \varepsilon' \leq 0 \\ c_d = c_{min}, if\ \varepsilon' > 0 \\ \varepsilon' = (\ddot{z}_1^2 - b^2\dot{z}_1^2) \\ b = a + E^{q(H_1 - d_m)}, b \leq b_m \end{cases} \tag{6}$$

where the E and q are constants to adjust the rise rate of b, b_m is the upper limit of b, and the d_m is the threshold of safe-stage in three-stages control. The $E^{q(H_1-d_m)}$ is a growth function which approach to 0 when $H_1 < d_m$ with suitable E and q, and it will increase rapidly to b_m if $H_1 > d_m$.

When the H_1 is in the safe-stage range, $H_1 < d_m$, the damping follows the rule in Eq.(5). When the vibration increases and H_1 come into middle-stage range, $H_1 > d_m$,

which makes the b increase rapidly to be large enough, and ε' will be less than 0 until H_1 get back to safe-stage again. With the $\varepsilon' < 0$, Eq.(6) tends to choose large damping during vibration, which aims to reduce the relative displacement of seat suspension back to safe range by sacrificing comfort. After $H_1 < d_m$ again, the algorithm will switch to safe-stage and care about comfort more.

Simulation with random road excitation is shown as follows:

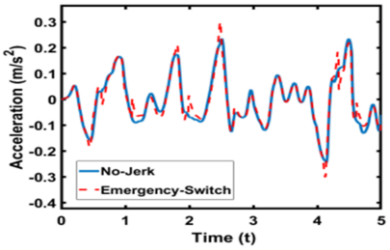


Figure 6. Acceleration of seat surface

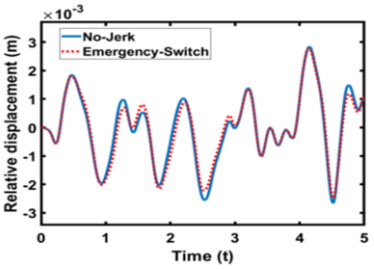


Figure 7. Relative displacement of seat surface

In this test, d_m is assumed to be 1mm. Noticed that some small jerks appear in acceleration signal, which means the emergency-switch acts and switches to large damping. In Fig.7, the emergency-switch curve can be found approaching 0 slowly compared with no-jerk control after being larger than d_m, and therefore the seat suspension can reduce the amplitude without huge jerks before danger-stage.

Eq.(6) is used in safe-stage and middle-stage, and d_m is the boundary of them two. If the input is so high that the H_1 cross the threshold of middle-stage d_d, the three-stages control will increase the damping to upper limit, which can reduce the H_1 as fast as possible to avoid the End-Stop impact. To sum up, the three-stages control can be defined as follows:

$$\begin{cases} c_d = c_{min} + K\varepsilon', if\ [\varepsilon' \le 0\ \cap\ H_1 < d_d] \\ c_d = c_{min}, if\ [\varepsilon' > 0\ \cap\ H_1 < d_d] \\ c_d = c_{max}, if\ H_1 \ge d_d \end{cases} \tag{7}$$

4. Laboratory Investigation of the Semi-active Seat Vibration Control

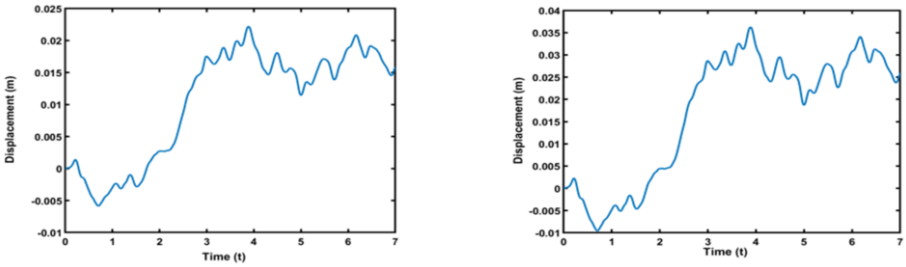
4.1. Investigation of Vibration Isolation Performances

In analysis of performance, three-stages control is compared with PID control built in Simulink, and the parameters used are shown as Table 2. Several kinds of input are used in simulation to reflect the different driving conditions.

In analysis of vibration reducing performance, D level and E level road excitation are used, and the car speed is assumed 60km/h, which is shown in Fig.8.

Table 2. Parameters for algorithms in simulation

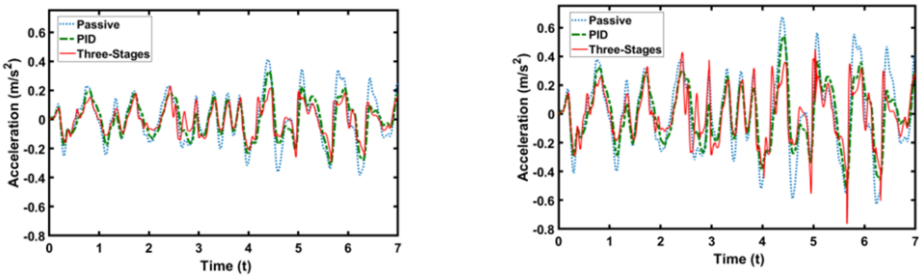
Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
P of PID	30	d_m (mm)	1	D of PID	0	K	5000
I of PID	100	d_d (mm)	2.5	E	154	q	1×10^5
Damping Max(Ns/m)	8000	Damping Min(Ns/m)	200				



(a)D level road excitation (b) E level road excitation

Figure 8. Random road excitation

Because of the limit of damper, damping of seat system is limited between 200 to 8000 Ns/m, and the acceleration of seat surface in time domain are shown as follows:



(a) D level road (b) E level road

Figure 9. Acceleration of seat surface

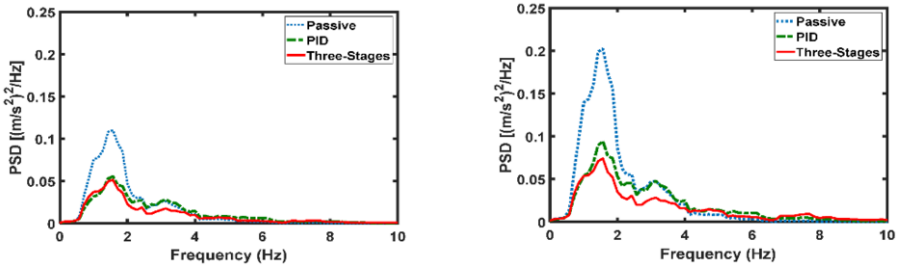
Table 3. RMS of acceleration with different algorithm

Control	RMS of Acceleration/(m·s^{-2})		Improvement	
	D Level	E Level	D Level	E Level
Passive	0.1649	-	0.2693	-
PID	0.1259	24%	0.2057	24%
Three-Stages	0.1186	29%	0.2037	25%

Note that three-stages provided a remarkable attenuation benefit during the simulation. It reduces the acceleration more effective than PID with the same limit of damper, which adversely affects the comfort of passengers. With the road excitation level increasing, some jerks appear in the acceleration signals, which is caused by the emergency mechanism action when the End-Stop happened. In order to analyze the improvement of three-stages, the root mean square (RMS) of these three acceleration signals are calculated as Table 3.

RMS of three-stages is 0.1186 m·s^{-2} which is reduced by 29% compared with passive seat. And it is also better than PID that the improvement is 24%. The similar

results can be found in E level road. All the conclusions can be proved by the power spectral densities (PSD) shown as Fig.10.



(a) D level road (b) E level road

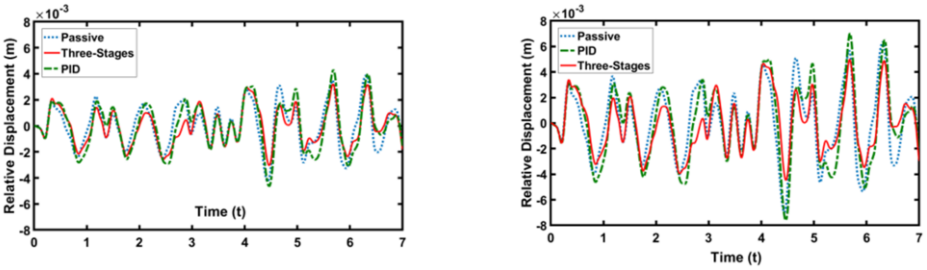
Figure 10. PSD of acceleration signals

Noticed that energy at resonance frequency is attenuated significantly, and better performance is shown by three-stages at high frequency (2-4Hz) than PID, which is consistent with the time domain analysis.

As shown in Fig.8, it is easy to find there are not any violent shock or long time resonance during the test, which causes the relative displacement of seat suspension is too small that can not reach the middle-stage or danger-stage designed for application, as a result, in order to analyze their performance, d_m and d_d are reset in this test to ensure all the stages will act.

In this simulation, the d_m of three-stages is set at 1mm to ensure there is enough time for middle-stage to act on the suspension, and the d_d is set at 2.5mm to avoid the End-Stop impact, which is assumed to happen when the relative displacement up to 3mm from start.

The relative displacement signals of them are shown as follows:



(a) D level road (b) E level road

Figure 11. Relative displacement of seat suspension

Table 4. RMS of relative displacement with different algorithm

Control	RMS of Displacement/(m)		Improvement	
	D Level	E Level	D Level	E Level
Passive	0.0016	-	0.0027	-
PID	0.0018	-12.5%	0.0029	-7%
Three-Stages	0.0014	12.5%	0.0022	19%

The three-stages is effective in reducing relative displacement, which is important to safety of seat and human during vibration. Noticed that better performances of three-stages appear at the high amplitude parts of vibration, as the result, the End-Stop impact assumed to happen at ±3mm is reduced obviously.

RMS of these signals are calculated as Table.4. It is interesting to find that the three-stages reduces the relative displacement while the PID increases that, which is because the PID is designed to reduce the acceleration and the action time may not exactly reduce the relative displacement in the same time.

4.2. Investigation of End-Stop Avoiding Performances

To further analyze the performance of three-stages in avoiding the End-Stop impact, the step excitation is used in simulation. It is designed high and wide enough to make the relative displacement of seat suspension reach the limit, which is measured ±4cm to the balance state, and causes End-Stop impact. The step excitation is shown as follows:

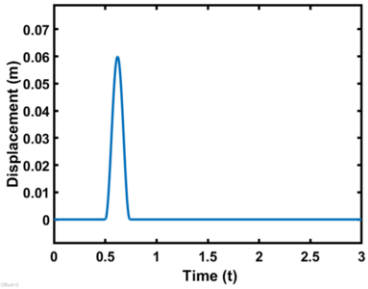


Figure 12. Step excitation for End-Stop

In addition, in order to evaluate the real performance of three-stages in application, the d_m and d_d are increased to 0.01 and 0.025 meters as reality. The relative displacement of passive seat, PID and three-stages are shown in Fig.13:

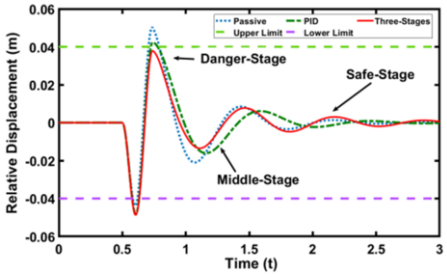


Figure 13. Relative displacement of seat surface.

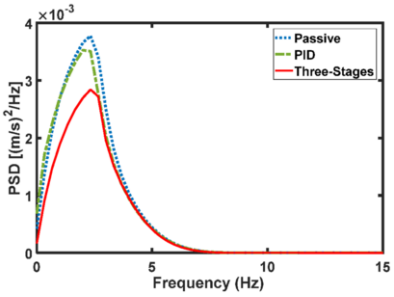


Figure 14. PSD of relative displacement signals

It is clear to find that when the shock acting on the seat bottom, the relative displacement increases to 4cm rapidly and End-Stop impact comes. After that, the emergency-switch of three-stages acts and reduces the amplitude as fast as possible, as the result the seat with three-stages control avoids the following impact while the second one acts on the PID and passive seat. When the seat suspension gets into lower stages, three-stages reduces the damping and the relative displacement reducing gets slowly to make the passengers feel comfortable. Although note that the three-stages reduces the relative displacement to the upper limit of suspension exactly, which means it can not avoid the End-Stop impact caused by larger shock, it can attenuate the impact more efficiently than others in simulation. That will be favorable to the safety of human and seat, and this also can be found clearly from PSD of them like Fig.14.

5. Conclusion

This study introduced a three-stages control logic for semi-active seat with MR damper. The transfer function shows that there is a static point in frequency-domain of seat suspension, and different damping are needed on both sides of it to reduce the vibration magnitude, which needs the frequency-range selector as $(\ddot{z}^2 - \alpha^2\dot{z}^2)$. During the application of this frequency-range selector, some jerks appears in simulation and they are reduced greatly with no-jerk improvement, which changes the binary logic of algorithm into continuously logic. And the emergency mechanism is proposed to avoid the End-Stop impact. The End-Stop impact is caused by different sitting postures, and adding the emergency-switch parameters ($b = a + E^{q(H_1 - d_m)}$) into algorithm does really good jobs to avoid it. At last, random road and shock excitation are used in simulation to act on the bottom of seat suspension model, and the results show that the three-stages control can attenuate both the acceleration of seat surface and displacement of seat suspension in different conditions (the improvement can reach around 25%) and that means the passengers or divers will be more comfortable and safe.

Acknowledgements

This research is supported by graduate scientific research and innovation foundation of Chongqing, China (Grant No. CYB19009)

References

[1] Choi SB, Han YM. Vibration control of electrorheological seat suspension with human-body model using sliding mode control. Journal of Sound and Vibration. 2007 Jun 5; 303 (1/2): pp.291-404.
[2] Jun W, Meihuan W, Yi Q. Approaches to predicting the vertical transmissibility of a suspension seat with a seated subject. Journal of Vibration and Acoustics-Transactions of the ASME. 2020 Aug 1; 142 (4): p 041012.
[3] Margolis DL. A Procedure for Comparing Passive, Active, and Semiactive Approaches to Vibration Isolation. Journal of the Franklin Institute-Engineering and Applied Mathematics. 1983; 315(4): pp 225-238.
[4] Karnopp DC. Design Principles for Vibration Control Systems Using Semiactive Dampers. Journal of Dynamic Systems Measurement and Control-Transactions of the ASME.1990;112: pp 448-455.
[5] Stabway R, Sproston JL, Servrns NG. Non-linear modeling of an electrorheological vibration dampers. Journal of Electrostatics. 1987 Dec; 20 (2): pp 167-184.
[6] Xuebin S, Ahmadian M, Southward S, Miller LR. An adaptive semiactive control algorithm for magnetoreological suspension systems. Journal of Vibration and Acoustics-Transactions of the ASME. 2005 Oct; 127 (5): pp 493-502.
[7] Savaresi SM, Spelta C. A single-sensor control strategy for semi-active sispensions. IEEE Transactions on Control Systems Technology. 2009 Jan; 17: pp 143-152.
[8] Mehdi A, Xubin S, Steve CS. No-jerk skyhook control methods for semiactive suspension. Journal of Vibration and Acoustics-Transactions of the ASME. 2004 Oct; 126: pp 580-584.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 369
J. Xu et al. (Eds.)

doi:10.3233/ATDE220250

Panoramic Image Stabilization Algorithm Base on Distance Transformation and Image Pyramid

Suming LIU [a, 1], Lin CHAI [a] and Lizuo JIN [a]
[a] *School of Automation, Southeast University, Nanjing, China*

Abstract. The vehicle panoramic scanning imaging system has the characteristics of single-camera splicing imaging and 360° full field view, while the traditional electronic image stabilization algorithm only applies to random camera vibration. In this regard, this paper uses the distance transform algorithm to extract the obvious area of the distance feature for template matching, and cooperates with the image pyramid to accelerate the matching. Finally, in the image fusion stage, the linear weighted fusion algorithm is used to obtain the final stitched panoramic image. Experimental results show that the algorithm used in this paper improves the image stabilization quality and takes into account the speed of image stitching, making the human eye more comfortable and stable when observing the video sequence target. This algorithm basically meets the needs of engineering applications and provides a guarantee for subsequent target detection and positioning.

Keywords. feature matching; motion compensation; image registration; electronic image stabilization

1. Introduction

The on-board monitoring system needs to monitor the environment in a 360-degree scene to find the target and observe the situation in time. Multiple cameras which used for shooting from different angles will cause high consumption, inconvenient installation and maintenance, and a risk of blind spots. In the existing technology, a method of stitching the continuous shooting images of a single camera from different angles into a panoramic picture is generally adopted. This method achieves the purpose of using a single camera to achieve a large-scale panoramic image. However, most of the existing turntable systems use fixed pitch angle and horizontal angle for directional shooting. During the rotation process, the uniform rotation of the horizontal orientation and the fixed position cannot be guaranteed. Therefore, the panoramic image stitching will be blurred due to the unstable encoder value during the shooting process, which will affect the look and subsequent processing.

In order to make image stable, the image sequence is usually calibrated. This process is called image stabilization. After decades of technological development, image stabilization technologies are mainly divided into three types: mechanical image stabilization, optical image stabilization and electronic image stabilization[1].

[1] Suming LIU: School of Automation, Southeast University, l_suming@163.com

Mechanical image stabilization uses the sensor and servo system to correct the slight deviation of the image. This method has strict requirements in terms of sensor accuracy and external factors, and needs to be used with multiple instruments to produce better results. Electronic image stabilization mainly uses two-dimensional image information to calculate the offset, and performs transformations such as translation and rotation on the images between frames. This process is simpler and lower in cost. By estimating the global motion between the two frames of images, the blurred picture is corrected, and finally the effect of image stabilization is achieved. Therefore, electronic image stabilization is currently a more widely used image stabilization method.

The classic methods of electronic image stabilization [2] mainly include: optical flow method, projection algorithm, feature point matching algorithm, block matching algorithm, and bit plane matching algorithm. The most important step in electronic image stabilization is image registration. The ransac algorithm calculates model parameters by extracting samples of preliminary matching points, which will result in a large number of iterations and a large amount of operations. In 2020, a new feature descriptor: BEBLID is proposed. Its matching accuracy is better than ORB algorithm, and the calculation is faster. At this stage, the above methods for panoramic image stitching can guarantee a better stitching effect. However, due to the huge number of feature points detected in the image feature detection process, it consumes a lot of space and time, and cannot achieve real-time performance in practical applications. This paper proposes a new two-step image stabilization method. First, the distance transformation algorithm is used to obtain image features, and the coarse template matching is performed on the regions with obvious distance features, and then the image pyramid is used to perform matching from coarse to fine. This method can greatly reduce the calculation time and achieve a fast and accurate image stabilization effect.

2. Panoramic image stabilization method

As mentioned above, the current technology has problems such as errors in stitching code value of a single camera, slow matching speed of the traditional method, and obvious boundary ghosting. Figure 1 is the structure flowchart of the method model.

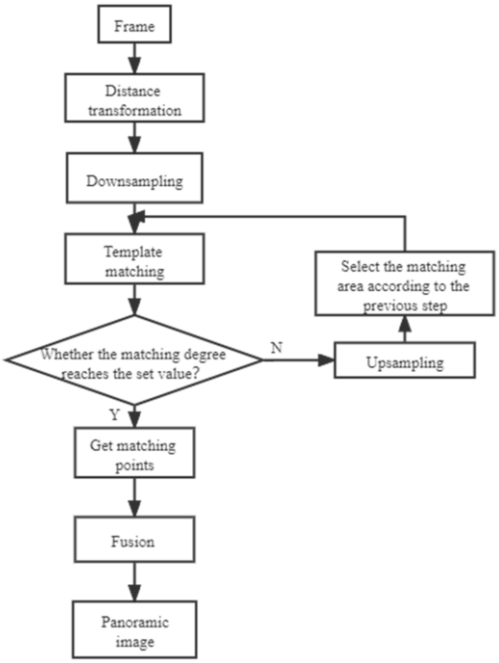


Figure1. Flowchart of panoramic image stabilization method.

2.1. Pre-compensation

The process of electronic image stabilization is actually to correct the image offset of two adjacent frames. The main methods include image feature methods, image gray information analysis and frequency-based methods. There is a jitter phenomenon in the pitch angle value of the system. The effect shown in Figure 1 is that there is a position difference in the vertical direction. If no processing is performed, it will affect the accuracy of subsequent motion estimation.

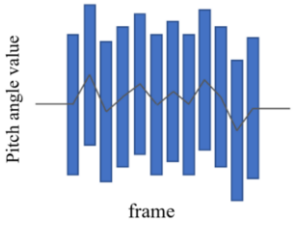


Figure 2. Difference in pitch angle value of turntable systems.

- Calculate the cumulative value of the absolute value between the ref frame and the i frame. The formula is:

$$sum = \sum_{j=ref}^{i} |Pit(j) - Pit(j-1)| \quad (1)$$

- Calculate the offset compensation amount $offset(i)$ of the i column based on the ref frame:

$$offset(i) = \frac{r}{\omega} \cdot (Pit(i) - Pit(ref)) \tag{2}$$

r is the number of rows of the original image, ω is the field of view range in the pitch direction. The code disk value information can be used for preliminary compensation.

2.2. Image feature extraction based on distance transform

In this paper, distance changes are used to extract matching features. Use distance changes to extract matching features. The distance transform algorithm [3] can calculate the distance from the non-zero-point pixel to the nearest zero-point pixel in an image, which can be used in the fields of skeleton extraction, diagonal matching and image stitching. The transformation changes the gray-scale intensity of points in the foreground area to show the distance from each point to the nearest boundary, and is more sensitive to outliers. The farther the point is from the border, the brighter it is.

The transformation result mainly depends on different distance measures. There are three distance measures: Euclidean distance, chessboard distance, and Manhattan distance. Euclidean distance formula is shown in Eq. (3). Chessboard distance formula is shown in Eq. (4). Manhattan distance formula is shown in Eq. (5).

$$disf\big(p(x_1, y_1), p(x_2, y_2)\big) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{3}$$

$$disf\big(p(x_1, y_1), p(x_2, y_2)\big) = |x_1 - x_2| + |y_1 - y_2| \tag{4}$$

$$disf\big(p(x_1, y_1), p(x_2, y_2)\big) = \max\left(|x_1 - x_2|, |y_1 - y_2|\right) \tag{5}$$

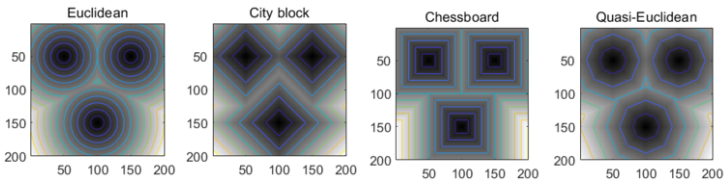


Figure 3. Visualization of four distance metrics

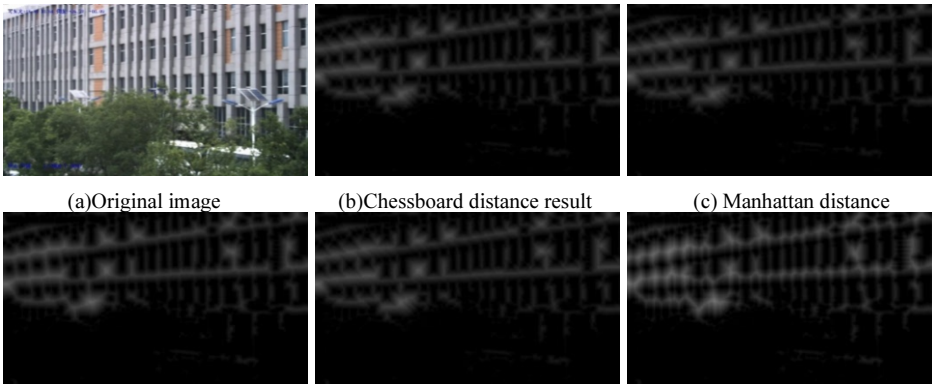
Through continuous corrosion operations until all the foreground areas are corroded, the distance from each foreground pixel to the foreground center skeleton pixel is obtained according to the order of corrosion, and different gray values are set according to the different distance values. Figure 4 shows the calculation results using different distance metrics.

There is also a chamfer distance algorithm [4] that requires less calculation. There are three types of point pixels in the image: internal points, external points, and isolated points. If the pixel point is a target pixel and the four neighborhoods are also target pixels, then the point is an internal point. If the four neighborhoods are background pixels, it is an isolated point, and the other area points are boundary points. The isolated point remains unchanged. The core is to use two small partial masks to traverse the image. Using mask 1, start from the upper left corner to traverse the array from left to right and from top to bottom, and update the element corresponding to point P in the mask:

$$F(P) = \min_{q \in mask1} \{F(P), D(P,q) + F(q)\} \quad (6)$$

Using mask 2, starting from the lower right corner, traversing from right to left, and from bottom to top, update the value of the element corresponding to point P in the mask:

$$F(P) = \min_{q \in mask2} \{F(P), D(P,q) + F(q)\} \quad (7)$$



(a)Original image (b)Chessboard distance result (c) Manhattan distance

(d) Euclidean distance (e) Chamfer distance (f) label pixel

Figure 4. Distance transform processing results

The final update of the array is the result of the distance transformation. It can be seen from the results that the Euclidean distance transformation result retains more image features. But the calculation time of Euclidean distance is too long. Comprehensive consideration of the effect and calculation time to choose the chamfering distance transformation method.

2.3. Image rough matching process

Since rotation has a small effect on jitter, it is not necessary to perform rotation and homography of two frames of image. Therefore, only the $x(k)$ and $y(k)$ directions of the motion components are analyzed during the processing. In this paper, a template matching algorithm [6] is used to match the two frames of objects. Template matching is given a template (usually a small image area) to find the area with the highest matching degree in the target image. It is widely used in target tracking, target detection and image stitching. The similarity measurement between the template and the sub-window of target image is the main part of template matching, and the commonly used method is pixel-by-pixel comparison, such as SAD, CSAD and SV-NCC.

The region with more obvious features in the image after the distance transformation in the previous step is selected as the matching template. The matching method adopts the normalized square difference matching method, which is defined as follows:

$$R(x,y) = \frac{\sum_{x',y'}(T(x',y') - I(x+x',y+y'))^2}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}} \tag{8}$$

The normalized cross-correlation considers the reference image and the current sub-image, and the formula is shown in Eq. (9).

$$C_N(r,s) = \frac{C(r,s)}{\sqrt{A(r.s)\cdot B}} = \frac{\sum_{(i,j)\epsilon R} I(r+i,s+j)\cdot R(i,j)}{[\sum_{(i,j)\epsilon R} I^2(r+i,s+j)]^{1/2}\cdot[\sum_{(i,j)\epsilon R} R^2(i,j)]^{1/2}} \tag{9}$$

It can be seen from the matching results that the regions with high similarity can also be performed well.

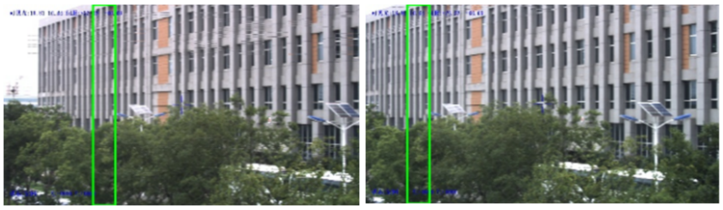


Figure 5. Matching results

2.4. Accelerate the precise matching process with image pyramid

The image pyramid can solve the problem of high computational complexity caused by direct template matching. Image pyramid refers to a collection of images with different resolutions, which is a kind of multi-scale expression of images. This acceleration strategy is mainly used for template matching. First, the template and image are down-sampled with the same number of layers. After the matching points are obtained, a layer-by-layer matching search is performed between candidate matching points, and finally the coordinates of the 0th layer are obtained. Search for images from small images and gradually reduce the range of image matching, this method can reduce the amount of calculation. Use the rough matching of the bottom image of the pyramid, and then search for matching layer by layer from coarse to fine. Compared with the original point-by-point matching, the matching range to the next layer can be reduced a lot, which effectively reduces the computing time.

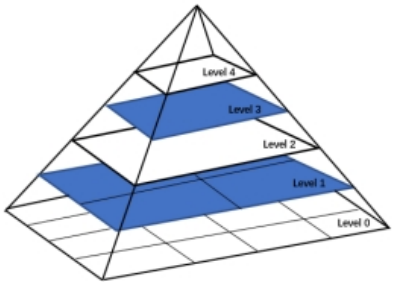


Figure 6. Schematic diagram of the pyramid

Direct matching takes about 73.27ms, while using image pyramids for matching takes about 12.93ms. This paper use image pyramids to match from coarse to fine to achieve algorithm acceleration.

2.5. Regional fusion

After the image stabilization of each frame is completed, there will be obvious stitching gaps and dislocations when directly stitching. In this paper, the weighted average fusion operation is adopted to fade in and fade out, and weight coefficients are respectively assigned to pixels in the overlapping area. The weight coefficient is related to the location. The expression is as follows:

$$I(x,y) = \begin{cases} I_1(x,y), (x,y) \in I_1 \\ \omega_1 \times I_1(x,y) + \omega_2 \times I_2(x,y), (x,y) \in (I_1 \cap I_2) \\ I_2(x,y), (x,y) \in I_2 \end{cases} \tag{10}$$

x_{max}represents the right boundary of the fusion, x_{min}represents the left boundary of the fusion.

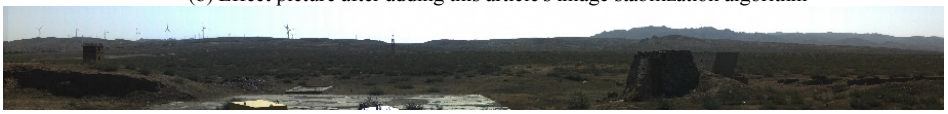
In summary, this article uses distance transformation to extract image information, template matching combined with image pyramid for feature matching, and image fusion algorithm, which can effectively improve the matching speed and imaging effect.

3. Experimental results and analysis

It can be seen from Figure 7 that due to the unstable fusion of the code disk value, the stitched image will be blurred and ghosted. The algorithm in this paper effectively improves the stitching effect.

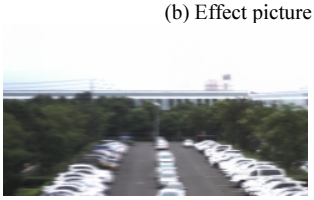
(a) Traditional stitching algorithm effect

(b) Effect picture after adding this article's image stabilization algorithm



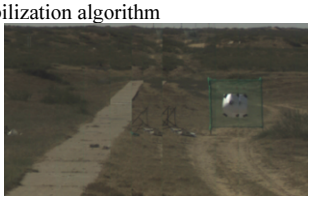
(a) Traditional stitching algorithm effect

(b) Effect picture after adding this article's image stabilization algorithm

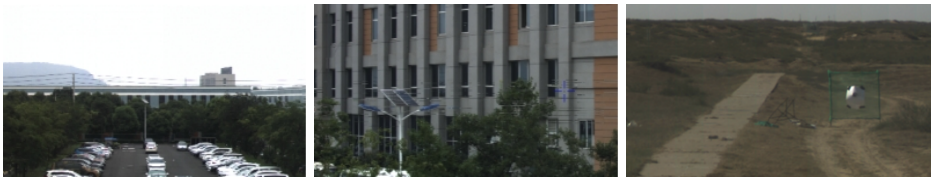


(a) Obvious ghosting

(b) Obvious ghosting



(c) Obvious seams



(d)No obvious ghosting (e) No obvious ghosting (f)No Obvious seams

Figure 7. Comparison between the traditional algorithm and the algorithm in this paper

3.1. Image stabilization evaluation index

The objective evaluation performance indicators of electronic image stabilization mainly include image stabilization accuracy and frame processing speed. The image stabilization accuracy is evaluated based on the error between the original frame image and the compensated frame image. Assuming that the image is completely compensated without active motion, the result of the difference operation of the image before and after the adjacent frame processing should be 0, but in actual situations, it cannot be completely compensated due to algorithm errors and noise. Therefore, the commonly used quantitative evaluation method is mainly peak signal-to-noise ratio (PSNR).

3.1.1. Image stabilization accuracy

PSNR can be used as an evaluation index to evaluate the consistency of image sequences, defined as follows:

$$PSNR = 10lg\frac{255^2}{MSE(P_1,P_0)} \tag{11}$$

Where $MSE(P_1, P_0)$ is the mean square error, defined as:

$$MSE(P_1,P_0) = \frac{1}{M\times N}\sum_{0\le i<N}\sum_{0\le j<M}(P_1(i,j) - P_0(i,j))^2 \tag{12}$$

Where M is the number of rows of the image, N is the number of columns of the image, P_0 is the reference frame, and P_1 is the frame corresponding to P_0 after the image is stabilized. When the result of the reference frame and the stabilized frame are closer, the image stabilization effect will be better. Ideally $MSE(P_1, P_0)$ is 0. $PSNR$ is inversely proportional to MSE, that is, the larger the $PSNR$ value, the better the image stabilization effect.

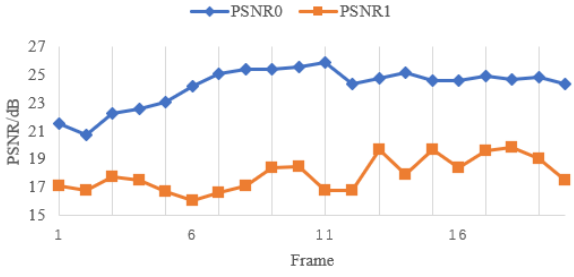


Figure 8. Comparison of $PSNR$ values for different algorithms

$PSNR0$ is the matching result after adding our method, and $PSNR1$ is the matching result according to the code disk value. It can be seen from Figure 8 that the $PSNR$ after the image stabilization algorithm can be increased by about 7% on average, indicating that the method in the article can achieve a better image stabilization effect.

3.1.2. Image stabilization speed

The running speed of the algorithm also has a certain impact on the final effect of image stabilization. On the premise of ensuring the stability and accuracy of the system, increasing the frame processing speed is the standard that the system design follows. The algorithm adopted in this paper takes about 26.66ms to calculate the motion model, which ensures the speed of the system.

Table 1. Comparison of the average PSNR and average algorithm time-consuming

Method	Time (ms)	PSNR(dB)	Standard Deviation(dB)
Original angle value	22.54	17.63	1.36
SURF	73.41	26.27	2.89
Ours	26.66	25.08	1.27

4. Conclusion

Aiming at the characteristics of panoramic image stitching, this paper proposes the use of distance transformation algorithm to extract features for template matching, and the use of image pyramids to reduce image registration time. In terms of fusion, the weighted fusion method of gradual in and gradual out is adopted to seamlessly connect the regions. This algorithm has realized the improvement of the image stabilization speed. The experimental results show that the method proposed in the article can effectively improve the quality of electronic image stabilization, and can also achieve better results for systems with poor turntable effects. With the requirement for better results and development of image process technology, subsequent research will improve the compensation algorithm, focusing on the movement of moving objects or backgrounds.

References

[1] Matsushita Y, Ofek E, Ge W, et al. Full-frame video stabilization with motion inpainting. IEEE Transactions on Pattern Analysis & Machine Intelligence. IEEE Computer Society; 2006 June; 28(7): pp 1150-1163.
[2] Szeliski R. Image Alignment and Stitching: A Tutorial. Foundations and Trends in Computer Graphics and Vision. 2006 Oct; 2(1): pp 1-104.
[3] Paik JK, Park YC,Kim DW. An adaptive motion decision system for digital image stabilizer based on edge pattern matching. IEEE Transactions on Consumer Electronics. IEEE; 1992 Jan; 38(3): pp 607-616.
[4] Felzenszwalb PF, Huttenlocher DP. Distance Transforms of Sampled Functions. Theory of Computing; 2004 Aug; 8(19): pp 415-428.
[5] Borgefors G. An improved version of the chamfer matching algorithm. 7 Int. Conf. Pattern Recognition, Montreal. 1984.
[6] Gunilla Borgefors. Distance transformations in digital images. Computer vision, graphics, and image processing. 1986 Jan; 34(3): pp 344-371.
[7] Rosenfeld A , Vanderbrug GJ . Coarse-Fine Template Matching. IEEE Transactions on Systems Man & Cybernetics. IEEE; 1977; 7(2): pp 104-107.

[8] Szeliski R, Sing BK. Direct methods for visual scene reconstruction. Proceedings IEEE Workshop on Representation of Visual Scenes (In Conjunction with ICCV'95);1995; pp 26-33.
[9] Cheng K, Shichao L, Rong L , et al. Video Stabilization via Prediction with Time-Series Network and Image Inpainting with Pyramid Fusion. Chinese Journal of Electronics. Chinese Institute of Electronics; 2021 Oct; 30(6): pp 1103-1110.
[10] Matsushita Y, Ofek E, Ge W, et al. Full-frame video stabilization with motion inpainting. IEEE Transactions on Pattern Analysis and Machine Intelligence.IEEE; 2006 July; 28(7): pp 1150–1163.
[11] Hu W, Chen CH, Su YJ, et al. Feature based real-time video stabilization for vehicle video recorder system. Multimedia Tools and Applications. Springer; 2018 Mar; 77(5): pp 5107–5127.
[12] Marcenaro L, Vernazza G, Regazzoni C S. Image stabilization algorithms for video-surveillance applications. International Conference on Image Processing. IEEE; 2001.
[13] Levin A, Zomet A, Peleg S, et al. Seamless image stitching in the gradient domain. Lecture Notes in Computer Science. 2004; 3024: pp 377-389.

Proceedings of the 1st International Conference on New Materials, Machinery and Vehicle Engineering 379
J. Xu et al. (Eds.)

doi:10.3233/ATDE220251

Application of Gray Predictive Theory in Vibration Isolation of Vehicle Semi-Active Seat Suspension

Zhiyuan ZHANG[a,1], Yuxuan LIANG[a] and Xiong DENG[a]

[a]*State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, China*

Abstract. Rapid response time is critical to semi-active suspension control system, but magneto-rheological (MR) damper has a relative low response time, which will damage the effect of vibration control. Gray predictive theory can use former sampled signal to predict latter signal to get an advanced output according to control strategy. This advantage can compromise the time delay of MR damper. In order to verify the effectiveness of this method, make comparation on simulation between skyhook control and gray-skyhook control about vibration control of vehicle seat suspension, the simulation result shows better performance from gray-skyhook control. It indicates gray predictive can decrease the negative influence of low response time of MR damper.

Keywords. Gray predictive, semi-active suspension, response time, simulation

1. Introduction

Vehicle seat suspension is an important subsystem of vehicle to control vibration. The commonly used seat suspension mainly includes passive seat suspension, semi-active seat suspension and active seat suspension. Semi-active seat suspension owing to the advantages of less energy consumption, simple structure and good control effects is widely studied. The hardware of semi-active seat suspension mainly consists of mass, stiffness component and damper, and the software of semi-active seat suspension is semi-active control strategy. Skyhook control[1] is a classical semi-active control and it is widely applied in vibration isolation about vehicle seat suspension. However, in reality the low response time of MR damper has a negative influence on vibration isolation when used skyhook control. In order to reduce response time of MR damper, some scholars[2-3] change MR damper structure or material then get a good effect.

Deng[4] proposed gray predictive theory, it aims at uncertain system that some information is known and some information is unknown. Gray predictive theory can use known information to draw valuable information then predict unknown information. Because of this characteristic, it is wildly applied to various filed[5-6].

[1] Corresponding Author, Zhiyuan Zhang, College of Mechanical and Vehicle Engineering, Chongqing University, Shapingba District, Chongqing, China; E-mail: 3282996569@qq.com.

In this article, build a model of two degree-of-freedom (DOF) seat suspension. A gray-skyhook control is designed to make comparations to skyhook with time delay and skyhook without time delay. The time delay simulates low time response of MR damper. Then compare effect of vibration isolation among the three control strategies. Finally, make evaluation about effect of vibration control and get a conclusion.

2. The dynamic model of vehicle seat suspension

2.1. The characteristics of MR damper

MR damper can change the damper force by ways of changing its operating current. As the actuator of semi-active seat suspension needs to be known its dynamic characteristics. MTS test system can provide a sine excitation and acquire damper force and displacement signals. MTS test system is showed in figure 1. Acquired signals via processing can draw a profile of force-velocity as shown in figure 2.

Figure 1. Damper force test system

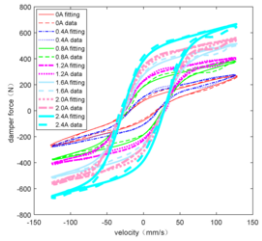


Figure 2. Force-velocity profile for each current

Figure 2 shows obvious hysteresis phenomenon in force-velocity profile. In order to build damper model correctly, the commonly used model includes Bingham model, Bouc-Wen model and Hysteretic model[7]. In this article, Hysteretic model is applied for building damper model. The Hysteretic model expression shows in expression 1.

$$F_d = c\dot{x} + kx + \alpha \cdot tanh[\beta\dot{x} + \delta \cdot sign(x)] + f_0 \quad (1)$$

Where, c is damping coefficient, k is stiff coefficient, α, β and δ are all shape coefficient, f_0 is compensation force. c, k and α relate to current i. $c = c_2 i^2 + c_1 i + c_0$,$k = k_2 i^2 + k_1 i + k_0$,$\alpha = \alpha_2 i^2 + \alpha_1 i + \alpha_0$.

Therefore, the damper model can be got by identifying these coefficients which relate to current, displacement and velocity. To identify coefficient, least square identification is an effective method. The identified result as shown in table 1. In order to verify the accuracy of coefficients, plot force-velocity profile and compare to tested data. figure 2 shows built damper model approaches the real damper.

Table 1. The identified coefficients of Hysteretic model

Coefficient	Value	Coefficient	Value
c_0	0.5	α_0	83.2
c_1	0.66	α_1	110
c_2	-0.37	α_2	188
k_0	2.1	β	0.0208
k_1	0.23	δ	0.877
k_2	-1.3	f_0	1.2

According to the expression of damper model, the inverse damper model can be deduced. It is critical for designing control strategy.

$$i = \begin{cases} \frac{-M+\sqrt{M^2-4LN}}{2L}, & \dot{x} \geq 0 \\ \frac{-M-\sqrt{M^2-4LN}}{2L}, & \dot{x} < 0 \end{cases} \tag{2}$$

Where:$L = c_2\dot{x} + k_2x + \alpha_2 \tanh[\beta\dot{x} + \delta sign(x)]$, $N = c_0\dot{x} + k_0x + \alpha_0 \tanh[\beta\dot{x} + \delta sign(x)] - f_0$, $M = c_1\dot{x} + k_1x + \alpha_1 \tanh[\beta\dot{x} + \delta sign(x)]$.

2.2. Two DOF seat suspension model

The vehicle seat suspension can be simplified as a two DOF model in figure 3. The explanation of dynamic parameters is listed in table 2.

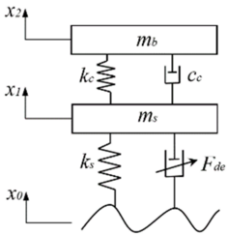


Figure 3. Two DOF vehicle seat suspension

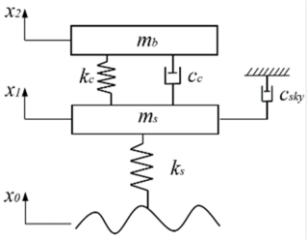


Figure 4. Ideal skyhook control model

The dynamic equations of two DOF model is established by:

$$\begin{cases} m_b\ddot{x}_2 = -c_c(\dot{x}_2 - \dot{x}_1) - k_c(x_2 - x_1) \\ m_s\ddot{x}_1 = c_c(\dot{x}_2 - \dot{x}_1) + k_c(x_2 - x_1) - F_d(t) - k_s(x_1 - x_0) \\ \dot{F}_d(t) = -\beta F_d(t) + \beta F_{din}(t) \end{cases} \tag{3}$$

Where, $F_{din}(t)$ is required damper force and $F_d(t)$ is the real damper force in real time. β relates to response time of damper force. the response time of damper force is about 30ms, so value of β is 30π.

Table 2. The dynamic parameters of two DOF seat suspension

Parameter	Value	Parameter	Value
Damping of cushion/c_c	0.95Ns/mm	Human mass/m_b	70kg
Stiff of cushion/k_c	31.5N/mm	Seat suspension mass/m_s	8kg
Stiff of spring/k_s	7.3N/mm		

3. The design of semi-active control algorithm

3.1. The traditional skyhook controller

The ideal skyhook control model as shown in figure 4. Skyhook damping parameter is c_{sky}, and its value is 2Ns/mm. Skyhook control algorithm is established in expression 4.

$$F_{din} = \begin{cases} F_{max},\ \dot{x}_1(\dot{x}_1 - \dot{x}_0) > 0 \ and \ F_{din} > F_{max} \\ c_{sky}\dot{x}_1,\ \dot{x}_1(\dot{x}_1 - \dot{x}_0) > 0 \ and \ F_{din} < F_{max} \\ F_{min},\ \dot{x}_1(\dot{x}_1 - \dot{x}_0) \le 0 \end{cases} \quad (4)$$

3.2. The design of gray-skyhook control algorithm

Gray predictive technique can be expressed by ordinary differential equations. The common form of gray predictive is GM(n, m), where, m represents the number of variables and n represents the order of gray model differential equations. With increase of order of differential equations and the number of variables, the calculations become more complex. However, the accuracy of prediction become worse. Therefore, GM(1,1) is widely used for gray prediction. The specific method is introduced from expression 5 to expression 10.

The initial discrete data is $x^{(0)} = \left[x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n-1), x^{(0)}(n)\right]$. from $x^{(0)}(1)$ to $x^{(0)}(n)$ is known, the target is unknown data $x^{(0)}(n+1)$.

There are some negative numbers in the initial discrete data, but AGO (Accumulated Generating Operation) process needs all numbers are positive. So $x^{(0)}$ should be mapped and become positive. Mapping expression is $x_m^{(0)}(k) = \exp\left[r \cdot x^{(0)}(k)\right]$, where, the value of r is 2. Mapped discrete data can be expressed as $x_m^{(0)} = \left[x_m^{(0)}(1), x_m^{(0)}(2), \dots, x_m^{(0)}(n-1), x_m^{(0)}(n)\right]$

AGO process is established in expression 5, then, $x_m^{(1)}$ can be got in expression 6.

$$x_m^{(1)}(k) = \sum_{i=1}^{k} x_m^{(0)}(i) \quad (5)$$

$$x_m^{(1)} = \left[x_m^{(1)}(1), x_m^{(1)}(2), \dots, x_m^{(1)}(n-1), x_m^{(1)}(n)\right] \quad (6)$$

1-order ordinary differential equation is established in expression 7.

$$\frac{dx_m^{(1)}}{dt} + cx_m^{(1)} = u \quad (7)$$

In order to solve 1-order ordinary differential equation, parameter c and u must be got. c and u can be solved in expression 8.

$$[c \quad u]^T = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{Y_n} \quad (8)$$

Where:

$$\boldsymbol{Y_n} = \begin{bmatrix} x_m^{(0)}(2) \\ x_m^{(0)}(3) \\ \dots \\ x_m^{(0)}(n) \end{bmatrix},\ \boldsymbol{B} = \begin{bmatrix} -\alpha_k x_m^{(1)}(2) + (\alpha_k - 1)x_m^{(1)}(1) & 1 \\ -\alpha_k x_m^{(1)}(3) + (\alpha_k - 1)x_m^{(1)}(2) & 1 \\ \dots & \dots \\ -\alpha_k x_m^{(1)}(n) + (\alpha_k - 1)x_m^{(1)}(n-1) & 1 \end{bmatrix}$$

The value of α_k is 0.5. Then the solution of expression 7 is solved. The next steps are IAGO (Inverse Accumulated Generating Operation) process and inverse mapping process. IAGO process is showed in expression 9 and inverse mapping process is showed in expression 10. Finally, the predicted data ($\hat{x}^{(0)}(k+1)$) can be evaluated.

$$\hat{x}_m^{(0)}(k+1) = \hat{x}_m^{(1)}(k+1) - \hat{x}_m^{(1)}(k) \tag{9}$$

$$\hat{x}^{(0)}(k+1) = \frac{1}{r} ln\left[\hat{x}_m^{(0)}(k+1)\right] = \frac{1}{r} ln\left\{(1-e^c)\left[x_m^{(0)}(1) - u/c\right]e^{-ck}\right\} \tag{10}$$

The gray predictive technique has been introduced, then it is easy to design gray-skyhook control strategy. The gray-skyhook control strategy is showed in figure 5. Apply gray predictive technique into predicting the excitation displacement signal (x_0) and seat floor displacement signal (x_1). Then differentiate them and get the excitation velocity signal ($\dot{x}_0$) and seat floor velocity signal ($\dot{x}_1$). Thus, skyhook control can use the predicted velocity signal to make an early output current to damper.

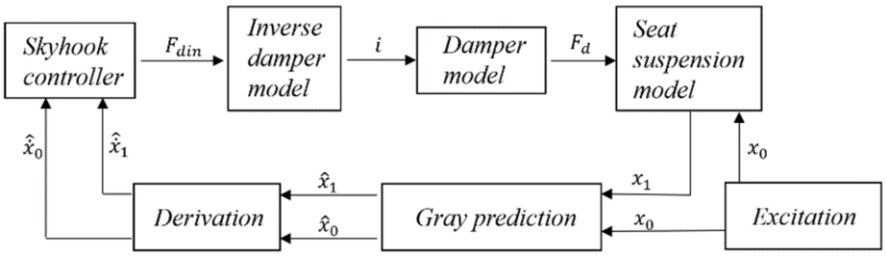
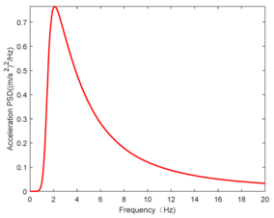


Figure 5. The logic diagram of gray-skyhook control

4. Simulation and analysis

Make simulation to verify the accuracy of predicted signal. the number of initial discrete signal is set as 4. Consider in the frequency of sampled signal that the sample time is set as 20ms. It means signal which is in 20ms later can be got in advance. The excitation signal is chosen from GB 8419-87, the acceleration PSD of excitation is showed in figure 6 and the displacement signal in time domain is showed in figure 7(a). The target is predicting the excitation displacement signal and seat floor displacement signal. The verified results are showed in figure 7, predicted signal is very closed to the real signal.



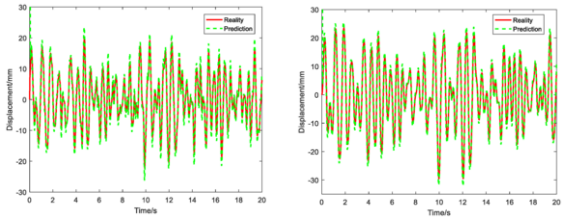


Figure 6. Acceleration PSD of excitation. **Figure 7.** Excitation (a) and seat floor (b) displacement signal

The accuracy of predicted signal is guaranteed, then make simulation about different control strategies which includes passive suspension (no control), ideal skyhook, gray-skyhook and skyhook with time delay. The evaluation standard of effect of vibration isolation usually considers in amplitudes of human body displacement and Root Mean Square (RMS) of human body acceleration. Figure 8 and figure 9 shows the simulated results of human body displacement and acceleration respectively.

Table 3. The effect of vibration isolation about human body acceleration

Control strategies	Acceleration RMS	Decline percentage compared to passive suspension
Passive suspension	2.087m/s^2	/
Ideal skyhook	1.477m/s^2	29.2%
Gray-skyhook	1.498m/s^2	28.2%
Skyhook with time delay	1.625m/s^2	22.1%

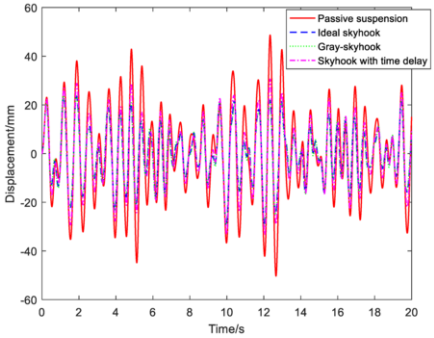


Figure 8. The vibration displacement of human body

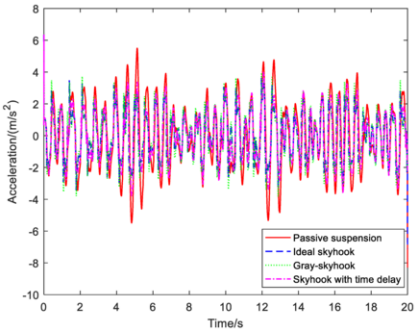


Figure 9. The vibration acceleration of human body

Value of human body acceleration RMS is listed in table 3. Simulated results of displacement and acceleration RMS of human body both indicates the controlled effect of gray-skyhook is closed to ideal skyhook and better than skyhook with time delay.

5. Conclusions

The response time of damper force really can make a negative effect on vibration isolation about skyhook control. Gray predictive technique can predict unhappened signal accurately. Gray-skyhook control can decrease the influence of time delay of damper force and get good performance in vibration isolation of vehicle seat suspension.

Acknowledgment

This work was financially supported by graduate scientific research and innovation foundation of Chongqing, China (Grant No. CYB19009, No. CYB21012). These supports are gratefully acknowledged.

References

[1] Karnopp DC, Crosby MJ, Harwood RA. Vibration control using semi-active force generators. ASME Journal of Engineering for Industry. 1974; 96(2): pp .619-626.
[2] Strecker Z, Roupec J, Mazurek I, Machacek O, Kubik M and Klapka M. Design of magnetorheological damper with short time response J. Intell. Mater. Syst. Struct. 2015; 26: pp. 1951–1958.
[3] Goncalves FD, Ahmadian M and Carlson JD. Investigating the magnetorheological effect at high flow velocities Smart Mater. Struct. 2015; 15: pp. 75–85.
[4] Deng JL. Introduction to gray system theory. J. Gray system 1 (1). 1989: pp. 1-24.

[5] Rogayye K, Karamollah B, Samad N, et al. A New Energy-Efficient Multipath Routing in Internet of Things Based on Gray Theory. International Journal of Information Technology & Decision Making. 2020; 19: pp. 37.
[6] Yuan, Qiang S, Yu Z, et al. Prediction and analysis of bearing vibration signal with a novel gray combination model. Advances in Mechanical Engineering. 2020; 12(5)
[7] Kwok NM, Ha QP, Nguyen TH, et al. A novel hysteretic model for magnetorheological fluid dampers and parameter identification using particle swarm optimization. Sensors and Actuators A: Physical. 2006; 132(2): pp. 441–451.

doi:10.3233/ATDE220252

Real-Time Target Detection in Closed Park Based on Multi-Sensor Information Fusion

Hongxia GAO [a], Qidi SUN [a,1], Yanqiu XIAO [a], Guangzhen CUI [a], Yanqi LIU [a] and Weili ZHANG [a]
[a] *College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou, China*

Abstract. Automatic driving technology has an important application prospect in park tourism, sanitation and other aspects. We took the intelligent vehicles in the closed park as the carrier and developed an intelligent sensing system based on the fusion of lidar and camera data. Firstly, the camera and lidar are jointly calibrated, and the data acquisition thread is created for timestamp registration to realize the spatio-temporal registration between heterogeneous sensors. Secondly, the visual target detection based on SSD network, get 2D target detection box. Then, point cloud target is detected based on PointPillars algorithm to get the depth and velocity message of obstacle. Finally, a fusion model is constructed to the camera lidar feature information to obtain the size, location and speed information of objects around the autonomous vehicle. The effectiveness of target detection of the proposed intelligent vehicle perception system is verified by a real vehicle test on a real road using Apollo autonomous driving development kit.

Keywords. Closed park, Multi-sensor, Information fusion, Target detection

1. Introduction

At present, automatic vehicles are equipped with sensors such as cameras, radar and lidar. However, the camera is affected by the complex environment, such as weather, light and other factors, and the target is often missed or the detection result is greatly different from the actual situation. Although lidar has high recognition accuracy and long distance, it is easily affected by environmental information[1]. Therefore, the information collected by a single sensor is limited and unreliable. Thus the intelligent vehicle perception system needs to coalesce multiple sensors to jointly perception the surrounding environment of the vehicle and obtain more accurate information about obstacles.

Environmental perception target detection algorithms mainly include 2D image target detection, 3D point cloud target detection, camera and lidar fusion object detection means. 2D image detection is mainly divided into two kinds. One method is based on region recommended candidate box extraction, such as Mask R-CNN[2], R-FCN[3]. The other is based on regression, which integrates the classification and location into a neural network, such as RetinaNet[4], CenterNet[5]. Lidar can obtain

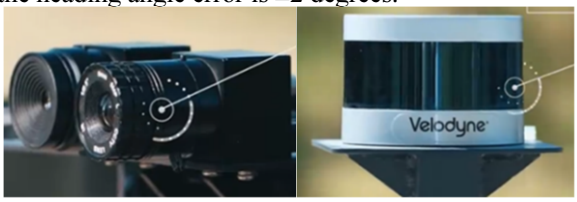
[1] QiDi SUN, College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou, China; E-mail: cindy08020616@163.com.

three-dimensional spatial information of driving environment and has advantages in target ranging and speed measurement. Point cloud target detection is primarily classified into three groups. One is to convert the original point cloud into voxel format and then input it into the network for detection, such as Sencond[6]. The second method is to convert the point cloud data into a two-dimensional image and feed it into the network for detection, such as PointPillars[7]. The third, such as PointNet[8], directly learns features from the original point cloud and then implements target detection. Research on multi-sensor data fusion technology, such as Chen[9] proposed a multi-view three-dimensional object detection network on account of multi-modal data fusion. The convolution layer was used to extract the point cloud features from aerial view and front view separately, and then the features are projected back to av,fv and the feature layer of the image. Finally, the features of the three branches were fused. Jason Ku[10] improved MV3D and proposed a 3D target detection method based on view aggregation, which removed the intensity information in the planform and front view of the lidar point cloud and boost the speed. Zheng[11] proposed a vehicle detection method based on the fusion laser point cloud and image information. Through the target detection and tracking of the original data of the image and the point cloud, the Hungarian algorithm is used to make the optimal matching of the detection results.

According to the comparison of existing research, By comparing existing studies, SSD combines YOLO's regression idea and Faster R-CNN anchoring mechanism to achieve more accurate positioning, and performs well in multi-scale target detection. PointPillars uses pillars to represent point cloud, extracts features from the pillars by maximum pooling operation, compresses the extracted features into bird's view map, and inputs them into 2D convoluted neural network for feature learning, which greatly improves the detection speed. Therefore, SSD and PiontPillars methods were selected as the basic framework of vision and point cloud detection in this paper. Finally, a fusion model was constructed to combine the feature information of camera lidar.

2. Hardware system

This paper uses a 12mm leopard camera with a resolution of 1920*1080 and a 6mm usb camera with a resolution of 640*480, as shown in Figure 1 (a). The 16-line velodyne lidar has a maximum recognition range of 50 m. The vertical measurement angle range is 30° (+15° to -15°), a vertical angular resolution of 2°, a horizontal measurement angle range of 360°, as shown in Figure 1 (b). The 16-line lidar is fixed on the top of the vehicle with a height H=1.5m to the ground and placed horizontally. The camera is installed at the front beam of the trolley structure frame and installed horizontally. The pitch angle is 0-2 degrees downward (the downward tilt is less than 2 degrees, and cannot be upward), the roll angle error is ±1 degree (the level of left and right sides), and the heading angle error is ±2 degrees.



(a) Cameras (b) Lidar

Figure 1. Vehicle sensor

3. Spatial and temporal registration

3.1. Coordinate system definition

On the autonomous vehicle, the lidar and vehicle are rigidly connected, and the relative attitude and displacement between the two are fixed. Therefore, the data points obtained by lidar scanning have a unique position coordinate corresponding to it in the vehicle coordinate system. Similarly, the camera also has a displacement position coordinate in the vehicle coordinate system, so there is a fixed coordinate transformation between the lidar and the camera. The coordinate system relationship is shown in Figure 2.

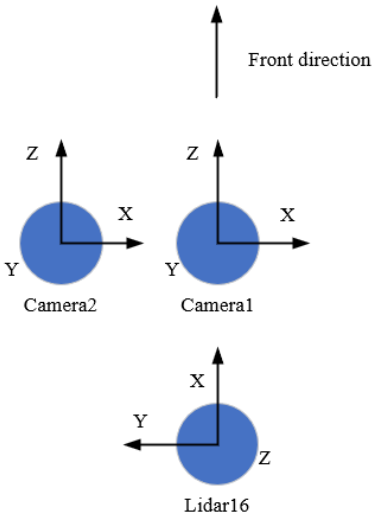


Figure 2. Camera and lidar coordinate system definition

3.2. Space synchronization

Monocular camera calibration obtains intrinsic and extrinsic parameters. The inner parameters mainly include focal distance, pixel center point, distortion coefficient and so on. Camera internal parameters can be obtained by Zhang Zhengyou calibration method. The operation method is based on ubuntu18-ros-melodic and its packaging box. The checkerboard calibration board and calibration result is shown in Figure 3 .

Figure 3. Calibration of monocular camera

Joint calibration obtains camera data and lidar point cloud data at the same time through real-time road data collection. The collected data are preprocessed to obtain the calibration result within the allowable error range, as shown in Figure 4.

Figure 4. Joint calibration of camera and lidar

3.3. Time synchronization

Each sensor collects data at different sampling frequencies. In order to ensure the accuracy and validity of detection data, time registration is required for data collected by different sensors at the same time. The measurement frequency of the camera used in this paper is 15Hz, and the data interval of each frame is 66.7ms. The measurement frequency of lidar is 10HZ, and the data interval of each frame is 100ms. In this paper, the sensor with long sampling period is used as the benchmark to carry out time stamp registration in a backward compatible way, as shown in Figure 5.

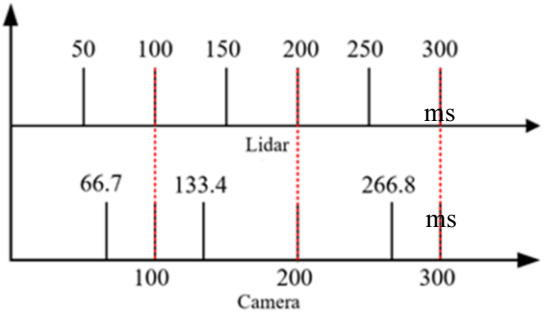


Figure 5. Time registration

4. Target detection based on multi-sensor information fusion

4.1. Visual object detection

SSD adopts the idea of YOLO regression and directly returns the boundary box and classification probability of the target. Meanwhile, it follows the strategy based on candidate box in Faster R-CNN and uses a large number of anchors to improve detection precision. SSD extracts feature maps of different scales for detection. Large scale feature maps can be used to detect small objects, while small scale feature maps can be used to detect large objects. After VGG16 network, multiple convolution layers are connected. This network structure fulfils the fundamental demand of intelligent vehicle perception system.

4.2. Point cloud object detection

Pointpillars is a new point cloud encoder that learns features without relying on a fixed encoder. It can use all the information about the point cloud to operate directly on the pillars, without manually adjusting the vertical box. Only uses 2D convolution in the

network, instead of 3D convolution. The whole network is composed of three parts. The point clouds are transformed into sparse pseudo-images by means of pillars. Using 2D network for feature learning. Bbox regression using SSD detection headers. This method has obvious speed advantage and has improved accuracy compared with the 3D target detection method which only uses point cloud as input.

4.3. Target detection based on feature information fusion

The visual object detection part based on SSD completes the classification detection of obstacles around intelligent vehicles, and obtains the size and category message of the target. Meanwhile, the lidar also obtains the three-dimensional space position and velocity information of the target. The feature information of the object detected by the two methods is coalesced to obtain the classification of the object as well as the three-dimensional space and velocity information. The flow chart of fusion method is shown in Figure 6.

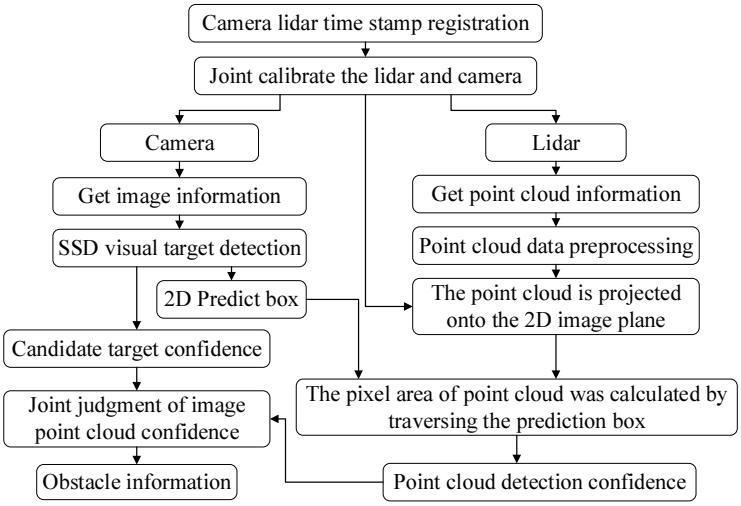


Figure 6. Fusion mechanism

4.4. Experiment

So as to test the accuracy and the validity of this way, a real-time perception experiment was carried out in the closed park based on Apollo autonomous driving development kit. The actual driving situation of intelligent vehicles was observed and the performance of the target detection system was evaluated according to the data collected during the experiment.

By making a virtual lane line and setting the start point and end point, the intelligent vehicle can independently perceive and predict the environment. The experiment result is shown in Figure 7. It can be seen that the environment perception system designed in this paper can accurately detect obstacles information surrounding the automatic vehicle, including the target size, location and speed, to meet the needs in the process of intelligent vehicle driving.

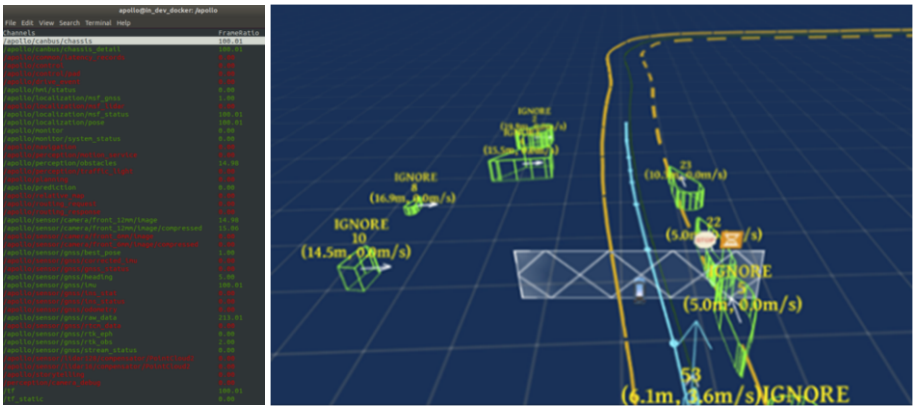


Figure 7. Real-time target detection results

5. Conclusion

Aiming at the environmental characteristics of the closed park, an intelligent vehicle perception system based on camera lidar information fusion was designed to obtain the location and speed information of obstacles around the intelligent vehicle. Based on Apollo autonomous driving development kit, the real vehicle test was carried out on a real road to complete the autonomous perception obstacle avoidance function of autonomous driving in a closed park, which verified the effectiveness of the intelligent vehicle perception system in this paper.

References

[1] Badue C, Guidolini R , Carneiro R V , et al. Self-driving cars: A survey. Expert Systems with Applications, 2020; p113816.
[2] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN". IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017; pp 2980-2988.
[3] Dai J , Li Y , He K , et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks. Curran Associates Inc, 2016; pp 379-387.
[4] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017; (99): pp 2999-3007.
[5] Zhou X, Wang D, Krhenbühl P. Objects as Points. 2019.
[6] Yan Y, Mao Y-X, Li B. SECOND: Sparsely Embedded Convolutional Detection. SENSORS, 2018; 18(10).
[7] Lang A H, Vora S, Caesar H, et al. PointPillars: Fast Encoders for Object Detection From Point Clouds. IEEE Conference on Computer Vision and Pattern Recognition, 2019; pp 12689-12697.
[8] Charles R. Qi, Hao S, Kaichun M, et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentatio. IEEE Conference on Computer Vision and Pattern Recognition, 2017: pp 77-85.
[9] Xiaozhi C, Huimin M, Ji W, et al. Multi-View 3D Object Detection Network for Autonomous Driving. IEEE Conference on Computer Vision and Pattern Recognition, 2017; pp 6526-6534.
[10] Jason K, Melissa M, Jungwook L, et al. Joint 3D Proposal Generation and Object Detection from View Aggregation. IEEE International Conference on Intelligent Robots and Systems, 2018; pp 1-8.
[11] Shaowu Z, Weihua L, Jianyao H. Vehicle detection in the traffic environment based on the fusion of laser point cloud and image information. Chinese Journal of Scientific Instrument, 2019; 40(12): pp 143-151.

doi:10.3233/ATDE220253

A Quasi-Optimal Energy Management Strategy for Hybrid Energy Vehicles Using Temporal Characteristic

Yao LU [a], Weirong LIU [b], Yue WU [a], Heng LI [b], Yongjie LIU [a], Jun PENG [b], Zhiwu HUANG [a,1]
[a] *School of Automation, Central South University, Changsha, China*
[b] *School of Computer Science and Engineering, Central South University, Changsha, China*

Abstract. For electric vehicles with supercapacitor and battery storage devices, it is a challenge to achieve real-time global energy optimization and extend the battery lifetime simultaneously. In this paper, a quasi-optimal strategy to allocate energy by combining dynamic programming with long short term memory (LSTM) network is proposed for a semi-active battery/supercapacitor hybrid energy storage system. In this work, the cost function of the vehicle operating cost is constructed, including battery degradation and energy consumption. The optimal current reference is determined by solving the optimization problem by dynamic programming algorithm. According to the result of the solution, we construct the corresponding training data set for LSTM training. The trained LSTM generates the quasi-optimal solution for the SC decision, which is utilized to distribute the load power afterward. Simulation on UDDS data set illustrates that the designed strategy can improve the energy efficiency and extend the battery lifetime while guaranteeing the SC SoC constraints.

Keywords. electric vehicle, hybrid energy storage system, energy management strategy, long short term memory, vehicle operating cost.

1. Introduction

Battery life is critical to the range of an electric vehicle (EV) [1], and the growing EV industry requires batteries to have long cycle lifetime. Frequent acceleration and deceleration of EVs will adversely affect battery life. The hybrid energy storage system (HESS) based on battery/supercapacitor (SCs) can give full play to their respective advantages and effectively reduce the adverse impact of high-frequency current components on battery life [2]. In HESSs, the energy management strategy (EMS) can affect storage system life and even vehicle operating safety [3].

In existing studies, EMSs have the following two categories according to different design ideas, i.e., rule-based methods [4] and optimized-based methods [5]-[6]. The strategy based on adaptive filters is a special rule-based method [4], but this strategy relies on experience and cannot achieve optimal energy management. In Contrast,

[1] Corresponding author: email: hzw@csu.edu.cn

dynamic planning (DP) is one of the most effective strategies to find global optimal programs, but it is not possible to provide online solutions. For the above problems, a near-optimal strategy from DP [5] is proposed. In [6], EMS combined with DP and neural networks is proposed, but only battery loss is considered.

To this end, a quasi-optimal EMS using DP and long short term memory (LSTM) is proposed for semi-active battery/SC HESSs to increase battery life and decrease energy consumption. First, establish the objective of the operating cost of the vehicle, including battery degradation costs and power costs. Second, solve it by DP to obtain the SC current reference value. Third, a data set is constructed, in which five state variables related to the optimization problem are taken as the input of LSTM and SC current reference value is taken as the output. And the approximate optimal reference of the LSTM output SC current after training. Compared with traditional neural networks, LSTM memorizes sequence data through internal parameters, which is more suitable for data sets with temporal characteristics [7].

This paper is organized as follows. Section 2 presents the hybrid energy system of electric vehicles. In Section 3, we solve the optimization problem by DP and construct data sets for training LSTM. Section 4 presents the simulation results, and analyses the comparative results between different EMSs. Section 5 concludes this paper.

2. Electric vehicle energy architecture

This paper uses a semi-active topology, in which the bi-directional DC / DC converter is connected in series with the SC package, the battery pack is directly connected to the DC bus [8]. In this case, the SC voltage is smaller, the battery pack voltage is equal to the DC bus voltage. This topology effectually extends the range of the SC input voltage, and the energy efficiency in the SC is significantly improved.

3. Design of energy allocation strategy

3.1. HESS optimization problem

Battery life is the first focus of the cost function for the HESS optimization problem. In order to achieve an extended battery life, the optimization problem of the vehicle operating cost is constructed in the battery life. This optimization problem mainly considers the battery degradation cost and the electricity cost, expressed by $Cost_{loss}(k)$ and $Cost_e(k)$, respectively. The HESS operation cost is given as

$$Minimize\left\{\sum_k \frac{\left[Cost_{loss}(k)+Cost_e(k)\right]}{L_c}\right\} \tag{1}$$

$$\mathrm{Cost}_e(k)=\frac{price_e T_s}{3600}[P_{sc}(k)+P_{bat}(k)], \tag{2}$$

$$Cost_{loss}(k)=\frac{C_{bat}V_{bat}price_{bat}}{0.2\times1000}\times9.78\times10^{-4}\frac{\left|I_{bat}(k)\right|T_s}{3600\times2}\times e^{-(\frac{15162-1516C_{rate}}{0.849R(\left|285.75-T_{bat}(k)\right|+265)})}Q_{loss}(k-1)^{-0.1779}. \quad (3)$$

Where $Price_{bat}$ and $Price_e$ are battery and electricity prices which are set to 1050CNY/kWh and 0.6CNY/kWh. L_c denotes the driving cycle distance. $P_{sc}(k)$ and $P_{bat}(k)$ represent SC power and battery power, $Q_{loss}(k-1)$ is the battery degradation at time step $k-1$. The calculation formula of capacity loss of battery is given in [10].

For the above optimization problems, the constraint mainly includes the current size of the battery/SC and the SoC ranges

$$I_{bat,\min}\le I_{bat}(k)\le I_{bat,max}\quad I_{sc,\min}\le I_{sc}(k)\le I_{sc,max}\quad SoC_{sc,\min}\le SoC_{sc}(k)\le SoC_{sc,max}. \quad (4)$$

Where the subscripts of min and max represent the minimum and maximum values of each variable.

3.2. Quasi-optimal reference for LSTM

To address the above issue, the DP algorithm is adopted to get the global optimal solution. Battery current is denoted as the state variable $x(k)$, SC output current as the control variable $u(k)$, and the corresponding model is $x(k+1)=g(x(k),u(k))$.We rewrite the cost function (3) more compactly as $J=\sum_{k=0}^{N-1}Cost(x(k),u(k))$.

The dynamic programming algorithm adopts reverse calculation, and the complex optimization problem can be transformed into the following sub-problems to solve.

The step $k-1$:

$$J^*_{N-1}(x(N-1))=\min_{u(N-1)}[Cost(x(N-1),u(N-1))+G(x(N))] \quad (5)$$

The step k:

$$J^*_{\mathrm{k}}(x(k))=\min_{u(k)}\left[Cost(x(k),u(k))+J^*_{\mathrm{k+1}}(x(k+1))\right] \quad (6)$$

Where $J^*_k(x(k))$ represents the minimum cost of the storage state $(x(k))$ at the time step k, and $G(x(N))$ represents the last cost. The dynamic algorithm is used to recursively calculate up to $J^*_k(x(1))$, so as to determine the optimal power allocation strategy.

The first step of building LSTM is to construct a data set composed of input and target data required by the LSTM. Due to demand current I_l, the battery current at the previous time step I_{bat_p}, the SoC of the SC SoC_{sc} are key factors affecting the results of the DP solution. Besides, the velocity v, and acceleration a can provide

information about driving trends, improve energy management performance. Therefore, five features described above are used as input, and the SC reference current $I_{sc,ref}$ is used as the output.

Using the neural network for energy management can not only realize real-time energy management but also approximate the optimal energy management strategy. In the neural network, all layers of the neural network are connected, and the neurons in each layer exist in isolation without connection. Therefore, the traditional neural network with this structure does not perform well in solving problems with temporal characteristics. Compared with the traditional neural network, LSTM can remember sequence data through internal parameters, which is suitable for hybrid energy management.

3.3. energy management and current control

After training, LSTM can output the SC expected current $I_{sc,ref}$, and adjust the duty cycle D_{uc} of energy converter by generating PWM signal, to realize nearly optimal energy management.

4. Simulation and verification

Reasonable selection of energy storage device can ensure the performance of the vehicle and reduce the initial cost and operation cost. In this paper, the battery pack is 2-parallel 125-serial configured with LiFePO4 battery cell and the SC pack is set to 6-parallel 11-serial MAXWELL cells. The simulation parameters are shown in Table 1.

The UDDS driving cycle is used to simulate the proposed strategy. In this paper, $I_{bat,\min}$ and $I_{bat,max}$ are set from -200A to 200A. $I_{sc,\min}$ and $I_{sc,max}$ are set from -60A to 120A, $SoC_{sc,\min}$ and $SoC_{sc,max}$ are set from 0.5 to 0.9. The SC SoC is set to 0.84 at the beginning.

Table 1. The simulation parameters

Parameter	Battery	SC
Nominal voltage(cell)	3.2V	2.7V
Capacitance(cell)	-	140F
Nominal capacity(cell)	60Ah	-
Nominal voltage(pack)	400V	297V
Nominal capacity(pack)	120Ah	-
Capacitance(pack)	-	76.36F
Internal resistance(pack)	~93.75mΩ	~14.67mΩ

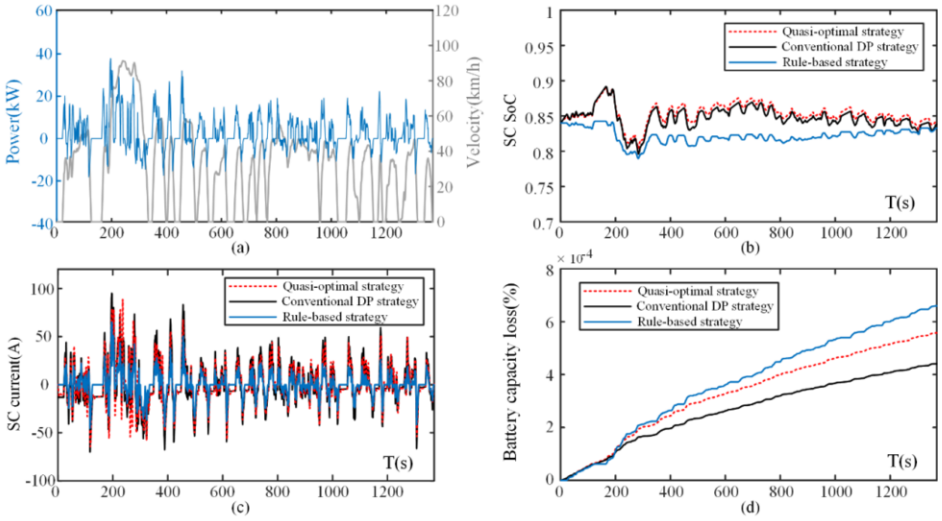


Figure 1. Simulation comparisons with different EMSs under UDDS driving cycle. (a) velocity and demand power, (b) SC SoC, (c) SC current, (d) battery capacity loss.

The designed quasi-optimal strategy is compared with the rule-based near-optimal strategy [5] to verify its performance, as shown in Figure 1. Figure 1 (a) is the velocity and demand power of UDDS data set. As can be seen from Figure 1 (b), SoC of SC under the three strategies can all maintain the initial value to provide good endurance. As shown in Figure 1 (c), the quasi-optimal strategy provides a relatively small battery current, which is not much different from the effect of conventional DP. Thus it can be considered as an approximate optimal energy strategy. In Figure 1(d), the battery degradation under the quasi-optimal strategy is less than that under the rule-based near-optimal strategy.

Table 2. Total costs for different EMSs.

Strategy	Q_{loss} (%)	Battery degradation (CNY/100km)	Total cost (CNY/100km)	Cost comparison
Conventional DP strategy	0.000441	9.26	14.28	benchmark
Quasi-optimal strategy	0.000557	11.71	16.75	17.30%
Rule-based strategy	0.000659	13.86	18.90	32.35%

Table 2 shows the battery degradation and total cost of existing EMSs. Compared with the rule-based near-optimal strategy, the quasi-optimal strategy can effectively prolong the battery life and reduce the battery degradation cost by 15.5%. In addition, for HESS operating costs, this strategy reduced total costs by 11.4% in comparison. The simulation results show that the quasi-optimal strategy can effectively prolong the battery life, reduce the cost of battery replacement, and improve the economy of the vehicle.

5. Conclusion

In this paper, a quasi-optimal EMS is presented based on DP and LSTM for a hybrid energy system of electric vehicles. To optimize the vehicle operating consumption, a cost minimum problem is constructed and solved by DP. The results obtained by DP

are used to train LSTM. The trained LSTM can provide a quasi-optimal SC current reference value in real time. Simulation conducted on UDDS driving cycle show that compared with rule-based DP strategy, this strategy reduces battery degradation by 15.5.%, and cut down the total cost of 11.4%.

Acknowledgments

The authors would like to acknowledge that this work is financially supported by the National Natural Science Foundation of China (Grant No. 61803394, 61672539, 61672537).

References

[1] S. F. Tie and C. W. Tan, "A review of energy sources and energy management system in electric vehicles," Renewable and Sustainable Energy Reviews, vol. 20, pp. 82 – 102, 2013

[2] S. Zhang, R. Xiong, and J. Cao, "Battery durability and longevity based power management for plug-in hybrid electric vehicle with hybrid energy storage system," Applied Energy, vol. 179, pp. 316–328, 10 2016.

[3] R. Xiong, H. Chen, C. Wang, and F. Sun, "Towards a smarter hybrid energy storage system based on battery and ultracapacitor - a critical review on topology and energy management," Journal of Cleaner Production, vol. 202, pp. 1228 – 1240, 2018.

[4] A. Florescu, B. Seddik, I. Munteanu, A. Bratcu, and A. Rumeau, "Adaptive frequency-separation-based energy management system for electric vehicles," Journal of Power Sources, vol. 280, 04 2015.

[5] Z. Song, H. Hofmann, J. Li, X. Han, and M. Ouyang, "Optimization for a hybrid energy storage system in electric vehicles using dynamic programing approach," Applied Energy, vol. 139, pp. 151 – 162, 2015.

[6] J. Shen and A. Khaligh, "A supervisory energy management control strategy in a battery/ultracapacitor hybrid energy storage system," IEEE Transactions on Transportation Electrification, vol. 1, no. 3, pp. 223–231, Oct 2015.

[7] Siami-Namini S, Tavakoli N, Namin A S. A comparison of ARIMA and LSTM in forecasting time series[C]//2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018: pp. 1394-1401.

[8] J. Cao and A. Emadi, "A new battery/ultracapacitor hybrid energy storage system for electric, hybrid, and plug-in hybrid electric vehicles," IEEE Transactions on Power Electronics, vol. 27, no. 1, pp. 122–132, Jan 2012.

[9] Z. Song, J. Li, J. Hou, H. Hofmann, M. Ouyang, and J. Du, "The battery-supercapacitor hybrid energy storage system in electric vehicle applications: A case study," Energy, vol. 154, pp. 433–441, 2018.

[10] J. Shen, S. Dusmez, and A. Khaligh, "Optimization of sizing and battery cycle life in battery/ultracapacitor hybrid energy storage systems for electric vehicle applications," IEEE Transactions on Industrial Informatics, vol. 10, no. 4, pp. 2112–2121, Nov 2014.

doi:10.3233/ATDE220254

Research on Automobile Driving Anti-Skid System Based on Throttle Control

Shuchen LIU[a,1], Ying ZHANG[b], Ningning REN[a], Jianying TIAN[a]
[a] *Shandong Institute of Commerce & Technology, Jinan, Shandong, China*
[b] *Henan Institute of Technology, Xinxiang, Henan, China*

Abstract. The anti-skid system of the automobile drive adopts the control of the engine torque and/or braking torque, which can be used to prevent the wheel from slipping on slippery roads and improve the acceleration performance, driving stability and maneuverability of the automobile. This article mainly solves how to realize the anti-skid driving of the car by controlling the throttle. For the purpose of improving automobile driving performance, its anti-skid control algorithm includes control gain adjustment and road adhesion coefficient estimation. Through various automobile tests on three low-adhesion roads, the anti-skid control algorithm is effectively proved.

Keywords. driving antiskid control system; throttle control; Anti-sliding control algorithm; attachment coefficient

1. Introduction

The development of electronic technology and control technology has promoted the development of active safety systems for automobiles. Automobile Active Safety System-Drive Anti-Slip Control System effectively improves the acceleration performance and driving stability of the car on low-attachment roads. The drive anti-skid control system can be divided into two types: anti-skid control and trajectory control. The anti-skid control system improves the acceleration performance by preventing the automobile from excessively slipping when starting or accelerating when the automobile is on a low-attachment or separated road surface[1]. On the other hand, the trajectory control system prevents excessive lateral acceleration during the steering acceleration process to ensure that the automobile follows the ideal trajectory.

The driving anti-skid control system can be realized by an engine control method or a brake control method. For some systems, the above two control methods can be performed at the same time to optimize automobile performance. However, due to economic reasons, most drive anti-skid control systems only use one control method. This paper mainly studies the drive anti-skid control system that realizes engine torque control by adjusting the throttle opening[2].

[1] Corresponding Author, Shuchena Liu, Shandong Institute of Commerce & Technology, Jinan, Shandong , 250000; E-mail:20110689@sict.edu.cn

2. TCS control principle

Driving force and adhesion coefficient are the two determinants of automobile acceleration performance. Especially in the case of low adhesion roads, it is more important to maintain the maximum adhesion coefficient. Among the factors that affect adhesion coefficient, the factors that are easy to adjust during automobile operation are Slip rate, it is defined as follows: $\lambda = 1 - \frac{V_v}{V_w} \times 100\%$ (1), λ represents the slip rate, V_w expresses the speed of the driving wheel, V_v indicates the speed of the automobile.

The basic method of the anti-skid control system is shown below. According to the definition of slip rate (formula (1-1)), the target driving wheel speed is determined by formula(2):

$$V_w^* = V_v / \left(1 - \lambda^*\right) \tag{2}$$

Among them, represents the best slip rate of the current road conditions, and the automobile speed V_v is estimated by the wheel speed of the non-driven wheels. The is measured by the wheel speed sensor, and the wheel speed sensor is also the main part of the anti-lock brake control system. The wheel speed difference is defined as $e = V_w^* - V_w$. The PI control method is adopted to minimize the wheel speed difference. That is, the PI controller calculates and determines the throttle opening to achieve the desired ideal slip rate. Through the accelerator pedal transmission mechanism to achieve the adjustment of the throttle opening.

3. TCS anti-skid control algorithm

Figure 1-2 is a flowchart of the control decision part. This part monitors the change of slip rate. The slip rate is mainly determined by the measurement data of 7 sensors. The monitoring function is performed during the anti-skid control process and the process of normal driving conditions. The seven measurement data are: accelerator, engine revolution sensor and four wheel speed sensors[3] [4]. The accelerator pedal position sensor measures the accelerator pedal opening set by the accelerator pedal position θ_{APS}, throttle Position Sensor measures the throttle opening set by the accelerator pedal drive system $\theta\left(=\theta_{TPS}\right)$. Note that the fully closed and fully open throttle positions are defined as 0° and 90° respectively. Therefore, during the anti-skid control operation, θ less than θ_{APS}, but in the inactive state, the two angles are equal. On the other hand, ω_{e-rpm} represents the engine speed, V_{fl} and V_{fr} represents the speed of the front left wheel and the front right wheel, V_{rl} and V_{rr} represents, respectively. A lower target slip rate corresponds to a road type with a lower adhesion coefficient.

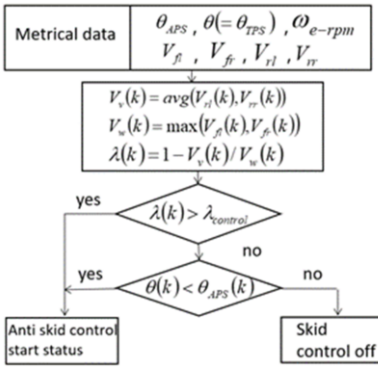


Figure 1. Flow chart of control decision-making part

The average value of the rear wheel speed is the automobile speed (front-wheel drive automobile). Since the slip phenomenon can only occur on one driving wheel (for example: separated road surface), the larger wheel speed of the driving wheels is the driving wheel speed, which is used to calculate the slip rate[5]. From the automobile speed and the driving wheel speed measurement value, the slip rate formula (1) is obtained. When the anti-skid control function is not activated, if the slip rate $\lambda(k)$ at time k is greater than the preset value $\lambda_{control}$, the anti-skid control is activated. Once the anti-skid control is activated, the slip rate will decrease rapidly. In this case, the slip ratio $\lambda(k)$ is smaller than $\lambda_{control}$. On the other hand, at the beginning of the initial control section, the throttle opening is forcibly closed for θ_1 (approximately 4°) T_0 seconds to reduce excessive slip. Because reducing the throttle opening too quickly may affect the comfort of the occupants on the car, it is necessary to gradually reduce the throttle angle to θ_1 in seven steps. At the beginning, even when the accelerator opening is rapidly reduced, the speed of the driving wheels first shows an upward trend due to the time delay caused by the electric motor and the powertrain[6]. In this phase, the speeds are measured, the calculation formula of the average deceleration α_1 is:

$$\alpha_1 = \left(V_{w,\max} - V_{w,\min}\right)/T_1 \tag{3}$$

Among them, T_1 represents the time between the maximum speed and the minimum speed. Note: T_1 is an (in this study, T_s =20m). According to the deceleration α_1, the road can be divided into 3 types. For details, please refer to Table 1. For example, the throttle opening is reduced to θ_1 a point that causes the wheel speed to decelerate faster on road surface C (higher adhesion coefficient) than on road surface A.

After the initial control part ensures the stability of the automobile by reducing the excessive slip of the driving wheels, the control system enters the main control part to improve the acceleration performance of the automobile. First, use the formula (2) with the target slip rate λ^* to calculate the determined target wheel speed V_{w^*} Then, use

the PI controller to calculate and determine the throttle opening to minimize the wheel speed difference $e = V_{w^*} - V_w$.

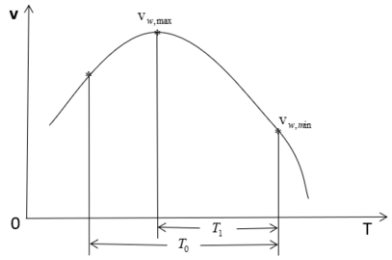


Figure 2. Maximum wheel speed and minimum wheel speed calculated by deceleration

Table 1. Surface conditions of various road types

Pavement quality	Pavement quality 1	Pavement quality 2	Pavement quality 3
Friction coefficient	0.05~0.2	0.2~0.4	0.4~0.6

$$\theta(k) = \theta(k-1) + (K_p + K_1)e(k) - K_p e(k-1) \tag{4}$$

Among them, K_p and K_1 respectively represent proportional gain and integral gain. Pay attention to the adjustment K_p value to ensure K_p decreases as the automobile speed increases. The reason is as follows: For larger automobile speeds, the wheel speed difference e is also relatively large. Therefore, in the case of a larger K_p value, the calculated throttle opening degree may have a larger difference. Due to the large change in engine output torque, it may affect the comfort of the occupants on the car.

4. Anti skid control system implementation

Figure 3 is the configuration diagram of the test automobile used in this test. As shown in Figure 3, its system includes anti-skid controller (with anti-skid control function), throttle transmission system, and a data acquisition board (from All sensors collect data). Use a notebook computer as the main controller to calculate the target throttle opening value and perform.

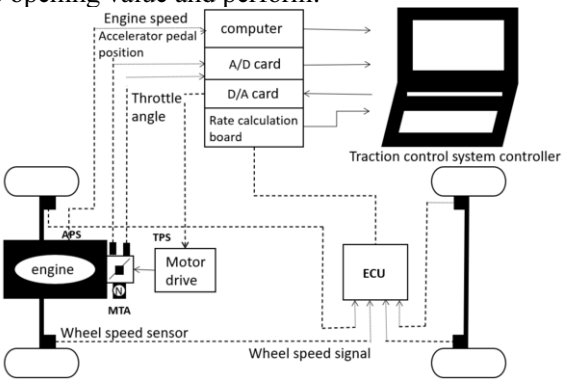


Figure 3. Traction control system anti-skid control system configuration diagram

The throttle drive system of the traction control system integrates the ordinary throttle mechanism. It can be used as an ordinary throttle body part to connect to the accelerator pedal, and can also be used transmission system. In the throttle drive system, the throttle valve is directly connected to the accelerator pedal, and the throttle opening in the fully closed state (θ =0°) and the throttle opening in the fully open state (θ =90) °) to rotate between. The throttle valve position must be within the limits of the throttle fully open and fully closed positions set by the accelerator pedal, and within the limits, the throttle valve can rotate freely until it reaches the angle. Therefore, the main purpose of the throttle transmission at an angular position controlled by the anti-skid controller. In order to drive the DC servo motor effectively, a pulse width modulation signal generator and a pulse width modulation amplifier based on the H-bridge configuration must be configured and connected to the motor. so the period between pulses can be measured and the speed can be calculated.

In order to successfully e of a automobile, an acceleration test method must be adopted[7]. In the process of using this method, the operator presses the accelerator pedal and then releases the brake to make the automobile suddenly accelerate from a stopped state, and control the engine speed to 4000 rpm. Measure the speed and trajectory within 10 seconds, and use these measurement data to evaluate the acceleration performance of controlled automobiles and uncontrolled automobiles. The three road types listed in Table 1 were tested in this experiment. Road type A simulates a frozen road, road type B simulates a snowy road, and road type C simulates a wet asphalt road. The weight of the test automobile is about 1500kg, including the operator and the total weight of the test equipment, about 1650kg.

5. Test results and analysis

Carry out automobile tests on various road types with different adhesion coefficients listed in Table 1. Some typical test results are as follows. Figure 4 illustrates the response of the automobile on road type B when the anti-skid control mechanism is not activated. For safety reasons, the throttle opening is limited to approximately 20°. When the engine speed reaches about 4000 rpm, quickly connect the engine and the transmission mechanism. As shown in Figure 4, the driving wheels slip excessively and cannot produce enough traction. Therefore, the average acceleration is only 0.10g in about 10 seconds, and the automobile speed reaches 35km/h in about 10 seconds.

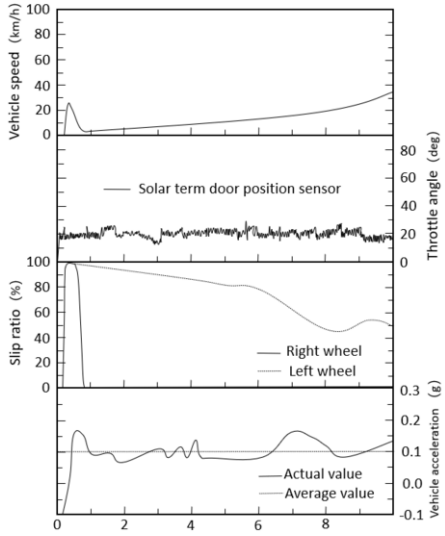


Figure 4. The anti slip control system is not activated and the vehicle is running

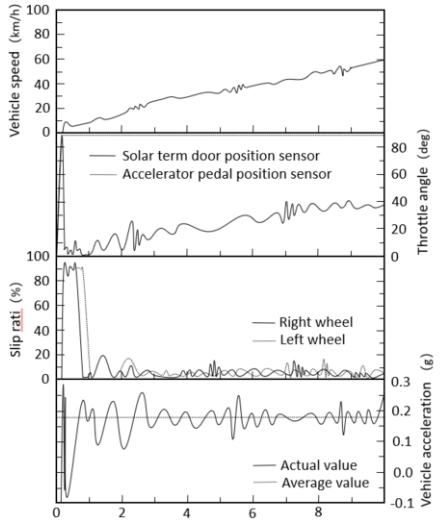


Figure 5. Anti skid control system start vehicle running status

Due to the high sensitivity of the driving wheels to low-adhesion roads, a large speed difference between the left and right wheels still occurs even on a road with a balanced surface. As shown in Figure 4, partly due to the differential speed, the speed of the left front wheel is much greater than that of the right front wheel. The speed difference will lead to side-to-side swaying phenomenon (automobile swaying phenomenon often occurs when turning), so the directional stability of the automobile is reduced.

Figure 5 The automobile response when the anti-skid control logic is activated During the entire test, the accelerator pedal was forced to be at the maximum acceleration position[8]. This is the most unsatisfactory working condition, so this

method is rarely adopted in the actual environment, especially on low-adhesion roads. Due to this extreme condition, the driving wheel has excessive slip, resulting in a slip rate as high as 90%. However, due to the activation of the anti-skid control function, the throttle opening rapidly decreased to increased to achieve acceleration performance[9] [10]. Compared with the automobile test results without TCS control (as shown in Figure 4), the average acceleration becomes 0.18g, an increase of 80%. The test used methods to reduce excessive slip and maintain the optimum slip rate corresponding to the optimum traction. The automobile speed reached 60km/h in about 10 seconds.

As shown in Figure 5, the acceleration of the automobile has changed significantly (there is no major change in Figure 4). Since the throttle valve angle of the controlled automobile reaches the fully open position, the torque output of the engine reacts violently, which in turn causes a large change in the longitudinal motion of the automobile[11]. When the automobile starts, if the acceleration of the automobile changes significantly, it may affect the driving comfort. The above test conditions are extreme conditions. In actual acceleration conditions, almost no one may step on the position, so this problem is unlikely to occur.

The road type estimated by the initial control part at the beginning of the anti-skid control matches the actual road type. The test data of road type A and separated road surface show that the test results are similar, and the acceleration performance of controlled automobiles is significantly better than that of uncontrolled automobiles. The speed of the uncontrolled automobile on road type A is 22km/h, and the speed of the controlled automobile is 47km/h. The speed of the uncontrolled automobile on the separation road is 28kw/h, and the speed of the controlled automobile is 57knw/h.

6. Conclusion

This paper has obtained and verified the anti-skid control algorithm based on engine throttle control through a series of automobile tests.

- The acceleration performance and stability of automobiles using anti-skid control algorithms are significantly better than those of non-controlled automobiles. Road recognition, target slip rate adjustment and control gain decision-making are effective for various road types and have strong applicability to different automobile speeds.
- On low-adhesion roads, the side-to-side swing phenomenon caused by the unbalanced power distribution leads to an increase in the accident rate. The use of anti-skid control systems can significantly reduce this phenomenon.

References

[1] Yamada K, Hashiguchi M, Ito M. Traction ControlSystem-Simulation Analysis of the Control Sys - tem. International Journal of Automobile Design. 1991; 12(1): pp 89-96
[2] Wrong JY. Theory of Ground Automobiles. John Wiely & Sons; 1978.

[3] Göhring E. Tractor/Sermitrailer Anti-Lock Performance and Compatibility asseen by the Commercial Automobile Manufacturer. IME-Conference "Anti-Lock Braking Systems for RoadAutomobiles". London. 1985.
[4] Ise I, Fujita K, Inoue Y, Masutonmi S. The Lexus Traction Control (TRAC) System . SAE Paper900212,1990
[5] Mischke A. Antiblockiersystem f ü r Nutzfahrzeuge (Anti-Lock System for conmercial Automobiles ahd Bu - ses). VDI - Berichte ,1981; p 418.
[6] Tanaka T, Isoda K, Ohsaki M, Shigehara. Traction control System for Improved Driving Safety. SAE. 1991; p 912583.
[7] Göhring E. Elktonsch pnumatiche Gtileschatug EPS für vistuie mechanische NuztarzeueSchaltgetriebe (Electronic Pneumatic Gear-Shift EPS for Multi-SpeedMechanical Transmissions for Commercial Automobiles) Automobil- Industrie 6/88.
[8] Song JB, Hong DW. HILS Implementation of TCS Slip Control System Based on Throttle Adjust - ment Approach . Journal of Korea Society of Automotive Engineers. 1998; 6(3): pp .45-53.
[9] Song JB, Kim HJ, Byeon KS. Position Control System for Engine Throttle Actuator. Proc . of the 2nd Asian Control Conference. 1997; 3: pp .383-386.
[10] Bader S. EPS - Elektronishch pneumatische Schaltung - Wirtschaftlich fahren mit erhöhter aktiver Sicherheit durch Elektronik (EPS Electronic -Pneumatic Gear -Shift- Cost Effective Operation with Enhanced Active Safety through Electronics). VDI- Berichte, 1986; p 612.
[11] SAE. SAE Handbook,1996.

doi:10.3233/ATDE220255

Research on SLAM Method of Intelligent Vehicle Based on Lidar-IMU Combination

Yanqi LIU[1, a], Yanqiu XIAO [a], Guangzhen CUI [a], Qidi SUN [a] and Weili ZHANG [a]
[a]*College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou*

Abstract. Simultaneous localization and Mapping is a key technology in the field of unmanned vehicle technology. Aiming at the present stage laser SLAM technology of gravity vector drift and elevation estimation error problem, based on the carrying of inertial measurement unit (IMU) and laser radar as experiment platform, Apollo suite with NDT method as the theoretical basis, in the Cyber RT operating system environment realize the car to the scene at the same time localization and map building. Firstly, the time stamp registration of laser and IMU is realized by calibrating the lidar. Secondly, IMU pre-integration is introduced to eliminate the motion distortion of point cloud. Then, reliable plane and edge features are obtained by feature extraction, and the depth information of environmental features is obtained. Finally, the current pose is estimated by NDT algorithm. The experimental results show that the map constructed by the combined SLAM method based on NDT algorithm has high precision and strong stability.

Keywords. Lidar, Inertial Measurement Unit, NDT, SLAM

1. Introduction

Camera and lidar are the two most frequently used sensors in intelligent vehicles, but the sensitivity of camera to illumination and viewpoint changes makes the mapping results unreliable. Lidar range measurement is accurate, data collection is not affected by environmental lighting, and the error model is clear. Its high resolution allows the fine details of the environment to be captured from a distance in a large aperture range, and it can efficiently complete the mapping task even at night. Therefore, real-time state estimation of intelligent vehicle by using 3D lidar is the focus of intelligent driving research at the present stage [1].

LIDAR-SLAM is the application of LiDAR to collect point cloud data from the surrounding environment, Based on the Iterative Closest Point[2],Correlation Scan Match[3] or Normal Distribution Transformation[4] to process point cloud data, solve the current pose, and finally output the environment map. With the maturity of lidar technology, SLAM technology turns to the practical stage. The Gmapping model[5] proposed by Giorgio, effectively solves the particle dissipation problem of particle filter, but this model is extremely dependent on external odometer. To solve this problem, Stefan[6] proposed Hector-SLAM model that does not rely on odometer data,

[1] YanQi LIU, College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Zhengzhou, China; E-mail: liuyanqi0816@163.com.

but it is difficult to deal with map closed-loop. In the KITTI Open, Zhang published LOAM[7] models suitable for 3D LiDAR and extensive outdoor scenes, but lacked closed-loop detection. With the rapid development of autonomous driving technology, SLAM technology has been gradually applied, and the current research is more inclined to solve the problems existing in practical application. Google released the open source SLAM model Cartographer[8] in 2016, which mainly applied 2D LiDAR indoor SLAM estimation to effectively solve the problems of indoor localization and map reconstruction. Shan et al. proposed LEGO-LOAM [9]model based on LOAM model, which adopted ground segmentation method and point cloud clustering method to make the extracted feature points more effective. Qi Yutao [10] used the scale-invariant feature method to optimize the initial positioning problem of NDT algorithm and improve the success rate of point cloud positioning.

To sum up, the primary problem to be solved in unmanned driving is vehicle positioning. Only after the position of the vehicle is determined in the high-precision map according to the positioning information, a reasonable driving route can be planned for the unmanned vehicle, and vehicle control and attitude adjustment can be carried out according to the current position. In order to further improve the robustness of 16-line lidar equipment in outdoor road environment, this paper firstly calibrated the lidar based on Apollo development kit to complete point cloud data collection. Secondly, the motion of the initial point cloud is corrected by IMU observations. Then, feature extraction was used to obtain reliable plane and edge features. Finally, NDT optimization algorithm was used to estimate the current pose, and Apollo development kit was used for experimental verification.

2. Data processing

2.1. Coordinate system definition

According to the installation position of lidar and IMU in Apollo suite, the radar coordinate system, IMU coordinate system and vehicle coordinate system are established. Through the position relations of different coordinate systems, the coordinate conversion relations of laser points are obtained. The coordinate system position relations are shown in Figure 1.

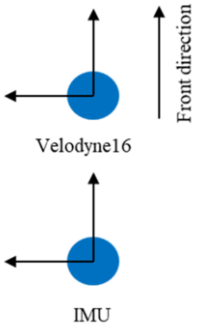
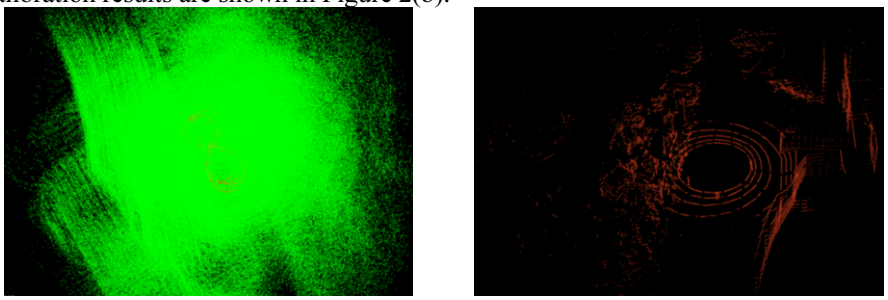


Figure 1. Sensor coordinate system position

2.2. Sensor calibration

Real-time data acquisition is carried out through fuel-client for radar calibration, and point cloud data is recorded in octagonal loop based on lidar and GNSS data, as shown in Figure 2a. After calibration, point cloud calibration data is processed by BOS, and the calibration results are shown in Figure 2(b).



(a) Point cloud calibration process (b) Point cloud calibration results

Figure 2. Lidar calibration

2.3. Point cloud information preprocessing

In this paper, the observation results of IMU are used as the initial value of lidar motion estimation to correct the point cloud [11], and the algorithm flow chart is shown in Figure 3.

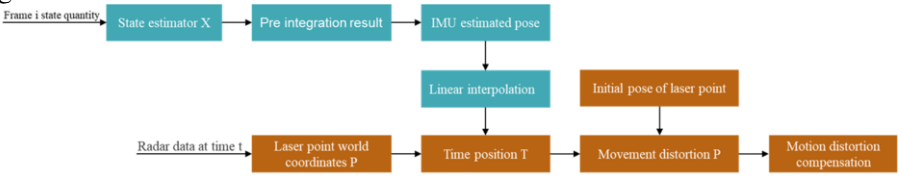


Figure 3. Point cloud motion correction

The original data ω_j and α_j of IMU can directly obtain an estimated pose through time integration, but the results of such calculation often have great observation noise. Therefore, according to the pre-integral estimation method, IMU state is defined as:

$$X_j^W = \left[p_j^W \quad v_j^W \quad q_j^W \quad b_a^T \quad b_\omega^T \right] \tag{1}$$

The state quantity of IMU is obtained by discrete estimation of the current IMU data and the state quantity of frame i, and its expression is as follows:

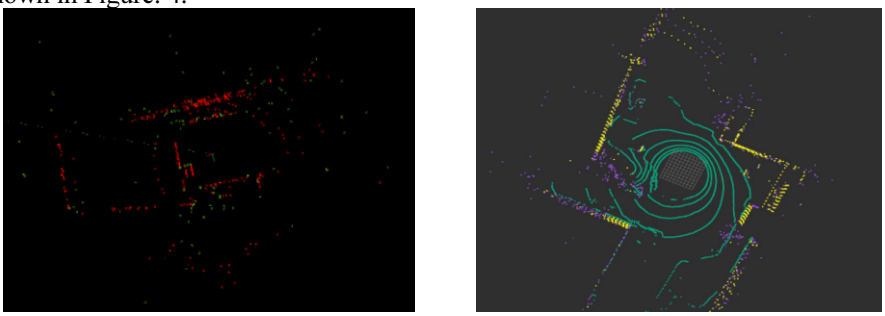
$$p_j = p_i + \sum_{k=i}^{j-1} \left[v_k \Delta t + \frac{1}{2} g^W \Delta t^2 + \frac{1}{2} R_k \left(a_k - b_a \right) \Delta t^2 \right] \tag{2}$$

$$v_j = v_i + g^W \Delta t_{ij} + \sum_{k=1}^{j-1} R_k \left(a_k - b_a \right) \Delta t \tag{3}$$

$$q_j = q_i \otimes \coprod_{k=i}^{j-1} \delta q_k = q_i \otimes \prod_{k=i}^{j-1} \begin{bmatrix} \frac{1}{2}\Delta t(\omega_k - b_\omega) \\ 1 \end{bmatrix} \tag{4}$$

For IMU data, a linear interpolation method is adopted for Lidar data of every frame. All laser points of the current frame are converted to the Lidar coordinate system at the starting time, and the pose transformation between the current frame and the starting frame is subtracted for each laser point, so as to compensate the motion distortion caused by its own motion.

The reliability of feature points and the accuracy of feature extraction can be improved by segmenting the pre-processed point cloud. The points with the same characteristics are divided into clusters. The grounding point is a special type of cluster. Then, the clusters with points less than 40 in the cluster are removed to leave more geometric targets in the sampling points and effectively remove the influence of noise in the environment. The changes before and after point cloud information processing are shown in Figure. 4.



(a) Unprocessed point clouds (a) Point cloud after processing

Figure 4. Point cloud information processing

3. Research on Lidar-IMU combined positioning method based on NDT algorithm

3.1. Lidar -IMU optimization method

Due to the sparse ground feature points of LiDAR data, direct matching can easily lead to wrong estimation of gravity vector direction. Therefore, the algorithm adopts the joint optimization method based on LiDAR data constraint and IMU state constraint, which can effectively constrain the gravity vector direction in pose estimation. Meanwhile, LiDAR data constraint method can also eliminate the high frequency noise in IMU data. The data processing process is shown in Figure 5.

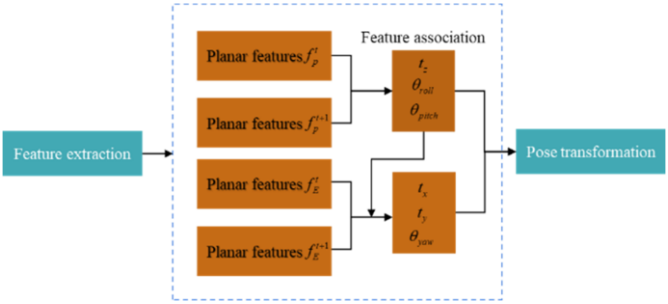


Figure 5. Combined Lidar-IMU estimate

Since LiDAR data itself does not have velocity and acceleration observation data, the target of optimization is 6-dOF pose T, then

$$T_k^L = \left(t_x, t_y, t_z, \theta_{roll}, \theta_{pitch}, \theta_{yaw}\right) = \left(p_k, q_k\right) \tag{5}$$

Where, L represents the current carrier coordinate system; p_k Represents the current location information; q_k Represents the current Angle information. Construct the joint optimization equation:

$$\min_{T_k^L} \frac{1}{2}\left\{E_L^2 + E_B^2\right\} \tag{6}$$

Finally, L-M method is adopted for iterative calculation, and the initial value of iteration T_k^L is the estimated position and attitude of the current IMU state. In this way, the number of iterations can be effectively reduced, and the wrong LiDAR feature matching caused by the assumption of zero initial value or uniform initial value can be eliminated, which is more consistent with the real motion.

3.2. Research on COMBINED SLAM based on NDT algorithm

Based on NDT algorithm, motion pose estimation of 16-thread Velodyne Lidar and six-axis IMU was carried out. Due north direction could be obtained through GNSS, and the yaw Angle of the final output was the included Angle with due north direction. The continuous frame laser point cloud is registered by NDT, and the registration result is used as the initial value of the point cloud registration of the next frame to obtain the current pose, and then the Transform relative to the initial frame can be calculated. Finally, the point cloud of the current frame is aligned to the origin of coordinates, and down-sampling is added into the map frame by frame to complete the map construction. The overall process is shown in Figure. 6.

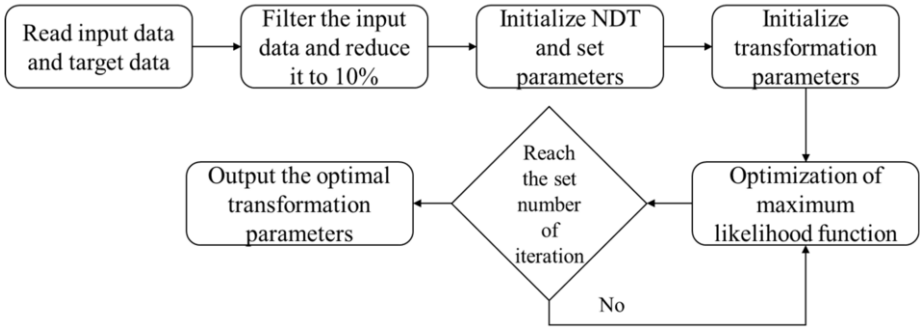


Figure 6. NDT algorithm flow

4. Lidar - IMU combined SLAM test

In order to test the accuracy and effectiveness of the algorithm, a real-time perception experiment was carried out in the closed park based on Apollo autonomous driving development kit to observe the actual driving situation of intelligent vehicles and evaluate the mapping results according to the data collected during the test.

Based on Apollo autonomous driving development kit, the real car test was carried out in the actual scene, information was collected and preprocessed for the park environment, and then the processed point cloud was registered through NDT, and finally the segmented map was obtained.

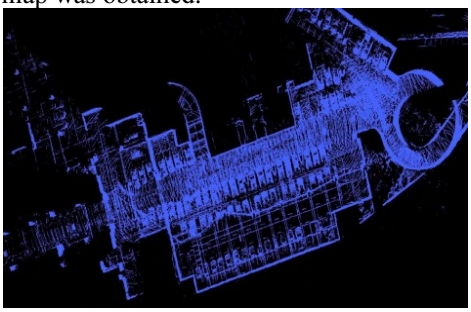


Figure 7. Results of Lidar-IMU navigation combined mapping

As can be seen from Figure 7, compared with traditional feature point method, NDT has the advantage of fast matching speed, which can be understood as uniform sampling, and can retain relatively complete details of point cloud. In structured environment with many trees, such as the scene of empty square, appropriate feature points can be quickly extracted, resulting in excellent image construction effect.

5. Conclusion

Aiming at the environment characteristics of the closed park, a laser-inertial navigation combined SLAM method for intelligent vehicle was proposed to obtain the information of the surrounding environment of intelligent vehicle. The actual scene is tested based on Apollo development suite, and the results show that the NDT algorithm has high precision and strong stability.

References

[1] HE Jia, Rong Hui, et al. A review of baidu and Google's development of driverless cars[J].Automotive Electrical Appliances, 2017, (12): 19-21.
[2] C Yang, Medioni G. Object modeling by registration of multiple range images[J]. Image and Vision Computing, 2002, 10(3): 145-155.
[3] OIson-E B. Real-time correlative scan matching [C]//Proc of IEEE International Conference on Robotics and Automation. Piscataway, NJ:IEEE Press, 2009: 4387-4393. 2019: 4387-4393.
[4] Strasser-W Biber P. The normal distributions transform: a new approach to laser scan matching [C]//Proc of IEEE//RSJ International Conference on Intelligent Robots and Systems. Piscataway, NJ: IEEE Press, 2003: 2743-2748.
[5] Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with rao-blackwellized particle filters[J]. IEEE transactions on Robotics, 2007, 23(1): 34-46.
[6] Kohlbrecher S, Von Stryk O, Meyer J, et al. A flexible and scalable SLAM system with full 3D motion estimation[C]//2011 IEEE international symposium on safety, security, and rescue robotics. IEEE, 2011: 155-160.
[7] Zhang J, Singh S. Low-drift and real-time lidar odometry and mapping[J]. Autonomous Robots, 2017, 41(2): 401-416.
[8] Hess W, Kohler D, Rapp H, et al. Real-time loop closure in 2D LIDAR SLAM[C]//2016 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2016: 1271-1278.
[9] Shan T, Englot B. Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 4758-4765.
[10] YuTao Qi. Research and Improvement of NDT algorithm for initial location based on Point Cloud Scale Invariant feature [D]. Lanzhou University, 2020.
[11] Bogoslavskyi I, Stachniss C. Efficient online segmentation for sparse 3d laser scans[J]. PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 2017, 85(1): 41-52.

Subject Index

This page intentionally left blank

Author Index