J. Meas. Sci. Instrum., 2025, 16(1): 119-127. http://jmsi.publish.founderss.cn/
DOI: 10.62756/jmsi.1674-8042.2025012 jmsi@nuc.edu.cn

Non-cooperative target extraction in complex industrial environment based on image segmentation

WU Xiaojun*, WANG Peng, ZHAO He, YU Xianzhe, LI Tiancheng

School of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

***Corresponding author:** WU Xiaojun (wuxiaojun@xauat.edu.cn)

Received: October 25, 2023 **Revised:** January 12, 2024 **Accepted:** January 18, 2024

Abstract: In complex industrial scenes, it is difficult to acquire high-precision non-cooperative target pose under monocular visual servo control. This paper presents a new method of target extraction and high-precision edge fitting for the wheel of the sintering trolley in steel production, which fuses multiple target extraction algorithms adapting to the working environment of the target. Firstly, based on obvious difference between the pixels of the target image and the non-target image in the gray histogram, these pixels were classified and then segmented in intraclass, removing interference factors and remaining the target image. Then, multiple segmentation results were merged and a final target image was obtained after small connected regions were eliminated. In the edge fitting stage, the edge fitting method with best-circumscribed rectangle was proposed to accurately fit the circular target edge. Finally, PnP algorithm was adopted for pose measurement of the target. The experimental results showed that the average estimation error of pose angle γ with respect to the z-axis rotation was 0.234 6°, the average measurement error of pose angle α with respect to the x-axis rotation was 0.170 3°, and the average measurement error of pose angle β with respect to the y-axis rotation was 0.227 5°. The proposed method has practical application value.

Key words: digital image processing; industrial environment; non-cooperative target; pose measurement

0 Introduction

Highly-intelligent technology is important for upgrading of industrial production. Although some machines can automatically work with pre-defined actions, they are expected to be more flexible. Visual servo technology may effectively solve this problem. Currently, monocular vision systems have been applied to three-dimensional pose measurement in intelligent industrial production owing to their low cost, flexible application, and high positioning precision[1,2]. With no need for stereo matching of binocular vision systems, they only require target acquisition cameras in actual applications and are featured as non-contact and wide collection[3]. Generally, the monocular vision system is divided into cooperative target monocular vision system and non-cooperative target monocular vision system. The former marks the targets in advance to extract feature points, while the latter is usually for the targets that are not convenient to mark, which makes the extraction of feature points relatively difficult[4].

Currently, the monocular vision technology can meet the positioning precision demands of most industrial scenes and its system is simple and flexible, so it has attracted much attention from relevant researchers. Miao et al.[5] proposed a monocular vision positioning algorithm based on the least squares Gaussian kernel support vector machine, which transforms the nonlinear mapping of image coordinates and world coordinates into the linear mapping. Mei et al.[6] proposed a method to calculate the relative pose between camera and target through the spatial geometric coordinates of special feature points of non-cooperative targets, which reduces the computation and maintains the measurement precisions without interference. Lin et al.[7] proposed a binocular vision pose measurement based on the spatial non-cooperative targets with distinguishing features. The movement state of the target in a space-like weightlessness environment was realized through a ground air-floating experimental platform, and pose measurement and ground verification test were carried out in the local area of the spacecraft with significant characteristics. Wang et al.[8] put forward a pose measurement algorithm determined by the topological structure of multiple feature points. In the large-angle motion of the targets, multiple feature points enable pose calculation with enough feature points, therefore, it is more precise than that with few feature points. Furthermore, it elaborates on the relationship between the

selection of feature points and the positioning precision. Xu et al. [9] estimated the vehicle's pose by combining the monocular camera with the vehicle's 3D model, and improved precision measurement by using the voting method to minimize the key point errors. The above studies results show that although pose precision and computational efficiency are important for monocular vision system, the extraction of high-quality feature points is also a key factor for pose measurement.

For monocular vision technology, the extraction of feature points emphasizes segmented extraction for the targets to be detected, involving the traditional target extraction with feature extraction as the core and the deep learning with satisfactory effects in various fields[10,11]. However, since the deep learning requires abundant learning samples and hardware equipment with higher requirements[12], in a special environment, the traditional mode is an optimal choice of target extraction. For example, Guo et al. [13] put forward a new method integrating hierarchical clustering and image segmentation, which solves the under-segmentation to a certain degree, but it could not fully avoid interference factors at the target edge in a complex industrial environment. Tian et al.[1] put forward an edge extraction method to achieve high-precision target acquisition by classifying images with *k*-means and conducting intraclass segmentation with threshold segmentation. However, the selection of cluster center is random, making intraclass segmentation effect unstable and the stability requirement of complex industrial equipment hard to be met. Although a simple threshold value method could achieve better segmentation results under an ideal status, the optimum effects are hard to be realized due to the changes of the ray of light and background environment[14,15]. In a complex background, the image segmentation based on inter-class variance involves a large amount of computation, resulting in poor timeliness and poor robustness[16].

Comprehensive evaluation shows that the existing extraction methods are difficult to acquire the targets stably in a complex industrial environment. In this study, we integrated multiple modes to realize target extraction and edge fitting. Firstly, The target pixels were clearly distinguished in the histogram and then classified. Afterwards, they were segmented in the class with the interference factors removed and the targets retained. Finally, multiple segmentation images were combined into a final target image. In case of edge defects, a best-circumscribed rectangle for edge fitting based on current situation and the circular feature of the target could be employed to collect feature points for pose measurement.

1 Methods

Multiple tests proved that an individual method fails to comprehensively consider global information and local information to achieve accurate extraction of complex industrial targets. In this study, we proposed a new method for target extraction, segmentation and edge fitting, which integrates multiple image processing algorithms. As shown in Fig.1, the input image of the target is divided into three classes: background, foreground and highlight in gray histogram, and the intraclass pixels are segmented with Otsu threshold value. Then, the relevant results are combined to obtain the final target image after removing small connected regions. During the edge fitting period, the best-circumscribed rectangle is adopted to perform the accurate fitting on the ellipse's target edge.

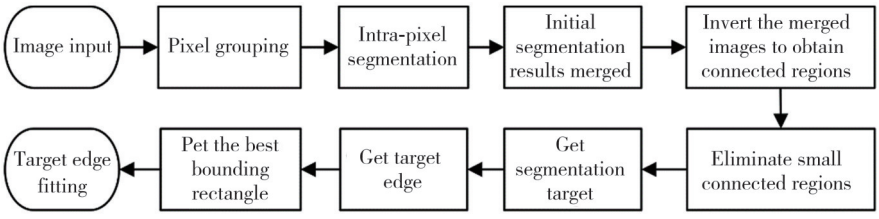


Fig. 1 Target extraction, image segmentation and edge fitting processes

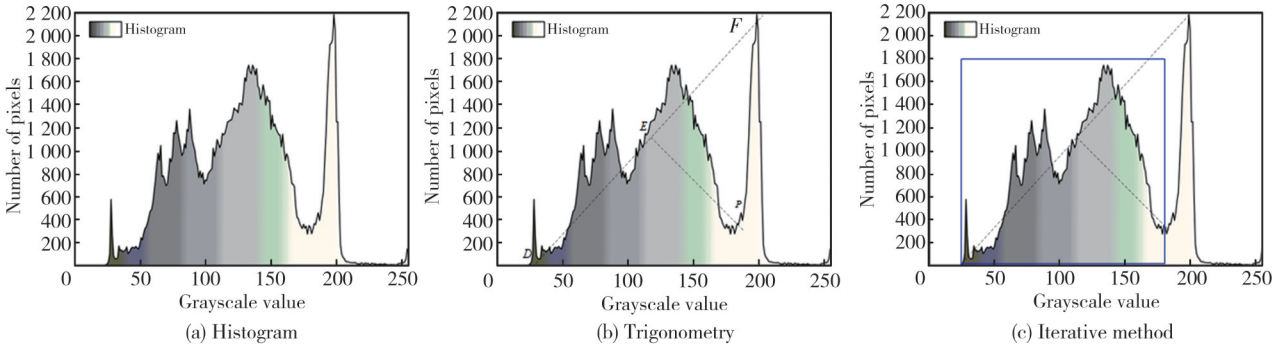
Since traditional methods rely on global information or lack of local information, the targets in a complex and varied industrial environment could not be extracted effectively. The innovation of our method lies in strong robustness on strong background environment and dazzling light. The images are divided into the background, foreground, and highlight based on the gray histogram of global information. Different categories of pixels could be segmented in the class with the Otsu algorithm.

The highlighted parts from complicated illumination and other non-target pixels' interference are eliminated, following by small non-target connected regions being eliminated. Finally, the images are reversed to obtain the target. Such a method is characterized by rapid processing speed and satisfactory effects.

The ellipse fitting with the best-circumscribed rectangle adapting to the target could avoid edge burr effectively, and improve the instantaneity precision of pose measurement.

1.1 Pixel grouping

The gray histogram reflects the distribution of pixel values of the image, and the valley between two peaks is the general dividing point between the background and the foreground according to the principle of threshold segmentation. As shown in Fig.2, the two valleys with the largest difference between peak and valley in the image are taken as the dividing pixels, which avoids the loss of pixels of the target. Based on the histogram, the pixels of the image are grouped into background, foreground and highlight.



(a) Histogram (b) Trigonometry (c) Iterative method

Fig. 2 Calculation and grouping of pixels via gray histogram

In Fig.2(b), feature points are $D(X_b, Y_b)$, $E(X_f, Y_f)$, $F(X_{max}, Y_{max})$ and $P(x_0, y_0)$. The largest difference between peak and valley comes from points $D(X_b, Y_b)$ and $F(X_{max}, Y_{max})$, and then we can get

$$\frac{Y - Y_b}{Y_{max} - Y_b} = \frac{X - X_b}{X_{max} - X_b}. \tag{1}$$

Let $Y_{max} - Y_b = A$, $X_b - X_{max} = B$, $Y_b X_{max} - X_b Y_{max} = C$, Eq.(1) can be rewritten as

$$Ax_0 + By_0 + C = 0, \tag{2}$$

where X_b is the gray value of background pixels as the abscissa in gray histogram, Y_b is the number of background pixels as the ordinate in gray histogram, X_f is the gray value of foreground pixels as the abscissa in gray histogram, Y_f is the number of forground pixels as the ordinate in gray histogram, X_{max} is the gray value of highlighted pixels as the abscissa in gray histogram, and Y_{max} is the number of highlighted pixels as the ordinate in gray histogram.

The area with the gray value corresponding to the farthest distance between the pixel points in the histogram and the straight line that belongs to highlighted area shown in Fig.2(b) is expressed as

$$\begin{cases} d = \left| \dfrac{Ax_0 + By_0 + C}{\sqrt{A^2 + B^2}} \right|, \\ Ax_0 + By_0 + C \leqslant 0. \end{cases} \tag{3}$$

Besides, for a point is under a straight line, its gray value is determined, therefore, it is taken as the classification threshold of the highlighted area and denoted as T_2.

After the highlighted area is segmented, the remaining pixels belong to the foreground or background, as shown in the box in Fig. 2(c). The threshold segmentation is conducted for the image, with the largest difference between the average gray levels of the foreground area, the background area and the whole image, which can be expressed by regional variance as

$$\begin{cases} W_b(T^t) = \sum_{i=0}^{T_1 - 1} p(i), \\ \mu_b = \dfrac{\sum_{i=0}^{T_1 - 1} p(i) i}{W_b(T^t)}, \end{cases} \tag{4}$$

$$\begin{cases} W_f(T^t) = \sum_{i=T_t}^{T_2 - 1} p(i) = 1 - W, \\ \mu_f = \dfrac{\sum_{i=T_t}^{T_2 - 1} p(i) i}{W_f(T^t)}, \end{cases} \tag{5}$$

$$\mu_0 = \mu_f W_f + \mu_b W_b, \tag{6}$$

$$\sigma^2 = W_f W_b (\mu_b - \mu_f), \tag{7}$$

where $W_b(T^t)$ is the distribution probability of the background pixels; T^t is the classification threshold of the pixels; $p(i)$ indicates the distribution probability of the pixels at the gray level i; μ_b means the average gray value of the background pixels; μ_f refers to the average gray value of the foreground pixels; and σ^2 is the variance between two categories of pixels.

For the background and foreground threshold points, the variance between two categories of pixels is calculated and the threshold corresponding to the pixel with the largest intraclass variance is the best classification point, thus obtaining the background and foreground classification threshold as T_1.

1.2 Target extraction

The sintering trolley is transportation equipment for iron

and steel production enterprises to transport ore powder. The sintering trolley works in an environment with high dust intensity and heavy oil pollution, which increases the difficulty of identifying non-cooperative targets.

Fig.3(a) shows a pallet wheel covered with a cloud of dust and oil stain. Since it works in a semi-closed environment, there are strong light facing the lens in the wheel gap in the sunshine and searchlights interfering with the lens in the workshop at night-time. To solve the above problems, the images are divided into the background, foreground, and highlighted pixels according to the histogram information, as shown in Fig.3(b). It can be clearly observed that the target edge contains three categories of pixels. After threshold segmentation, only the target area containing the highlight and foreground is reserved while the background is removed, as shown in Fig. 3(c). Furthermore, the highlight is removed and the target area with some small connecting areas is reserved, as shown in Fig. 3(d). Fig. 3(e) shows the target area removing the background. Finally, the complete target edge is obtained via edge extraction and edge connection, as shown in Fig.3(f).

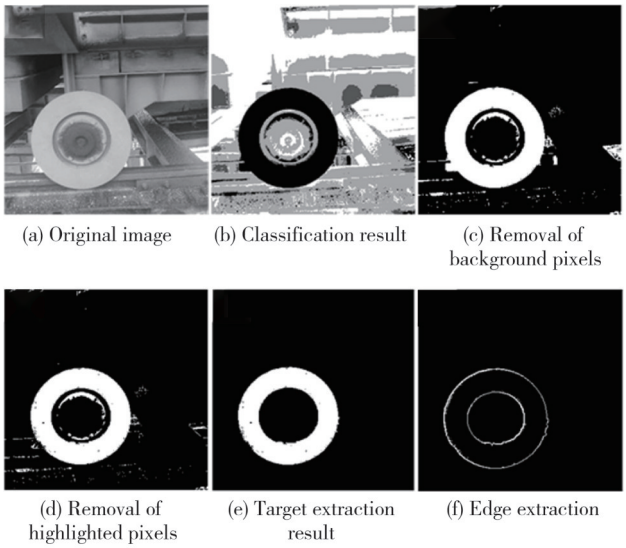


Fig. 3 Image segmentation

Fig. 4 presents the segmentation effects of different algorithms on the target.

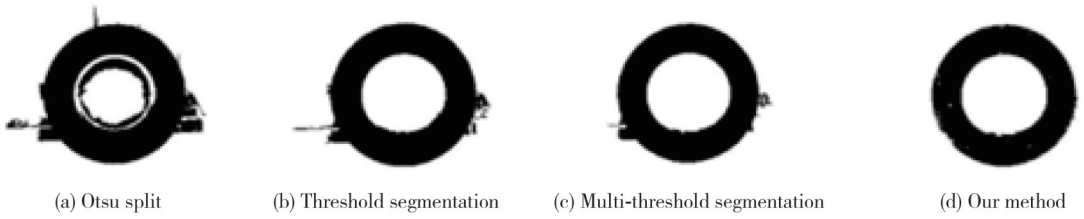


Fig. 4 Segmentation results using different methods

It can be seen that the image in Fig. 4(a) does not shield the highlighted area and contains a similar background, the image in Fig. 4(b) shields the interference from similar background but fails to suppress the highlight area, and the image in Fig. 4(c) suppresses the similar background but lacks of consideration of local information and effective processing of edge highlight. Compared with the above methods, our method achieves better segmentation effects by integrating multiple solutions to classify the image first based on global information and then extract the segmentation target based on local information, whereas other methods could not avoid the interference from external factors because they segment the images only considering global information, as shown in Fig.4(d).

1.3 Best-circumscribed rectangle

Since the target image obtained by proposed image segmentation method has incomplete edge and burrs, a reconstruction method based on the best-circumscribed rectangle and ellipse of edges is proposed considering target-oriented particularity, timeliness and robustness in actual applications, avoiding the error caused by false edges in the detection, as shown in Fig.5.

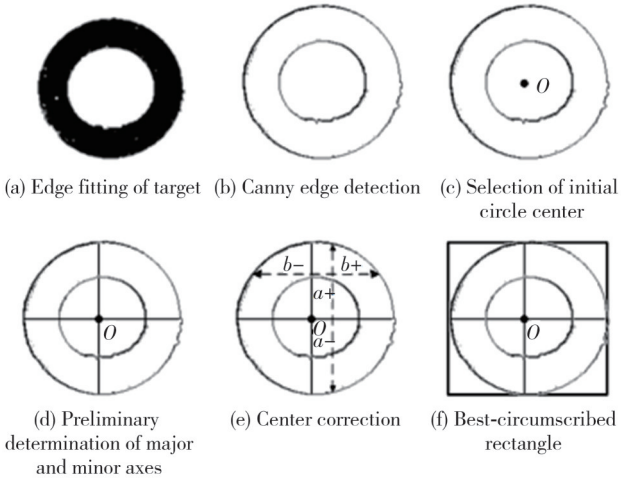


Fig. 5 Ellipse fitting

The use of the Canny edge detection operator for extraction of target edge improves system precision, and

edge pixels are connected to improve elaborate degree of the edges.

For the edge obtained, the pre-selected center coordinates of the fitting ellipse can be calculated according to the coordinates of its edge pixel point $P(x_i,y_i)(i=1,2,\cdots,n)$, which is expressed as

$$\begin{cases}\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i\\ \bar{y}=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i\end{cases}\Rightarrow(x_i,y_i), \tag{8}$$

where (x_i,y_i) are the initial values of the ellipse center and $(\bar{x},\bar{y})$ are the mean values of the coordinates of edge pixels. Hence, the pre-selected major and minor axes of the fitting ellipse could be calculated.

Considering the research object is the pallet wheel continuously running at a constant speed on the circular track, its projection on the plane can be regarded as a circle rotating around an axis perpendicular to its plane. The major and minor axes of the ellipse projected are parallel to the coordinate axes, and the length and width of the circumscribed rectangle equal the corresponding major and minor axes of the ellipse, respectively. It can be expressed as

$$\begin{cases}x'=\bar{x},\\ y'=\bar{y},\end{cases} \tag{9}$$

where x' is the major axis to be fitted, and y' is the minor axis to be fitted.

After calculating the distance from each edge point to the major and minor axes of the ellipse, the central position of the ellipse could be corrected according to the dispersion of distance values. Thus, the interference of edge burrs and mutilation can be removed according by calculating the distance from the half-axis pixel points to the major and minor axes of the ellipse.

$$\begin{cases}a^{+}=\sqrt{(y_i-\bar{y})^2+{x_i}^2},(y_i>\bar{y}),\\ a^{-}=\sqrt{(\bar{y}-y_i)^2+{x_i}^2},(y_i<\bar{y}),\end{cases} \tag{10}$$

$$\begin{cases}b^{+}=\sqrt{(x_i-\bar{x})^2+{y_i}^2},(x_i>\bar{x}),\\ b^{-}=\sqrt{(\bar{x}-x_i)^2+{y_i}^2},(x_i<\bar{x}),\end{cases} \tag{11}$$

where a^{+} is the distance from the upper edge point of the ellipse to the minor axis, a^{-} is the distance from the lower edge point of the ellipse to the minor axis, b^{+} is the distance from the right edge point of the ellipse to the major axis, and b^{-} is left edge point of the ellipse to the major axis.

Let $M=a^{+}-a^{-}$, $N=b^{+}-b^{-}$, we get the matrix as

$$\begin{bmatrix}M\\N\end{bmatrix}=\begin{bmatrix}a^{+}-a^{-}\\b^{+}-b^{-}\end{bmatrix}. \tag{12}$$

Let the median of M and N be $\begin{bmatrix}M_{0.5}\\N_{0.5}\end{bmatrix}$, the deviation of M and N from the median value is calculated by

$$\begin{bmatrix}\Delta M\\\Delta N\end{bmatrix}$$

$$\begin{bmatrix}\Delta M\\\Delta N\end{bmatrix}=\begin{bmatrix}M\\N\end{bmatrix}-\begin{bmatrix}M_{0.5}\\N_{0.5}\end{bmatrix}. \tag{13}$$

The median of ΔM and ΔN are denoted as $\begin{bmatrix}\Delta M_{0.5}\\\Delta N_{0.5}\end{bmatrix}$, and then we can get

$$\begin{bmatrix}T_l\\T_s\end{bmatrix}=\begin{cases}\begin{bmatrix}M_{0.5}\\N_{0.5}\end{bmatrix}+n\begin{bmatrix}\Delta M_{0.5}\\\Delta N_{0.5}\end{bmatrix}, & \begin{bmatrix}M\\N\end{bmatrix}>\left(\begin{bmatrix}M_{0.5}\\N_{0.5}\end{bmatrix}+n\begin{bmatrix}\Delta M_{0.5}\\\Delta N_{0.5}\end{bmatrix}\right),\\ \begin{bmatrix}M_{0.5}\\N_{0.5}\end{bmatrix}-n\begin{bmatrix}\Delta M_{0.5}\\\Delta N_{0.5}\end{bmatrix}, & \left(\begin{bmatrix}M_{0.5}\\N_{0.5}\end{bmatrix}-n\begin{bmatrix}\Delta M_{0.5}\\\Delta N_{0.5}\end{bmatrix}\right)>\begin{bmatrix}M\\N\end{bmatrix},\\ \begin{bmatrix}M\\N\end{bmatrix}, & \left(\begin{bmatrix}M_{0.5}\\N_{0.5}\end{bmatrix}+n\begin{bmatrix}\Delta M_{0.5}\\\Delta N_{0.5}\end{bmatrix}\right)>\begin{bmatrix}M\\N\end{bmatrix}>\left(\begin{bmatrix}M_{0.5}\\N_{0.5}\end{bmatrix}-n\begin{bmatrix}\Delta M_{0.5}\\\Delta N_{0.5}\end{bmatrix}\right),\end{cases} \tag{14}$$

$$(X_0,Y_0)=\left(\left(x'+\frac{T_l}{2}\right),\left(y'+\frac{T_s}{2}\right)\right), \tag{15}$$

where T_l is the corrected value of the major axis, T_s is the corrected value of the minor axis, and (X_0,Y_0) are the center coordinates of the ellipse after error calibration.

As shown in Fig. 5(f), the major axis and minor axis of the ellipse are finally determined within a circumscribed rectangle, which avoids the interference of discrete edge points and simplifies the calculation.

Fig. 6 shows the experimental results with our proposed method. It can be seen that the fitting center point of the target edge is accurately in the center of the actual target, and the target edge with accurate contour is completely circumscribed in the rectangle, which means that the edge fitting is successful and the error from segmentation algorithm is removed.

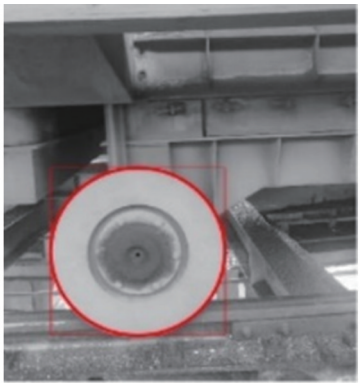
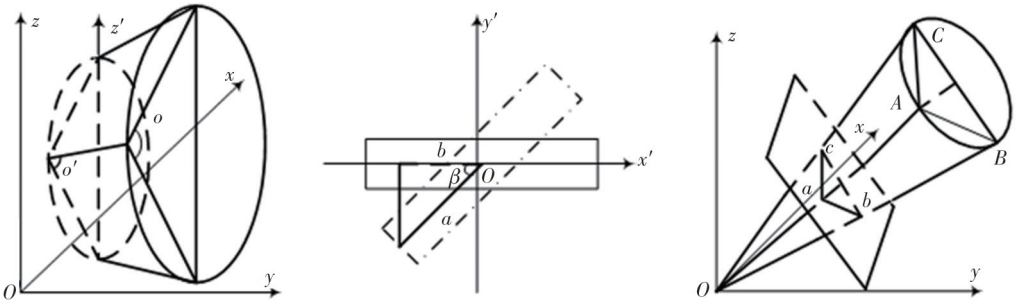


Fig. 6 Experimental results

2 Pose estimation

If the angle between a circular target and a camera exists, its planar imaging is elliptical. For the convenience of calculation, the top view can be simply regarded as the intersection of two straight lines, as shown in Fig.7.

In Fig.7(a), the target is the wheel rotating around z axis. To put it simply, the intersection angles with z axis and x axis are 0 degree and the angle with y axis is the rotation angle of wheel, as shown in Fig.7(b). Pose estimation uses PnP (perspective n points), which is a method for solving 3D to 2D point pair motion. It describes how to estimate the pose of a camera when n 3D space points and their projection positions are known. That is, how to solve the pose of the camera when n 3D space points are given. In this study, we use P3P, which is based on more than 3 non-collinear feature points. Here, the center point O of the ellipse, the intersection point A of the minor axis and the edge, and the intersection points B and C of the major axis and the edge are 3D space points, and a, b and c are 2D projections points, as shown in Fig.7(c).



(a) Wheel projection (b) Top view of wheel projection (c) P3P pose estimation

Fig. 7 Schematic diagram of pose estimation

According to the law of cosines, we can get

$$\begin{cases} OA^2+OB^2-2\cdot OA\cdot OB\cdot\cos\angle aob=AB^2, \\ OA^2+OC^2-2\cdot OA\cdot OC\cdot\cos\angle aoc=AC^2, \\ OB^2+OC^2-2\cdot OA\cdot OC\cdot\cos\angle boc=BC^2. \end{cases} \tag{16}$$

Let $y=\dfrac{OB}{OC}$, and $y=\dfrac{OA}{OC}$, we can get

$$\begin{cases} x^2+y^2-2xy\cos\angle aob=\dfrac{OB^2}{OC^2}, \\ x^2+1-2x\cos\angle aoc=\dfrac{AC^2}{OC^2}, \\ y^2+1-2y\cos\angle boc=\dfrac{BC^2}{OC^2}. \end{cases} \tag{17}$$

Let $u=\dfrac{AB^2}{OC^2}$, $v=\dfrac{BC^2}{AB^2}$, and $w=\dfrac{AC^2}{AB^2}$, we can get

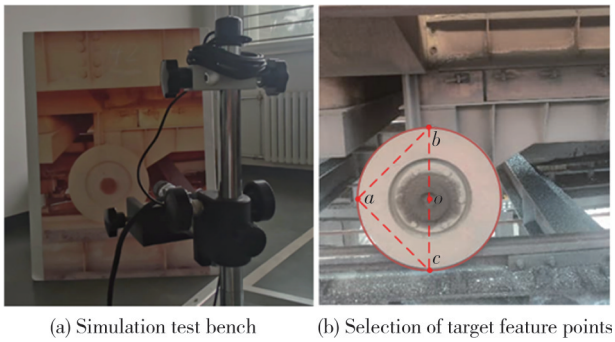
$$\begin{cases} x^2+y^2-2xy\cos\angle aob=u, \\ x^2+1-2x\cos\angle aoc=uw, \\ y^2+1-2y\cos\angle boc=uv. \end{cases} \tag{18}$$

The lengths of AB, AC and BC can be determined according to the coordinates of input points, and then three cosine values can be evaluated according to the plane coordinates. After the unknowns are solved according to Eq.(18), pose estimation can be achieved.

3 Experimental verification

3.1 Experimental platform

The experimental platform was set up for image acquisition and target simulation. Basler acA1300-75gc was adopted as the image acquisition system and the world coordinates of the target were $o(0,0,0)$, $a(0,200,1\,000)$, $b(-200,0,1\,000)$, and $c(0,200,100)$ acquired before simulation, as shown in Fig.8.

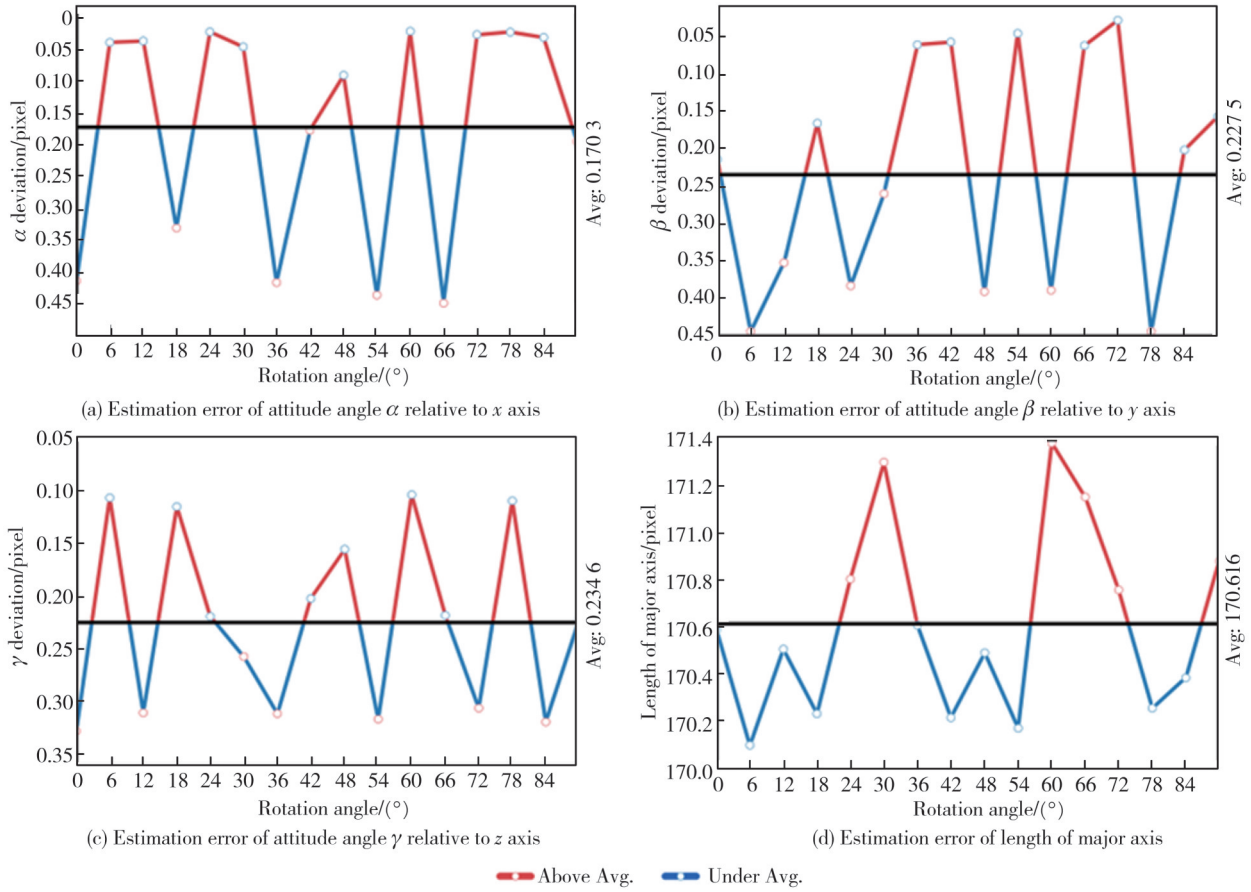


(a) Simulation test bench (b) Selection of target feature points

Fig. 8 Experimental platform

3.2 Experimental results and discussion

To verify the effectiveness of the proposed method, the experiment was conducted. The target simulator rotated 6° each time and 15 times each test. Thus, the relative pose estimation could be obtained between the image acquisition system and the target simulator using P3P pose estimation, and the results are shown in Fig.9.



(a) Estimation error of attitude angle α relative to x axis

(b) Estimation error of attitude angle β relative to y axis

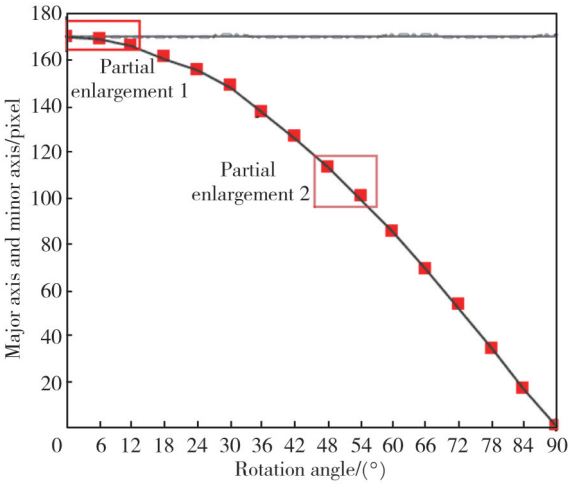
(c) Estimation error of attitude angle γ relative to z axis

(d) Estimation error of length of major axis

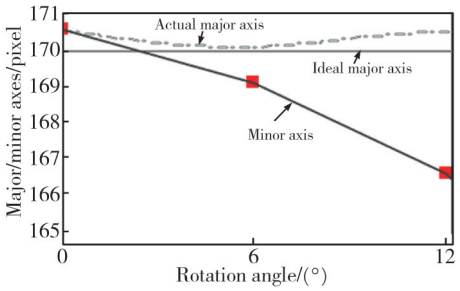
Fig. 9 Estimation errors of attitude angles varying with rotation angles using P3P pose estimation

It can be observed from Fig. 9 that there are certain deviations of the attitude angles by using P3P pose estimation. Among them, the estimation error relative to x axis ranges from 0.024 8° to 0.447 3°, with the average of 0.170 3°; the estimation error relative to y axis ranges from 0.027 1° to 0.443 1°, with the average of 0.227 5°; and the estimation error relative to z axis ranges from 0.103 5° to 0.318 8°, with the average of 0.234 6°. Fig. 9 (d) shows the estimation error of the length of major axis a, which is a fixed value ideally but actually has a change between 170.099 23 pixels and 171.386 24 pixels.

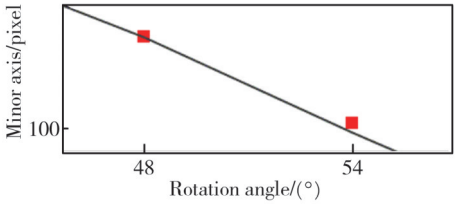
Fig.10 shows the curves of the fitting major and minor axes of the ellipse varying with the rotation angle, and the black curve forms the actually-measured data. In the experiment, the ideal state of the major axis should be a fixed value, but it actually varies from 170.099 231 14 pixels to 171.386 246 95 pixels, meeting the precision requirements of the major axis. In partially enlarged figures, the length of the minor axis should theoretically be the cosine product of major axis's length and rotation angle, but actually has a deviation, which will eventually lead to certain errors in the calculation of rotation angles.



(a) Over change trend



(b) Partial enlargement 1



(c) Partial enlargement 2

Fig. 10 Change trend of major and minor axes

4 Conclusions

In this study, we proposed a new method integrating multiple algorithms for the non-cooperative target extraction in a complex industrial environment, achieving pixel classification first and then target extraction considering global information and local information. The computation was decreased in the edge fitting process and the self-adaption of various paragraphs was realized in the algorithm. The best circumscribed rectangle was used to fit the edge of the elliptical target, which significantly reduces the time complexity and avoids the errors effectively due to edge burrs. The multiple experiments using the P3P pose estimation algorithm show that the proposed method has strong robustness.

Acknowledgement

This work was supported by Key Research and Development Projects in Shaanxi Province (No. 2021GY-265), and Xi'an University Talent Service Enterprise Project (No.2020KJRC0049).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

[1] TIAN G B, WANG J, WANG B W. Monocular camera non-cooperative target extraction and pose detection. Infrared and Laser Engineering, 2021, 50(12): 536-544.

[2] WANG X J, YU C, KAI Z. Methods of monocular pose measurement based on planar objects. Optics and Precision Engineering, 2017, 25(1): 274-280.

[3] LAO D B, ZHANG H J, XIONG Z, et al. Automatic measurement method of attitude based on monocular vision. Acta Photonica Sinica, 2019, 48(3): 315001.

[4] JIN Z M, WANG L, LIU H H, et al. Monocular-based pose estimation of non-cooperative space targets using EKF and EKPF//2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, October 17-19, 2020, Chengdu, China. New York: IEEE, 2020: 46-51.

[5] MIAO R L, LI J L, JIU H F, et al. Research on monocular-vision indoor location technology for olfactory-based searching robots. Journal of Huazhong University of Science and Technology (Nature Science Edition), 2019, 47(4): 19-24.

[6] MEI J P, WANG H, ZHANG D, et al. Dynamic target tracking algorithm of a high-speed parallel robot based on monocular vision. Journal of Tianjin University, 2020, 53(2): 138-146.

[7] LIN T T, JIANG H, LI R H, et al. Non-cooperative target visual pose measurement and ground verification method. Journal of Dalian Jiaotong University, 2020, 41(3): 34-40.

[8] PENG W, ZHOU Q T, SUN C K. Study of pose estimation based on multiple feature points topological determination. Infrared and Laser Engineering, 2017, 46(5): 517001.

[9] XU L Z, FU Q W, TAO W, et al. Monocular vehicle pose estimation based on 3D model. Optics and Precision Engineering, 2021, 29(6): 1346-1355.

[10] ZHANG C Y, ZHAO Q, PHILIP CHEN C L, et al. Deep compression of probabilistic graphical networks. Pattern Recognition, 2019, 96: 106979.

[11] SUN X Y, ZHU X J, WANG P Y, et al. A review of robot control with visual servoing//2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems, July 19-23, 2018, Tianjin, China. New York: IEEE, 2018: 116-121.

[12] WU C R, ZHANG S Y, HE Z X. A rapid detection method of complex background ellipse based on gradient classification. Journal of Zhejiang University Engineering, 2018, 52(5): 943-950.

[13] GUO X G, WANG J, CHENG C. Image segmentation algorithm combining hierarchical clustering algorithm and graph-based segmentation algorithm. Journal of National University of Defense Technology, 2022, 44(3): 194-200.

[14] SHEN X J, ZHANG H, CHEN H P, et al. Fast recursive multi-threshold segmentation algorithm. Journal of Jilin University (Engineering and Technology Edition), 2016, 46(2): 528-534.

[15] LI D J, GUO S, YANG L. Yarn defect detection based on improved image threshold segmentation algorithm. Journal of Textile Research, 2021, 42(3): 82-88.

[16] XIAO M Y, LI X F. Multi-scale 3D Otsu thresholding algorithm based on Gaussian decomposition. Journal of Jilin University (Engineering and Technology Edition), 2017, 47(1): 255-261.

基于图像分割的复杂工业环境非合作目标提取算法

吴晓君*, 王　鹏, 赵　鹤, 余显喆, 李天成

西安建筑科技大学 机电工程学院, 陕西 西安 710055

摘　要： 在复杂工业场景中，使用单目视觉伺服控制很难获得高精度非合作目标姿势。为此，以钢铁生产过程中使用的烧结小车滚轮为研究对象，提出一种实现目标提取和高精度边缘拟合的新方法，它由多种适用于提取目标工作环境的算法融合而成。首先，基于灰度直方图中目标图像像素与非目标图像像素的明显差异，对这些像素进行分类，然后在类中进行分割，以去除干扰因素并保留目标图像。其次，对多个分割结果进行合并，消除小连通区域后得到最终目标图像。在边缘拟合阶段，提出了最优边界矩形的边缘拟合方法，以精确拟合圆形目标边缘。最后，采用PnP算法进行目标位姿检测。由实验结果可知，姿态角γ相对于z轴旋转的平均测量误差为0.234 6°，姿态角α相对于x轴旋转的平均测量误差为0.170 3°，姿态角β相对于y轴旋转的平均测量误差为0.227 5°。实验结果证明，该方法具有实际应用价值。

关键词： 数字图像处理；工业环境；非合作目标；姿势估计

引用格式： WU Xiaojun, WANG Peng, ZHAO He, et al. Non-cooperative target extraction in complex industrial environment based on image segmentation. Journal of Measurement Science and Instrumentation, 2025, 16(1): 119-127. DOI: 10.62756/jmsi.1674-8042.2025012