A Novel Information Fusion Method for Vision Perception and Location of Intelligent Industrial Robots

Shoufeng Jin[1], Qiangqiang Lin[1], Jian Yang[2], Yu Bie[3, *], Mingrui Tian[4], Zhixiong Li[5]

[1]*Department of Mechanical Engineering, Xi'an Polytechnic University,*
No. 95, Jinhua North Road, Xincheng District, Xi'an, Shaanxi 710048, China
[2]*Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham,*
Ningbo China,199 Taikang East Road, Ningbo 315100, China
[3]*School of Chemical Engineering, Kunming University of Science and Technology,*
727 Jingming South Road, Chenggong District, Kunming, Yunnan, 650504, China
[4]*Shaanxi Provincial Key Laboratory of Chang'an University of Highway Construction Machinery,*
No. 95, Jinhua North Road, Xincheng District, Xi'an, Shaanxi 710064, China
[5]*School of Engineering, Ocean University of China,*
No. 238 Songling Road, Tsingdao 266100, China
bieyujuli@163.com

Abstract—**An improved SURF (Speeded-Up Robust Feature）algorithm is proposed to deal with the time-consuming and low precision of positioning of industrial robot. Hessian matrix determinant is used to extract feature points from the target image and a multi-scale spatial pyramid is constructed. The location and scale value of feature points are determined by neighbourhood non-maximum suppression method. The direction of feature points is defined as directional feature descriptors by the binary robust independent elementary feature (BRIEF). The progressive sample consensus (PROSAC) is used to carry out second precise matching and remove mismatching points based on the Hamming distance. Then, an affine transformation model is established to describe the relationship between the template and target images. Centroid coordinates of the target can be obtained based on the affine transformation. Comparative tests were carried out to demonstrate that the proposed method can effectively improve the recognition rate and positioning accuracy of the industrial robots. The average time consuming is less than 0.2 s, the matching accuracy is 96 %, and the positioning error of the robot is less than 1.5 mm. Therefore, the proposed method has practical application importance.**

Index Terms—**Industrial robot; SURF algorithm; BRIEF descriptor; Affine transformation; Centroid coordinates.**

I. INTRODUCTION

Traditional industrial robots function directly based on offline programming or teaching control and complete the designated actions by following the preset instructions, so that

the labor force can be reduced with the increase of productivity [1], [2]. However, traditional industrial robots are not flexible enough to adapt to the working object's changes in shape, size, and position, so that the capacity of recognition rate and flexibility in dealing with different tasks degrade severely under these conditions [3], [4]. Hence, industrial robots must be intelligent to meet the needs of identifying and locating multiple operation targets equipped with multiple sensors.

With the development of visual sensor and computer vision theory, environment perception and target cognition of industrial robots have received much research attention [5], [6]. Amintoosi *et al*. [7] applied a shape feature vector matching algorithm to workpiece images to automatically sort workpieces. Vinividyadharan and Subusurendran [8] used a fuzzy algorithm for workpiece recognition. Geometric parameters of the workpieces were used as the identification feature. Lowe *et al*. [9] proposed a visual feature vector matching method to extract the features of workpieces. The established feature vector was used to achieve the identification of the workpieces. Ojala *et al*. [10] proposed the SIFT feature extraction approach and used a template matching method to identify the workpiece features. Then, the morphological method was adopted to obtain the workpiece feature point in two-dimensional (2D) space to improve the workpiece identification performance. Uijlings *et al*. [11] combined the Hough linear detection, geometric transformation and SIFT to eliminate the target rotation and extract useful features for target recognition. Ren *et al*. [12] used the artificial neural network (ANN) and SIFT to extract the image features and calculate the Euclidean distance between the target image feature and the extracted feature to realize the object recognition. Fezenswalb *et al*. [13] proposed the SURF-BRISK-MSAC algorithm to calculate the

feature similarity to eliminate false feature matching points in workpiece identification. These machine vision technologies make it possible for industrial robots with intelligent vision ability in the complex industrial scene. However, due to industrial scene illumination and complex background disturbance, the size and shape of operational objectives may decrease the vision perception and location performance of intelligent industrial robots. It is still a challenging task to accurately identify and locate the objectives using machine vision.

In this paper, a job target recognition and location method is proposed to improve the flexibility and robustness of robots in varying tasks. In this new method, the SURF descriptor and the BRIEF descriptor with rotation invariance are used to extract the feature points of the job target. Then, neighbours of Hamming are used for the feature matching. Considering that the noise in the feature matching process may cause false matching, the PROSAC algorithm is used to double the matching points of the Hamming distance after the initial matching. The best matching feature points can be obtained to determine the target coordinates, which can be compared with the calibration coordinates to realize the identification and positioning of industrial robots.

II. SYSTEM COMPONENTS

An intelligent industrial robot system, similar to the four degree-of-freedom (DOF) SCARA robot, is constructed (Fig. 1). Planar joints are used to realize the functions of positioning and orientation in this robot. The robot consists of one loading device with maximum 3 Kg load, one servo motor system, one visual system with a AVT GE 1050 industrial camera (a resolution of 1024×1024, frame rate 59 FPS at this resolution, progressive scanning mode and global exposure mode suitable for acquisition of moving targets, 25 mm optical lens, and LED light), a 360 degree rotating table driven by a stepper motor plus a harmonic reducer. The camera is installed in the robot arm end. A signal processing software is installed in a PC to collect the camera images and the open source development board Arduino mega 2560R3 is used to control the robot arm.
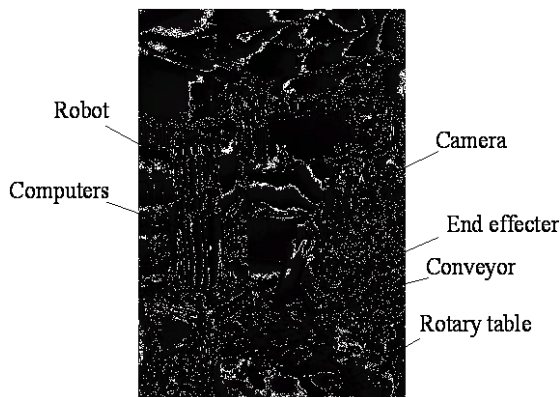

Fig. 1. An intelligent Industrial Robot System.

The identification of job objectives is the prerequisite for the task of intelligent industrial robots to complete the crawling and assembling. The existing recognition methods mainly include template-based matching algorithm and feature-based matching algorithm. Template-based matching

algorithm is susceptible to external environment interference and, thus, is not robust. The feature-based matching algorithm has strong anti-jamming effects on image scaling, rotation, affine transformation, and illumination, which can reduce the influence of noise and disturbances. As a popular feature-based matching algorithm, SIFT [14] is well-recognized in feature matching, but its computational complexity is high. SURF [15] can address this problem. In this paper, a SURF-based method is proposed. Theories of the proposed method are described as follows.

A. Feature Point Extraction

SURF transforms the image and the Gaussian second-order differential template into the addition and subtraction of the integral image. In image I, a point (x, y) is selected. Hessian matrix of the point scale σ is defined as

$$H(x,\sigma) = \begin{vmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{vmatrix}, \qquad (1)$$

where $L_{xx}(x,\sigma)$ and $L_{xy}(x,\sigma)$ are respectively the convolution of the derivative of the Gauss function in the x and y directions, and $L_{yy}(x,\sigma)$ is a two-dimensional convolution of the gray value of the midpoint of the image.

Then, the convolution of template and image is transformed into a Box Filter operation as shown in Fig. 2 and Fig. 3.
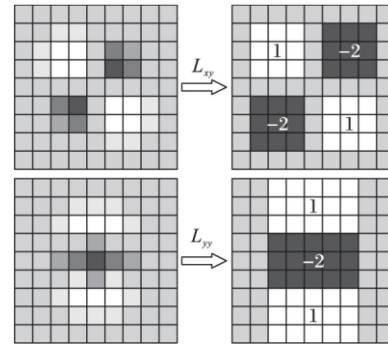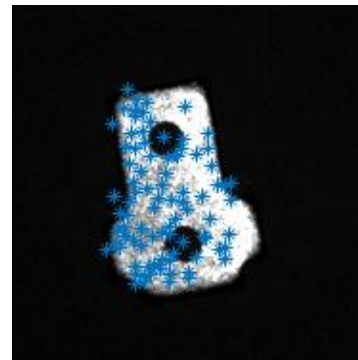

Fig. 2. Box filter.

Fig. 3. Feature points.

By setting 9×9 template as the minimum scale space value of the image filtering, the Hessian matrix in (1) can be simplified as follows

$$Det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2, \qquad (2)$$

where D_{xx}, D_{xy}, and D_{yy} are the results from the convolution

of template and image, and $L_{xy}(x,\sigma)$, $L_{yy}(x,\sigma)$, and $Det(H_{approx})$ are the response values of around point x from the Box Filter.

B. The Establishment of Scale Space

In order to obtain spots of different scales, it is necessary to establish a spatial scale pyramid of the image. As a result of the use of box filtering and integral images, SIFT does not need to build the image pyramid directly, but uses an indirect method instead. The response image of Hessian matrix is calculated using different size of box filter. Then, the 3D image of the response image is used to suppress the non-maximum value of the neighbourhood [16], [17]. The Box filter is constructed as shown in Fig. 2. The characteristic point of the Hessian matrix is compared with the other 26 values of the 3D scale space of the point. When the characteristic point value is greater than or less than the other 26 points, the point is the candidate feature point. Then, by the scale space and image space interpolation calculation, a stable feature point position and scale values can be obtained. By the method, 125 feature points are extracted from the template image as shown in Fig. 3.

C. BRIEF Describes the Feature Point

In order to increase direction information to the extracted feature points, the direction of the offset vector between the gray level of the feature points and the center of mass is defined by the gray scale centroid method as the direction of the feature points [18]. The neighbourhood moments are defined by

$$m_{pq} = \sum_{xy} x^p y^q I(xy),\tag{3}$$

where $I(x, y)$ is the gray value of point (x, y), p and q are the coefficients of the moments.

The centroid is expressed by

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right),\tag{4}$$

where m_{00} is the zero moment, m_{01} and m_{10} are the first moments. θ is the angle between the feature point and the centroid and can be defined as the direction of the feature point

$$\theta = \arctan\left(\frac{m_{01}}{m_{10}}\right) = artcan\left(\frac{\sum_{x,y} yI(x,y)}{\sum_{x,y} xI(x,y)}\right).\tag{5}$$

The BRIEF descriptor is described by the BRIEF descriptor with rotation invariance [19]. The working principle of the BRIEF feature point is to select several pixel pairs in the neighbourhood of the feature points randomly. In order to solve the problem of noise interference, two 5×5 sub-windows are randomly selected in the 31×31 pixel neighbourhood with the feature point as the center [20]. The binary is obtained by using the criterion τ assignment

$$\tau(p;x,y) = \begin{cases} 1, p(x) < p(y), \\ 0, p(x) \geq p(y), \end{cases}\tag{6}$$

where $p(x)$ and $p(y)$ are the gray functions of neighbourhood at points x and y.

N (x, y) pixel positions are randomly selected in the neighbourhood, the binary assignment shown in (6) is repeated to obtain a descriptor of the N-dimensional binary string of BRIEF as shown in (7)

$$f_n(p) = \sum_{1 \leq i \leq m} 2^{i-1} \tau(p; x_i, y_i).\tag{7}$$

The centroid direction information of the feature point obtained by equation (3) is added to the descriptor. Each feature point in Fig. 3 is analyzed and encoded with a 256-bit binary encoding as shown in Fig.4.
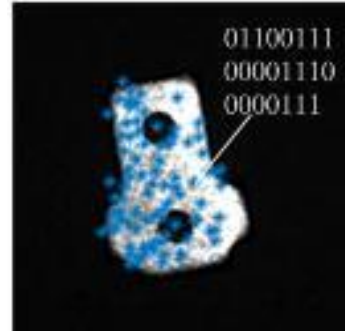

Fig. 4. Bit binary encoding for feature points.

D. Feature Point Matching

Using the feature points extracted by SURF and BRIEF, the Hamming distance between the two key points is calculated, and the nearest neighbour distance is used to measure the similarity [21]. The descriptors of the two feature points are derived from the feature descriptors obtained by

$$K_1 = x_0 x_1 \cdots x_{225}, \quad K_2 = y_0 y_1 \cdots y_{225},\tag{8}$$

and the Hamming distance between the two feature descriptors can be expressed as

$$D(K_1, K_2) = \sum_{i=0}^{i=255} x_i \oplus y_i,\tag{9}$$

where x and y are binary values and $\oplus$ denotes XOR logic operation.

A large value of $D(K_1, K_2)$ will lead to a large number of corresponding bits between the two feature descriptors and a small similarity between the two feature descriptors.

Wheel bracket is chosen as the template in Fig. 5. The captured image at the job site is shown in Fig. 6. The images shown in Figs. 5–6 are matched by the above-mentioned algorithm. The wheel bracket is completely identified as shown in Fig. 7. In details, similarity between feature points and matching points can be estimated by Hamming distance. Matching time consuming is set at 100 ms in order to meet the

real-time requirement of the system. With the total pair number 82 and the primary pair number 73, the matching accuracy is expressed as the ratio of the primary matching logarithm and the total matching logarithm and is equal to 89%. Obviously, some mismatching points exist in the matching of template image and operation target image, which come from the interference of noise and disturbance components. Errors of false matching may be the result of neighbouring Hamming distance, which affects the final matching accuracy.

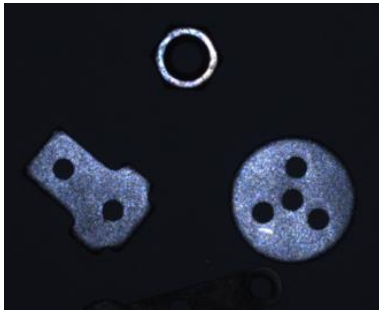Fig. 5. A template of wheel bracket.
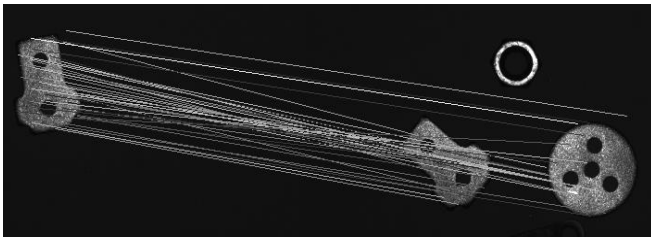

Fig. 6. Job Target Chart.

Fig. 7. Matching result.

E. PROSAC Algorithm for Secondary Fine Matching

Aiming at the mismatching points that affect the matching accuracy shown in Fig. 7, the Hamming distance is used as the similarity measure of the feature points and matching points. Due to noise and external interference, the nearest neighbour Hamming distance may result in false matching pairs. These points are called external points. The existence of external points seriously affects the matching accuracy of the algorithm. Therefore, to perform secondary matching of Hamming distance from the initial matching feature points, the PROSAC is used. In the PROSAC processing, RANSAC first extracts the samples from all the data [22]. Then, PROSAC divides the samples by descending the quality of the samples and extracts the samples from the data subsets with higher quality to find the optimal estimate solution, thus eliminating the false match. The result of the secondary matching is shown in Fig. 8. After the second precise

matching, the feature points are further optimized. The matching time is reduced to 180 ms and the matching accuracy is 96 %.

III. JOB TARGET LOCATION ALGORITHM

Job target positioning is based on the identification of the target to determine the exact location of the coordinates. The general use of the target coordinates of the centroid is to characterize its position information. According to the intelligent industrial robot system (Fig. 1), there is a translational, rotational, and proportional change between the template image and the target image. Therefore, the affine transformation parameters are calculated according to the affine model. According to the centroid and affine transformation parameters of the target template, the centroid position coordinates of the image target in a complex environment can be obtained.
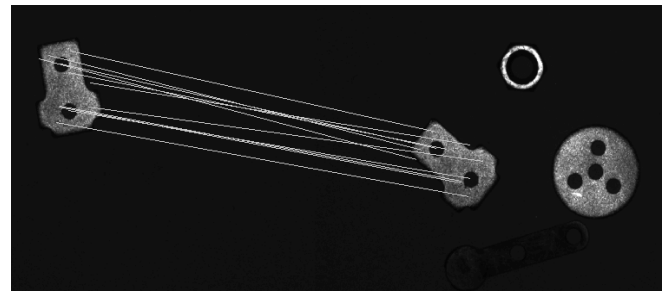

Fig. 8. Secondary matching result.

### A. The Centroid Coordinates of the Job Target Template

The target template image is preprocessed as shown in Fig. 9. In this process, the edge contour information of the target is extracted by Canny operator. The background is filled by morphological algorithm. Thus, the complete shape of the target, as shown in Fig. 9(c), is obtained by removing the isolated small targets and burrs outside the target area. To realize determining the exact target location, an 8-neighbourhood labelling method is used to label the target image and two-dimensional centroid coordinates of the labelled region are calculated as

$$x = \sum_{i,j\in\Omega} \frac{i}{n}, \; y = \sum_{i,j\in\Omega} \frac{j}{n} \qquad (10)$$

where i and j are the horizontal and vertical coordinates of the target image pixels, n is the total number of pixels, and Ω is the set of pixels belonging to the same target image. The centroid coordinates of the template image obtained from the centroid calculation are shown in Fig. 9(d).
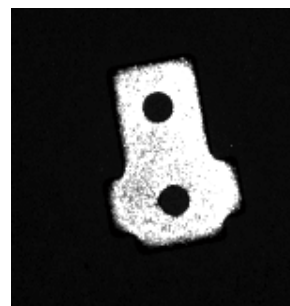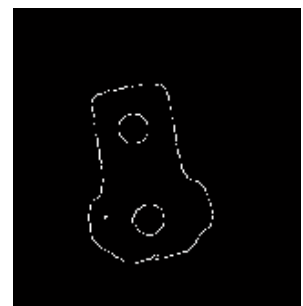
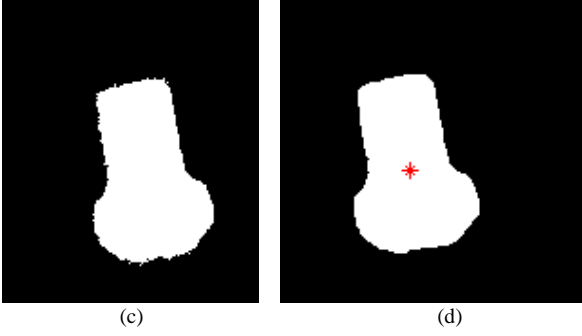(a)   (b)

Fig. 9. Template positioning: (a) image enhancement; (b) edge detection; (c) background fill; (d) target setting.

B. Affine Model Parameter Calculation

Considering the translation and rotation of template and target images, the 4-parameter affine model is used to calculate the affine transformation parameter. The transformation matrix is

$$A = \begin{vmatrix} s\cos(\theta) & -s\sin(\theta) & t_x \\ s\sin(\theta) & s\cos(\theta) & t_y \\ 0 & 0 & 1 \end{vmatrix}, \quad (11)$$

where s is the scale change parameter, t_x is the amount of translation in the x direction, t_y is the amount of translation in the y direction, and θ is the rotation angle.

The mapping between the point (x_p, y_p) on the template image and the point (x_q, y_q) on the target image is:

$$\begin{vmatrix} x_q \\ y_q \end{vmatrix} = \begin{vmatrix} s\cos(\theta) & -s\sin(\theta) \\ s\sin(\theta) & s\cos(\theta) \end{vmatrix} \begin{vmatrix} x_p \\ y_p \end{vmatrix} + \begin{vmatrix} t_x \\ t_y \end{vmatrix}. \quad (12)$$

By setting $m_{11} = -m_{22} = s\cos\theta$, $m_{13} = t_x$, $m_{12} = -m_{21} = s\sin\theta$, $m_{23} = t_y$, it follows that:

$$\begin{cases} x_q = m_{11}x_p + m_{12}y_p + m_{13}, \\ y_q = m_{21}x_p + m_{22}x_p + m_{23}. \end{cases} \quad (13)$$

Let us find the numbers of 6 positions $[m_{11}, m_{12}, m_{13}, m_{21}, m_{22}, m_{23}]$, at least three pairs of the matching points. Suppose (p_i, p_j) and (q_a, q_b) are the two pairs of matching feature points corresponding to the feature set P and Q extracted by the template and target images, respectively, and can be obtained by least squares (LSM) [23].

Suppose that a matching point pair is taken in a frame image $(x'_1, y'_1), (x'_2, y'_2), \cdots, (x'_n, y'_n)$, it can lead to the corresponding matching point pair in the corresponding matching graph $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. The relationship between them is:

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{vmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \end{bmatrix} = \begin{vmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{vmatrix}, \quad \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{vmatrix} \begin{bmatrix} m_{21} \\ m_{22} \\ m_{23} \end{bmatrix} = \begin{vmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{vmatrix}. \quad (14)$$

The mean square error RMSE is defined as

$$I_{RMSE} = \sum_{i=1}^{n} \left[\frac{(m_{11}x_i + m_{12}y_i + m_{13} - x'_i)^2 +}{+(m_{21}x_i + m_{22}y_i + m_{23} - y'_i)^2} \right] / n. \quad (15)$$

C. Calculate the Centroid of the Target Image

According to the centroid coordinate position of the obtained template image, the position of the centroid coordinate of the target in the target image is calculated by the affine transformation model parameter into (13) as shown in Fig. 10.
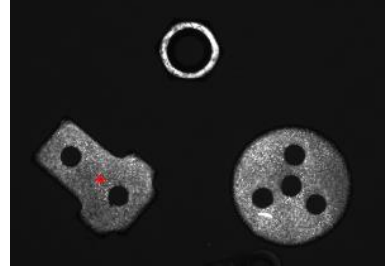

Fig. 10. The centroid coordinates of the job target.

IV. EXPERIMENT ANALYSIS

The intelligent industrial robot system shown in Fig. 1 is used to carry out the experimental analysis of the proposed recognition localization algorithm. The template image and the job target image are extracted, matched, and identified, and the acquired target coordinates are combined with the calibration parameters of the visual system. The results demonstrate the assembly ability of the robot.

In the test, a metal structure with a thickness of 2.0 mm is a target part to be grabbed by a sucker-type grabbing mechanism. The two-dimensional position information of the target is calculated in the process of locating the target.

A. Identify Matching Experiments

Fig. 11 shows the logic process in identifying and positioning targets. Feature points are detected, firstly, by SURF algorithm and their main directions are extracted using gray centroid method. Thus, directional BRIEF descriptors are constructed to describe the feature points. By combing Hamming distance and PROSAC algorithm, mismatching points are eliminated in the second precise matching procedure.
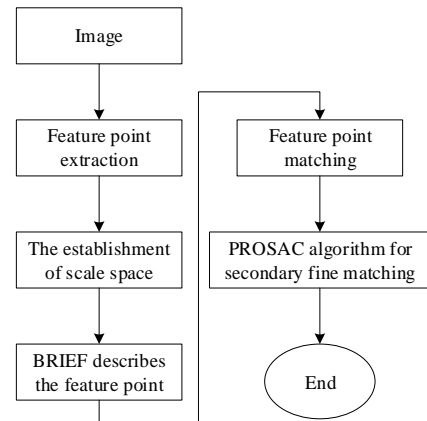

Fig. 11. Flow chart of identifying and positioning targets.

In the experiment, the classical SIFT, SURF, and the proposed algorithm are compared in the same scenario. The experimental results are shown in Fig. 12. In addition, the computational time of the proposed algorithm is less than 0.2 s, which is much smaller than that of SURF and SIFT. High computing efficiency of the proposed method makes it competent for real-time industrial production.
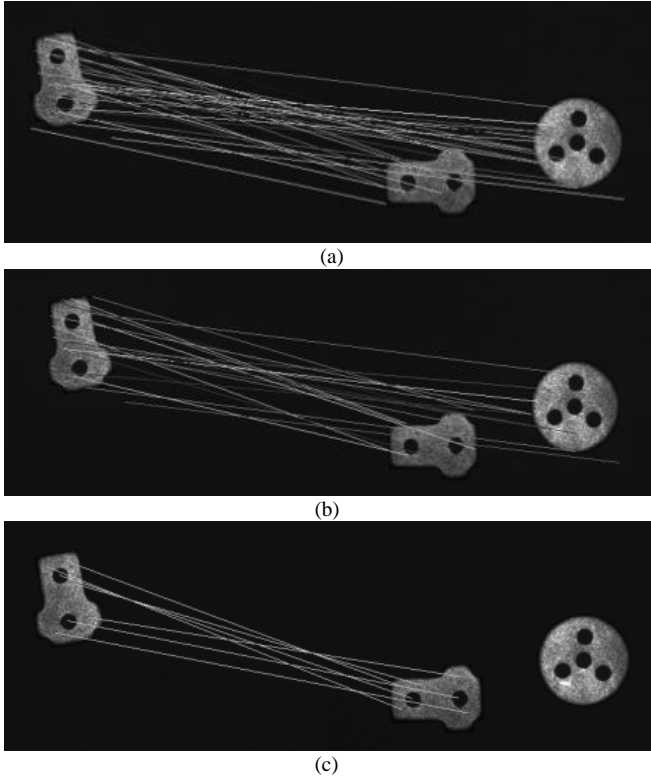

(a)

(b)

(c)

Fig. 12. Comparison results: (a) SIFT; (b) SURF; (c) the proposed algorithm.

TABLE I. TIME CONSUMPTION (MS).

	SIFT	SURF	The proposed algorithm
1	248.32	223.45	180.34
2	267.51	209.18	185.60
3	267.45	230.91	174.68
4	265.34	224.78	198.15
5	278.91	245.76	187.14
6	274.23	246.46	184.89
7	289.01	252.14	183.76
Mean	270.01	233.24	184.94

In addition, the recognition results are shown in Fig. 13 in case of occlusion or overturn for target. It can be seen from the graph that this algorithm exhibits many characteristics, such as scale and rotation invariance under various disturbances, and satisfies the requirements of low time-consuming.
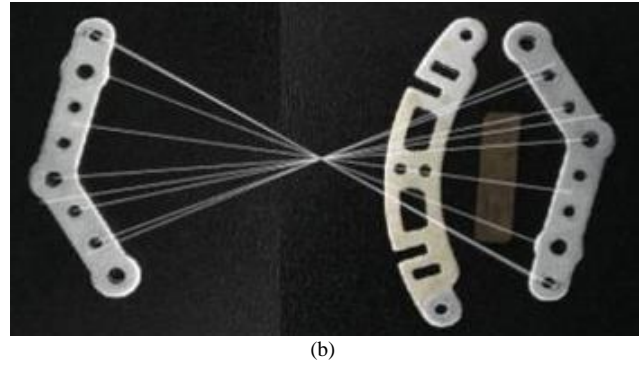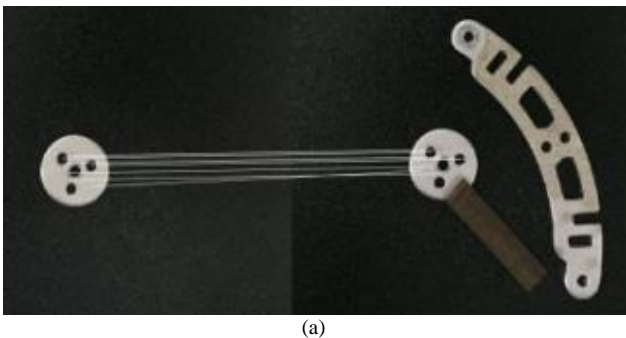

(a)

(b)

Fig. 13. Recognition results: (a) Target under occlusion; (b) Target flipping.

B. *Positioning of Job Objectives*

Complex workpieces with regular shape and irregular shape are identified and positioned to verify the proposed algorithm. Target is included by green box in Fig. 14, which also shows the positioning experimental results. Combined with the calibration parameters of the visual system, the operating centroid of the job objective can be obtained (Table II). The positioning error in the table is less than 1.5 mm and the largest error is 4.2 %. As a result, the positioning accuracy of the proposed method is satisfactory.
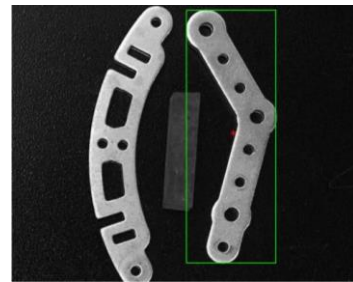

Fig. 14. Target image.

Fig. 15. Positioning results.

TABLE II. JOB TARGET COORDINATES OF THE CENTROID.

	Image coordinates (pixels)		Calculated (mm)		Actual value (mm)	
	x	y	x	y	x	y
1	81	179	16.89	37.32	17	37
2	102	103	21.27	21.48	21	21
3	109	357	22.73	74.43	23	74
4	57	244	11.88	50.87	12	51

V. CONCLUSIONS

An intelligent industrial robot system with visual perception was constructed with four degrees of freedom SCARA robot and AVT GE 1050 industrial camera in this paper. A new matching and positioning method based on the combination of SURF, BRIEF, Hamming distance, and affine transform combination was proposed to eliminate/reduce the

influence of industrial complex environment on the recognition and positioning of the target image. The experimental validation results demonstrated that (1) the metal structural part with the thickness of 2.0 mm was chosen as a target and its image is obtained exactly under rotation and occlusion, which verifies the effectiveness of the presented algorithm, (2) the feature points can be extracted by SURF and BRIEF, and the Hamming distance of pseudo feature points matching algorithm using PROSAC can improve the matching speed and accuracy of the feature points, and, (3) through the transformation between the template image and the target image, a 4-parameter affine transformation model was established to obtain less than 1.5 mm of the positioning error. The results showed that the proposed method is able to provide the position information of the target for robot grasping and assembly in practical applications. In the future works, the proposed algorithm will be improved by increasing feature points, expanding application scope, and reducing the complexity to obtain better accuracy of the workpiece recognition.

REFERENCES

[1] A. Glowacz, W. Glowacz, Z. Glowacz et al., "Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals", *Measurement*, vol. 113, pp. 1–9, 2018. DOI: 10.1016/j.measurement.2017.08.036.

[2] M. Gupta, S. Kumar, L. Behera et al., "A novel fusion framework for robust human tracking by a service robot", *Robotics & Autonomous Systems,* vol. 94. pp. 134–147, 2017. DOI: 10.1016/j.robot.2017.05.001.

[3] A. Glowacz, "Fault diagnosis of single-phase induction motor based on acoustic signals", *Mech. Sys. Sig. Process*, vol. 117, pp. 65–80, 2019. DOI: 10.1016/j.ymssp.2018.07.044.

[4] N. Ye, Z. Wang, R. Malekian, R. Wang, T. Zhao, D. Andriukaitis, A. Valinevicius, D. Navikas, V. Markevicius, "A hidden Markov model combined with RFID-based sensors for accurate vehicle route prediction", *International Journal of Ad Hoc and Ubiquitous Computing,* vol. 23, no. 1/2, pp. 124–133, 2016. DOI: 10.1504/IJAHUC.2016.078473.

[5] L. Bodenhagen, A. R. Fugl, A. Jordt et al., "An adaptable robot vision system performing manipulation actions with flexible objects", *IEEE T. Autom. Sci. Eng.*, vol. 11, no. 3, pp. 749–765, 2014. DOI: 10.1109/TASE.2014.2320157.

[6] T. Stipancic, B. Jerbic, and P. Curkovic, "A context-aware approach in realization of socially intelligent industrial robots", *Robot. Cim-Int. Manuf.*, vol. 37, pp. 79–89, 2016. DOI:10.1016/j.rcim.2015.07.002.

[7] A. Laucka, D. Andriukaitis, "Research of the Defects in Anesthetic Masks", Radioengineering, vol. 24, no. 4, pp. 1033–1043, 2015. DOI: 10.13164/re.2015.1033.

[8] M. Amintoosi, M. Fathy, and N. Mozayani, "A fast image registration approach based on SIFT key-points applied to super-resolution", *Imaging Sci. J.*, vol. 60 , no. 4, pp. 185–201, 2012. DOI: 10.1179/1743131X11Y.0000000015.

[9] Vinividyadharan and Subusurendran, "Automatic Image registration Using SIFT-NCC", *I. J. Comp. Appl.*, vol. ACCTHPCA, no. 4, pp. 29–32, 2012.

[10] G. Lowe, "Distinctive image features from scale-invariant key points", *Int. J. Comput. Vision.*, vol. 60, no. 2, pp. 91–110, 2004. DOI: 10.1023/B:VISI.0000029664.99615.94.

[11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture class if cation with local binary patterns", *IEEE T. Pattern Anal.*, vol. 24, no. 7, pp. 971–987, 2002. DOI: 10.1109/tpami.2002.1017623.

[12] R. Uijlings, A. Van De Sande, T. Gevers et al., "Selective search for object recognition", *Int. J. Comput. Vision*, vol. 104, no. 2, pp. 154–171, 2013. DOI: 10.1007/s11263-013-0620-5.

[13] S. Ren, K. He, B. Girshick et al., "Faster R-CNN: towards real-time objection detection with region proposal networks", *IEEE T. Pattern Anal.*, vol. 39, no. 6, pp. 1137–1149, 2017. DOI: 10.1109/TPAMI.2016.2577031.

[14] Z. Hossein-Nejad and M. Nasri, "An adaptive image registration method based on SIFT features and RANSAC transform", *Comp. Elec. Eng.*, vol. 62. pp. 524–537, 2016. DOI: 10.1109/ICCSP.2016.7754318.

[15] F. Fezenszwalb, B. Girshick, D. Mcallester et al., "Objection detection with discriminatively trained part based models", *IEEE T. Pattern Anal.*, vol. 32, no. 9, pp. 1627–1645, 2010. DOI: 10.1109/mc.2014.42 .

[16] K. Simonyan, A. Vedaldi, and A. Zissermani, "Learning local feature descriptors using convex optimization", *IEEE T. Pattern Anal.*, vol. 36, no. 8, pp. 1573–1585, 2014. DOI: 10.1109/TPAMI.2014.2301163.

[17] K. Brown, H. Gang, S. Wind, "Discriminative learning of local image descriptors", *IEEE T. Pattern Anal.*, vol. 33, no. 1, pp. 43–57, 2011. DOI: 0.1109/TPAMI.2010.54.

[18] C. Strecha, A. Bronstein, M. Bronstei et al., "LDAH: improved matching with smaller descriptors", *IEEE T. Pattern Anal.*, vol. 43, no. 1, pp. 66–78, 2012. DOI: 10.1109/tpami.2011.103 .

[19] Y. Liu, H. Liu, F. L. Ni et al., "New self-calibration approach to space robots based on hand-eye vision", *J. Cent. South Univ.*, vol. 18, no. 4, pp. 1087–1096, 2011. DOI: 10.1007/s11771-011-0808-1.

[20] S. Liu, G. Wang, H. Zhang et al., "Design of robot welding seam tracking system with structured light vision", *Chin. J. Mech. Eng-En.,* vol. 23, no. 4, pp. 436–442, 2010. DOI: 10.3901/cjme.2010.04.436.

[21] W. Wang, G. Wu, Y. Bai et al., "Hand-eye-vision based control for an inspection robot's autonomous line grasping", *J. Cent. South Univ.*, vol. 21, no. 6, pp. 2216–2227, 2014. DOI: 10.1007/s11771-014-2173-3.

[22] M. An, Z. Jiang, and D. Zhao, "High speed robust image registration and localization using optimized algorithm and its performances evaluation", *J. Syst. Eng. Electron.*, vol. 21, no. 3, pp. 520–526, 2010. DOI: 10.3969/j.issn.1004-4132.2010.03.026.

[23] C. Shi, J. Yang, L. Liu et al., "A massively parallel keypoint detection and description(MP-KDD) algorithm for high-speed vision chip", *Sci. China*, vol. 57, no. 10, pp. 1–12, 2014. DOI: 10.1007/s11432-014-5174-9.