Binocular stereovision omnidirectional motion handling robot

Bo Tang1 and Li Jiang2

Abstract
Binocular stereovision has become one of the development trends of machine vision and has been widely used in robot recognition and positioning. However, the current research on omnidirectional motion handling robots at home and abroad is too limited, and many problems cannot be solved well, such as single operating systems, complex algorithms, and low recognition rates. To make a high-efficiency handling robot with high recognition rate, this article studies the problem of robot image feature extraction and matching and proposes an improved speeded up robust features (SURF) algorithm that combines the advantages of both SURF and Binary Robust Independent Elementary Features. The algorithm greatly simplifies the complexity of the algorithm. Experiments show that the improved algorithm greatly improves the speed of matching and ensures the real-time and robustness of the algorithm. In this article, the problem of positioning the target workpiece of the robot is studied. The three-dimensional (3-D) reconstruction of the target workpiece position is performed to obtain the 3-D coordinates of the target workpiece position, thereby completing the positioning work. This article designs a software framework for real-time 3-D object reconstruction. A Bayesian-based matching algorithm combined with Delaunay triangulation is used to obtain the relationship between supported and nonsupported points, and 3-D reconstruction of target objects from sparse to dense matches is achieved.

Keywords
Binocular stereovision, omnidirectional motion, handling robot, SURF feature, 3-D reconstruction

Date received: 30 January 2020; accepted: 25 April 2020

Introduction
In the process of human cognition of the world, we perceive external information mainly through observation. Eighty percent of the information is obtained through vision. With the rapid development of industrial technology, traditional industrial production methods can no longer meet the requirements of the times. It has great application value in industrial production. Machine vision works by simulating human vision functions.1 Industrial charge-coupled device cameras are used to capture images of target scenes or objects. Different image processing methods are used in the corresponding optical imaging system to obtain the various results we want.

Many research teams at home and abroad have conducted in-depth research on design methods and control algorithms of handling robots. Clavien et al.2 introduced the design of an independent contact region-based motion controller for AZIMUT-3, using new parameters of the

1 School of Mechatronic Engineering, Xi’an Technological University, Xi’an, China
2 School of Health Services Management, Xi’an Medical University, Xi’an, China

Corresponding author:
Bo Tang, School of Mechatronic Engineering, Xi’an Technological University, Xi’an 710021, Shaanxi, China.
Email: tangbo@xatu.edu.cn
motion state space and using the time scaling principle originally developed for the manipulator trajectory, designed the motion controller. Yorozu and Takahashi proposed a real-time omnidirectional automatic obstacle avoidance method based on sensors. This method is based on the synchronous control of the translation of the movable gap and the coverage area and the effective rotational movement, thereby achieving an autonomous omnidirectional mobile robot. Chang et al. proposed a model and output feedback tracking control method for omnidirectional rehabilitation training walker. The results prove that choosing the proper design parameters will cause the root mean square of the tracking error to be arbitrarily small.

Wang et al. investigated a new omnidirectional laser speckle rheology (LSR) catheter for mapping the mechanical properties of luminal organs without the need for rotational motion. Evaluations performed by the authors in test phantoms with circumferentially varying mechanical properties demonstrate the ability of LSR catheters to perform accurate mechanical assessment of luminal organs. Hoang and Jo proposed a method based on an omnidirectional camera and a laser rangefinder to estimate the pseudo three-dimensional (3-D) motion of a vehicle using a geometrically constrained minimum parameter set. Yang et al. proposed a new type of omnidirectional visual inspection system for detecting morphological defects. Experimental results show that the system can perform high-quality measurement and reconstruction on the 3-D space of the pipeline and has achieved good detection results in terms of speed and accuracy. Zhang et al. propose a space-saving steering system for four-wheel steering and four-wheel drive electric vehicles with higher maneuverability and flexibility. In this article, we use model-based methods to optimize task-specific readability models to generate clear robot behavior. Busch et al. developed and controlled a new low-cost omnidirectional interactive mobile robot platform with force feedback function that can assist and guide the patient’s hand during the treatment process. The literature studied the problem of stably controlling motion and tracking the trajectory of a mobile robot with three rollers and solved the stability and tracking of various programmed motions and trajectories of the robot. Andreyev and Peregudova proposed a possible method to build a robot capable of participating in competitions, which can be used as a reference for current and new teams. The literature model, analyze, and construct soft robots with locomotive capabilities controlled by body deformation. Finally, a prototype of the soft robot is established, and experiments and comparisons are performed to prove the validity of the theory.

Many research teams at home and abroad have conducted in-depth research on the theory of binocular stereovision and its application in different fields. The literature proposed a forward design optimization method of binocular stereovision system based on high-dimensional function modeling and optimization, and the feasibility of the method was verified by experiments. The literature propose a self-calibrating single-lens 3-D video extensometer for noncontact, nondestructive, and high-precision strain measurement to prove the effectiveness of the proposed self-calibrating single-lens 3-D video extensometer. Wang et al. designed calibration targets suitable for these two cameras. It can be seen from the “binocular” stereovision system reconstruction experiment results that the calibration data provided by this method can meet the system requirements. Similarity measurement is an essential step; however, there is always a trade-off between accuracy and computational efficiency. The literature used a high-speed binocular stereovision system to establish a combustion field diagnostic method and demonstrated the feasibility of the method. Yue-fei et al. proposed a method based on stereovision, which can accurately measure the twist of helicopter rotor blades and simplify the operation process. In Xiong et al., the test results show that the reconstruction by binary speckle has better performance. Tang et al. achieved a selective baseline for stereovision by selecting two images from a one-dimensional light field image to view a close-range object using a single master lens. Koyama et al. discuss the occlusion problem and provide a solution for self-occlusion and occlusion between objects. The results show that the algorithm has superior performance and can recognize 80% or less occluded objects. The literature proposed a binocular image stereo-matching algorithm for dynamic image sequences. The results show that the proposed algorithm can effectively suppress the noise generated by stereo matching of dynamic binocular vision images. Hui et al. proposed a pedestrian detection algorithm based on binocular stereovision. Experimental results show that this method can effectively use binocular visual information to avoid monocular visual false alarms, improve detection accuracy, and increase detection speed. Jin et al. established a kinematic model for a system with three segments connected by nonstandard connection types and their trajectories. The comparison of simulation results shows that the improved A* algorithm optimizes the path quality better than A* algorithm in complex environments. Xu and Liu analyzed the availability of all visual cues with the help of semantic information and proposed a new visual odometer method. Experimental results confirm that the new method provides effective improvements in accuracy and robustness in complex dynamic scenes.

Although the application of robots is becoming more and more widespread, they are used in various fields of industrial production. However, the working environment of most industrial robots is determined in advance, taught through a preprogrammed program, and then run according to a fixed trajectory. In general, the pose of the target workpiece is uncertain, such as occlusion and random placement. In addition, the actual pose and ideal pose of the target workpiece usually have certain errors. The existence of these errors may, as a result, the robot cannot
work as expected, causing the robot to make a mistake. In general, the target workpiece conveying system in the omnidirectional motion conveying system does not have a fixed beat. There may be multiple target workpieces and the order is also changed. In this way, the fixed teaching industrial robot cannot meet the production requirements. Therefore, the traditional robot preteaching mode has limited the flexibility and flexibility of omnidirectional motion handling to a large extent, and binocular stereovision technology can just solve this problem.

Based on binocular stereovision, this article uses improved image feature extraction and matching algorithm of speeded up robust features (SURF) algorithm and combines point-based 3-D reconstruction method to achieve 3-D reconstruction of target workpiece position, obtain 3-D coordinates of target workpiece position, and complete the robot. This article studies the problem of robot image feature extraction and matching and proposes an improved SURF algorithm that combines the advantages of both SURF and Binary Robust Independent Elementary Features (BRIEF). Experiments show that the improved algorithm greatly improves the matching speed, ensures the real-time performance of the algorithm, and retains the robustness of the traditional SURF algorithm. This article studies the positioning of the target workpiece of the robot and combines the internal and external parameters obtained from the camera’s stereo calibration and the 3-D reconstruction method based on points to 3-D reconstruct the target workpiece position to obtain the 3-D coordinates of the target workpiece position to complete the positioning work. This article designs a software framework for real-time 3-D object reconstruction. A Bayesian-based matching algorithm combined with Delaunay triangulation is used to obtain the relationship between supported and non-supported points, and 3-D reconstruction of target objects from sparse to dense matches is achieved.

**Method**

**Robot feature extraction model based on improved SURF algorithm**

In this article, the improved SURF algorithm uses the binary descriptor BRIEF operator at the feature description stage. The core theory of the operator is to randomly select a number of points around the feature points to make up the image block and then binarize its gray value. Combine the obtained result into a binary string and use it as the feature descriptor of the feature point. Any bit of the BRIEF descriptor is obtained by randomly selecting two binary points for comparison. Define a binary comparison criterion $\tau$ of an image neighborhood $p$, as shown in the following formula, whose size is $S \times S$

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

Among them, $x$ and $y$ are any two pixels in the image neighborhood $P$, and $p(x)$ and $p(y)$ are the gray values at $x$ and $y$, respectively. There are many different methods for selecting nd $(x, y)$ pixel pairs from $S \times S$-sized image blocks. Calonder experiments use different sampling methods to study them and test the following sampling methods

1. The sampling points $x_i, y_i$ are uniformly distributed $\left[-\frac{S}{2}, \frac{S}{2}\right]$ and are uniformly distributed throughout the sampling area.
2. Both the sampling points $x_i$ and $y_i$ have a Gaussian distribution $\left[0, \frac{1}{\pi S^2}\right]$.
3. $x_i$ obeys the Gaussian distribution of $\left[0, \frac{1}{\pi S^2}\right]$, and $y_i$ obeys the Gaussian distribution $\left[x_i, \frac{1}{\pi S^2}\right]$. First Gaussian sampling is performed for $x_i$ at the origin and then $y_i$ is Gaussian sampling at the center for $x_i$.

Final experiments proved that the second method yielded the best results, so the BRIEF descriptor used the second method. The descriptors generated by the above method have no direction, so to obtain the feature points with rotation invariance, this article uses the gray centroid method to obtain the feature point main direction $\theta$ for the feature points extracted by the traditional SURF algorithm. For any feature point $p$, the moment of its neighborhood pixels can be expressed by the following formula

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y)$$ (2)

**Omnidirectional azimuth reconstruction algorithm based on target positioning**

1. Gaussian mixture model

Generally, different sampling parts in the image correspond to different Gaussian models. The single Gaussian distribution density function is shown in the following formula

$$g(x, u, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left[ -\frac{1}{2} (x - u)^T \Sigma^{-1} (x - u) \right]$$ (3)

Among them, $u$ is the center point of the pixel, and $\Sigma$ describes the covariance matrix. The method of establishing Gaussian mixture model (GMM) is as follows:

1. Initialize GMM parameters: $\Theta = \{w_1, w_2, \ldots, w_M, u_1, u_2, \ldots, u_M, \Sigma_1, \Sigma_2, \ldots, \Sigma_M\}$. Human experience gives the initial weight of each GMM, as shown in $w^0_1 = w^0_2 = \cdots = w^0_M = \frac{1}{M}$. Then use the K-means algorithm to find the centers of modified GMM density functions, as shown in $u^0_1 = u^0_2 = \cdots = u^0_M$. 

$$Tang and Jiang$$
2. Iteratively approximate the final parameters by using expectation maximization algorithm.

3. GMM is obtained from $M$ single Gaussian functions by weighting and averaging. The result is shown in:

$$p(x) = w_1 g(x; u_1, \Sigma_1) + w_2 g(x; u_2, \Sigma_2) + \cdots + w_M g(x; u_M, \Sigma_M).$$

2. GrabCut algorithm description

The GrabCut algorithm uses two incomplete labeling and iterative graph cutting strategies to reduce the amount of subsequent user interaction operations. For image $y = (y_1, y_2, \cdots, y_N)$, the opacity $a_\alpha = \arg \min_a E(\alpha, k, \theta, y)$. Gibbs energy functions are shown by $H E(\alpha, k, \theta, y) = U(\alpha, k, \theta, y) + V(\alpha, y)$ and $U(\alpha, k, \theta, y) = \sum_n D(\alpha, k, \theta, y)$. The data items are defined as shown in the following formula:

$$D(\alpha_n, k_n, \theta_n, y_n) = -\log p(y_n | \alpha_n, k_n, \theta_n, y_n) - \log \pi(\alpha_n, k_n)$$

(4)

where $\pi(\alpha, k)$ is the weight of each Gaussian model.

The main implementation process of the GrabCut algorithm is as follows: (1) First determine a rectangular frame and then use the Trimap algorithm to mark the entire image. The pixels outside the rectangular frame constitute the background area $T_B$ of the image and remember the unknown area $T_y = \overline{T_B}$ of the image and the foreground area of the image is $T_F = \phi$; (2) the opacity of the image is then set according to related theories, where the opacity $a_p = 0$ of the pixels in the $T_B$ area is recorded and the opacity $a_p = 1$ of the pixels in the $T_U$ area is set; (3) $T_B$ the pixels in the region are used to establish the initial background GMM color data model, and the pixels in the $T_U$ region are used to establish the initial foreground GMM color data model; (4) the GMM labels of all pixels in the studied image are updated in real time; (5) the real-time update the GMM parameter $\theta$; (6) use the max-flow/min-cut algorithm to process the Gibbs energy function to minimize its value; and (7) the algorithm continues to execute until the iteration abort condition is reached, otherwise it returns (4).

**Stereo-matching algorithm for omnidirectional motion based on Bayes theory**

By scaling the local surface shape index map of the 3-D target, the multiscale characteristics of the target surface are simulated, and feature points are defined at each scale, so that stable features can be extracted regardless of the target’s scale change. The Gaussian function can produce smooth characteristics that meet the special criteria of scale-space requirements. As the Gaussian parameter $\sigma$ decreases, the details of the data features appear in turn.

For any image $f(x, y, z)$, the result after Gaussian smoothing can be expressed as:

$$L(x, y, \sigma) = G(x, y, \sigma) * f(x, y), G(x, y, \sigma) = \frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}$$

(5)

Since the information in the Gaussian scale space is redundant, a Gaussian scale-space consisting of a set of Gaussian smooth images is first established and then a point-to-point subtraction operation is performed on adjacent scales to generate a Gaussian differential (difference of Gaussian) scale space, that is $D(x, y, z, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$.

To get key descriptor $(L_{ii}, R_{ii})$, you need to meet:

$$\frac{D_{LR1}}{D_{LR2}} < \tau$$

(6)

Among them, $D_{LR1}$ represents the point $R_i$ closest to $L_i$ in the figure, $D_{LR2}$ represents the point $R_i$ closest to $L_i$ in the figure, and $\tau$ represents the threshold. When the above formula is satisfied, this point is considered to be the best matching point. Under this condition, mismatch will still occur. Using random sample consensus (RANSAC) operator to remove mismatch, we can get accurate support point matching pair $L_i = (l_1, l_2, \cdots, l_m)$ and $R_i = (r_1, r_2, \cdots, r_n)$.

Let the image pair be $[imagel, imager]$, the set of matched support points is $S = (s_1, s_2, \cdots, s_n)$, and let each support point $s_m = (u_m, v_m, d_m)$, $(u_m, v_m)$ represents the image coordinates, and the parallax of the $d_m$ support point pair and use this value as the prior probability. Taking the collected left image as the reference image and using $O_n^l$ and $O_n^r$ to represent the remaining nonsupporting points except the supporting points in the left and right images, the joint probability distribution is:

$$p(d_n, O_n^l, O_n^r, S) \propto p(d_n | S, O_n^l)p(O_n^l | S, O_n^r, d_n)$$

(7)

where $p(d_n | S, O_n^l)$ represents the prior distribution function and $p(O_n^l | O_n^r, d_n)$ represents the image likelihood function. Interpolate each triangle in the Delaunay triangle mesh obtained using the support points to get the plane:

$$\mu_i(O_n^l) = a_iu_n + b_iv_n + c_i$$

(8)

where $i$ represents the pixel value corresponding to the plane. The Bayesian theory is used to obtain the posterior probability distribution of Markov random fields by using the disparity data of different regions:

$$P(D|I) = \exp(-E)$$

(9)

where $D$ is the parallax distribution of the image, $I$ is the stereo image pair, and $E$ is the global energy function.
**Experiment**

*Data sources*

The binocular camera used in this article is a small-scale industrial digital camera with Giga Giga network; the two cameras are fixed and installed in parallel on the axis. The camera is connected to the computer through two Giga network cables, and the corresponding software development kit (SDK) is installed to complete the camera setting, so that the images collected by the camera can be transferred into the computer with high fidelity. In the actual experiment, the installation of binocular camera cannot be completely parallel, so it is necessary to calibrate it and adjust the camera to achieve the optimal layout.

To ensure the accuracy of the experiment, the chessboard calibration board of visco is selected in this experiment. The corresponding model is cp-400-o-30, and the calibration board is 12/9 square array. The size of each square is 30 mm/30 mm, and the error accuracy is within ±0.2 mm. In this experiment, the binocular camera collects 17 images of calibration plate with rotation and offset and saves the corresponding calibration images. The corner points of each pair of calibration plate images collected by the left and right cameras are extracted. The corner points are mainly the inner corner points of the calibration plate, namely the most central 6 × 9 line corner points. See Table 1 for the specific parameters of 17 images.

**Selection of positioning system for handling robot**

1. **Robot motion system control**

After obtaining the calibration relationship of the system camera and successfully obtaining the 3-D coordinates of the target workpiece position, this article uses a binocular stereovision-based visual servo system control method to guide the robot to grab the target workpiece. The basic principle of binocular stereovision-based visual servoing system is to estimate the target pose in 3-D Cartesian space by extracting the characteristics of the target image, combining the known target geometric model and the internal and external parameters of the camera. Based on the pose deviation, a visual controller is designed to obtain the desired movement speed of the end of the manipulator in Cartesian space. The Jacobian matrix of the robot is used to calculate the movement speed of the joint space. The robot controls the movement of the robot.

2. **Error analysis and compensation system**

In the process of motion, industrial robots generate inertial forces of movement, which cause the rods of the robot to deform elastically, causing vibration, and affecting the positioning accuracy of the robot to a certain extent. This article analyzes the elastic vibration error, adopts the binocular stereovision-based elastic vibration error compensation strategy, and uses the elastic vibration error model of the handling robot to solve the deformation of the flexible rods in each position. The least square method is used to correct the trajectory, the correction result is as the compensation amount of the ideal joint, and the influence of the flexible deformation of the rod on the positioning accuracy of the industrial robot is further reduced.

**Data sources and equipment**

The omnidirectional motion handling robot positioning system based on binocular stereovision studied in this article uses a six-axis industrial robotic arm based on YCD-I six-degree-of-freedom mechanical servo system. Through corresponding theoretical research, the robot’s positioning of the workpiece is finally realized. The hardware system includes a six-axis robot, HNY-CV-002 parallel binocular camera, light source, computer, and workpiece.

Binocular stereovision is the simultaneous acquisition of images of the same target by two cameras, which have strict requirements on the calculation speed and transmission speed of image data. This article uses the HNY-CV-002 parallel binocular camera (see Table 2). It uses a universal serial bus 3.0 communication interface with a frame rate of 30 frames/s and a transmission speed greater than 100 megabytes/s. The main parameters of HNY-CV-002 camera are presented in Table 2.

**Experimental environment and software implementation process**

This article uses Visual Studio 2010 software platform to realize the vision positioning technology of natural stacked workpieces and combines Halcon 12 and Open CV 3.1.0

---

**Table 1. Image parameters of checkerboard calibration board.**

| Number | Size (mm) | Error accuracy (mm) | Speckle size (pixel) | Speckle number |
|--------|-----------|---------------------|----------------------|----------------|
| 2      | 30 × 30   | 0.11                | 640 × 640            | 3000           |
| 3      | 30 × 30   | 0.15                | 800 × 640            | 3000           |
| 2      | 30 × 40   | −0.09               | 800 × 640            | 3700           |
| 3      | 30 × 40   | 0.12                | 800 × 800            | 3100           |
| 3      | 40 × 30   | −0.12               | 800 × 800            | 3400           |
| 3      | 40 × 30   | −0.05               | 800 × 960            | 3200           |
machine vision algorithm library to write and verify the algorithm of the design system of this article. At the same time, the entire software system is integrated into the software operation interface using the Qt graphical user interface (GUI), which improves the degree of system visualization. Second, the camera driver and SDK development were redeveloped with Halcon software. The algorithm implementation and system GUI design were in C++ and Qt languages, respectively.

The entire system is divided into a template part and a recognition part. The template part mainly uses the point cloud of a single workpiece as a template, and the recognition part is the recognition of the target workpiece in the point cloud image to be measured. The interface is divided into three operation modules: camera calibration module, 3-D reconstruction module, and target recognition and positioning module. The camera calibration module uses a binocular camera to collect the image of the calibration plate, extracts the corners of the central part of the checkerboard calibration plate, and outputs the calibration result. The 3-D reconstruction part reconstructs the speckle map of the collected workpiece and successively outputs supporting point cloud, dense point cloud, and point cloud package map. Target recognition and positioning according to the input point cloud template of a single workpiece, the point cloud image to be measured is identified and located, and the recognition result and pose information under coarse and fine extraction are output.34

The software implementation process of this system is shown in Figure 1.
Results and discussions

Analysis of robot feature extraction results

(1) Contrast test of scale + rotation change

From the experimental results in Figure 2, it can be seen that, in terms of matching accuracy, the matching accuracy of the two algorithms is not much different, and both remain around 90%; in terms of the time consumed for feature description, the improved binary descriptor used in SURF algorithm in this article compared to traditional SURF, the operating speed is increased by about 1.62 times.

(2) Contrast test of image blur

In the image blur change comparison experiment, Figure 3(a) is the matching image between the original image and the 3 × 3 Gaussian template after the blur change and experimental result data. Figure 3(b) is the original image and the 5 × 5 Gaussian template matching map between blurred images and experimental result data. From the above experimental results, it can be known that under fuzzy changes, both algorithms can maintain a high matching accuracy, both of which are above 90%. In particular, the matching accuracy of the improved SURF algorithm in this article is above 95%, which is higher than the traditional SURF algorithm about 3.6 percentage points. In terms of algorithm operation speed, compared with the traditional SURF algorithm, the operation speed of the improved SURF algorithm in this article is increased by about 1.73 times.

(3) Comparative test of light changes

Figure 4 is the image feature matching results of the two algorithms under different lighting conditions. Figure 4(a) and (b) is the experimental data results under 20% and 40% lighting changes, respectively. From the results obtained in Figure 4(a) and (b), it can be seen that when the light intensity is increased, the number of matching pairs is
reduced, from 179 to 134 and from 96 to 71, which reduces the matching by about 25%. Yes, but the matching accuracy of the two algorithms is relatively stable, both of which are above 90%. In addition, compared with the traditional SURF algorithm, the matching accuracy of the improved SURF algorithm is improved by about 2.51 percentage points. In terms of algorithm running speed, compared with the traditional SURF algorithm, the improved SURF algorithm in this article runs 1.86 times faster.

In this paper, the improved SURF algorithm is described by replacing the feature point description method in the traditional SURF algorithm with the binary descriptor BRIEF operator. The improved SURF algorithm still uses the traditional SURF algorithm for feature point detection at the feature point detection stage, so that the obtained feature points have scale-invariant features; and the gray point centroid method is used to calculate the main direction of the feature points to make the detected features points have directionality, which makes the algorithm rotation invariant. In the feature point description stage, the binary descriptor BRIEF operator is used for feature description. This description method has fast calculation speed and reduces the algorithm’s operation time. Finally, through a large number of experiments, it is proved that compared with the traditional SURF algorithm, in the application scenario where image changes exist, the improved SURF algorithm not only improves the real-time performance of the algorithm but also maintains the same performance as the traditional SURF. The SURF algorithm can meet the fast and accurate matching of the target workpiece image by the omnidirectional motion handling robot.

**Analysis of robot target positioning and reconstruction results**

According to the binocular stereo camera calibration and 3-D reconstruction theory of target workpiece feature points described above, this section designs a 3-D reconstruction error analysis experiment to judge and analyze the binocular stereo camera calibration accuracy and 3-D reconstruction of target workpiece feature points. The binocular stereovision system studied in this article uses the HNY-CV-002 parallel binocular camera with a baseline distance of 55 mm. The camera’s left and right camera coordinate system only has translation in the x-axis direction, and the internal parameters of the left and right cameras basically the same. The experiments performed 3-D reconstruction on the corner points of the extracted feature map. The accuracy of the 3-D reconstruction method was verified by calculating the distance between two adjacent corner points in the reconstructed feature map. The length of each square in the actual black and white grid is 32 mm. Calculate the distance between adjacent corner points in the black and white grids in the 3-D reconstruction result and then perform the same with the side length of the grid in the actual black and white grid comparative analyses, the experimental results are presented in Table 3. The error curve is shown in Figure 5.

From the results in Table 3 and Figure 5, it can be known that the average error of the reconstruction results obtained by the 3-D reconstruction method used in this article is 0.276 mm when the actual side length of the grid is 32 mm. It can be inferred that the 3-D reconstruction method has high accuracy and can meet the requirements of the flexible loading and unloading robot for grasping the target workpiece. By analyzing the errors, we can know that the reasons for the above errors are as follows: (1) The inevitable errors generated during the stereo calibration of the binocular camera, such as the errors in the extraction of the corner points of the black and white grid plates during calibration. (2) During the process of collecting images through the camera, due to the influence of environmental factors, such as light and noise interference. (3) The camera installation problem is impossible for the left and right cameras to be absolutely parallel.

This article mainly studies the positioning and reconstruction of the target workpiece position. First, the stereo-matching technology of binocular stereovision is introduced. Second, the segmentation method is used to eliminate the influence of background factors to obtain the

| Number | Calculated | Actual value | Error between the two |
|--------|------------|--------------|-----------------------|
| 1–2    | 31.85      | 32           | 0.15                  |
| 2–3    | 32.58      | 32           | 0.58                  |
| 3–4    | 31.87      | 32           | 0.13                  |
| 4–5    | 32.20      | 32           | 0.20                  |
| 5–6    | 32.45      | 32           | 0.45                  |
| 6–7    | 32.29      | 32           | 0.29                  |
| 7–8    | 32.13      | 32           | 0.13                  |
| 8–9    | 31.57      | 32           | 0.43                  |
| 9–10   | 31.88      | 32           | 0.12                  |
foreground target. Then, the foreground target in the left and right images workpiece is stereo matching. This article uses the improved SURF feature stereo-matching method to obtain the stereo-matching result. Then, by using the point-based 3-D reconstruction method, the target workpiece position is 3-D reconstructed to obtain the 3-D coordinates of the target workpiece position. Finally, the accuracy of the 3-D reconstruction method is verified by 3-D reconstruction error experiments.

**Analysis of robot stereo-matching results**

In the experiment, two pairs of binocular cameras and a digital speckle projection were used. The generated digital speckle was irradiated to the surface of the object by the projector, and the left and right camera pictures were collected by the camera. In this experiment, a gypsum portrait is used as the target object, because there is more complex texture information in the gypsum portrait and using this as a target workpiece can better reflect the superiority of the algorithm.

First, after the binocular camera is calibrated, the target object is placed at a designated position on the work platform, and a speckle projector projects a digital speckle image onto it. After the images are collected by the binocular camera, the support points and parallax are calculated using SURF feature detection and epipolar constraint criteria. Then, the disparity value is used as the prior probability value of the Bayes dense match. Due to the uniqueness of Delaunay triangles, combined with the dense matching algorithm of Bayes theory, disparity values of nonsupporting points are obtained. Finally, the rotation matrix and translation matrix obtained by camera calibration are used to obtain the spatial 3-D coordinates of the unsupported points to obtain the dense point cloud image. The reconstructed point cloud image is stitched to obtain a relatively complete object 3-D point cloud model.

In Figure 6 it can be seen that the point cloud information of the support point reconstruction using the SURF and epipolar constraint rules can clearly describe the contour information of the target object, and the reconstructed edge information is relatively smooth. The point cloud information obtained by the Bayes matching algorithm clearly expresses the details and texture features of the target object, has higher accuracy, and reduces the mismatch rate, and its accuracy is 0.1 mm.

By comparing the SURF and digital speckle reconstruction method (referred to as method 2 for short) and the reconstruction method described in this article (referred to as method 1), the comparison is made from two aspects of running time and logarithm of matching points. It can be seen from the figure that although method 1 is 0.04 s slower than method 2, the number of matching point pairs obtained by method 1 reconstruction method is 16 times that of method 2, and the obtained 3-D point cloud density is greater, which is more beneficial to the object.

In this article, binocular stereovision reconstruction methods are compared and different binocular stereo-matching methods are introduced. Aiming at the problem that most binocular stereo-matching algorithms cannot meet the requirements of reconstruction accuracy and real-time performance, a SURF feature detection algorithm is used to extract feature points from speckle images, and the Bayes stereo-matching algorithm and Delaunay triangle method are used to obtain the denseness of the target object. The disparity information is combined with camera calibration parameters to obtain a dense point cloud image. Experimental verification shows that the algorithm can obtain high-density point cloud information with a slight loss of speed and fully characterize the texture and contour information of the target object.

**Conclusions**

Based on binocular stereovision, this article uses improved image feature extraction and matching algorithm of SURF algorithm and combines point-based 3-D reconstruction method to achieve 3-D reconstruction of target workpiece position, obtain 3-D coordinates of target workpiece position, and complete the robot.

This article studies the problem of robot image feature extraction and matching and proposes an improved SURF algorithm that combines the advantages of both SURF and BRIEF. The feature description phase of the algorithm uses
binary descriptor BRIEF operator for feature description, which greatly simplifies the complexity of the algorithm. At the same time, Hamming distance is used to match feature points. Ratio detection method and RANSAC algorithm are used to eliminate false matches. It is shown that the improved algorithm greatly improves the matching speed, ensures the real-time performance of the algorithm, and retains the robustness of the traditional SURF algorithm.

In this article, the problem of positioning the target workpiece of the robot is studied. The stereo coordinate matching method based on the improved SURF algorithm is used to obtain the image coordinates of the matching point pair. Then, the image coordinates of the target workpiece position marker points are calculated based on the average method. Finally, the camera stereo is combined. The internal and external parameters obtained by calibration and point-based 3-D reconstruction can reconstruct the position of the target workpiece and obtain the 3-D coordinates of the position of the target workpiece, so as to complete the positioning work.

This paper also uses a combination of binocular camera and laser speckle to design a software framework for real-time 3-D object reconstruction. A Bayesian-based matching algorithm combined with Delaunay triangulation is used to obtain the relationship between supported and unsupported points. Under the premise of less loss of reconstruction efficiency, the accuracy of point cloud reconstruction is improved.

**Declaration of conflicting interests**
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Key Research and Development Program of Shaanxi (Program No.2019GY-123).

**ORCID ID**
Bo Tang @ https://orcid.org/0000-0003-3544-9687

**References**

1. Chai G. Dynamic optimized formation transform method of agricultural mobile robot based on cooperative location. Revista de la Facultad de Agronomia de la Universidad del Zulia 2019; 36(4): 76–82.

2. Clavien L, Lauria M and Michaud F. Instantaneous centre of rotation based motion control for omnidirectional mobile robots with sidewards off-centred wheels. Robot Auton Syst 2018; 106: 58–68.

3. Yorozu A and Takahashi M. Obstacle avoidance with translational and efficient rotational motion control considering movable gaps and footprint for autonomous mobile robot. Int J Control Autom 2016; 14(5): 1352–1364.

4. Chang H, Sun P and Wang S. Output tracking control for an omnidirectional rehabilitative training walker with incomplete measurements and random parameters. Int J Syst Sci 2017; 48(3): 1–12.

5. Wang J, Hosoda M, Tshikudi DM, et al. Intraluminal laser speckle rheology using an omni-directional viewing catheter. Biomed Opt Exp 2017; 8(1): 137.

6. Hoang VD and Jo KH. A simplified solution to motion estimation using an omnidirectional camera and a 2-D LRF sensor. IEEE T Ind Inform 2017; 12(3): 1064–1073.

7. Yang Z, Lu S, Wu T, et al. Detection of morphology defects in pipeline based on 3D active stereo omnidirectional vision sensor. IET Image Process 2018; 12(4): 588–595.

8. Zhang Z, Zhang X, Pan H, et al. A novel steering system for a space-saving 4WS4WD electric vehicle: design, modeling, and road tests. IEEE Trans Intell Transp Syst 2016; 99(1): 1–14.

9. Glorieux E, Franciosa P and Ceglarek D. End-effector design optimisation and multi-robot motion planning for handling compliant parts. Struct Multidiscip Optim 2018; 57(3): 1377–1390.

10. Busch B, Grizou J, Lopes M, et al. Learning legible motion from human–robot interactions. Int J Soc Robot 2017; 9(5): 765–779.

11. Yildiz I. A low-cost and lightweight alternative to rehabilitation robots: omnidirectional interactive mobile robot for arm rehabilitation. Arab J Sci Eng 2017; 43(1): 1–7.

12. Tsai S, Xue Y, Huang P, et al. Establishing a criteria system for green production. P I Mech Eng B Eng J 2014; 229(8): 1395–1406.

13. Andreyev AS and Peregudova OA. The motion control of a wheeled mobile robot. J Appl Mathe Mech 2016; 79(4): 316–324.

14. Costa PJ, Moreira N, Campos D, et al. Localization and navigation of an omnidirectional mobile robot: the robot@factory case study. IEEE Rev Iberoam De Tecnol Del Aprendiz 2016; 11(1): 1–9.

15. Fei Y and Xu H. Modeling and motion control of a soft robot. IEEE T Ind Electron 2017; 64(2): 1737–1742.

16. Hacene N and Mendil B. Fuzzy behavior-based control of three wheeled omnidirectional mobile robot. Int J Autom Comput 2019; 16(2): 163–185.

17. Ding Y, Fan J and Song T. Optimal forward design method for the binocular stereo vision inspection system. Chin J Sci Instrum 2016; 37(3): 650–657.

18. Campagnoli C, Croom S and Domini F. Stereovision for action reflects our perceptual experience of distance and depth. J Vis 2017; 17(9): 21.

19. Shao X, Eisa MM, Chen Z, et al. Self-calibration single-lens 3D video extensometer for high-accuracy and real-time strain measurement. Opt Exp 2016; 24(26): 30124.

20. Wang Q, Yin Y, Zou W, et al. Measurement error analysis of binocular stereo vision: effective guidelines for bionic eyes. IET Sci Meas Technol 2017; 11(7): 829–838.

21. Zhao M, Li H and Tong Y. Research on the calibration method of “bi-binocular” stereo vision system. Optl Tech 2017; 43(5): 385–393.

22. Li Y, Cheng S and Sun M. Accelerating strategy for stereo matching in binocular vision system. Optl Tech 2018; 44(2): 216–220.
23. Tang Y, Li L, Wang C, et al. Real-time detection of surface deformation and strain in recycled aggregate concrete-filled steel tubular columns via four-ocular vision. Robot CIM-Int Manuf 2019; 59: 36–46.

24. Yue-fei X, Xi-qiao H, Mo-qi L, et al. A new combustion diagnostic method based on stereo vision technology. J Propul Technol 2017; 38(3): 673–679.

25. Xiong BS, Luo YH, Huang JP, et al. Measurement method of torsion of helicopter rotor blades based on stereo vision. Yingyong Kexue Xuebao/J Appl Sci 2018; 36(6): 950–957.

26. Tang Q, Liu X, Wu Y, et al. Analysis of speckle characteristics in three-dimensional reconstruction based on digital speckle. Chin J Lasers 2018; 45(10): 1–8.

27. Koyama S, Onozawa K, Tanaka K, et al. Multiocular image sensor with on-chip beam-splitter and inner meta-micro-lens for single-main-lens stereo camera. Opt Exp 2016; 24(16): 18035.

28. Priya L and Anand S. Object recognition and 3D reconstruction of occluded objects using binocular stereo. Cluster Comput 2018; 21(8): 1–10.

29. Zhang Z, Guo WX, Su L, et al. Stereo matching algorithm for robot binocular images in real time. T Beijing Ins Tech 2017; 37(9): 975–981.

30. Hui J, Yang Y, Hui Y, et al. Research on identify matching of object and location algorithm based on binocular vision. J Comput Theor Nanos 2016; 13(3): 2006–2013.

31. Jin Z, Zhao M, Zhang R, et al. Optimized aggregated channel features pedestrian detection algorithm based on binocular vision. Tianjin Daxue Xuebao 2016; 49(12): 1225–1230.

32. Xu Y and Liu R. Path planning for mobile articulated robots based on the improved A* algorithm. Int J Adv Robot Syst 2017; 14(4): 1–10.

33. An L, Zhang X, Gao H, et al. Semantic segmentation–aided visual odometry for urban autonomous driving. Int J Adv Robot Syst 2017; 14(5): 1–11.

34. Qin S. Identification of agricultural weeds based on image processing. Revista de la Facultad de Agronomia de la Universidad del Zulia 2019; 36(6): 122–128.