REGULAR PAPER

A High Precision and Fast Alignment Method Based on Binocular Vision

Han Gao[1,3] · **Fei Shen**[1,2,3] 🅾 · **Feng Zhang**[1,2,3] · **Zhengtao Zhang**[1,2,3]

Abstract

Alignment tasks for precision electronics manufacturing require high accuracy and low time consumption. However, in the current industrial environment, multiple servo alignment operations are often required to achieve the desired accuracy targets, which is time-consuming. In this paper, a high precision, fast alignment method based on binocular vision is proposed, which allows the accurate movement of the workpiece to the target position in only one alignment operation, without the need for a standard calibration board. Firstly, a calibration method of the telecentric lens camera is proposed based on an improved nonlinear damped least-squares method to establish the relationship between the image coordinate system and the local world coordinate system in the binocular vision system. Secondly, in order to transform the coordinates from the local world coordinate system to a unified coordinate system with the platform's rotation center as the origin, an angle constraint-based rotation center calibration method is proposed. Thirdly, a two-stage feature point detection method based on shape matching is proposed to detect the feature points of the workpiece. Based on these, the position and pose of the workpiece are obtained. Then the alignment commands are calculated based on the current and the target position and pose of the workpiece, enabling the accurate alignment to be accomplished in one operation. Finally, taking the mobile phone's cover glass alignment task as an example, a series of calibration and alignment experiments were carried out. The experiments and results show that the alignment errors are within ± 0.020 mm and the time taken to calculate alignment commands is less than 20 ms, which demonstrates the effectiveness of the proposed method.

Keywords Visual alignment · Binocular vision · Camera calibration · Image processing

1 Introduction

Vision-based alignment is one of the critical techniques in the domain of vision inspection and assembly, which can be widely applied in fields of electronic equipment, semiconductor and robotics. For example, the precise alignment task of the optical fibers is performed by a closed-loop control based on the telecentric stereo microvision [1]. In [2], the assembly task of the slice micropart in 3-D space is completed by the serial assembly with three microscopic cameras and a laser triangulation measurement instrument (LTMI). Although vision-based alignment has the advantage of high-speed, non-contact, high accuracy, and flexibility, the alignment methods are open to further development to improve precision and robustness.

In general, a vision-based alignment system consists of a vision system and a motion system. Firstly, the vision system measures the position and pose of the workpiece by recognizing features in the captured images. Secondly, the alignment commands are obtained by the controller based on the deviation of alignment. Finally, the motion system completes the alignment process based on the alignment commands from the controller. Therefore, the overall performance of the vision-based alignment system depends on the accuracy of the alignment method's crucial parts, which are the measurement of the workpiece pose and position, the calibration accuracy of alignment system parameters, and the control strategy of the alignment process.

✉ Fei Shen
  fei.shen@ia.ac.cn

1   Research Center of Precision Sensing and Control,
    Institute of Automation, Chinese Academy of Sciences,
    Beijing 100190, People's Republic of China

2   CASI Vision Technology CO., LTD., Luoyang 471000,
    People's Republic of China

3   University of Chinese Academy of Sciences, Beijing 100049,
    People's Republic of China

To measure the workpiece's pose and position accurately, serial steps need to be carried out. Firstly, the calibration of the vision system is required to map the coordinate of a point from the image coordinate system to the world coordinate system. Secondly, the image processing method should be used to recognize features in the captured images. Finally, the position and pose of the workpiece are obtained based on the calibration result of the vision system and the workpiece's features. Depending on the number of cameras used in the vision system, the measurement methods of the workpiece's position and pose can be mainly divided into two categories which are the monocular vision method [3–6] and the multi-vision method [7–10]. Furthermore, due to advantages such as high resolution and less distortion, telecentric lenses are widely used for non-contact measurement of the workpiece's position and pose.

To achieve high precision of the alignment task, the control strategy has been widely adopted in the alignment process. The feedback control-based alignment method can achieve the desired accuracy after several control periods with the closed-loop control strategy and the deviation of alignment as the feedback amount. As a hot research topic in the feedback control-based alignment method, the visual servo technology is utilized to complete the alignment task in a coarse-to-fine manner [11–13]. Y. Ma et al. [11] proposed a coordinated pose alignment strategy with two microscopic cameras to realize pose alignment. In [12], a vision-based system is proposed to automatically complete the watch hand's precise alignment. S. Kwon et al. [13] proposed an alignment system with a visual servo to accomplish the coarse-to-fine alignment task. The vision system is designed to recognize the alignment marks, and an observer-based is designed for the display visual alignment tasks. With the robustness to the environmental variation and the achievable high precision, the visual servo-based alignment method has been adopted in the field of microassembly and micro-manipulation, where the precision requirement is high, with speed being a secondary requirement.

However, visual servo-based alignment methods must be completed after several control cycles, which means that a long time is required before the desired alignment accuracy is achieved. For vision-based alignment and inspection tasks in manufacturing lines, the alignment task is only one part of the production [14]. In order not to have an impact on subsequent production, the alignment task must be accomplished in a few tens of milliseconds in one shot alignment operation. In this case, the visual servo-based alignment method with several control periods is no more suitable. Therefore, in this paper, a high precision and fast alignment method is proposed to accomplish the alignment task after one-shot alignment operation, with a binocular vision system as the vision subsystem of the
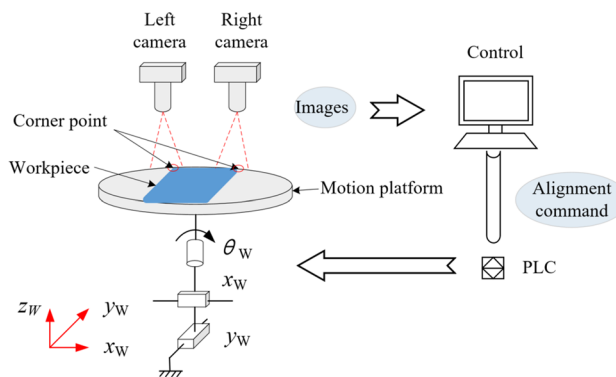

Fig. 1 The schematic of alignment system

alignment system. We discuss how to improve the performance of the key parts of the alignment method.

Firstly, to establish the relationship between the image coordinate system and the local world coordinate system, an improved nonlinear damped least-squares calibration method for the telecentric lens camera is proposed to speed up the convergence of the camera calibration process. Secondly, to complete the alignment task in one shot operation, a world coordinate system with the rotation center of the rotation platform as the origin needs to be obtained in advance to unify the local world coordinate systems of the binocular vision system. Thus, an angle constraint-based rotation center calibration method is proposed through the active rotation of the motor three times. Thirdly, a two-stage feature point detection method based on shape matching is proposed to obtain the feature point of the workpiece robustly. Finally, a series of experiments are conducted on an alignment system to verify the effectiveness of the proposed alignment methods.

The rest of this paper is organized as follows. In Sect. 2, the binocular vision-based alignment system structure is introduced, and the coordinate systems are established. In Sect. 3, the proposed alignment methods are presented, including an improved nonlinear damped least-squares calibration method for the telecentric lens camera, an angle constraint-based rotation center calibration method, and the calculation method for the alignment commands. In Sect. 4, the experiment results and error analysis are shown. Finally, the conclusion and the suggestions for further work are given in Sect. 5.

2 Alignment System and Coordinate Systems

2.1 Binocular Vision-based Alignment System

As shown in Fig. 1, a binocular vision-based alignment system is designed to complete the alignment task, which

consists of a motion system, a binocular vision system and a control system.

(1) The motion system consists of a motion platform with a two dimensional translation platform and a rotation platform. The alignment task is completed by the movement of the motion platform, on which the workpiece is fixed.

(2) The binocular vision system, designed to measure the position and pose of a workpiece by capturing images of the workpiece, consists of two telecentric lens cameras. The cameras are mounted with the optical axis direction orthogonal to the motion platform.

(3) The control system is designed to calculate the alignment commands according to the images received from the vision system and to drive the motion platform to complete the alignment task through a programmable logic controller(PLC).

2.2 Establishment for the Coordinate Systems

As shown in Fig. 2, some relevant points need to be labeled before establishing the coordinate systems. The rotation center of the rotation platform is denoted as p_{WO}. The left and the right corner of the workpiece are labeled as p_{WL} and p_{WR}, respectively. The coordinate systems mentioned in this paper are established when the motion platform is in the reset state. $o_W x_W y_W z_W$ is a unified world coordinate system with coordinate axes parallel to the directions of the motion platform's movement and p_{WO} as its origin. $o_{WL} x_{WL} y_{WL} z_{WL}$ and $o_{WR} x_{WR} y_{WR} z_{WR}$ are local-world coordinate systems with their axes parallel to the unified coordinate system $o_W x_W y_W z_W$'s axes. p_{WL} and p_{WR} are set as the coordinate origin of the $o_{WL} x_{WL} y_{WL} z_{WL}$ and $o_{WR} x_{WR} y_{WR} z_{WR}$, respectively. $o_{CL} x_{CL} y_{CL} z_{CL}$ and $o_{CR} x_{CR} y_{CR} z_{CR}$ are labeled as camera coordinate systems of the left and
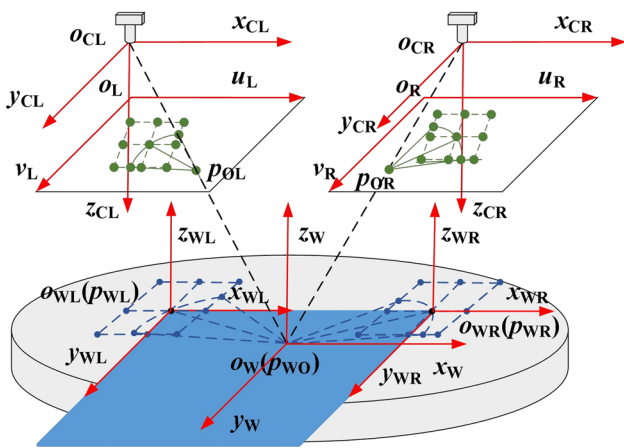
the right cameras, respectively, whose optic axes coincide with $o_{CL} z_{CL}$ and $o_{CR} z_{CR}$ correspondingly. $o_L u_L v_L$ and $o_R u_R v_R$ are image coordinate systems of the left and the right vision systems, respectively. $o_L u_L$, $o_L v_L$, $o_R u_R$ and $o_R v_R$ are parallel to $o_{CL} x_{CL}$, $o_{CL} y_{CL}$, $o_{CR} x_{CR}$ and $o_{CR} y_{CR}$, respectively.

3 Fast Alignment Method Based on Binocular Vision

As shown in Fig. 3, the alignment method based on the binocular vision proposed in this paper is divided into two stages. The vision system is calibrated in Stage I, whose main tasks include the camera calibration based on an improved Levenberg–Marquardt algorithm and the calibration of the rotation center with an angle-based constraint. In Stage II, the alignment control command is calculated through the following three steps.

1) With the images of the workpiece as input, the fast feature point detection method is utilized to get the image coordinates of the feature points on the workpiece, such as the corner points.

2) Based on the calibration model established in Stage I, the image coordinates of the feature points are transformed to the unified world coordinates.

3) Combined with the target position and pose, the alignment command is calculated to rectify the position and pose of the workpiece to the target in one operation.

We will detail the calibration method of the vision system in Sect. 3.1, the establishment of the alignment model and the calculation of the alignment command in Sect. 3.2, and a two-stage feature point detection method based on shape matching in Sect. 3.3, which is a critical factor for high accuracy vision-based alignment.
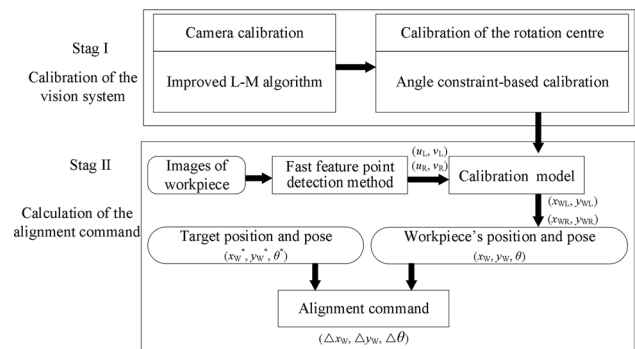

Fig. 2 Coordinate systems of the binocular vision system

Fig. 3 Flowchart of the alignment method

3.1 Calibration for the Vision System

The calibration of the vision system consists of two components: the calibration of the camera and the calibration of the rotation platform's rotation center. The former is used to obtain the relationship between the image coordinate system $o_L u_L v_L$, $o_R u_R v_R$ and the local world coordinate system $o_{WL} x_{WL} y_{WL} z_{WL}$, $o_{WR} x_{WR} y_{WR} z_{WR}$, while the latter is used to calculate the rotation center for achieving the alignment task in one operation. Firstly, the telecentric lens camera model is established, and an improved Levenberg–Marquardt algorithm is used to obtain the camera model's calibration parameters. Then an angle-based constrained calibration method is proposed to overcome the sensitive problem of measuring the rotation center in industrial scenarios.

3.1.1 Telecentric Lens Camera Model

Take the left camera as an example. Notice that the coordinate system has been established in Sect. 2.2, the relationship between (u_L, v_L) and (x_{CL}, y_{CL}, z_{CL}) can be given as follows.

$$\begin{bmatrix} u_L \\ v_L \\ 1 \end{bmatrix} = \begin{bmatrix} k_L & 0 & 0 \\ 0 & k_L & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{CL} \\ y_{CL} \\ 1 \end{bmatrix} (z_{CL} \in [z_0, z_0 + \Delta z]) \quad (1)$$

where k_L is the magnification factor of the telecentric lens, z_0 is the location where the telecentric image is sharpest, and Δz is the telecentric depth. Set $z_{WL} = 0$, the relationship between (x_{CL}, y_{CL}, z_{CL}) and (x_{WL}, y_{WL}, z_{WL}) is a combination of the rotation transformation and the translation transformation as follows.

$$\begin{bmatrix} x_{CL} \\ y_{CL} \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11L} & r_{12L} & r_{13L} & p_{xL} \\ r_{21L} & r_{22L} & r_{23L} & p_{yL} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{WL} \\ y_{WL} \\ 0 \\ 1 \end{bmatrix} \quad (2)$$

Due to the rotation angles between the left camera coordinate system and the world coordinate system can be expressed as $\theta_{zL}, \theta_{yL}, \theta_{xL}$, the parameters of the rotation transformation can be described as follows.

$$\begin{aligned} & r_{11L} = \cos\theta_{zL}\cos\theta_{yL}, r_{21L} = \sin\theta_{zL}\cos\theta_{yL}, \\ & r_{12L} = -\sin\theta_{yL}\cos\theta_{xL} + \cos\theta_{zL}\sin\theta_{yL}\sin\theta_{xL}, \\ & r_{22L} = \cos\theta_{zL}\cos\theta_{xL} + \sin\theta_{zL}\sin\theta_{yL}\sin\theta_{xL}, \\ & r_{13L} = \sin\theta_{zL}\sin\theta_{xL} + \cos\theta_{zL}\sin\theta_{yL}\cos\theta_{xL}, \\ & r_{23L} = -\cos\theta_{zL}\sin\theta_{xL} + \sin\theta_{zL}\sin\theta_{yL}\sin\theta_{xL}, \end{aligned} \quad (3)$$

Then the left camera model can be given by

$$[u_L \ v_L \ 1]^T = M_L [x_{WL} \ y_{WL} \ 0 \ 1]^T \quad (4)$$

where M_L is the homography matrix of the left camera model.

$$M_L = \begin{bmatrix} k_L & 0 & 0 \\ 0 & k_L & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11L} & r_{12L} & r_{13L} & p_{xL} \\ r_{21L} & r_{22L} & r_{23L} & p_{yL} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Similarly, the right camera model can be given by

$$[u_R \ v_R \ 1]^T = M_R [x_{WR} \ y_{WR} \ 0 \ 1]^T \quad (5)$$

where M_R is the homography matrix of the right camera model.

$$M_R = \begin{bmatrix} k_R & 0 & 0 \\ 0 & k_R & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11R} & r_{12R} & r_{13R} & p_{xR} \\ r_{21R} & r_{22R} & r_{23R} & p_{yR} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3.1.2 Camera Calibration Based on an Improved Levenberg–Marquardt Algorithm

For the calibration of the left camera, the parameters in Eq. (4) need to be calculated. Considering these parameters as the unknown variables, Eq. (4) can be rewritten as

$$\begin{cases} u_{Li} - k_L(r_{11E}x_{WLi} + r_{12L}y_{WLi} + p_{xL}) = 0 \\ v_{Li} - k_L(r_{21L}x_{WLi} + r_{22L}y_{WLi} + p_{yL}) = 0, i = 1, 2, ..., m \end{cases} \quad (6)$$

Then the $2m$ equations with unknown variable $x_L = [\theta_{zL}, \theta_{yL}, \theta_{xL}, p_{xL}, p_{yL}, k_L]$ can be formed as

$$f_L(x_L) = \begin{bmatrix} f_{L1}(x_L) \\ f_{L2}(x_L) \\ ... \\ f_{L2m}(x_L) \end{bmatrix} = 0 \quad (2m \geq 6) \quad (7)$$

where $f_{L2i-1}(x_L) = u_{Li} - k_L(r_{11L}x_{WLi} + r_{12L}y_{WLi} + p_{xL})$, $f_{L2i}(x_L) = v_{Li} - k_L(r_{21L}x_{WLi} + r_{22L}y_{WLi} + p_{yL})$, $i = 1, 2, ..., m$. The nonlinear least-squares method is used to solve the aforementioned equations with the cost function $P_L(x_L)$ as

$$P_L(x_L) = \frac{1}{2} \sum_{i=1}^{2m} \left\| f_{Li}(x_L) \right\|_2^2 \quad (8)$$

Thus, the minimum value of $P_L(x_L)$ is the solution of $f_L(x_L) = 0$, named as x_L^*.

$$x_L^* = \min_{x_L \in R^6} P_L(x_L) = \min_{x_L \in R^6} \frac{1}{2} f_L(x_L)^T f_L(x_L) \quad (9)$$

With a descent step h_L^k, the first-order Taylor expansion of $f_L(x_L)$ around x_L^k is brought into $P_L(x_L)$.

$$P_L(x_L^k + h_L^k) \approx \frac{1}{2}f_L^T(x_L^k)f_L(x_L^k) + h_L^{kT}J_L^Tf_L^T(x_L^k) + \frac{1}{2}h_L^{kT}J_L^TJ_L \tag{10}$$

where J_L is the Jacobi matrix $J_{ijL} = \partial f_{iL}/\partial x_{jL}$. Based on the Levenberg–Marquardt algorithm(L-M), the iteration formula is given as follows.

$$h_L^k = -(J_L^TJ_L + \mu_L I)^{-1}J_L^Tf_L(x_L^k) \tag{11}$$

$$x_L^{k+1} = x_L^k + h_L^k \tag{12}$$

where μ_L is a damping parameter. Notice that the L–M algorithm is a trust region algorithm, the standard L-M updates μ_L by a ratio factor according to the performance of the cost function's decrease.

Its relative success notwithstanding, the standard L–M algorithm maybe sluggish, especially when the algorithm moves to a canyon with a large aspect ratio in the parameters space. Therefore, instead of requiring cost reduction in each descent step, [15] uses the cosine similarity between adjacent steps as an acceptance criterion and finds a faster converging path with increasing cost tolerance. However, this algorithm does not solve the parameter evaporation problem. When the algorithm gets lost in the plateau of the parameter space, some parameters are pushed to infinity, which also means that the ratio of adjacent steps will increase significantly.

Therefore, an improved updating strategy for the damping factor μ_L is proposed to alleviate the parameter evaporation problem and speed up the convergence of the L-M algorithm. $p_a = \cos(h_L^{k-1}, h_L^k)$ and $p_b = \min\{h_L^{k-1}, h_L^k\}$ / $\max\{h_L^{k-1}, h_L^k\}$ are calculated to obtain the damping acceptance criterion named as ζ.

$$\zeta = p_a p_b \tag{13}$$

The damping factor μ_L is updated with different strategies according to the distribution of ζ. Specifically, when $\zeta > \zeta_{th}$, it means that the L–M algorithm has accepted the current descent step. In this case, the damping factor μ_L needs to be smaller to obtain a more accurate result by decreasing with the factor m. Conversely, when $\zeta \leq \zeta_{th}$, it indicates that the L–M algorithm has refused the current descent step. Thus, the damping factor μ_L needs to be larger to expand the trust region by increasing with the factor m. In this paper, m is set as 2 and the ζ_{th} is set as 0.9. The above update strategy of the damping factor is given as

$$\mu_L^k = \begin{cases} \mu_L^{k-1} / m & \zeta > \zeta_{th} \\ \mu_L^{k-1} * m & \zeta \leq \zeta_{th} \end{cases} \tag{14}$$

Notice that the damping factor for obtaining the descent step h_L^k should be μ_L^{k-1} since μ_L^k is updated based on the h_L^{k-1} and h_L^k. Thus, the iteration formula is given as follows.

$$h_L^k = -(J_L^TJ_L + \mu_L^{k-1}I)^{-1}J_L^Tf_L(x_L^k) \tag{15}$$

$$x_L^{k+1} = x_L^k + h_L^k \tag{16}$$

To sum up, the calibration method of a camera can be described as follows.

1) Obtain m points with (x_{Wi}, y_{Wi}, z_{Wi}) and (u_i, v_i) by the active translational movement of the motion platform. Then 2 m equations are formed according to Eq. (7).
2) At the beginning of the iteration, the initial value of $x_L^k(k=0)$ can be preset as a non-zero vector. Then, the maximum value of the diagonal matrix $J_L^TJ_L$ is chosen to be the initial value of the damping factor denoted as μ_L^0. Thus, $h_L^k(k=0)$ can be calculated according to Eq. (12).
3) Then, the unknown vector x_L^k is solved by applying the iteration Eqs. (14)–(16). Specifically, in each iteration ($k > = 1$), the descent step h_L^k is obtained according to Eq. (15). x_L^k is updated by applying the iteration formula (16). Finally, the damping factor μ_L^k is updated according to the Eq. (14). The iterative process stops until the 2-norm of the two adjacent vectors' error becomes smaller than the threshold value. Thus, the homography matrix M_L can be calculated according to the calibration parameters of the left camera.

3.1.3 Calibration for the Rotation Center

To complete the alignment task for the workpiece which is fixed on the motion platform randomly in one operation, the rotation center of the rotation platform should be calibrated firstly. As shown in Fig. 4, the motion platform is rotated with an angle α three times. The left corner point and the right corner point are named as p_{WLi} and p_{WRi}, respectively. The image coordinates of p_{WLi} and p_{WRi} in the image coordinate system $o_Lu_Lv_L$ and $o_Ru_Rv_R$ are given as (u_{Li}, v_{Li}) and $(u_{Ri}, v_{Ri})(i=1,2,3)$, respectively. The local world coordinates of p_{WLi} and p_{WRi} in the local world coordinate $o_{WL}x_{WL}y_{WL}z_{WL}$, $o_{WR}x_{WR}y_{WR}z_{WR}$ can be calculated by the calibration model illustrated in Sect. 3.1 which are expressed as (x_{WLi}, y_{WLi}) and (x_{WRi}, y_{WRi}) (i=1,2,3). These local world coordinates of p_{WLi} and p_{WRi} are utilized to calibrate the rotation center of the rotation platform.

The fitting of a circle is usually calculated by the least-squares method. Then the accuracy depends on the distribution of the fitting points on the circle. I. Kåsa [16] proved that the regular placement of the data points along a circle could improve the performance of the circle fitting. Zhu Jia et al.
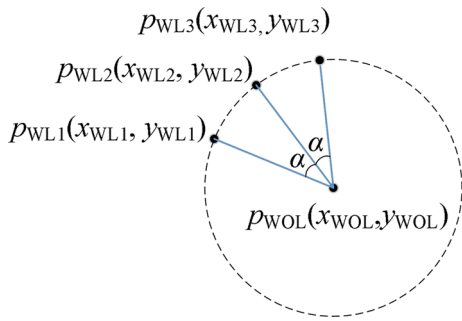
Fig. 4 The angle-based constrained calibration of the rotation center in the left camera's view

[17] demonstrated that when the center angle α is small, the transmission factor of the measurement error of the circle center and radius will increase rapidly with the decrease of the center angle α. For the case of the rotation platform of the alignment system cannot be rotated by a large angle, since it would cause the workpiece to deviate out of the field of view, the least-squares method will lead to non-negligible errors for the calculation of the rotation center. To solve this problem, we proposed a angle constraint-based calibration method for the rotation center.

Firstly, the central angle α is used as prior knowledge to calculate the candidate coordinates of the rotation center, and then the best solution with the minimum standard deviation is picked out as the coordinate of the rotation center. As shown in Fig. 4, taking the calibration of the rotation center in the left camera's view as an example, when the left corner p_{WL} moves from (x_{WL1}, y_{WL1}) to (x_{WL2}, y_{WL2}) with a step angle α, the rotation center $p_{WOL}(x_{WOL}, y_{WOL})$ can be calculated according to the following equation.

$$\begin{cases} x_{WL2} = x_{WOL} + \cos\alpha(x_{WL1} - x_{WOL}) - \sin\alpha(y_{WL1} - y_{WOL}) \\ y_{WL2} = y_{WOL} + \sin\alpha(x_{WL1} - x_{WOL}) + \cos\alpha(y_{WL1} - y_{WOL}) \end{cases}$$
(17)

Thus, the solution of the rotation center in the world coordinate system is given by

$$\begin{cases} x_{WOL} = \dfrac{x_{WL1} + x_{WL2}}{2} + \dfrac{(y_{WL1} - y_{WL2})\sin\alpha}{2 - 2\cos\alpha} \\ y_{WOL} = \dfrac{y_{WL1} + y_{WL2}}{2} + \dfrac{(x_{WL2} - x_{WL1})\sin\alpha}{2 - 2\cos\alpha} \end{cases}$$
(18)

Since that the rotation center can be obtained by any two points with their corresponding angle is known, C_n^2 candidate solutions can be obtained by using n points $p_{WLj} = (x_{WLj},$

$y_{WLj})$, $p_{WRj} = (x_{WRj}, y_{WRj})$ ($j = 1, 2, ..., n$). After calculating the standard deviation of the Euclidean distance between each candidate rotation center and all points, candidate solutions with the minimum standard deviation are taken as the best solutions, which are given as $p_{WOL}(x_{WOL}, y_{WOL})$ and $p_{WOR}(x_{WOR}, y_{WOR})$.

$$\begin{cases} p_{WOL} = \underset{p_{WOLi}}{\arg\min} \dfrac{1}{n} \sqrt{\sum_{j=1}^{n} (p_{WOLi} - p_{WLj})^2} \\ p_{WOR} = \underset{p_{WORi}}{\arg\min} \dfrac{1}{n} \sqrt{\sum_{j=1}^{n} (p_{WORi} - p_{WRj})^2} \end{cases}$$
(19)

$$i = 1, 2, ..., C_n^2, \quad j = 1, 2, ..., n$$

As shown in Fig. 2, $o_W x_W y_W z_W$ is the unified world coordinate system with its axes parallel to the motion platform's moving directions and rotation center of the rotation platform p_{WO} as the origin. Thus, based on Eqs. (4), (5) and (19), the coordinate of p_{WL} and p_{WR} in $o_W x_W y_W z_W$ can be given as follows.

$$\begin{cases} x_{WL} = \dfrac{r_{22L}(u_L - k_L p_{xL}) - r_{12L}(v_L - k_L p_{yL})}{k_L(r_{11L}r_{22L} - r_{12L}r_{21L})} - x_{WOL} \\ y_{WL} = \dfrac{r_{11L}(v_L - k_L p_{yL}) - r_{21L}(u_L - k_L p_{xL})}{k_L(r_{11L}r_{22L} - r_{12L}r_{21L})} - y_{WOL} \\ x_{WR} = \dfrac{r_{22R}(u_R - k_R p_{xR}) - r_{12R}(v_R - k_R p_{yR})}{k_R(r_{11R}r_{22R} - r_{12R}r_{21R})} - x_{WOR} \\ y_{WR} = \dfrac{r_{11R}(v_R - k_R p_{yR}) - r_{21R}(u_R - k_R p_{xR})}{k_R(r_{11R}r_{22R} - r_{12R}r_{21R})} - y_{WOR} \end{cases}$$
(20)

So far, we show how to calculate the unified world coordinates of a point according to its image coordinates. Next, we will introduce how to calculate the alignment command according to these coordinates.

3.2 Calculation for the Alignment Command

The position and pose of a workpiece can be represented by a reference point $p_W(x_W, y_W)$ and a title angle θ. In our alignment system, the middle point of the corner points is taken as the reference point of the workpiece. The angle of the line formed by the corner points is used as the title angle θ. Thus, the position and pose $\{x_w, y_w, \theta\}$ of the workpiece to be measured is given by

$$\begin{cases} x_{\mathrm{W}} = \frac{1}{2}\left(x_{\mathrm{WL}} + x_{\mathrm{WR}}\right) \\ y_{\mathrm{W}} = \frac{1}{2}\left(y_{\mathrm{WL}} + y_{\mathrm{WR}}\right) \\ \theta = \arctan\left(\dfrac{y_{\mathrm{WR}} - y_{\mathrm{WL}}}{x_{\mathrm{WR}} - x_{\mathrm{WL}}}\right) \end{cases} \tag{21}$$

Similarly, the target position can be calculated as $\{x_w{}^*, y_w{}^*, \theta^*\}$. For the calculation of the alignment command, as shown in Fig. 5, the workpiece is rotated with an angle $\Delta\theta$ to the transitional position which is parallel to the target and then the translation amount is obtained. Then, the alignment command $\{\Delta x_w, \Delta y_w, \Delta\theta\}$ is given by

$$\Delta\theta = \theta^* - \theta \tag{22}$$

$$\begin{bmatrix} \Delta x_{\mathrm{W}} \\ \Delta y_{\mathrm{W}} \end{bmatrix} = \begin{bmatrix} x_{\mathrm{W}}^* \\ y_{\mathrm{W}}^* \end{bmatrix} - \begin{bmatrix} \cos\Delta\theta & -\sin\Delta\theta \\ \sin\Delta\theta & \cos\Delta\theta \end{bmatrix} \begin{bmatrix} x_{\mathrm{W}} \\ y_{\mathrm{W}} \end{bmatrix} \tag{23}$$

3.3 A Fast Feature Point Detection Method

The detection accuracy of feature points is a crucial factor affecting the calibration and alignment accuracy. In addition, the acquisition of feature points is often the most time-consuming step in the alignment operation. However, the accurate feature point detection is difficult to be completed since the gray transition bands would make the accurate edge detection be difficult, as shown in Fig. 6. These undesirable factors are usually found in alignment tasks and are mainly caused by the position of the light source and the limitation of the platform movement.

Thus, a fast feature point detection method is proposed in this paper. As shown in Fig. 7, the method consists of three
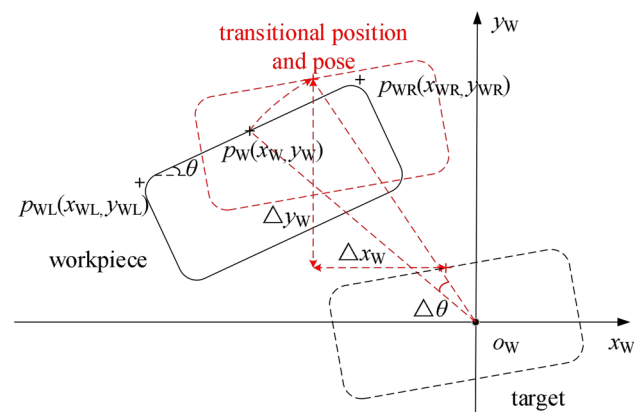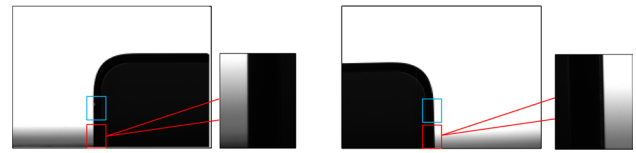

Fig. 6 Gray transition bands in the workpiece image (marked with red bounding boxes) are caused by inhomogeneous illumination and can affect the accuracy of edge detection adversely. The corner detection method proposed in this paper locates well-lit regions without gray transition bands (marked with blue bounding boxes) firstly, and then calculates the edges of the workpiece in these regions

steps, including object detection, edge extraction, and feature point extraction. Well-lit regions without gray transition bands are located in object detection step in a fast and robust manner to reduce the effects of inhomogeneous illumination and image noise. Then, the accurate edge detection is performed in these regions to ensure the detection accuracy of the corner point.

The alignment task of the mobile phone's cover glass is taken as an example to illustrate the method clearly. Since the corner points of mobile phone's cover glass are taken as the feature points and are utilized to recognize the reference point $p_{\mathrm{W}}(x_{\mathrm{W}}, y_{\mathrm{W}})$, the region of interest denoted as ROI in the input image containing straight edges is located through a coarse-to-fine shape matching method in Step 1. Then, the edge extraction is accomplished by the line fitting in Step 2. In Step 3, the left and the right corner point are obtained by calculating the intersection of the lines in the left and the right vision system, respectively. The middle point of these corner points is used as the reference point.

The premise of accurately obtaining the image coordinates of feature points is to complete the object detection quickly and robustly to adapt to the random placement of the workpiece. In this paper, a coarse-to-fine shape matching method is proposed to accomplish the object detection. The
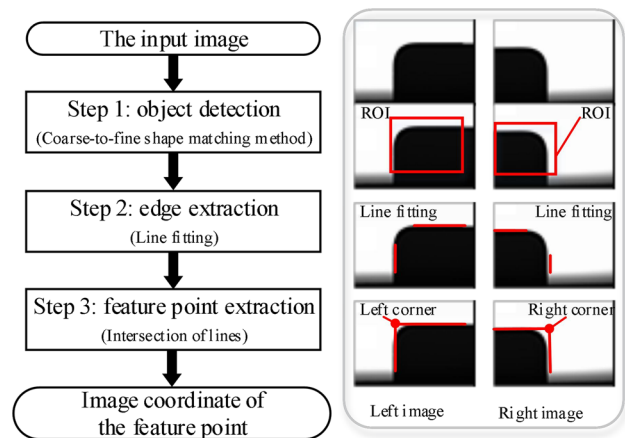

Fig. 5 The calculation of the alignment command

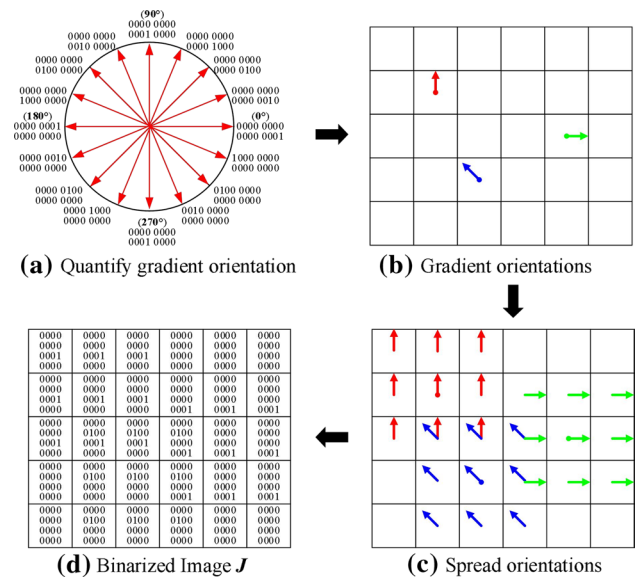**Fig. 7** Flowchart of the feature point detection method

main contribution of this method lies in the adoption of a coarse-to-fine strategy to achieve the compatibility of acceleration and accuracy. Specifically, after the template image and the input image are both downsampled, the branch-and-bound scheme and gradient spread method are used for the coarse matching to speed up the shape matching process. Furthermore, to obtain a more accurate result, the images with a higher resolution compare to the images in Stage I are adopted as inputs to carry out the fine matching based on the coarse matching results.

Stage I Coarse Shape Matching Based on the Branch-and-bound Scheme and the Gradient Spread Method.

The image matching task can be interpreted as the process of finding the best affine transformation among a large number of possible affine transformations. This process is time-consuming, so that an appropriate strategy should be found to speed up the search. Fast-Match method [18] utilized the branch-and-bound scheme for accelerating with the sum of absolute differences (SAD) as the similarity measurement. However, the SAD is calculated based on image blocks' gray values, which may be influenced by the illumination variation and cause the matching task' failure in some situations. In contrast, edge features exist widely in images and are insensitive to the illumination variation. Therefore, in this paper, edge features are considered to accomplish the image matching task for the adaptability of illumination variation. To realize this idea, the gradient direction is extracted as the descriptor for the edge features with the absolute cosine value of the gradient directions between the template and input image as the similarity measurement. Furthermore, the gradient spread process is adopted, and the branch-and-bound scheme is utilized to speed up the search of affine transformations. In addition, since edge features are utilized to accomplish the image matching, this matching method can be categorized as a shape matching method.

The gradient spread process is proposed by the LINE-MOD method [19] to keep the matching task invariant to the small translations and deformations. This process enhances the smoothness of images by diffusing the gradient direction of an edge point to points within its local neighborhood. Specifically, as shown in Fig. 8, a binarized image *J* storing the gradient direction codes is utilized to represent the input image *I* and is obtained by extracting, quantizing, encoding the gradient directions of edge points. The gradient spread process is then utilized in this process, as shown in Fig. 8c.

After the gradient spread process is utilized and the binarized image *J* of the input image *I* is established, the coarse shape matching is performed according to Algorithm 1. Firstly, the template image **Θ** and the input image *I* are downsampled

(a) Quantify gradient orientation　　**(b)** Gradient orientations

(d) Binarized Image *J*　　**(c)** Spread orientations

Fig. 8 **a** The gradient direction is quantized to 16 directions and encoded to a 16-bit code accordingly. **b** Gradient directions of the input image are extracted and quantized. **c** The gradient spread direction of one point is defined as the recording of all the gradient directions of its neighborhoods in a radius of T/2 (T = 3). **d** The gradient spread directions are encoded to 16-bit codes, and a binarized image *J* is constructed

to reduce the size of the search parameters in the pursuit of reducing time consumption. Secondly, in the preparation stage, a series of variables are established to use the branch-and-bound scheme and the gradient spread process to accelerate the shape matching. In detail, to calculate the similarity measurement quickly, the binarized image *J* of the input image is established within a neighborhood $[- T/2, T/2] \times [- T/2, T/2]$. A response table τ is also precomputed to save the maximum cosine similarity between any possible gradient direction in the template image **Θ** and any possible direction code in the binarized image *J*. Then an affine transformation net N_0 containing all possible affine transformations is constructed for searching the best affine transformation. Thirdly, the candidate affine transformation net N_C containing n candidate affine transformations is searched by the branch-and-bound scheme. By using a parameter δ to control the search precision, the size of the candidate affine transformation net N_C is gradually reduced to n, which means that n expected candidate affine transformations are obtained finally. In addition, with the size of the candidate transformation network N_C decreasing, the length T of the neighborhood decreases by the factor of δ accordingly. Algorithm 1 is depicted as follows.

Algorithm 1 Stage I Coarse Shape Matching

Input: the template image Θ, the input image I.

Output: the n candidate affine transformations.

Step 1: Downsample the template image Θ and the input image I.

Step 2: Establish the binarized image J of the input image I, the response table τ, the affine transformation net N_0.

Step 3: Search n candidate affine transformations by the branch-and-bound scheme.

    3.1 Define an iteration variable k, a parameter δ for controlling the precision, the candidate affine transformation net's maximum size m and an expected size n, a decay factor d. Initialize $\delta_0 = 1.0$, $m=10$, $n = 3$, $d = 0.8$.

    3.2 Let k=0 and repeat the following process while $m > n$.

        a. Select all affine transformations from N_k with the step of $\delta_k * T$ and calculate the similarity measurement. Record all affine transformations and the corresponding similarity measurement.

        b. Construct the refined transformation net N_k' with the top $[\delta_k * m]_{th}$ maximum similarity measures from N_k.

        c. Expand the net N_k' to the net Q_k in the local neighborhood with a radius of $d\delta_k * T/2$.

        d. Set $k = k+1$, $\delta_k = d\delta_k$, $N_{k+1} = Q_k$, $m = Size(N_k')$ and reconstruct the image J with a neighbor radius of $d\delta_k * T/2$.

Return: the candidate affine transformation net $N_C = N_k'$.

Stage II Fine Shape Matching

Since the gradient spread directions of one point is defined as the recording of all the gradient direction of its neighborhoods in a radius $T/2$, the error of shape matching is proportional to the neighborhoods radius. Therefore, based on the coarse matching results, a step-by-step search is performed to search for the best affine transformation.

As shown in Algorithm 2, the template image Θ and the input image I are both downsampled with a higher resolution than the images used in Stage I. Then the candidate affine transformation net N_C is expanded to a net Q_C to find the best transformation named T_{Best}.

Algorithm 2 Stage II Fine Shape Matching

Input: the template image Θ, the input image I, the candidate affine transformation net N_C.

Output: the best affine transformation T_{Best}.

Step 1: Downsample the template image Θ and the input image I with a higher resolution than the images used in Stage I.

Step 2: Expand the candidate affine transformation net N_C to the net Q_C with a radius of [-5, 5] × [-5, 5].

Step 3: A step-by-step search is performed to find the best affine transformation named as T_{Best}.

Return: the best affine transformation T_{Best}.

---

4 Experiments and Analysis

A series of experiments are conducted to verify the alignment method proposed in this paper. The alignment system is established, and the calibration results containing the calibration for the camera and the rotation center of the rotation platform are illustrated in Sect. 4.2. Alignment experiments are carried out to verify the calibration and alignment accuracy in Sect. 4.3. Finally, to compare the binocular vision system with the monocular vision system, the contrast experiments are carried out in Sect. 4.4.

4.1 Alignment System

The alignment system is composed of a motion platform and a vision system. As shown in Fig. 9, the motion platform
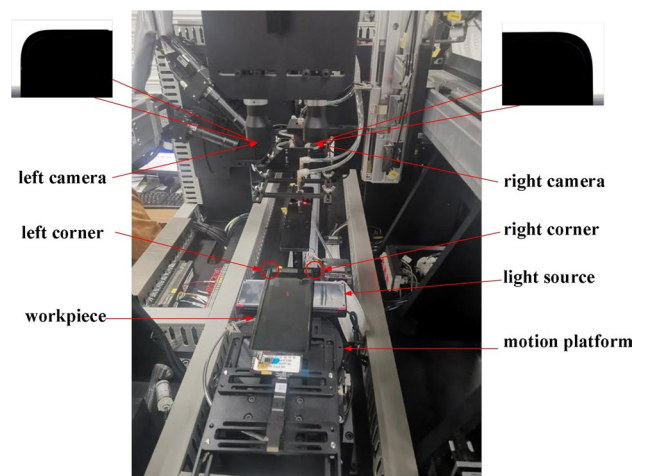

Fig. 9 Workstation

consists of a two-dimensional translation platform and a rotation platform. Then a binocular-based vision system with two telecentric lens cameras is established to measure the position and pose of the workpiece by capturing the images containing the feature points. The binocular vision system is mounted with optic axes orthography to the motion platform. Taking the alignment task of the mobile phone's cover glass as an example, the left and the right corners of the workpiece are taken as the feature points. Then, the images containing the left and the right corner of the workpiece are captured by camera 1 and camera 2, respectively.

In the binocular based vision system, two Basler alA3800-8gm GigE cameras (image size: 3840×2748 *pixel*; pixel size: $1.67\ \mu m \times 1.67\ \mu m$) are mounted orthogonal to the motion platform. The motion platform uses a two-dimensional translation motor LMP-20C20 and a rotation motor FOI170-Z10-A00-N01 from the LinkHou Corporation to compose the translation platform and the rotation platform. The repeatability and resolution of the translation motor LMP-20C20 are $\pm 2\ \mu m/\pm 2\ \mu m$ and $0.5\ \mu m/0.5\ \mu m$, respectively. The repeatability and absolute accuracy of the rotation motor FOI170-Z10-A00-N01 are $\pm 0.0398°$ and $\pm 0.398°$, respectively.

4.2 Calibration Result for the Vision System

4.2.1 Calibration Result for the Telecentric Lens Camera

According to the proposed calibration method given in Sect. 3.1, several pairs of points need to be obtained by the active movement of the motion platform to calibrate the left camera, the right camera, and the rotation center of the rotation platform. The movement amount is taken as the coordinates of the local world coordinate since the accuracy of the motion platform is high enough, and the local world coordinate system is established on the axis of the motion platform. Thus, the motor platform is controlled to move according to the commands $(\Delta x,\ \Delta y) = \{(-3,3),\ (0,3),\ (3,3),\ (3,0),\ (0,0),\ (-3,0),\ (-3,-3),\ (0,-3),\ (3,-3)\}(mm)$ and then the images containing the left and the right corner of the workpiece are captured.

Taking the coordinates of the corners into the aforementioned Eqs. (4)–(5) in Sect. 3.1, the calibration parameters for the left camera are obtained as $[\theta_{zL},\ \theta_{yL},\ \theta_{xL},\ p_{xL},\ p_{yL},$

$k_{L}] = [-1.20°,\ -10.23°,\ 181.63°,\ 16.07,\ 10.95,\ 92.36]$ with a pixel equivalent $10.82\ \mu m/pixel$. The calibration parameters for the right camera are obtained as $[\theta_{zR},\ \theta_{yR},\ \theta_{xR},\ p_{xR},$ $p_{yR},\ k_{R}] = [-1.20°,\ -9.46°,\ 180.87°,\ 21.33,\ 12.52,\ 92.02]$ with a pixel equivalent $10.86\ \mu m/pixel$. Therefore, the homography matrix between the image pixel coordinate systems $o_{L}u_{L}v_{L},\ o_{R}u_{R}v_{R}$ and the camera world coordinate systems $o_{WL}x_{WL}y_{WL}z_{WL},\ o_{WR}x_{WR}y_{WR}z_{WR}$, denoted as $M_{L},\ M_{R}$, respectively, are given by

$$M_{L} = \begin{bmatrix} 0.0909 & -0.0159 & 0.0165 & 1.4842 \\ -0.0019 & -0.0923 & 0.0026 & 1.0113 \\ 0 & 0 & 0 & 0.0924 \end{bmatrix}$$

$$M_{R} = \begin{bmatrix} 0.0907 & -0.0149 & 0.0151 & 1.9628 \\ -0.0019 & -0.0920 & 0.0014 & 1.1521 \\ 0 & 0 & 0 & 0.0920 \end{bmatrix}$$

4.2.2 Calibration Result for the Rotation Center

According to the angle-based constrained calibration method of the rotation center, the motion platform is rotated clockwise with the angle $\alpha = 3°$ three times to attain the local world coordinates of the left and the right corner, denoted as $p_{WLi}(x_{WLi},\ y_{WLi})$, $p_{WRi}(x_{WRi},\ y_{WRi})(i = 1,2,3)$. According to the Eqs. (14)–(15) proposed in Sect. 3.3, the coordinates of the rotation center in the view of the left and the right camera are obtained as $p_{WOL}(x_{WOL},\ y_{WOL}) = (34.80,\ -85.21)$ and $p_{WOR}(x_{WOR},\ y_{WOR}) = (-37.16,\ -84.49)$.

In addition, to evaluate the calibration and the alignment error of the angle-based constrained calibration method proposed in this paper and the least-squares-based calibration method, the contrast experiments are implemented in Sect. 4.4.

4.3 Verified Experiments for the Calibration Accuracy and the Alignment Accuracy

To verify the calibration accuracy and alignment accuracy, experiments were conducted with the workpiece is adsorbed on the motion platform in a random pose. Under

Table 1 Movements amount of the motion platform

k	Δx_w^*	Δy_w^*	$\Delta \theta^*$	k	Δx_w^*	Δy_w^*	$\Delta \theta^*$
1	0	0	0	21	0	0	− 3
2	1	1	0	22	1	1	1
3	1	− 1	0	23	1	− 1	1
4	1	2	0	24	1	1	− 1
5	− 1	2	0	25	− 1	− 1	1
6	− 1	− 3	0	26	− 1	− 1	− 1
7	2	− 3	0	27	− 1	2	1
8	2	3	0	28	− 1	2	− 1
9	− 2	3	0	29	2	2	1
10	− 2	− 1	0	30	2	2	− 1
11	− 1	− 1	0	31	− 2	2	− 1
12	0	0	1	32	1	1	2
13	0	0	1.5	33	1	1	− 2
14	0	0	2	34	− 1	1	2
15	0	0	2.5	35	− 1	1	− 2
16	0	0	3	36	− 1	2	2
17	0	0	− 1	37	− 1	2	− 2
18	0	0	− 1.5	38	2	2	2
19	0	0	− 2	39	2	2	− 2
20	0	0	− 2.5	40	− 2	2	2

(a)   **(b)**

(c)   **(d)**
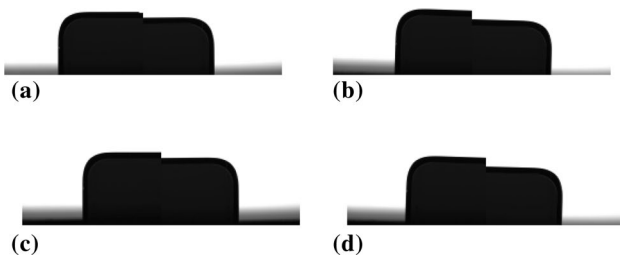
Fig. 10 Schematic of the data with the position and pose $\{\Delta x_w^*(mm), \Delta y_w^*(mm), \Delta \theta^*(°)\}$. **a** (0,0,0) **b** (0,0,2) **c** (2,3,0) **d** (− 1,1,2)

the condition of keeping the relative attitude of the workpiece and the motion platform unchanged, a large packet of data is obtained through the active motion of the motion platform, as shown in Table 1. As shown in Fig. 10, each slice of data is composed of images containing the left and the right corners of the workpiece. The left and the right corners of the workpiece are denoted as $p_{li}(u_{li}, v_{li})$ $p_{ri}(u_{ri}, v_{ri})$ in the image coordinate system $o_L u_L v_L$, $o_R u_R v_R$, and $p_{li}(x_{wli}, y_{wli})$ $p_{ri}(x_{wri}, y_{wri})$ in the world coordinate system $o_{WL} x_{WL} y_{WL} z_{WL}$, $o_{WR} x_{WR} y_{WR} z_{WR}$. Due to the high precision of the motion platform, these movements can be used to calculate the true value of the theoretical position deviation of the workpiece. Furthermore, to verify the method's robustness, the above experimental process is repeated after changing the relative attitude between the workpiece and the motion platform.

(Contain 10 translations, 10 rotations, 20 translation-and-rotations.)

### 4.3.1 Verified Experiments for the Telecentric Lens Camera's Calibration Accuracy

M_L and M_R are the calibration results of the telecentric lens cameras, which can be utilized to map the image coordinates of corners to the world coordinates. As shown in Table 1, if any two slices of data have the same value of $\Delta \theta^*$, they are considered to be parallel. Each pair of the data satisfying the parallel relationship is picked out, and the world coordinates of the corners are calculated using M_L and M_R. In these data pairs, the world coordinates of the left corner are denoted as (x_{wli}, y_{wli}), (x_{wlj}, y_{wlj}), $i, j \in 1,2,\ldots,C_n^2$, $i \neq j$. Then the difference between pair the data pair (i, j) is written as $d_{lk}(d_{xlk}, d_{ylk}) = (x_{wli} - x_{wlj}, y_{wli} - y_{wlj})$. The theoretical distance of the left corner between the data pair (i, j) is given by $d_{lk}^*(d_{xlk}^*, d_{ylk}^*) = (\Delta x_{wi} - \Delta x_{wj}, \Delta y_{wi} - \Delta y_{wj})$. Thus, the calibration error $\{d_{xlerror}, d_{ylerror}\}$ of the left camera can be obtained by

$$\begin{cases} d_{xl_{error}} = d_{xlk}^* - d_{xlk} \\ d_{yl_{error}} = d_{ylk}^* - d_{ylk} \end{cases} \tag{24}$$

The calibration error $d_{xrerror}$, $d_{yrerror}$ of the right camera is obtained in the same way. Seventy pairs of data's calibration error are given in Fig. 11, and the calibration accuracy for the left camera and right camera are within ± 0.020 mm. Furthermore, another data package is also obtained when
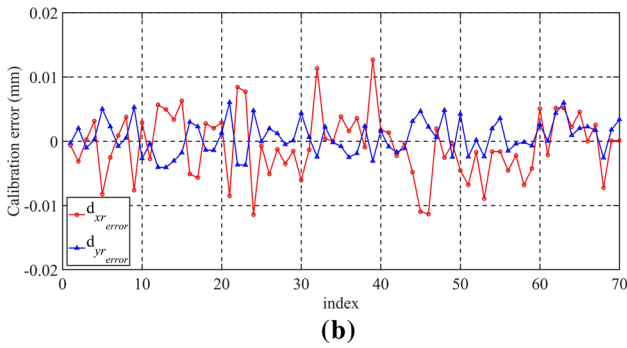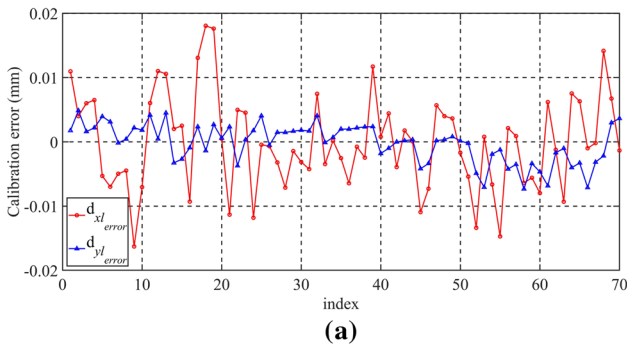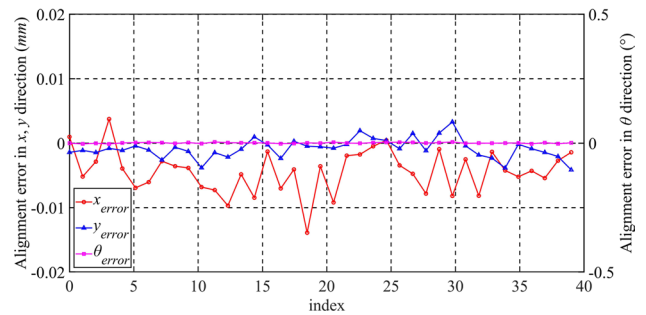
Fig. 11 The calibration error when the workpiece is adsorbed on the motion platform in a random attitude. **a** the calibration error {$d_{xlerror}$, $d_{ylerror}$} of the left camera. **b** the calibration error {$d_{xrerror}$, $d_{yrerror}$} of the right camera
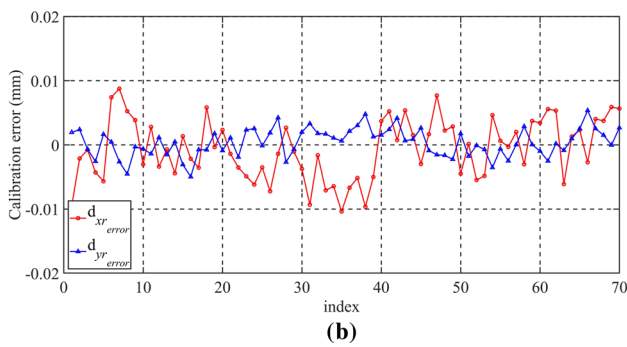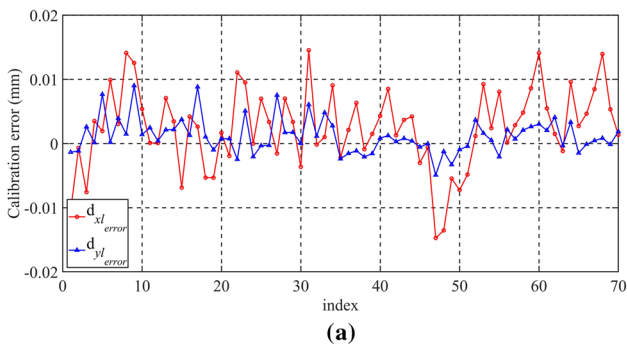
Fig. 12 The calibration error when the relative attitude between the workpiece and the motion platform is changed. **a** the calibration error {$d_{xlerror}$, $d_{ylerror}$} of the left camera. **b** the calibration error {$d_{xrerror}$, $d_{yrerror}$} of the right camera
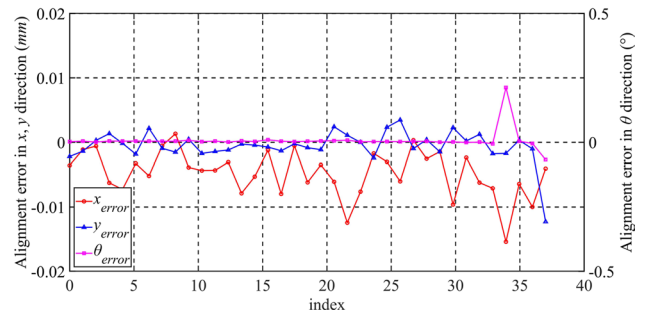
Fig. 13 The alignment error {x_{error}, y_{error}, θ_{error}} when the workpiece is adsorbed on the motion platform in a random attitude

Fig. 14 The alignment error {x_{error}, y_{error}, θ_{error}} when the relative attitude between the workpiece and the motion platform is changed

the relative attitude between the workpiece and the motion platform is changed. As shown in Fig. 12, the calibration accuracy for the left camera and right camera are also within ± 0.020 *mm*, which indicates that the calibration method is robust to the variation of the relative attitude between the workpiece and the motion platform.

4.3.2 Verified Experiments for the Rotation center's Calibration Accuracy

According to Eqs. (20)–(23), the rotation center calibration accuracy plays a crucial role in the alignment error. Thus, the rotation center calibration accuracy is evaluated by calculating the alignment error. Based on Table 1, the alignment error can be calculated by the difference between the alignment command and the movement amount. Specifically, the unified world coordinates of the left and the right corner are obtained according to Eq. (20) using their image coordinates. The target position is set as {x_w^*, y_w^*, θ^*} = {0, 0, 0}. Then, the alignment commands {Δx_w, Δy_w, $\Delta\theta$} can be calculated according to Eqs. (21)–(23). Take the movement amount of the motion platform as the standard value, the alignment error {x_{error},

y_{error}, θ_{error}} in the x, y, θ direction can be obtained as follows.

$$\begin{cases} x_{error} = \Delta x_W^* - \Delta x_W \\ y_{error} = \Delta y_W^* - \Delta y_W \\ \theta_{error} = \Delta \theta_W^* - \Delta \theta \end{cases} \qquad (25)$$

where {Δx_W^*, Δy_W^*, $\Delta \theta^*$} is the movement amount of the motion platform.

Forty slices of data's alignment errors are given in Fig. 13. The alignment error x_{error}, y_{error} is attained within ± 0.020 mm, and the θ_{error} is attained within $\pm 0.25°$. Meanwhile, when the relative attitude between the workpiece and the motion platform is changed, the alignment errors of another package of data are also obtained, as shown in Fig. 14. The alignment experiments results indicate that the achievable calibration accuracy is reasonable. In addition, the average calculating time for an alignment command is only 20 ms, which can fully meet the real-time requirements of the industrial application.

4.4 Contrast Experiments

A series of contrast experiments are carried out to prove the advantage of the method proposed in this paper. Firstly, the performance of the two mentioned calibration methods of the rotation center are compared, which are the least-squares-based calibration method and the angle constraint-based calibration method. Secondly, to verify whether a multi-vision system would improve the alignment accuracy, we implement a contrast experiment to accomplish the alignment task based on a binocular vision system and a monocular vision system, respectively.

4.4.1 Contrast Experiment for the Calibration Method of the Rotation Center

According to Eqs. (20)–(23), the calibration accuracy of the rotation center directly influences the alignment error. Therefore, comparing the alignment accuracy makes it possible to evaluate which algorithm has better higher accuracy. Specifically, to compare the least-squares-based calibration method and the angle constraint-based calibration method, their calibration results are taken as the rotation centers p_{WOL}
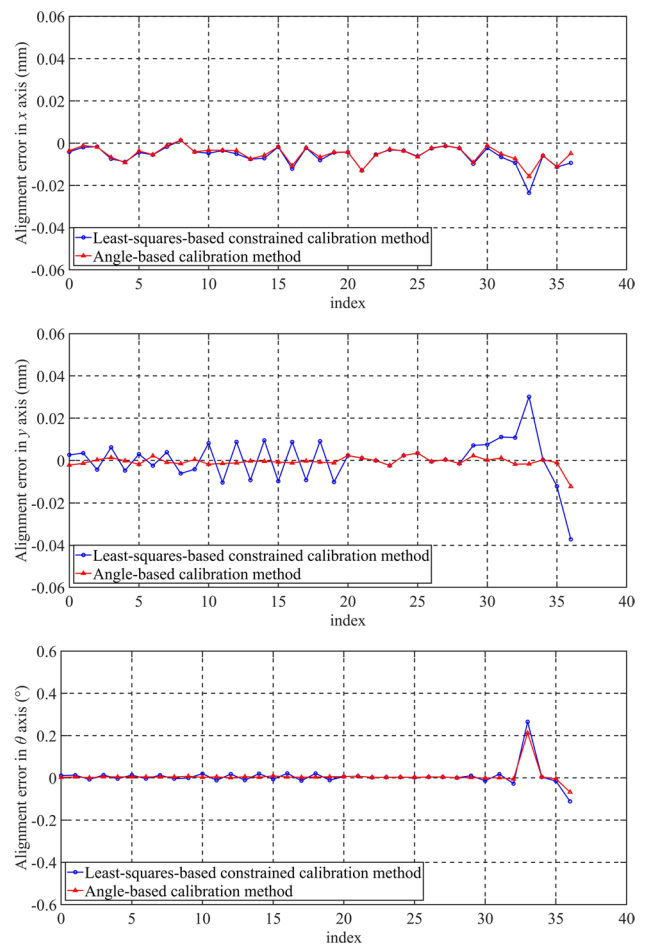

Fig. 15 The alignment errors of the two calibration methods of the rotation center

and p_{WOR}, respectively. Similarly to the verified experiments illustrated in 4.3.2, forty slices of data's alignment errors of two calibration method are given in Fig. 15. Furthermore, the mean absolute error (MAE) and the range of the alignment errors are obtained to characterize the accuracy of the two calibration methods, respectively, as shown in Table 2. The experiment results show that the angle constraint-based calibration method proposed by this paper outperforms the least-squares-based calibration method in alignment accuracy.

Table 2 Alignment errors with calibration method of the rotation center

Method	MAE-x	MAE-y	MAE-θ	R-x	R-y	R-θ
Least-squares based	0.005	0.007	0.011	0.024	0.067	0.038
Angle constraint (*Ours*)	0.005	0.001	0.011	0.017	0.015	0.028
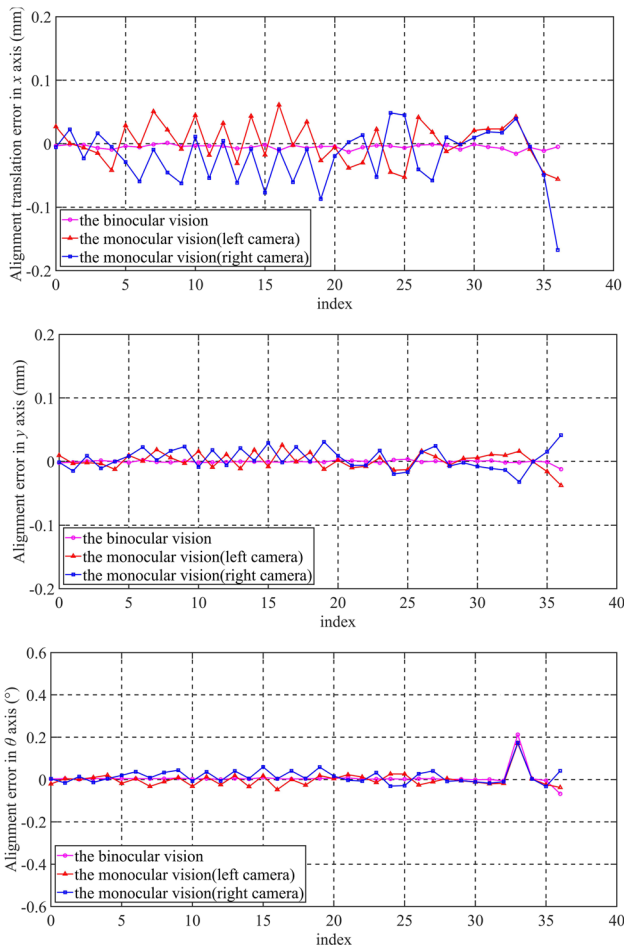
Fig. 16 The alignment error in the x, y, θ direction for the different types of vision system

4.4.2 Contrast Experiment for the Vision System

To explore the influence of the vision system on alignment accuracy, different types of vision system are utilized to perform the alignment task. A binocular vision system composed of left and the right cameras is utilized to perform the alignment task. Then, a monocular vision system is formed by the left camera or the right camera, respectively, to perform the same alignment task.

Figure 16 shows that the binocular vision-based system performs better than the monocular vision-based system. Furthermore, the MAE and the range of the alignment error in Table 3 also indicate that the binocular vision-based system achieves higher accuracy than the monocular vision-based system. The main reason lies in that the monocular vision-based system calculates the angle of the workpiece using the region of interest from a single image, whereas the binocular vision-based system utilizes the left and the right images. In other words, it means that the binocular vision-based alignment method can utilize more image information and achieve higher alignment accuracy, especially for the alignment task with a larger workpiece size.

5 Conclusions

A high precision and fast alignment method based on binocular vision is proposed to accomplish the alignment task in one operation. A calibration method for the telecentric lens camera based on an improved nonlinear damped least-squares method is proposed to speed up the convergence of the calibration process. Meanwhile, an angle constraint-based calibration method for the platform's rotation center is proposed to pursue higher precision. Furthermore, to detect the feature point more robust, a two-stage feature point detection method based on shape matching is presented. Experiments conducted on an alignment system demonstrate the effectiveness of the proposed methods. The alignment error is within $\pm 0.020\ mm$ and the time taken to calculate the alignment command is less than 20 ms. Future work will focus on expanding proposed methods to adapt four cameras for high-precision alignment tasks of larger size workpieces.

Table 3 Alignment errors with different types of vision system

| Vision system	MAE-x	MAE-y	MAE-θ	R-x	R-y	R-θ
Monocular (left)	0.027	0.010	0.026	0.117	0.064	0.218
Monocular (right)	0.034	0.013	0.022	0.216	0.074	0.206
Binocular (*Ours*)	0.005	0.005	0.011	0.017	0.016	0.280

References

1. Chen, Z., Zhou, D., Liao, H., & Zhang, X. (2016). Precision alignment of optical fibers based on telecentric stereo microvision. *IEEE/ASME Transactions on Mechatronics, 21*(4), 1924–1934.

2. Shen, F., Wu, W., Yu, D., Xu, D., & Cao, Z. (2015). High-precision automated 3-D assembly with attitude adjustment performed by LMTI and vision-based control. *IEEE/ASME Transactions on Mechatronics, 20*(4), 1777–1789.

3. Tamadazte, B., Piat, N. L. F., & Dembélé, S. (2011). Robotic micromanipulation and microassembly using monoview and multiscale visual servoing. *IEEE/ASME Transactions on Mechatronics, 16*(2), 277–287.

4. Gu, Q., Aoyama, T., Takaki, T., & Ishii, I. (2015). Simultaneous vision-based shape and motion analysis of cells fast-flowing in a microchannel. *IEEE Transactions on Automation Science and Engineering, 12*(1), 204–215.

5. Kyriakoulis, N., & Gasteratos, A. (2010). Color-based monocular visuoinertial 3-D pose estimation of a volant robot. *IEEE Transactions on Instrumentation and Measurement, 59*(10), 2706–2715.

6. Li, C., & Gao, X. (2018). Adaptive contour feature and color feature fusion for monocular textureless 3D object tracking. *IEEE Access, 6*, 30473–30482.

7. Ma, Y., Liu, X., & Xu, D. (2020). Precision pose measurement of an object with flange based on shadow distribution. *IEEE Transactions on Instrumentation and Measurement, 69*(5), 2003–2015.

8. Shen, F., Qin, F., Zhang, Z., Xu, D., Zhang, J., & Wu, W. (2021). Automated pose measurement method based on multivision and sensor collaboration for slice microdevice. *IEEE Transactions on Industrial Electronics, 68*(1), 488–498.

9. Lins, R. G., Givigi, S. N., & Kurka, P. R. G. (2015). Vision-based measurement for localization of objects in 3-D for robotic applications. *IEEE Transactions on Instrumentation and Measurement, 64*(11), 2950–2958.

10. Liu, S., Xu, D., Liu, F., Zhang, D., & Zhang, Z. (2016). Relative pose estimation for alignment of long cylindrical components based on microscopic vision. *IEEE/ASME Transactions on Mechatronics, 21*(3), 1388–1398.

11. Ma, Y., Liu, X., Zhang, J., Xu, D., Zhang, D., & Wu, W. (2020). Robotic grasping and alignment for small size components assembly based on visual servoing. *The International Journal of Advanced Manufacturing Technology, 106*(11), 4827–4843.

12. Kwon, S., Jeong, H., & Hwang, J. (2012). Kalman filter-based coarse-to-fine control for display visual alignment systems. *IEEE Transactions on Automation Science and Engineering, 9*(3), 621–628.

13. Wang, P., Shen, S., Lu, H., & Shen, Y. (2019). Precise watch-hand alignment under disturbance condition by microrobotic system. *IEEE Transactions on Automation Science and Engineering, 16*(1), 278–285.

14. Golnabi, H., & Asadpour, A. (2007). Design and application of industrial machine vision systems". *Robotics and Computer-Integrated Manufacturing, 23*(6), 630–637.

15. Transtrum, M. K., Sethna, J. P. (2012) Improvements to the Levenberg- Marquardt algorithm for nonlinear least-squares minimization. arXiv preprint arXiv

16. Kåsa, I. (1976). A circle fitting procedure and its error analysis. *IEEE Transactions on Instrumentation and Measurement, 25*(1), 8–14.

17. Zhu, J., Li, X. F., Tan, W. B., Xiang, H. B., & Chen, C. (2009). Measurement of short arc based on center constraint least square circle fitting". *Optics and Precision Engineering, 17*(10), 2486–2492.

18. Korman, S., Reichman, D., Tsur, G., Avidan, S. Fast-match: Fast affine template matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2331–2338).

19. Hinterstoisser, S., Cagniart, C., Ilic, S., Sturm, P., Navab, N., Fua, P., & Lepetit, V. (2012). Gradient response maps for real-time detection of textureless objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(5), 876–888.

Han Gao received the B.Sc. degree from Jilin University of China, Changchun, China, in 2019. He is currently a master candidate in the Research Center of Precision Sensing and Control, IACAS. His research interests include neural networks, computer vision and image processing.
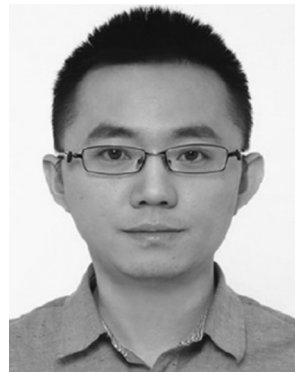
Fei Shen received the B.Sc. and M.Sc. degrees from Xidian University, Xi'an, China, and Beijing Institute of Technology, Beijing, China, in 2007 and 2009, respectively, and the Ph.D. degree in control science and engineering from IACAS, Beijing, China, in 2012. He is currently a Professor in the Research Center of Precision Sensing and Control, IACAS. His research interests include visual inspection, robot vision control and micro-assembly.

Feng Zhang received the B.Sc. degree in control engineering from Qingdao University, Qingdao, China, in 2006, and the Ph.D. degree in control science and engineering from IACAS, Beijing, China, in 2012. He is currently a Vice Professor in the Research Center of Precision Sensing and Control, IACAS. His research interests include robot control, robot vision control and micro-assembly.

Zhengtao Zhang received the B.Sc. and M.Sc. degrees from China University of Petroleum, Dongying, China, and Beijing Institute of Technology, Beijing, China, in 2004 and 2007, respectively, and the Ph.D. degree in control science and engineering from IACAS, Beijing, China, in 2010. He is currently a Professor in the Research Center of Precision Sensing and Control, IACAS. His research interests include visual measurement, micro-assembly and automation.