High-precision robotic assembly system using three-dimensional vision

Shaohua Yan[1,2], Xian Tao[1,2] and **De Xu[1,2]**

Abstract

The design of a high-precision robot assembly system is a great challenge. In this article, a robotic assembly system is developed to assemble two components with six degree-of-freedoms in three-dimensional space. It consists of two manipulators, a structured light camera which is mounted on the end-effector aside component A to measure the pose of component B. Firstly, the features of irregular components are extracted based on U-NET network training with few labeled images. Secondly, an algorithm is proposed to calculate the pose of component B based on the image features and the corresponding three-dimensional coordinates on its ellipse surface. Thirdly, the six errors including two position errors and one orientation error in image space, and one position error and two orientation errors in Cartesian space are computed to control the motions of component A to align with component B. The hybrid visual servoing method is used in the control system. The experimental results verify the effectiveness of the designed system.

Introduction

With the development of technology, the demand for high-precision assembly in industrial manufacturing and space exploration is increasing.[1–3] Industrial assembly devices are generally divided into two categories. One is the specific translation and rotation mechanism.[4,5] For example, Luo et al.[4] used a linear drive mechanism for precision threading operations. The translation error and rotation error of the platform reached 3 μm and 0.005°, respectively. Yu et al.[5] used the feature constraint relationship between components to control translation and rotation devices completing component assembly simulation. However, the working range of specific translation and rotation mechanisms is small, and its flexibility is low. The other is based on a general manipulator.[6,7] For example, Wang et al.[6] added an elastic displacement device to the manipulator to achieve peg-in-hole assembly, which improved the success rate of each assembly. Meng et al.[7] realized precise robot assembly for large-scale spacecraft components based on computer-aided design models of aircraft components and key geometric features located by ranging sensors and binocular vision. Generally, a manipulator has six degree-of-freedoms (DOFs). Therefore, it is very helpful for manipulator-based assembly systems to realize high-precision assembly with six DOFs in three-dimensional (3D) space.

In the robotic assembly system, the target pose is usually measured with vision-based methods.[8,9] Generally, the

[1] Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

Corresponding author:
Xian Tao, School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100080, China.
Email: taoxian2013@ia.ac.cn

point feature, line feature, and circle feature are employed in the pose estimation methods. For example, in Liu et al.,[10] the end position of the dispensing needle was obtained through the point feature and then the precision dispensing operation was completed. In Liu et al.,[11] line features were used to measure the pose of a long cylindrical component. In Liu and Xu,[12] a fast and effective circle detection algorithm was proposed for target position estimation. However, the above pose estimation methods require two or three cameras placed in different directions. Sun et al.[13] measured the target pose with a camera based on the projection relationship between the circle and the ellipse. But the accuracy of this method relies on the ellipse fitting. Another kind of target pose measurement method with the structured light camera is becoming popular.[14–16] For example, Kim et al.[14] accurately estimated the surface normal vector of the target based on a structured light camera and then completed the object-grasping task. In Satorres et al.,[15] the relative position relationship between the manipulator and the object was obtained through the 3D data in the 3D camera. Litvak et al.[16] assembled randomly distributed components based on the depth camera and convolutional neural network, and the success rate reached 91%. Therefore, pose measurement based on a structured light camera is a better choice.

Visual servoing methods are very popular in many applications including automatic assembly systems.[17,18] They are classified as image-based visual servoing,[19] position-based visual servoing,[20] and hybrid visual servoing method.[21] Xu et al.[19] proposed an image-based visual servoing method, in which point features and line features are used for position control and attitude control, respectively. Image-based visual servoing has certain robustness to camera calibration errors and robot model errors. Through comparative experiments on position-based and image-based visual servoing systems, Peng et al.[20] found that position-based visual servoing has a faster convergence speed. What's more, some advanced control methods for tracking control of mechanical servo systems help improve convergence speed. For example, Deng and Yao[22] designed a high-performance tracking controller without velocity measurement in electrohydraulic servomechanisms, which achieves asymptotic tracking performance when facing time-invariant modeling uncertainties. Aiming at mechanical servosystems with mismatched uncertainties, Deng and Yao[23] proposed a novel recursive robust integral of the sign of the error control method, which achieves excellent asymptotic tracking performance. Therefore, it is necessary to combine the advantages of image-based visual servoing and position-based visual servoing methods to realize the precision assembly of two components.

The purpose of this article is to achieve precise assembly of irregular components. A robotic assembly system is developed to assemble two components with six DOFs in 3D space, which consists of two manipulators and a

structured light camera. Image-space information and 3D space information acquired by structured light cameras are effectively combined to measure the pose of component B. Considering the advantages of image-based and position-based visual servoing methods, this article proposes a hybrid visual servoing method with higher convergence speed and accuracy. The manipulators can control the components of different initial positions and postures for automatic assembly. The main contributions of this paper are as follows:

1. A robotic assembly system with two manipulators is developed to assemble two components with six DOFs in 3D space. The hybrid visual servoing method combining errors in Cartesian space and image space is used in the control system.
2. A feature extraction algorithm for the images of irregular components is proposed, which is based on U-NET network training with few labeled images.
3. The pose of component B is calculated from the image features and the corresponding 3D coordinates on its ellipse surface.

The rest of this article is organized as follows. The first part describes the assembly task and system. Secondly, an image feature extraction and pose measurement method is proposed. Then presents a hybrid visual servoing method to align the two components. The details of the automated assembly process are also introduced. The experiments and results with the proposed assembly method are given. Finally, this article is concluded.

Assembly task and system

Assembly task

The two components to be assembled are shown in Figure 1. They are metal connectors with an outer diameter of about 43 mm, which are divided into component A and component B. As shown in Figure 1(a), the left side is component A and the right side is component B. There are five groove areas on the inner side of component B, as shown in Figure 1(b). The positions and sizes of the grooves are unevenly distributed. Correspondingly, there are five protruding areas on the upper surface of component A, as shown in Figure 1(c).

When assembling, it is necessary to align the groove area of component B to the protruding area of component A with six DOFs, including 3D position and three-direction angles. Our task is to realize the precise assembly of these two components.

Assembly system

The automated precision assembly system is designed as given in Figure 2. Manipulator 1 is a seven-DOF robot with a clamping device and component A connected to it. A
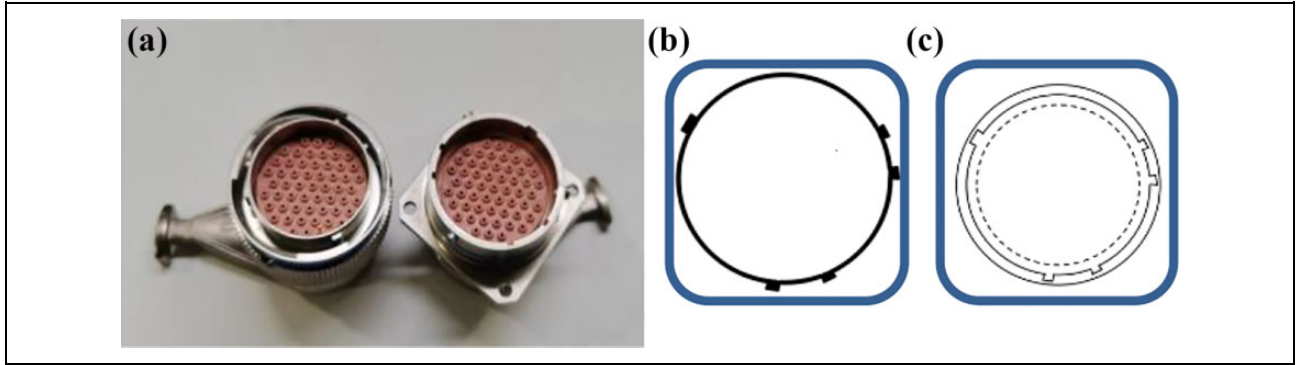
Figure 1. Components: (a) components A and B, (b) component A and its surface structure, and (c) component B and its surface structure.
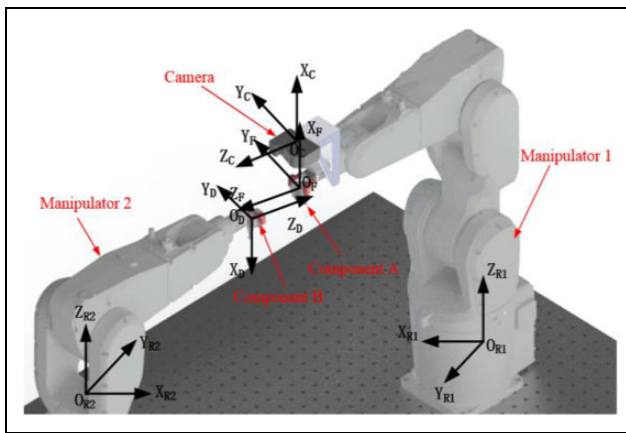

Figure 2. Assembly system configuration.

structured light camera is fixed at the end of manipulator 1. Manipulator 2 is a universal robot (UR3) with a gripping device and component B connected to it.

Manipulator 1 can translate along and rotate around the X, Y, and Z axes to align component A to component B. The poses of manipulator 1 and manipulator 2 can be adjusted to initialize the pose of component B in the structured light camera. The computer can control the entire assembly process including image capture with the camera, image processing, feature extraction, pose estimation, and alignment and insertion of the two components.

The coordinates are established as shown in Figure 2. $O_{R1}X_{R1}Y_{R1}Z_{R1}$ is the base frame of manipulation 1, $O_{R2}X_{R2}Y_{R2}Z_{R2}$ is the base frame of manipulation 2, $O_DX_DY_DZ_D$ is the end-effector frame of manipulation 2, $O_CX_CY_CZ_C$ is the camera frame, and $O_FX_FY_FZ_F$ is the end-effector frame of manipulation 1. The camera is carefully adjusted so that the axes of the camera frame are as parallel to the axes of the end-effector frame of manipulation 1 as possible.

Image feature extraction

Elliptic ring region extraction

Figure 3 shows the image of component B captured by the structured light camera. To get the current pose of component B in the camera frame, its inherent features such as ring circles should be extracted. As shown in Figure 3(a), there is noise in the gray image of component B, which leads to the disturbance on the edges.

There will be a large error in detecting the ring contour of the component through edge detection and ellipse fitting. Another method is to obtain the ring area via threshold segmentation.

But the gray value of the ring area is not evenly distributed due to the influence of light. Therefore, it is difficult to accurately segment the ring area with threshold segmentation.

Therefore, this article uses data labeling and deep learning methods to solve the problem of inaccurate feature extraction. As shown in Figure 3(b), the elliptical ring area on the surface of component B is marked, the outside of the ring is an ellipse, and the inside is an ellipse containing the edge of the groove. A U-NET network is designed and its structure diagram is shown in Figure 4. It includes a contraction path for capturing semantics and an asymmetrical expansion path for precise positioning. The contracted path part consists of four convolutional layers and pooling layers for down-sampling, and the extended path part consists of four deconvolutional layers and convolutional layers for up-sampling. This U-NET network is trained with the labeled data. Then it is used to segment the ring area from the image of component B. As shown in Figure 3 (c), the elliptical ring area containing the groove information is accurately extracted.

Groove feature extraction

General methods cannot effectively detect the groove features on the ring. Therefore, the inner and outer ellipses are combined to detect the groove feature.

As shown in Figure 5(a), the contour of the elliptical ring area is output by the U-NET network. The two contours containing most edge points of the inner and outer sides are considered as the inner ellipse and the outer ellipse of the ring. Then the least square method is used to fit the inner ellipse parameter equations (1) and outer ellipse parameter equations (2), respectively. The ellipse fitting result is shown in Figure 5(b)
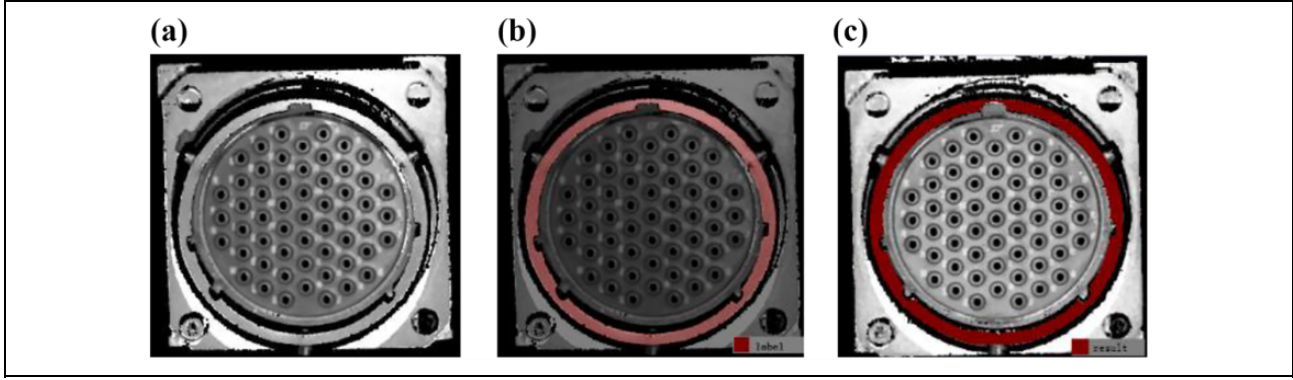
Figure 3. Image of component B: (a) original image, (b) image with manually marked ring area, and (c) image with segmented ring area. In (b) and (c), the ring area is indicated with red color.
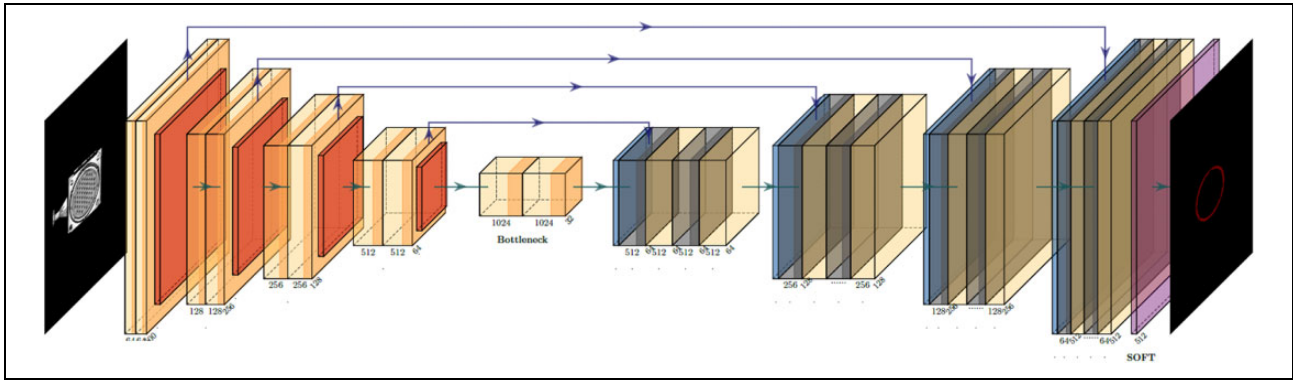

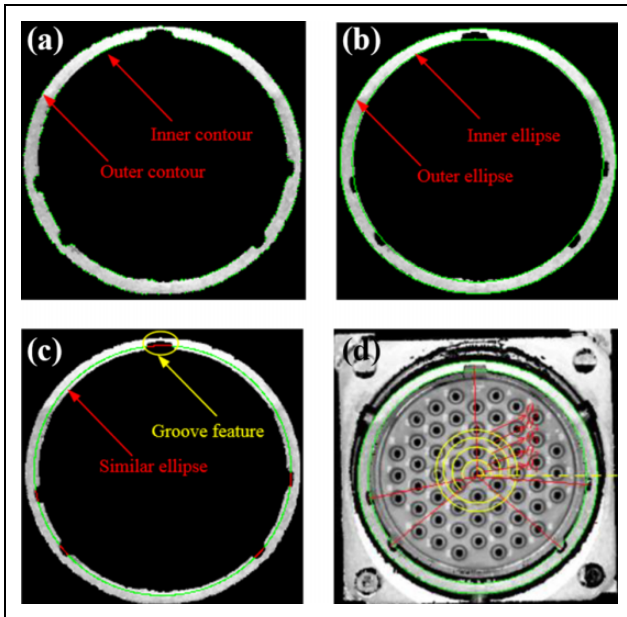**Figure 4.** U-NET network structure diagram.

Figure 5. Groove feature extraction process: (a) contours detection, (b) ellipse fitting, (c) groove feature extraction, and (d) image with extracted groove features.

$$\begin{cases} u_{\text{in}} = u_0 + a_{\text{in}}\cos\theta_0\cos\theta/2 - b_{\text{in}}\sin\theta_0\sin\theta/2 \\ v_{\text{in}} = v_0 + a_{\text{in}}\sin\theta_0\cos\theta/2 + b_{\text{in}}\cos\theta_0\sin\theta/2 \end{cases} \quad (1)$$

where (u_0, v_0) is the pixel coordinate value of the center point of the ellipse, $(u_{\text{in}}, v_{\text{in}})$ is the pixel coordinate of the point on the inner ellipse, a_{in} and b_{in} are the long and short axis lengths of the inner ellipse, θ_0 represents the initial angle of the ellipse, and $\theta \in (0, 2\pi)$ is the parameter variable

$$\begin{cases} u_{\text{out}} = u_0 + a_{\text{out}}\cos\theta_0\cos\theta/2 - b_{\text{out}}\sin\theta_0\sin\theta/2 \\ v_{\text{out}} = v_0 + a_{\text{out}}\sin\theta_0\cos\theta/2 + b_{\text{out}}\cos\theta_0\sin\theta/2 \end{cases} \quad (2)$$

where (u_{out}, v_{out}) is the pixel coordinate of the point on the outer ellipse and a_{out} and b_{out} are the long and short axis lengths of the outer ellipse.

According to the inner and outer ellipse equations, similar ellipse parameter equations (3) passing through the groove area are obtained

$$\begin{cases} u_e = u_0 + (a_{\text{in}} + k(a_{\text{out}} - a_{\text{in}}))\cos\theta_0\cos\theta/2 \\ \qquad - (b_{\text{in}} + k(b_{\text{out}} - b_{\text{in}}))\sin\theta_0\sin\theta/2 \\ v_e = v_0 + (a_{\text{in}} + k(a_{\text{out}} - a_{\text{in}}))\sin\theta_0\cos\theta/2 \\ \qquad + (b_{\text{in}} + k(b_{\text{out}} - b_{\text{in}}))\cos\theta_0\sin\theta/2 \end{cases} \quad (3)$$

where (u_e, v_e) is the pixel coordinate of the point on the similar ellipse and $k \in (0, 1)$ represents the coefficient of similar ellipse close to the outer ellipse.

The parameter angle θ in the similar ellipse equation (3) is gradually increased to find the continuous point along the similar ellipse where the pixel value is significantly different from the ring area. The corresponding parameter angle set $(\theta_{11}, \theta_{12}, \cdots \theta_{1k})$ is recorded. After traversing $\theta \in (0, 2\pi)$, we can get five parameter angle sets

$$\{(\theta_{11}, \theta_{11}, \cdots \theta_{1k_1}), (\theta_{21}, \theta_{21}, \cdots \theta_{2k_2}), \cdots (\theta_{51}, \theta_{51}, \cdots \theta_{5k_5})\} \quad (4)$$

Feature extraction results via searching along similar ellipses are shown in Figure 5(c). Finally, the average $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ of five parameter angle sets is considered as the angle of each groove. The results of feature extraction on the original image are shown in Figure 5(d).

Automatic assembly

Automatic assembly is divided into three parts, namely the desired image capture stage, the camera alignment stage, and the component insertion stage. The whole assembly process is given in Figure 6.

The desired image capture stage is mainly to obtain the desired image and the displacement of the manipulator between the alignment and insertion positions via one assembly manually controlled. The desired image features are extracted from the desired image. During the camera alignment stage, the features of component B in image space and Cartesian space are acquired. A hybrid visual servoing control method is designed for precise alignment. In the component insertion stage, component A is translated by displacements D_1 and D_2. Then component A is inserted into component B.

Desired image capture stage

As shown in stage A in Figure 6, manipulator 1 is manually controlled to complete one assembly. Manipulator 1 is translated the given displacement D_1 along the z-axis in its end-effector frame to move component A away from component B. Manipulator 1 is translated along the x-axis in its end-effector frame until the camera can capture the image of component B. The displacement along the x-axis is recorded as D_2. This state is called the camera alignment state.

The images captured in the camera alignment state are considered as the desired image. The elliptical ring area containing the groove of the desired image is extracted by trained U-NET network. The image coordinates $\{(u_1, v_1), (u_2, v_2), \cdots (u_n, v_n)\}$ of sampled points in the ellipse ring area are obtained. Corresponding, the 3D coordinates $\{(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_n, y_n, z_n)\}$ in the camera frame are recorded. The random sample
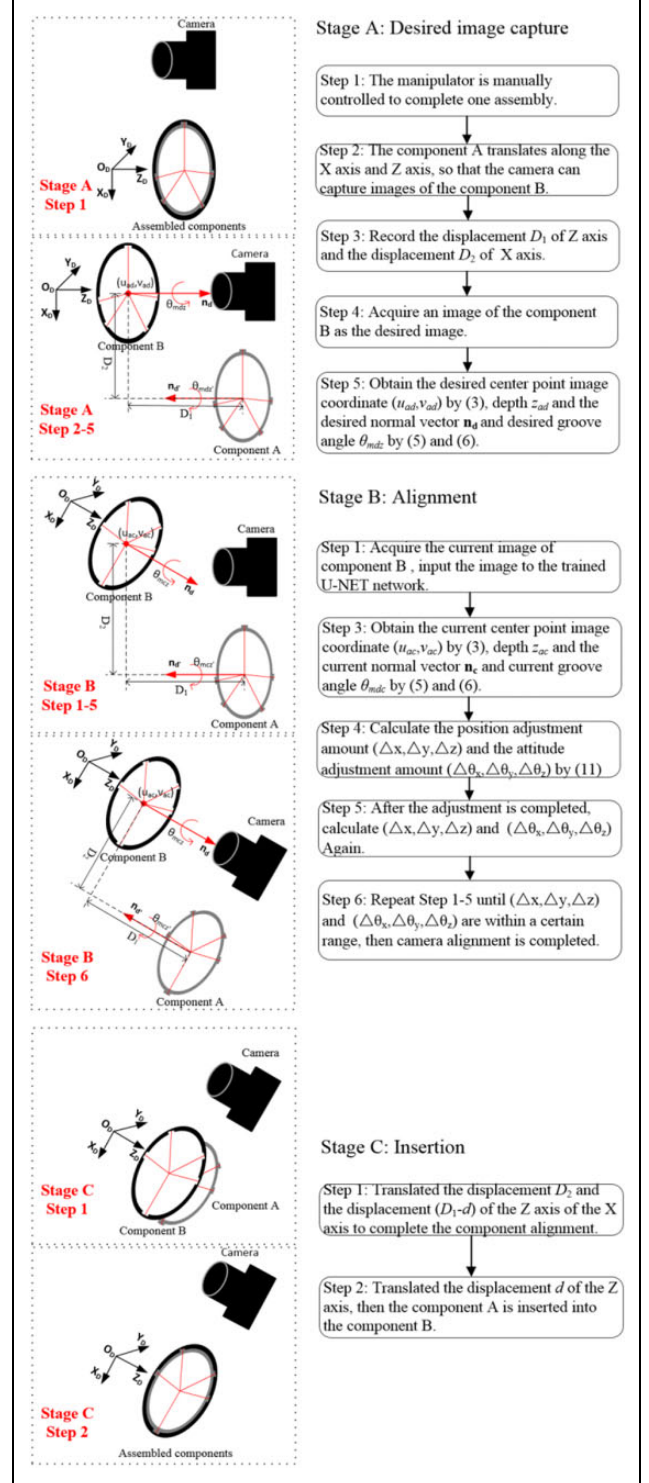

Figure 6. The program flow chart of the assembly procedure.

consensus algorithm is used to fit the ring area plane (4) of component B

$$a_d x + b_d y + c_d z + e_d = 0 \quad (5)$$

where a_d, b_d, c_d, and e_d are the parameters of the fitting plane.
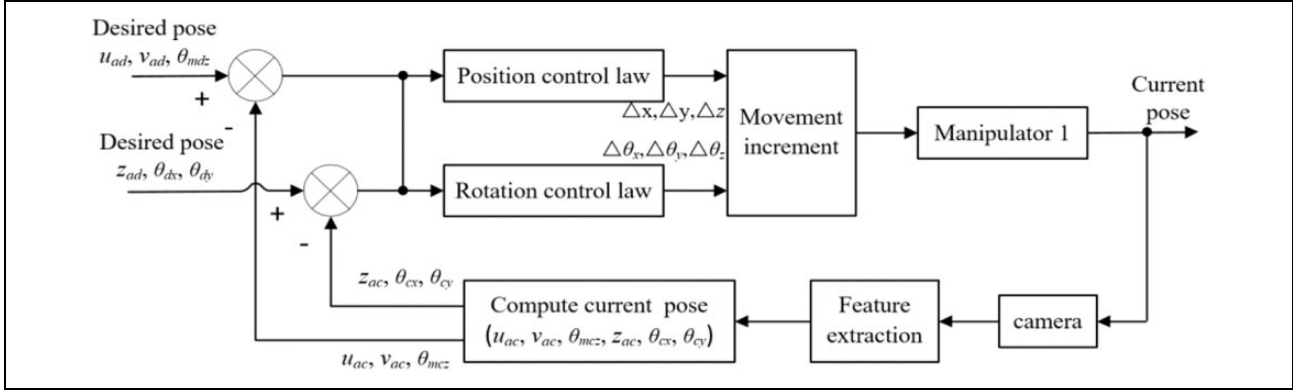
Figure 7. Block diagram of automatic control system.

The desired normal vector $[a_d, b_d, c_d]^T$ is obtained. The desired normal vector is normalized to the desired unit normal vector $\mathbf{n}_d$

$$\mathbf{n}_d = \frac{[a_d \quad b_d \quad c_d]^T}{\sqrt{a_d^2 + b_d^2 + c_d^2}} = [n_{dx} \quad n_{dy} \quad n_{dz}]^T \quad (6)$$

The desired posture angle θ_{dx} and θ_{dy} are calculated with the desired plane unit normal vector by formula (6). Posture angle θ_{mdz} is an angle sequence, which contains groove angle information. It is obtained by the above groove feature extraction algorithm

$$\begin{cases} \theta_{dx} = \mathrm{asin}\, n_{dx} \\ \theta_{dy} = \mathrm{asin}\, n_{dy} \\ \theta_{mcz} = [\theta_{c1}, \theta_{c2}, \theta_{c3}, \theta_{c4}, \theta_{c5}]^T \end{cases} \quad (7)$$

The desired center point image coordinate $P_d = (u_{ad}, v_{ad})$ of component B is obtained through ellipse fitting. Correspondingly, the 3D coordinate $P_{ad} = (x_{ad}, y_{ad}, z_{ad})$ in the camera frame is read from the 3D camera.

In this way, the desired features P_d, θ_{dx}, and θ_{dy} are acquired in image space, and the desired features $\mathbf{n}_d$, P_{ad}, and θ_{mdz} are acquired in Cartesian space.

Camera alignment stage

The current image of component B is acquired in real time. According to the above method of feature extraction, the current features $P_c = (u_{ac}, v_{ac})$, θ_{cx}, and θ_{cy} are acquired from the current image, and the current features $\mathbf{n}_c = [a_c, \quad b_c, \quad c_c]^T$, $P_{ac} = (x_{ac}, y_{ac}, z_{ac})$, and $\theta_{mcz} = [\theta_{c1}, \theta_{c2}, \theta_{c3}, \theta_{c4}, \theta_{c5}]^T$ are acquired in Cartesian space, as described in the "Desired Image Capture Stage" section.

A hybrid visual servoing control system is designed, in which the features from image space and Cartesian space are combined to realize the alignment between component B and camera. The block diagram of the hybrid visual servoing automatic control system is shown in Figure 7.

The pose of the end-effector of manipulator 1 is adjusted in its end-effector frame according to formula (11). The features from image space are used to control translations along the x-axis and y-axis and rotation around the z-axis. The features from Cartesian space are used to control translation of the end-effector along the z-axis and rotations around the x-axis and y-axis

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta\theta_x \\ \Delta\theta_y \\ \Delta\theta_z \end{bmatrix} = \begin{bmatrix} k_1(u_{ac} - u_{ad}) \\ k_1(v_{ac} - v_{ad}) \\ k_2(z_{ac} - z_{ad}) \\ k_2(\theta_{cx} - \theta_{dx}) \\ k_2(\theta_{cy} - \theta_{dy}) \\ k_2\Delta\theta_{mz} \end{bmatrix} \quad (8)$$

where k_1 and k_2 are coefficients and $\Delta\theta_{mz}$ is the best angle error calculated by $\theta_{mcz} = [\theta_{c1}, \theta_{c2}, \theta_{c3}, \theta_{c4}, \theta_{c5}]^T$ and $\theta_{mdz} = [\theta_{d1}, \theta_{d2}, \theta_{d3}, \theta_{d4}, \theta_{d5}]^T$.

As shown in stage B in Figure 6, the camera alignment state is achieved after hybrid visual servoing control. At this point, the errors between the current pose and desired pose approach 0. The displacement between component A and component B along the z-axis in the end-effector frame of manipulator 1 is D_1. The displacement between component A and component B along the x-axis in the end-effector frame of manipulator 1 is D_2.

Component insertion stage

In the component insertion stage, component alignment and component insertion are completed. At first, as shown in stage C in Figure 6, component A is translated the displacement $(D_1 - d)$ along the z-axis and the displacement D_2 along the x-axis in the end-effector frame of manipulation 1, where d is a small displacement. After ensuring the safety of assembly, component A is translated the displacement d along the z-axis in the end-effector frame of manipulation 1. Then component A is inserted into component B. The entire assembly is completed precisely and efficiently.

Experiments and results

Experiment system

An experiment system was established according to the scheme given in the "Assembly System" section, as shown in Figure 8. In this experiment system, there were two manipulators including one seven-DOF robotic arm and one six-DOF manipulator. Manipulator 1 had a clamping device and component A connected to it. Manipulator 2 was a UR3 (universal robots company) manipulator with a gripping device and component B connected to it. A structured light camera was fixed at the end of manipulator 1. The structured light camera was LMI Gocator3210 (LMI technologies company) binocular snapshot sensor. The resolution of the camera in the x-axis and y-axis directions is 60–90 μm, the field of view is 71×98 mm–100×154 mm, and the working distance is 164 mm.
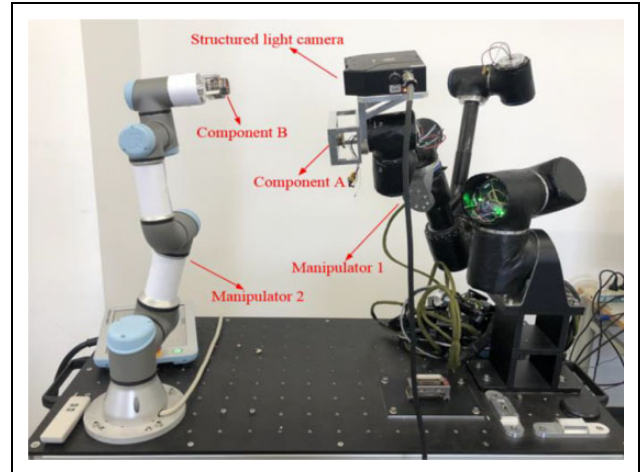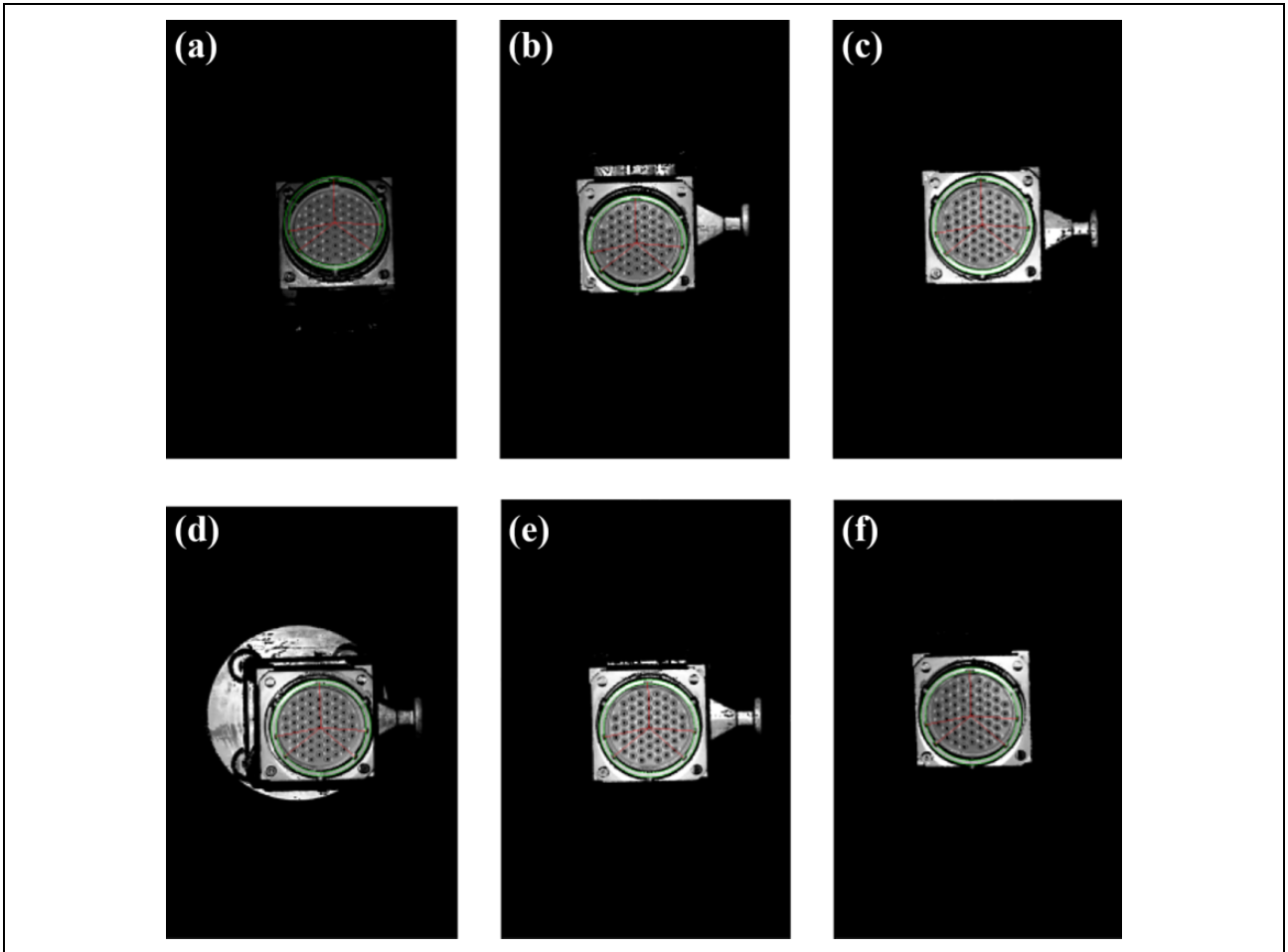

Figure 8. Experiment system.

Figure 9. Feature extraction results of images at different angles and distances: (a) image after rotating around the positive directions of the x-axis, (b) image after rotating around the negative directions of the x-axis, (c) image after rotating around the positive directions of the y-axis, (d) image after rotating around the negative directions of the y-axis, (e) image after translating along the positive directions of the z-axis, and (f) image after translating along the negative directions of the z-axis.

Table 1. Prior information.

Desired center point image coordinates P_{ad} (pixel)	(627.08, 1233.53)
3D coordinates of desired center point P_{ad} (mm)	(0.17, 21.72, −180.04)
Desired attitude angle θ_{dx} (°)	2.72
Desired attitude angle θ_{dy} (°)	−0.86
Groove angle θ_{mdz} (°)	(10.26, 39.06, 138.60, 174.96, 270.54)
The displacement D_1 of the evacuation (mm)	180
The displacement D_2 of the captured image (mm)	135
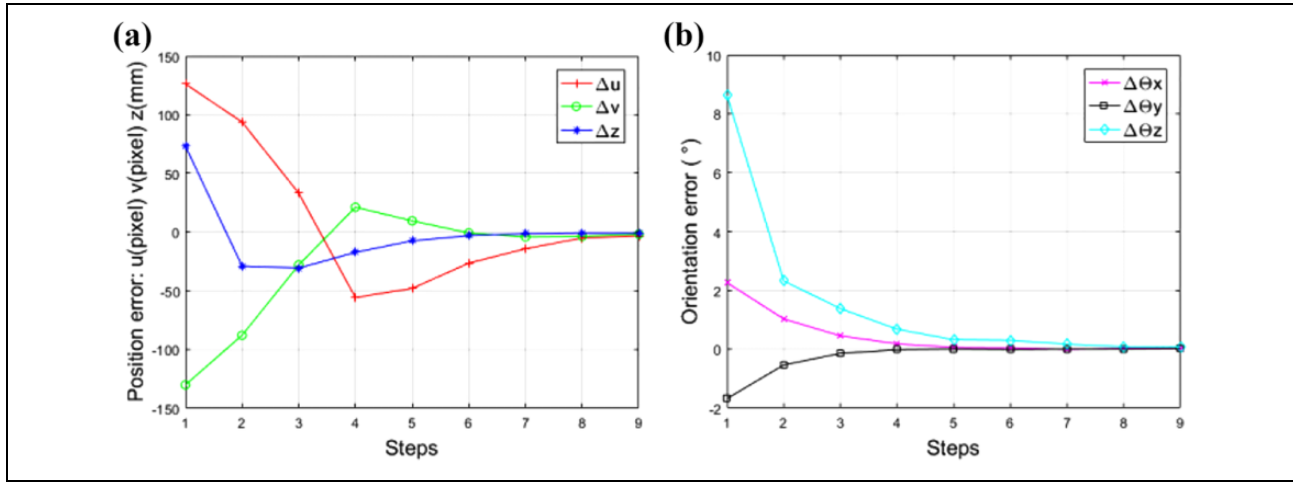
Figure 10. Error curves with the proposed method: (a) position error of component B and (b) orientation error of component A.

U-Net network and feature extraction results

The training set for the U-NET network consisted of 60 images with different angles and distances and 600 images generated by data augmentation. Each image was a gray image obtained by the structured light camera in an actual environment. The size of the original images was 1251 × 1925 pixels, which were resized to 512 × 512 pixels when training U-NET network.

The new images of different angles and distances were input into the trained U-NET network for test. The feature extraction experiments with the method described in the "Groove Feature Extraction" section were conducted. The extracted features for the images captured at different angles and distances are shown in Figure 9. Figure 9(a) and (b) are the feature extraction results of the image after rotating around the positive and negative directions of the x-axis, respectively. Figure 9(c) and (d) are the feature extraction results of the image after rotating around the positive and negative directions of the y-axis, respectively. Figure 9(e) and (f) are the feature extraction results of the image after translating along the positive and negative directions of the z-axis, respectively. It can be seen from Figure 9 that the five grooves on the ring area are all accurately extracted.

Automatic assembly

Before the assembly experiment, the desired features of component B had been obtained by the method in the "Groove Feature Extraction" and "Desired Image Capture Stage" sections. The coefficient k of the similar ellipse was equal to 0.25. The prior information obtained in desired image capture stage is presented in Table 1.

In the assembly experiments, the poses of component A and component B were initialized randomly within a certain range, and the structured light camera obtained the current image of component B in real time. The current features of component B were obtained by the method in the "Groove Feature Extraction" and "Desired Image Capture Stage" sections. The errors between the current features and the desired features were used as the input of hybrid visual servoing system. The coefficients k_1 and k_2 in the hybrid visual servoing system were both set to 0.6. The error curves of component B between the current pose and the desired pose are shown in Figure 10. It can be seen that after about eight steps, the position error and orientation error have approached 0.

The trajectory of component B in image space during the assembly process is shown in Figure 11(a). It can be seen that the center point image coordinates of component B are gradually approached the desired center point image coordinates. The trajectory of component B in Cartesian space during the assembly process is shown in Figure 11(b). It can be seen that the center point 3D coordinates of component B are gradually approached the desired center point 3D coordinates.
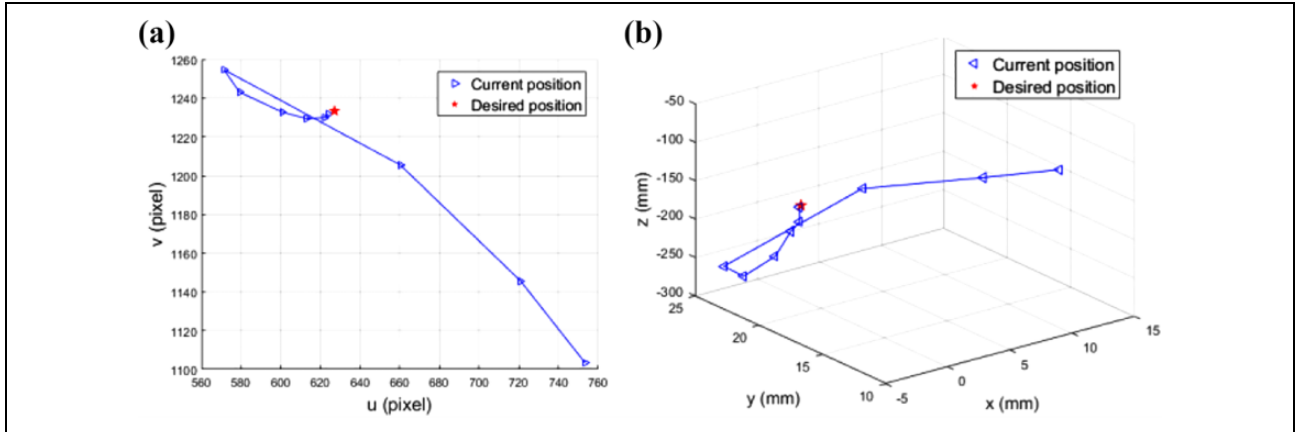
Figure 11. The trajectory of component B in assembly: (a) trajectory in image space and (b) trajectory in Cartesian space.
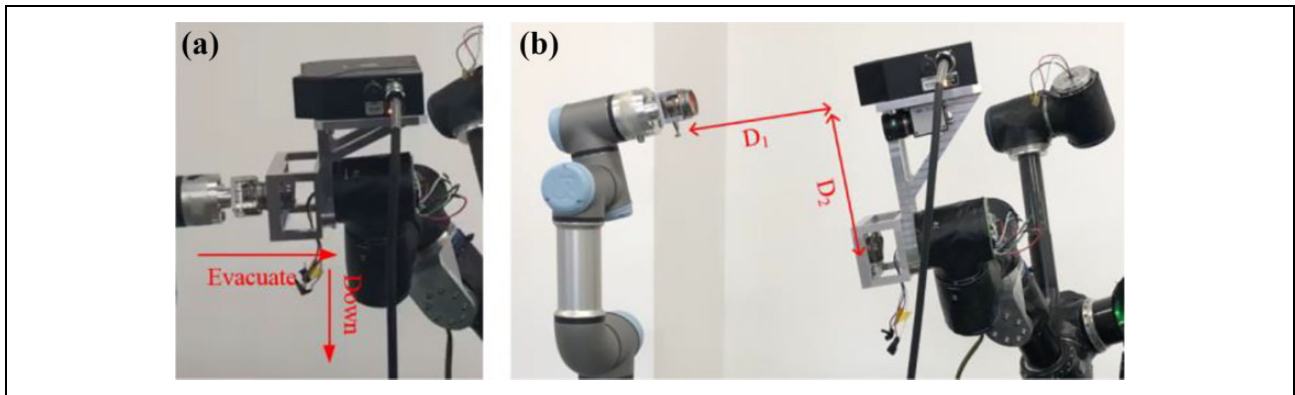

Figure 12. Desired image capture stage: (a) the direction of movement of the end-effector of the manipulator and (b) the displacement D_1 of the evacuation.

The actual scenes of the desired image capture stage are shown in Figure 12. As shown in Figure 12(a), the manipulator was manually controlled to complete one assembly. Manipulator 1 was translated the given displacement D_1 along the z-axis in its end-effector frame to move component A away from component B. Manipulator 1 was translated along the x-axis in its end-effector frame until the camera could capture the image of component B. The displacement along the x-axis was recorded as D_2.

After the desired features had been obtained, we initialized the poses of component A and component B, as shown in Figure 13(a). As shown in Figure 13(b), the camera alignment state was achieved after hybrid visual servoing control.

As shown in Figure 14(a), after component A had moved up D_2, it was aligned with component B. Then component A was translated the displacements (D_1-d) along the z-axis in the end-effector frame of manipulation 1, where d was equal to 3 mm. After ensuring the safety of assembly, component A was translated the displacement d along the z-axis in the end-effector frame of manipulation 1. Then component A was inserted into component B, as shown in Figure 14(b).
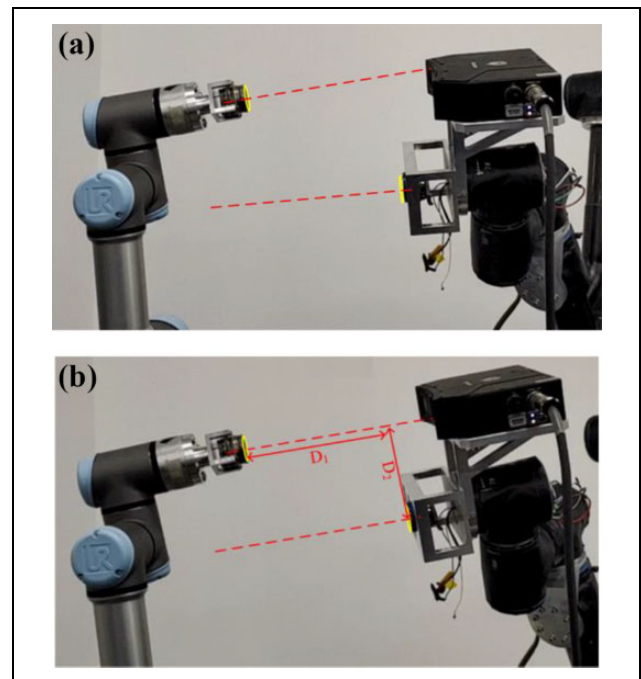

Figure 13. The camera alignment stage: (a) initial state and (b) camera alignment state.

The total time cost in one assembly was about 18 s: it was as follows, camera alignment 16 s and component insertion 2 s. Fifty assembly experiments were conducted, and all were successful. It can be found the alignment and insertion achieved good results.

Comparative experiments

The position-based method in ref.[20] was selected as the comparative method. The position-based visual servoing control was realized according to formula (9), and the features were all from Cartesian space

Figure 14. The component insertion stage: (a) translation D_2 and (b) component insertion state.

$$
\begin{bmatrix}
\Delta x \\
\Delta y \\
\Delta z \\
\Delta \theta_x \\
\Delta \theta_y \\
\Delta \theta_z
\end{bmatrix}
=
\begin{bmatrix}
k_a(x_{ac} - x_{ad}) \\
k_a(y_{ac} - y_{ad}) \\
k_a(z_{ac} - z_{ad}) \\
k_b(\theta_{cx} - \theta_{dx}) \\
k_b(\theta_{cy} - \theta_{dy}) \\
k_b \Delta \theta_z
\end{bmatrix}
\tag{9}
$$

where the difference from formula (8) was that x_{ac}, x_{ad}, y_{ac}, and y_{ad} were obtained by directly reading the 3D coordinates of the desired point and current point in the camera, and $\Delta \theta_z$ is calculated by 3D coordinates.

The coefficients k_a and k_b in equation (9) were both set to 0.6. A series of comparative experiments were well conducted. Component A was also well aligned with component B in orientation and position and was successfully inserted into component B to form an assembled component with the method in ref.[20] In one experiment with the comparative method, the error curves of component B between the current pose and the desired pose are shown in Figure 15. It can be seen that after about 10 steps, the position error and orientation error have approached 0. The error curves of the comparative method oscillate more times, and our method has a faster convergence speed.

The errors and steps of eight groups of comparative experiments in orientation alignment and position alignment were listed in Table 2. It can be found that the errors of our method are in a smaller range. Because the method in ref.[20] will suddenly have a large error in a certain dimension, our proposed method is more steady.

Conclusions

A robotic assembly system with two manipulators is designed to assemble two components with six DOFs in 3D space. A feature extraction algorithm for the images of components is designed with the U-NET network. A hybrid visual servoing method combining the errors in image
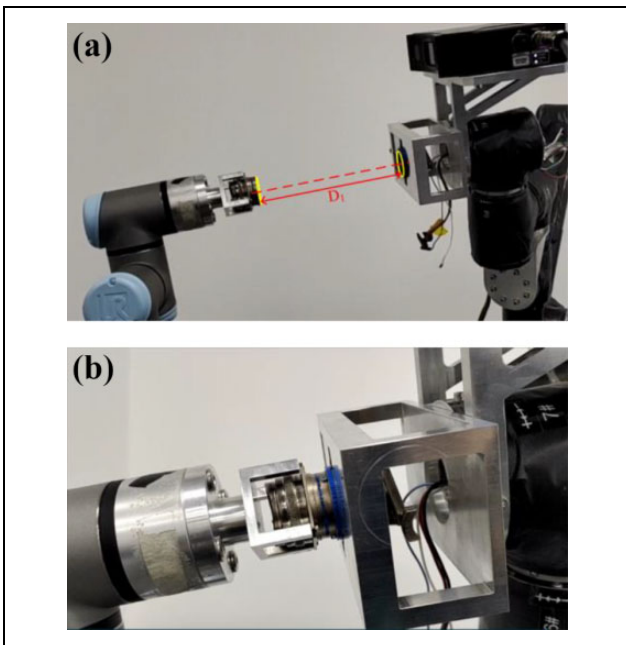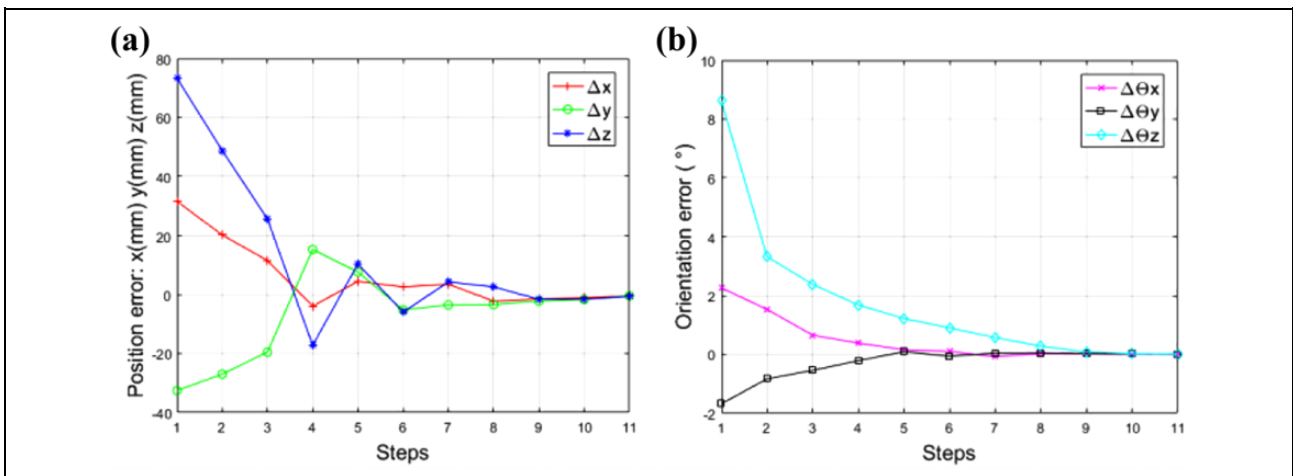

Figure 15. Error curves with the comparative method: (a) position error of component B and (b) orientation error of component A.

Table 2. The errors and steps in camera alignment.

| No. | (Δx, Δy, Δz) (mm) and ($\Delta\theta_x$, $\Delta\theta_y$, $\Delta\theta_z$) (degree) | | | Steps | |
| | | After camera alignment | | | |
	Initial	Proposed method	Method in ref.[20]	Proposed method	Method in ref.[20]
1	(19.41, 10.42, 28.62) (3.96, −0.96, 20.65)	(0.15, −0.16, 0.15) (0.05, 0.06, 0.73)	(0.23, −0.17, 0.21) (0.06, 0.04, 0.89)	7	9
2	(−21.72, 26.43, −42.40) (1.03, 0.66, 10.23)	(0.17, 0.05, 0.26) (0.05, 0.06, 0.23)	(0.21, −0.07, 0.15) (0.06, 0.01, 1.52)	8	10
3	(30.07, −8.59, −7.68) (−3.12, −1.36, 5.03)	(0.15, −0.14, 0.23) (−0.01, 0.04, 0.41)	(−0.11, −0.18, −0.06) (0.02, −0.02, 0.07)	7	7
4	(−5.05, 4.26, −40.56) (1.64, 1.87, −9.74)	(0.05, 0.23, 0.26) (0.01, 0.01, 0.68)	(−0.01, 0.10, 0.16) (−0.01, 0.01, 0.03)	6	7
5	(−12.11, 10.72, 7.80) (3.86, 0.42, 30.89)	(−0.03, 0.06, 0.29) (−0.01, 0.01, 0.03)	(0.05, 0.09, 0.16) (0.01, 0.02, 0.19)	6	8
6	(1.21, 17.32, 60.63) (6.46, −0.71, −10.51)	(0.11, 0.02, 0.04) (0.02, 0.02, 0.51)	(0.13, 0.07, 0.14) (−0.01, 0.01, 0.29)	6	7
7	(−9.15, 26.29, 92.58) (9.87, 0.03, 4.62)	(0.04, 0.08, −0.05) (−0.01, 0.01, 0.13)	(0.13, 0.14, 0.17) (−0.02, 0.02, 0.25)	7	7
8	(−10.08, 10.53, −76.86) (−8.87, −2.15, 10.09)	(0.04, 0.12, 0.13) (−0.01, 0.01, 0.48)	(0.25, 0.14, 0.23) (−0.02, 0.01, 0.53)	8	9

space and Cartesian space is proposed. Three DOFs are controlled in image space, which are the center's position on the image plane and the rotation of component B around the z-axis. The other three DOFs are controlled in Cartesian space, which are the depth and the rotations around the x-axis and y-axis.

A series of complete assembly experiments have been conducted in a real environment. The pose error is reduced to a small range in a few steps, and the success rate in 50 assembly experiments is 100%. Subsequently, a series of comparative experiments to compare the proposed method with the method in ref.[20] are well-conducted. The error curves of the method in ref.[20] oscillate more times, and our method has a faster convergence speed. The errors of our method are in a smaller range. Our method can improve the steadiness and efficiency of the alignment process. The alignment process of component A to component B is converged fast and accurately with our method.

In the future, we will pay more attention to more intelligent assembly control methods.

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

ORCID iDs

Xian Tao https://orcid.org/0000-0001-5834-5181
De Xu https://orcid.org/0000-0002-7221-1654

References

1. Tsenev V. Robot assembly with flexible automatic control according to INDUSTRY 4.0. In: *IEEE XXVIII international scientific conference electronics (ET)*, Sozopol, Bulgaria, 12–14 September 2019, pp. 1–4. DOI: 10.1109/ET.2019.8878551.

2. Zeng F, Xiao J, and Liu H. Force/torque sensorless compliant control strategy for assembly tasks using a 6-DOF collaborative robot. *IEEE Access* 2019; 7: 108795–108805.

3. Yu Y, Xu Z, Lv Y, et al. Design and analysis of space docking mechanism for on-orbit assembly with application to space telescopes. In: *IEEE international conference on mechatronics and automation (ICMA)*, Changchun, China, 5–8 August 2018, pp. 1867–1871. DOI: 10.1109/ICMA.2018.8484668.

4. Luo Y, Chen M, Wang X, et al. Precision assembly system based on position-orientation decoupling design. In: *2nd world conference on mechanical engineering and intelligent manufacturing (WCMEIM)*, Shanghai, China, 22–24 November 2019, pp. 685–688. DOI: 10.1109/WCMEIM48965.2019.00145.

5. Yu H, Ma T, Wang M, et al. Feature-based pose optimization method for large component alignment. In: *4th international conference on control, robotics and cybernetics (CRC)*, Tokyo, Japan, 27–30 September 2019, pp. 152–156. DOI: 10.1109/CRC.2019.00039.

6. Wang S, Chen G, Xu H, et al. A robotic peg-in-hole assembly strategy based on variable compliance center. *IEEE Access* 2019; 7: 167534–167546.

7. Meng S, Ruiqin H, Lijian Z, et al. Precise robot assembly for large-scale spacecraft components with a multi-sensor system. In: *5th international conference on mechanical, automotive and materials engineering (CMAME)*, Guangzhou, China, 1–3 August 2017, pp. 254–258. DOI: 10.1109/CMAME.2017.8540181.

8. Lei Y, Xu J, Zhou W, et al. Vision-based position/impedance control for robotic assembly task. In: *Chinese control conference (CCC)*, Guangzhou, China, 27–30 July 2019, pp. 4620–4625. DOI: 10.23919/ChiCC.2019.8865406.

9. Taptimtong P, Mitsantisuk C, Sripattanaon K, et al. Multi-objects detection and classification using vision builder for autonomous assembly. In: *10th International conference of information and communication technology for embedded systems (IC-ICTES)*, Bangkok, Thailand, 25–27 March 2019, pp. 1–4. DOI: 10.1109/ICTEmSys.2019.8695970.

10. Liu S, Xu D, Li Y, et al. Nanoliter fluid dispensing based on microscopic vision and laser range sensor. *IEEE Trans Ind Electron* 2017; 64(2): 1292–1302.

11. Liu S, Xu D, Liu F, et al. Relative pose estimation for alignment of long cylindrical components based on microscopic vision. *IEEE/ASME Trans Mechatron* 2016; 21(3): 1388–1398.

12. Liu S and Xu D. Fast and accurate circle detection algorithm for porous components. *J Electr Eng Electron Technol* 2014; 03(01): 1–8.

13. Sun S, Yin Y, Wang X, et al. Robust landmark detection and position measurement based on monocular vision for autonomous aerial refueling of UAVs. *IEEE Trans Cybern* 2019; 49(12): 4167–4179.

14. Kim J, Nguyen H, Lee Y, et al. Structured light camera base 3D visual perception and tracking application system with robot grasping task. In: *IEEE international symposium on assembly and manufacturing (ISAM)*, Xi'an, China, 30 July–2 August 2013, pp. 187–192. DOI: 10.1109/ISAM.2013.6643524.

15. Satorres M, Gómez O, Gámez G, et al. Visual predictive control of robot manipulators using a 3D ToF camera. In: *IEEE international conference on systems, man, and cybernetics*, Manchester, UK, 13–16 October 2013, pp. 3657–3662. DOI: 10.1109/SMC.2013.623.

16. Litvak Y, Biess A, and Bar-Hillel A. Learning pose estimation for high-precision robotic assembly using simulated depth images. In: *International conference on robotics and automation (ICRA)*, Montreal, QC, Canada, 20–24 May 2019, pp. 3521–3527. DOI: 10.1109/ICRA.2019.8794226.

17. Chaumette F and Hutchinson S. Visual servo control, part I: basic approaches. *IEEE Robot Autom Mag* 2006; 13: 82–90.

18. Chaumette F and Hutchinson S. Visual servo control, part II: advanced approaches. *IEEE Robot Autom Mag* 2007; 14: 109–118.

19. Xu D, Lu J, Wang P, et al. Partially decoupled image-based visual servoing using different sensitive features. *IEEE Trans Syst Man Cybern Syst* 2017; 47(8): 2233–2243.

20. Peng Y, Jivani D, Radke RJ, et al. Comparing position- and image-based visual servoing for robotic assembly of large structures. In: *IEEE 16th international conference on automation science and engineering (CASE)*, Hong Kong, China, 20–21 August 2020, pp. 1608–1613. DOI: 10.1109/CASE48305.2020.9217028.

21. Corke P and Hutchinson SA. A new hybrid image-based visual servo control scheme. In: *Proceedings of the 39th IEEE conference on decision and control (Cat. No. 00CH37187)*, Sydney, NSW, 12–15 December 2000, pp. 2521–2526, vol. 3. DOI: 10.1109/CDC.2000.914182.

22. Deng W and Yao J. Extended-state-observer-based adaptive control of electrohydraulic servomechanisms without velocity measurement. *IEEE/ASME Trans Mechatron* 2020; 25(3): 1151–1161.

23. Deng W and Yao J. Asymptotic tracking control of mechanical servosystems with mismatched uncertainties. *IEEE/ASME Trans Mechatron* 1–1. DOI: 10.1109/TMECH.2020.3034923.