

Construction of Vision-based Manipulation System for 3D Industrial Objects

Guangtao Zhao, Ying Jia and Zhicai Ou

Abstract— Many applications call for robots to perform the tasks that the robot can accurately grasp and assemble many kinds of industrial parts. Using robots to perform these tasks can afford human safety as well as the high productivity. In this paper, an industrial robot system aiming to assemble the industrial parts with high precision is presented. The paper focuses on the design of the robot control system and the precise extraction of the position-and-orientation of the industrial parts. Finally, experiments are presented to show the efficiency of the system, in which sets of automotive parts, such as axletrees, bearings, pistons and pins are grasped from the conveyor and assembled together precisely.

I. INTRODUCTION

The application of vision-based industrial robotic assembly system is becoming increasingly widespread, especially in the capability of object grasping and assembly. In application of industrial assembly line, stability, effective manipulation and operation speed are very important to complete the whole task. This fact necessitates the high performance for the industrial assembly robotic system. And a typical industrial robot system should include the following functions: object identification, grasping, path planning and assembly. Aiming to the above, an industrial robot system with high precision is presented and this paper is focused on the design of the control system and the extraction of 3D position information for the target object.

In the present industrial application of robotic assembly system, works in visual control can be classified into three main categories based on the control structure and the definition of the error signal. The first is the 3D visual control structure[1]. In this structure, the input of the error signal is calculated in three-dimension Cartesian coordinate. The features extracted from the image are utilized to estimate the pose of the target with respect to the camera. The second category is the 2D visual control structure [2] [3] & [4]. In this structure, the control error function is defined in the image space. The control input is defined neither in task space coordinates nor in joint coordinates since a closed loop scheme is performed directly in image plane. For this method, it is necessary to compute the image Jacobian matrix which establishes the relationships between the changes of the object and the changes of the corresponding image features. Different from the above two techniques, the third approach which is called 2-1/2-D

visual control structure[5] is defined by the control error function, in part in the image space and in part in the Cartesian space.

Compared with the control structure mentioned above, a simplified image-based visual control system will be presented in this paper. With the assumption of a fixed height between the camera and the object, this visual control system provides a fast computation to meet the real-time operation requirement. What's more, the shift position of the object in the image frame is proportional to that in the world frame.

On the other hand, the grasping manipulation is one of the crucial steps for the assembly system. Considering the subsequent assembly operation with high precision, the industrial parts need to be grasped with a proper position and a proper orientation and this requires the accurate localization information for the target object. Therefore, it is necessary to precisely figure out the position and orientation of the parts to be grasped. Regarding to this, much work has been presented. Kefalea [6] presented an approach to localize the objects by combining the evidence provided by different visual cues, edges and colors. A method was presented by Speth *et al.* [7], who used the images of the target from different points of view to recover critical 3D information such as size, location and pose. Besides using the visual information, Haidacher and Hirzinger [8] proposed an approach for object localization by utilizing the contact measurements. Despite this, there are still problems and challenges for reliability and accuracy of the parts localization.

In this paper, a robotic system for assembling different kinds of industrial parts is proposed. The target of this system is to grasp the industrial parts, and further to put them together with high precision (0.1mm). Considering the constraints and requirements of the actual assembly line, a simplified and efficient visual control system and a 3D grasping model are adopted to achieve fast, stable and efficient grasp and assemblage.

This paper is organized as follows. Section II gives a brief introduction to the robotic assembly system. In Section III, a simplified and efficient visual control system is analyzed and in Section IV the efficient visual-based grasping model is presented. Section V gives the other important compositions of the system. Experimental results are showed in Section VI and conclusions are drawn in the last Section.

II. SYSTEM FUNCTIONAL DESCRIPTION

The practical target of the system is to make the system

Guangtao Zhao is with the Institute of Automation, Chinese Academy of Science.

Ying Jia is with the College of Science, Minzu University of China.

Zhicai Ou is with the Institute of Automation, Chinese Academy of Sciences.

RoboticVisionTech EX2014
ABB v. RoboticVisionTech
IPR2023-01426

able to search, identify, grasp and assemble the different industrial parts by itself. The robotic system consists of a six d.o.f. FANUC robot arm endowed with a camera-in-hand, a two-fingered hand, a conveyor and some designed fixtures. The hardware framework of the system is depicted by Fig. 1.

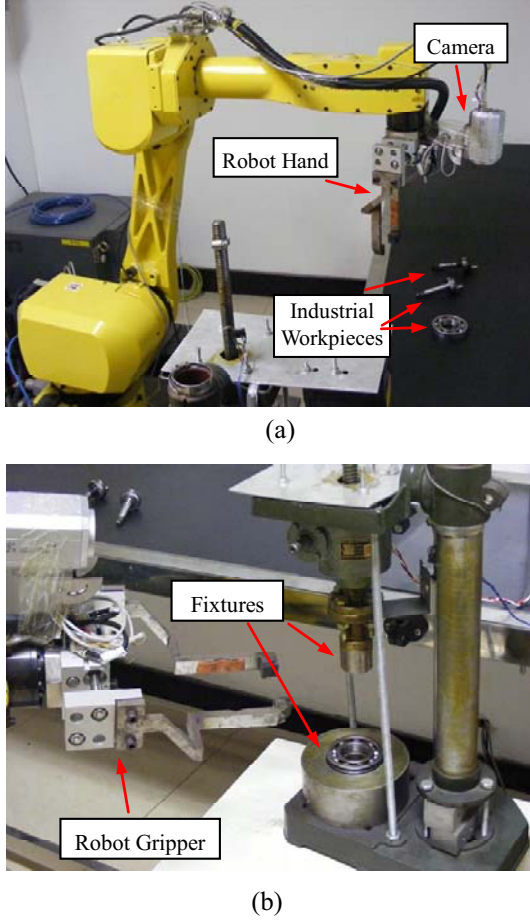


Fig. 1. The Hardware Framework of the Assembly System. (a) shows the robot, the workpieces placed on the conveyor and the camera. (b) the fixtures for helping the assembly operation

In Fig. 2, the whole software framework of the robotic assembly system is described. The manipulation process includes the following steps: (1) To search above the conveyor for the industrial parts and classify them; (2) To obtain the position and orientation of the target parts; (3) To

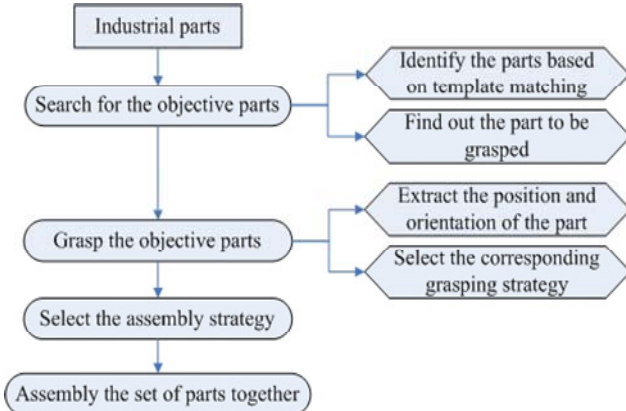


Fig. 2. The framework of the software in the system

choose the related grasping strategy according to the pose information of the object; (4) To control the robot to grasp the target parts with proper strategy; (5) To perform the corresponding assembly task aiming at different parts. In this paper, we focus on the construction of the system and the localization of the 3D target parts.

III. IMAGE MOMENTS AS VISUAL FEATURES FOR THE VISION CONTROL STRUCTURE

In the industrial assembly, the important performance is high speed and robustness which make the robot system more complicated. In the industrial system, however, there are some types of properties which help to make the construction of the robot system simple and efficient: (1) There are often only a few kinds of industrial parts to be identified. (2) The accurate geometric parameters of industrial parts are usually known. (3) The industrial parts often have distinctive features (holes and corners).

Image moments were widely utilized for pattern recognition in the past. In our assembly system, the image moments are used because they can give a generic and geometric representation of the target part whatever the shape of the part is simple or complex. The key point to use the image moments is to try to avoid the image singularities that occur when redundant image features are utilized.

The analytical form of the image Jacobian matrix about the image moments was determined in [9]. Here, the visual features using image moments are selected to manipulate the piston and peg, the bearing and axletree.

A. Image Moments and Their Interaction Matrix

For a binary or segmented image, let $R(t)$ be the observed part image at time t . Then, the image moment of the target part is defined by

$$m_{ij}(t) = \iint_{R(t)} I(x, y) dx dy \quad (5)$$

where $I(x, y) = x^i y^j$.

If the centroid of the object are $x_g = m_{10} / m_{00}$ and $y_g = m_{01} / m_{00}$, the centered moments with respect to the centroid of the target object is

$$\mu_{ij}(t) = \iint_{R(t)} (x - x_g)^i (y - y_g)^j dx dy \quad (6)$$

Here, we take into account the relationship between the time variation of the moments and the relative kinematic screw $V = (v, \omega)$, where $v = [v_x, v_y, v_z]$ and $\omega = [\omega_x, \omega_y, \omega_z]$ represent the translational and rotational velocity, respectively [10]

$$\dot{m}_{ij} = L_{m_{ij}} v, \quad \dot{\mu}_{ij} = L_{\mu_{ij}} V \quad (7)$$

where $L_{m_{ij}}$ is the interaction matrix related to the origin and $L_{\mu_{ij}}$ the interaction matrix with respect to the centered moments.

B. Controller Design

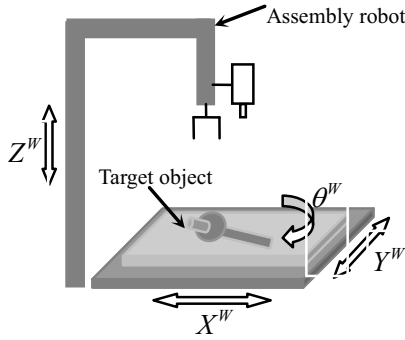


Fig. 3. The camera is not just over the workpiece but should obtain the position and orientation of the workpiece at a glance.

The manipulation of the industrial parts is achieved by the use of a three-degree-of freedom stage with two translation and one rotation motion. To provide visual information, three moments-based visual features are selected. They are x_g and y_g , which are the center of gravity of an industrial part in the image, and the object orientation α . These three visual features are given as follows:

$$x_g = \frac{m_{10}}{m_{00}}, \quad y_g = \frac{m_{01}}{m_{00}}, \quad \alpha = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \quad (8)$$

Through the development of the image moments, we obtain the following relationship between the velocity of the stage motion and the feature change obtained in image coordinates

$$\begin{bmatrix} \dot{x}_g \\ \dot{y}_g \\ \dot{\alpha} \end{bmatrix} = \begin{bmatrix} -f/Z_g & 0 & y_g \\ 0 & -f/Z_g & -x_g \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ \omega_z \end{bmatrix} \quad (9)$$

where z_g is the depth of the gravity center. Since the interaction matrix is upper triangular, it presents partially decoupling properties. Therefore, the image singularities problem is solved since the jacobian matrix is full rank all the time.

In addition, due to image processing error and model computation error, the open-loop visual control is not accurate and robust enough to meet the industrial requirement. Thus, visual servo which is a close-loop control is adopted in our system (showed in Fig. 4).

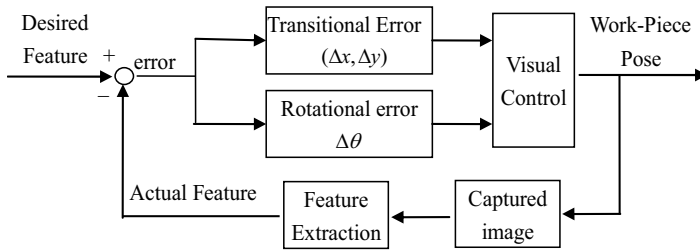


Fig. 4. Visual control configuration

IV. VISION-BASED OBJECT MODELING METHOD

Position and orientation information of the industrial parts are often needed to decide how to grasp them and operate them. In addition, the precision of the obtained information has a great impact on the grasping speed and accuracy of the robot system. Therefore, how to obtain the information efficiently and quickly is one of the key issues in the system.

Because of the central perspective projection, it is not accurate to directly compute the orientation of the target parts based on the image points. In our system, the camera lens is always parallel to the conveyor so the points in the conveyor plane and its corresponding image points meet the relation of simple-ratio invariability. Since that, an arbitrary angle constructed by the points in the conveyor plane is equal to the corresponding angle constructed by their image points. In addition, the accurate geometric parameters of industrial parts are known and the camera lens is calibrated ahead of time so the distance between the image plane and the conveyor is known too. Let Ω denote the set of all 3-D vertices of the industrial parts, ω the set of corresponding image points in the plane, π_p and π_c the plane of the conveyor. The following procedure is adopted:

1. To classify the point set Ω into $\{x_i\}$ and $\{y_j\}$ in order to make

$$\sum_i \{x_i\} \in \pi_c \quad \text{and} \quad \sum_j \{y_j\} \notin \pi_c. \quad (1)$$

Here, the position of image points of $\{x_i\}$ is not necessary to change.

2. Aiming at every point in $\{y_j\}$, apply the following 3D grasping model to rectify its position.

Based on the above procedure, an example is analyzed. Two vertices A_w and B_w of a part are like in Fig.3 and the robot's task is to grasp it from the point A_w in the vertical direction. As depicted in Fig. 5, the points T_c , A_c , B_w and O_c are in the plane π_c , and the points T_p , A_p , B_p and O_p are in the plane π_p . Additionally, O_p is the principal point of the image, O_L is the focus, O_c is the intersection point of the optical axis and the plane π_c .

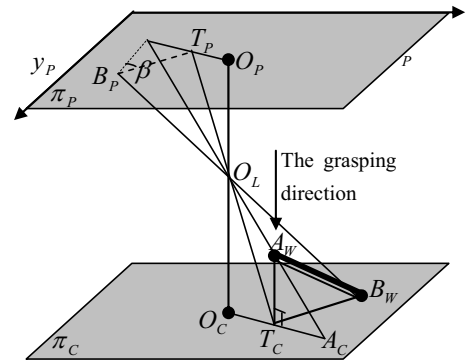


Fig. 5. Illustration of obtaining the orientation of the industrial parts

Based on Fig. 5, it is obvious that the image of the line segment $A_w B_w$ is $A_p B_p$. Generally, we have the orientation of $A_p B_p$ as the orientation of $A_w B_w$. However, in the assembly system, the correct grasping orientation is identical with the orientation $T_p B_p$ and the error angle is the angle β .

For this problem, the point T_p is not known. Therefore, the key point of this problem is to accurately find out the point T_p in the image.

Based on the geometric knowledge, we find out that

$$r_i = \frac{A_p O_p}{A_c O_c} = \frac{O_p T_p}{O_c T_c} \propto \quad (2)$$

Here r_i is related to the distance of camera lens to the conveyer and it can be computed by camera calibration. Thus, the grasping direction can be computed as follows.

$$\alpha = \arctan\left(\frac{y_{Tp} - y_{Bp}}{x_{Tp} - x_{Bp}}\right) \quad (3)$$

Why does the robot system result in the error angle β ? Through analyzing, the following reason is found out. Due to the central perspective projection, the image of the $A_w B_w$ is not the $T_p B_p$ but the $A_p B_p$. If the camera focal length is f and the distance between the image plane and the conveyor plane is d , the error angel β is:

$$\beta = \arccos\left(\frac{B_w T_c \cdot (d - A_w T_c)}{\sqrt{(B_w T_c)^2 \cdot (d - A_w T_c)^2 + (A_w T_c)^2 \cdot (O_c T_c)^2}}\right) \quad (4)$$

Based on the equation (5), the influencing factors are:

- The vertical distance $A_w T_c$ between the point A_w and the plane π_c .
- The distance of the point O_c and T_c .
- The distance between the lens and the conveyor.

Through the analysis, the distance $A_w T_c$ and the distance $O_c T_c$ are proportional to the error angel. However, the distance between the lens and the conveyor is the contrary to the above two points. But, the image of the part to be grasped becomes small when the distance become longer. Therefore, a proper distance is chosen to meet our requirement.

V. THE OTHER IMPORTANT COMPOSITIONS OF THE SYSTEM

In the assembly system, there are still some important compositions except for the whole system configuration. They are the following three sections.

A. Hand-Eye Calibration

To accurately control the displacements of the robot, the relative position and orientation (transformation matrix) between the camera and the robot hand must be computed before the application of the assembly system. This problem

is known as the hand-eye calibration.

Based on [11], the hand-eye calibration can be resolved by moving the robot more times and observing the corresponding motion of the camera and the robot at the same time.

B. Searching for and Matching Objects with the Hausdorff Distance

Before the robot grasps the objective part, the important tasks are to recognize them fast and efficiently.

In our robotic assembly system, the matching method based on the Hausdorff distance has been widely applied in recent years. This method is quite tolerant of small position errors such as those that occur with edge detectors and other feature extraction methods. Given two finite point sets $A = \{a_1, \dots, a_p\}$ and $B = \{b_1, \dots, b_q\}$, the Hausdorff Distance is defined as

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (10)$$

where $h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$.

In our assembly robotic system, the modified Hausdorff distance [12] is used as the matching model to recognize the different industrial parts. It is defined as

$$h(A, B) = \frac{1}{N_A} \sum_{a \in A} \min_{b \in B} \|a - b\| \quad (11)$$

Where N_A is the number of the sets A .

C. The Assembly of the Industrial Parts

In the system, two kinds of devices for the parts assembly are designed to achieve the assembly task. Regarding to the assembly of the axletree and bearing, it is enough to only apply the mechanical devices. However, it is difficult to assemble the piston and peg because of interference fit. In this case, the principle of attractive region is used to meet the requirement with high precision as 0.1mm.

VI. EXPERIMENTAL RESULTS

Many experiments about the assembly of automotive parts are tested in our designed robotic system. The sets of the automotive parts include the bearing-axletree set and the piston-peg set. In our system, a Nikon color CCD camera is mounted on the robot end-effector. The acquired image is 704×576 pixels, and contains a lot of distortion and blur.

According to the specific task, the whole manipulation is divided into two stages: the off-line stage and the on-line stage.

- At the off-line stage, the following important steps should be completed. They are: (1) the hand-eye calibration which finds out the transformation matrix between the robot gripper and the camera in order to accurately control the hand; (2) to specify the base position of the automotive parts to be grasped and establish their templates of the objective parts.
- At the on-line stage, the motion of the robot is

controlled by the closed-loop visual system. The manipulation for the robotic system to grasp the target parts includes the following four steps: (1) To search for the parts to be operated; (2) To recognize different kinds of the parts; (3) To obtain the position and orientation information of the parts; (4) To operate the parts based on the vision feedback and pre-defined strategies.

In this section, the experiments of bearing-axletree and piston-peg assembly are taken as the examples so as to illustrate the whole assembly system and validate the presented vision-based modeling method.

A. Off-line Preparations before Experiment

Considering the actual requirement, the CCD camera is mounted about 300mm from the conveyor. Respectively, we take the base images of the automotive parts: bearings, axletrees, pistons and pegs.

Using the method proposed by [13], we obtain the camera distortion parameters which are $k_1 = 1.280739e-007$ and $k_2 = -2.366564e-012$. In order to finish hand-eye calibration, we move the robot twice and observe the changes about the position of the camera and the changes of the robot end-effector. So as to obtain a unique solution, the two robot end-effector movements must have different axes of rotation and their angles of rotation must not be 0 or π . Let A_1 and B_1 be the first movement of the camera and robot end-effector, and let A_2 and B_2 be the second movement of the camera and robot end-effector. The numerical values of the A_1 , B_1 , A_2 , B_2 are:

$$A_1 = \begin{bmatrix} 0.9653 & 0.2598 & -0.0271 & -48.4978 \\ -0.2606 & 0.9648 & -0.0341 & 1.1852 \\ 0.0172 & 0.0400 & 0.9991 & -7.8113 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 0.9659 & 0.2588 & 0.0000 & 0 \\ -0.2588 & 0.9659 & 0.0000 & 0 \\ 0.0000 & 0.0000 & 1.0000 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0.9993 & 0.0139 & 0.0333 & -20.8673 \\ -0.0187 & 0.9889 & 0.1477 & -13.2461 \\ -0.0308 & -0.1482 & 0.9885 & 23.4041 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0.9848 & 0.0000 & -0.1736 & 20 \\ 0.0000 & 1.0000 & 0.0000 & -20 \\ 0.1736 & 0.0000 & 0.9848 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus, with the two equations $A_1X = XB_1$ and $A_2X = XB_2$ [14], the hand-eye calibration result is:

$$X = \begin{bmatrix} -0.2167 & 0.9717 & -0.1405 & -19.3037 \\ -0.9736 & -0.2105 & 0.0840 & 179.3834 \\ 0.0520 & 0.1071 & 0.9865 & -11.6820 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

B. On-line Operation of the Robot System

Based on the assembly procedure showed in Fig. 2, two kinds of assembly experiments are taken as the examples to validate the efficiency of the whole assembly system. The following is the details.

1) Searching for and Classifying Industrial Parts

As mentioned in section IV, the parts recognition method based on the modified Hausdorff Distance is applied in the searching process. While the robot moves the camera to scan the conveyor for finding target parts, the system performs template matching to classify the parts.

2) Obtaining the Orientation by Vision-based Model

Based on the new 3D grasping model proposed in this paper, the orientation of the axletree is obtained. In Fig. 6, two axletrees are placed randomly on the conveyor and the white lines stands for their orientation accurately through the model.

Through many experiments, the relation between the error angle β and the impact factors is proved.

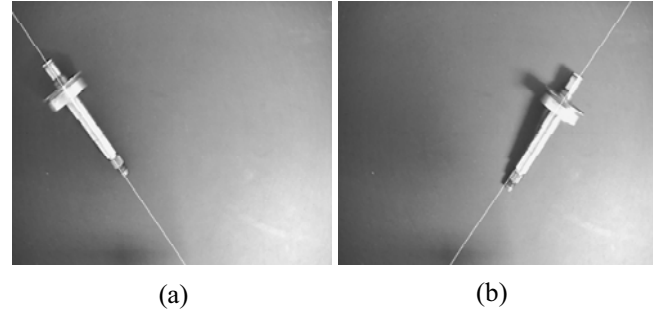


Fig. 6. Orientation of the axletree by applying the vision-based model. The light lines in the two figures indicate the orientations of the axletrees.

3) Obtaining the Position with Image Moments

In this robotic system, it is important to obtain the corresponding position relationship between the present position and the base position in their images. Fig. 7 shows the centers of the axletree, bearing, and piston in the images. In experiments, the error is about 5mm. But it doesn't influence the grasp and assembly of the part.

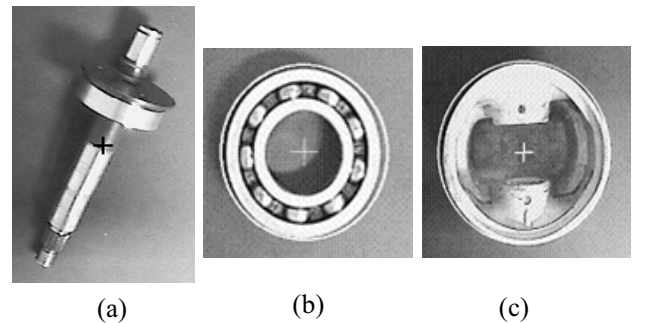


Fig. 7. The center of the automotive parts by applying image moment. (a) the axletree. (b) the bearing. (c) the piston

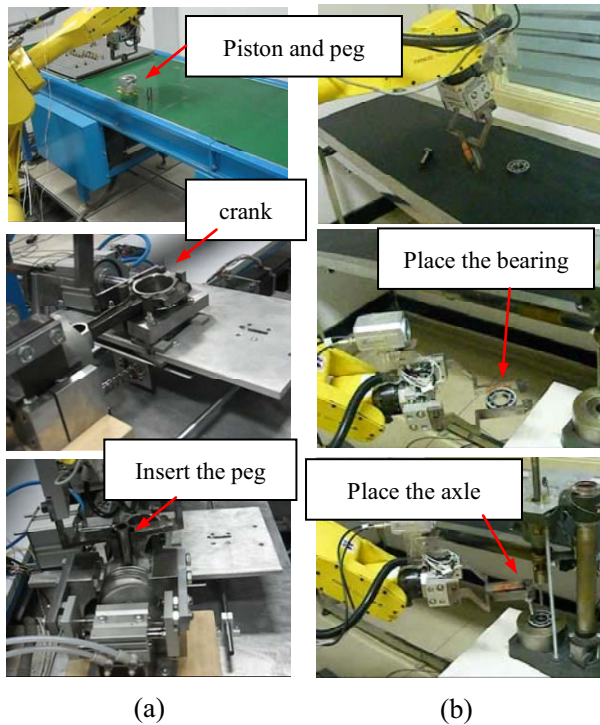


Fig. 8. The axletree-bearing grasping experiments

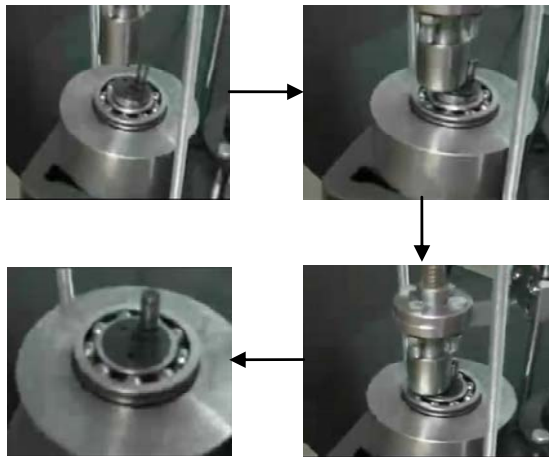


Fig. 9. The axletree-bearing assembly experiments

4) Grasping and Assemble the Industrial Parts

The whole assembly process is illustrated in Fig. 8. The sets of bearing-axletree and piston-peg are assembled with all the techniques mentioned in this paper: searching for the industrial parts, classifying the parts, localizing the axletree and the bearing, grasping the parts and assembling them. Fig. 8 shows the process of the parts grasping. Fig. 9 shows the assembly of the different parts.

C. Brief Summery of the Assembly System

In the assembly system with high precision, the assembly experiments of the axletree-bearing and the piston-peg are performed successfully by the presented methods and procedure.

Further experiments show that (1) The success rate is 75 percent without visual feedback information; (2) The

success rate reaches to 90 percent when using the feedback information once.

VII. CONCLUSION

In this paper, an industrial robotic assembly system is presented to complete the grasp and assembly operation with high precision as 0.1mm. In order to achieve the above task, a simplified and efficient visual control structure is proposed to achieve the fast and robust control. In addition, a novel 3D grasping model is designed to accurately obtain the position and orientation of the parts. What's more, the presented method is general for different kinds of the 3D industrial parts. In the end, the experiments about grasping and assembling of the bearing and axletree are presented to prove the efficiency of the designed robotic assembly system.

REFERENCES

- [1] W.J. Wilson, C.C. Williams Hulls and G.S. Bell, "Relative End-effector Control Using Cartesian Position Based Visual Servoing", *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 684-696. 1996
- [2] N.J. Cowan and D. E. Chang, "Geometric Visual Servoing", *IEEE Transactions on Robotics*, vol. 21, no. 6, pp. 1128-1138. 2005
- [3] P.I. Corke and S.A. Hutchinson, "A New Partitioned Approach to Image-Based Visual Servo Control", *IEEE Transactions on Robotics and Automation*, vol. 17, no. 4, pp. 507-515. 2001,
- [4] R. Kelly, R. Carelli, O. Nasisi, B. Kuchen and F. Reyes, "Stable Visual Servoing of Camera-in-Hand Robotic Systems", *IEEE/ASME Transactions on Mechatronics*, vol. 5, no. 1, pp. 39-48. 2000
- [5] E. Malis, F. Chaumette and S. Boudet, "2½D Visual Servoing", *IEEE Transactions on Robotics and Automation*, vol. 15, no.2 pp. 238-250. 1999
- [6] E. Kefalea, "Object Localization and Recognition for a Grasping Robot", *IECON '98*, vol. 4, pp. 2057 - 2062, 1998.
- [7] J. Speth, A.Morales, P.J.Sanz, "Vision-Based Grasp Planning of 3D Objects by Extending 2D Contour Based Algorithms", *IROS*, pp. 2240 - 2245, 2008.
- [8] S. Haidacher, G. Hirzinger, "Estimating Finger Contact Location and Object Pose from Contact Measurements in 3D Grasping", *ICRA*, vol. 2, pp. 1805 - 1810, 2003.
- [9] F. Chaumette, "Image moments: A general and useful set of features for visual servoing," *IEEE Trans. Robot.*, vol.20, no.4, pp.713-723, Aug. 2004
- [10] Espiau B, Chaumette F, Rives P, "A new approach to visual servoing in robotics", *IEEE Transactions on Robotics and Automation*. 8(3):313-326, 1992
- [11] Y.C. Shiu and S. Ahmad, "Calibration of Wrist-Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form $AX=XB$ ", *IEEE Trans. Robotics and Automation*. vol.5 no.1. pp.16-29. 1989.
- [12] Kwon Oh-Kyu, Sim Dong-Gyu, Park Rae-Hong, "Robust Hausdorff distance matching algorithms using Pyramidal structures", *Pattern Recognition*, 34(7): 2005-2013, 2001
- [13] Guangtao Zhao, Hong Qiao, Zhicai Ou, "A method for calibrating camera lens distortion with cross-ratio invariability in welding seam system", *IEEE Conference of Intelligent Robotics and Applications*
- [14] Shiu Y.C. and Ahmad S. Calibration of Wrist-Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form $AX=XB$. *IEEE Trans. Robotics and Automation* 1989; 5(1): 16-29.