ORIGINAL ARTICLE SPECIAL ISSUE: Intelligent Robotics (1/2) Guest edited by Yi-Hung Liu

Hand-Eye Coordination for Robotic Assembly Tasks

Wen-Chung Chang* and Chia-Hung Wu

Department of Electrical Engineering, National Taipei University of Technology, Taiwan

(**Received** 5 October2012; **Accepted** 25 October 2012; **Published on line** 1 December 2012)

*Corresponding author: wchang@ntut.edu.tw

DOI: 10.5875/ausmt.v2i4.162

Abstract: This paper addresses issues related to the design and implementation of robotic assembly tasks. Specifically, we consider automatic assembly systems with real-time visual sensing for the back shells of cellular phones. Typically, industrial assembly tasks are accomplished using either the look-then-move open-loop or the look-and-move closed-loop control approach. For either approach, successful assembly requires that issues concerned with task accuracy must be considered based on camera calibration parameters. For the hand-eye robotic system to operate in real-time, one must adopt an appropriate control structure to maximize task efficiency. Simple and repetitive assembly tasks can be performed quickly through look-then-move open-loop controls. However, relatively slower look-and-move closed-loop control approaches are better suited for complex tasks or those requiring greater flexibility. To accomplish automatic assembly tasks with real-time visual sensing, either eye-to-hand or eye-in-hand vision must be employed. The proposed vision-based control approaches for back shell assembly tasks are likely to have real potential in industrial manufacturing applications.

Keywords: Binocular Vision; Hand-Eye Coordination; Robotic Assembly; Robotic Manipulator; Vision-Based Control

Introduction

In modern industry, robotic manipulators play a critical role in factory automation. Given the scale of mass production, these robots are designed to perform repetitive tasks. Therefore, the capability of robotic manipulation is somewhat restricted by the designated motion behavior. For robotic manipulators to undertake tasks with autonomy and flexibility, they rely on sensors such as cameras to navigate unknown environments in real-time. By processing real-time visual data, robotic manipulators can thus accomplish autonomous tasks precisely and robustly.

Recent years have seen considerable research into vision-based controls for robotic manipulators. Typical approaches to vision-based control can be classified as: i) Cartesian-based [1], ii) modified Cartesian-based [2], and iii) image-based [3]. In Cartesian-based controls, vision is employed to estimate the robot's pose relative to the environment in Cartesian space. Such a control strategy can be directly implemented whenever the robot comes into contact with the environment. A set of 3-D points for the object, together with their corresponding projections on the 2-D image plane are required to determine the relative pose between the object and the camera. The equations that relate these two data sets are nonlinear in nature and hence increase the difficulty of the problem especially if real-time pose estimation is desired. Haralick *et al.* [4] proposed using iterative weighted least-squares techniques to solve the pose estimation problem. Wu *et al.* [5] used feature points in images to reconstruct the actual environment and plan the task trajectory. Quach *et al.* [6] used a projection matrix to implement tracking control for a robotic manipulator. In [7], an application of visual servoing in medical robotics was presented using compensated periodic heart and breathing motion. The modified Cartesian-based approach only differs in the way errors are encoded for controller design. It has an advantage over the conventional Cartesian-based

approach in that high degree-of-freedom motion tasks can possibly be executed with precision. Specifically, observed features are first computed in the image space to form the required set of features before being reconstructed in Cartesian space to define the encoded error.

Most image-based control approaches provide more robust performance than position-based approaches in terms of camera calibration and robot modeling errors. Chang *et al.* [8] used cameras to capture environmental information to help guide robot movements. Chang [9] also proposed a strategy to implement path tracking control in the absence of correspondence information related to the desired trajectory between the two cameras. Buttazzo *et al.* [10] used a five-axis robotic manipulator to automatically grasp 3-D objects in different heights. Using an eye-in-hand framework, a CCD camera was moved a fixed distance along the z-axis. The camera geometry was then evaluated to estimate the object's depth and height. Lippiello *et al.* [11] proposed an eye-in-hand visual servo system to allow a robotic manipulator to grasp an unknown object, repeatedly applying a fast surface reconstruction algorithm to produce the corresponding local grasp planning. To measure the distance and error between an object and the gripper, Luo *et al.* [12] integrated a laser range finder in a manipulator control system equipped with a camera allowing for feature detection and object recognition.

When using vision for control, the coordinate relation between the camera and robotic manipulator must be known *a priori*. Work in [9, 13-15] presents the stability and convergence property of visual servo control systems. Chang [16] also proposed an eye-to-hand structure that allowed for the reconstruction of the normal vector in an unknown 3-D surface with binocular vision and the projection of laser cross light. Given that tasks are appropriately encoded with available visual measurements, the positioning tasks can be accomplished precisely despite the approximate calibration of the hand-eye systems.

Wen-Chung Chang received his B.S. degree in Control Engineering from National Chiao Tung University, in 1988, his M.S. degree in Engineering from the Department of Aeronautics and Astronautics, Stanford University, in 1992, and his M.S., M.Phil., and Ph.D. degrees in Electrical Engineering from Yale University, in 1993, 1996, and 1997, respectively. Upon returning to Taiwan in 1997, he joined the Department of Electrical Engineering, National Dong Hwa University as an Assistant Professor. Since August 2001, he has been with the National Taipei University of Technology, where he is currently an Associate Professor of Electrical Engineering. His research interests include vision-based control, robotics, intelligent space, visual tracking, adaptive control, and real-time control applications.

Chia-Hung Wu is currently a doctoral student in Department of Electrical Engineering at National Taipei University of Technology. His research interests focus on robotic manipulation, visual servoing, and force-reflection systems.

This paper addresses both eye-to-hand and eye-in-hand vision approaches. With the calibrated stereo vision system, the 3-D pose of the back shell and the phone body can both be determined. Either the look-then-move open-loop control law or look-and-move closed-loop control law is then synthesized to accomplish the required assembly task.

The remainder of this paper is organized as follows. The following section introduces the proposed robotic assembly system configuration. The eye-to-hand and eye-in-hand vision approaches are then compared, with a description of the camera and hand-eye calibration. The look-then-move open-loop and look-and-move closed-loop approaches are addressed before the concluding remarks.

System Description

This paper addresses issues involved in automatically installing the back shell of cellular phones. Figure 1 illustrates the control task. The system is composed of an industrial six-degree-of-freedom robotic manipulator equipped with a vacuum absorption device.

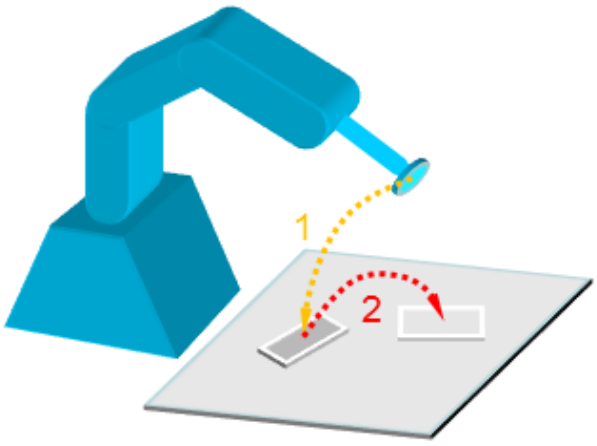

Figure 1. The automatic back shell robotic assembly task is accomplished in two stages: the vacuum absorption device holds the back shell and then positions it onto the cellular phone.

Performing such a robotic assembly task autonomously requires sensors to track the relative positions and orientations of the cellular phone and its back shell. Visually measuring an open space without actual contact is thus an appropriate sensing approach for the control tasks. Measuring objects in 3-D space can be accomplished using a two-camera setup. Both cameras must first be calibrated to establish the coordinate transformation matrix between the tool and the cameras in a process referred to as *hand-eye calibration*. To extract object information from real-time images, images from the two cameras are processed to

detect corner features. These are then used to reconstruct the detected features, 3-D positions and orientations of the back shell and the cellular phone. This allows for the computation of the target poses to reach the back shell and the cellular phone. The robotic tool can then be driven to grasp the back shell by using either the *look-then-move open-loop* or *look-and-move closed-loop* approaches. The back shell is then positioned on top of the cellular phone. This two step process is shown in Figure 1. Since absolute precision is not necessary in the first step, open-loop control could be selected to speed up the assembly process. When the second step is completed, the back shell is positioned onto the cellular phone, and the assembly task is complete. Figure 2 shows an example of the robotic assembly task with eye-in-hand vision. In this example task, a back shell is held and then positioned precisely onto the cellular phone using offline calibrated cameras and hand-eye relations.

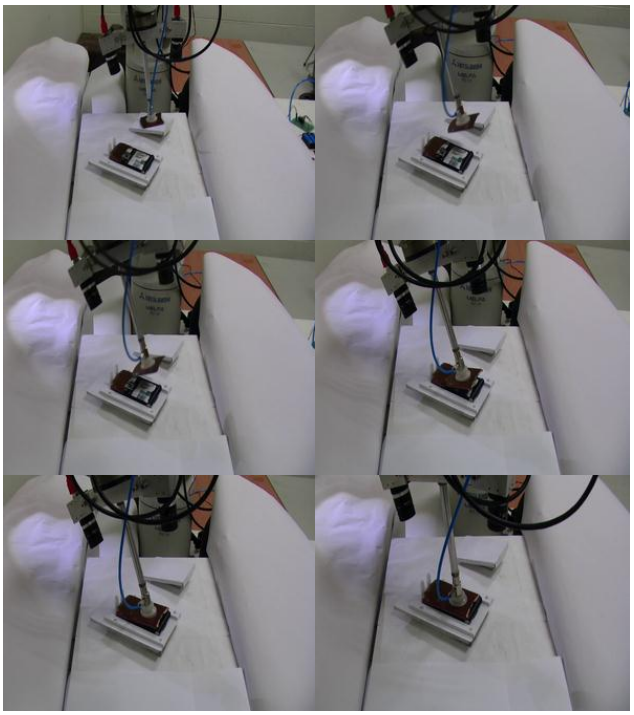

Figure 2. The automatic back shell assembly task is performed using an eye-in-hand vision-based control system. Two-camera real-time vision is used to visually guide a vacuum absorption device to pick up the back shell and then precisely position it onto the cellular phone. The six images, from left to right and top to bottom, illustrate the back shell assembly process.

If the cameras and the hand-eye calibration are both precisely calibrated, one can simply apply the look-then-move open-loop control approach to perform both steps. However, if precise calibration is not possible, ensuring precise task completion may require a look-and-move closed-loop control approach, such as a binocular visual servo control approach. For example, the pose control task could be performed through a sequence of two controllers. The first controller drives the left side of the back shell to the cellular phone using the left camera, at which point the second controller is placed in the feedback loop to drive the right side of the back shell to the cellular phone using the right camera while maintaining the left side of the back shell in target position. Completion of both controller tasks ensures the precise positioning of the back shell on the cellular phone. As long as the tasks are appropriately encoded with the available visual measurements such that the encoded error is zero which implies that the original tasks have been accomplished with precision, control task precision can be guaranteed even given only approximate offline camera and hand-eye calibration.

Eye-to-Hand or Eye-in-Hand Vision

Either eye-to-hand or eye-in-hand vision can be used for automatic assembly tasks with real-time visual sensing. Either configuration allows a robotic manipulator to locate and reconstruct the positions and orientations of a cellular phone and its back shell in the work space. However, due to the fact that the binocular field of view appears to be different for eye-to-hand and eye-in-hand configurations, the motion of lifting and assembling the back shell must be completed using distinct control approaches. Thus, the actual workspace for assembly tasks will be a subset of the manipulator workspace. Nevertheless, in each candidate approach, the robotic manipulator must be driven to grasp the back shell and position it precisely onto the cellular phone.

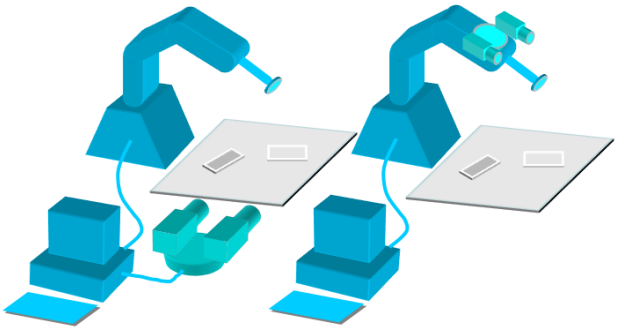

Figure 3. The proposed eye-to-hand {left} and eye-in-hand {right} vision-based control systems for automatic robotic assembly tasks. Real-time observed images are processed to generate appropriate control commands for the manipulator, equipped with a vacuum absorption device, to perform automatic robotic assembly tasks.

Figure 3 illustrates both eye-to-hand and eye-in-hand vision-based control systems for automatic robotic assembly tasks. In the eye-to-hand configuration, the cameras are fixed in the workspace whereas, in the eye-in-hand configuration the cameras are mounted on the end-effector and thus move with the manipulator.

The former approach provides a larger workspace but with lower precision due to the increased distance between the cameras and the assembly parts. The latter approach allows for increased task precision since the cameras are positioned closer to the assembly parts. However, as the distance between the cameras and the assembly parts closes, the binocular field of view effect is cancelled, leaving two uniocular fields of view instead, thus limiting depth assessments.

Calibration

For the proposed hand-eye vision-based control system to automatically perform robotic assembly tasks, both camera parameters [17, 18] and coordinate transformations between the tool and the cameras must be calibrated offline. Task precision apparently depends on the calibration results if the open-loop control approach is used. A closed-loop visual servoing approach could resolve this precision control issue, allowing control tasks to be accomplished with precision even if the system is only approximately calibrated. Nevertheless, for both the look-then-move open-loop and look-and-move closed-loop approaches, the cameras, either eye-to-hand or eye-in-hand, and hand-eye relations must be calibrated offline.

Camera Calibration

Based on the calibration procedure presented in [17, 18], for either system we first estimate the intrinsic and extrinsic parameters of both cameras. Consider the intrinsic matrix of a camera:

$$\mathbf{M}=\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where f_x and f_y are the focal lengths, and c_x and c_y are the image centers.

$$\begin{bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \boldsymbol{\xi}_3 & \boldsymbol{\psi} \end{bmatrix} \tag{2}$$

is the extrinsic matrix of a camera where $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, and $\boldsymbol{\xi}_3$ are rotation vectors, and $\boldsymbol{\psi}$ is a translation vector. A homography denoted by $\mathbf{H}$ can be estimated from the image of a calibration board where all points are assumed with a zero coordinate in the z-axis of the world frame as follows:

$$\mathbf{H}=\begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix}=\frac{1}{\lambda}\mathbf{M}\begin{bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \boldsymbol{\psi} \end{bmatrix}, \tag{3}$$

where λ is an arbitrary scalar. Since $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are mutually orthogonal, one can see that

$$\mathbf{h}_1^T\mathbf{M}^{-T}\mathbf{M}^{-1}\mathbf{h}_2=0, \tag{4}$$

$$\mathbf{h}_1^T\mathbf{M}^{-T}\mathbf{M}^{-1}\mathbf{h}_1=\mathbf{h}_2^T\mathbf{M}^{-T}\mathbf{M}^{-1}\mathbf{h}_2. \tag{5}$$

Let

$$\mathbf{B}=\mathbf{M}^{-T}\mathbf{M}^{-1}=\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}. \tag{6}$$

In the light of Equation (6), one can see that

$$\mathbf{h}_i^T\mathbf{B}\mathbf{h}_j=\begin{bmatrix} h_{i1}h_{j1} \\ h_{i1}h_{j2}+h_{i2}h_{j1} \\ h_{i2}h_{j2} \\ h_{i3}h_{j1}+h_{i1}h_{j3} \\ h_{i3}h_{j2}+h_{i2}h_{j3} \\ h_{i3}h_{j3} \end{bmatrix}^T\begin{bmatrix} b_{11} \\ b_{12} \\ b_{22} \\ b_{13} \\ b_{23} \\ b_{33} \end{bmatrix}. \tag{7}$$

It follows from Equations (4), (5), and (7) that

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11}-\mathbf{v}_{12})^T \end{bmatrix}\mathbf{b}=0. \tag{8}$$

If l images of the calibration board are observed, stacking l such equations as in Equation (8), one can see that

$$\mathbf{V}\mathbf{b}=0, \tag{9}$$

where $\mathbf{V}$ is a $2l\times 6$ matrix. If $l\geq 3$, solution $\mathbf{b}$ exists and the camera parameters can be determined accordingly.

The lens distortion can be modeled by the following equations with parameters q_1 and q_2 [19, 20].

$$\hat{x}_c=x_c+x_c\left[q_1\left(x_c^2+y_c^2\right)+q_2\left(x_c^2+y_c^2\right)^2\right], \tag{10}$$

$$\hat{y}_c=y_c+y_c\left[q_1\left(x_c^2+y_c^2\right)+q_2\left(x_c^2+y_c^2\right)^2\right], \tag{11}$$

where $(\hat{x}_c,\hat{y}_c)$ and (x_c,y_c) respectively denote the camera coordinates with and without distortion.

The coordinate transformation from the calibration board frame $\{B\}$ to either the left camera frame $\{L\}$ or the right camera frame $\{R\}$, involving rotational and

translational parameters, can be established accordingly as follows:

$$ {}^{L}\mathbf{p} = {}^{L}_{B}\mathbf{R}\,{}^{B}\mathbf{p} + {}^{L}\mathbf{p}_{BORG}, \tag{12} $$

$$ {}^{R}\mathbf{p} = {}^{R}_{B}\mathbf{R}\,{}^{B}\mathbf{p} + {}^{R}\mathbf{p}_{BORG}, \tag{13} $$

where the superscripts denote the frame where the coordinates of the point are specified. The coordinate transformation between the left and right camera frames can thus be established as follows:

$$ {}^{L}_{R}\mathbf{R} = {}^{L}_{B}\mathbf{R}\left({}^{R}_{B}\mathbf{R}\right)^{-1}, \tag{14} $$

$$ {}^{L}\mathbf{p}_{BORG} = -{}^{L}_{B}\mathbf{R}\left({}^{R}_{B}\mathbf{R}\right)^{-1}{}^{R}\mathbf{p}_{BORG} + {}^{L}\mathbf{p}_{BORG}. \tag{15} $$

Hand-Eye Calibration

To transform the coordinates in the two-camera frame to the base frame of a manipulator, hand-eye calibration procedures are needed to determine the transformation matrix. Without loss of generality, the two-camera frame can be defined as the left camera frame. In the case of eye-to-hand configuration, one wants to determine the transformation matrix ${}^{0}_{L}\mathbf{T}$ from the left camera frame to the base frame of the manipulator. As with the eye-in-hand configuration, one must further determine the transformation matrix ${}^{T}_{L}\mathbf{T}$ from the left camera frame to the tool frame.

The following procedure describes how the transformation matrices ${}^{0}_{L}\mathbf{T}$ and ${}^{T}_{L}\mathbf{T}$ can be determined for the eye-in-hand configuration. A similar procedure can be performed to determine the transformation matrix ${}^{0}_{L}\mathbf{T}$ for the eye-to-hand configuration. In both configurations, the transformation matrix that specifies the hand-eye relation is the transformation from the two-camera frame $\{C\}$ to the tool frame $\{T\}$ where the two-camera frame $\{C\}$ is defined as the left camera frame $\{L\}$. In particular, the hand-eye calibration can be established as follows:

1. The forward kinematics of the six-degrees-of-freedom manipulator can be used to establish the coordinate transformation between the tool frame and the base frame.
2. Maintain the robotic manipulator at a fixed pose where both cameras can see pre-selected non-coplanar points. The tool's z-axis is set parallel to that of the base frame. Compute the transformation matrix ${}^{0}_{T}T$ based on available manipulation kinematics.
3. Reconstruct the 3-D coordinates

$$ {}^{L}\mathbf{p}_{i} = \left[\begin{array}{cccc} {}^{L}x_{i} & {}^{L}y_{i} & {}^{L}z_{i} \end{array}\right]^{T},\ i \geq 4 \tag{16} $$

of the non-coplanar points based on the two-camera image coordinates. By moving the tool tip so that it touches each of the points, one can determine their 3-D coordinates

$$ {}^{0}\mathbf{p}_{i} = \left[\begin{array}{cccc} {}^{0}x_{i} & {}^{0}y_{i} & {}^{0}z_{i} \end{array}\right]^{T},\ i \geq 4 \tag{17} $$

in the base frame of the manipulator according to available manipulation kinematics. Based on the 3-D coordinates of the points in both the left camera frame and the base frame of the manipulator, one can thus determine the rotation matrix and translational vector as follows:

$$ \left[\begin{array}{c} {}^{0}\mathbf{p}_{i} \\ 1 \end{array}\right] = {}^{0}_{L}\mathbf{T}\left[\begin{array}{c} {}^{L}\mathbf{p}_{i} \\ 1 \end{array}\right], \tag{18} $$

where

$$ {}^{0}_{L}\mathbf{T} = \left[\begin{array}{cc} {}^{0}_{L}\mathbf{R} & {}^{0}\mathbf{p}_{LORG} \\ 0 & 1 \end{array}\right]. \tag{19} $$

Let

$$ {}^{0}_{L}\mathbf{R} = \left[\begin{array}{ccc} {}^{0}_{L}\mathbf{r}_{1} & {}^{0}_{L}\mathbf{r}_{2} & {}^{0}_{L}\mathbf{r}_{3} \end{array}\right]. \tag{20} $$

By stacking the columns in the rotation matrix and the translational vector as the parameter vector, Equation (18) can be re-written as follows:

$$ \left[\begin{array}{cccc} {}^{L}x_{i}I_{3} & {}^{L}y_{i}I_{3} & {}^{L}z_{i}I_{3} & I_{3} \end{array}\right]\left[\begin{array}{c} {}^{0}_{L}\mathbf{r}_{1} \\ {}^{0}_{L}\mathbf{r}_{2} \\ {}^{0}_{L}\mathbf{r}_{3} \\ {}^{0}\mathbf{p}_{LORG} \end{array}\right] = {}^{0}\mathbf{p}_{i}. \tag{21} $$

Therefore, one can compute the least square estimate of the parameter vector by stacking at least four sets of data corresponding to the pre-selected non-coplanar points.

4. The fixed transformation matrix from the left camera frame to the tool frame can thus be determined as

$$ {}^{T}_{L}\mathbf{T} = {}^{T}_{0}\mathbf{T}\,{}^{0}_{L}\mathbf{T}, \tag{22} $$

where ${}^{T}_{0}\mathbf{T}$ is determined by the available manipulator kinematics.

Look-then-Move Open-Loop Approach

For the hand-eye robotic system to operate in real-time, one must adopt an appropriate control structure to maximize task efficiency. Simple and repetitive assembly tasks can be performed quickly by using look-then-move open-loop control approaches. Two-camera vision is used to determine the position and orientation of both the back shell and the cellular phone. Based on the visually-estimated pose, the set-point pose control command is sent to the manipulator for the rapid completion of the assembly task. However, task precision is impacted by the accuracy of both pose estimation and manipulator positioning. Estimation accuracy depends on the calibration precision while the positioning accuracy is typically determined by manipulator performance. An industrial manipulator using offline perfectly-calibrated two-camera vision can provide task accuracy to a resolution up to individual pixels. By using two-camera vision, one can reconstruct the individual corner features of the back shell. The four corners are then used to calculate center point and normal vector. The control procedure is listed below.

1. Reconstruct four corner points from their two-camera image coordinates.
2. Determine center of the back shell in the left and right image planes by locating the intersecting point of two diagonal lines. Use these two image points to reconstruct the 3-D position of the shell center $\mathbf{r}_0$ in the left camera frame.
3. Establish a coordinate vector $\vec{\mathbf{x}}^*$ by taking a vector from $\mathbf{r}_0$ to the middle point of the shorter edge segment. Vector $\vec{\mathbf{y}}^*$ can be established by taking a vector from $\mathbf{r}_0$ to the middle point of the longer edge segment.
4. The normal vector to the back shell can be computed by taking the cross product of the vectors $\vec{\mathbf{x}}^*$ and $\vec{\mathbf{y}}^*$. That is, $\vec{\mathbf{z}}^* = \vec{\mathbf{x}}^* \times \vec{\mathbf{y}}^*$.
5. Convert the 3-D coordinates of $\mathbf{r}_0$ in the left camera frame to the base frame of the manipulator. $\left({}^0x \quad {}^0y \quad {}^0z\right)$ is the desired position of the tool tip in the base frame $\{0\}$.
6. The desired orientation of the tool is thus defined based on $\vec{\mathbf{x}}^*$, $\vec{\mathbf{y}}^*$, and $\vec{\mathbf{z}}^*$. That is, the rotation matrix of the tool frame relative to the base frame can be determined by the corresponding $X-Y-Z$ fixed angle set $\{\alpha \quad \beta \quad \gamma\}$.
7. To guarantee completion of the positioning task, the four corners of both the cellular phone and the back shell must be observed in the two-camera images. By the above methods, one can calculate the center of the cellular phone and its back shell. The desired position and orientation of the tool can then be determined as a pose control command for the manipulator in a look-then-move open-loop motion.

$$\mathbf{u} = \left[{}^0x \quad {}^0y \quad {}^0z \quad \alpha \quad \beta \quad \gamma\right]^T. \tag{23}$$

Look-and-Move Closed-Loop Approach

Simple and repetitive assembly tasks can be performed effectively by using look-then-move open-loop control approaches. However, comparatively slower look-and-move closed-loop control approaches must be adopted for complex tasks requiring flexibility. With real-time visual measurements, the robotic system is deigned to reduce visual errors in real time, thus allowing for the completion of control tasks even when the controlled object and target are both in motion. Meanwhile, the positioning accuracy can be guaranteed at resolutions up to individual pixels even when the vision system and the hand-eye relations are only roughly calibrated. In this real-time visual servo approach, one must first define an appropriate encoded error based on available visual measurements in the following form:

$$\mathbf{e} = G(\mathbf{f}) - G(\mathbf{f}^*), \tag{24}$$

where G is the two-camera model and $\mathbf{f}$ and $\mathbf{f}^*$ denote the observed features and their desired positions. The encoded error $\mathbf{e}$ is defined such that the encoded error being driven to zero implies that the original control task has been accomplished with precision. Differentiating (24) with respect to time t, one can see that

$$\dot{\mathbf{e}} = \frac{\partial G(\mathbf{f})}{\partial \mathbf{f}}\dot{\mathbf{f}} = \mathbf{J}(\mathbf{f})\dot{\mathbf{f}}, \tag{25}$$

where $\mathbf{J}(\mathbf{f})$ is the image Jacobian matrix that requires knowledge of the depth information about features $\mathbf{f}$. With the pre-determined kinematic relation between the observed features rigidly attached to the controlled cellular phone device and the control command sent to the manipulator,

$$\dot{\mathbf{f}} = \mathbf{M}\mathbf{u}. \tag{26}$$

One can thus see the following differential equation representing the look-and-move closed-loop system:

$$\dot{\mathbf{e}} = \mathbf{J}(\mathbf{f})\mathbf{M}\mathbf{u} = \mathbf{L}(\mathbf{f})\mathbf{u}. \tag{27}$$

Using an appropriate depth estimation method, any control law that could drive the encoded error $\mathbf{e}$ to zero would successfully accomplish the automatic assembly task. As an example, the following visual servo control law drives the encoded error $\mathbf{e}$ to zero and could be used for control:

$$\mathbf{u} = -k\mathbf{L}^{+}(\mathbf{f})\mathbf{e}, \tag{28}$$

where the superscript + denotes the pseudo-inverse and k is a positive gain constant.

Conclusion

This paper presents vision-based control approaches for hand-eye robotic assembly tasks. Specifically, corner features in the images of a cellular phone and its back shell are extracted so that 3-D positions and normal vectors of the two objects can be calculated by two-camera vision. Adopting the look-then-move open-loop control approach with precisely calibrated cameras and hand-eye relation, allows for assembly tasks to be completed with resolution up to individual pixels. On the other hand, look-and-move closed-loop visual servo controllers would be capable of accomplishing the task with precision even if the cameras and hand-eye relation are not precisely calibrated because closed-loop control is employed with appropriate task encoding. Once the encoded error is driven to zero by the proposed visual servo controllers, the assembly task has been accomplished with precision up to pixel resolution. In practical situations, the look-then-move open-loop approach typically fails to maintain task precision without precise calibration, whereas the look-and-move closed-loop one appears to be capable of achieving tasks with precision despite only approximate calibration of the camera parameters. The proposed vision-based control approaches for back shell assembly tasks are likely to have real potential in industrial manufacturing applications.

Acknowledgement

This research was supported by the Industrial Technology Research Institute, Taiwan, under grant B301AA4110-FY101. The authors would like to thank Dr. Chun-Lung Chang and Dr. Po-Huang Shieh for their valuable input.

References

[1] W. J. Wilson, "Relative end-effector control using cartesian position based visual servoing," *IEEE Transactions on Robotics and Automation,* vol. 12, no. 5, pp. 684-696, 1996.
doi: 10.1109/70.538974

[2] W. C. Chang, "Visual servoing of rigid robots with modified cartesian-based encoding," in *American Control Conference*, Chicago, IL, 2000, vol. 3, pp. 2018-2022.
doi: 10.1109/ACC.2000.879555

[3] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Transactions on Robotics and Automation,* vol. 8, no. 3, pp. 313-326, 1992.
doi: 10.1109/70.143350

[4] R. M. Haralick, H. Joo, C. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim, "Pose estimation from corresponding point data," *IEEE Transactions on Systems, Man and Cybernetics,* vol. 19, no. 6, pp. 1426-1446, 1989.
doi: 10.1109/21.44063

[5] G. Lau and D. Rahija, "Design and implementation of a distributed real-time image processing system," in *The 1st International Conference on Engineering of Complex Computer Systems*, Ft. Lauderdale, FL, 1995, p. 266.
doi: 10.1109/ICECCS.1995.479340

[6] N. H. Quach and L. Ming, "Visual based tracking of planar robot arms: A scheme using projection matrix," in *IEEE International Conference on Robotics, Intelligent Systems and Signal Processing*, ChangSha, China, 2003, vol. 1, pp. 588-593.
doi: 10.1109/RISSP.2003.1285640

[7] A. A. Rizzi and D. E. Koditschek, "An active visual estimator for dexterous manipulation," *IEEE Transactions on Robotics and Automation,* vol. 12, no. 5, pp. 697-713, 1996.
doi: 10.1109/70.538975

[8] W. C. Chang and M. L. Chai, "Real-time vision-based contour following with laser pointer," in *IEEE International Conference on Robotics and Automation (ICRA)*, Taipei, Taiwan, 2003, vol. 2, pp. 2549-2554.
doi: 10.1109/ROBOT.2003.1241976

[9] W. C. Chang, "Binocular vision-based 3-F trajectory following for autonomous robotic manipulation," *Robotica,* vol. 25, no. 05, pp. 615-626, 2007.
doi: 10.1017/S0263574707003505

[10] G. C. Buttazzo, B. Allotta, and F. P. Fanizza, "Mousebuster: A robot for real-time catching," *IEEE Control Systems,* vol. 14, no. 1, pp. 49-56, 1994.
doi: 10.1109/37.257894

[11] V. Lippiello, F. Ruggiero, and L. Villani, "Floating visual grasp of unknown objects," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2009, pp. 1290-1295.
doi: 10.1109/IROS.2009.5354350

[12] R. C. Luo, M. H. Li, H. L. Jhu, and J. W. Chen, "Multi-sensor based object grasping with eye-in hand laser ranger," in *The 35th Annual Conference of IEEE Industrial Electronics (IECON)*, 2009, pp. 2289-2294.
doi: 10.1109/IECON.2009.5415315

[13] W. C. Chang, "Precise positioning of binocular eye-to-hand robotic manipulators," *Journal of Intelligent and Robotic Systems,* vol. 49, no. 3, pp. 219-236, 2007.
doi: 10.1007/s10846-007-9135-z

[14] W. C. Chang, "Stereo vision-based trajectory following without correspondence information," in *European Control Conference*, Cambridge, UK, 2003.
Available:
http://www.nt.ntnu.no/users/skoge/prost/proceedings/ecc03/pdfs/454.pdf

[15] W. C. Chang, "Hybrid force and vision-based contour following of planar robots," *Journal of Intelligent and Robotic Systems,* vol. 47, no. 3, pp. 215-237, 2006.
doi: 10.1007/s10846-006-9074-0

[16] W. C. Chang, T. M. Hung, and C. W. Cho, "Hybrid tracking control of a binocular vision-based robotic manipulator," in *CACS Automatic Control Conference*, Taipei, Taiwan, 2006.

[17] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *The Seventh IEEE International Conference on Computer Vision*, Kerkyra, 1999, vol. 1, pp. 666-673 vol.661.
doi: 10.1109/ICCV.1999.791289

[18] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 11, pp. 1330-1334, 2000.
doi: 10.1109/34.888718

[19] D. C. Brown, "Close-range camera calibration," *Photogrammetric Engineering,* vol. 37, no. 8, pp. 855-866, 1971.
Available:
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.14.6358

[20] G. Q. Wei and S. D. Ma, "Implicit and explicit camera calibration: Theory and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 16, no. 5, pp. 469-480, 1994.
doi: 10.1109/34.291450