Image-Based Visual Servoing Control of a SCARA Robot

Sung Hyun Han*
Division of mechanical Engineering and Automation, Kyungnam University
Man Hyung Lee
School of Mechanical Engineering, Pusan National University
Hideki Hashimoto
Institute of Industrial Science, University of Tokyo, Japan

In this paper, we present a new approach to visual feedback control using image-based visual servoing with stereo vision. In order to control the position and orientation of a robot with respect to an object, a new technique is proposed using binocular stereo vision. The stereo vision enables us to calculate an exact image Jacobian not only around a desired location but also at other locations. The suggested technique can guide a robot manipulator to the desired location without providing a priori knowledge such as the relative distance to the desired location or the model of an object even when the initial positioning error is large. This paper describes a model of stereo vision and how to generate feedback commands. The performance of the proposed visual servoing system is illustrated by experimental results and compared with conventional control methods for an assembly robot.

Key Words : Visual Feedback, Image-Based Visual Servoing, Intelligent Robot, SCARA Robot

1. Introduction

Visual servoing is the fusion of results from many fundamental areas including high-speed image processing, kinematics, dynamics, control theory, and real-time computing. It has much in common with research on active vision and structure from motion, but is quite different from the often described use of vision in hierarchical task-level robot control systems. Many of the control and vision problems are similar to those encountered by active vision researchers who are building "robotic heads". However, the task in visual servoing is to control a robot to manipulate its environment using vision as opposed to just observing the environment.

There are mainly two ways to put visual feedback into practice. One is called look-and-move, the other visual servoing. The former is a method which transforms the position and orientation of an object obtained by a visual sensor into those in the world frame fixed to an environment and guides the arm of the manipulator to a desired location in the world frame (Allen, 1991 ; Hashimoto, 1991).

In this method, precise calibration of the manipulator and camera system is needed. Visual servoing on the other hand uses the Jacobian matrix which relates the displacement of an image feature to the displacement of a camera motion and performs closed-loop control regarding the feature as a scale of the state. Therefore, we can construct a servo system based only on the image, and can achieve robust control against calibration errors because there is no need to calculate the corresponding location in the world frame (Hashimoto, 1991 ; Chaumette, 1991 ; Hashimoto.

* Corresponding Author,
E-mail : shhan@kyungnam.ac.kr
TEL : +82-551-249-2624 ; **FAX** : +82-551-249-2617
Division of Mechanical Engineering and Automation, Kyungnam University, 449, Weolyoung-Dong, Masan 631-701, Korea. (Manuscript **Received** march 10, 2000; **Revised** April 27, 2000)

1992 ; Han, 1996). A hand-eye system is often used in visual feedback, and there are two ways of arranging the system: placing a camera and a manipulator separately, and placing the camera at the tip of the manipulator. The former motion strategy of the manipulator becomes more complicated than the latter. In the latter, it is easy to control the manipulator using visual information because the camera is mounted at the manipulator tip. In this paper, we deal with the latter method. In conventional approaches, some researchers have presented methods to control the manipulator position with respect to the object or to track the feature points of an object using a hand-eye system as an application of visual servoing (Chaumette, 1991 ; Hashimoto, 1992). These methods maintain or achieve a desired relative position between the camera and the object by monitoring feature points on the object from the camera (Bernard, 1992 ; Weiss, 1987).

However, these tasks have been adrieved by a hand-eye system with monocular vision, and it is necessary to compensate for the loss of information because the original three-dimensional information of the scene is reduced to two-dimensional information on the image. For instance, we must add information on the three-dimensional distance between the feature point and the camera in advance, or use a model of the object stored in memory. In addition, the problem of the manipulator position failing to converge to a desired value arises depending on the way of selecting feature points or when the initial positioning error is not small. It is because some elements of the image Jacobian cannot be computed with only image information, and substituting approximate values at the desired location for them may result in large errors at other locations (Hager, 1996 ; Sundareswaran, 1996).

This paper presents a method to solve this problem by using binocular stereo vision. The use of stereo vision can lead to an exact image Jacobian not only around a desired location, but also at other locations. The suggested technique places a robot manipulator to the desired location without providing a priori knowledge such as the relative distance to the desired location or the model of an object, even if the initial positioning error is large. This paper deals with the modeling of stereo vision and how to generate feedback commands. The performance of the proposed visual servoing system is evaluated by simulations and experiments, and obtained results are compared with the conventional control methools for a SCARA robot.

2. Visual Feedback Control

Visual servo systems typically use one of two camera configurations: end-effector mounted, or fixed in the workspace.

The first, often called an eye-in-hand configuration, has the camera mounted on the robot's end-effector. Here there exists a known, often constant, relationship between the pose of the camera(s) and the pose of the end-effector. Figure 1 represents the image-based visual servo structure.

We define the frame of a hand-eye system with stereo vision and use a standard model of the stereo camera whose optical axes are set parallel to each other and perpendicular to the baseline. The focal points of the two cameras are set apart at a distance d on the baseline, and the origin of the camera frame $\sum_c$ is located at the center of these cameras.

The image plane is orthogonal to the optical axis and set apart at a distance f from the focal point of the camera, and the origins of the frames of the left and right images, $\sum_l$ and $\sum_r$, are located at the intersecting point of the two optical axes and the image planes. The origin of the

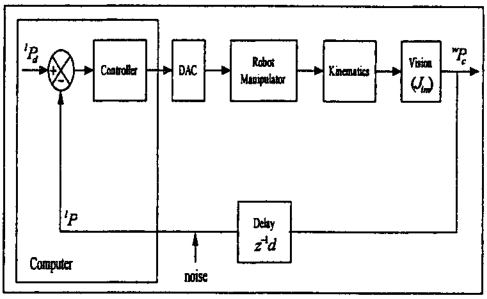


Fig. 1 Block diagram of imaged-based visual servoing system

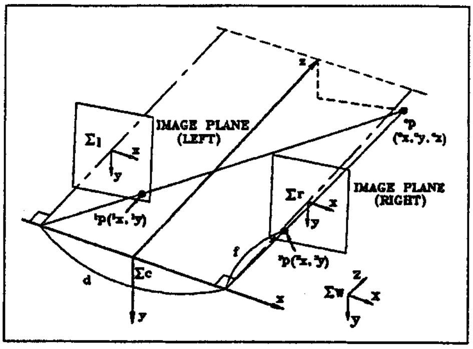


Fig. 2 The coordinates system of stereo vision model

world frame $\sum_w$ is located at a certain point in the world. The x, y, and z axes of the coordinate frames are shown in Fig. 2.

Now let ${}^lp=({}^lx, {}^ly)$ and ${}^rp=({}^rx, {}^ry)$ be the projections onto the left and right images of a point p in the environment, which is expressed as ${}^cp=({}^cx{}^cy{}^cz)^T$ in the camera frame. Then the following equation is obtained (see Fig. 2):

$$ {}^lx\ {}^cz=f({}^cx+0.5d) \qquad (1\text{-a})$$
$$ {}^rx\ {}^cz=f({}^cx-0.5d) \qquad (1\text{-b})$$
$$ {}^ly\ {}^cz=f^c\ y \qquad (1\text{-c})$$
$$ {}^ry\ {}^cz=f^c\ y \qquad (1\text{-d})$$

Suppose that the stereo correspondence between the feature points of the left and right images are found. In visual servoing, we need to know the precise relation between the velocity of the moring camera and the velocity of the feature points in the image, because we generate a feedback command of the manipulator based on the velocity of the feature points in the image.

This relation can be expressed in matrix form called the image Jacobian. Let us consider n feature points $p_k(k=1, \cdot\cdot, n)$ on the object, with coordinates in the left and right images denoted ${}^lq_k({}^lx_k, {}^ly_k)$ and ${}^rp_k({}^rx_k, {}^ry_k)$, respectively. Also define the current location of the feature points in the image ${}^Ip$ as

$$ {}^Ip=({}^lx_1\ {}^rx_1\ {}^ly_1\ {}^ry_1 \cdots {}^lx_n\ {}^rx_n\ {}^ly_n\ {}^ry_n)^T \qquad (2)$$

where each element is expressed with respect to the virtual image frame $\sum_p$.

First, to simplify matters, let us consider the case when the number of the feature points is one. The relation between the velocity of the feature point in image ${}^I\dot{p}$ and the velocity of the camera frame ${}^c\dot{p}$ is given as

$$ {}^I\dot{p}={}^IJ_c{}^c\dot{p} \qquad (3)$$

where ${}^IJ_c$ is the Jacobian matrix which relates the two frames. Now let the translational velocity components of the camera be σ_x, σ_y, and σ_z, and the rotational velocity components be w_x, w_y, w_z. We can the express the camera velocity V as

$$ \begin{aligned} V&=[\sigma_x\ \sigma_y\ \sigma_z\ w_x\ w_y\ w_z]^T \\ &=[{}^cv_c\ {}^cw_c]^T \end{aligned} \qquad (4)$$

The velocity of the feature point seen from the camera frame ${}^c\dot{p}$ can then be written

$$ \begin{aligned} {}^c\dot{p}&=\frac{d^cp}{dt} \\ &=\frac{d}{dt}{}^cR_w({}^wp-{}^wp_c) \\ &={}^cR_w\{-{}^ww_c\times({}^Wp-{}^wp_c)\} \\ &\quad+{}^cR_w({}^W\dot{p}-{}^w\dot{p}_c) \end{aligned} \qquad (5)$$

where ${}^cR_w$ is the rotation matrix from the camera frame to the world frame and ${}^wp_c$ is the location of the origin of the camera frame expressed in the world frame. As the object is assumed to be fixed with respect to the world frame, ${}^w\dot{p}=0$. The relation between ${}^c\dot{p}$ and V is

$$ \begin{aligned} {}^c\dot{p}&={}^cR_w\{-{}^ww_c\times({}^wp-{}^wp_c)\}-{}^cR_w{}^w\dot{p}_c \\ &=-{}^cw_c\times{}^cp-{}^c\dot{p}_c \\ &=\begin{bmatrix} -w_y{}^cz+w_z{}^cy-\nu_x \\ -w_z{}^cx+w_x{}^cz-\nu_y \\ -w_x{}^cy+w_y{}^cx-\nu_z \end{bmatrix} \end{aligned} \qquad (6)$$

Therefore, substituting Eq. (6) into Eq. (3), we have the following equation:

$$ \begin{aligned} {}^I\dot{p}&={}^IJ_c\ {}^c\dot{p} \\ &=J\ V \end{aligned} \qquad (7)$$

In Eq. (7), the matrix J which expresses the relation between velocity ${}^I\dot{p}$ of the feature point in the image and moving velocity V of the camera is called the image jacobian.

From the mathematical model for stereo vision, Eq. (1), the following equation can be obtained:

$$ 2{}^cx({}^lx-{}^rx)=d({}^lx+{}^rx) \qquad (8)$$
$$ {}^cy({}^lx-{}^rx)={}^ly\ d={}^ry\ d \qquad (9)$$
$$ {}^cz({}^lx-{}^rx)=f\ d \qquad (10)$$

The above discussion is based on the case of one feature point. In practical situations, however, visual servoing is realized by using multiple feature points. When we use n feature points, image Jacobians $J_1, \cdots, J_n$ are given from the coordinates of feature points in the image. By combining them, we express the image Jacobian (J_{im}) as

$$J_{im}=[J_1 \cdots\cdots J_n]^T \tag{11}$$

It is therefore possible to express the relation of the moving velocity of the camera and the velocity of the feature points even in the case of mutiple feature points, i.e.,

$${}^I\dot{p}=J_{im}\ V \tag{12}$$

where we suppose that the stereo and temporal correspondence of the feature points are found.

In the case of monocular vision, the image Jacobian J has the following form:

$$J=f\begin{bmatrix} -\frac{1}{{}^cx} & 0 & \frac{{}^cx}{{}^cx^2} & \frac{{}^cx{}^cy}{{}^cz} & -(1+\frac{{}^cx^2}{{}^cx^2}) & \frac{{}^cy}{{}^cz} \\ 0 & -\frac{1}{{}^cx} & \frac{{}^cy}{{}^cx^2} & 1+\frac{{}^cy^2}{{}^cz^2} & -\frac{{}^cx{}^cy}{{}^cx^2} & -\frac{{}^cx}{{}^cz} \end{bmatrix} \tag{13}$$

We now introduce the positional vector of the feature point in the image of the monocular vision system using the symbol ${}^mP=({}^mx,\ {}^my)$. This is the projection of the point expressed as ${}^cP=({}^cx\ {}^cy\ {}^cz)^T$ in the camera frame onto a the image frame of the monocular vision, and has the following relation:

$${}^mx=f\ {}^cx\ {}^cz^{-1} \tag{14a}$$

$${}^my=f\ {}^cy\ {}^cz^{-1} \tag{14b}$$

Substituting Eqs. (14a) and (14b) into Eq. (13) yields another expression of the image Jacobian for monocular vision:

$$J=f\begin{bmatrix} -\frac{1}{{}^cz} & 0 & \frac{{}^mx}{{}^cz} & \frac{{}^mx{}^my}{f} & -\frac{{}^mx^2+f^2}{f^2} & {}^my \\ 0 & -\frac{f}{{}^cx} & \frac{{}^my}{{}^cx} & 1+\frac{{}^my^2+f^2}{f^2} & -\frac{{}^mx{}^my}{f} & -{}^mx \end{bmatrix} \tag{15}$$

The disparity which corresponds to the depth of the feature point is included in J in the case of stereo vision. The s-term expressed in the camera frame ${}^cz$ is included in J in the case of monocular vision. In visual servoing, the manipulator is controlled so that the feature points in the image reach their respective desired locations.

We define an error function between the current location of the feature points in image ${}^Ip$ and the desired location ${}^Ip_d$ as

$$E=Q({}^Ip-{}^Ip_d) \tag{16}$$

where Q is a matrix which stabilizes the system. Then the feedback law is defined as follows:

$$V=-GE \tag{17}$$

where G corresponds to the feedback gain.

To realize visual servoing, we must choose Q so that error convergence is satisfied:

$$\begin{aligned} \dot{E}&=\frac{\partial E}{\partial t} \\ &=Q\ \frac{\partial^I p}{\partial t} \\ &=Q^I\dot{p} \\ &=Q\ J_{im}\ V \\ &=-G\ Q\ J_{im}\ E \end{aligned} \tag{18}$$

We use the pseudo-inverse matrix of the image Jacobian J_{im} for Q to make QJ_{im} positive, and to prevent the input from becoming extremely large, i.e.,

$$Q=J^+_{im}=(J^T_{im}J_{im})^{-1}J^T_{im} \tag{19}$$

Therefore the feedback command is given as

$$V=-G\ J^+_{im}({}^Ip-{}^Ip_d) \tag{20}$$

Figure 3 shows a block diagram of the control scheme described by Eq. (20). Note that the feedback command v is sent to the robot controller, and both the transformation of u to the desired velocity of each joint angle $\dot{q}_d$ and its velocity servo are accomplished in the robot

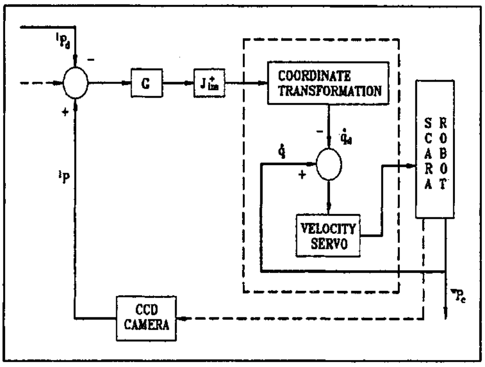


Fig. 3 Block diagram of visual feedback system

controller as show in Fig. 3.

Futhermore, as J_{im} is a $4n\times 6$ matrix and the pseudo-inverse matrix J_{im}^{+} is a $6\times 4n$ matrix, a feedback command Eq. (20) of 6 degrees of freedom is obtained.

3. Experiments

We have compared visual servoing using monocular vision with that using stereo vision by experiments. Fig. 4 represents the experimental set-up. In Fig. 4 two samsung DSP vision boards based-on TMS320C31 chips were used.

In the experiment, the feature points of the object are the four corners of a square whose side dimension is 300 mm. we also use four feature points for the stereo vision case. Parameters are set as follow: the focal length, f=16 mm, baseline d=130 mm, sampling time 50 msec, gain $\lambda=1$, desired location ${}^{c}P_{d}=(100\ 100\ 500)^{T}$mm, desired orientation in Euler angle $(\varphi,\ \theta,\ \psi)=(0,\ 0,\ 0)$ rad, initial error $(-50\ -50\ -50)^{T}$mm in the translation and $(\varphi,\ \theta,\ \psi)=(20,\ 20,\ 20)$ rad in the orientation.

We select the four corners of a rectangle whose size is 200×200 mm as the feature points and set the translational error as $(-150\ -150\ -450)$ mm, with the other values the same as before. The error between the desired location and the current location of the feature points for the monocular and stereo vision cases are shown in Fig. 5.

Next, we show the results for a different choic of feature points. In Fig. 5, we can see that the result diverges in the case of monocular vision, but converges in the case of stereo vision. This is because the image Jacobian is fixed at the desired location in the case of monocular vision. Therefore, a correct feedback command cannot be generated when the initial error is large. On the other hand, the image Jacobian can be updated at every instant in the case of the stereo vision; thus it is possible to generate a correct feedback command which assures stability of visual servoing.

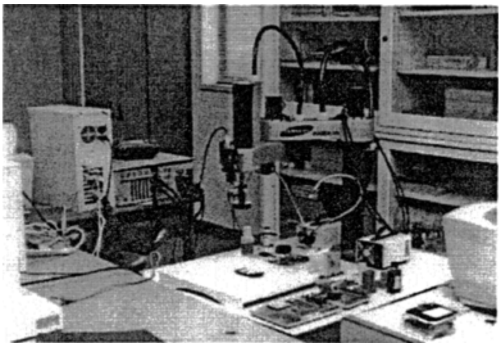
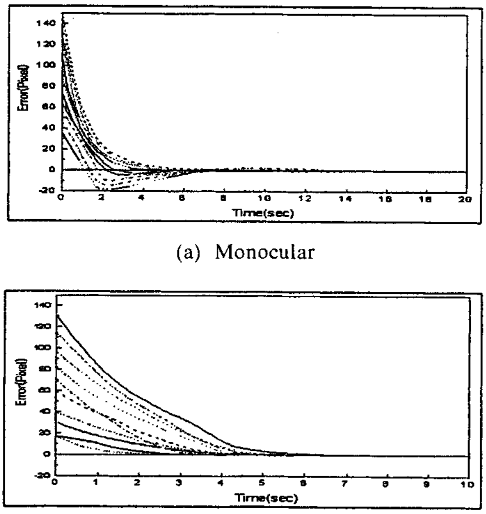


Fig. 4 Experimental equipment set-up.

In experiments, we used a four-axis samsung SCARA Robot (SM5 Model) with a stereo camera attached to the end of the arm. The feature points are three circular planes of 20 mm radius on three corners of an equilateral triangle, with each side of length 87 mm and placed on the board. Precise calibration has not been done for the stereo camera attached to the ends.

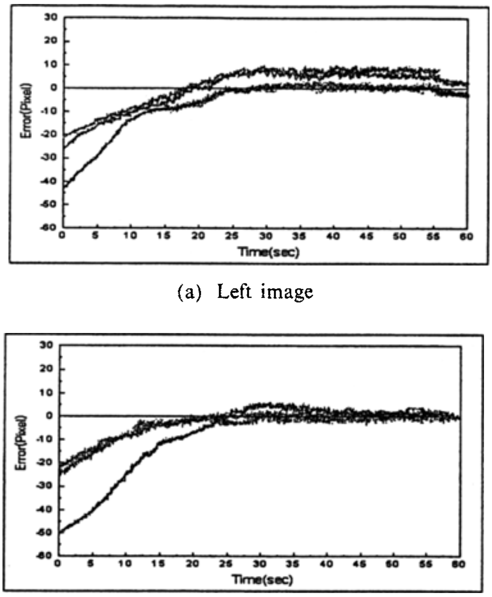
Two stereo images were taken and transformed to binary images in real time and in parallel by two image input devices, and the coordinates of the gravitational center of each feature point was calculated in parallel by two transputers. The stereo correspondence of the feature point was prarided in the first sampling. However, the stereo and temporal correspondence of the feature points in the succeeding sampling were found automatically by searching a nearby area where there were feature points in the previous sampling frame. The coordinates of the feature points were sent to



(a) Monocular

(b) Stereo

Fig. 5 Positional Error in x and y axes in the case of the stereo and monocular vision



(a) Left image

(b) Right image

Fig. 6 Position Error in x and y axes

a transputer for motion control, from which a feedback command was calcuated for the robot. The result was sent to the robot controller via an RS-232C, and the robot was controlled by a velocity servo system in the controller.

The sampling period for visual servoing was about 50 msec; 16 msec for taking a stereo images, about 1 msec for calculating the coordinates of the feature points, 3 msec for calculating feedback command, and about 16 msec for communicating with the robot controller. If feedback input is sent to the robot controller without using the RS-232C, faster visual servoing can be realized.

The desired location was $(0, 0, 500)^T$mm, the desired orientation in Euler angle, $(\phi, \theta, \psi) = (0, 0, 0)$ degree, and the initial error $(50, 50, 50)^T$mm for translation. The other parameters were the same as in the simulation. The errors of the current and desired locations of the feature points are shown in Fig. 6. From these experimental results, we can see that the manipulator converges toward the desired location even if calibration is not precise.

4. Conclusion

This paper proposes a new method of visual servoing with stereo vision to control the position and orientation of an assembling robot with respect to an object. The proposed technique with stereo vision can lead to an exact image Jacobian not only around a desired location but also at the other locations. This technique places a robot manipulator to the desired location without giving a priori knowledge such as the relative distance to the desired location or the model of an object, even if the initial positioning error is large. By using the proposed visual servoing with stereo vision, the image Jacobian can be calculated at any position. Therefore, neither shape information nor desired distance of the target object is required. Also the stability of visual servoing is assured even when the initial errors are very large. The proposed method overcomes several problems associated with visual servoing using monocular vision.

To use this visual servoing method in practical tasks, there still exist many problems such as determining the appropriate number of feature points to reduce noise or quantization error, and methods for choosing the feature points.

References

Allen, P. K., Toshimi, B. and Timcenko, A., 1991, "Real-Time Visual Servoing," *In Proceeding of the IEEE International Conference on Robotics and Automation*, pp. 851~856.

Bernard, E., Francois, C. and Patrick, R., 1992, "A new Approach to Visual Servoing in Robotics," *IEEE Transactions on Robotics and Automation*, pp. 313~326.

Chaumette, F., Rives, P. and Espiau, B., 1991, "Positioning of a Robot with Respoect to an Object, Tracking It and Estimating Its Velocity by Visual Servoing," *In Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2248~2253.

Han, S. H., 1996, "Implementation of Robust Adaptive Control for Robotic Manipulator Using

TMS320C30," *KSME Journal*. Vol. 10, No. 4, pp. 413～422.

Hashimoto, K., Kimoto, T. Edbine, T. and Kimura, H., 1991, "Manipulator Control with Image-Based Visual Servo," *In Proceedings of IEEE International Conference on Robotics and Automation*, pp. 2267～2272.

Hashimoto, K., Edbine, T. and Kimura, H., 1992, "Dynamic Visual Feedback Control for a Hand-Eye Manipulator," *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1863～1868.

Hager, G., Hutchinson, S. and Corke, P. I., 1996, "A Tutorial on Visual Servo Control," *IEEE Trans. Robotics & Automation*, Vol. 12, No. 5, pp. 651～670.

Sundareswaran, V., Bouthemy, P. and Chaumette, F., 1996, "Exploiting Image Motion for Active Vision in a Visual Servoing Framework," *International Journal of Robotics Research*. Vol. 15, No. 6, pp. 629～645.

Weiss, L. E., Sanderson, A. C. and Neuman, C. P., 1987, "Dynamic Sensor-Based Control of Robots with Visual Feedback," *IEEE Journal of Robotics and Automation*, pp. 404～417.