Proceedings of the 13th
KACC, October 1998

Real-Time Control of a SCARA Robot by Visual Servoing with the Stereo Vision

S. H. Han*, M. H. Lee**, K. Son**, M. C. Lee**, J. W. Choi**, and J. M. Lee***

* School of Mechanical Engineering, Kyungnam Univ., Masan, Korea. 631-701
Tel: +82-551-249-2624, Fax: +82-551-243-8133,
E-mail: shhan@kyungnam.ac.kr
** School of Mechanical Engineering, Pusan National Univ.,Pusan, Korea.
Tel:+82-51-510-2331, Fax:+82-51-512-9835
E-mail: mahlee@hyowon.cc.pusan.ac.kr
*** School of Electrical Engineering & ERC/Net Shape & Die Manufacturing, Pusan National Univ.,Pusan, Korea.
Tel:+82-51-510-2378, Fax:+82-51-512-9835
E-mail: jmlee@hyowon.cc.pusan.ac.kr

Abstract

This paper presents a new approach to visual servoing with the stereo vision. In order to control the position and orientation of a robot with respect to an object, a new technique is proposed using a binocular stereo vision. The stereo vision enables us to calculate an exact image Jacobian not only at around a desired location but also at the other locations. The suggested technique can guide a robot manipulator to the desired location without giving such priori knowledge as the relative distance to the desired location or the model of an object even if 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 the simulation and experimental results and compared with the case of conventional method for a SCARA robot.

1. Introduction

When a manipulator is used to grasp or to put an object at a certain place, we must move an end-tip of the manipulator to a desired location. As it is very difficult to adapt the manipulator to continuously changing environments with only inner sensors, we need to use outersensors to detect the error. For instance, human beings mainly use their vision to detect an error and control the movement of the hand. In general, this kind of control which uses vision is called visual feedback.[1,2] There are mainly two ways to put the visual feedback into practice. One is called look-and-move and the other is visual servoing. The former is the 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.

In this method, precise calibration of a manipulator and camera system is needed. On the contrary, visual servoing uses the Jacobian matrix which relates the displacement of an image feature to the displacement of a camera motion and performs a 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 have a robust control against the calibration error because there is no need to calculate the corresponding location in world frame.[2]. A hand eye system is often used in visual feedback and there are two ways of arranging the system. One is placing a camera and a manipulator separately; the other is placing the camera at the end-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 a visual information because the camera is mounted on the manipulator end-tip. In this paper, we deal with the latter method. In the conventional works, some researches have presented methods to control the manipulator position with respect to the object or to track the feature points on an object using a hand eye system as the application of visual servoing.[3] These methods maintain or accomplish a desired relative position between the camera and the object by monitoring feature points on the object from the camera.[4,5]

However, these have been all done by the hand eye system with monocular visions and it is necessary to compensate for the loss of information because the original three-dimensional information of the scene is reduced to two-dimension information on the image. For instance, we must add an information of the three-dimension distance between the feature point and the camera in advance or use a model of object stored in the memory. Besides, a problem that the manipulator position fails 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 the information of the image and substituting approximate values at the desired location for them may result in large errors at the other locations.

This paper presents a method to solve this problem by using a binocular stereo vision. The use of stereo vision can lead to an exact image Jacobian not only at around a desired location but also at the other locations. The suggested technique places a robot manipulator to the desired location without giving such priori knowledge 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 modeling of stereo vision and how to generate feedback commands. The performance of the proposed visual servoing system was evaluated by the simulations and experiments and obtained results were compared with the conventional case for a SCARA robot.

2. Visual Servoing System

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

Figure 1 represents the schematic diagram of a suggested visual servoing system. In Figure 1, two DSP vision boards (MVB03) are used, which were made Samsung Electronics Company in Korea based-on the TMS320C30 chips.

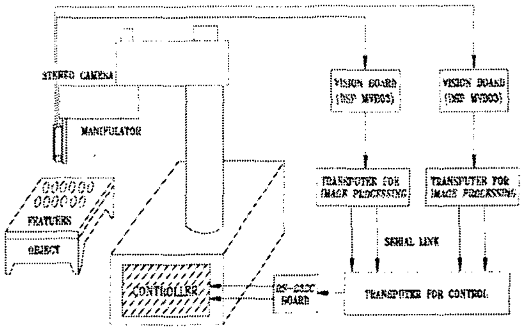


Figure 1. Schematic diagram of visual servoing system.

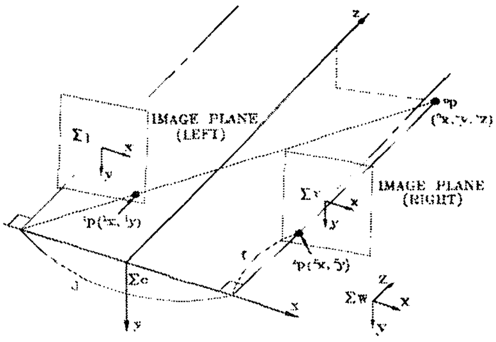


Figure 2. The coordinates system of stereo vision model.

An image plane is orthogonal to the optical axis and apart at distance f from the focal point of a camera and the origins of frame of the left and right images , Σ_l and Σ_r, are located at the intersecting point of the two optical axes and the image planes. The origin of the 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 Figure 2.

Now let ${}^l p = ({}^l x, {}^l y)$ and ${}^r p = ({}^r x, {}^r y)$ be the projections onto the left and right images of a point p in the environment, which is expressed as ${}^c p = ({}^c x \, {}^c y \, {}^c z)^T$ in the camera frame. Then the following equation is obtained (see Figure 2).

$$
\begin{aligned}
{}^l x \; {}^c z &= f\,({}^c x + 0.5d) && \text{(1-a)}\\
{}^r x \; {}^c z &= f\,({}^c x - 0.5d) && \text{(1-b)}\\
{}^l y \; {}^c z &= f\,{}^c y && \text{(1-c)}\\
{}^r y \; {}^c z &= f\,{}^c y && \text{(1-d)}
\end{aligned}
$$

Suppose that the stereo correspondence of feature points between the left and right images are found. In the visual servoing, we need to know the precise relation between the moving velocity of camera and the velocity of feature points in the image, because we generate a feedback command of the manipulator based on the velocity of feature points in the image. This relation can be expressed in a matrix form which is called the image Jacobian.

Let us consider n feature points $p_k(k=1,\cdots,n)$ on the object and the coordinates in the left and right images are ${}^l p_k({}^l x_k, {}^l y_k)$ and ${}^r p_k({}^r x_k, {}^r y_k)$, respectively. Also define the current location of the feature points in the image ${}^I p$ as

$$
\begin{aligned}
{}^I p = ({}^l x_1 \; {}^r x_1 \; {}^l y_1 \; {}^r y_1 \; {}^r y_1 \\
\cdots \; {}^l x_n \; {}^r x_n \; {}^l y_n \; {}^r y_n)^T
\end{aligned}
\tag{2}
$$

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

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

$$
{}^I \dot{p} = {}^P J_c \, {}^c \dot{p} \tag{3}
$$

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

$$
\begin{aligned}
V &= [\,\sigma_x \;\; \sigma_y \;\; \sigma_z \;\; w_x \;\; w_y \;\; w_z\,]^T \\
&= [\,{}^c v_c \;\; {}^c w_c\,]^T
\end{aligned}
\tag{4}
$$

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

$$
\begin{aligned}
{}^c \dot{p} &= \frac{d\,{}^c p}{dt} \\
&= \frac{d}{dt}\,{}^c R_w\,({}^w p - {}^w p_c) \\
&= {}^c R_w \{-\,{}^w w_c \times ({}^W p - {}^w p_c)\} \\
&\qquad + {}^c R_w ({}^W \dot{p} - {}^w \dot{p}_c)
\end{aligned}
\tag{5}
$$

where ${}^c R_w$ is the rotation matrix from the camera frame to the world frame and ${}^w p_c$ is the location of the origin of the camera frame written in the world frame.

As the object is assumed to be fixed into the world frame, ${}^w \dot{p} = 0$. The relation between ${}^c \dot{p}$ and V is

$$
\begin{aligned}
{}^c \dot{p} &= {}^c R_w \{-\,{}^w w_c \times ({}^w p - {}^w p_c)\} \\
&\quad - {}^c R_w \, {}^w \dot{p}_c \\
&= -\,{}^c w_c \times {}^c p - {}^c \dot{p}_c \\
&= \begin{bmatrix} -w_y\,{}^c z + w_x\,{}^c y - \nu_z \\ -w_z\,{}^c x + w_x\,{}^c z - \nu_y \\ -w_x\,{}^c y + w_y\,{}^c x - \nu_z \end{bmatrix}
\end{aligned}
\tag{6}
$$

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

$$ {}^I\dot{p} = {}^I J_c\ {}^c\dot{p} = J\ V \qquad (7) $$

In Eq. (7) 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 model of the stereo vision Eq. (1), the following equation can be obtained.

$$ {}^c x\ (2\,{}^l x - {}^r x) = d\,{}^l x + {}^r x \qquad (8) $$

$$ {}^c y(\,{}^l x - {}^r x) = {}^l y\ d = {}^r y\ d \qquad (9) $$

$$ {}^c z\ ({}^l x - {}^r x) = f\ d \qquad (10) $$

Above discussion is based on the case of one feature point. In practical situation, however, the visual servoing is realized by using plural feature points. When we use n feature points, image Jacobian $J_1, \cdots, J_n$ are given from the coordinates of feature points in the image. By combining them, we express the image Jacobian as

$$ J_{im} = [J_1 \cdots\cdots J_n]^T \qquad (11) $$

Then, it is possible to express the relation of the moving velocity of the camera and the velocity of the feature points even in the case of plural feature points, that is,

$$ {}^I\dot{p} = J_{im}\ V \qquad (12) $$

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

We now introduce the positional vector of the feature point in the image of monocular vision using the symbol ${}^m P = ({}^m x,\ {}^m y)$. This is the projection of the point expressed as ${}^c P = ({}^c x\ {}^c y\ {}^c z)^T$ in the camera frame into the image frame of the monocular vision, and has the following relation.

$$ {}^m x = f\ {}^c x\ {}^c z^{-1} \qquad (13) $$

$$ {}^m x = f\ {}^c y\ {}^c z^{-1} \qquad (14) $$

A disparity which corresponds to the depth of the feature point is included in J in the case of the stereo vision, but s-term expressed in the camera frame ${}^c z$ is included in J in the case of the monocular vision. In the stereo vision, as all elements of J can be described in the image frame, a feedback command can be computed with only the image information. On the other hand, variables in the image frame and the depth ${}^c z$ in the camera frame are combined together in J for the monocular vision. Therefore, it is unable to accomplish feedback control with only information of image and it is necessary to employ either modeling of the object or the desired location in the depth direction ${}^c z$ in the camera frame. To cope with this problem, the visual servoing with the monocular vision conventionally uses a pseudo-inverse matrix of the image Jacobian fixed at the desired point. In the stereo vision, however, there is no need to consider the stability problem, because it is possible to calculate a correct value at every sampling period. Furthermore, to determine the relative location and orientation uniquely, at least four feature points are required in the case of the monocular vision. On the other hand, three feature points are sufficient for the case of the stereo vision.

In the 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 ${}^I p$ and the desired location ${}^I p_d$ as

$$ E = Q({}^I p - {}^I p_d) \qquad (15) $$

where Q is a matrix which stabilizes the system. Then the feedback law is defined as following equation

$$ V = -G\,E \qquad (16) $$

where G corresponds to a feedback gain.

To realize the visual servoing, we must choose Q so that convergence is satisfied with the error system can be satisfied with

$$\dot{E} = \frac{\partial E}{\partial t} = -G\,Q\,J_{im}\,E \qquad (17)$$

We use pseudo-inverse matrix of the image Jacobian J_{im} for Q to make $Q\,J_{im}$ positive and not to make an input extremely large, that is,

$$Q = J_{im}^{+} = (J_{im}^{T} J_{im})^{-1} J_{im}^{T} \qquad (18)$$

Therefore, the feedback command is given as

$$V = -G\,J_{im}^{+}\,({}^{I}p - {}^{I}p_d) \qquad (19)$$

Figure 3 shows a block diagram of the control scheme described by Eq. (19).

Note that the feedback command u is sent to the robot controller and both the transformation of u to the desired velocity of each joint angle and its velocity servo are accomplished in the robot controller as show in the Figure 3.

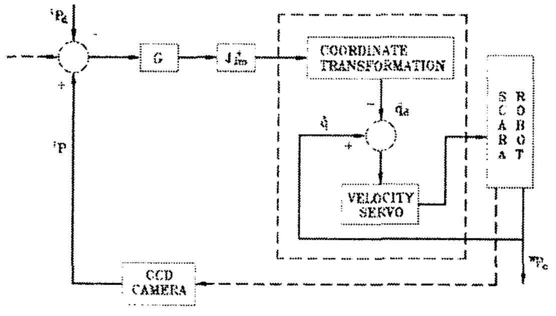


Figure 3. Block diagram of the visual feedback system.

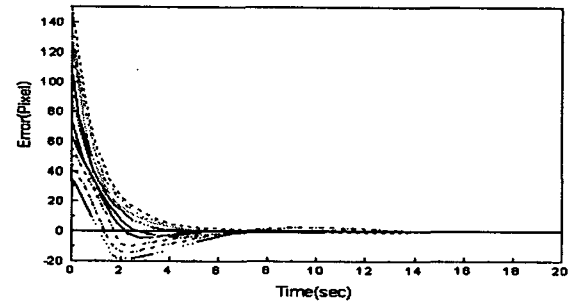
Futhermore, as J_{im} is a $4n \times 6$ matrix and pseud.o-inverse matrix J_{im}^{+} is a $6 \times 4n$ matrix, a feedback command Eq. (19) of 6 degrees of freedom is obtained.

3. Simulation and Experiment

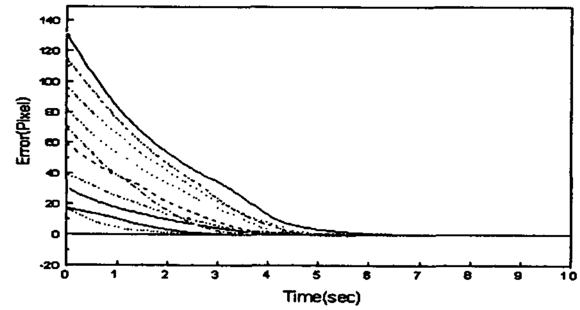
3.1 Simulation

We have compared the visual servoing using the monocular vision with that using the stereo vision by the simulation. In the simulation, feature points of an object are the four corners of a square whose side dimension is 300mm. In the same condition, we used four feature points even in the stereo vision. Each parameter is that the focal length f=16mm, baseline d=130mm, sampling time T=50msec, gain $\lambda=1$, a desired location ${}^{c}P_d = (100\ 100\ 500)^T$mm, a desired orientation in Euler angle $(\varphi, \theta, \psi) = (0,0,0)$rad, and initial error is $(-50\ -50\ -50)^T$mm in the translation. The error between the desired location and the current location of feature points in cases of the monocular and stereo visions are shown in Figure 4.

In Figure 5, we can see that the result diverges in the case of the monocular vision, but converges in the case of the stereo vision.



(a) Monocular



(b) Stereo

Figure 4. Positional error in x and y axes.

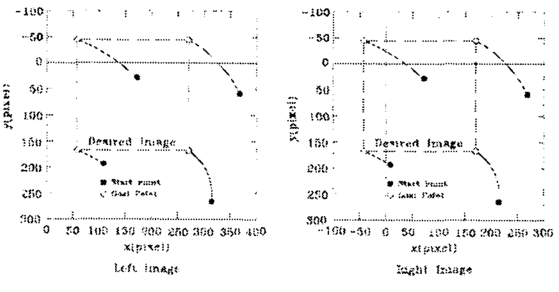


Figure 5. Trajectories of the feature points on the images.

3.2 Experiments

In experiments, we used a four-axis SCARA Robot (SM5 Model) made in Korea with a stereo camera attached to the end tip of the arm. The feature points are three circular planes of 20mm radius on three corners of a equilateral triangle, one side 87mm and are placed on the board. Precise calibration had not been done for the stereo camera attached to the end-tips.

The sampling period of visual servoing was about 50msec. Details were 16 msec for taking a stereo images, about 1 msec for calculating the coordinates of the feature points, 3msec for calculating feedback command, about 16msec for communicating with the robot controller.

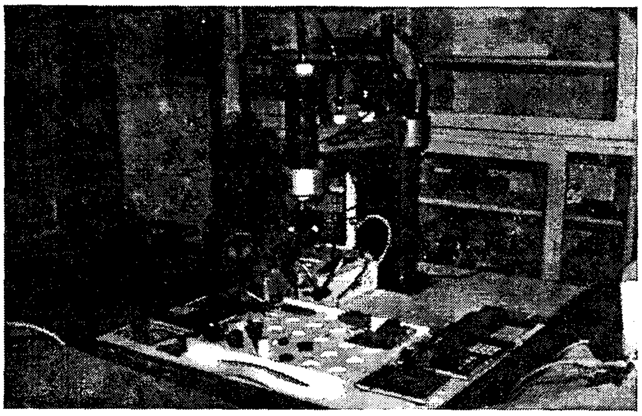
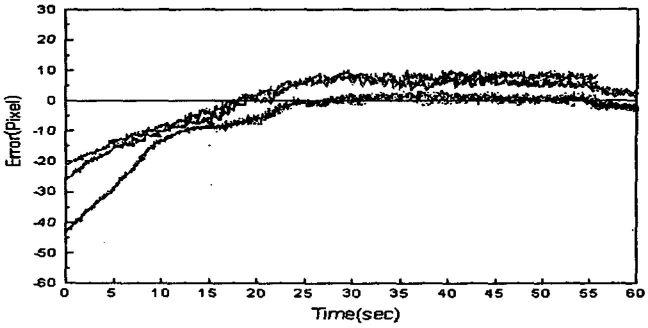


Figure 6. The Experimental equipment set-up.

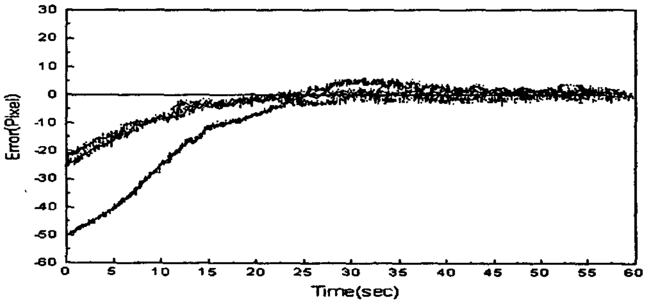
If we send a feedback input to the robot controller without using RS-232C, the faster visual servoing can be realized. From Figure 7, we can see that the manipulator converges toward a desired location even if the calibration is not precise.

4. Conclusion

This paper proposes a new method of visual servoing with the stereo vision to control the position and orientation of a SCARA robot with respect to an object.This visual servoing method overcomes the several problems associated with the visual servoing with the monocular vision. By using this stereo vision, the image Jacobian can be calculated at any position. So neither shape information nor desired distance of the target object is required. Also the stability of visual servoing is assured even when the initial error is very large. By simulation and experiment, the effectiveness of proposed method has been illustrated.



(a) Left image



(b) Right image

Figure. 7 Position error in x and y axes.

5. References

[1] Weiss, L. E., Sanderson, A. C., and Neuman, C. P., "Dynamic sensor-based control of robots with visual feedback". *IEEE Journal of Robotics and Automation*, (1987).

[2] Bernard, E., Francois C., and Patrick R., "A new approach to visual servoing in robotics". *IEEE Transactions on Robotics and Automation*, 8(3) (1992).

[3] Allen, P. K., Toshimi, B., and Timcenko, A., "Real-time visual servoing". *In Proceeding of IEEE International Conference on Robotics and Automation*, (1991).

[4] Chaumette, F., Rives, P., and Espiau, B., "Positioning of a robot with respect 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, (1991).

[5] Hashimoto, K., Kimoto, T., Edbine, T., and Kimura, H., "Manipulator control with image-based visual servo". *In Proceedings of IEEE International Conference on Robotics and Automation*, pp. 2267-2272, (1991).

ACKNOWLEDGE

This study was supported by Samsung Electronics Co., Ltd. as a part of G7 project supervised by Ministry of Commerce, Industry and Energy in korean government, ERC/Net Shape & Die Manufacturing at Pusan National University.