'88 KACC 1988. 10. 21~22

3-D Position Estimation for Eye-in-Hand Robot Vision

°Won Jang, Kyungjin Kim, Myung Jin Chung, Zeungnam Bien

Dept. of EE. KAIST
P.O.Box 150
CHEONG-YANG, SEOUL, KOREA

Abstract

"Motion Stereo" is quite useful for visual guidance of the robot, but most range finding algorithms of motion stereo have suffered from poor accuracy due to the quantization noise and measurement error.

In this paper, 3-D position estimation and refinement scheme is proposed, and its performance is discussed. The main concept of the approach is to consider the entire frame sequence at the same time rather than to consider the sequence as a pair of images. The experiments using real images have been performed under following conditions : hand-held camera, static object. The result demonstrate that the proposed nonlinear least-square estimation scheme provides reliable and fairly accurate 3-D position information for vision-based position control of robot.

1. INTRODUCTION

Consider a camera attached on the hand of a manipulator as shown in Fig.1 and a robotic task, such as grasping arbitrarily located object in 3-D space. After recognizing a desired object from a digitized image, a robot manipulator should approach via visual feedback control[1]. In this case, a robot system needs accurate depth information, which is indispensable not only for recognizing a object but also for precise control of the end-effector with respect to the object. Stereo vision can provide such information, but "motion stereo" may be more favorable.

We are interested in motion stereo for a number of reasons. Especially, the motion of the camera is not restricted to a limited lateral displacement and therefore the end-effector can move freely and reach any configuration. Though tracking a dynamic image is not so trivial task, we can improve the accuracy of the range data by increasing the distance between consecutive image frames assuming known camera motion. In robotics, the camera will be moving under computer control and hence the motion parameter of the camera will be known. Further, in a lot of case, robot motion to reach a object is slow enough, this may allow the system to refine the poor data using multiple measurements.

Several researches have been performed on extraction and refinement of depth information in practical situation. In early work of Nevatia[2], we can find that the use of a number of progressive, closely spaced images has the advantage of reliability over the useof just two views for stereo and potential saving of computational effort. T.D.Williams[3] proposed a range finding technique from the camera motion in real world scene. But, usually, these techniques may suffered from noise and therefore are not always suited for robotic task requiring precision.

A main concept of our approach is to consider the entire frame sequence at the same time rather than to consider the sequence as a set of pairs of imges. In this paper, we confine our discussion to extraction and refinement of 3-D data for vision-based manipulator control under following conditions; 1) calibrated hand-held camera in arbitarary but known motion, 2) static environment with respect to the base frame of the manipulator, 3) some preselected image points can be tracked during visual feedback control without "missing part problem"[9].

2. MODELING OF HAND-EYE COORDINATION

Let us consider a $n \times 1$ joint vector θ_i which represents current configuration of the manipulator consisting each joint angles. Suppose that B_X denotes A homogeneous vector which describes a point P in terms of the Base Frame $\{B\}$[1], then a description of P relative to the Wrist Frame $\{W\}$[2] at i-th step can be calculated by using following kinematic relation[6],

$${}^{W}X_i = {}^{W}_{B}T \cdot {}^{B}X = {}^{B}_{W}T^{-1}(\theta_i) \cdot {}^{B}X$$

where ${}^{B}X = ({}^{B}X\ ,\ {}^{B}Y\ ,\ {}^{B}Z\ ,1)^T$ and ${}^{B}_{W}T$ is a homogeneous 4×4 matrix which describe $\{W\}$−frame relative to { B } and is a function of θ_i.

Now, let us describe the point P in terms of the Camera Frame { C }. In this paper, the Camera Frame { C } can be defined as follows : 1) Z-axis of { C } is parallel to the optical axis of the camera, 2) the X and Y axis of { C } are parallel to the sides of the image, and 3) the origin of the { C } coincides with the center of the iamge plane. Suppose that ${}^{W}_{C}T$ denotes a constant homogeneous transformation which describes { C }-frame relative to { W }-frame and is known, a equation of the kinematic model which describes the point P in terms of { C } at i-th-step is

$${}^{C}X_i = {}^{W}_{C}T^{-1} \cdot {}^{B}_{W}T^{-1}(\theta_i) \cdot {}^{B}X \tag{2}$$

[1] The Base Frame : { B } is located at the base of the manipulator. It is affixed to a non-moving part of the robot, sometimes called link 0.

[2] The Wrist Frame : { W } is affixed to the last link of the manipulator and is defined relative to { B } as ${}^{B}_{W}T$.

Our observation model should represent a process of the image formation, i,e it maps a point ${}^C X_i$ defind in { C }-frame to a point u_i on the image plane. Among many lens model describing image formation, we have adopted a Gaussian optics[11]. Therefore, a porjected location of ${}^C X_i$ on the image plane at i-th step can be computed as

$$u_i = (u_i \ , v_i)^T \tag{3}$$

$$= (f \cdot {}^C X_i/({}^C Z_i - f), f \cdot {}^C Y_i/({}^C Z_i - f))^T$$

Where ${}^C X_i = ({}^C X_i \ , {}^C Y_i \ , {}^C Z_i \ , 1)^T$, and f is a focal length of the camera.

In practical situation, we are used to compute u_i from the location on the digitized image frame instead of direct measurement[8]. Let U_i denote the location obtained from image processing and the coordinate of the lower left corner of the image frame be (0,0). Then the relation between u_i and U_i for noise-free case can be described as

$$U_i = (U_i \ , V_i)^T = h \ (\ {}^C X_i \ ) \tag{4}$$

$$= (K_u \cdot u_i + O_u \ , K_v \ , v_i + O_v)^T$$

where K_u and K_v are horizontal and vertical scale factor of the vision system. $(O_u \ , O_V)^T$ indicates the center coordinate of the image frame. For our system, 512 resolution, K_u , K_u , O_u, and O_v are determined to be 81.256, 75.638, 254, and 252, respectively.

3. 3-D POSITION

In this chapter we show to incorporate the system equation into an algorithm for estimating 3-D position from successive image frames obtained at different camera positions. Let us recall the observation model, and consider the effect of the measurement noise. Suppose that U_i denotes an image point computed from noisy and quantized images at i-th step, then

$$U_i = h(\ {}^C X_i) + n_i \tag{5}$$

where $n_i = (n_{u,i} \ , n_{v,i})^T$ represents a random noise vector. Let us assume that $n_{u,i}$ and $n_{v,i}$ have same probability desity function, but statistically independent, and successive noise terms are uncorrelated. So far, ${}^C X_i$ have been assumed to represent the unknown parameter. Since there are fewer measurements than parameter, it is, in general, not possible to determine ${}^C X_i$, without assigning additional structure to the behavior of the parameter.

Treating ${}^B X$ as the unknown parameter, the problem can be reduced to the non-linear time invariant form. With Eq. (2), Eq. (5) become

$$U_i = h(\ {}^C X_i) + n_i \tag{6}$$

$$= h(\ {}^W_C T^{-1} \cdot {}^B_W T^{-1} \ (\theta_i) \cdot {}^B X) + n_i$$

where θ_i is a joint vector with known value. Then, total system equation in which m measurements are available can be described as

$$\Omega_m = H_m \ (\ {}^B X) + N_m \tag{7}$$

$$\text{Where } \Omega_m = \begin{bmatrix} U_1 \\ U_2 \\ \cdot \\ \cdot \\ \cdot \\ U_m \end{bmatrix}$$

$$\text{and } H_M = \begin{bmatrix} h(\ {}^B X \ , \theta_2) \\ h(\ {}^B X \ , \theta_2) \\ \cdot \\ \cdot \\ \cdot \\ h(\ {}^B X \ , \theta_m) \end{bmatrix}$$

have 2mx1 dimension.

Now, we can summarize the problem as following; Suppose that the parameter ${}^B X$ described by Eq. (6) is related to m measurements Ω_m according to Eq. (7), find the best estimate of ${}^B X$ for given Ω_m in the sense of least square, i.e, find the value ${}^B X$ that minimize the cost function J_m

$$J_m = [\ \Omega_m - H_m(\ {}^B X) \ ]^T \cdot$$

$$Q \cdot [\ \Omega_m - H_m(\ {}^B X)] \tag{8}$$

Where Q means 2mx2m weighting matrix. Eq. (8) can be rewrited as

$$J_m = \sum_{i=1}^{m} w_{u,i} \ , (U_i - h_u(\ {}^B X \ , \theta_i))^2$$

$$+ \sum_{i=1}^{m} w_{v,i} \ , (V_i - h_v(\ {}^B X \ , \theta_i))^2 \tag{9}$$

In this paper, a rectangular weighting function is used. So $w_{u,i} = w_{v,i} = 1.0$, for all i.

4. FINDING A SOLUTION

As discussed in chapter3, 3-D position can be estimated by finding the minimum of the cost function. For very simple case, such as two views with lateral camera motion, a closed form solution can be derived[4]. But, in general, one is compelled to solve this problem numerically due to its nonlinearity. In this paper, SIMPLEX search technique[9] is used.

Let us inspect the figure of the cost function. Fig.2 shows a typical response surface of J around the extremum when { B } and { C } are parallel and just two views are available. It appears that J has a unique minimum and varies smoothly. The J for the case of 8 views is shown in Fig.3.

5. EXPERIMENTS

The overall experimental system includes a gray level video-rate frame grabber modeled as FD-5000 Gould with

512 by 512 resolution, a miniature CCD camera with 16 mm focal length, PUMA 560 manipulator equipped with its VAL controller, and PDP 11/34 mini-computer to process image data. It is remarked that the necessary data and command are transferred via an RS-232-C serial port.

Fig.4 shows a photograph of the robot system in operation. A typical scene of the blocks within work volume are displayed in Fig.5 We used three blocks - A,B, and C - whose top surfaces are triangle, trapezoid,a nd square. The height of A, B, and C are 10, 18, and 30 mm, respectively.

Firstly, we have performed an simple experiment to extract the distances between the camera per one step was 1.0 mm along X-axis of the Base Frame, and the locations of the vertex have been tracked and recorded as shown in Table-1. SIMPLEX minimiztion procedure was repeated to determine the distances between the objects and camera using from just 2 frames to entire 65 frames. Fig.6 and 7 show good convergence of the estimated values as the number of available data increase. The dotted, dashed, and solid lines represent the estimated distances of A, B, and C, respectively. After about 30 steps, the estimated Z distance looks quite reliable and distinguishable as shown in Fig.6 . It means that if camera moves more than 3 cm, we can obtain fairly accurate range data. Fig.7 shows the estimated values of the X and Y location of the object A with respect to the Base Frame. In general, the fluctuation of the X and Y data are much less than that of Z data. The initial searching position of this procedure is around $^{B}(0,0,300)$, and the search can be terminated when the change of the value of the cost function for successive steps becomes less than 10^{-4}. It is interesting to note that almost all searchs are terminated within around 100 iterations. Fig.8 shows a example search which is terminated at 139-th step. 't' marks in Fig.9 represent the results of stereometric calculations[10]. The estimated distances derived by averaging operation[11] and proposed algorithm are represented by dotted and solid lines, respectively.

Next, the robot manipulator was programmed to move 5.0 mm per one step along Z-axis of the Base Frame. In this case, the focus of the camera was adjusted manually because of no automatic focusing mechanism in our camera assembly. After 15 steps, 7.5 cm, a quite reliable result can be obtained as shown in Fig.10 from the measured data as Table-2.

6. FURTHER RESEARCHES

We have presented a 3-D position estimation scheme using multiple measurements for Eye-in-hand Robot Vision System. Though proposed algorithm with kinematic and obsevation model provides a rather reliable and accurate data, a more exact noise modeling, which describes a probabilistic behavior of image formation and iamge processing, is an area for further research. We are studying a better algorithm for position estimation, which containing a Gaussian error model and lens abberation, together with some probabilistic simulations to compare the performance of each approachs.

Further, major aim of our system is not only to extract a 3-D position or to show the efficacy of this approach but to control the robot motion along desired path without aids of external position data. For this, we are also studying a real-time image tracking algorithm using a hard-wired vision processor and a effective servoing strategy.

REFERENCES

[1] W.T.Miller, "Sensor-Based Control of Robot Manipulator Using a General Learning Algorithm", IEEE J Robot, and Automat., vol. RA-3, No.2, April 1987.

[2] R. Nevatia, "Depth Measurement by Motion Stereo", Computer Graphics Image Processing 5, pp.203-214, 1976.

[3] T.D.Williams, "Depth from Camera Motion in Real World Scene", IEEE Trans, Pattern Anal. Machine Intell., vol. PAMI-2, No.6, pp.511-516, Nov. 1980.

[4] R.O.Duda and P.E.Hart, PATTERN CLASSIFICATION AND SCENE ANALYSIS, Wiely, 1973, Chap. 10.

[5] R.A.Jarvis, "A Perspective on Range Finding Techniques for Computer Vision", IEEE Trans, Pattern Anal., Machine Intell., vol. PAMI-5, No.2, pp.122-139, Mar. 1983.

[6] J.J.Craig, Introduction to Robotics, Addison-wesley, 1986, Chap.3.

[7] M.Born and E.Wolf, Principles of Optics, 6-th ed., Pergamon Press, 1980, Chap.4.4.

[8] S.Y.Nof, Handbook of Industrial Robotics, John Wiely and Sons, 1985, Chap. 16.

[9] R.W.Daniels, An Introduction to Numerical Method and Optimization Techniques, North-Halland, 1987, Chap. 8.

[10] E.S.McVey and J.W.Lee, "Some Accuracy and Resolution aspects of Computer Vision distance Measurements", IEEE Trans., Pattern Anal., Machine Intell., vol.PAMI-4, No.6, pp.646-649, Nov. 1982.

[11] J.Amat, A.casals and V.Llario, "Improving Accuracy and Resolution of a Motion Stereo Vision System", Proc., IEEE Int. Conf. Robotics and Automation., pp.634-638, 1986.

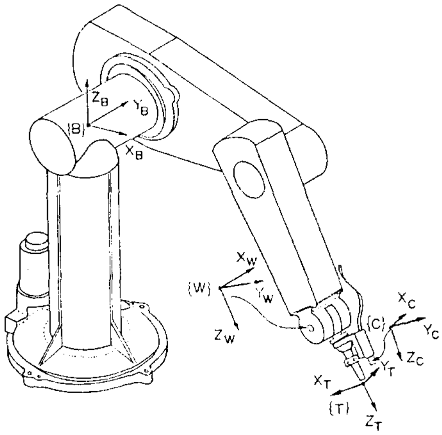

Fig.1.Relevant frames of PUMA 560 robot system and hand-held camera.

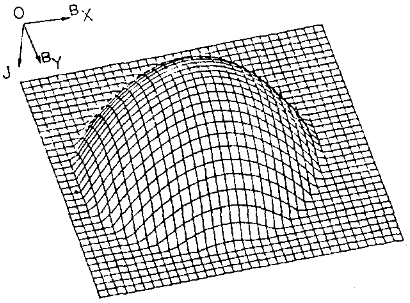

Fig.2.Inverted plot of the response surface(2-view). The best values for X and Z are those where the J value is the lowest.

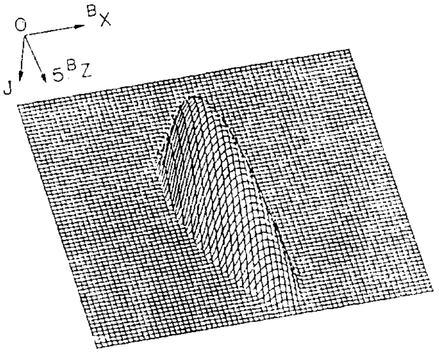

Fig.3.Representation of the response surface(8-view case).

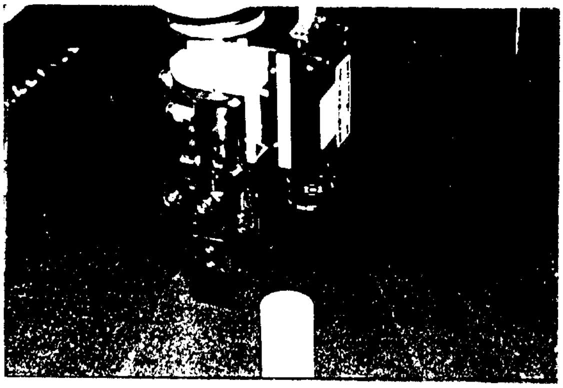

Fig.4.Photograph of the experimental robot manipulator and hand-held camera in operation.

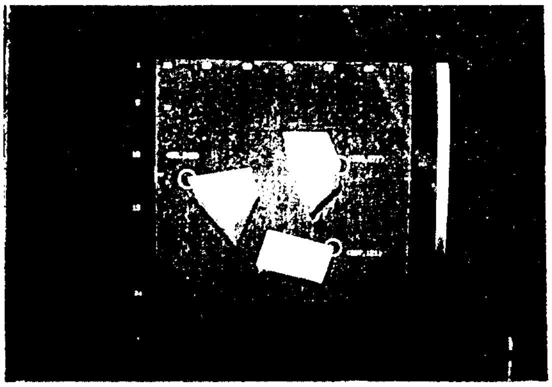

Fig.5.Typical scenes of the blocks and tracked feature points.

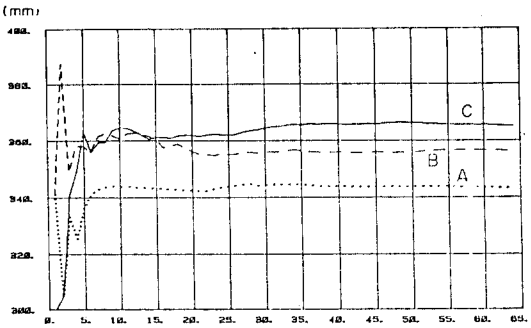

Fig.6.Estimated Z distances versus number of the measurements (lateral motion). The distance of the work table on which the blocks are placed is 375.0 mm.

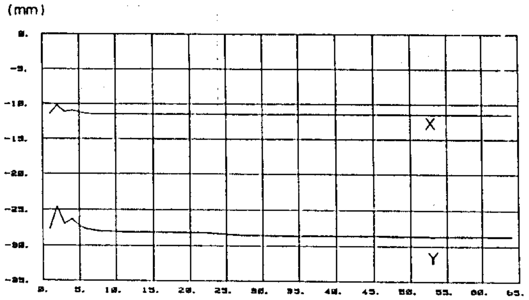

Fig.7.Estimated X and Y value versus number of data.

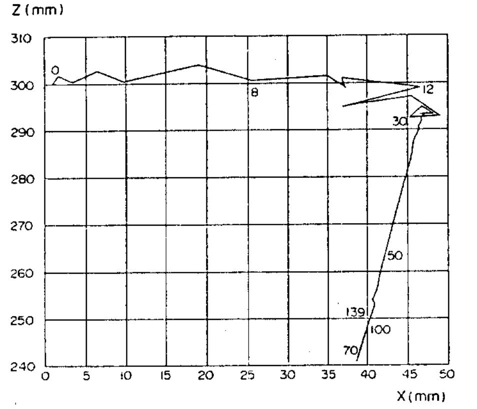

Fig.8.An example SIMPLEX searching path terminated at 139-th step. The initial point is (0,0,300).

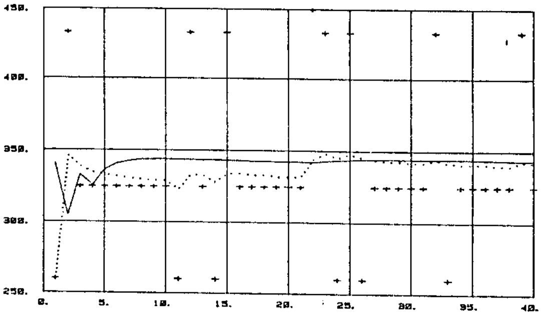

Fig.9.Estimated distances with stereometric and proposed algorithm.

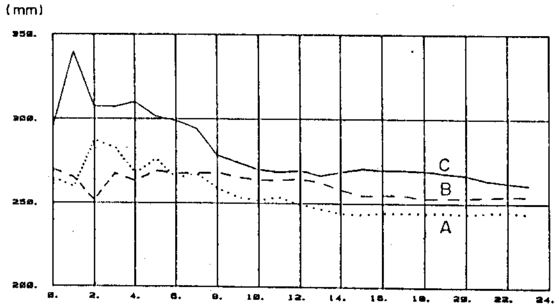

Fig.10.Imaging geometry of motion stereo showing measurement uncertainty.

Table-1.Measured locations of the vertex of A, B, and C at successive steps during lateral camera motion.

steps	A	B	C
1	(464,148)	(496,299)	(267,204)
2	(459,148)	(493,299)	(264,204)
3	(456,148)	(489,299)	(261,205)
4	(452,148)	(485,299)	(257,204)
5	(448,148)	(481,299)	(252,204)
⋮	⋮	⋮	⋮
64	(213,151)	(256,302)	(33,208)
65	(209,151)	(252,302)	(28,207)

Table-2.Measured locations of the vertex of A, B, and C during approaching motion.

steps	A	B	C
1	(324,150)	(361,291)	(152,210)
2	(326,147)	(364,290)	(151,208)
3	(327,145)	(365,291)	(150,208)
4	(326,145)	(366,292)	(147,207)
5	(328,142)	(368,292)	(145,207)
⋮	⋮	⋮	⋮
24	(357, 99)	(410,308)	(104,190)
25	(359, 96)	(423,309)	(102,189)