LOCATION OF PLANAR TARGETS IN THREE SPACE FROM MONOCULAR IMAGES

Karin Cornils and Plesent W. Goode
NASA Langley Research Center

Presented at the 1987 Goddard Conference on Space Applications of Artificial Intelligence (AI) and Robotics

Greenbelt, Maryland
May 13-14, 1987

LOCATION OF PLANAR TARGETS IN THREE SPACE FROM MONOCULAR IMAGES

By

Karin Cornils
Plesent W. Goode
NASA Langley Research Center

ABSTRACT

Many pieces of existing and proposed space hardware that would be targets of interest for a telerobot can be represented as planar or near-planar surfaces. Examples include the biostack modules on the Long Duration Exposure Facility, the panels on Solar Max, large diameter struts, and refueling receptacles. Robust and temporally efficient methods for locating such objects with sufficient accuracy are therefore worth developing.

Two techniques that derive the orientation and location of an object from its monocular image are discussed and the results of experiments performed to determine translational and rotational accuracy are presented. Both the "quadrangle projection" and "elastic matching" techniques extract three space information using a minimum of four identifiable target points and the principles of the perspective transformation. The selected points must describe a convex polygon whose geometric characteristics are prespecified in a data base.

The rotational and translational accuracy of both techniques was tested at various ranges. This experiment is representative of the sensing requirements involved in a typical telerobot target acquisition task. Both techniques determined target location to an accuracy sufficient for consistent and efficient acquisition by the telerobot.

INTRODUCTION

Simple and computationally efficient methods for locating targets in 3-space are necessary for real-time automatic control of manipulators. One class of techniques having application to a broad range of sensor-based control problems is that of four point location algorithms. Placing identifiable points on space hardware to enhance it as a manipulator target is feasible. The targets are man made objects whose components and structural measurements are well-documented. Typical points that could be extracted are those that can be derived from the moments of the planar or near-planar surfaces of bolt heads, fueling receptacles and large-diameter struts. The perspective projections of these points through a lens system onto an image sensor can be compared to their known interdistances, and the location of the object on which they lie relative to the sensor can be determined. The image sensor used in both the referenced studies and in this study is the solid-state camera.

Previous work includes a closed form solution developed by Haralick [1] that assumes a rectangular configuration of the four points. Implementation of Haralick's Algorithm in real-time robot vision systems has been accomplished at the NASA Langley Research Center, The National Bureau of Standards, and Martin Marietta Denver Aerospace. Results using this algorithm were reported by Myers et al [2] and Wolfe et al [3]. Hung et al [4] developed an algorithm that directly computes the 3-D coordinates of the vertices of a quadrangle relative to the camera frame. Goode and Cornils [5] adapted the theory developed by Hung, Yeh, and Harwood to the real-time control of manipulators. An algorithm developed by Goode [6] using four or more points to approximate a convex shape to determine a target's orientation and location was applied to closed-loop manipulator control. This paper summarizes the techniques developed in [5] and [6], and reports the results of an experiment designed to determine the rotational and translational accuracy of the two methods.

Two Four-Point Algorithms

The objective of both methods is to resolve the three dimensional location of objects having planar or minimally curved surfaces relative to the camera's axis frame. The camera's axis frame, for the purpose of the following discussion, is defined in figure 1. It is rotationally coincident with the frame of the manipulator's end effector, but translationally offset by (-15, 80, -190) millimeters (figure 1). Also, equations are presented in sufficient detail to allow implementation, but are summarized without extensive derivation. A complete reference list is provided for further investigation of equation development.

The quadrangle projection method determines the location and orientation of a planar or near-planar object from any four points on the object that describe a quadrangle. Given the inter-vertex distances of the quadrangle and the optical parameters of the lens/camera system, the rotational and translational displacements between the object and camera can be uniquely determined. Hung et al [4] prove that there exists a unique vector, K, which relates the target quadrangle and its image such that

$$\langle T \rangle = K \langle I \rangle \qquad (1)$$

where <I> is the quadrangle <I0, I1, I2, I3> that is the projection of the target quadrangle <T0, T1, T2, T3> on the image plane (figure 2). Each vertex Tj (j = 0,1,2,3) of the quadrangle <T> is a three component vector (tjx, tjy, tjz) representing the three dimensional location of the vertex. Each vertex Ij (j = 0,1,2,3) of the quadrangle <I> is a three component vector (ijx, ijy, ijz) representing the two dimensional location of the target's projection on the image plane and the distance of the image plane from the camera (i.e. the focal length of the lens). The K vector, (k0, k1, k2, k3), can be found using the following system of equations:

$$I3 = (k0/k3)(1-alpha-beta)(I0) + (k1/k3)(alpha)(I1) + (k2/k3)(beta)(I2) \qquad (2)$$

which can be solved for k0/k3, k1/k3, and k2/k3. The component k3 is computed from:

$$k3 = ||T0 - T3||/||(k0/k3)(1 - alpha - beta)(I0) - I3|| \quad (3)$$

For each probable target it is necessary to determine and specify the alpha and beta parameters based upon the inter-vertex distances of the target quadrangle. Let P0, P1, P2 and P3 be the two dimensional coordinates of the target quadrangle's vertices relative to the target's reference frame. Then:

$$alpha = -((p0x)(p3y - p2y) + (p2x)(p0y - p3y) + (p3x)(p2y - p0y))/D(P)$$

$$beta = ((p0x)(p3y - p1y) + (p1x)(p0y - p3y) + (p3x)(p1y - p0y))/D(P) \quad (4)$$

where

$$D(P) = (p0x)(p2y - p1y) + (p1x)(p0y - p2y) + (p2x)(p1y - p0y)$$

This information is sufficient to solve for the three dimensional positions of the target quadrangle vertices relative to the camera frame. The quadrangle orientation, described by the normal to the plane occupied by the quadrangle, is determined by substituting the coordinates of any three vertices into the general equation of the plane. Solving the system of equations gives the following explicit expressions for the orientation vector in terms of the quadrangle vertices derived above:

$$Ax' = ((t1y)(t2z) - (t1z)(t2y) + (t0z)(t2y) - (t0y)(t2z) + (t0y)(t1z) - (t0z)(t1y))/D(T)$$

$$Ay' = ((t1z)(t2x) + (t1x)(t2z) + (t0x)(t2z) - (t0z)(t2x) + (t0z)(t1x) - (t0x)(t1z))/D(T)$$

$$Az' = ((t1x)(t2y) - (t1y)(t2x) + (t0y)(t2x) - (t0x)(t2y) + (t0x)(t1y) - (t0y)(t1x))/D(T)$$

where (5)

$$D(T) = (t0x)((t1y)(t2z) - (t1z)(t2y)) + (t0y)((t1z)(t2x) - (t1x)(t2z)) + (t0z)((t1x)(t2y) - (t1y)(t2x))$$

and Ax, Ay, and Az are determined from Ax', Ay', and Az' by normalizing by the magnitude of the vector (Ax', Ay', Az'). This vector along with three others comprise an homogeneous transform matrix called an NSAP matrix [7], [8]. This matrix completely describes the target's location in the camera's axis frame. The approach vector (Ax, Ay, Az) is the orientation vector derived above. The sliding vector (Sx, Sy, Sz) is related to the slope of the base of the quadrangle with respect to the camera frame. It is composed of the x, y, and z components of the vector, T1 - T0, normalized by its length. The vector, (Nx, Ny, Nz), is the cross product of the approach and sliding vectors. The position vector, (Px, Py, Pz), is simply the components of the selected point of approach on the target quadrangle. The intersection of the diagonals is commonly chosen.

The second method is based on the elastic matching [9] approach to pattern recognition and has application to shape decomposition, object recognition, and object location. It is an adaption of the linear programming technique of goal programming to the nonlinear problem of elastic matching [6]. Conceptually, elastic matching can be explained by envisioning a transparent reference image

overlaying a goal image. The reference image is then warped or distorted to conform to the goal image by locally matching corresponding regions in the two images. The reference image is a flexible template that is modelled as a system of equation pairs where each equation pair represents a linear combination of patterns that a point in the reference image can describe in moving to a point in the goal image (figure 3). The amount of displacement each pattern contributes to the distortion is determined by identifying the values of the parameters, Ai and Bi, associated with each of the distortion patterns. The parameter values are derived by minimizing the absolute differences between corresponding reference and goal image points without violating the pattern constraints. This type of problem is easily modelled mathematically using the linear programming technique of goal programming [10]. The computational procedure of the Simplex Algorithm most efficiently resolves the optimal values of the model's parameters.

The elastic matching technique has been used to recognize objects with planar or minimally curved surfaces and to locate them in three dimensions [6]. The discussion here concerns the location of an object, once it has been recognized, and four or more points of known geometric relationship extracted from its image. The object is represented in a data base as a reasonably convex set of points whose values describe the object in an orientation and location normal to and centered on the optical axis of the camera, and a distance equal to the focal length of the lens, along the axis. This is the distorted reference image used to match the extracted image. The three dimensional location of the target object can be derived from the parameters, A0 through A3 and B0 through B3. Equations (6) through (9) show the geometric significance of the parameters.

$$A0 = X' - X \qquad : \text{translation} \tag{6}$$

$$B0 = Y' - Y$$

$$A1 = -(1 - \text{gain}) \qquad : \text{zoom} \tag{7}$$

$$B1 = -(1 - \text{gain}) \qquad \text{gain} = X'/X \text{ or } Y'/Y$$

$$A2 = (X' - X)/Y \qquad : \text{rotation about z (the} \tag{8}$$

$$B2 = (Y' - Y)/X \qquad \text{optical axis)}$$

$$A3 = -(1 - \text{gain})/Y \qquad : \text{perspective information} \tag{9}$$

$$B3 = -(1 - \text{gain})/X$$

Parameters A4, A5, B4, and B5 yield shape information and are used to aid object recognition [6]. Equations (10) through (12), which are based on properties of the perspective transformation [11], show the parameters' relationship to the range, rotation about y, and rotation about x respectively of the target object relative to the camera's axis system.

$$\text{range} = ((f)(Wo)(2 - A1))/((1 - A1)(Ws)) \tag{10}$$

where f is the focal plane distance of the camera/lens system, Wo is the target width, and Ws is the camera's image sensor width,

$$\tan Ry = (f)(A3)/(A1) \tag{11}$$

where Ry is the rotation about the y axis (pitch), and

$$\tan Rx = (f)(B3)/(B1) \tag{12}$$

where Rx is the rotation about the x axis (yaw).

The construction of the reference image model in the goal programming format is detailed in [6] and [10].

Test Apparatus

The experimental test fixture consists of an optical bench, a six degree-of-freedom articulator, a planar target, and a solid-state camera. The basic concept was to construct a stable system with enough flexibility to accomodate a range of easily-measured rotations and translations about a common point, simultaneously sampling and storing the results. The camera is fixed and the articulator is rigidly mounted at either of two range settings such that, at the zero initialization position, the z axis through the center of the target board and the focal axis of the camera coincide. The orientations of the articulator are set up to rotate about the point described by the intersection of the target plane in its initial position and the focal axis of the camera. Following calibration at each of the range settings, the angular and translational displacements can be dialed in with precisions of 1/1000 of an inch and 1/360th degree on the articulator.

The target consists of four white points of 0.24 inches (6 mm) radius, mounted on a dark background, forming the cornerpoints of an isosceles trapezoid with bases of 4.6 inches (117.5 mm) and 3.0 inches (76 mm) and height of 4.1 inches (105 mm). The camera is a solid-state, CCD, light-sensing system that outputs RS 170 standard video. The camera has a spatial resolution of 384 x 491, a 45 dB signal-to-noise ratio, and is fitted with a 0.63 inches (16 mm) focal length lens. However, the spatial resolution and intensity range limits reside with the image processor's 320x240 pixel image memory and four bit (16 shades) gray level. Computation is performed on a microcomputer with the following capability: 16 bit word size, a program memory capacity of 64 kilobytes, and a data memory of 256 kilobytes.

Test Procedure

The experiment is conducted with the target mounted at a distance of one meter from the camera and then at one-half meter from the camera. To isolate the response of the vision system to the various rotations, the initial experiment involves rotating the target about each of the axes individually.

For each axis, the target is rotated through a range of plus-or-minus 60 degrees, beyond which processing becomes impractical. The rotations are sampled at 10-degree intervals and the translations at .025 inches (1 mm) intervals over a range of plus-or-minus 0.3 inches (7.6 mm). The translations are taken at these intervals to give an indication of the response to translational displacements of 1-millimeter increments. The experiment is repeated at one-half meter, both to test the response of the system to variation in image and target point size, and to indicate response to a large scale differential translation. The target's point coordinates are sampled 30 times for each displacement and processed by both location algorithms. The translational and rotational solutions of both algorithms are then processed to find the mean, standard deviation, and confidence limits of each calculated displacement.

In order to determine the effect of rotational and translational errors in combination, an experiment in vision driven acquisition of a cylindrical strut was conducted. The camera, mounted on the end effector of a six degree-of-freedom manipulator (figure 1), and the strut were placed in random orientations relative to each other. The location algorithms were then used to correct the trajectory of the end effector and update the orientation and location of the strut until the strut was acquired by the end effector. Each acquisition sequence was initiated at a distance of approximately 0.75 meters between the camera lens and the strut.

CONCLUSION AND RESULTS

Two four-point location algorithms have been discussed and an experiment to determine their accuracy has been described. Results are displayed in the graphs of figures (4) and (5). The displacement quantities are presented in degrees and inches because the articulator was gradated in those units of measurement.

The results obtained contain the effects of errors inherent in the vision system and the test apparatus. The primary error sources are the spatial resolution of the image acquisition and processing subsystem, and the size of the target points. The uncertainty of the location of a target point is controlled by the 0.35 inches (8.8 mm) by 0.26 inches (6.6 mm) image sensor area and the 320 horizontal by 240 vertical pixel array of the image processor's image memory. These dimensions determine that a one-pixel or 0.001 inches (.0275 mm) change in sensor image corresponds to a 0.03 inches (0.8 mm) uncertainty of target point location in the x-y plane, given a target distance of 0.5 meter and a lens focal length of 0.63 inches (16 mm). Range uncertainty is 0.13 inches (3.3 mm) and rotation uncertainty about the x and y axes is 1.8 degrees with the target used. Rotation about the z axis is sensitive to 0.44 degrees of displacement.

The quadrangular projection technique produces consistent results accurate to the inherent error of the system. The averaged absolute translational

errors are 0.072 inches (1.8 mm) and 0.012 inches (0.3 mm) at 1.0 and 0.5 meters respectively, and the mean absolute rotational errors are 0.86 and 0.44 degrees at those distances. The instantaneous accuracies can be expected to fall within the confidence limits computed for each displacement. The worst case standard deviation for rotational displacements is 1.96 degrees at 0.5 meter and 2.8 degrees at 1.0 meter. The maximum standard deviations for translational displacements are 0.02 inches (0.51 mm) at 0.5 meter and 0.06 inches (1.5 mm) at 1.0 meter.

The elastic matching technique data was collected at the 0.5 meter distance only. The mean rotational error using the elastic matcher was 3.3 degrees and the worst case standard deviation was 4.5. There were differences between the actual and computed rotations of as much as 7 degrees. The translations computed using the elastic matcher were as accurate as those derived using the quadrangular projection technique. The primary reason for the inconsistency of the rotational performance of the elastic matcher is the small number of points involved in the match. The matcher actually approximated the quadrangle quite well, with each model point displaced to within two or three pixels of the actual image point, but did not always identify the appropriate parameter for the geometric condition. Constraints were lacking to ensure a unique solution. The consistency of the matcher is directly proportional to the number of points being matched, consistently demonstrating rotational accuracy when 16 points are being matched. It also performed well in a manipulator servo experiment [6] which involved accurate positioning of the manipulator tool with respect to a rectangular target.

Both techniques have successfully driven an end effector to acquire a target to an accuracy of 0.5 degree rotational error and 0.02 inch (.5 mm) translational error at acquisition. This accuracy is consistent regardless of the relative orientation of the camera/end effector and the target at the initiation of the acquisition sequence. Combinations of rotational displacement have little effect on the ultimate accuracy of these techniques when they are used in a servo mode. A more rigorous experiment is now being devised to test the algorithm accuracies at both one meter and at one-half meter.

The quadrangular projection method is a more consistent location technique than the elastic matcher when the target consists of four points. Accuracies to within 1 degree and 0.1 inch (2.54 millimeters) at 1 meter are more than sufficient for accurate closed-loop control of a manipulator. When used in a servo mode, the measurement accuracy and noise immunity of both methods increases as the manipulator approaches the target.

REFERENCES

1. Myers, D. R., Juberts, M., and Leake, S. A.: "Enhanced Telemanipulator Operation using a Passive Vision System," in Proceedings of IEEE Conference on Man and Cybernetics, Tucson, Arizona, Nov. 1985.

2. Haralick, R. M., "Determining Camera Parameters from the Perspective Projection of a Rectangle," Technical Note, Virginia Polytechnic Institute, Blacksburg, Virginia, June 1982.

3. Wolfe, W. J., "A Vision System for Recognizing and Tracking Known Objects," in Proceedings of JPL Space Telerobotics Workshop, Pasadena, California, January 20-22, 1987.

4. Hung, Y., Yeh, P., Harwood, D., "Passive Ranging to Known Planar Point Sets," Proceedings of the IEEE International Conference on Robotics and Automation, pp. 80-85, 1985.

5. Goode, P. W., Cornils, K.: "Monovision Techniques for Telerobots," in Proceedings of JPL Space Telerobotics Workshop, Pasadena, California, January 20-22, 1987.

6. Goode, P. W.: "A Multifunction Recognition Operator for Teleroootic Vision," Proceedings of the AIAA Guidance, Navigation, and Control Conference, August, 1986.

7. Paul, R. P.: Robot Manipulators: Mathematics, Programming, and Control, MIT Press, Cambridge, Massachusetts, 1981.

8. Lee, C.S.G.: "Robot Arm Kinematics, Dynamics, and Control," IEEE Computer, Vol 15, No. 12, December 1982.

9. Widrow, B.: "The Rubber Mask Technique," Pattern Recognition, Vol 5, 1973, pp. 175-211.

10. Hillier, F. S. and Lieberman, G. J., Introduction to Operations Research, 3rd edition, Holden-Day, San Francisco, 1980.

11. Haralick, R. M., "Using Perspective Transformations in Scene Analysis," Computer Graphics and Image Processing, 13, 191-221, 1980.

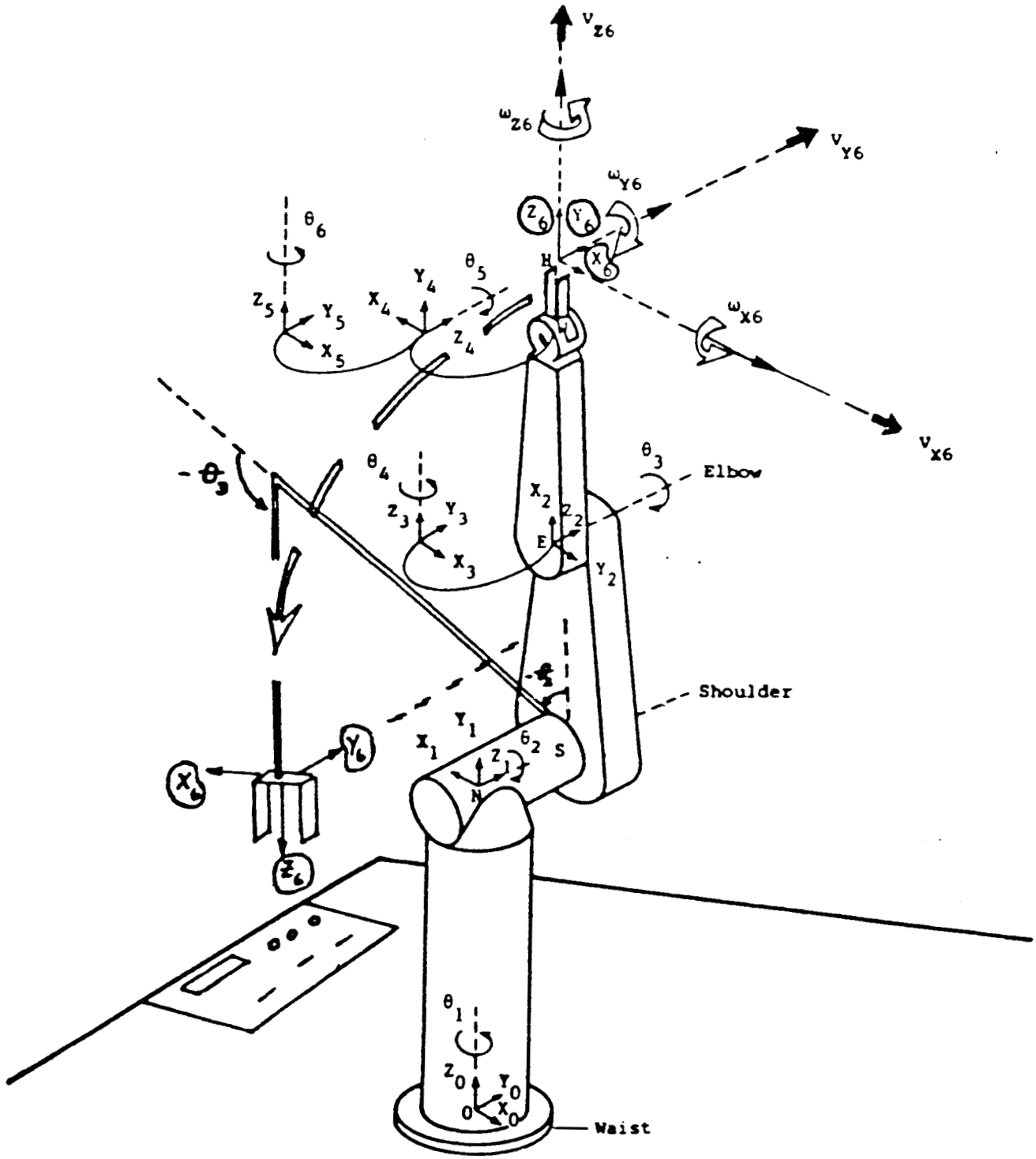


Figure 1. - Manipulator Axis Frames

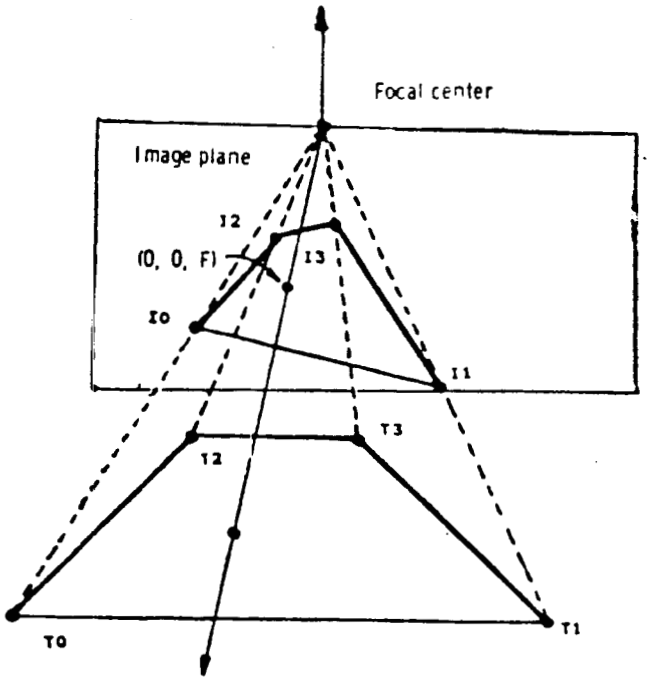


Figure 2. - Quadrangle Projection

A_0, B_0 : Translation

A_1, B_1 : Gain

A_2, B_2 : Rotation in X-Y plane

A_3, B_3 : Perspective of triangular shape information

A_4, B_4 : Semicircular shape information

A_5, B_5 : Elliptical shape information

$f(x, y)$: Model function

$f(x', y')$: Image function

$$X + A_0 + A_1X + A_2Y + A_3XY + A_4(X^2 - Y^2) + A_5XY^2 = X'$$

$$Y + B_0 + B_1Y + B_2X + B_3XY - B_4(Y^2 - X^2) + B_5YX^2 = Y'$$

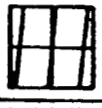
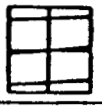
Distortion pattern	Distortion term
	XY
	XY^2
	Y
	Y^2-X^2
	YX^2
	X
	X^2-Y^2

Figure 3. - Elastic Matcher

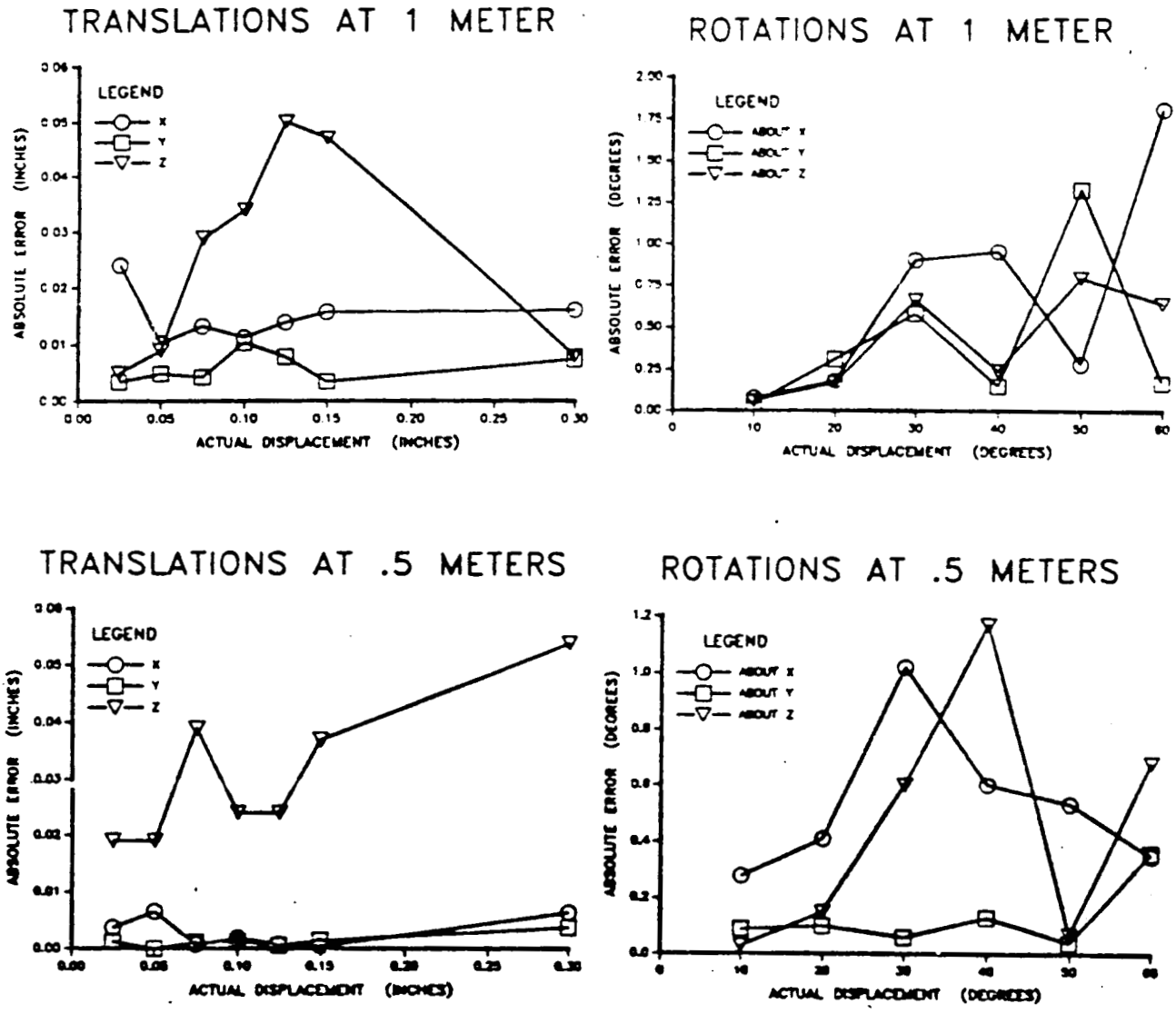
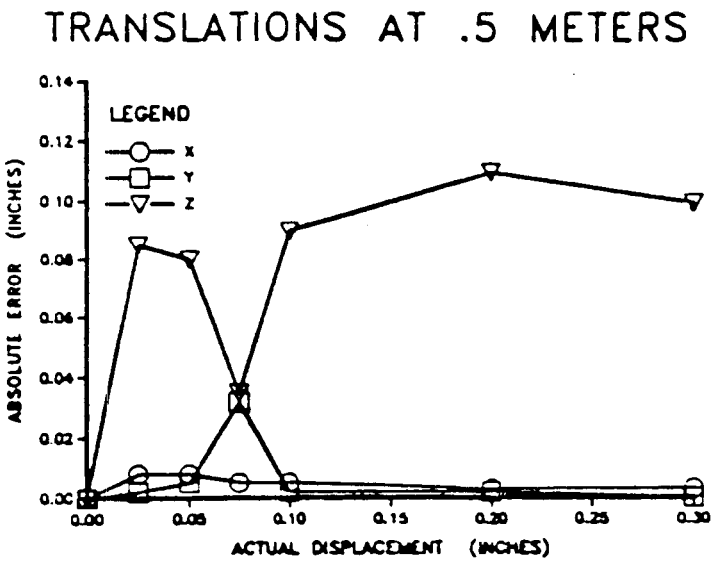


Figure 4. - Displacement Error; Quadrangle Projection Algorithm



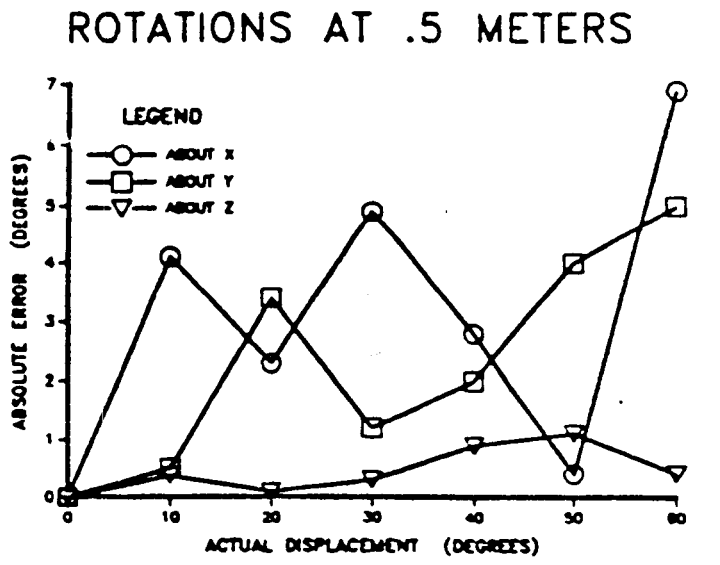


Figure 5. - Displacement Error; Elastic Matcher Algorithm