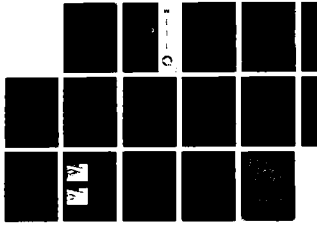
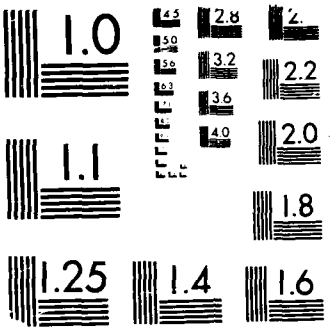
AD-A193 890 3-D ROAD STRUCTURE FROM MOTION STEREO(U) MARYLAND UNIV 1/1
COLLEGE PARK CENTER FOR AUTOMATION RESEARCH M ASADA
APR 87 CAR-TR-286 ETL-0471 DACA76-84-C-0004

UNCLASSIFIED F/G 17/5 NL





ETL-0471

AD-A193 090

(4)

3-D Road Structure from Motion Stereo

Minoru Asada

Center for Automation Research
University of Maryland
College Park, MD 20742-3411

April 1987

Prepared for
U.S. ARMY CORPS OF ENGINEERS
ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060-5546

88

E T L

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS N/A
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) CAR-TR-286 CS-TR-1839	5. MONITORING ORGANIZATION REPORT NUMBER(S) ETL-0471
6a. NAME OF PERFORMING ORGANIZATION University of Maryland 6b. OFFICE SYMBOL (If applicable) N/A	7a. NAME OF MONITORING ORGANIZATION Army Engineer Topographic Laboratories
6c. ADDRESS (City, State and ZIP Code) Center for Automation Research College Park, MD 20742-3411	7b. ADDRESS (City, State and ZIP Code) Fort Belvoir, VA 22060-5546
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Defense Advanced Research Projects Agency 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DACA76-84-C-0004
8c. ADDRESS (City, State and ZIP Code) 1400 Wilson Blvd. Arlington, VA 22209	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 62301E PROJECT NO.: TASK NO.: WORK UNIT NO.:
11. TITLE (Include Security Classification) 3-D Road Structure from Motion Stereo	
12. PERSONAL AUTHOR(S) Minoru Asada	
13a. TYPE OF REPORT Technical 13b. TIME COVERED FROM _____ TO N/A	14. DATE OF REPORT (Yr., Mo., Day) April 1987 15. PAGE COUNT 29
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	
17	07		Computer Vision, Motion Stereo, 3D Shape Recovery
17	08		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This paper presents a new method for reconstructing the 3-D structure of road boundaries from consecutive images. First, we present a method for estimating depth information by applying a motion stereo method to consecutive images, given an estimate of the interframe motion. The relation between depth, motion and disparity is investigated, since the accuracy of the depth depends on the disparity range. Next, the error of the estimated road structure due to quantization errors and motion estimation errors is examined. Finally, a representation for road boundaries is proposed that makes explicit the error of the road edge location in 3-D space. Experimental results are shown for an input image sequence taken by the ALV simulator robot in the Center for Automation Research at the University of Maryland.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Rosalene Holecheck	22b. TELEPHONE NUMBER (Include Area Code) (202) 355-2767 22c. OFFICE SYMBOL CEETL-RI-T

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

1. INTRODUCTION

There exist a variety of 3-D shape recovery techniques in computer vision based on photometric constraints, stereoscopy, pattern regularity, surface structure constraints and motion information [1]. These methods should play a key role in robotics applications as the complexity of robot tasks increases. In systems such as visual navigation systems [2–5], some assumptions or constraints are needed to accurately obtain 3-D information since "3-D shape from ...." methods are very sensitive to noise. In man-made environments, the flatness of the floor and prevalence of vertical edges can be used to infer the 3-D motion of the robot [2], or knowledge of the camera attitude can be used to recover 3-D structure using stereo vision [3]. Many outdoor robots have used monocular inverse perspective techniques to recover the 3-D road structure. They employed assumptions about, for example, road width, camera position and smoothness of the road surface [4, 5]. Those methods are very sensitive to even small deviations. Little attention has been directed to the problem of estimating the effect of various errors (calibration, discrepancies between model assumptions and true conditions, etc.) on 3-D reconstruction [6].

In this paper, we propose a new method for reconstructing the 3-D structure of road boundaries from consecutive images. First, we present a method for estimating depth information; this is a motion stereo method applied to consecutive images, given an estimate of the interframe motion. The relation between depth, motion and disparity is investigated, since the accuracy of the depth depends on the disparity range. Next, the error of the estimated road structure

due to quantization errors and motion estimation error is examined. Finally, a representation for road boundaries is proposed that makes explicit the error of the road edge location in 3-D space. Experimental results are shown for an input image sequence taken by the Autonomous Land Vehicle (ALV) simulator robot [5].

2. THE PRINCIPLE OF MOTION STEREO USED IN THE METHOD

Our motion stereo analysis is based on the following assumptions.

(1) The road boundaries in consecutive images are detected and the correspondence between frames is established.

(2) The robot motion is known.

Algorithms [5, 7] have been developed for road boundary detection in images. Our program uses the output of one such algorithm. Also, given the interframe motion, determining the correspondence between road boundaries in consecutive frames is straightforward. Internal sensors can be employed for accurately estimating the motion of the robot.

We outline the depth recovery algorithm as follows. Given the interframe motion, the epipolar line in the second image corresponding to any point P in the first image can be determined by projecting the epipolar plane determined by P and the two camera centers onto the second image [8]. We can, however, determine this epipolar line more directly. The known motion consists of a translation $T=(U,V,W)$ and a rotation $R=(\alpha,\beta,\gamma)$. The displacement vector of a feature point in the first frame consists of two vectors t (translation) and r (rotation) in

the image plane. While r is uniquely determined when R is known, t has one degree of freedom—length. The direction of t is constrained so as to be toward the focus of expansion (the intersection of T with the image plane [9]). Thus, we can determine the epipolar line as follows:

(1) Transform the feature point P_1 in the first frame by the rotation r. The displaced location is denoted by P_1' (see Figure 1).

(2) Draw a line from P_1' to the FOE.

(3) The line segment P_1'-FOE is the desired epipolar line. The corresponding point in the second frame must lie on this line segment (from P_1' to FOE) when $W<0$ or on the half line from P_1' to Q (see Figure 1) when $W>0$.

This method for constructing the epipolar line does not require calculation of the equation of the epipolar plane in 3-D space. Furthermore, the constraint (3) is not explicit in previous methods [8].

The translation vector $t=(d_x,d_y)$ is easily determined by finding the intersection between the epipolar line segment and the road boundary in the second frame (see Figure 1). The 3-D coordinates of the feature point can be recovered from the translation vector (d_x,d_y). For simplicity, we assume that the first frame has already been corrected for the known rotation R; therefore, only the translation (U,V,W) needs be considered. The camera translation (U,V,W) can be interpreted as a translation $(-U,-V,-W)$ of the feature point with respect to a fixed camera. Therefore, the coordinates of the feature point (x_1,y_1) in the first frame and (x_2,y_2) in the second frame are given as

$$x_1=\frac{fX}{Z},\ y_1=\frac{fY}{Z},\ x_2=\frac{f(X-U)}{Z-W},\ \text{and}\ \ y_2=\frac{f(Y-V)}{Z-W} \tag{2.1}$$

where f denotes the focal length of the camera and (X,Y,Z) represent the 3-D coordinate of the point in the first camera coordinate system. From eqns(2.1)

$$d_x=x_2-x_1=-\frac{fU}{Z-W}+\frac{fXW}{Z(Z-W)}$$

and

$$d_y=y_2-y_1=-\frac{fV}{Z-W}+\frac{fYW}{Z(Z-W)} \tag{2.2}$$

Finally, we obtain the 3-D coordinates:

$$Z=-\frac{f\,U}{d_x}+W(\frac{x_1}{d_x}+1)\ \text{ or }\ Z=-\frac{f\,V}{d_y}+W(\frac{y_1}{d_y}+1) \tag{2.3}$$

and

$$X=\frac{x_1Z}{f},\qquad Y=\frac{y_1Z}{f},$$

The recovery of road structure from motion stereo does not, of course, require prior knowledge of camera height, tilt angle or road width as do most monocular inverse perspective methods.

3. PROPERTIES OF THE DISPARITY FOR A MOBILE ROBOT

We consider how the disparity relates to the tilt angle and the motion. Figure 2 shows the parameters of our computer simulation. The image size is 512×480. The visual angle of the camera is approximately 33 degrees, corresponding to about 800 pixels of equivalent focal length. The tilt angle of the

camera is approximately 23 degrees. The camera height is approximately 50 mm; we use this length as unit length in the simulations. In the following, we set $U=V=0$ to simplify the analysis; these would be zero if the robot moved straight on a flat surface.

(a) **Disparity vs. Tilt Angle**

When $U=V=0$, the disparity (d_x, d_y) is given by

$$d_x=\frac{fXW}{Z(Z-W)}=\frac{x_1W}{Z-W},\quad d_y=\frac{fYW}{Z(Z-W)}=\frac{y_1W}{Z-W}.$$

Here we need consider only the disparity d_y because the tilt angle does not affect X and U. If we assume a flat surface and no tilt ($\theta=0$), $Y=h$ (camera height); therefore

$$d_y=\frac{fhW}{Z(Z-W)}. \tag{3.1}$$

If the camera has a tilt angle θ, then the camera has an apparent translation component in the direction of the y-axis, although the robot has no translation component in the vertical direction ($V=0$); therefore, eqn(3.1) becomes

$$d_y=\frac{fhW}{(h\sin\theta+Z\cos\theta)(h\sin\theta+Z\cos\theta-W\cos\theta)}. \tag{3.2}$$

(In the following, (U,V,W) refers to the robot translation. See Appendix I for the relation between (U,V,W) and equation (3.2).) Figure 3 displays the relationship between disparity and tilt angle for the curve when $W=0.1$ ($h=1$: unit length). The horizontal axis corresponds to the distance δ between the camera and the ground, and the vertical axis to disparity d_y. The bar graphs below these curves

represent the extent of the visual fields, which depend on both θ and the vertical field of view of the camera. Based on these graphs, we can make the following observations.

1) For a fixed δ, the tilt angle does not have much effect on d_y.

2) As θ increase, the visual field is compressed towards the camera, and near the camera, where disparity d_y is large, accurate depth information can be obtained.

In the following, the tilt angle is fixed to 23 degrees: this scales the vertical visual field from 1.4 to 8.7 units in front of the camera when $W=0.1$.

(b) **Disparity vs. Motion**

In an ordinary stereo vision system (e.g. [10]), where the camera's lines of sight are parallel and the base line is perpendicular to the line of sight, disparity d_x is proportional to the length of the base line (U):

$$d_x = -\frac{fU}{Z} \qquad (V=W=0). \tag{3.3}$$

Next, we examine the relationship between disparity and motion W. From (3.2), the relationship between disparity and motion when $W=w$ is

$$d_y^w = \frac{fhw}{Z'(Z'-w\cos\theta)}$$

where $Z'=h\sin\theta+Z\cos\theta$. If we increase the motion w by a factor of n ($W=nw$)

$$d_y^{nw} = \frac{fhnw}{Z'(Z'-nw\cos\theta)}.$$

If F is the ratio of disparity for velocity nw to disparity at velocity w, then

$$F=\frac{d_y^{nw}}{d_y^w}=n+\frac{n(n-1)w\cos\theta}{Z'-nw\cos\theta}. \tag{3.4}$$

Figure 4 shows curves of F in terms of n with $w=0.1$. We can make the following observations.

1) From this graph and (3.4), $F\approx n$ when $w\ll Z'$.

2) In an ordinary stereo system, the common visual field of the two cameras changes from 70% to 30% when the base line changes from $U=0.1$ to $U=0.2$. In motion stereo, the common visual field of two frames changes only from 95% to 84% when the motion changes from $W=0.1$ to $W=0.2$.

4. ERROR ANALYSIS OF ESTIMATED DEPTH IN 3-D SPACE

The estimated depth information has errors due both to quantization errors of edge locations and to estimation errors of robot motion. In this section, we consider the errors in the depth information in 3-D space.

4.1. Depth Error due to Quantization Error in Edge Location

Since the disparity is determined from the difference of the corresponding feature points in two frames, inaccuracy of the disparity results from localization errors of feature points in the image plane. If a point has a quantization error e_{yi} in the i^{th} frame ($i=1,2$), the disparity has a total quantization error e_y given by

$$|e_y|=|e_{y1}|+|e_{y2}|.$$

Figure 5 shows a side view of the depth error due to quantization error of edge location. An actual point must exist in a rectangle $PQRS$ (a solid in 3-D space)

which is obtained as a common region of two thick rays from the first and the second camera positions. Since PR is very small compared to the length QS, we approximate this rectangle as a "stick" QS in 3-D space.

An important property of the depth errors due to quantization errors is that they are independent of one another.

4.1.1. Relative Depth Errors

We can estimate the relative errors in the depth information by our method as follows.

From (2.3),

$$Z = -\frac{fV}{d_y} + \frac{Wy_2}{d_y} = \frac{1}{d_y}(-fV + Wy_2).$$

If the disparity error is uniformly distributed in the interval $[-e_y, +e_y]$, then the maximal overestimated depth Z_o and the minimal underestimated depth Z_u are given by

$$Z_o = \frac{1}{d_y - e_y}(-fV + Wy_2) \quad \text{and} \quad Z_u = \frac{1}{d_y + e_y}(-fV + Wy_2).$$

ΔZ, the expected error in Z, is thus

$$\Delta Z = \frac{Z_o - Z_u}{2} = \frac{e_y}{d_y^2 - e_y^2}(-fV + Wy_2).$$

Therefore, the relative error $\left|\frac{\Delta Z}{Z}\right|$ is given by

$$\left|\frac{\Delta Z}{Z}\right| = \frac{d_y e_y}{d_y^2 - e_y^2}.$$

When $e_y \ll d_y$, $\left|\frac{\Delta Z}{Z}\right| \approx \frac{e_y}{d_y}$: the relative error of the depth value is inversely proportional to the disparity value.

4.1.2. Results of Simulation

We applied our method to synthetic road boundaries. In the following simulation, we used the same camera parameters as in Section 3 (tilt angle is fixed at 23 degrees). Figure 6(a) shows the road boundaries extracted from first frame (+) and second frame (○) with translation (U,V,W)=(0,0,0.4) and rotation (α,β,γ)=(0,–5.0°,0) (steering to the right). The locations of feature points on the road boundary are quantized in a 512×480 pixel image. Figure 6(b) shows the 3-D road structure in the world coordinate system from the top and side. Large circles in the top view show the locations of the robot in the first frame (+) and in the second frame (small ○). Figure 7 shows the results of reconstruction. Short line segments (which we will refer to as "sticks") with small circles at both ends represent intervals of edge location in 3-D when e_y=1 pixel. The true location (+) lies on the corresponding stick. The nearer the point is to the viewer, the shorter is the length of the stick. The relative depth error was very nearly inversely proportional to disparity as mentioned above.

4.2. Depth Error due to Estimation Error of Robot Motion

So far we have reconstructed the depth information assuming that the motion parameters are accurate. In practice, however, the motion parameters might not be very accurate. Therefore, analysis of the effect on the estimated 3-D structure caused by the errors in the motion parameters is very important. Here, we consider the inaccuracy of the translational motion and do not deal with that of rotation. We discuss the effect of the inaccuracy of rotational motion in Section 5.

When the translation components of the robot motion have some errors, their effect on the depth value is not local as in Section 4.1 but global to the scene because the translational components (U, V, W) are common to all points in the scene. Figures 8 show the perspective and side views of this process. Obviously, the error in the vertical component ΔV of translation has a large effect on the estimated depth because the angle between the y-axis of the camera and the line of sight to the point is large.

4.2.1. Experimental Results

We used the ALV simulator robot in the Computer Vision Laboratory at the University of Maryland to input road images. The motion of the ALV can be simulated by the robot arm which is program controllable from a VAX-11/785 (see [5] for more detail). The input images are quantized into 512×480 pixel images. Figures 9(a) and (b) show the input road images where the black squares beside the road are used to determine the location of the FOE from consecutive

images. The most serious problem is camera calibration. We obtained the internal camera parameters (the focal length and the image center) by moving the camera a known direction and distance (see Appendix II for more detail). The road boundaries are easily obtained by [5] or [7] and are fitted to line segments. Top views of the road boundary space curves estimated from various components of translation are shown in Figures 10(a), (b) and (c) for U, V and W with various errors. Each component changes from -5% to $+5\%$ of the traveled distance such that the traveled distance is constant (1 inch). From Figure 10(b), we can see that the V component has a large effect on the depth value as mentioned above in Figure 8.

Although our method does not require knowledge of road width, camera height or tilt angle of the camera, these values are very useful for verifying the result of the method. From the estimated range of depth information, we can obtain the following parameters:

Camera height: from 1.9 inches to 2.3 inches (actually, about 2 inches)

Tilt angle: from 20° to 24° (actually, about 23°)

Road width: from 2.4 inches to 3 inches (actually, about 2.6 inches). As a result, the error range is very large although the actual values are included in this range.

5. DISCUSSION

We have described a new method for reconstructing the depth information of a road boundary from motion stereo and for representing the error of a 3-D road structure in 3-D space. The depth errors due to inaccuracy of motion esti-

mation are especially important because the errors in the translational components change the depth values of all points in the same way. In this paper, we have not considered the effect of the rotational movement on the depth value. A rotational motion about the X- or Y-axis of 1° causes a motion of 10 pixels in the image plane if the same focal length lens is used (about 800 pixels of equivalent focal length). Displacement due to a rotation around the Z-axis depends on the distance from the image center and the maximum displacement is 4 or 5 pixels at the edge of the image. Therefore, an error in rotation estimation can cause serious errors in disparity. As mentioned in Section 4, the error range of depth information is relatively large if we use only two frames. Therefore, we need to use many frames and/or other sensor outputs in order to refine the depth information. The final goal of our research is to obtain a more accurate 3-D road structure by fusing the error ranges of depth information from many more frames and/or from range sensors in 3-D space.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Larry S. Davis and Mr. Daniel DeMenthon for helpful comments and discussions.

Appendix I: DISPARITY WITH TILT ANGLE

When the camera has some tilt angle, the translation of the robot is not equivalent to that of camera. We denote the translation of the robot by (U_w, V_w, W_w) and that of the camera by (U_c, V_c, W_c). The relations between them are given by

$$U_c = U_w, \quad V_c = W_w \sin\theta + V_w \cos\theta, \quad W_c = W_w \cos\theta - V_w \sin\theta,$$

where θ is the tilt angle of the camera. Similarly, denote the camera height and the distance to the point from the viewer of the robot by (Z_w,h_w) and that of the camera by (Z_c,h_c). The relations between them are given by

$$Z_c = Z\cos\theta + h\sin\theta, \quad h_c = -Z\sin\theta + h\cos\theta.$$

Substituting (U_c, V_c, W_c, Z_c, h_c) instead of (U,V,W,Z,h) in eqn(2.2), we obtain eqn(3.2).

Appendix II: CAMERA CALIBRATION FOR MOTION STEREO

A. Basic Assumptions

Camera calibration in the context of 3-D machine vision is the process of determining the internal camera geometric and optical characteristics (internal camera parameters) and/or the 3-D position and orientation of the camera frame relative to a certain world coordinate system (external camera parameters). (The current state of the art is well described in [11].) We consider obtaining only the internal camera parameters using the motion of the robot arm to which the camera is attached, since the motion stereo method does not need external parameters to obtain depth information.

The camera calibration method described here is based on the following assumptions.

(1) There is no lens distortion: Although it exists even if we use a very accurate lens, we do not consider lens distortion here in order to simplify the method.

(2) The motion parameters of the robot arm are known accurately: We used an American Robot MERLIN robot arm on which a small solid state CCD black and white camera is attached. Theoretically, this robot can position its arm to within .001 inch of a previously defined point ([12]). Therefore, we can expect that the motion parameters are accurate enough for our purpose.

(3) The geometrical relation between the camera and the robot arm is known: The camera position and orientation relative to the robot arm have been obtained ([5]) using the geometrical parameters of the road simulation board. Of course, there is a slight difference between the true and estimated camera axes

(we call the optical axis of the camera the camera axis); therefore, the image origin of the estimated camera axis is different from the true one. However, there is an interesting remark in [11] that the fitting between the observation and model is still good although different setups yield different values of the image origin for the same camera and lens. Although we cannot prove theoretically that the image origin has a negligible effect on the accuracy of the final 3-D measurement, we can compute the expected depth error due to this by simulation. Figure A-1 shows the side view of the true and estimated camera axes. A point P in 3-D space is projected on the true image plane at y_P^t, but we observe this point on the estimated image plane as y_P^e. The relation between these y-coordinates is given as follows when the angle between these camera axes is θ:

$$y_P^e = \frac{y_P^t f}{f\cos\theta - y_P^t \sin\theta},$$

where f denotes the focal length. Similarly, the relation between the true disparity dy^t and the estimated disparity dy^e of two points P and Q in 3-D space is given by

$$dy^e = dy^t \frac{f^2\cos\theta}{(f\cos\theta - y_P^t\sin\theta)(f\cos\theta - y_Q^t\sin\theta)}.$$

Figures A-2(a) and (b) show the relative error of the depth value estimated from the above equation when the translation W in the z-axis direction is 0.25 and 0.5 (the dimension is the same as in Section 4), respectively. The horizontal axis is for the distance from the camera to the point, the vertical axis for the relative error of depth. The hatched region shows the error due to the quantization error

estimated in Section 4.1. From these figures, we can regard the error due to the orientation error of the camera axis as negligible as long as the angle between the true and estimated camera axes is very small, for example, less than 2° (1° corresponds to about a 10 pixel shift of the image origin).

B. Determining the Internal Camera Parameters

From the viewpoint of optical flow analysis, we can determine the internal camera parameters by moving the camera a known distance and known direction. When the camera moves in 3-D space with a translation (U,V,W), the coordinates of the Focus of Expansion (x_{FOE},y_{FOE}) are given as follows ([3]):

$$x_{FOE}=\frac{fU}{W},\quad y_{FOE}=\frac{fV}{W} \tag{A.1}$$

Thus, we can determine the image origin and the focal length by moving the camera straight $(U=V=0)$ and diagonally $(U=V=W)$.

(a) Image origin

First, we determine the coordinates of the image origin (c_x,c_y) by moving the camera with a translation (.0, .0, 4.0) (in the following, the dimension of the translation is an "inch"). Figure A-3 shows the position of the image origin which is obtained as the FOE of the translation. The specification of the feature points and their correspondences between two frames are performed by hand. The image size is 512×480 and the obtained coordinates of the image origin are $(c_x,c_y)=(273.59, 233.96)$.

(b) Focal length

Theoretically, determining the focal length is straightforward, but there are two unknowns: aspect ratio r of the quantized pixel in the vertical direction to the pixel in the horizontal direction, and roll angle γ between the true and estimated camera coordinate systems. Figure A-4(a) and (b) are flow images with translations (4.0, .0, .0) and (.0, 4.0, .0), respectively, and they show the existence of the roll angle between the true and estimated camera coordinate systems. Then, in order to obtain the focal length f_x in the horizontal direction and f_y in the vertical direction ($f_x = r f_y$), we determine the aspect ratio r and the roll angle γ as follows:

(1) Make four flow images with translations $(\pm t, \pm t, t)$ $(t \neq 0)$ and determine the four coordinates of the FOE for each translation. Their positions and coordinates are denoted as P, Q, R, S, and (x_i, y_i) $(i = P, Q, R, S)$ in Figure A-5.

(2) Determine r so that angle$(<POQ) = \frac{\pi}{2}$ (or $<QOR$, $<ROS$, $<SOP = \frac{\pi}{2}$).

From the scalar product of two vectors $\overline{OP}$ and $\overline{OQ}$:

$$(x_P - c_x) \times (x_Q - c_x) + r(c_y - y_P) \times r(c_y - y_Q) = 0.$$

(3) Obtain the roll angle γ as a difference angle between $\frac{\pi}{4}$ and angle$(<XOP)$. The angle$(<XOP)$ is given by

$$\text{angle}(<XOP) = \tan^{-1}\left(\frac{c_y - y_P}{x_P - c_x}\right).$$

(4) Determine the focal lengths f_x and f_y by rotating P (or Q,R,S) by the roll angle γ.

Figures A-6(a), (b), (c) and (d) show the flow images with translations ($\pm$2.5, $\pm$2.5, 2.5) and their FOEs. Compared with the FOE of the straight motion (Figure A-4), the variance of the intersections with flow is large. As a result, we obtain the parameters as follows:

$$r = 0.787,\ \gamma = -1.26°,\ f_x = 614.5,\ f_y = 780.7.$$

C. Concluding remarks

We have described a camera calibration method for motion stereo. In this method, we obtained the internal camera parameters: the image origin and the focal length. Since we used the relative values of the robot motion, that means we did not measure absolute values, and since the estimated depth includes various errors as described in Section 4, to estimate the accuracy of the calibrated parameters seems difficult. From our experiment, we can conclude that:

(1) To obtain a more accurate focal length, we need to input many more images with various translations and determine the FOE with a smaller variance. As mentioned above, the variance of the FOE with diagonal translation is large, say 10 pixels, because the angles between flows are small.

(2) We should consider lens distortion. The difference between the coordinates of the FOE obtained from the robot motion parameters and that from feature point matching in two frames is about 4% of the traveled distance. Since the motion parameters of the robot arm are very accurate, this error seems to occur due to quantization error and lens distortion. If we make a calibration table for lens distortion, we could obtain a better estimate of depth value.

REFERENCES

[1] B.K.P. Horn, *Robot Vision*, MIT Press, Cambridge, Massachusetts, 1986.

[2] S. Tsuji, Y. Yagi, and M. Asada, "Dynamic scene analysis for a mobile robot in man-made environment", IEEE Int. Conf. Robotics and Automation, pp. 850–855, 1985.

[3] S. Tsuji, J. Y. Zheng, and M. Asada, "Stereo vision of a mobile robot: world constraints for image matching and interpretation", IEEE Int. Conf. Robotics and Automation, pp. 1594–1599, 1986.

[4] D. DeMenthon, "A zero-bank algorithm for inverse perspective of a road from a single view", IEEE Int. Conf. Robotics and Automation, pp. 1444–1449, 1987.

[5] A.M. Waxman, J. LeMoigne, L.S. Davis, E. Liang, and T. Siddalingaiah, "A visual navigation system", IEEE Int. Conf. Robotics and Automation, pp. 1600–1606, 1986.

[6] M. Asada and S. Tsuji, "Shape from projecting a stripe pattern", IEEE Int. Conf. Robotics and Automation, pp. 787–792, 1987.

[7] L.S. Davis, T. Kushner, J. LeMoigne and A. Waxman, "Road boundary detection for autonomous vehicle navigation", *Optical Engineering*, **25**, pp. 409–414, 1986.

[8] T. Echigo and M. Yachida, "A fast method for extracting of 3-D information using multiple stripes and two cameras", 9th IJCAI, pp. 1127–1130, 1985.

[9] K. Prazdny, "Determining the instantaneous direction of motion from optical flow generated by a curvilinearly moving observer", *Computer Graphics and Image Processing*, **17**, pp. 238–248, 1981.

[10] G. Xu, S. Tsuji, and M. Asada, "A motion stereo method based coarse-to-fine control strategy", *IEEE Trans. PAMI* **PAMI-9**, pp. 332–336, 1987.

[11] R.Y. Tsai, "An efficient accurate camera calibration for 3-D machine vision", CVPR86, pp. 364–374, 1986.

[12] American Robot Corporation, "Merlin robot system operator and user guide", Manual No. SMT-2.0-0184, revision 2.1, 1984.

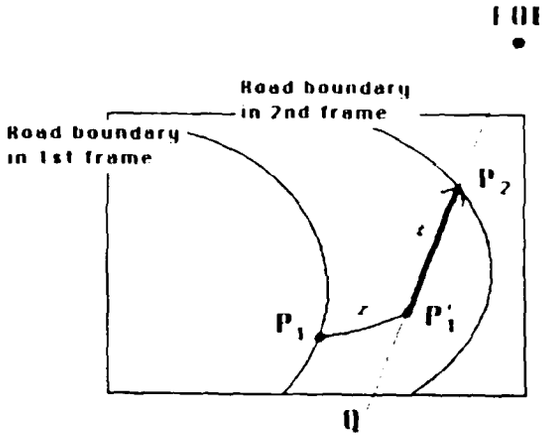


Fig.1 Epipolar line in the motion stereo

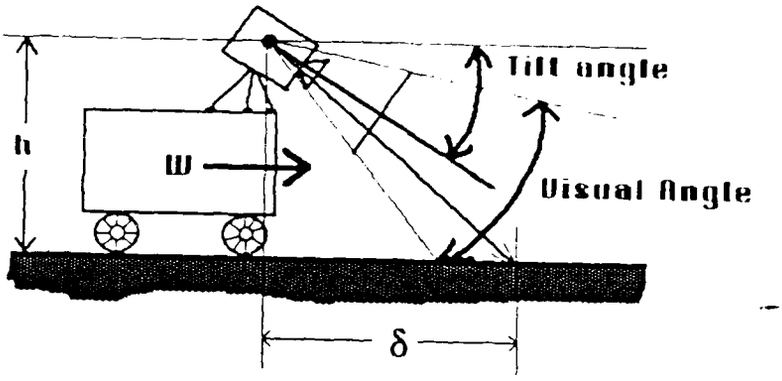


Fig.2 Geometrical parameters of ALV

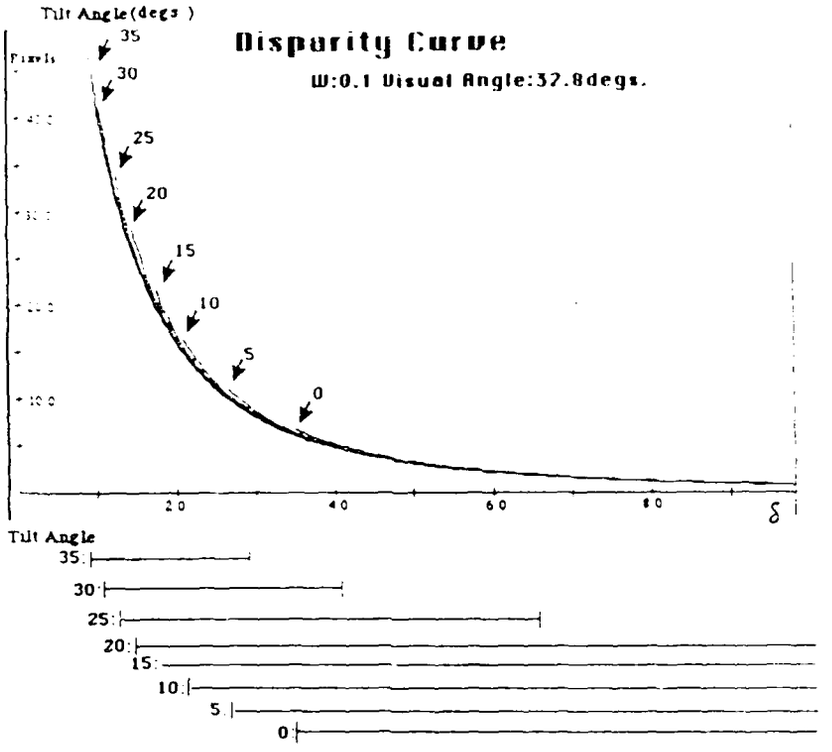


Fig.3 Disparity curves v.s. tilt angles

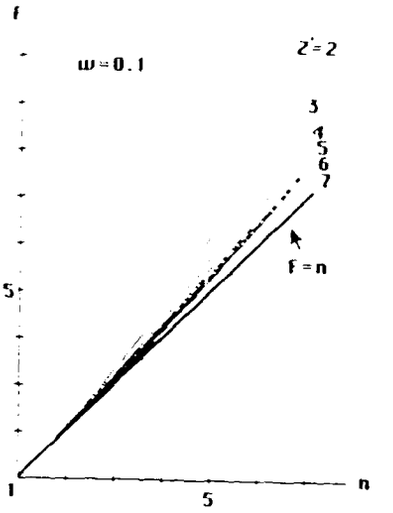


Fig.4 Factor curves in terms of motion parameters

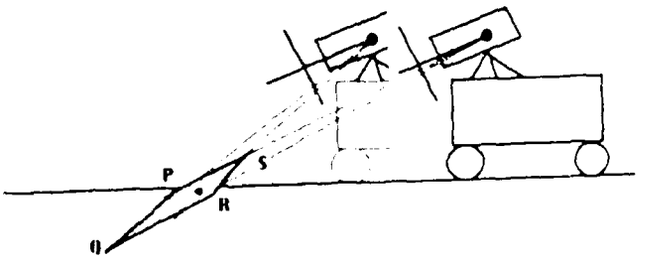


Fig.5 Side view of depth errors due to quantization error of edge location

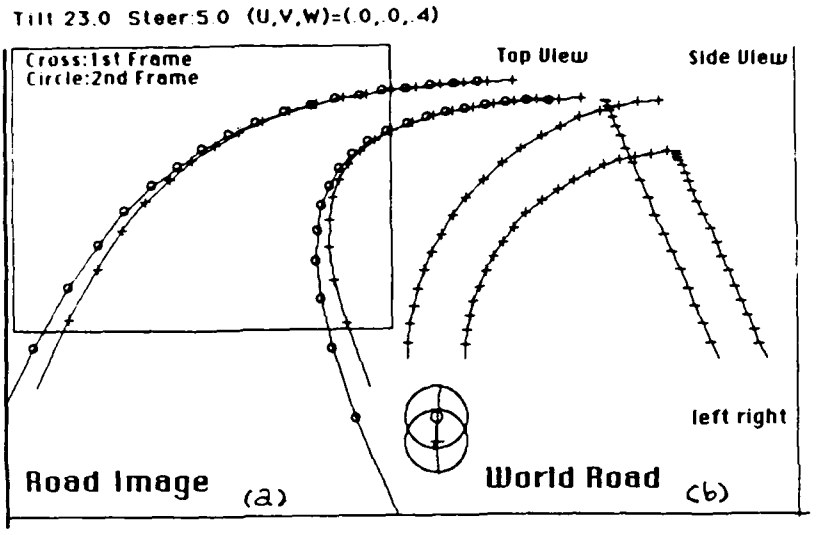


Fig.6 Road boundaries with translation and rotation

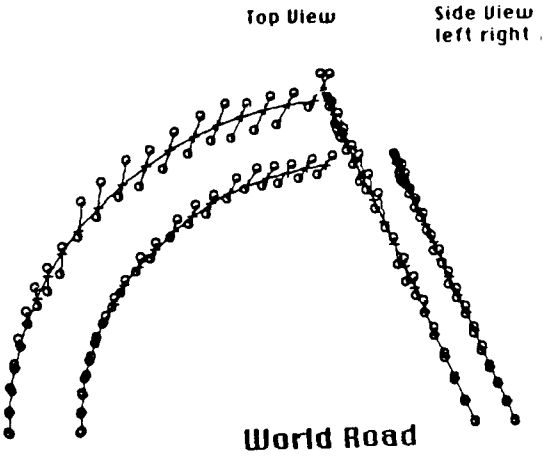


Fig.7 Simulation result for Fig.6

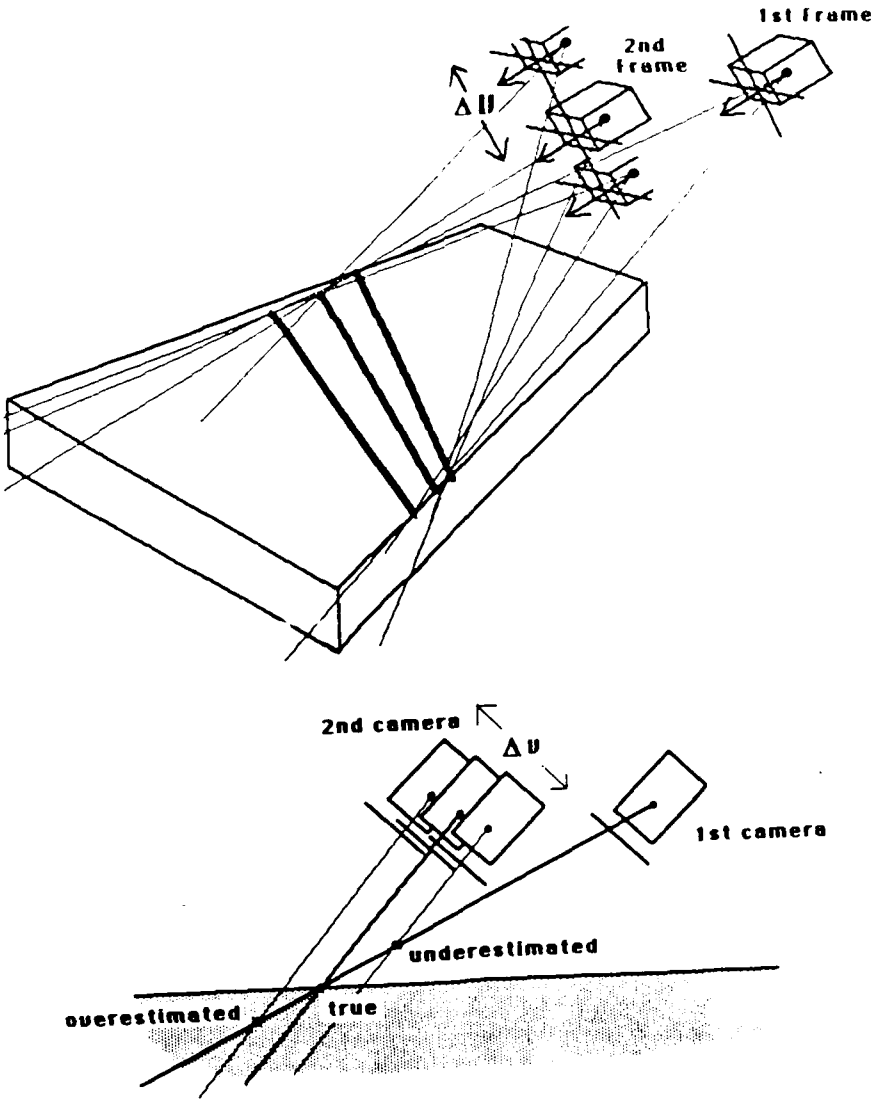
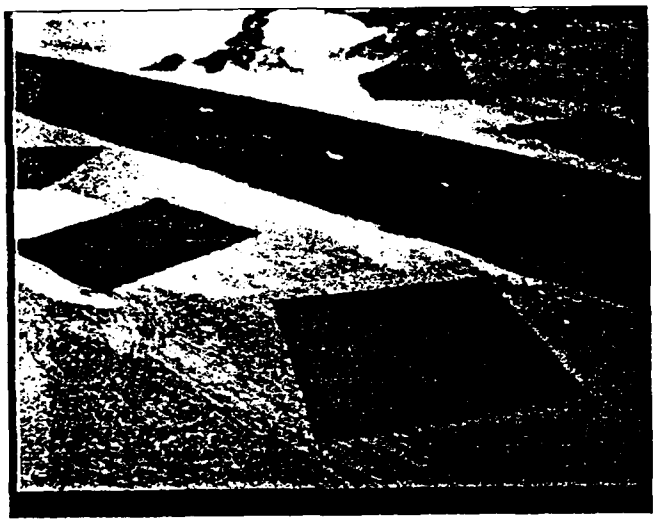
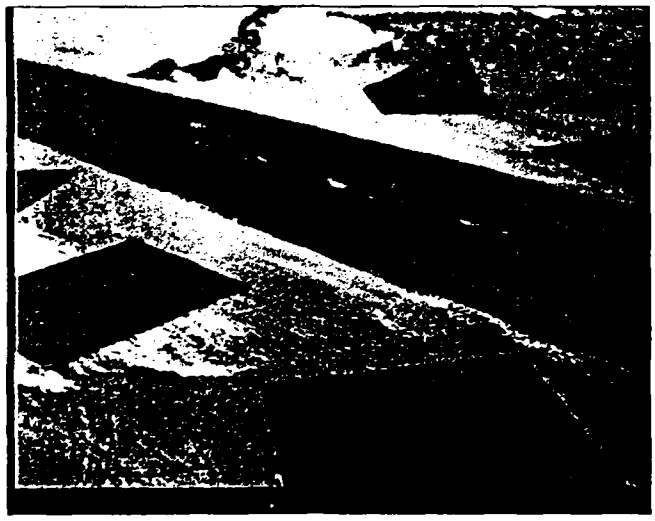


Fig.8 Space curves of road boundary estimated from erroneous translation



(a)



(b)

Fig.9 Input images taken by ALV simulator

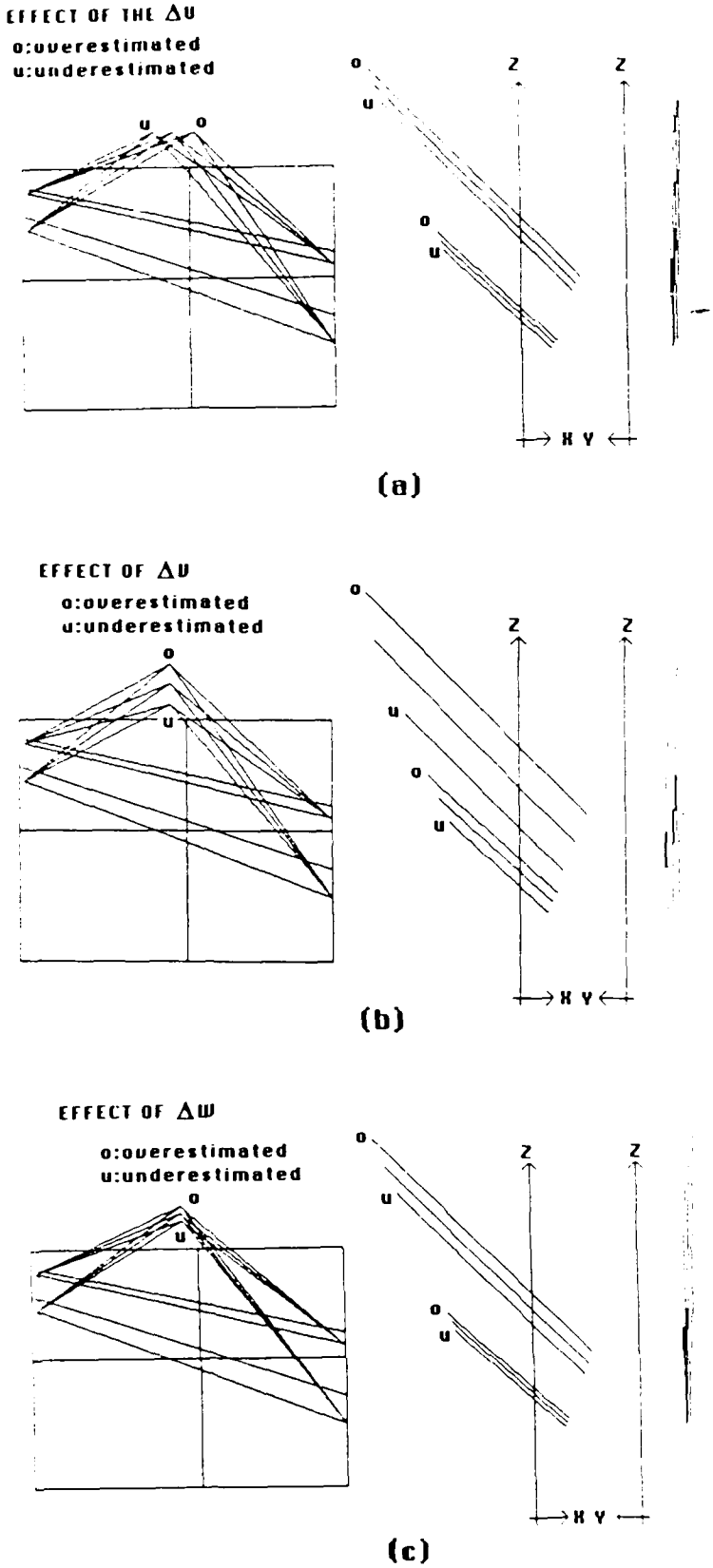


Fig.10 Experimental results.

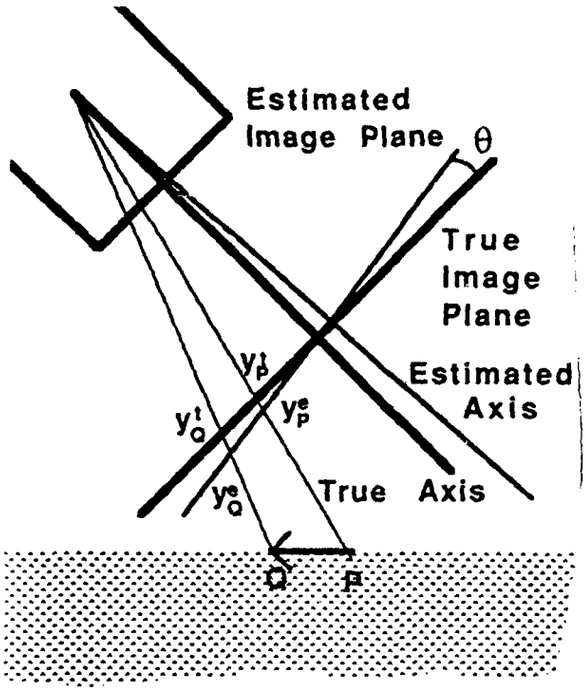


Fig.A-1 The relation between the true and estimated camera axes

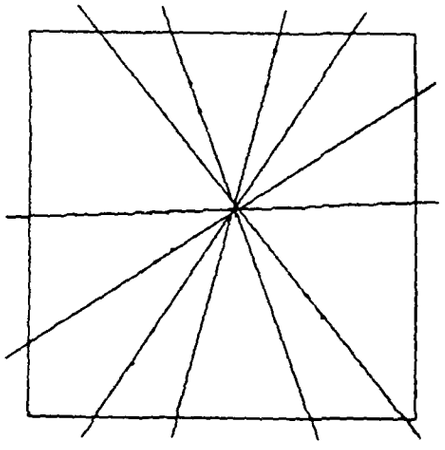


Fig.A-3 The image origin determined by straight motion

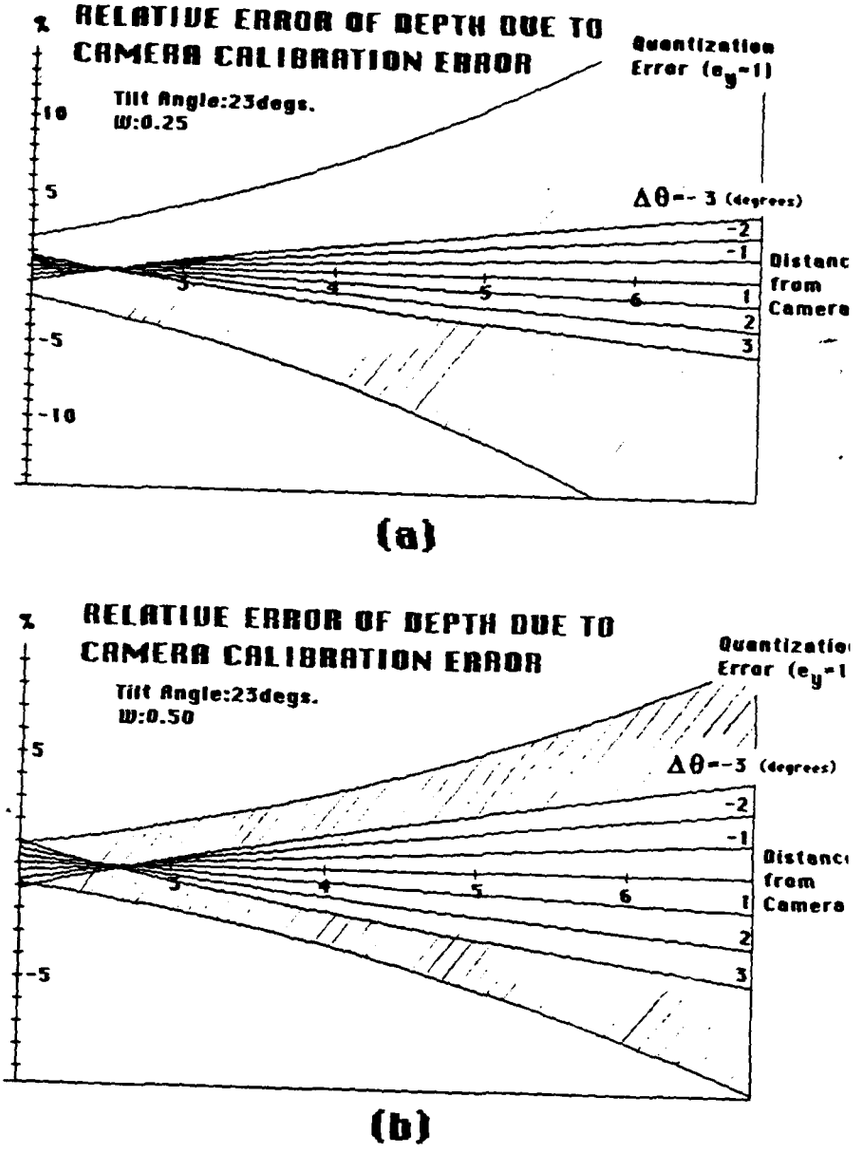


Fig.A-2 Relative errors of depth due to the difference angles between the true and estimated camera axes

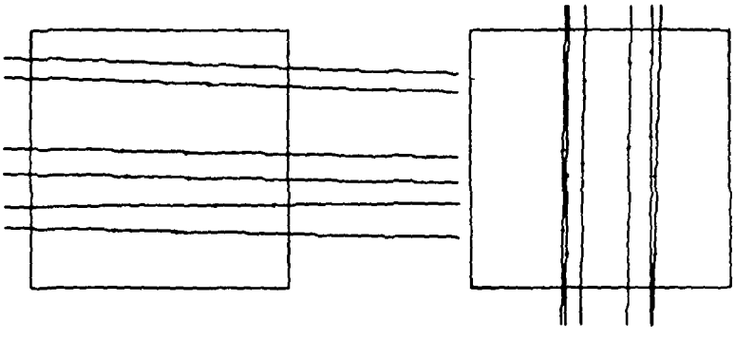


Fig.A-4 Flow images with vertical and horizontal translations

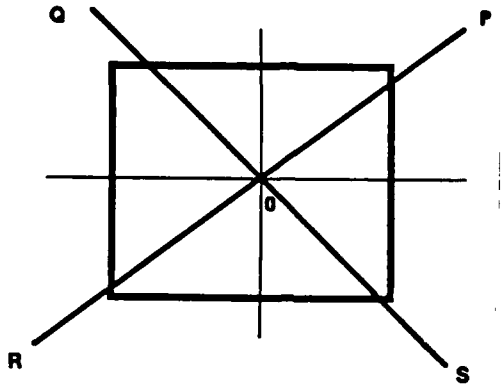


Fig.A-5 Four FOEs by four diagonal translations

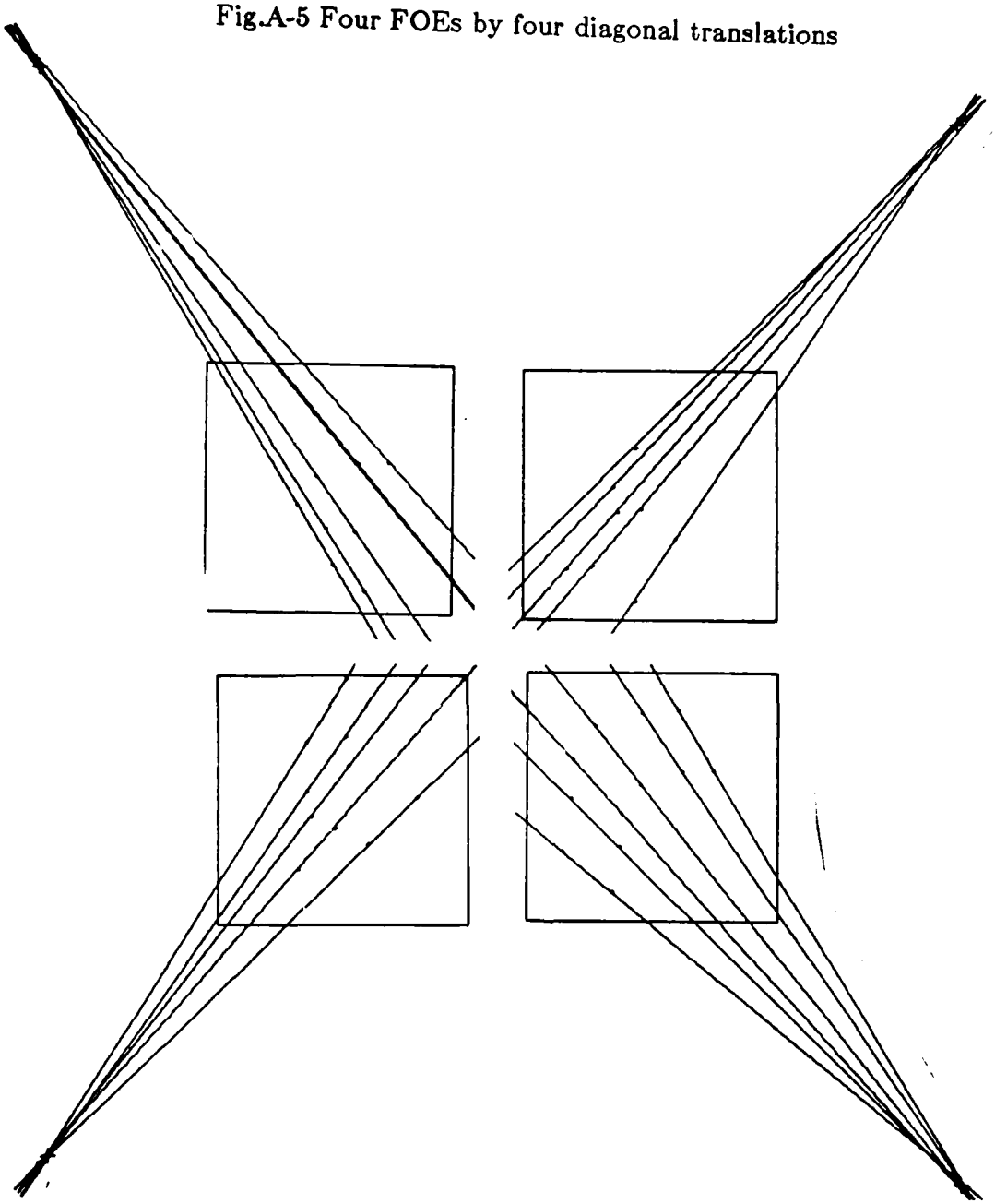


Fig.A-6 Four FOEs from experiments

END

DATE

FILMED

DTIC

JULY 88