

RSD-TR-11-83

**MOTION PARAMETER ESTIMATION: DECOUPLING
THE SCALE CHANGE PARAMETER¹**

**Edward J. Delp
Paul H. Eichel**

July 1983

**Center for Robotics and Integrated Manufacturing
Robot Systems Division
COLLEGE OF ENGINEERING
THE UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109**

**THE UNIVERSITY OF MICHIGAN
ENGINEERING LIBRARY**

-
Supported by the Center for Robotics and Integrated Manufacturing
of Engineering, University of Michigan, and the Air Force Office of
Scientific Research under Grant F49620-82-C-0089.

en 8m

UNRQ755

Abstract

Estimating the motion parameters of moving objects in temporal sequences of images is an important problem in time-varying imagery. Because of its computational simplicity, the method of linear least squares is attractive for the case of two-dimensional motion with scale change i.e. rigid objects moving in a plane parallel to the image plane plus zooming. The use of least squares for parameter estimation is not limited to a spatial gradient based motion detector. Two other pixel based motion detectors are shown to be equivalent to the spatial gradient based detector for this application. Since such motion involves four parameters, the least squares method entails inversion of a 4×4 matrix. We show that the scale change parameter may be decoupled from the others thus decomposing the problem into a one-dimensional and a three-dimensional estimation problem. Both computation speed and accuracy are enhanced by this decomposition.

I. Introduction

The method of linear least squares has been applied to the problem of estimating motion parameters for the case of two-dimensional motion plus scale change [3],[12],[14],[15]. The object(s) to be tracked may translate and rotate in a plane parallel to the image plane and translate along an axis perpendicular to the image plane (scale change or zooming). The latter three of the above references discuss the use of a spatial-gradient based motion detector to obtain local velocity information. We show that two other motion detectors, the $\frac{d}{dt} \nabla^2 G$ operator of Marr and Ullman [10], and difference pictures, are mathematically similar to spatial-gradient detectors with respect to the information they furnish regarding the local velocity field. The least squares method of estimating the four motion parameters proceeds in identical fashion regardless of the specific motion detector used.

We furthermore show that the scale change parameter may be decoupled from the other three parameters. Thus the four dimensional linear least squares problem may be decomposed into the one dimensional problem of estimating the scale change and the three dimensional problem of finding the parameters of translation and rotation in the plane. This result has important implications regarding the speed and accuracy of the calculations.

II. Linear Least Squares Motion Estimation

In this section we develop the equations of motion. It will be seen that they are linear in the motion parameters, thus facilitating the use of the method of least squares. The reader is referred to [16] for an alternative, but equivalent, development based on the affine transform or to [6] for another approach.

When a rigid object is constrained to move in a plane, the velocity at any point of the object may be decomposed into a translational component, \vec{V}_t , and an angular component, $\vec{\omega}$, such that:

$$\vec{V}(\vec{x}) = \vec{V}_t + \vec{\omega} \times \vec{r} \quad (1)$$

where $\vec{V}_t = (V_{t_x}, V_{t_y}, 0)$, $\vec{\omega}$ is the axis of rotation (perpendicular to the image plane), \vec{r} is a vector from the intersection of this axis with the image plane to the point $\vec{x} = (x, y)$, and " \times " denotes a vector cross product [1]. Without loss of generality, we may take the axis of rotation to pass through the origin, so that \vec{r} is now identically the positional vector \vec{x} .

Zooming, or change of scale, is a result of translation along the coordinate axis perpendicular to the image plane. Depending on sign, such motion contributes a radial velocity component at each point of the object that is proportional to both the scale change coefficient, α , and the distance of the point from the origin, $|\vec{x}|$. This radial component is directed toward the origin for motion away from the camera and vice versa:

$$\vec{V}_r = \alpha \vec{x} \quad (2)$$

Here, positive α indicates a positive scale change or motion toward the camera.

Thus, the net velocity at a point in the image plane is given by:

$$\vec{V}(\vec{x}) = \vec{V}_t + \vec{\omega} \times \vec{x} + \alpha \vec{x} \quad (3)$$

Three types of motion detectors are common in the literature, one based on the spatial gradient, another based on the $\frac{d}{dt} \nabla^2 G$ operator, and the third on difference pictures, do not completely specify the velocity field $\vec{V}(\vec{x})$ at every point, \vec{x} , in the image. Rather, they yield only scalar quantities. We show that

for all three, the detector output is actually the inner product of the velocity field with some other vector field. The magnitude and direction of this other field may be estimated, however, enabling the motion parameters to be estimated using the method of linear least squares.

The first of the motion detectors considered here is based on the spatial gradient. When the interframe change in pixel intensity at a point \vec{x} is expanded in a Taylor series involving the spatial derivatives of the image intensity function $f(\vec{x})$, and only the first term of the series is retained, the resulting approximation may be stated in vector notation as [2],[8],[11]:

$$\Delta f(\vec{x}) = -\vec{V}(\vec{x}) \cdot \nabla f(\vec{x}) \quad (4)$$

where $\vec{V}(\vec{x})$ is the local velocity at \vec{x} , $f(\vec{x})$ is the image intensity function, $\nabla f(\vec{x})$ is the spatial gradient of the intensity function, and $\Delta f(\vec{x})$ is the change in the intensity between frames. Dividing Equation (4) by the magnitude of the gradient we have:

$$\vec{V} \cdot \vec{n} = -\frac{\Delta f(\vec{x})}{|\nabla f|} \quad (5)$$

where \vec{n} is a unit vector in the direction of the gradient. The direction as well as the magnitude of the gradient may be estimated by a gradient operator such as the Sobel Operator, and since Δf is easily found by subtraction, Equation (5) may be used to infer the component of the local velocity vector in the direction \vec{n} , whose direction cosines are known. Note that Equation (5) does not completely specify the magnitude and direction of \vec{V} at the point \vec{x} [17].

The $\frac{d}{dt} \nabla^2 G$ motion detector of Marr and Ullman [10] is based on the $\nabla^2 G$ static edge detector of Marr and Hildreth [9]. When an image is convolved with the $\nabla^2 G$ operator mask, the contours of zero-crossings of the resulting image

correspond to edges in the original image. Because of the monotonic nature of the convolution output at the zero-crossing, the sign of the time derivative, $\frac{d}{dt}(\nabla^2 G * f)$, (where $*$ denotes convolution) unambiguously specifies the direction of motion, i.e. whether the contour is moving to its "right" or "left". The time derivative may be approximated by subtracting a pair of convolved images of a temporal sequence.

The magnitude of the time derivative along with knowledge of the slope of the convolved image at the zero-crossing can be used to infer, to a first order approximation, the component of velocity normal to the zero-crossing contour much like the case of the spatial gradient technique:

$$\vec{v} \cdot \vec{n} = \frac{\frac{d}{dt}(\nabla^2 G * f)}{m} \quad (6)$$

where $\frac{d}{dt}(\nabla^2 G * f)$ is estimated as indicated above and m is the slope of the convolved image perpendicular to the zero-crossing contour. As in the case of the spatial derivative case [17], the complete velocity field cannot be specified by this method, a situation referred to by Marr and Ullman as the aperture effect [10].

Difference pictures may also be used as a motion detector. When two frames of a temporal sequence are subtracted, pixel by pixel, and the resultant difference image is thresholded, non-zero difference regions are generated at the boundaries of moving, high contrast objects [7]. The "width" of such regions is related to the velocity component normal to the high contrast boundary. That is, where the local velocity vector is perpendicular to a high contrast edge, the resulting difference region is of greatest width. If the velocity vector is tangent to the edge, no difference region appears (i.e. it is of zero width). Thus, the

difference region width is proportional to $\vec{V}(\vec{x}) \cdot \vec{n}$ where \vec{n} is a unit vector normal to the edge at \vec{x} . As with the other two motion detectors, difference pictures provide information about the component of velocity in some direction (the normal to the boundary) and we can estimate that direction from the orientation of surrounding elements of the boundary [3].

From the above, it is evident that all three motion detectors yield the same information concerning the velocity field, namely the inner product of the velocity with some unit vector whose direction is known. Letting \vec{n} denote this unit vector, the component of velocity along \vec{n} is, using Equation (3):

$$\vec{V} \cdot \vec{n} = \vec{V}_t \cdot \vec{n} + \vec{\omega} \times \vec{x} \cdot \vec{n} + \alpha \vec{x} \cdot \vec{n} \quad (7)$$

Expanding this equation, we see that:

$$d = (n_x V_{t_x} + n_y V_{t_y}) + (n_y x - n_x y) \omega + (n_x x + n_y y) \alpha \quad (8)$$

Here, $d = \vec{V} \cdot \vec{n}$ is the detector output, $\vec{n} = (n_x, n_y, 0)$ the direction of the local vector quantity obtained from the motion detector, $\vec{\omega} = (0, 0, \omega)$, $\vec{V}_t = (V_{t_x}, V_{t_y}, 0)$, and $\vec{x} = (x, y, 0)$. When the motion detector is based on the gradient, $n_x = \frac{1}{|\nabla f|} \frac{\partial f}{\partial x}$ and $n_y = \frac{1}{|\nabla f|} \frac{\partial f}{\partial y}$, where $|\nabla f| = \left[\frac{\partial f}{\partial x} \right]^2 + \left[\frac{\partial f}{\partial y} \right]^2$; for the $\frac{d}{dt} \nabla^2 G$ operator, \vec{n} is a unit vector normal to the zero-crossing contour; and for the difference pictures, \vec{n} is normal to the difference region boundary.

We observe that Equation (8) is linear in the four unknowns $[V_{t_x}, V_{t_y}, \omega, \alpha]$ and may be readily solved knowing the detector output for four or more image points. Let \vec{y} be a column vector of the parameters we wish to estimate: $\vec{y} = [V_{t_x} \ V_{t_y} \ \omega \ \alpha]^T$. For four or more image points, $\vec{x}_i = (x_i, y_i)$; $i=1,2,\dots,n$, with corresponding motion detector outputs, d_i , and unit vector direction cosines,

n_{x_i} and n_{y_i} , we form the set of equations:

$$d = A \bar{y} \quad (9)$$

where $d = [d_1 \ d_2 \ \dots \ d_n]^T$ and the rows of A are (from Equation (8)): $A_i = [n_{x_i} \ n_{y_i} \ (n_{y_i}x_i - n_{x_i}y_i) \ (n_{x_i}x_i + n_{y_i}y_i)]$. The linear least squares estimate of the parameter vector \bar{y} is then given by:

$$\bar{y} = (A^T A)^{-1} A^T d \quad (10)$$

Computationally, more accurate numerical answers can be achieved using the Moore-Penrose generalized inverse of A rather than Equation (10). This approach is furthermore necessary when $A^T A$ is singular [13]. As discussed in [18], the actual motion parameters of the object itself may only be determined to within a common scale factor except for the rotation parameter.

It may be noted in passing that although the three detectors discussed here are equivalent from the standpoint of finding local velocity components, they are not equal with regard to computational complexity and accuracy. The first two are accurate only where the interframe displacements of moving objects are small because of the first order approximation to the Taylor series. Exactly the opposite is true of difference pictures which can tolerate large displacements, but small movements suffer from spatial quantizing error. Spatial gradient based motion detectors have serious drawbacks where multiple moving objects are present [4]. The other two, however, can be used to partially segment the image before parameter estimation is attempted, greatly facilitating the task of tracking multiple objects. See [3] for a discussion of these ideas.

III. Decoupling the Scale Change Parameter

In this section we show that, given the motion parameter estimation problem of Section II, it is possible to solve for the scale change parameter, α , independently of the other three parameters which may be found using linear least square estimation.

Using the Divergence Theorem (or Gauss' Theorem), we have:

$$\iiint_R \nabla \cdot \vec{F} d\tau = \iint_S \vec{F} \cdot \vec{n} d\sigma \quad (11)$$

where \vec{F} is a vector field and the surface integral is performed over the surface S defining the volume R . Equation (11) is true as long as \vec{F} and its partial derivatives are continuous in R and on S and if S is piecewise smooth [5]. If the vector field \vec{F} has no components in the direction of the z axis, we have a similar result in one less dimension:

$$\iint_A \nabla \cdot \vec{F} dx dy = \int_C \vec{F} \cdot \vec{n} dt \quad (12)$$

where C is the closed boundary of region A .

The velocity field, \vec{V} , of the image plane satisfies the requirements for Equation (12) to hold. To see this, we have from Equation (3):

$$\begin{aligned} \vec{V} &= \vec{V}_i + \vec{\omega} \times \vec{x} + \alpha \vec{x} \\ &= (V_{i_x} - \omega y + \alpha x) \hat{i} + (V_{i_y} + \omega x + \alpha y) \hat{j} + 0 \hat{k} \end{aligned} \quad (13)$$

Since V_{i_x} , V_{i_y} , ω , and α are constants, $\vec{V}(\vec{x})$ is continuous at every point \vec{x} in the image plane. Furthermore, the first partial derivatives of \vec{V} are:

$$\frac{\partial \vec{V}}{\partial x} = \alpha \hat{i} + \omega \hat{j}$$

$$\frac{\partial \vec{V}}{\partial y} = -\omega \hat{i} + \alpha \hat{j} \quad (14)$$

All higher order partials are zero. Finally, \vec{V} clearly has no z-axis component.

Now, since $\vec{V} \cdot \vec{n}$ is precisely the motion detector output, it follows that a detector's output integrated about a closed contour equals the divergence of \vec{V} integrated over the interior of that contour. Taking the divergence of Equation (13):

$$\begin{aligned} \nabla \cdot \vec{V} &= \hat{i} \cdot \frac{\partial \vec{V}}{\partial x} + \hat{j} \cdot \frac{\partial \vec{V}}{\partial y} + \hat{k} \cdot \frac{\partial \vec{V}}{\partial z} \\ &= 2\alpha \end{aligned} \quad (15)$$

Substituting Equation (15) into (12) we obtain the result:

$$\begin{aligned} \int_C \vec{V} \cdot \vec{n} \, dl &= 2\alpha \iint_A dx \, dy \\ &= 2\alpha(\text{Area}) \end{aligned} \quad (16)$$

The scale change parameter may therefore be estimated by integrating the motion detector's output around a closed contour and dividing by twice the enclosed area. For discrete images, this integration is replaced by a summation over all points of the boundary. Where the $\frac{d}{dt} \nabla^2 G$ operator motion detector is used, such integration paths arise naturally since the $\nabla^2 G$ static operator generally gives rise to closed zero-crossing contours. Spatial gradient based motion detectors may be used in a two-step process to identify an object's boundary (a closed contour) and then to evaluate $\vec{V} \cdot \vec{n}$ around that boundary.

Difference picture based motion detection is particularly well suited to a rapid estimation of α using Equation (16). As noted in the previous Section, the "width" of a difference region is equal to the component of velocity normal to the boundary. Figure 1a illustrates that $\int \vec{V} \cdot \vec{n} \, dl$ along a straight boundary

segment is therefore exactly equal to the area of the difference region. This is not quite true for curved boundaries as shown in Figure 1b. For this section of boundary of constant curvature (radius of curvature = R) subtending a central angle of φ , we make a change of variables to polar coordinates: $dl = r d\vartheta$ and evaluate:

$$\begin{aligned} \int_a^b \vec{V} \cdot \vec{n} dl &= \int_0^\varphi (\alpha R) R d\vartheta \\ &= \varphi R^2 \alpha \end{aligned} \quad (17)$$

But the area of the difference region is:

$$A_r = \frac{\varphi}{2\pi} (\pi(\alpha R + R)^2 - \pi R^2) = \varphi R^2 \left(\alpha + \frac{\alpha^2}{2} \right) \quad (18)$$

Comparing Equations (17) and (18), the difference region area is a good approximation of the line integral when α is small, even for curved boundaries. Thus, α may be estimated by merely summing the areas of the difference regions (with opposite signs for newly covered and uncovered regions), and dividing by twice the area enclosed by the boundary. Furthermore, it will be noted that summing the difference regions in the above manner is the same as finding the change in area, ΔA , of the object area, A . Thus:

$$\alpha = \frac{\sum(\text{areas of difference regions})}{2A} = \frac{\Delta A}{2A} \quad (19)$$

a result which could be used with other motion detectors as well.

Once α is estimated, we can remove its contribution to the motion detector output (using Equation (8)):

$$d - (n_x x + n_y y) \alpha = (n_x V_{t_x} + n_y V_{t_y}) + (n_y x - n_x y) \omega \quad (20)$$

for three or more points and solve for $[V_{t_x}, V_{t_y}, \omega]$ in usual fashion with three dimensional linear least squares.

IV. An Example

The use of this technique for estimating the motion parameters of moving objects in real scenes is illustrated in Figure 2. In this example, the motion detector used is the difference picture detector described in [3]. The principal difference between that example and this one is the way in which the scale change parameter is estimated. In this case, the parameter is estimated using Equation (19) instead of the four-dimensional linear least squares method. The newer method results in significantly faster execution.

Figures 2a and 2b are taken from a sequence of images in which the two prominent foreground objects are translating, rotating, and moving toward the camera. These movements are synthetically produced so that the precise motion parameters are known. At the image plane, the motion from frame 2a to 2b corresponds to a 5.0 pixel translation to the right, 3.0 pixel translation down, a counterclockwise rotation of 0.05 radians, and a scale change of +0.07. Figure 2c is the difference image produced from the two sequence frames. The motion parameters are then estimated from this difference image using Equations (19) and (20), and the resulting parameter estimates are 4.54 pixels to the right, 2.45 pixels down, rotation of 0.052 radians (ccw), and scale change of +0.071.

The accuracy of these estimates is illustrated in Figure 2d. The first frame of the sequence is synthetically moved according to the estimated parameters and the resulting predicted second frame is subtracted from the actual second frame (Figure 2d). This difference picture indicates the discrepancy between the actual velocity field at every point and the velocity field predicted from the estimated parameters. Since the difference regions of Figure 2d are at the

most one pixel in width, we see that the velocity field is being estimated very accurately at every point of the image.

V. Conclusions

Three pixel-based motion detectors have been shown to be equivalent to the problem of least squares estimation of motion parameters for the case of two-dimensional motion with scale change. They are: the spatial gradient based detector, the $\frac{d}{dt} \nabla^2 G$ operator, and difference pictures. Furthermore, the four parameter estimation problem can be decomposed into two parts. The scale change parameter may be rapidly estimated independently of the rotation and translation parameters by integrating the motion detector output around a closed path (e.g. the target object's boundary). In the case of the difference picture motion detector, this may be approximated by merely summing the areas of the difference regions. Having estimated the scale change parameter in this way, the other three parameters are then estimated by a three-dimensional linear least squares projection.

VI. References

- [1] Becker, R. A., *Introduction to Theoretical Mechanics*, McGraw-Hill, 1954.
- [2] Cafforio, C. and F. Rocca, "Methods for measuring small displacements of television images," *IEEE Trans. on Info. Theory*, vol. IT-22, pp. 573-579, Sept. 1976.
- [3] Delp, E. J. and P. H. Eichel, "Estimation of motion parameters from local measurements of velocity components," *Proc. Twentieth Annual Allerton Conf. on Comm., Control, and Comp.*, Oct. 6-8, 1982, Monticello, Ill., pp. 135-142.
- [4] Fennema, C. L. and W. B. Thompson, "Velocity determination in scenes containing several moving objects," *Computer Graphics and Image Proc.*, vol 9, pp. 301-315, Apr. 1979.
- [5] Hildebrand, F. B., *Advanced Calculus for Applications*, Prentice-Hall, 1962.
- [6] Huang, T. S. and R. Y. Tsai, "Image sequence analysis; motion estimation," *Image Sequence Analysis*, T. S. Huang ed., Springer Verlag, New York, 1981, pp. 1-18.
- [7] Jain, R., and H. H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," *IEEE Trans. PAMI*, vol. PAMI-1, pp. 206-214, Apr. 1979.
- [8] Limb, J. O., and J. A. Murphy, "Estimating the velocity of moving images in television signals," *Computer Graphics and Image Proc.*, vol. 4, pp. 311-327, Dec. 1975.
- [9] Marr, D. and E. Hildreth, "Theory of edge detection, *MIT AI-Memo No. 518*, Apr. 1979.
- [10] Marr, D. and S. Ullman, "Directional selectivity and its use in early visual processing," *MIT AI-Memo 524*, June 1979.
- [11] Netravali, A. N. and J. D. Robbins, "Motion compensated television coding: part 1," *Bell Sys. Tech. Journal*, vol. 58, pp. 631-670, March 1979.
- [12] Newman, T., "Video target tracking by lie algebra techniques, *Proc. Workshop on Automatic Missile Tracking*, Redstone Arsenal, Alabama, Nov. 1979.
- [13] Noble, B. and J. Daniel, *Applied Linear Algebra*, Prentice-Hall Inc., 1977.

- [14] Prazdny, K., "Computing motions of planar surfaces from spatio-temporal changes in image brightness," *Proc. IEEE Conf. Pattern Recog. and Image Processing*, June 14-17, 1982, Las Vegas, Nevada, pp. 256-258.
- [15] Schalkoff, R. J., "Algorithms for a real-time automatic video tracking system," PhD dissertation, Univ. of Virginia, Charlottesville, Va., May 1979.
- [16] Schalkoff, R. J., and E. S. McVey, "A model and tracking algorithm for a class of video targets," *IEEE Trans. PAMI*, vol. PAMI-4, pp. 2-10, Jan. 1982.
- [17] Thompson, W. B. and S. T. Barnard, "Lower-level estimation and interpretation of visual motion," *Computer*, vol. 14, pp. 20-28. Aug. 1981.
- [18] Tsai, R. Y. and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," *Proc. IEEE Conf. Pattern Recog. and Image Processing*, June 14-17, 1982, Las Vegas, Nevada, pp. 112-118.

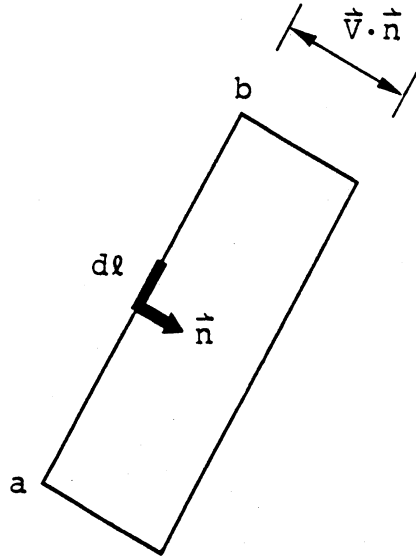


Figure 1a: $\int_C \vec{V} \cdot \vec{n} dl$ along a straight boundary.

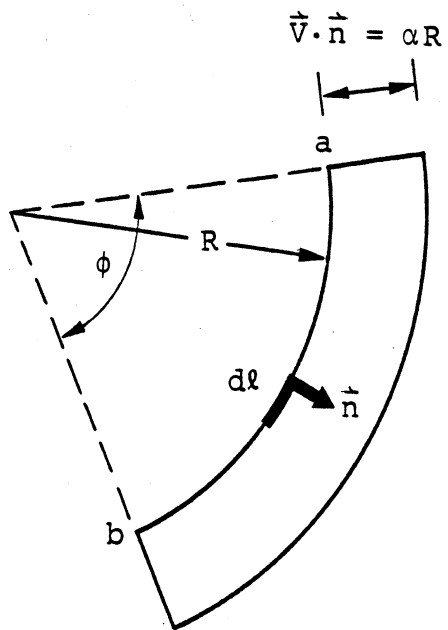


Figure 1b: $\int_C \vec{V} \cdot \vec{n} dl$ along a curve of constant radius.

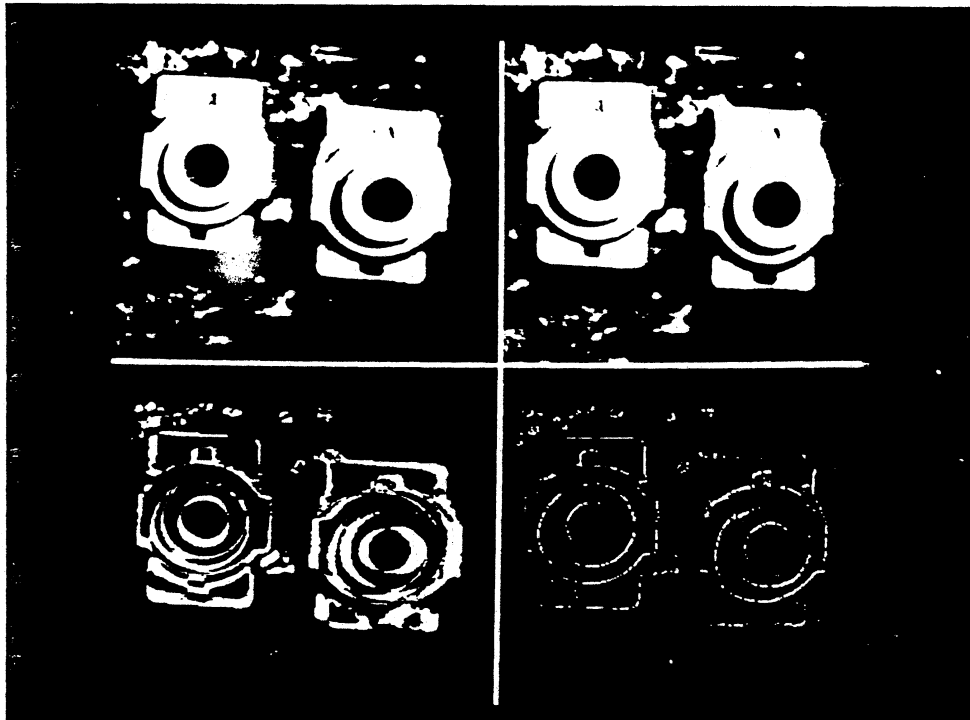


Figure 2:(a) Upper Left: First frame of the sequence. (b) Upper Right: Second frame of the sequence. (c) Lower Left: Difference image between above two frames. (d) Lower Right: Difference image between actual motion and estimated motion.

UNIVERSITY OF MICHIGAN



3 9015 02844 0686