Optical Flow Estimation: An Error Analysis of Gradient-Based Methods with Local Optimization

JOSEPH K. KEARNEY, WILLIAM B. THOMPSON, MEMBER, IEEE, AND DANIEL L. BOLEY, MEMBER, IEEE

***Abstract*—Multiple views of a scene can provide important information about the structure and dynamic behavior of three-dimensional objects. Many of the methods that recover this information require the determination of *optical flow*—the velocity, on the image, of visible points on object surfaces. An important class of techniques for estimating optical flow depend on the relationship between the gradients of image brightness. While gradient-based methods have been widely studied, little attention has been paid to accuracy and reliability of the approach.**

Gradient-based methods are sensitive to conditions commonly encountered in real imagery. Highly textured surfaces, large areas of constant brightness, motion boundaries, and depth discontinuities can all be troublesome for gradient-based methods. Fortunately, these problematic areas are usually localized can be identified in the image. In this paper we examine the sources of errors for gradient-based techniques that locally solve for optical flow. These methods assume that optical flow is constant in a small neighborhood. The consequence of violating in this assumption is examined. The causes of measurement errors and the determinants of the conditioning of the solution system are also considered. By understanding how errors arise, we are able to define the inherent limitations of the technique, obtain estimates of the accuracy of computed values, enhance the performance of the technique, and demonstrate the informative value of some types of error.

***Index Terms*—Computer vision, dynamic scene analysis, error analysis, motion, optical flow, time-varying imagery.**

I. Introduction

THE velocity field that represents the motion of object points across an image is called the optical flow field. Optical flow results from relative motion between a camera and objects in the scene. Most methods which estimate image motion lie within two general classes. *Gradient-based* approaches utilize a relationship between the motion of surfaces and the derivatives of image brightness [2], [3], [6], [14], [11], [13], [15], [17], [18], [19], [21]. *Matching* techniques locate and track small, identifiable regions of the image over time. A third approach that has recently received attention examines the dynamic variation of image structures such as contours [22].

Manuscript received July 19, 1985; revised September 17, 1986. Recommended for acceptance by W. E. L. Grimson. This work was supported by the Air Force Office of Scientific Research under Contract F49620-83-0140.

J. K. Kearney is with the Department of Computer Science, Cornell University, Ithaca, NY 14853, on leave from the Department of Computer Science, University of Iowa, Iowa City, IA 52242.

W. B. Thompson and D. L. Boley are with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455.

IEEE Log Number 8612805.

The appropriateness of a motion estimation technique depends on the nature of the processes that will interpret the estimates. Each approach has its strengths and weaknesses. Since the success of the interpretive and subsequent processes depends on the properties the motion measurements, it is very important to understand the characteristics of motion estimation techniques. The performance dependencies of motion estimation techniques have rarely been examined. In this paper we analyze one method to determine the sources of error and methods to cope with error.

For many problems gradient-based methods offer significant advantages over matching techniques. Matching techniques are highly sensitive to ambiguity among the structures to be matched. Optical flow can be accurately estimated for only highly distinguishable regions. This means that flow can only be determined at a sparse sampling of points across the image. Furthermore, it is computationally impractical to estimate matches for a large number of points. The gradient-based approach allows optical flow to be simply computed at a more dense sampling of points than can be obtained with matching methods.

Gradient-based techniques avoid the difficult task of finding distinguishable regions or points of interest. The gradient approach leads to algorithms which are characterized by simple computations localized to small regions of the image. These techniques can be applied over the entire image. As we shall see in the analysis that follows, the gradient technique is also sensitive to ambiguous areas—it is impossible to locally determine the motion of a homogeneous region. However, gradient-based estimates are typically available over a greater area than those obtained reliably by matching. In addition, the loss of precision for gradient-based estimates in ambiguous areas can be quantified. Accuracy measurements can be used to weight the contribution of motion estimates in further analysis or to filter poor estimates from the flow field. These accuracy measurements can be obtained as a by-product of the flow estimation process and require little additional computation.

While gradient-based methods have been widely studied, little attention has been paid to the accuracy and reliability of the approach. A major difficulty with gradient-based methods is their sensitivity to conditions commonly

0162-8828/87/0300-0229$01.00 © 1987 IEEE

encountered in real imagery. Highly textured surfaces, motion boundaries, and depth discontinuities can all be troublesome for gradient-based methods. Fortunately, these problematic areas can be identified in the image. In this paper we examine the conditions that lead to errors, methods to reduce errors, and the estimation of measurement errors for one class of gradient-based techniques. By understanding how errors arise we are able to define the inherent limitations of the gradient-based technique, obtain estimates of the accuracy of computed values, enhance the performance of the technique, and demonstrate the informative value of some types of errors.

II. The Gradient Constraint Equation

The gradient constraint equation relates optical flow—velocity on the image (u, v)—and the image brightness function $I(x, y, t)$. The common assumption of gradient-based techniques is that the observed brightness—intensity on the image plane—of any object point is constant over time. Consequently, any change in intensity at a point on the image must be due to motion. Relative motion between the object and camera will cause the position of a point located at point (x, y) at time t to change position on the image over a time interval δt. By the constant brightness assumption, the intensity of the object point will be the same in images sampled at times t and $t + \delta t$. The constant brightness assumption can be formally stated as

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t). \quad (1)$$

Expanding the image brightness function in a Taylor's series around the point (x, y, t) we obtain

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + h.o.t. \quad (2)$$

A series of simple operations leads to the gradient constraint equation:

$$0 = I_x u + I_y v + I_t \quad (3)$$

where

$$I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t}.$$

A detailed derivation is given by Horn and Schunck [6].

III. Gradient-Based Algorithms

The gradient constraint equation does not by itself provide a means for calculating optical flow. The equation only constrains the value of u and v to lie on a line when plotted in flow coordinates.

The gradient constraint is usually coupled with an assumption that nearby points move in a like manner to arrive at algorithms which solve for optical flow. Groups of neighboring constraint equations are used to collectively constrain the optical flow at a pixel. Constraint lines are combined in one of three ways. Methods of *local optimization* [13], [15], [17], [18], [21] solve a set of constraint lines from a small neighborhood as a system of equations. *Global optimization* [5], [6], [19] techniques minimize an error function based upon the gradient constraint and an assumption of local smoothness of optimal flow variations over the entire image. The *clustering* approach [2], [3] operates globally, looking for groups of constraint lines with coinciding points of intersection in flow space.

We will examine the local optimization technique in detail. Some implications of this analysis for other methods will be discussed in the summary section. The analysis is further extended to global methods in [8], [9].

IV. Local Optimization

The method of local optimization estimates optical flow by solving a group of gradient constraint lines obtained from a small region of the image as a system of linear equations. Two constraint lines are sufficient to arrive at a unique solution for (u, v). More than two equations may be included in the system to reduce the effects of errors in the constraint lines. The solution to the overdetermined system may be found by any of a number of error minimization techniques.

We will examine errors in the solution of two equation systems. In practice one should solve an overdetermined system by some method of best fit, such as least squares. The analysis presented here is extended to over-determined systems in [7].

The sources of errors for two equation systems will be examined in Section V. Error reduction techniques and methods to determine the accuracy of optical flow estimates will be discussed in Section VI. The techniques presented in Section VI were implemented and tested with real imagery. The results of this analysis are presented in Section VII.

V. Error Analysis

The pair of equations which we will solve to estimate optical flow at point $\boldsymbol{p}_i = (x_i, y_i, t_i)$ is

$$(i)\ I_x^{(i)} u + I_y^{(i)} v = -I_t^{(i)}$$
$$(j)\ I_x^{(j)} u + I_y^{(j)} v = -I_t^{(j)} \quad (4)$$

where the gradients I_x, I_y, and I_t in equations i and j are evaluated at $\boldsymbol{p}_i$ and a nearby point $\boldsymbol{p}_j$.

The gradients in the system (4) are estimated from discrete images and will be inaccurate due to noise in the imaging process and sampling measurement error. Also, the values of (u, v) at $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ are assumed to be the same. The formulation will be incorrect to the extent that optical flow differs between the two points. The error in the optical flow estimate will depend on the measurement errors in the gradient estimates, the local variation in optical flow, and the error propagation characteristics of the linear system. A system which is very sensitive to small perturbations in the constraint equations is said to ill-conditioned.

We will examine the factors that contribute to gradient measurement errors and consider how violations of the constant flow assumption lead to errors in the estimated flow vector. Next, we will examine the factors that determine the conditioning of the linear system. Finally, we will discuss how all these factors together determine the accuracy of optical flow estimates.

A. Gradient Measurement Error

The estimates of the intensity gradients I_x, I_y, and I_t will be corrupted by errors in the brightness estimates and inaccuracies introduced by sampling the brightness function discretely in time and space. The error in the brightness function is random and results from a variety of sources such as channel noise and quantization of brightness levels. We assume that the brightness error is approximately additive and independent among neighboring pixels. The gradient, estimated from changes in the brightness estimates, will contain a component of random error which is distributed like the error in the brightness function. The random component of the gradient error will be additive and independent of the magnitude of the gradient to the extent that the brightness noise is additive.

The brightness function is sampled discretely in time and space and this will introduce a systematic measurement error into the estimates $\hat{I}_x$, $\hat{I}_y$, and $\hat{I}_t$ of the gradients. The gradient sampling error depends on the second and higher derivatives of the brightness function. To examine the sampling error in $\hat{I}_x$ we expand the brightness function evaluated at $(x + \Delta x, y, t)$ around the point (x, y, t) producing

$$\begin{aligned} &I(x + \Delta x, y, t) \\ &\quad = I(x, y, t) + I_x \Delta x + I_{xx} \Delta x^2 + h.o.t. \end{aligned} \tag{5}$$

where I_x, I_{xx} are the partial derivatives of brightness in the x direction evaluated at (x, y, t). Rearranging terms we obtain an estimate for the brightness gradient in the x direction:

$$\begin{aligned} \hat{I}_x &= \frac{I(x + \Delta x, y, t,) - I(x, y, t)}{\Delta x} \\ &= I_x + I_{xx} \Delta x + h.o.t. \end{aligned} \tag{6}$$

The error $\epsilon_{I_x(sampling)}$ is defined as $\hat{I}_x - I_x$, the difference between the computed and true values. From (6), we obtain the approximate relationship

$$\epsilon_{I_x(sampling)} \approx I_{xx} \Delta x. \tag{7}$$

Likewise, the sampling error in the estimates of I_y and I_t are approximately given by

$$\epsilon_{I_y(sampling)} \approx I_{yy} \Delta y \tag{8}$$

$$\epsilon_{I_t(sampling)} \approx I_{tt} \Delta t. \tag{9}$$

The sampling error for the spatial gradients depends upon the spatial resolution of the camera, Δx and Δy, and the second spatial derivatives of the brightness function, I_{xx}, I_{yy}. The sampling error for the temporal gradient, $\epsilon_{I_t(sampling)}$, is influenced by the frame rate, Δt, and the higher order derivatives of the brightness function over time.

We can express $\epsilon_{I_t(sampling)}$ purely in terms of spatial derivatives and motion. Differentiating the gradient constraint equation (3) with respect to x, y, and t we obtain the following three equations:

$$I_{xx} u + I_x \frac{\partial u}{\partial x} + I_{yx} v + I_y \frac{\partial v}{\partial x} = -I_{tx} \tag{10}$$

$$I_{xy} u + I_x \frac{\partial u}{\partial y} + I_{yy} v + I_y \frac{\partial v}{\partial y} = -I_{ty} \tag{11}$$

$$I_{xt} u + I_x \frac{\partial u}{\partial t} + I_{yt} v + I_y \frac{\partial v}{\partial t} = -I_{tt}. \tag{12}$$

Where the second derivatives of the brightness function exist and are continuous, the left-hand side of (10) and (11) can be substituted for I_{xt} and I_{yt} in (12). Collecting terms we see that

$$\begin{aligned} -[u \quad v] \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} - u \cdot \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \end{bmatrix} \begin{bmatrix} I_x \\ I_y \end{bmatrix} \\ - v \cdot \begin{bmatrix} \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{bmatrix} \begin{bmatrix} I_x \\ I_y \end{bmatrix} + \begin{bmatrix} \frac{\partial u}{\partial t} & \frac{\partial v}{\partial t} \end{bmatrix} \begin{bmatrix} I_x \\ I_y \end{bmatrix} = -I_{tt}. \end{aligned} \tag{13}$$

The first term in (13) depends upon optical flow while the rest of the left-hand side depends upon the derivatives of optical flow over time and space. If optical flow is approximately constant in a small neighborhood and approximately constant over time at each point on the image then

$$[u \quad v] \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \approx I_{tt}. \tag{14}$$

We have derived a constraint equation for second derivatives that is analogous to gradient constraint equation (3).

Without loss of generality, we can rotate our coordinate system so that the flow vector at a point lies along the x-axis. In the new coordinate system we have

$$u^2 I_{xx} \approx I_{tt}. \tag{15}$$

It is evident from (15) that the magnitude of I_{tt} depends upon the second derivative of the brightness function in the direction of motion and the magnitude of motion. The temporal derivative will be well estimated only where the spatial brightness function is nearly linear in the direction of motion.

In summary, the systematic errors in the gradients which make up the coefficients of (4) are given by (7), (8), and (9). In general, the systematic error in estimating I_t is influenced by the magnitude of optical flow and the derivatives of optical flow and the first and second spatial derivatives of brightness. When the x-axis of the coordinate system is aligned with motion and optical flow is nearly constant over time and space we can characterize the systematic error in the temporal derivative by

$$\epsilon_{I_t(sampling)} \approx u^2 I_{xx} \Delta t. \tag{16}$$

If there is significant optical flow, the error given by (16) can become quite large in regions which contain nonlinearities in the brightness function and substantially alter our estimate of I_t.

B. Nonuniformity in the Flow Field

The estimation scheme we have been analyzing assumes that velocity on the image plane is constant in some small neighborhood. This will be true only for very special surfaces and motions. When optical flow is not constant the method can provide a good approximation where flow varies slowly over small neighborhoods.

The true set of equations in (4) should actually be

$$\begin{aligned} I_x^{(i)} u + I_y^{(i)} v &= -I_t^{(i)} \\ I_x^{(j)}(u + \Delta u) + I_y^{(j)}(v + \Delta v) &= -I_t^{(j)} \end{aligned} \tag{17}$$

where the actual flow vectors at points p_i and p_j are (u, v) and $(u + \Delta u, v + \Delta v)$, respectively, and the gradients are estimated at points p_i and p_j. The difference between the true solution and our estimate can be treated as an error on the right-hand side of (17) by distributing the multiplication on the left-hand side of the system and rearranging terms as

$$\begin{aligned} I_x^{(i)} u + I_y^{(i)} v &= -I_t^{(i)} \\ I_x^{(j)} u + I_y^{(j)} v &= -I_t^{(j)} - \epsilon_{\Delta flow} \end{aligned} \tag{18}$$

where

$$\epsilon_{\Delta flow} = [I_x^{(j)}, I_y^{(j)}] \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}. \tag{19}$$

Thus, the error caused by violation of the constant flow assumption can be treated as an additional error in the estimate of I_t.

To examine the significance of this error, we will consider the size of $\epsilon_{\Delta flow}$ relative to $I_t^{(j)}$. But first we will convert to vector notation. Let

$$g_s^{(j)} = \begin{bmatrix} I_x^{(j)} \\ I_y^{(j)} \end{bmatrix}, \quad \omega = \begin{bmatrix} u \\ v \end{bmatrix} \text{ and } \Delta\omega = \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}. \tag{20}$$

For the constraint equation at p_j, we know from (17) and (19) that

$$\left| \frac{\epsilon_{\Delta flow}}{I_t^{(j)}} \right| = \frac{|g_s^{(j)} \cdot \Delta\omega|}{|g_s^{(j)} \cdot (\Delta\omega + \omega)|} \tag{21}$$

$$= \frac{\|\Delta\omega\| \cos\theta_1}{\|\Delta\omega + \omega\| \cos\theta_2} \tag{22}$$

where θ_1 is the angle between the gradient vector, $g_s^{(j)}$, and the local change in optical flow, $\Delta\omega$, and θ_2 is the angle between the gradient vector, $g_s^{(j)}$, and the vector $\Delta\omega + \omega$.

The relative error in I_t depends on the relative lengths of the vectors ω and $\Delta\omega$ and relative magnitudes of the cosines of the angles θ_1 and θ_2. In general, the orientations of the spatial gradient, optical flow, and the local change in optical flow will be independent. Therefore, we expect the relative error in I_t to be strongly related to the relative magnitudes of the flow and change of flow vectors.

In most scenes, flow will vary slowly over most of the image. At surface boundaries we can expect to frequently find discontinuities in optical flow due to discontinuities in motion or depth. Here, the variation in flow will contribute a substantial error and flow estimates will usually be quite poor. However, much of the image will consist of smoothly varying surfaces. When neighboring image points lie on the same smooth surface, flow will generally be similar and hence, the error contributed by variations in flow will be small.

We will consider an example which allows arbitrary three-dimensional translation of a planar surface to demonstrate the important factors influencing the error contributed by variations in optical flow. We consider two neighboring image points that lie on a surface translating with velocity (U, V, W) in three-dimensional space (see Fig. 4). Let the surface be defined by the planar equation

$$Z(X, Y) = R + \alpha X + \beta Y. \tag{23}$$

In Appendix A we derive the following approximate bound

$$\frac{\|\Delta\omega\|}{\|\omega\|} \leq \tan\gamma \left(\|(\alpha, \beta)\| + \frac{|W|}{\|(U, V)\|} \right) \tag{24}$$

where

$$\tan\gamma = \frac{\|(\Delta x, \Delta y)\|}{f}. \tag{25}$$

The angle γ is the angle subtended by $(\Delta x, \Delta y)$ with a focal length of f; this is simply the size of the neighborhood measured in degrees of visual angle. The length of the change-of-flow vector relative to the length of the flow vector depends upon the size of the neighborhood, the slope of the surface viewed, and the ratio of velocity along the line of sight to velocity perpendicular to the line of sight.

Recall that the value given by (24) represents a rough measure of the proportion of error on the right-hand side contributed by variations in optical flow. If the neighborhood is small we expect random errors in the temporal gradient to usually be larger than the error caused by flow variation. The gradient measurement errors discussed in the last section may lead to a much larger degradation. So, for most of the image the error caused by variation in flow should not constitute a problem. However, at surface boundaries optical flow can change dramatically, especially when object motions are allowed. Here, the local optimization result will be a very poor measure of optical flow.

C. Conditioning

The accuracy of the estimates $\hat{u}$ and $\hat{v}$ will depend on the measurement errors in the gradient constraint equations and the error propagation characteristics of the linear system. When a system of linear equations is very sensitive to small errors in the coefficients or right-hand side it is said to be ill-conditioned. If the spatial intensity gradients change slowly, then the linear system will contain constraint lines that are nearly parallel. As a consequence, the system will be nearly singular and small errors in the gradient measurements may result in large changes in the estimated flow value. We will find that the conditioning of the linear system largely depends upon nonlinearities in the brightness function which are perpendicular to the brightness gradient.

If the gradients are known exactly and optical flow is constant then

$$G\omega = -b \tag{26}$$

where

$$G = \begin{bmatrix} I_x^{(i)} & I_y^{(i)} \\ I_x^{(j)} & I_y^{(j)} \end{bmatrix}, \quad \omega = \begin{bmatrix} u \\ v \end{bmatrix} \text{ and } b = \begin{bmatrix} I_t^{(i)} \\ I_t^{(j)} \end{bmatrix}. \tag{27}$$

As before, the rows of G and b are taken from a point p_i and its neighbor p_j. The vector ω will be in error to the degree that the gradient measurements are inaccurate and optical flow varies between points p_i and p_j. The previous section showed that the error accrued when u and v are not constant is the same as that which would be obtained if the b vector is suitably modified as in (18). This error will be absorbed on the right-hand side of (26). Thus, the system which is actually solved is

$$(G + E)(\omega + \delta\omega) = -(b + \delta b) \tag{28}$$

where

$$E = \begin{bmatrix} \epsilon_{I_x^{(i)}} & \epsilon_{I_y^{(i)}} \\ \epsilon_{I_x^{(j)}} & \epsilon_{I_y^{(j)}}' \end{bmatrix},$$

$$\delta b = \begin{bmatrix} \epsilon_{I_t^{(i)}} \\ \epsilon_{I_t^{(j)}} \end{bmatrix} \text{ and } \delta\omega = \begin{bmatrix} \epsilon_u \\ \epsilon_v \end{bmatrix}. \tag{29}$$

The errors in the spatial and temporal gradients arise from both systematic and random measurement errors.

A number of measures of conditioning have been proposed [23]. The most widely used index of conditioning is the condition number, *cond*( ), which is defined as

$$cond(G) = \|G\| \, \|G^{-1}\| \tag{30}$$

for a matrix of coefficients G. The condition number roughly estimates the extent to which **relative** errors in the coefficients and the right-hand side are magnified in the estimate of optical flow. For the problem at hand, the conditioning of the matrix G is determined by the nature of the spatial brightness function over the interval (p_i, p_j).

The inverse of G can be directed calculated as

$$G^{-1} = \frac{1}{I_x^{(i)} I_y^{(j)} - I_x^{(j)} I_y^{(i)}} \begin{bmatrix} I_y^{(j)} & -I_x^{(j)} \\ -I_y^{(i)} & I_x^{(i)} \end{bmatrix} \tag{31}$$

$$= \frac{1}{\|g^{(i)}\| \, \|g^{(j)}\| \sin\phi} \begin{bmatrix} I_y^{(j)} & -I_x^{(j)} \\ -I_y^{(i)} & I_x^{(i)} \end{bmatrix} \tag{32}$$

where $g^{(i)}$ is the spatial gradient vector at p_i and ϕ is the angle between $g^{(i)}$ and $g^{(j)}$.

Before we can evaluate the condition number we must select a matrix norm. We will use the Frobenius norm.[1] We will continue to use the $\|\cdot\|_2$ norm to evaluate vector norms. From the definition of $\|\cdot\|_F$ and the results above we have

$$cond(G) = \frac{\|g^{(i)}\|^2 + \|g^{(j)}\|^2}{\|g^{(i)}\| \, \|g^{(j)}\| \sin\phi} \tag{33}$$

$$= \frac{1}{\sin\phi}\left(\frac{\|g^{(i)}\|}{\|g^{(j)}\|} + \frac{\|g^{(j)}\|}{\|g^{(i)}\|}\right). \tag{34}$$

The magnitude of *cond*(G) depends on the orientations and relative magnitudes of the two spatial gradient vectors. The value of *cond*(G) is minimized when the spatial gradients are perpendicular and have the same magnitude. As the spatial gradients become more nearly parallel the magnitude of *cond*(G) is increased, and hence, error propagation is worsened. Increases in the relative difference in the magnitudes of the spatial gradients also cause *cond*(G) to increase. The magnitude of this effect will not usually be important. If neither of the gradients is very small, then the relative sizes of the gradients will not differ enormously. The gradients will be poorly estimated where they are small, so for multiple reasons estimates will be error prone in these regions.

The most important factor determining conditioning is the angle between the gradients. Where the gradients are nearly parallel, conditioning will be a problem. Thus, if both points lie along a straight edge, we cannot obtain a solution. (This is an example of the aperture problem [5].)

Some higher derivatives of brightness must be large for there to be a significant change in gradient orientation over

[1] The Frobenius norm, $\|\cdot\|_F$, is defined as the square root of the sum of the squares of all the elements. The Frobenius norm can be used to bound the more familiar $\|\cdot\|_2$ norm [20]. It can be shown that

$$\frac{1}{\sqrt{2}}\|\cdot\|_F \le \|\cdot\|_2 \le \|\cdot\|_F.$$

a small neighborhood. Let Δg be the difference between the two gradient vectors. We can expand the gradient in a Taylor series

$$g^{(j)} = g^{(i)} + \begin{bmatrix} I_{xx}^{(i)} & I_{xy}^{(i)} \\ I_{yx}^{(i)} & I_{yy}^{(i)} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + h.o.t. \quad (35)$$

Consequently,

$$\Delta g \approx \begin{bmatrix} I_{xx}^{(i)} & I_{xy}^{(i)} \\ I_{yx}^{(i)} & I_{yy}^{(i)} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}. \quad (36)$$

The angle between the gradients depends on the component of Δg that is perpendicular to $g^{(i)}$. If optical flow is to be accurately estimated in a small region around $p^{(i)}$, then at least one component of the second derivative perpendicular to the gradient must be large. There must be at least some direction in which we can select a neighbor so that the gradient orientation $g^{(j)}$ will differ from $g^{(i)}$. The importance of variations in the spatial gradient for motion detection has been recognized by others. The change in the direction of the spatial gradient has been used to identify corners that can act as feature points for matching algorithms [10]. A similar corner detector has been used to identify a sparse set of locations where a local gradient-based algorithm might yield acceptable results [14]. The examination of conditioning provides an analytical basis for attending to variations in the spatial gradient and a quantitative measure of the error propagation characteristics of a local system of constraint equations.

D. Combining the Sources of Error

We now face a dilemma. We have just shown that some component of the second spatial derivative of the brightness function must be large to minimize error propagation. However, we earlier showed that sampling errors in the gradients were proportional to the magnitude of the second derivative. There is a tradeoff between the gradient measurement errors and conditioning. The problem would not be too serious if we were only concerned about errors in the spatial gradients. If we let the sampling interval be reasonably small with respect to the neighborhood from which we select our equations, we can potentially satisfy both goals—the gradient can change slowly from pixel-to-pixel but the total variation over the neighborhood can be large enough to allow acceptable conditioning.

A serious conflict can arise in the tradeoff between conditioning and sampling errors in the temporal derivative. Recall that the systematic measurement error in $\hat{I}_t$ is proportional to nonlinearities in the spatial brightness function (15). To achieve acceptable conditioning, the spatial gradients in the local set of constraint equations must differ significantly in orientation. For this to happen the spatial brightness function must be nonlinear in some direction. If optical flow is oriented in this direction, then the condition number and measurement errors will be inversely related. Increases in the magnitude of the second spatial derivatives will reduce the condition number and increase the measurement error. Note that there need not be a conflict; optical flow can be perpendicular to direction in which the gradient orientation is varying.

The problem is heightened by the sensitivity of gradient measurements where the flow vector is large. The systematic measurement error in the temporal derivative increases as the square of flow magnitude (15). Where flow is large, even small nonlinearities can contribute significant measurement errors. However, where object points are stationary or moving slowly, the measurement error in the temporal gradient will be negligible and most accurate estimates will be obtained when the orientation of the spatial gradients is highly variable.

As an illustration of the interplay between the concerns of conditioning and measurement error, consider an image painted with an isotropic texture. If the region is stationary then a large amount of detail will be desirable to minimize conditioning. If optical flow is significantly greater than zero, then too much detail will lead to unacceptably large measurement errors. A balance must be struck between these two sources of error.

The conditioning of G can be improved by using a large neighborhood. The risk in choosing neighbors over too great a distance is that the error due to nonconstant flow can become very large. If the neighbors lie on a single surface the contribution of errors due to nonconstant flow will usually grow slowly with neighborhood size. But if neighbors lie on different surfaces their motions may differ substantially. As neighborhood size is increased it becomes more likely that neighbors will lie across a surface boundary and the difference in optical flow will lead to significant errors.

The total error in the flow estimate is determined by the characteristics of the optical flow field, the nature of the brightness function, and the selection rule for constructing the linear system. The sources of error are summarized in Table I. These factors interact in a complex way to determine the accuracy of the local optimization scheme. Only where the contribution of these sources of error is balanced will good estimates be obtained.

VI. Algorithm Extensions Based Upon the Error Analysis

We next consider how knowledge about the causes of errors can be used to reduce errors and introduce techniques to judge the accuracy of estimates. The improvements in performance are based upon parameter selection and preprocessing of the image to extract the most information from a region while minimizing the intrusions of error. A method of iterative refinement [13], [16] is also described.

By examining the image sequence for the conditions which lead to errors we can judge the accuracy with which estimates can be made before the estimate is actually

TABLE I
THE SOURCES OF ERROR IN LOCAL ESTIMATES OF OPTICAL FLOW

Error Source	Determinants
1. Gradient Measurement Error	
(a) random (I_x, I_y, I_t)	(i) ↑↑ sensor noise
	(ii) ↑↑ quantization noise
(b) systematic (I_t)	(i) ↑↑ nonlinearities in the brightness function in the direction of optical flow
	(ii) ↑↑ optical flow magnitude
2. Non-constant Flow	(i) ↑↑ neighborhood size
	(ii) ↑↑ surface slant
	(iii) ↑↑ ratio of velocity along the line of sight to velocity perpendicular to the line of sight[†]
3. Ill-conditioning	(i) ↑↓ neighborhood size
	(ii) ↑↓ sin of the angle between the spatial gradient vectors
	(ii) ↑↑ relative difference in the magnitudes of the spatial gradient vectors

↑↑ error increases with determinant
↑↓ error decreases with determinant
† for translating surfaces

made. Examination of the flow estimate itself can provide additional information about the precision of the estimate. Together, *a priori* and *a posteriori* estimates of accuracy provide a useful heuristic for evaluating the precision of optical flow estimates.

A. Error Reduction Techniques

1) Smoothing: Blurring the image will lead to a smoother, more linear brightness function. Blurring will diminish the systematic error in the gradient estimates by reducing the second and higher derivatives of the spatial brightness function. Random errors will also tend to be reduced by the averaging. An unfortunate consequence of the smoothing is that the error propagation characteristics of the linear system will tend to be worsened. The smoothing will reduce the variation of the gradient, leading to a more ill-conditioned system. The loss of detail is desirable from the standpoint of gradient measurement but undesirable with respect to conditioning. Hence, blurring is most desirable in regions where the systematic error is predominant.

As noted in Section V-A, the systematic error in the gradients depends upon the nonlinearity of the brightness function over the sampling interval. For the temporal gradient, the systematic measurement error depends upon the linearity of the brightness function over the region which moves past a point of observation on the image and the variations of optical flow over time and space. Blurring will be most effective in portions of the image which undergo a significant motion and contain large nonlinearities in the brightness function. The degree of blurring should be sufficient to approximately linearize the brightness function over the region of translation.

The damage which blurring does to the conditioning of the linear system can be counterbalanced by increasing the size of the neighborhood over which the system is constructed. The risk incurred by enlarging the area from which the constraint equations are drawn is that the motions of the points may differ significantly, as could happen if points are on two different surfaces. The selection of the radius of blur and the neighborhood size must be made judiciously so as to avoid increasing the error in the solution vector.

2) Overdetermined Systems: Until this point we have ignored the problem of selecting the direction in which the neighbor is to be chosen to form the linear system. From our previous discussion of error propagation it is clear that the choice of direction can dramatically affect the error in the optical flow estimate. One way to circumvent the difficulty of choosing an appropriate direction is to construct an overdetermined set of equations from points in many directions. The overdetermined system can be solved by minimizing the residual over possible values of optical flow. The choice of the norm to be minimized and the minimization scheme may be an important determinant of the error, but are not analyzed here. As with two equation systems, conditioning will be important for overdetermined systems and conditioning will be related to the same characteristics of the image as in the two equation case. Another approach is to perform the analysis separately in a number of directions and then seek a consensus among the solutions [4].

3) Iterative Registration: If optical flow is known approximately then this knowledge can be used to reduce the error in the local optimization technique. We develop a more general form of the gradient constraint equation that solves for the difference between an approximate estimate and the actual flow. Our derivation abbreviates an analysis presented by Paquin and Dubois [16].

Consider the image sequence that samples the three-dimensional function. We actually estimate the displacement of a point between successive samples of the image sequence. If velocity is constant then the displacement observed on the image over the time interval Δt is $(u\Delta t, v\Delta t)$. Let $\boldsymbol{d}$ be a displacement vector in three-dimensional x, y, t-space. Let $\hat{\boldsymbol{d}}$ be an estimate of $\boldsymbol{d}$. Given a displacement estimate

$$\hat{\boldsymbol{d}} \equiv \begin{bmatrix} \hat{u}\Delta t \\ \hat{v}\Delta t \\ \Delta t \end{bmatrix} = \begin{bmatrix} x\text{-component of displacement} \\ y\text{-component of displacement} \\ t\text{-component of displacement} \end{bmatrix} \tag{37}$$

we can estimate optical flow by $(\hat{u}, \hat{v})$.

The vector $\hat{\boldsymbol{d}}/\|\hat{\boldsymbol{d}}\|$ is a unit vector in the direction of the estimated displacement. The gradient of I in this direction is

$$I_{\hat{d}} = \frac{\hat{\boldsymbol{d}}^T}{\|\hat{\boldsymbol{d}}\|}\begin{bmatrix} I_x \\ I_y \\ I_t \end{bmatrix} = \frac{1}{\|\hat{\boldsymbol{d}}\|}(I_x\hat{u} + I_y\hat{v} + I_t)\Delta t \quad (38)$$

$$= \frac{1}{\|\hat{\boldsymbol{d}}\|}(I_x\hat{u} + I_y\hat{v} - I_x u - I_y v)\Delta t \text{ (using(3))} \quad (39)$$

$$= \frac{1}{\|\hat{\boldsymbol{d}}\|}(I_x\delta u + I_y\delta v)\Delta t \quad (40)$$

where $\delta u = \hat{u} - u$ and $\delta v = \hat{v} - v$ are the errors in the estimated flow velocities. Finally, we get an expression that relates the error in the displacement estimate to measurable brightness gradients.

$$\|\hat{\boldsymbol{d}}\| I_{\hat{d}} = I_x\delta u\Delta t + I_y\delta v\Delta t = (I_x, I_y, 0)\cdot \Delta\boldsymbol{d} \quad (41)$$

where

$$\Delta\boldsymbol{d} = \hat{\boldsymbol{d}} - \boldsymbol{d} = \begin{bmatrix} \delta u\Delta t \\ \delta v\Delta t \\ \Delta t \end{bmatrix}. \quad (42)$$

We can compute an estimate of the quantity (41) by using the Taylor expression

$$I(x + \hat{u}\Delta t, y + \hat{v}\Delta t, t_1 + \Delta t) = I(x, y, t_1) + \|\hat{\boldsymbol{d}}\| I_{\hat{d}}. \quad (43)$$

Solving for $I_{\hat{d}}$ and combining with (41) yields the approximation

$$(I_x, I_y)\cdot\begin{bmatrix} \delta u \\ \delta v \end{bmatrix}\Delta t = I(x + \hat{u}\Delta t, y + \hat{v}\Delta t, t_1 + \Delta t) - I(x, y, t_1). \quad (44)$$

The new constraint equation (41) is a more general form of the gradient constraint equation. The more general form relates the gradient in an arbitrary direction to the spatial gradients and optical flow. If the displacement estimate is $(0, 0, \Delta t)$, then $I_{\hat{d}} = I_t$.

The general form of the gradient constraint equation can be used to refine an estimate $\hat{\boldsymbol{d}}$ by solving for $\Delta\boldsymbol{d}$. The spatial components of displacement, $(\delta u\Delta t, \delta v\Delta t)$, can be estimated with (44) using local optimization. This process can be performed iteratively to find successively better estimates of optical flow. An improvement can be expected, on the average, whenever successive registrations are closer to the true displacement vector:

$$\|\Delta\boldsymbol{d}_{i+1}\| \leq \|\Delta\boldsymbol{d}_i\| \qquad i = 1, 2, \cdots. \quad (45)$$

The improvement arises from successively better estimates of $I_{\hat{d}}$. As was demonstrated earlier in (13) the systematic error in the estimate of temporal derivative grows as the square of flow magnitude. The same relationship is true for directional derivative $I_{\hat{d}}$ and the flow difference in the general constraint equation.

Solving for the difference between an estimate of optical flow and the true optical flow is computationally equivalent to registering a portion of an image pair and estimating the change of position in the adjusted sequence. For this reason the technique has been called *iterative registration* [13]. The estimate of optical flow may be derived from estimates made at some previous time or from prior processing on a single frame pair.

Note that if the inequality of (45) does not hold, then the error might be expected to increase. If an estimate of optical flow is poor then the refinement effort may lead to an even larger error. The next section is devoted to methods to evaluate the quality of optical flow estimates. A measure of the accuracy of a flow estimate can be used to judge whether or not the estimate should be used for registration. Alternatively, the degree of registration can be based on the confidence put in the flow estimate, the more accurate the estimate is judged to be, the more that the frame pair should be adjusted in the direction of the estimate.

The iterative registration technique can be combined with variable blurring to produce a coarse-to-fine system for estimating optical flow [13]. Flow is roughly estimated with an image sequence which has been blurred so that the brightness function is approximately linear over areas the size of the maximum expected displacement. The coarse estimate of optical flow is used, at each point, to register a small region of the image at a finer level of resolution. This process is repeated at successively finer levels of resolution.

How much advantage can be gained from iterative registration? The spatial variation of optical flow will not be affected by registration. Thus, the error due to incompatibilities among equations in the linear system is unaffected by iterative registration. Also, the estimate of the directional gradient will contain some amount of random measurement error even if successive frames are in perfect registration. The propagation of these errors depends primarily upon the conditioning of $\|\boldsymbol{G}\|$, which is not influenced by registration. We cannot expect to reduce the error in $\hat{\boldsymbol{d}}$ below that caused by random error in $I_{\hat{d}}$ and nonconstant flow through iterative registration.

While performing a coarse-to-fine registration the degree of blurring at each stage should be appropriate to the expected error in optical flow at the next more coarse level of analysis. In the absence of knowledge about the motions of individual points the blurring must be performed uniformly across the image. While the error will, on the average, be reduced for points which translate significantly, the error will tend to be increased for points which are stationary or move very little. No benefit is obtained by linearizing the brightness function at stationary regions and the error propagation characteristics are worsened. Some of the accuracy lost at stationary regions during coarse processing might be recovered at finer levels but, in general, the best estimates could be obtained at a fine level without registration. In the next section methods are developed to estimate the accuracy of optical flow estimates. This information can be used in the coarse-to-fine system of iterative registration to judge whether an improvement has been obtained at each level. *A priori* estimates of the magnitude of flow are also developed in the next section. The iterative registration technique can be improved by adapting the technique to knowledge about the accuracy of estimates and the magnitude of motion.

B. Estimating Error

Many of the factors which lead to errors in the local optimization estimation technique can be identified and measured from the image. The error propagation characteristics of the linear system can be estimated from the matrix of spatial gradients. The degree to which relative errors are magnified is indicated by $cond(\boldsymbol{G})$. Regions of the image for which the propagation characteristics are poor will be very sensitive to small measurement errors in the gradients. The optical flow estimates obtained in these regions are likely to be inaccurate.

The systematic measurement error in $\hat{I}_t$ was shown to depend upon the linearity of the brightness function in the direction of motion (13). One way to measure of the nonlinearity of the brightness function is to compare the spatial gradients of brightness in successive frames [3], [13]. If $I_x(x, y, t)$ is significantly different from $I_x(x, y, t + \delta t)$ then it can be inferred that the estimate of the temporal gradient is likely to be in error.

Once an estimate has been obtained we can bound the error by referring back to the image. The following *a posteriori* error bound can be derived from (44):

$$\left\|(\delta u, \delta v)\right\| \Delta t \geq \frac{I(x + \hat{u}\Delta t, y + \hat{v}\Delta t, t_1 + \Delta t) - I(x, y, t_1)}{\left\|(I_x, I_y)\right\|}. \tag{46}$$

If the norm of the spatial gradient is not too small, this will provide a good measure of the magnitude of the error in the flow estimate.

If an overdetermined set of equations is used to estimate optical flow, then measurement errors in the gradients and incompatibilities among the constraint equations due to differential motion will be reflected in the residual of the solution. The residual vector can be estimated by

$$\boldsymbol{G}\hat{\boldsymbol{\omega}} + \boldsymbol{b} = \boldsymbol{r} \tag{47}$$

where $\hat{\boldsymbol{\omega}}$ is the estimated optical flow and $\boldsymbol{r}$ is the residual. A large residual indicates that substantial errors exist in the system and that the estimated flow vector is likely to be inaccurate.

The residual vector will be especially large at occlusion edges where the change in flow is discontinuous. It has been proposed that the residual error be used as an indication of the presence of an occlusion edge [21]. To be identifiable, the change in optical flow across an occlusion edge must lead to an error which is greater than that normally encountered from other measurement errors. A threshold on the residual must be established which will normally be exceeded only at significant discontinuities in the flow field. The error accrued from a change in the flow vector is equivalent to a measurement error on the right-hand side of the local optimization system. Since the equivalent error on the right-hand side is magnified by the size of the spatial gradients, the threshold for identifying large residual errors should be adaptive to the spatial gradients. Likewise, it was shown that the systematic measurement errors in the gradients were related to the second derivatives of brightness, so the threshold on the residual should depend upon the second derivatives, as well.

VII. Methods

The gradient-based approach is demonstrated with two versions of the local optimization technique. The first method implements a simple local optimization. The second method combines local optimization with iterative registration. Both methods assign a confidence to optical flow estimates.

A. Simple Local Optimization

The basic local optimization method performs a least squares minimization on an overdetermined set of gradient constraint equations to estimate optical flow at each point. Each image is first blurred with a Gaussian blurring function. The standard deviation of the blurring function used to collect the data presented here was about 2 pixels. The blurring serves to reduce the noise in the image and linearize the brightness function.

Constraint equations from a group of neighboring points are gathered to produce an overdetermined system of linear equations of the form

$$\boldsymbol{G}\boldsymbol{\omega} = -\boldsymbol{b} \tag{48}$$

where

$$\boldsymbol{G} = \begin{bmatrix} I_x^{(1)} & I_y^{(1)} \\ I_x^{(2)} & I_y^{(2)} \\ \vdots & \vdots \\ I_x^{(n)} & I_y^{(n)} \end{bmatrix}, \quad \boldsymbol{\omega} = \begin{bmatrix} u \\ v \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} I_t^{(1)} \\ I_t^{(2)} \\ \vdots \\ I_t^{(n)} \end{bmatrix}. \tag{49}$$

Each row of $\boldsymbol{G}$ and $\boldsymbol{b}$, is evaluated at a different point. To ensure that the equations are sufficiently distinct we selected neighbors from a 5×5 window centered around the point to be estimated.

In general, the overdetermined system (48) has no exact solution. An approximate solution is found by minimizing the residual vector $\boldsymbol{r}$, defined in (47). The flow estimate is chosen to be the vector $\boldsymbol{\omega}$ which minimizes some criteria function of $\boldsymbol{r}$. In our work we minimize $\|\boldsymbol{r}\|_2$ by letting

$$\hat{\boldsymbol{\omega}} = \boldsymbol{G}^{+}\boldsymbol{b} \tag{50}$$

where G^+ is the pseudo inverse of G [20]. Calculation of the pseudo inverse requires the inversion of the 2×2 matrix $G^t - G$. The inverse will not exist where the local gradients do not sufficiently constrain optical flow to allow for an exact solution. In this case the confidence of the flow estimate is set to zero and u and v are undefined.

A confidence is assigned to each flow estimate on the basis of:

1) an estimate of the measurement error in the temporal gradient,
2) an estimate of the conditioning,
3) the size of the residual vector $\boldsymbol{r}$, and
4) the *a posteriori* bound given by (46).

The importance of each of these factors in determining the accuracy of estimates is discussed above. That analysis does not, however, provide us with a formula for estimating the total error in the flow vector (u, v). We must find a means to combine several factors which each indicate the presence of conditions which can be lead to errors.

Recall how each factor outlined above relates to the error in (u, v). The systematic measurement error in the temporal gradient depends on the linearity of the brightness function. The change in the spatial gradients between successive frames provides an indication of the linearity of the brightness function over the region which has translated by a point [13]. To obtain an estimate of the contribution of this error to errors in $\hat{\boldsymbol{\omega}}$, we divide the magnitude of the change in the spatial gradients by the magnitude of the spatial gradient.

The error propagation characteristics of the linear system $G\hat{\boldsymbol{\omega}} = \boldsymbol{b}$ can be determined by examining the matrix of spatial gradients. If linear system is ill-conditioned, small measurement errors will tend to produce large errors in $(\hat{u}, \hat{v})$.

The residual vector indicates the degree to which the estimated flow vector jointly satisfies the system of constraint equations. But the value of the residual vector is not easy to interpret because the size of the residual is dependent on the overall magnitude of the brightness gradients. We normalize the residual by determining, for each equation, the minimum distance between the estimate and the equation. This is equal to the distance between the estimate and the constraint equation along a line perpendicular to the constraint equation that passes through the estimate. The average minimum distance is used as an index of the degree to which the equations are satisfied.

Once an estimate has been obtained, the *a posteriori* error bound given by (46) can be used to judge the accuracy of the estimate. In locations where this bound is large the computed optical flow vector is likely to be in error.

Each of the measurements described above provides an index of the expected error in the flow estimate. The four error estimates are not independent. The residual error and the *a posteriori* bound measure the accumulative error, from all sources, in the flow estimate. The variation in the spatial gradient and the conditioning of G measure conditions which are likely to lead to poor estimates: nonlinearity in the spatial brightness function is particularly troublesome for gradient measurement and the conditioning of G conveys the error propagation characteristics of the linear system. Even though the four estimates are not independent we found that they were best treated as separate sources of information and best combined multiplicatively. We examined a number of combination rules and found that the results were not highly sensitive to the particular rule for combining error estimates. A measure of confidence was obtained from the inverse of the error estimates. The confidence value can be interpreted as a rough measure of the likelihood that an optical flow estimate is correct.

B. Local Optimization with Iterative Registration

The simple method of local optimization can be extended by a method of iterative refinement. Flow estimates are used to register the frame pair on each successive iteration of the estimation procedure. It was earlier shown that the measurement error in the temporal gradient could be significantly reduced if the registration locally reduced the displacement of the image frames. Since the optical flow field will usually contain variations, the predicted registration will differ across the image. To obtain a consistent linear system, a small region of the first frame must be registered with the second frame on the basis of the predicted flow at the point for which optical flow is to be estimated. A system of linear equations is constructed from the registered region.

This process can be performed iteratively, using the optical flow estimated at the previous stage to register the frame pair on the next iteration. It is important to emphasize that, at each stage, the registration can only be expected to improve performance when the new registration is an improvement over the registration in the last iteration. Otherwise, the new estimate of optical flow will, in general, be worse then the previous estimate. Since it is desirable to register the image only where the flow estimates are believed to be correct, we register in proportion to the confidence in the flow estimate. A flow field of zero flow vectors is used to initialize the first iteration.

The iterative registration technique is employed with variable blurring to produce a coarse-to-fine system of analysis. Images are blurred with a Gaussian weighting function. In early iterations the standard deviation of the Gaussian weighting function is large. The standard deviation of the weighting function is reduced in each successive iteration. At each level, the radius of the blurring function should be large enough to guarantee that the brightness function is approximately linear over the maximum expected flow from the registered images.

The size of the neighborhood from which the constraint equations are selected must depend upon the amount which the images are blurred. At a coarse level of analysis there is little detail which distinguishes nearby points. To obtain sufficiently different constraint equations, the sep-

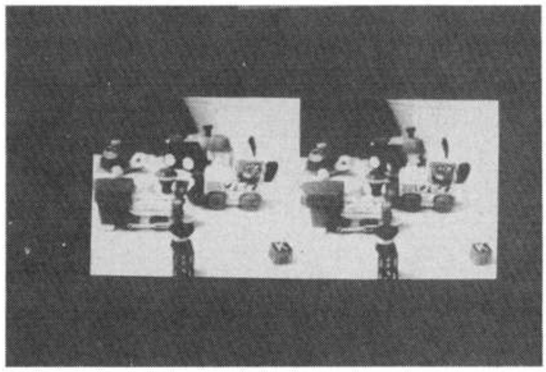
TABLE II
THE COARSE-TO-FINE ANALYSIS

Iteration	Blur Radius σ	Neighborhood Size
1	7	6
2	5	4
3	3.5	3
4	2	2

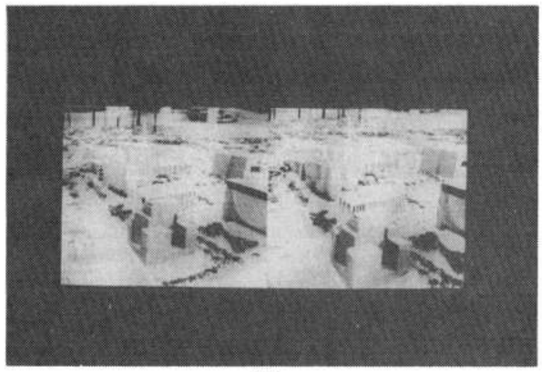
aration between observation points must be increased; otherwise, the conditioning of the linear system will degenerate.

Our system contains four iterations which correspond to four levels of coarseness. The neighbor size and the value of the standard deviation for the approximation to the Gaussian weighting function are given in Table II for each of the four iterations.

A difficulty with the coarse-to-fine system is that the flow estimates for stationary and slowly moving points made at coarse levels may be worse than the initially assumed zero vector. To ensure that the new flow estimate made at one level is not worse than the value input into the level, we examine the error bound given by (46) for both the initial and new estimates. If the error bound for the new estimate is significantly larger than the bound for the old estimate, it is ignored.



(a)



(b)

Fig. 1. Image sequences.

C. Results

The two methods described above were tested with the two image pairs presented in Fig. 1. In the first sequence the camera was stationary. The two toy trains in the center of the first image move toward each other in the second image. The second sequence simulates a view from an aircraft flying over a city. The scene is actually a model of downtown Minneapolis. (This picture originally appeared in Barnard's thesis [1].)

The optical flow fields obtained with the simple local optimization technique are shown in Fig. 2(a) and (b) for the moving trains and flyover scenes. Associated with each vector is a confidence in the correctness of the value. A threshold on confidence was established which produced a reasonably dense sampling of mostly correct values. Only vectors which exceeded the confidence threshold are displayed. The resulting field was too dense to clearly display the entire field. Consequently, only 20 percent of the vector fields are shown in Fig. 2.

The results of the coarse-to-fine method of iterative refinement are shown in Fig. 2(c) and (d). Confidence thresholds were established which produced vector densities which were comparable to that obtained with simple local optimization. Both techniques produce reasonably accurate results with the moving train sequence.

The two techniques are more easily distinguished on the basis of their performance with the flyover sequence. The simple local optimization method produces a large number of errors even for the relatively sparse sampling of vectors displayed in Fig. 2(b). The method of iterative registration generated many fewer errors in fields which are much more dense than that obtained with the simple local optimization approach.

Note the areas where very few vectors are displayed. Optical flow is poorly estimated in these regions and low values of confidence are assigned to the estimates obtained there. The problematic regions are usually fit into one or more of the following characterizations:

1) largely homogeneous regions,
2) highly textured regions which are moving, or
3) regions which contain large discontinuities in the flow field.

Optical flow estimates obtained in homogeneous areas are likely to be in error because of the poor conditioning of linear systems constructed in these regions. The temporal gradient is poorly measured in highly textured regions which undergo significant motion. In regions which contain large discontinuities in the flow field the temporal gradient is poorly estimated and the systems of equations from the region are likely to contain inconsistencies.

The success with which confidence estimates predict the accuracy of flow estimates is demonstrated in Fig. 3. The flow field produced by the simple local optimization technique with the moving trains sequence is displayed with a low threshold on confidence in Fig. 3(a) and a high threshold in Fig. 3(b). As before, only 20 percent of the vectors which exceed the threshold are displayed. Similar thresholds are shown for the method of iterative registra-

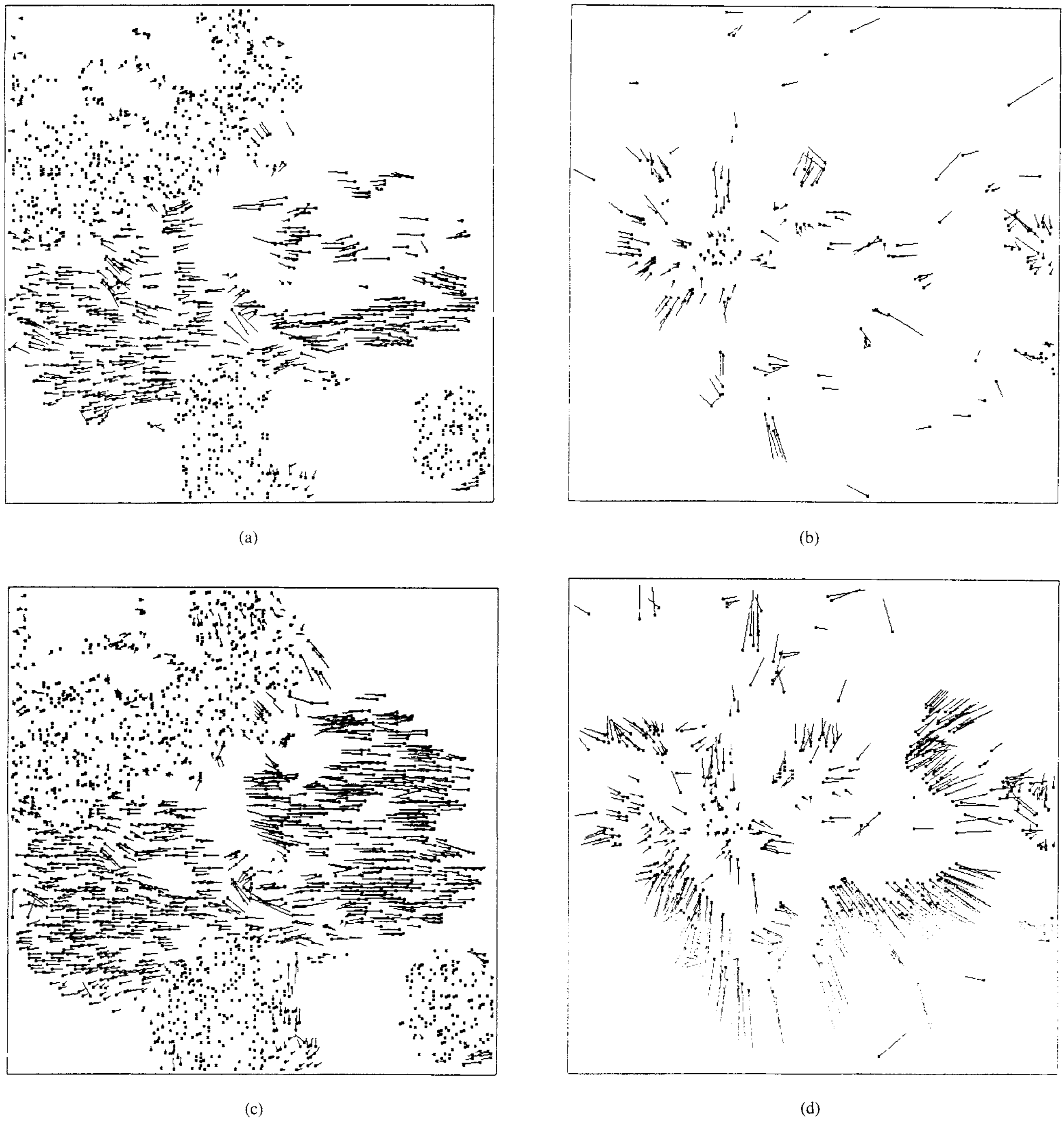


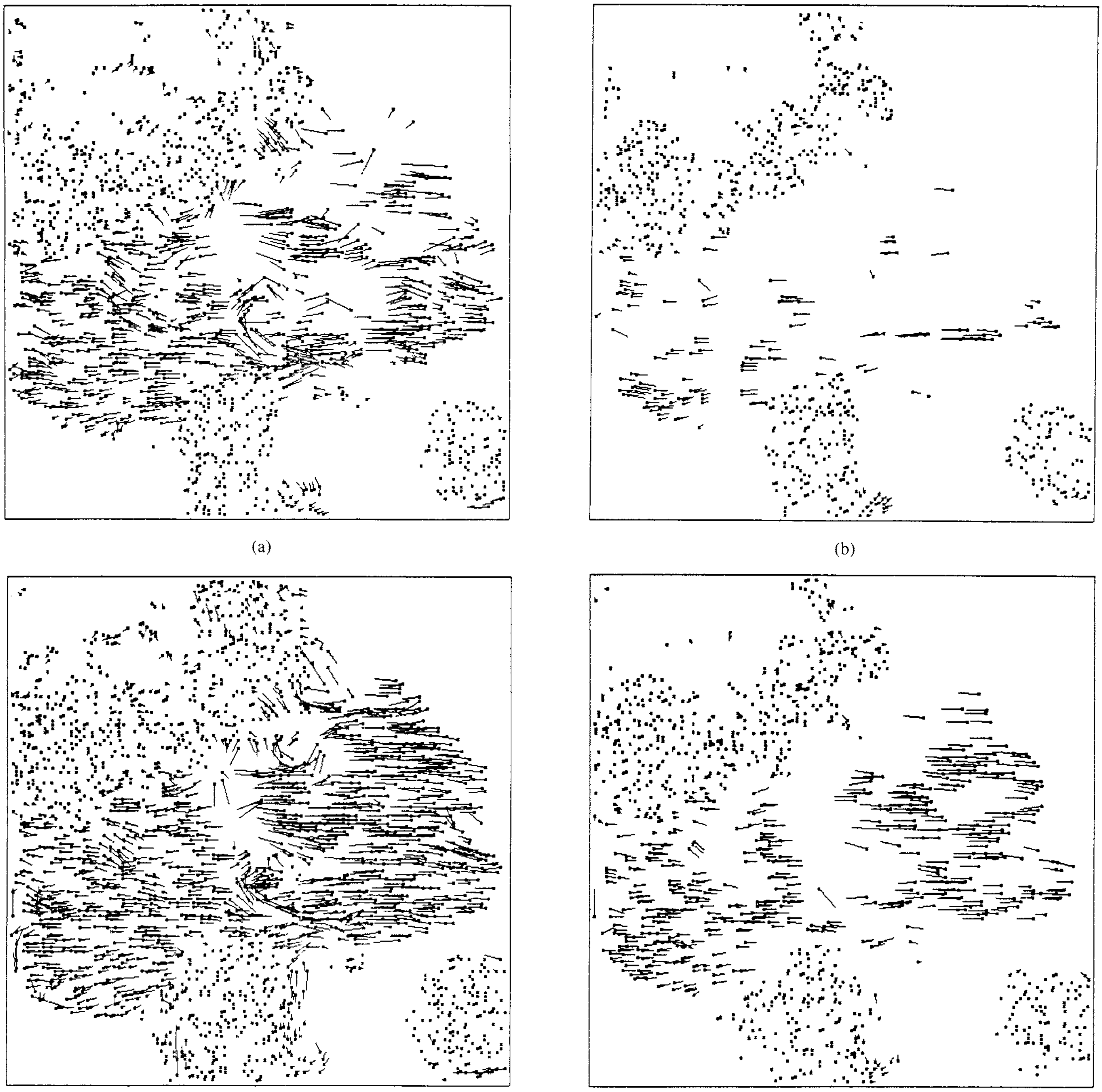
Fig. 2. Optical flow estimates.

tion in Fig. 3(c) and (d). For both methods confidence provides a reasonable index of the accuracy of flow estimates. A sparse sampling of accurate estimates exceeds the high confidence threshold. When the threshold is lowered, more dense fields are obtained with a significantly greater number of bad vectors.

VIII. Summary

Gradient-based methods that locally solve for optical flow suffer from three principal sources of error. The first difficulty is that the brightness gradients will be poorly estimated in regions that are highly textured. This problem is most significant for estimates of the temporal brightness gradient in moving regions. Secondly, variations in optical flow across the image violate the assumption of locally constant optical flow. The analysis presented here suggests that changes in optical flow will contribute a significant error only at discontinuities in the flow field. Finally, there must be sufficient local variation in the orientation of the brightness gradient to avoid poor error propagation characteristics associated with ill-conditioned systems.

The problems discussed here apply to other gradient-based methods, as well. All methods require that the gradient constraint equation be accurately determined. The



(a) (b) (c) (d)

Fig. 3. The accuracy of confidence estimate. Optical flow estimates exceeding low and high thresholds on confidence are displayed.

brightness gradients will be susceptible to the same measurement errors for all the methods. If the estimation technique can be structured to iteratively refine optical flow estimates then iterative registration may greatly improve performance. If the registration reduces the magnitude of the displacement of a region between successive frames, then the temporal brightness gradient will usually be better estimated.

Large variations in optical flow can be disruptive for any method that assumes smoothness or constancy of optical flow. The local and global methods rely on a similar smoothness assumption. Both methods require that flow vary slowly across the image. The locally constructed system of constraint equations is solved as if optical flow is constant over the neighborhood from which the constraint lines are collected. When optical flow is not constant the local method can provide a good approximation where flow varies slowly over small neighborhoods. The global method seeks a solution which minimizes local variation in flow. There are contrasting aspects in the performance of the two approaches that are directly related to the difference in the scope of interactions across the image.

The narrow focus of local methods leads to a major problem over portions of the image. In regions where the spatial gradients change slowly the local system will be poorly conditioned. This problem is reduced in the global optimization method. Information is propagated over the image, so regions which have insufficient local constraints will benefit from the estimates at surrounding regions.

The local and global methods share a common weakness. Where flow changes sharply estimates will be very inaccurate. The affect of these errors is limited by the neighborhood size in the local method. In contrast, global methods may propagate these errors throughout the image. While the global sharing of information is beneficial for constraint sharing, it is detrimental with respect to error propagation. The problem is quite severe. Without some capability to constrain interactions to separate regions that satisfy the smoothness assumption, global optimization methods are practically useless for most real imagery. In another paper we describe a global method whereby the influence of a point on its neighbor is proportional to the judged correctness of the information to be shared [9]. In this way the mutual constraint of neighbors is controlled and the propagation of errors limited.

The empirical results presented in this paper demonstrate the potential for gradient-based methods. Reasonably dense sets of accurate flow vectors were obtained for both image sequences. The best success was achieved with an approach that combined a coarse-to-fine analysis with iterative registration.

The results demonstrate the feasibility of measuring the quality of optical flow estimates. The local method is susceptible to a variety of problems and tends to produce very poor estimates in troublesome areas of the image. Without accurate estimates of confidence, good estimates cannot be distinguished from bad and the local techniques are of little use.

This work emphasizes the importance of understanding the mechanisms which underlie computational methods. An awareness of the strengths and weaknesses of methods and the way in which they operate can lead to adaptations and enhancements which are of great practical value.

Appendix A
Optical Flow Variations

Several papers have examined the relationship between the three-dimensional motion of objects and observers and the characteristics of the optical flow field. We will consider an example which allows arbitrary three-dimensional translation of a planar surface to demonstrate the important factors influencing changes in optical flow over the image.

Let the three-dimensional coordinate system be attached to the camera as in Fig. 4 which is redrawn from Longuet-Higgins and Prazdny [12]. All motion is associated with the camera. Let U, V, and W be the translational velocities of the observer in the X, Y, and Z directions. When motion is constrained to translation, the components of the three-dimensional velocity vector are

$$X' = -U \qquad Y' = -V \qquad Z' = -W. \tag{A.1}$$

Using a perspective projection, the position of an object point on the image is related to its three-dimensional position by

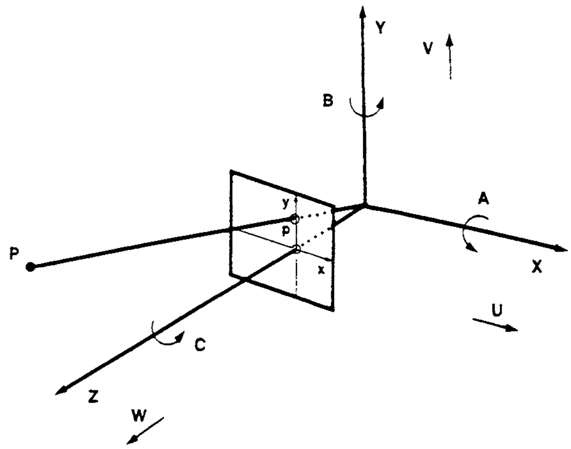


Fig. 4. The camera-based coordinate system.

$$x = \frac{fX}{Z} \quad y = \frac{fY}{Z} \tag{A.2}$$

where f is the focal length of the camera. Velocity on the image plane (u, v) at a point (x, y) is

$$u = x' \quad v = y'. \tag{A.3}$$

Substituting from (A.2) into the right-hand side of (A.3) and differentiating we obtain

$$u = f \cdot \left(\frac{X'}{Z} - \frac{XZ'}{Z^2}\right) = \frac{-fU + xW}{Z} \tag{A.4}$$

and

$$v = f \cdot \left(\frac{Y'}{Z} - \frac{YZ'}{Z^2}\right) = \frac{-fV + yW}{Z}. \tag{A.5}$$

Consider a point $\boldsymbol{P}_0$ on the surface of a rigid body which projects to $\boldsymbol{p}_0$ on the image. We orient the coordinate system so that $\boldsymbol{P}_0$ lies on the observer's line of sight.[2] The three-dimensional coordinates of $\boldsymbol{P}_0$ are $(0, 0, R)$ and the position of $\boldsymbol{p}_0$ on the image is $(0, 0)$. We assume that the surface is planar so that

$$Z(X, Y) = R + \alpha X + \beta Y \tag{A.6}$$

for points on the surface near $\boldsymbol{P}_0$.

Following Longuet-Higgins and Prazdny, we introduce the dimensionless coordinate

$$z = f \cdot \left(\frac{Z - R}{Z}\right) = \alpha x + \beta y. \tag{A.7}$$

The components of optical flow formalized in (A.4) and (A.5) can be rewritten as

$$u = \left(\frac{-fU + xW}{R}\right)\left(1 - \frac{z}{f}\right) \tag{A.8}$$

[2]This coordinate transformation is not strictly correct for a planar retina as pictured in Fig. 4. The change of coordinates can be justified in several ways. It can be assumed that the retina is globally spherical, but can locally be modeled as planar. Or, it can be assumed that the distance of $\boldsymbol{p}_0$ from the origin is sufficiently small relative to the focal length that the distortion introduced by the transform will be minimal. Or finally, we can simply restrict our attention to points along the line of sight.

and

$$v = \left(\frac{-fV + yW}{R}\right)\left(1 - \frac{z}{f}\right). \quad \text{(A.9)}$$

The surface is assumed to be planar, so the derivatives of u and v with respect to x and y are well defined. At the point p_0, where $x = y = z = 0$, u and v are

$$u = -\frac{fU}{R} \quad \text{and} \quad v = -\frac{fV}{R}. \quad \text{(A.10)}$$

The derivatives of u and v are given by

$$u_x = \frac{\alpha U + W}{R} \qquad u_y = \frac{\beta U}{R} \quad \text{(A.11)}$$

$$v_x = \frac{\alpha V}{R} \quad \text{and} \quad v_y = \frac{\beta V + W}{R} \quad \text{(A.12)}$$

since

$$z_x = \alpha \quad \text{and} \quad z_y = \beta. \quad \text{(A.13)}$$

Recall that the error incurred by assuming constant flow could be treated as measurement error in I_t, on the right-hand side of (18). The magnitude of this error, relative to I_t, is strongly dependent on the ratio of the magnitude of the change in optical flow to the magnitude of the flow vector. We can now express the ratio of change-of-flow to flow in terms of the three-dimensional parameters of shape and motion, and the viewing angle. The change in optical flow between two points separated by $(\Delta x, \Delta y)$ is

$$(\Delta u, \Delta v) = (\Delta x u_x + \Delta y u_y, \Delta x v_x + \Delta y v_y). \quad \text{(A.14)}$$

Inserting the appropriate terms from (A.11) and (A.12) into (A.14) and dividing by optical flow as given by (A.10), we arrive at an expression for the ratio of change-of-flow to flow at a point:

$$\frac{\left\|\begin{bmatrix}\Delta u\\ \Delta v\end{bmatrix}\right\|}{\left\|\begin{bmatrix}u\\ v\end{bmatrix}\right\|} = \frac{\left\|\begin{bmatrix}(\alpha U + W)\Delta x + \beta U \Delta y\\ \alpha V \Delta x + (\beta V + W)\Delta y\end{bmatrix}\right\|}{\left\|\begin{bmatrix}fU\\ fV\end{bmatrix}\right\|} \quad \text{(A.15)}$$

$$= \frac{\left\|\begin{bmatrix}U\\ V\end{bmatrix}\left([\alpha, \beta]\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}\right) + W\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}\right\|}{f\left\|\begin{bmatrix}U\\ V\end{bmatrix}\right\|} \quad \text{(A.16)}$$

$$\leq \tan\gamma\left(\left\|[\alpha, \beta]\right\| + \frac{\|W\|}{\left\|\begin{bmatrix}U\\ V\end{bmatrix}\right\|}\right) \quad \text{(A.17)}$$

where

$$\tan\gamma = \frac{\left\|\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}\right\|}{f}. \quad \text{(A.18)}$$

The angle γ is the angle subtended by $(\Delta x, \Delta y)$ with a focal length of f; this is simply the size of the neighborhood measured in degrees of visual angle. The length of the change-of-flow vector relative to the length of the flow vector depends upon the size of the neighborhood, the slope of the surface viewed, and the ratio of velocity along the line of sight to velocity perpendicular to the line of sight.

REFERENCES

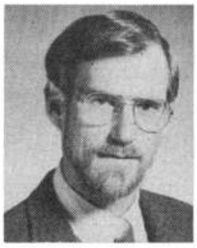
[1] S. T. Barnard, "The image correspondence problem," Ph.D. dissertation, Dep. Comput. Sci., Univ. Minnesota, Dec. 1979.
[2] C. Cafforio and F. Rocca, "Methods for measuring small displacements of television images," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 573–579, Sept. 1976.
[3] C. L. Fennema and W. B. Thompson, "Velocity determination in scenes containing several moving objects," *Comput. Graphics Image Processing*, vol. 9, pp. 301–315, Apr. 1979.
[4] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, June 1981.
[5] E. C. Hildreth, "Computations underlying the measurement of visual motion," *Artificial Intell.*, vol. 23, pp. 309–354, 1984.
[6] B. K. P. Horn and B. Schunck, "Determining optical flow," *Artificial Intell.*, vol. 17, pp. 185–203, 1981.
[7] J. K. Kearney, "Gradient-based estimation of optical flow," Ph.D. dissertation, Dep. Comput. Sci., Univ. Minnesota, Aug. 1983.
[8] J. K. Kearney and W. B. Thompson, "Gradient-based estimation of optical flow with global optimization," in *Proc. First Conf. Artificial Intell. Applications*, Denver, CO, Dec. 1984, pp. 376–380.
[9] —, "Bounding constraint propagation for optical flow estimation," in *Motion Understanding*, J. Aggarwal and W. Martin, Ed. Hingham, MA: Kluwer Academic, 1986.
[10] L. Kitchen and A. Rosenfeld, "Gray-level corner detection," Comput. Sci. Center, Univ. Maryland, College Park, Tech. Rep. TR-887, 1980.
[11] J. O. Limb and J. A. Murphy, "Estimating the velocity of moving images in television signals," *Comput. Graphics Image Processing*, vol. 4, pp. 311–327, Dec. 1975.
[12] H. C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," *Proc. Roy. Soc. London*, vol. B-208, pp. 385–397, 1980.
[13] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 5th Int. Joint Conf. Artificial Intell.*, Aug. 1981, pp. 674–679.
[14] H.-H. Nagel, "Displacement vectors derived from second-order intensity variations in image sequences," *Comput. Vision, Graphics, Image Processing*, vol. 21, pp. 85–117, 1983.
[15] A. N. Netravali and J. D. Robbins, "Motion-compensated television coding: Part I," *Bell Syst. Tech. J.*, vol. 58, no. 3, Mar. 1979.
[16] R. Paquin and E. Dubois, "A spatio-temporal gradient method for estimating the displacement field in time-varying imagery," *Comput. Vision, Graphics, Image Processing*, vol. 21, no. 2, pp. 205–221, Feb. 1983.
[17] R. J. Schalkoff, "Algorithms for a real-time automatic video tracking system," Ph.D. dissertation, Univ. Virginia, May 1979.
[18] R. J. Schalkoff and E. S. McVey, "A model and tracking algorithm for a class of video targets," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 2–10, Jan. 1982.
[19] B. G. Schunck and B. K. P. Horn, "Constraints on optical flow," in *Proc. IEEE Conf. Pattern Recognition and Image Processing*, Aug. 1981, pp. 205–210.
[20] G. W. Stewart, *Introduction to Matrix Computations.* New York: Academic, 1973.

[21] W. B. Thompson and S. T. Barnard, "Low-level estimation and interpretation of visual motion," *Computer*, vol. 14, Aug. 1981.
[22] A. M. Waxman and K. Wohn, "Contour evolution, neighborhood deformation, and global image flow: Planar surfaces in motion," *Int. J. Robotics Res.*, vol. 4, no. 3, pp. 95-108, Fall 1985.
[23] J. R. Westlake, *A Handbook of Numerical Matrix Inversion and Solution of Linear Equations*. New York: Wiley, 1968.

Joseph K. Kearney was born in Minneapolis, MN, in 1951. He received the B.A. degree in psychology from the University of Minnesota, Minneapolis, in 1975, the M.A. degree in psychology from the University of Texas, Austin, in 1979, and the M.S. and Ph.D. degrees in computer science from the University of Minnesota, Minneapolis, in 1981 and 1983, respectively.

He is an Assistant Professor of Computer Science at the University of Iowa, Iowa City. He is currently a Visiting Research Scientist in the Department of Computer Science at Cornell University, Ithaca, NY. His research interests include computer vision, robotics, and measurement theory.

Dr. Kearney is a member of the American Association for Artificial Intelligence, the Association for Computing Machinery, and the IEEE Computer Society.



William B. Thompson (S'72-M'75) was born in Santa Monica, CA, in 1948. He received the Sc.B. degree in physics from Brown University, Providence, RI, in 1970, and the M.S. and Ph.D. degrees in computer science from the University of Southern California, Los Angeles, in 1972 and 1975, respectively.

He is currently an Associate Professor in the Department of Computer Science of the University of Minnesota, Minneapolis. Previously, he was a member of the Image Processing Institute at the University of Southern California. His primary research interest is in the area of computer vision, with an emphasis on the development of models for the perception of spatial organization. As part of this work, he has been active in the development of techniques for analyzing visual motion. In addition, he is a principal in the expert problem solving research group at the University of Minnesota.

Dr. Thompson is a member of the American Association for Artificial Intelligence and the Association for Computing Machinery, and is an Associate Editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE.

Daniel L. Boley (M'84) received the B.A. degree in mathematics from Cornell University, Ithaca, NY, in 1974, and the M.S. and Ph.D. degrees in computer science from Stanford University, Stanford, CA, in 1976 and 1981, respectively.

He is currently an Assistant Professor of Computer Science at the University of Minnesota in Minneapolis. He has been a Visiting Fellow at the Center for Mathematical Analysis of the Australian National University, Canberra, Australia, and a Visiting Professor at the Department of Mathematics of the Politecnico di Milano, Milan, Italy. His research interests include the computation of matrix eigenvalue problems on both parallel MIMD and pipelined computer architectures, and numerical analysis in linear control theory and linear algebra.

Dr. Boley is a member of the Society for Industrial and Applied Mathematics and the Association for Computing Machinery.