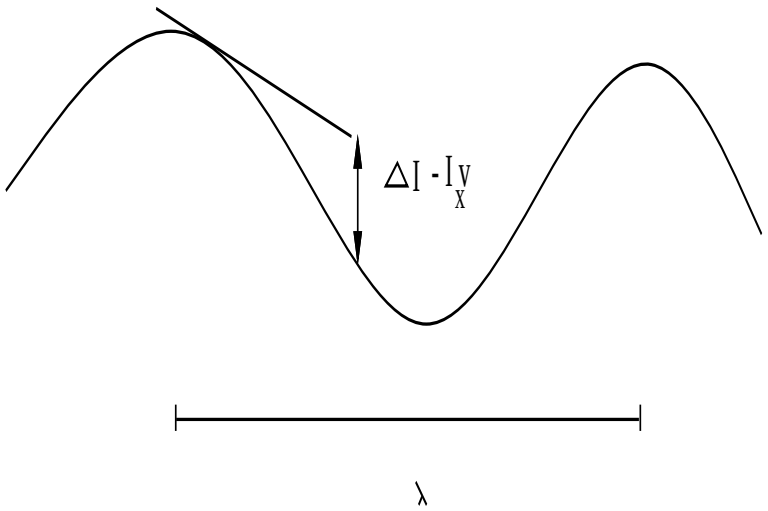

Figure 2: The linear slope approximation for a sinusoid is only valid for a fraction of a wavelength. The range of velocity estimates a gradient-based approach can reliably detect is limited by the spatial frequencies of the underlying intensity function.

to make (7) consistent is simply σ_3. Equation 8 is very similar to the standard least squares solution. The latter is obtained by setting σ_3 to zero.

Total least squares assumes the error in each element of the measurement matrix is independent and identically distributed (the error matrix is white). If this is not the case, total least squares can actually perform worse than standard least squares. This is similar to the requirement in standard least squares that the errors in the measurements be normalized. We can use prior estimates of the measurement variances to whiten the measurement matrix.

2.2 Systematic Errors due to large displacements

The systematic error term in the constancy equation (2) is often ignored. However this term can easily be larger than the stochastic error for relatively small displacements. The constraint equations makes a linear approximation to the underlying intensity function and is thus invalid for large displacements. The assumption breaks down quadratically in the displacements. If we operate in a single spatial dimension we can examine the relative magnitude of this term. If the signal is a simple sinusoid, it is obvious that the linear approximation is valid only for a fraction of the wavelength of the sinusoid. (Figure 2). If the wavelength,λ, of the stimulus is known, we can limit the acceptable displacement between time frames to some fraction of the wavelength.

$$|v| < \beta\lambda = 2\pi\beta/\omega \tag{9}$$

An natural limit is $\beta = 1/2$ since displacements greater than half a cycle would introduce aliasing. This limit on displacement as a function of stimulus frequency has a biological basis too. In random dot kinematograms it has been shown that the upper displacement, d_{max} for coherent motion detection falls off as the inverse of frequency [CB90, CJ85]. The exact value of β above is in doubt however. Multiple scales and orientations (as in our model) may confuse the issue. An interesting debate on the exact value of β can be seen in [BC90] and [LB90]. From a numerical point of view, Battiti et al. [BAK91] examined the systematic error implicit in gradient techniques which use finite differences to estimate partial derivatives. In the one dimensional case for a translating sinusoid of frequency ω, they

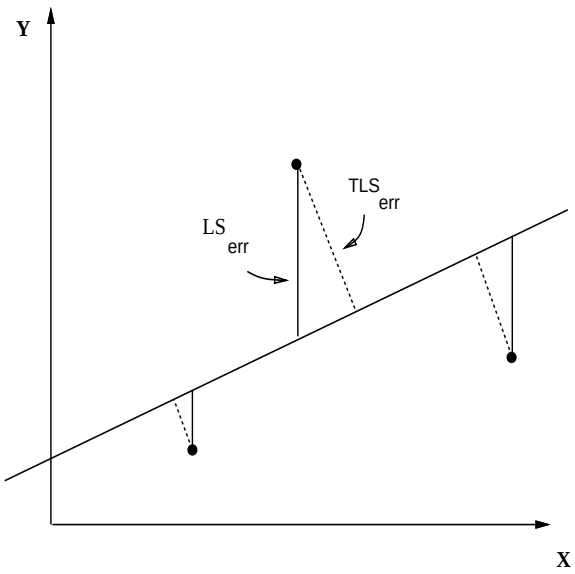

Figure 1: Difference between least squares and total least squares for fitting a line to a collection of points. Least squares assumes the errors are in the y variables and thus minimizes the vertical distance between the line and the points. Total least squares allows for errors in both x and y and thus minimizes the perpendicular distance between the line and the points.

The essential difference between classical least squares and total least squares can be made clear by a simple 1-D example. Suppose we wish to fit a line to a group of points, (x_i, y_i). In classical least squares we wish to find the values of the slope and intercept, (m, b) which minimizes the sum squared difference between the y_i and the predicted y.

$$\min_{b,m} \sum_i (y_i - mx_i - b)^2 \tag{6}$$

This minimizes the vertical distances between the line and the measurements y_i It assumes the variables x_i are error free and all noise is contained in the y_i. Total least squares allows for errors in the x_i variables too. It wishes to minimize the perpendicular distance between the line and the measured points (see Figure 1). This was referred to as *eigenvector fit* in [DH73]. The idea of allowing errors in all variables when fitting data has been around for some time [Pea01, Mad59]. The concept was extended to multivariate problems about 20 years ago [Spr69]. The connection to the singular value decomposition of the measurement matrix was pointed out by Golub and Van Loan [GV80] and Van Huffel and Vandewalle [VV91].

In the total least squares framework, (5) is usually written as

$$\left[\mathbf{A}|\vec{\mathbf{I}}_t\right] \begin{pmatrix} \vec{\mathbf{v}} \\ 1 \end{pmatrix} = \vec{\mathbf{0}} \tag{7}$$

The combined matrix $\left[\mathbf{A}|\vec{\mathbf{I}}_t\right]$ is referred to as the measurement matrix. This form recognizes that each entry in the measurement matrix is subject to noise. In total least squares, an estimate is found by making the smallest, in terms of its Frobenius norm, perturbation to the measurement matrix such that (7) has a solution [VV91]. This is in contrast to least squares where only the measurement vector $\vec{\mathbf{I}}_t$ is perturbed to find a solution. The estimate using total least squares is

$$\vec{\mathbf{v}} = -(\mathbf{A}^T\mathbf{A} - \sigma_3^2\mathbf{I})^{-1}\mathbf{A}^T\vec{\mathbf{I}}_t \tag{8}$$

where σ_3 is the smallest singular value of the measurement matrix. The Frobenius norm of the perturbation needed

v by assuming constancy of partial derivatives and other functions of the intensity. A third approach [LK81, CV90] is to assume the velocity field is locally constant and to combine constraint equations from neighboring pixels. A review of these and other approaches such as correlation and energy models can be found in [BFB92].

In our approach, we first convolve the image sequence with a set of linear spatio-temporal filters, $f_i(x, y, t)$. These are Gaussian derivatives of first or second order at a number of orientations and scales. These are the same filters used in some other approaches to early vision, such as in stereo and texture discrimination [JM92a, JM92b, MP90]. Each convolved image, $I_i = I * f_i$, has its own constraint equation of the form (2). This results in an overconstrained system of equations in the unknowns u and v.

$$\begin{pmatrix} I_{1x} & I_{1y} \\ I_{2x} & I_{2y} \\ . & . \\ I_{nx} & I_{ny} \end{pmatrix} \cdot \vec{\mathbf{v}} = \begin{pmatrix} -I_{1t} \\ -I_{2t} \\ . \\ -I_{nt} \end{pmatrix} + \vec{\mathcal{O}}(\vec{\mathbf{v}}^2) \tag{4}$$

The spatio-temporal partial derivatives, I_{ix}, I_{iy}, I_{it}, can be considered as the result of convolution of the image sequence I with linear, spatio-temporal filters since $I_{ix} = (I * f_i)_x = I * f_{ix}$. Defining the matrix $\mathbf{A}$ and vector $\vec{\mathbf{I}}_t$ the system of equations can be written as:

$$\mathbf{A} \cdot \vec{\mathbf{v}} = -\vec{\mathbf{I}}_t + \vec{\mathcal{O}}(\vec{\mathbf{v}}^2) \tag{5}$$

The spatial extent of the filters brings in information from neighboring pixels so the aperture problem exists only for degenerate cases.

The fundamental problem now is to solve the overconstrained system of equations (5) so as to obtain as accurate an estimate of $\vec{\mathbf{v}}$ as possible. We begin by analyzing the sources of error.

1. *Stochastic error.* In the presence of sensor noise, we expect that the measurements of I_{ix}, I_{iy}, I_{it}, the spatiotemporal derivatives of $I * f_i$, would be corrupted with noise. We will make the standard convenient assumption that sensor noise is independent from pixel to pixel and has a Gaussian distribution. This is analyzed further in subsection 2.1.

2. *Systematic error for large displacements.* The system of equations (5) is derived by neglecting second order terms, so we expect systematic errors whenever the local velocity is large. The magnitude of the error is dependent on a number of factors including the scale of the filter f_i being used and the local spatial frequencies present in the image neighborhood. This is analyzed further in subsection 2.2.

3. *Errors due to model failure.* In subsection 2.3 we group together the errors that arise due to violation of certain key assumptions of the differential approach: (a) Constancy of image brightness which is not strictly true whenever there is a significant specular component, and (b) that the optical flow field is locally constant over the support of the filters, which is not true if the filter support straddles a depth discontinuity or when there is a significant rotational or dilational component in the flow field.

2.1 Stochastic Error and Total Least Squares

If we knew that the errors were confined to the measurements of I_{ti}, i.e. the right hand side of the system (5), then the correct approach is well known from estimation theory. We find the classical weighted least squares solution which from the Gauss-Markov theorem is the best one can do. [1] The weight matrix can be determined by examining the covariance matrix of the filters f_{ti}.

However the classical least squares method makes the implicit assumption that the measurements on the left hand side I_{xi}, I_{yi} are error-free and that the errors are confined to the measurements on the right hand side I_{ti}. This assumption is not true, impelling us to use the *total least squares* method. Total least squares is also known as *orthogonal regression* or *errors-in-variables regression* [VV91].

[1] Of course, we all know that this is just a consequence of assuming that the sensor noise has a Gaussian distribution, an assumption rarely verified in practice. One appeals to the Central limit theorem and hopes for the best.

1 Introduction

A number of different approaches to recovering optical flow have been proposed: those based on correlation, energy and differential considerations. A recent survey is due to Barron et al. [BFB92, BFBB92] where the different approaches were compared on a series of synthetic and real images. They found that a phase-based approach by Fleet and Jepson [FJ90] performed the best numerically.

We have developed a new algorithm for computing optical flow in the differential framework which performs comparably to the Fleet and Jepson approach but with less computational cost and a higher density of estimates. We start with a multi-channel filtering of the intensity response, thus producing an overconstrained system of equations in the motion parameters. The convolution with a series of filters is a common starting point for a number of early vision tasks such as edge detection, stereo and texture discrimination [Can86, JM92a, JM92b, MP90]. In Section 2 we analyze the sources of error in the differential method as falling into 3 categories: (a) stochastic error due to sensor noise, (b) systematic error due to large displacements and (c) model error where the underlying model is violated. This analysis leads to an algorithm based on a robust version of total least squares. This algorithm is outlined in Section 3.

The implementation is described in Section 4. In Section 5 the algorithm is run on the series of synthetic and real sequences used by Barron et al. Thus a direct comparison between our work and others can be made. A high density of estimates was found for all sequences, implying that the "aperture problem" occurs rarely in most images. A confidence measure is available as a byproduct of the total least squares formulation. Through a simple experiment we demonstrate that this measure is related to the estimated accuracy of the motion vector. We also look at a scale pyramid implementation of the filter responses in this section to demonstrate that this more efficient method of computing multiple scale responses does not degrade the performance significantly.

Finally, in Section 6 we describe a parallel implementation on a multiple processor and examine the speedup of the algorithm. This demonstrates that a real-time parallel version of the algorithm may be possible.

2 The Differential Constraint Equation

A single constraint equation for the components of the velocity are derived from the constancy assumption. The constancy assumption assumes that some function of the brightness is constant between time frames.

$$I(x, y, t) = I(x + u\delta t, y + v\delta t, t + \delta t) \tag{1}$$

where $I(x, y, t)$ is the brightness or some function of the brightness at location (x, y) and time t. The vector field $\vec{\mathbf{v}} = (u, v)$ is the optical flow and is a function of image coordinates (x, y). We assume a differentiable brightness function and make a Taylor expansion of the right hand side of equation (1):

$$I_x u\delta t + I_y v\delta t + I_t \delta t = \mathcal{O}(\vec{\mathbf{v}}^2 \delta t^2) \tag{2}$$

where I_x,I_y and I_t are partial derivatives with respect to space and time, evaluated at the point (x, y, t). The right hand side represents the remaining terms of the Taylor expansion. This term contains products of higher spatial and temporal derivatives of the brightness function as well as higher powers of the displacements. By assuming a unit temporal delay, δt, this equation can be written in a more compact form:

$$\nabla I \cdot \vec{\mathbf{v}} + I_t = \mathcal{O}(\vec{\mathbf{v}}^2) \tag{3}$$

The right hand side is usually assumed small and set to zero. The validity of ignoring the right hand side of equation (2) is dependent on the spatial frequency content of the intensity pattern and the magnitude of the displacement.

This constraint equation has been used in motion detection for some time [FT79]. It consists of a single equation in the two unknowns which forms a single constraint line in velocity space. Any velocity on this line is valid for this single equation. This was called the "aperture problem" since for special cases the velocity could not be determined uniquely. Horn and Schunck [HS81] introduced a smoothness constraint in order to solve uniquely for displacement. A number of other authors [TP84, Nag87, UGVT88, VGT90, Sri90] produced two or more linear equations in u and

Robust Computation of Optical Flow in a Multi-Scale Differential Framework

Joseph Weber and Jitendra Malik

Department of Electrical Engineering and Computer Science
University of California at Berkeley
Berkeley, CA 94720
email: jweber@robotics.berkeley.edu, malik@robotics.berkeley.edu

November 9, 1992

Abstract

We have developed a new algorithm for computing optical flow in a differential framework. The image sequence is first convolved with a set of linear, separable spatiotemporal filters similar to those that have been used in other early vision problems such as texture and stereopsis. The brightness constancy constraint can then be applied to each of the resulting images, giving us, in general, an overdetermined system of equations for the optical flow at each pixel. There are three principal sources of error (a) stochastic due to sensor noise (b) systematic errors in the presence of large displacements and (c) errors due to failure of the brightness constancy model. Our analysis of these errors leads us to develop an algorithm based on a robust version of total least squares. Each optical flow vector computed has an associated reliability measure which can be used in subsequent processing. The performance of the algorithm on the data set used by Barron et al. (CVPR 1992) compares favorably with other techniques. In addition to being separable, the filters used are also causal, incorporating only past time frames. The algorithm is fully parallel and has been implemented on a multiple processor machine.

Keywords: optical flow, differential approach, brightness constancy assumption, total least squares, multi-channel filtering, robust statistics