`MIT Media Laboratory Vision and Modeling Technical Report #165.`

Computing Optical Flow Distributions Using Spatio-temporal Filters

Eero P. Simoncelli[†] and Edward H. Adelson[‡]

Vision and Modeling Group, Media Laboratory
[†]Department of Electrical Engineering and Computer Science
[‡]Department of Brain and Cognitive Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

email: eero@media-lab.media.mit.edu

November 15, 1990
Revised: March, 1991

Abstract

We describe the relationship between gradient methods for computing optical flow and filter-based spatio-temporal energy models of biological motion processing, revealing that these techniques are equivalent under certain conditions. We discuss extensions of these techniques which compute probability distributions of optical flow. The use of distributions allows proper handling of the uncertainties inherent in the optical flow computation, facilitating the combination with information from other sources. The probabilistic framework also leads to several useful extensions of the standard quadratic gradient solution. We use these extensions to compute optical flow for both a synthetic and a real image sequence.

This research was partially supported while the first author was a visiting scientist at the NASA-Ames Research Center.

1 Introduction

The recovery of motion information from visual input is an important task for both natural and artificial vision systems. Most models for the analysis of visual motion begin by extracting two-dimensional motion information. In particular, computer vision techniques typically compute two-dimensional optical flow vectors which describe the motion of each portion of the image in the image plane. Methods for the recovery of optical flow are often classified into two groups: those that match features between successive temporally discretized frames, and those that perform computations on the spatio-temporal gradient of image intensity.

In addition, several authors have proposed the use of spatio-temporal filters for analyzing motion [1, 2, 3]. Of interest in the present paper are some recent physiological models of motion processing based on quadratic combinations of the outputs of spatio-temporal filters [4, 5]. We refer to these as Spatio-Temporal Energy Models (STEM).

In this paper, we formalize the relationship between several of the standard least-squares gradient-based techniques and the spatio-temporal energy models. We describe the uncertainty inherent in the computation of optical flow through use of a simple Gaussian noise model, and we compute a maximum likelihood estimate solution. The resulting solution is an extension of the standard gradient solution. More generally, we suggest the usefulness of probability theory in the context of the problem of estimating optical flow. We test this model on both a synthetic and real image sequence, analyzing the errors for the synthetic sequence.

2 Gradient Methods

We write the image intensity signal as a function of position and time: $f(x, y, t)$. Then the standard gradient formulation of the optical flow problem is based on the assumption that the total derivative of the image intensity function must be zero at each position in the image and at every time:

$$\vec{f_s} \cdot \vec{v} + f_t = 0, \qquad (1)$$

where

$$\vec{f_s} = \begin{pmatrix} f_x \\ f_y \end{pmatrix},$$

and f_x, f_y, and f_t are the spatial and temporal derivatives of the image f, and $\vec{v}$ is the optical flow (at the position and time that the derivatives have been computed). We have left out the the spatial and temporal location parameters in order to simplify notation. This formulation assumes that changes in the image intensity are due only to translation of the local image intensity and not to changes in lighting, reflectance, etc. Furthermore, by formulating the constraint only in terms of first derivatives, we are implicitly approximating the image intensity as a planar function.

Typically, one would write a squared error function based on this total derivative constraint as follows:

$$E(\vec{v}) = \left[\vec{f_s} \cdot \vec{v} + f_t\right]^2. \tag{2}$$

To compute a Linear Least-Squares Estimate (LLSE) of $\vec{v}$ as a function of $\vec{f_s}$ and f_t, we set the gradient (with respect to $\vec{v}$) of this quadratic expression equal to the zero vector:

$$\nabla E(\vec{v}) = \mathbf{M} \cdot \vec{v} + \vec{b} = \vec{0}, \tag{3}$$

where

$$\mathbf{M} = \vec{f_s}\vec{f_s}^T = \begin{pmatrix} f_x^2 & f_x f_y \\ f_x f_y & f_y^2 \end{pmatrix}, \qquad \vec{b} = \begin{pmatrix} f_x f_t \\ f_y f_t \end{pmatrix}. \tag{4}$$

One immediate observation is that the matrix $\mathbf{M}$ is *always* singular (i.e., its determinant is always zero). Intuitively, this is to be expected since the solution is based on a planar approximation to the image surface at a point, and therefore suffers from the aperture problem. Equation (1) only places a constraint on the velocity vector in the direction of $\vec{f_s}$; that is on the component of flow *normal* to the spatial image orientation.

In addition to the planarity singularity, another possible source of singularity is due to phase-dependence. Consider an element of the matrix $\mathbf{M}$ evaluated at a point (without loss of generality we take $x = 0$). We can write this in terms of the Fourier transform of f as follows:

$$f_x^2(0) = \left| \int F(\vec{\omega}) \omega_x d\vec{\omega} \right|^2$$

It should be clear if the image consists of a drifting sine wave grating, the integration will evaluate to zero for certain phases of the sine wave. At these phases, we will not be able to compute a flow vector. This situation can be improved through use of quadrature pairs of filters for computing derivatives, but we will not discuss this issue here.

In order to eliminate the singularity problem, researchers have typically incorporated additional constraints in the error function. Horn and Schunk [6] applied a global smoothness constraint to the flow field in order to regularize the problem. One can also combine information locally using an "intersection of constraints" rule: Since we have a constraint on the normal component of velocity at each point, we can choose the velocity which is most consistent with all of the normal constraints in some small region. Implicitly, this is also a type of smoothness constraint, since we are assuming that the velocity vector is constant in the region. We do this by writing an error function based on the normal constraints from each point within a patch, indexed by a subscript $i \in \{1, 2, \ldots n\}$:

$$E(\vec{v}) \;=\; \sum_i \left[\vec{f}_s(x_i, y_i, t) \cdot \vec{v} + f_t(x_i, y_i, t)\right]^2 . \tag{5}$$

Computing the gradient of this expression with respect to $\vec{v}$ gives:

$$\nabla_v E(\vec{v}) = \sum_i \left(\mathbf{M}_i \cdot \vec{v} + \vec{b}_i\right) ,$$

with solution

$$\vec{v} = -\left(\sum_i \mathbf{M}_i\right)^{-1} \left(\sum_i \vec{b}_i\right) , \tag{6}$$

where we define $\mathbf{M}_i = \mathbf{M}(x_i, y_i, t)$ and $\vec{b}_i = \vec{b}(x_i, y_i, t)$ as in equation (4). In practice, we can also include a weighting function, w_i, in the summation in order to emphasize the information closer to the center of the averaging region. Thus, the velocity is computed from blurred quadratic functions of the spatial and temporal derivatives. We should note here that the matrix in the above equation may still be singular (despite the blurring). We will address this problem in section 4.

Two other observations should be made concerning the solution given above. The first is that it may be derived as a Taylor series approximation to the solution of a matching problem. We define an error function which is the mean squared error of the difference between two image patches at different times and positions:

$$\begin{aligned} E &= \sum_i \left[f(x_i + v_x, y_i + v_y, t+1) - f(x_i, y_i, t)\right]^2 \\ &\approx \sum_i \left[v_x f_x(x_i, y_i, t) + v_y f_y(x_i, y_i, t) + f_t(x_i, y_i, t)\right]^2 \\ &= \sum_i \left[\vec{f}_s(x_i, y_i, t) \cdot \vec{v} + f_t(x_i, y_i, t)\right]^2 , \end{aligned}$$

where we have expanded $f(x_i + v_x, y_i + v_y, t+1)$ as a first order Taylor series. This is the approach taken by Lucas and Kanade [7] in the context of stereo vision. The resulting error function is identical to that of equation (2).

It is also useful to write the solution in the frequency domain. Watson and Ahumada [8] and others have noted that the Fourier transform spectrum of an image undergoing rigid translation lies in a plane in the spatio-temporal frequency domain. Consider the energy function given in equation (2) with the summation occurring over the entire image:

$$\begin{aligned} E &= \sum |v_x f_x + v_y f_y + f_t|^2 \\ &= \sum |v_x F(\vec{\omega})\omega_x + v_y F(\vec{\omega})\omega_y + F(\vec{\omega})\omega_t|^2 \\ &= \sum [(v_x, v_y, 1)\cdot\vec{\omega}]^2 \cdot |F(\vec{\omega})|^2 \end{aligned}$$

where the sum on the first line is over all image pixels and the sums on the latter two lines are over all frequencies, $\vec{\omega}$. The solution of this equation is a planar regression solution, weighted by the image spectrum, $|F(\vec{\omega})|^2$. Thus, if one performs the blurring in the Lucas and Kanade solution over the entire image, this is equivalent to performing a planar regression weighted by the Fourier spectrum.

3 Spatio-Temporal Energy Models (STEM)

Several authors have developed physiologically-based models for the extraction and analysis of image motion. One popular set of models is based on spatio-temporal energy mechanisms [4, 5, 9]. In these models, oriented spatio-temporal subband filters are applied to the image and their squared outputs are combined to produce local measures of motion energy. Outputs of such mechanisms have been used by Heeger [5] and by Grzywacz and Yuille [9] to compute optical flow.

We show here that the basic gradient solution discussed in the previous section may be computed from the outputs of a particular STEM. The first observation is that the derivative operation on a sampled image may be written as a convolution. Since derivatives are only defined on continuous functions, the computation on a discrete function requires an intermediate interpolation step with a continuous function $C(\vec{r})$. The derivative of the interpolated function must then be re-sampled at the points of the original sampling lattice. The sequence of operations may be written for a one-dimensional signal as follows:

$$\begin{aligned}\frac{d}{dx}f(n) &\equiv \left[\frac{d}{dx}\left(\sum_m f(m)C(x-m)\right)\right]_n \\ &= \left[\sum_m f(m)\frac{dC}{dx}(n-m)\right],\end{aligned}$$

where we assume unit sample spacing in the discrete variables n and m for simplicity.

Thus, the derivative operation is defined as convolution with a filter which is the sampled derivative of some continuous interpolation function $C(\vec{r})$. One could use an "ideal" lowpass (sinc) function, or a gentler function such as a gaussian. Furthermore, one could imagine prefiltering the image to extract some spatio-temporal subband and performing the derivatives on this subband. The prefiltering operation can be combined associatively with the derivative operation in a way which is equivalent to convolving with the derivative of the prefiltering function. The prefilter should be symmetric or anti-symmetric so as to avoid altering the optical flow of the sequence.

Given the three directional derivatives along the x-, y-, and t- axes (as computed above), we can also compute derivatives in other orientations by taking linear combinations of the axis-oriented derivatives. For example, the spatial derivative of f at an angle of $\pi/4$ is computed as

$$f_{\vec{p}} = \frac{df}{d\vec{p}} = \frac{1}{\sqrt{2}}(f_x + f_y),$$

where

$$\vec{p} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Using this relationship, we can write the entries of $\mathbf{M}$ and $\vec{b}$ as defined in equation (4) in terms of squared directional derivatives:

$$\mathbf{M} = \begin{pmatrix} f_x^2 & \frac{1}{2}(f_{\vec{p}}^2 - f_{\vec{q}}^2) \\ \frac{1}{2}(f_{\vec{p}}^2 - f_{\vec{q}}^2) & f_y^2 \end{pmatrix}, \qquad \vec{b} = \begin{pmatrix} \frac{1}{2}(f_{\vec{l}}^2 - f_{\vec{r}}^2) \\ \frac{1}{2}(f_{\vec{d}}^2 - f_{\vec{u}}^2) \end{pmatrix}.$$

where we define

$$\begin{aligned} f_{\vec{q}} &= \tfrac{1}{\sqrt{2}}(f_x - f_y), \qquad f_{\vec{l}} = \tfrac{1}{\sqrt{2}}(f_x + f_t), \\ f_{\vec{r}} &= \tfrac{1}{\sqrt{2}}(f_x - f_t), \qquad f_{\vec{d}} = \tfrac{1}{\sqrt{2}}(f_y + f_t), \\ f_{\vec{u}} &= \tfrac{1}{\sqrt{2}}(f_y - f_t). \end{aligned}$$

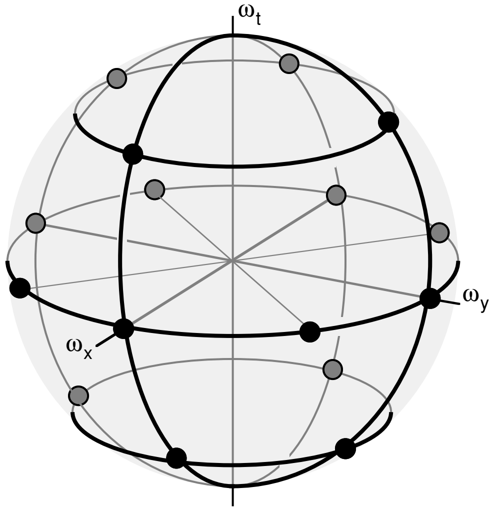


Figure 1: Illustration of the locations of the eight derivative filters in the spatio-temporal frequency domain. Each pair of dots on opposite sides of the sphere correspond to the central tuning frequency of one of the filters.

Thus, if the filters used to compute spatio-temporal energies are related as specified in the equation above, (i.e. they are derivatives of some common function C), then they can be combined in "opponent" fashion to compute optical flow using the standard gradient algorithms described in the previous section. The concept of motion opponency has been used by several authors to explain biological processing of motion information [10, 11, 4]. The computation described above requires eight different filters: four static (purely spatial) filters, and four motion (spatio-temporal) filters. These are illustrated in a diagram in figure 1.

4 Distributed Representations of Optical Flow

In this section, we discuss optical flow extraction as an estimation problem. In particular, we will work through flow computations resulting from several different simple noise models. There are many advantages to viewing the problem probabilistically. Optical flow fields are inherently uncertain: Errors occur because of image noise, lighting changes, low contrast regions, the aperture problem, and multiple motions in a single localized region. A probabilistic framework allows these uncertainties to be represented in the computations, and passed along to the next stage of computation. We should emphasize here that we do not just wish to provide a scalar "confidence" measure, but a two-dimensional probability distribution, capable of representing inhomogeneous *directional* uncertainties. Anandan [12] computed two-dimensional confidences based on a block-matching histogram, and Heeger [5] estimated two-dimensional covariances. Szeliski [13] has discussed the use of Bayesian techniques for a variety of problems in low-level vision.

Representing optical flow in terms of probability distributions also allows information from other sources, such as inertial sensors, to be combined properly with visual motion information, with the most certain information being weighted more heavily. Probability theory provides a tool for formalizing the problem of optical flow computation and characterizing various solutions according to their error distributions. Finally, models in terms of probability distributions may often serve as physiological motion-processing models, with distributions being described by the response of populations of cells [14].

Single Noise Source Model

The goal, then, is to compute an expression for the probability of the image velocity conditional on the image sequence. For the purposes of this paper, we will more specifically be concerned with a conditional probability based on the image gradient, ∇f:

$$\mathbf{P}(\vec{v} \mid \nabla f).$$

Consider the total derivative constraint in equation (1). In practice, there will be errors in the derivative computations due to camera and quantization noise, aliasing, imprecision in the derivative filters, etc. As mentioned earlier, even if the derivative measurements are error-free, the constraint in equation (1) may fail to be satisfied because of changes in lighting or reflectance, or the presence of multiple motions. We would like to account for both of these types of error. As is common in estimation theory, we decribe each of these types of uncertainty using additive gaussian noise. We begin by lumping together all of the uncertainty and describing it with a single additive gaussian noise variable n:

$$\vec{f_s} \cdot \vec{v} + f_t = n, \qquad n \sim N(0, \sigma_n).$$

This allows us to write an expression for the conditional probability:

$$\mathbf{P}(f_t \mid \vec{v}, \vec{f_s}) = \exp\left[-(\vec{f_s}^T \cdot \vec{v} + f_t)^T \sigma_n^{-1} (\vec{f_s}^T \cdot \vec{v} + f_t)\right].$$

In order to write down the desired conditional probability, we can use Bayes' rule to switch the order of the arguments:

$$\mathbf{P}(\vec{v} \mid \vec{f_s}, f_t) = \frac{\mathbf{P}(f_t \mid \vec{v}, \vec{f_s}) \cdot \mathbf{P}(\vec{v})}{\mathbf{P}(f_t)}.$$

For the prior distribution $\mathbf{P}(\vec{v})$, we can choose a zero-mean gaussian with variance Λ_p. Ignoring for the moment the normalization constant in the denominator, we can write the desired distribution:

$$\mathbf{P}(\vec{v} \mid \vec{f_s}, f_t) \quad \propto \quad \exp\left[-(\vec{f_s}^T \cdot \vec{v} + f_t)^T \sigma_n^{-1} (\vec{f_s}^T \cdot \vec{v} + f_t) - \vec{v}^T \Lambda_p^{-1} \vec{v}\right].$$

The resulting distribution contains two parameters (σ_n and Λ_p) which describe the variance of expected error in the constraint equation, and of the prior distribution on the optical flow vector $\vec{v}$, respectively. The distribution is gaussian, and the variance and mean may be derived using standard techniques:

$$\begin{aligned} \Lambda_{\vec{v}} &= \left[\vec{f_s} \sigma_n^{-1} \vec{f_s}^T + \Lambda_p^{-1}\right]^{-1} \\ &= \left[\frac{1}{\sigma_n}\mathbf{M} + \Lambda_p^{-1}\right]^{-1} \qquad (7) \\ \mu_{\vec{v}} &= -\Lambda_{\vec{v}} \vec{f_s} \sigma_n^{-1} f_t \\ &= -\frac{1}{\sigma_n} \Lambda_{\vec{v}} \vec{b} \qquad (8) \end{aligned}$$

where $\mathbf{M}$ and $\vec{b}$ are defined in equation (4).

If we wish to extract a flow vector from this probability distribution, the simplest choice is the Maximum Likelyhood Estimate (MLE). Since the distribution is gaussian, this will just be the mean. Notice that if σ_n is the identity, this is similar to the gradient solution mentioned in a previous section, but with the addition of the Λ_p^{-1} term which ensures the invertibility of the variance matrix. This is not really surprising, since the MLE of a gaussian distribution is equivalent to computing the LLSE.

If the solution is computed independently at each point (as written above), the mean will be (approximately) the *normal* flow vector, and the width (variance) of these distributions in the direction perpendicular to the normal direction will be determined by Λ_p. The variance in the normal direction will be determined by both Λ_p and the trace of $\mathbf{M}$ (i.e. the sum of the squared magnitudes of the spatial derivatives).

If the next stage of processing cannot make use of normal flow (along with variance information) as input, then we can combine information in small neighborhoods to compute actual flow. As shown in the previous section, this is done by combining constraints in the manner described by equation (5). If this is done in equation (8), and if we assume that the noise at each point in the neighborhood is independent, then the resulting mean and variance are:

$$\begin{aligned} \Lambda_{\vec{v}} &= \left[\sum_i \frac{w_i}{\sigma_n}\mathbf{M}_i + \Lambda_p^{-1}\right]^{-1} \\ \mu_{\vec{v}} &= -\Lambda_{\vec{v}}\left[\sum_i \frac{w_i}{\sigma_n}\vec{b}_i\right] \end{aligned} \tag{9}$$

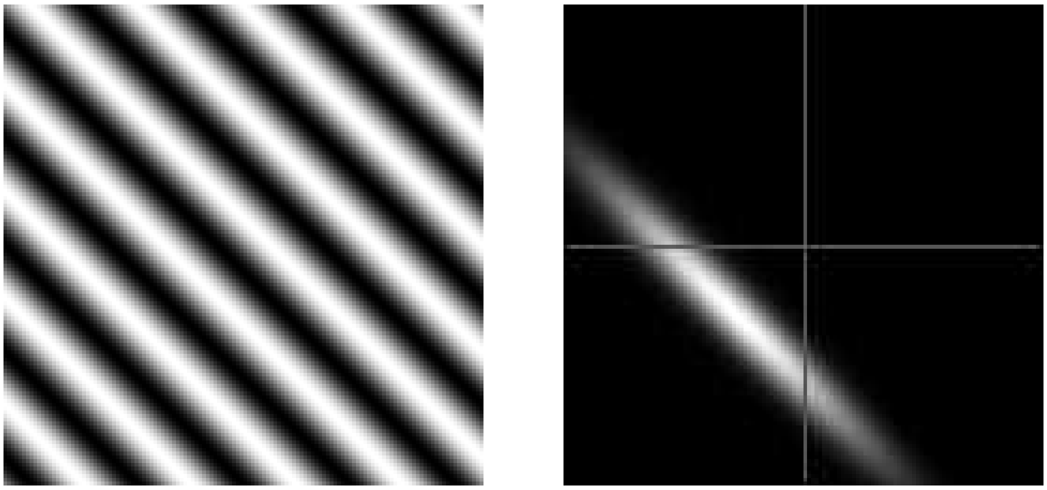
We have included in the expression a blur function w_i which weights the information in the center of the neighborhood more heavily.

Again, the MLE solution is similar to the Lucas and Kanade solution from the first section, except for the prior variance term which ensures matrix invertibility. Note that the inverse variance (certainty) increases roughly quadratically with the magnitude of the derivative measurements. To illustrate the behavior of the system, we consider the response to a drifting sinusoidal grating. One frame of the input image is shown in figure 2(a). The response of the system, at a pixel near the center of the image is shown in figure 2(b). Note that the ambiguity of the motion, which arises because the signal is really a *one-dimensional* signal, is captured by the elongated shape of the distribution.

An adaptive version of this algorithm could proceed by blurring over larger and larger regions until the magnitude of the variance (determinant of the variance matrix) was below some threshold. Since the variance matrix $\Lambda_{\vec{v}}$ describes a two-dimensional shape, this could even be done directionally, averaging pixels which lie in the direction of maximal variance until the variance in this direction was below a threshold.

Two Noise Source Model

A slightly more complicated noise model yields a potentially more useful solution. As mentioned earlier, even if the derivative measurements are error-free, the constraint in equation (1) may fail to be satisfied because of changes in lighting or reflectance, or the presence of multiple



(a) (b)

Figure 2: (a) One frame of a drifting sin grating sequence. The normal direction of the motion was down and to the left. The drift speed (in the normal direction) was 0.83 pixel/frame. (b) The response of the system computed over a patch near the center of the image. The peak of the response coincides with the actual normal velocity.

motions. To try account for these errors, we can include another, independent, gaussian noise variable which is subtracted directly from the flow vector:

$$\vec{f}_s \cdot (\vec{v} - \vec{n}_1) + f_t = n_2, \qquad n_i \sim N(0, \Lambda_i).$$

or

$$\vec{f}_s \cdot \vec{v} + f_t = \vec{f}_s \cdot \vec{n}_1 + n_2. \tag{10}$$

The noise term n_2 is the same as the noise in the previous case and describes the errors in the derivative measurements. These could be due to noise or aliasing in the original image sequence, imprecision in the filters, etc. The new noise term ($\vec{n}_1$), describes errors resulting from a failure of the planarity assumptions.

With both noise terms included, the resulting mean and variance of the optical flow distribution are

$$\begin{aligned} \Lambda_{\vec{v}} &= \left[\vec{f}_s(\vec{f}_s^T \Lambda_1 \vec{f}_s + \Lambda_2)^{-1} \vec{f}_s^T + \Lambda_p^{-1}\right]^{-1} \\ \mu_{\vec{v}} &= -\Lambda_{\vec{v}} \vec{f}_s (\vec{f}_s^T \Lambda_1 \vec{f}_s + \Lambda_2)^{-1} f_t \end{aligned}$$

If we choose Λ_1 to be a diagonal matrix, with diagonal entry σ_1, and write the scalar variance of n_2 as $\sigma_2 \equiv \Lambda_2$, then we can write this as:

$$\begin{aligned} \Lambda_{\vec{v}} &= \left[\vec{f}_s \left(\sigma_1 ||\vec{f}_s||^2 + \sigma_2 \right)^{-1} \vec{f}_s^T + \Lambda_p^{-1} \right]^{-1} \\ &= \left[\frac{\mathbf{M}}{\left(\sigma_1 ||\vec{f}_s||^2 + \sigma_2 \right)} + \Lambda_p^{-1} \right]^{-1} \end{aligned} \tag{11}$$

$$\begin{aligned} \mu_{\vec{v}} &= -\Lambda_{\vec{v}} \vec{f}_s \left(\sigma_1 ||\vec{f}_s||^2 + \sigma_2 \right)^{-1} f_t \\ &= -\Lambda_{\vec{v}} \cdot \frac{\vec{b}}{\left(\sigma_1 ||\vec{f}_s||^2 + \sigma_2 \right)} \end{aligned}$$

These results are similar to those in equation (8), except that the quadratic derivative terms in $\mathbf{M}$ and $\vec{b}$ are modified by a compressive nonlinearity. That is, for regions with low contrast (i.e. small $||\vec{f}_s||^2$), the σ_2 term dominates the divisor of $\mathbf{M}$. For large contrast regions, the σ_1 term tends to normalize the magnitude of the quadratic terms in $\mathbf{M}$. This seems intuitively reasonable: When the contrast (SNR) of the signal is low, an increase in contrast should increase one's certainty of the velocity estimate. But as the contrast increases above the noise level of the signal, the certainty should asymptotically reach some maximum value rather than continuing to rise quadratically. This matches the description of the noise terms given earlier: n_2 accounts for errors in the derivative measurements, and $\vec{n}_1$ accounts for failures of the constraint equation. The nonlinearity is illustrated in figure 3, where we have plotted the trace of the inverse covariance $\Lambda_{\vec{v}}^{-1}$ as a function of $||\vec{f}_s||^2$.

When we combine information over a small neighborhood using a weighting function w_i, the solution becomes:

$$\Lambda_{\vec{v}} = \left[\sum_i \frac{w_i \mathbf{M}_i}{\left(\sigma_1 ||\vec{f}_s(x_i, y_i, t)||^2 + \sigma_2 \right)} + \Lambda_p^{-1} \right]^{-1}$$

$$\mu_{\vec{v}} = -\Lambda_{\vec{v}} \cdot \sum_i \frac{w_i \vec{b}_i}{\left(\sigma_1 ||\vec{f}_s(x_i, y_i, t)||^2 + \sigma_2 \right)} \tag{12}$$

Here, the effect of the nonlinearity on the combination of information over the patch is to provide a sort of gain control mechanism. The previous solution in equation (9) weighted the

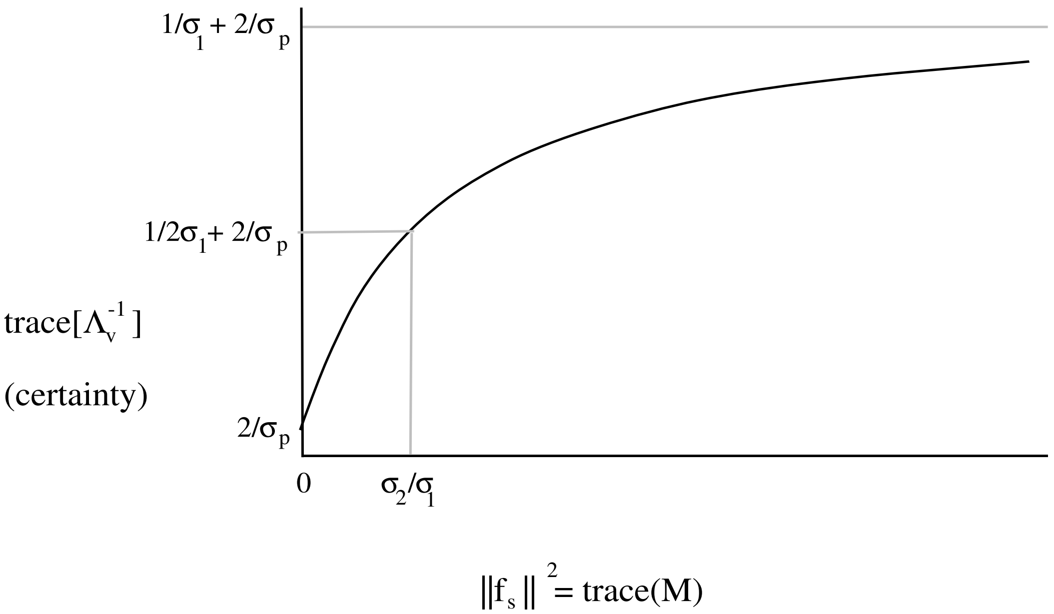


Figure 3: Plot of the nonlinearity which operates on the quadratic energy measurements in the solution given in equation (11).

information in a patch according to the derivative energies. If we ignore σ_2, the solution above normalizes the information, equalizing the contribution from each point in the neighborhood by the magnitude of the spatial gradient. Heeger [15] has found that the use of gain control on the energy images often improves the performance of the flow computation.

The solutions outlined above make use of the simple relationship between LLSE solutions and MLE solutions for gaussian random variables, and therefore are limited to being unimodal. In order to analyze multiple motions or transparency properly, one needs to incorporate more complicated distributions. Alternatively, one could combine flow information derived from different scales (spatial subbands) in a way that resulted in multi-modal distributions.

5 Examples

We computed optical flow on both synthetic and real image sequences using the technique defined by equation (12) combined with a multi-scale pyramid decomposition. The multi-scale approach is necessary since the gradient technique will fail if there is too much spatio-temporal aliasing (i.e., if the displacements being measured are greater than one half of a cycle of the highest spatial frequencies present in the pre-filtered image sequence). Similar multi-scale "warping" approaches have been used by Quam [16] and Anandan [12].

We first built a (spatial) "gaussian pyramid" [17] on each frame of the image sequence: Each frame was recursively blurred using a simple gaussian filter and subsampled by a factor of two in each spatial direction. This resulted in a set of images (a "pyramid") of different spatial resolution. We then computed the optical flow on the sequence of top level (lowest frequency) images using the computation specified by equation (12).

An upsampled and interpolated version of this coarse, low-resolution flow field was used to warp the sequence of images in the next pyramid level. The warping operation is defined as

$$f_{\text{warped}}(x, y) = f_{\text{original}}\left(x - v_x(x, y), y - v_y(x, y)\right),$$

where we used bi-cubic spline interpolation to evaluate f_{original} at fractional-pixel locations. Equation (12) was used to estimate the optical flow of the warped sequence, and this "optical flow correction" was composed with the previously computed optical flow to give a new optical flow estimate. This correction process was repeated for each level of the pyramid until the flow fields were at the resolution of the original image sequence.

In implementing equation (12), we used a set of sampled analytic derivatives of gaussians as derivative filters. The kernels had spatio-temporal dimensions 7x7x6. The noise parameters used were chosen empirically as follows: $\sigma_1 = 0.08, \sigma_2 = 1.0, \sigma_p = 2.0$. The solution seemed relatively insensitive to variations in these parameters. The averaging step was performed by separably applying a one-dimensional (spatial) weighting kernel: $w_i = (0.0625, 0.25, 0.375, 0.25, 0.0625)$.

We computed flow on a synthetic (texture-mapped) fly-through sequence of the Yosemite valley. Figure 4 shows a frame of the original image sequence. Figure 5 shows the corresponding frame of the actual flow field (computed using the three-dimensional motion parameters and the depth map). Figure 6 shows the recovered flow field.

To analyze the appropriateness of the noise model, we need to check that the covariance information adequately describes the errors in the flow estimates. Since the covariance information is difficult to display or analyze, we computed a "deviation" value $D(x, y, t)$ at each

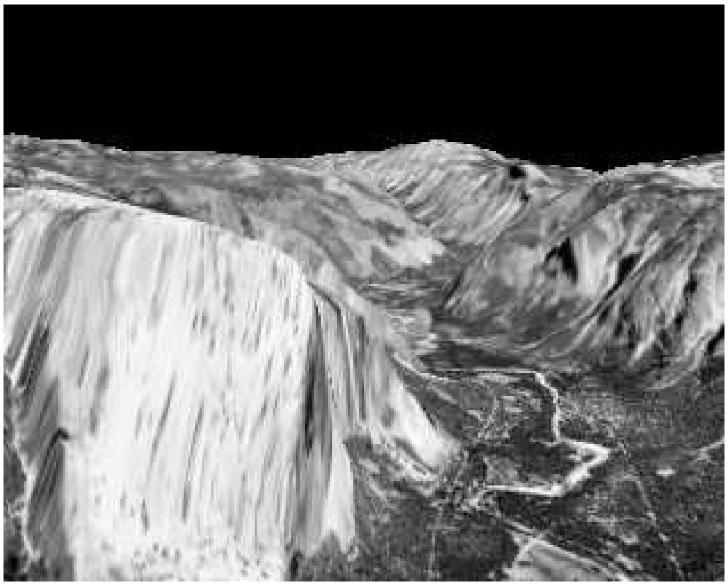


Figure 4: A frame from the original "Yosemite" fly-through sequence. The sequence was synthetically generated by Lyn Quam at SRI.

point:

$$D = \pm\sqrt{(\vec{v}_{\text{actual}} - \vec{v}_{\text{est}})^T \, \Lambda_{\vec{v}}^{-1} \, (\vec{v}_{\text{actual}} - \vec{v}_{\text{est}})},$$

where for simplicity we have omitted the positional arguments, (x, y, t), which parameterize each of the quantities in the equation. The normalized histogram of the values of D, is shown in figure 7. If the flow field errors were exactly modeled by the simple additive gaussian noise terms of equation (10), then this histogram would be in the shape of the function obtained by integrating a two-dimensional univariate gaussian over its angular coordinate:

$$h(r) \propto |r| \cdot e^{-r^2/2}.$$

For comparison, this function is plotted in figure 8. The error histogram is seen to qualitatively match, suggesting that the noise model is not unreasonable.

We also computed optical flow for a real sequence which was filmed from a helicopter flying above a valley. One frame from the original is shown in figure 9 and the corresponding frame of the recovered flow field is shown in figure 10.

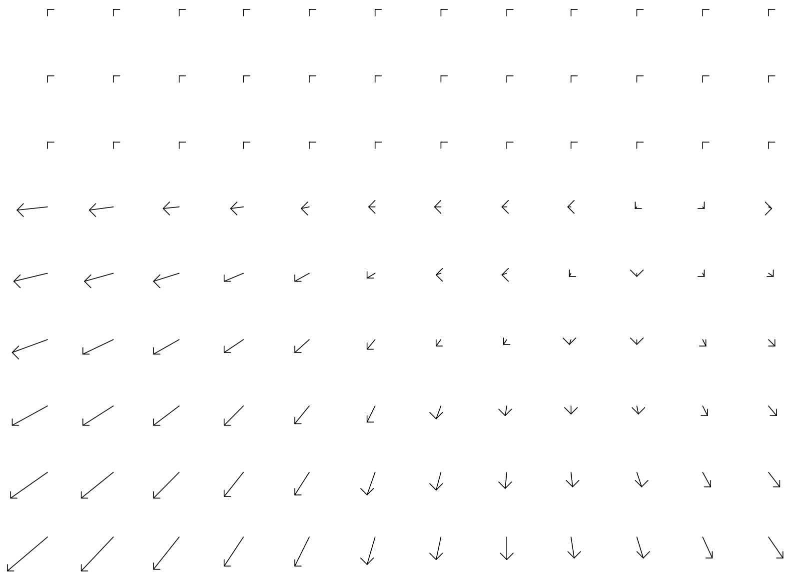


Figure 5: The actual optical flow field corresponding to the frame from the "Yosemite" sequence shown in the previous figure.

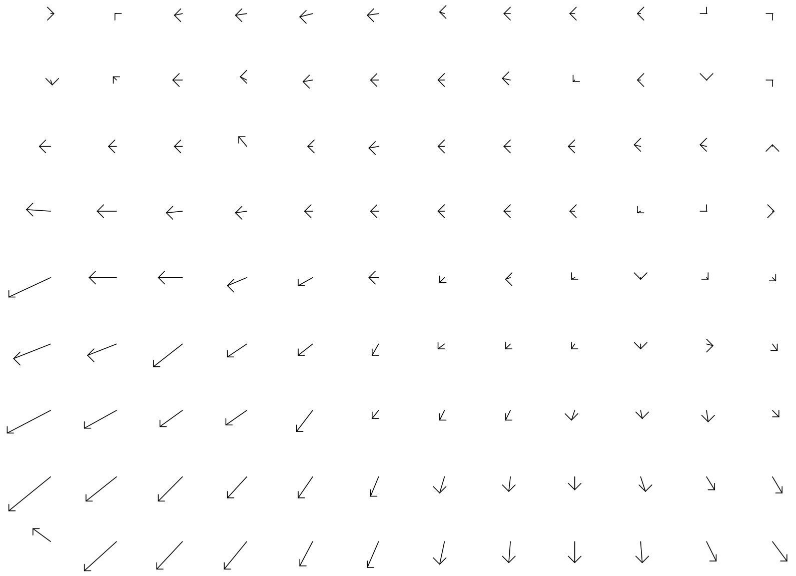


Figure 6: The recovered optical flow corresponding to the frame from the "Yosemite" sequence shown in the previous figure.

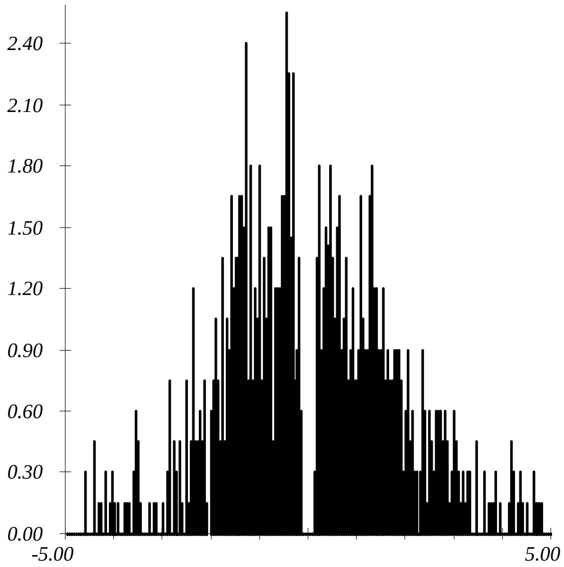


Figure 7: Histogram of the deviations D of the optical flow vectors (see text).

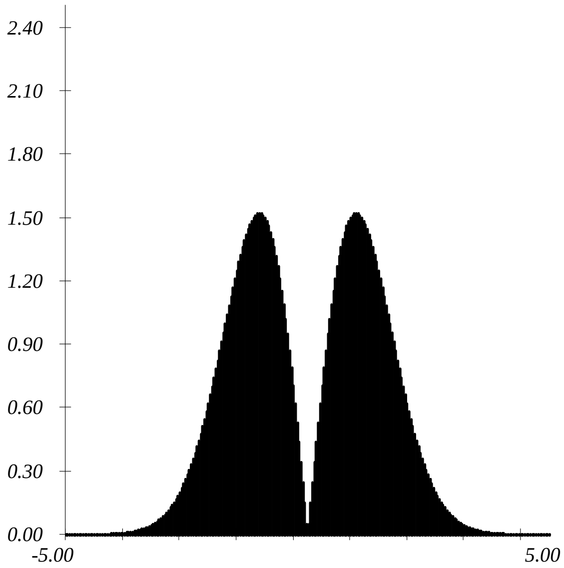


Figure 8: Ideal distribution of deviations D for the gaussian noise model (see text).

Figure 9: A frame from the original "Nap-of-the-earth" (NOE) fly-through sequence. The sequence was was proveded by the NASA Ames Research Center.