Presented at: IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii. June, 1991.

Probability Distributions of Optical Flow

Eero P. Simoncelli *† Edward H. Adelson *‡ David J. Heeger §

Abstract

Gradient methods are widely used in the computation of optical flow. We discuss extensions of these methods which compute probability distributions of optical flow. The use of distributions allows representation of the uncertainties inherent in the optical flow computation, facilitating the combination with information from other sources. We compute distributed optical flow for a synthetic image sequence and demonstrate that the probabilistic model accounts for the errors in the flow estimates. We also compute distributed optical flow for a real image sequence.

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 this paper, we describe a probabilistic formulation of the gradient approach to the optical flow problem. 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. 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:

$$\vec{f}_s \cdot \vec{v} + f_t = 0, \tag{1}$$

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 order to eliminate the singularity problem, researchers have typically incorporated additional constraints in the error function. Horn and Schunk [1] 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.

*Vision and Modeling Group, MIT Media Laboratory, 20 Ames Street, Cambridge, MA 02139.

†Department of Electrical Engineering and Computer Science. E-mail: eero@media-lab.media.mit.edu

‡Department of Brain and Cognitive Science. E-mail: adelson@media-lab.media.mit.edu

§NASA-Ames Research Center, MS262-2, Moffett Field, CA 94035. E-mail: heeger@bessel.arc.nasa.gov

We do this by writing an error function based on the normal constraints from each point within a patch, indexed by a subscript $i \in \{1, 2, \ldots n\}$:

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

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

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

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 3.

The solution given in equation (6) may also 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:

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

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 [2] in the context of stereo vision. The resulting error function is identical to that of equation (2).

3 Distributed Representations of Optical Flow

In this section, we discuss optical flow extraction as an estimation problem. There are many advantages to viewing the problem probabilistically. Optical flow fields are inherently uncertain 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 [3] computed two-dimensional confidences based on a block-matching histogram, and Heeger [4] estimated two-dimensional covariances. Szeliski [5] has discussed the use of Bayesian techniques for a variety of problems in low-level vision.

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 will describe each of these types of uncertainty with independent additive gaussian noise terms variable n_1 and n_2:

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

or

$$\vec{f}_s \cdot \vec{v} + f_t = \vec{f}_s \cdot \vec{n}_1 + n_2. \qquad (7)$$

The noise term n_2 describes the errors in the temporal derivative measurements. The other noise term ($\vec{n}_1$), describes errors resulting from a failure of the planarity assumptions.

Equation (7) describes the conditional probability, $\mathbf{P}(f_t \mid \vec{v}, \vec{f}_s)$. 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 choose a zero-mean gaussian with covariance Λ_p. The resulting distribution is gaussian, and the covariance and mean may be derived using standard techniques (i.e., completing the square in the exponent):

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

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:

$$\Lambda_{\vec{v}} = \left[\frac{\mathbf{M}}{\left(\sigma_1 \|\vec{f}_s\|^2 + \sigma_2\right)} + \Lambda_p^{-1}\right]^{-1} \qquad (8)$$

$$\mu_{\vec{v}} = -\Lambda_{\vec{v}} \cdot \frac{\vec{b}}{\left(\sigma_1 \|\vec{f}_s\|^2 + \sigma_2\right)}.$$

The Maximum A Posteriori (MAP) estimate is simply the mean, $\mu_{\vec{v}}$, since the distribution is gaussian. This solution is similar to that specified by equation (3). This is not really surprising, since computing the maximum of a gaussian distribution is equivalent to computing the LLSE. The differences are that 1) the prior variance Λ_p ensures the invertibility of the matrix, and 2) 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

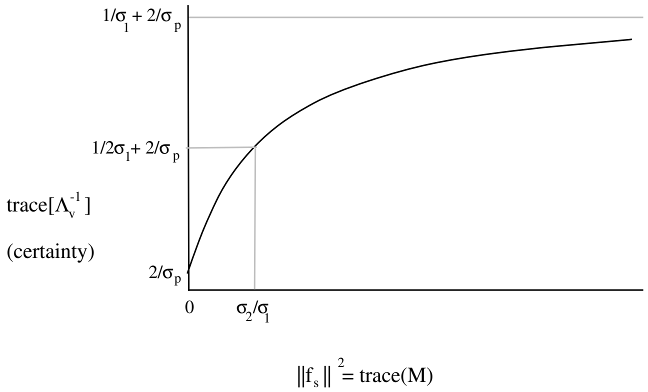

Figure 1: Plot of the nonlinearity which operates on the quadratic energy measurements in the solution given in equation (8).

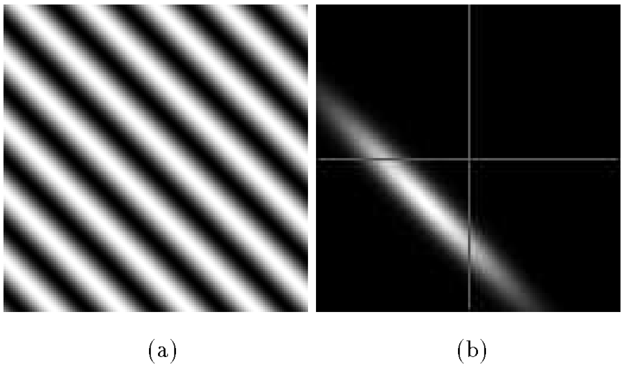

Figure 2: (a) One frame of a drifting sin grating sequence. The normal direction of the motion was down and to the left. The drift speed (in the normal direction) was 0.83 pixel/frame. (b) The response of the system computed over a patch near the center of the image. The peak of the response coincides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

$\|\vec{f}_s\|^2$), the σ_2 term dominates the divisor of $\mathbf{M}$. For large contrast regions, the $\sigma_1\|\vec{f}_s\|^2$ 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 our 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. The noise term n_2 accounts for errors in the derivative measurements. At low contrasts, these will be the dominant source of error. The term $\vec{n}_1$ accounts for failures of the constraint equation. At high contrasts, these will be the dominant source of error. The nonlinearity is illustrated in figure 1, where we have plotted the trace of the inverse covariance $\Lambda_{\vec{v}}^{-1}$ (i.e., the certainty of the estimate) as a function of contrast, $\|\vec{f}_s\|^2$.

If the solution is computed independently at each point (as written above), the mean will be (approximately) the *normal* flow vector, and the width 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 normal flow (along with variance information) is insufficient input for the next stage of processing, then we can combine information in small neighborhoods (analogous to equation (5)). If we assume that the noise at each point in the neighborhood is independent, then the resulting mean and variance are:

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

$$\mu_{\vec{v}} = -\Lambda_{\vec{v}} \cdot \sum_i \frac{w_i \vec{b}_i}{\left(\sigma_1\|\vec{f}_s(x_i,y_i,t)\|^2+\sigma_2\right)}. \qquad (9)$$

Here, the effect of the nonlinearity on the combination of information over the patch is to provide a sort of gain control mechanism. 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 [6] has found that the use of gain control in spatio-temporal energy models often improves the performance of the flow computation.

To illustrate the solution given in equation (9), we consider the response to a drifting sinusoidal grating. One frame of the input image is shown in figure 2(a). The resulting distribution, computed for a patch 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.

4 Examples

We computed optical flow on both synthetic and real image sequences using the technique defined by equation (9) 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 [7] and Anandan [3].

We first built a (spatial) "gaussian pyramid" [8] 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 (9).

An upsampled and interpolated version of this coarse,

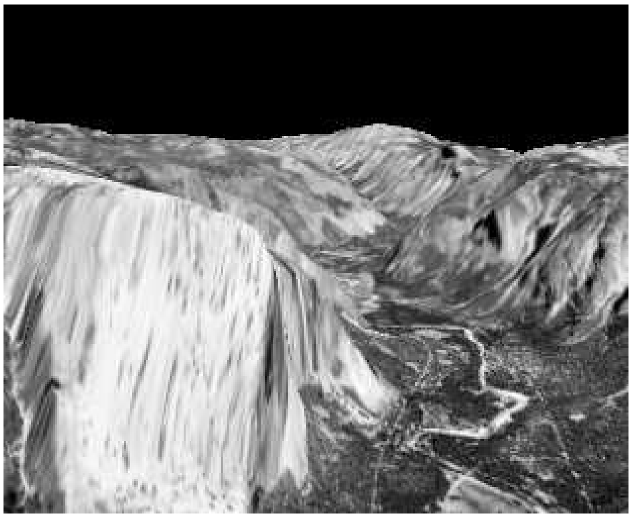

Figure 3: A frame from the original "Yosemite" fly-through sequence. The sequence was synthetically generated by Lyn Quam at SRI.

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 (9) 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 (9), 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 3 shows a frame of the original image sequence. Figure 4 shows the corresponding frame of the actual flow field (computed using the three-dimensional motion parameters and the depth map). Figure 5 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 at each 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

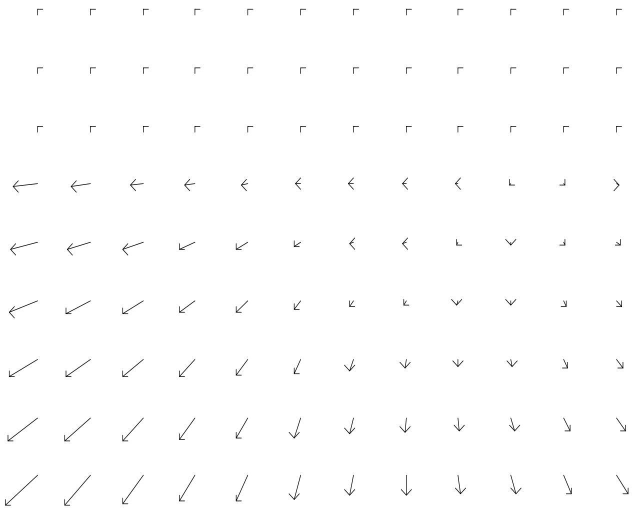

Figure 4: The actual optical flow field corresponding to the frame from the "Yosemite" sequence shown in the previous figure.

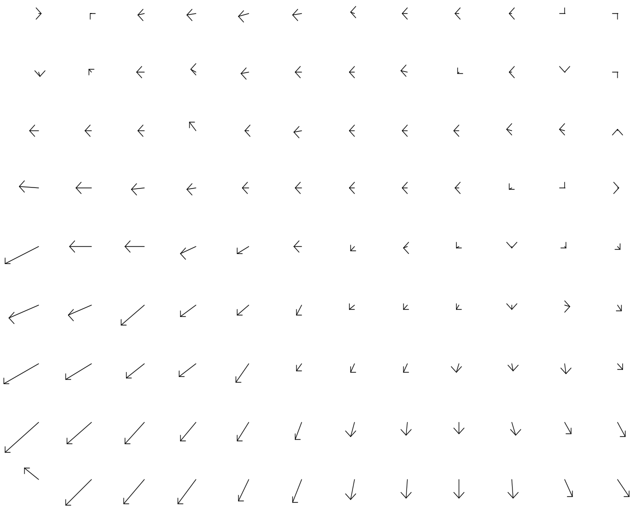

Figure 5: The recovered optical flow corresponding to the frame from the "Yosemite" sequence shown in the previous figure.

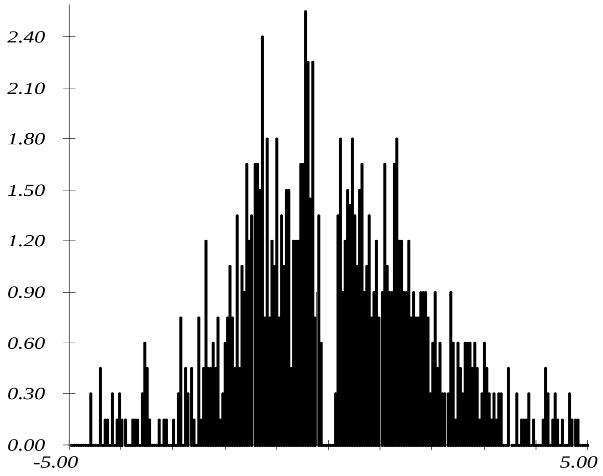

Figure 6: Normalized histogram of the deviations D of the optical flow vectors (see text).

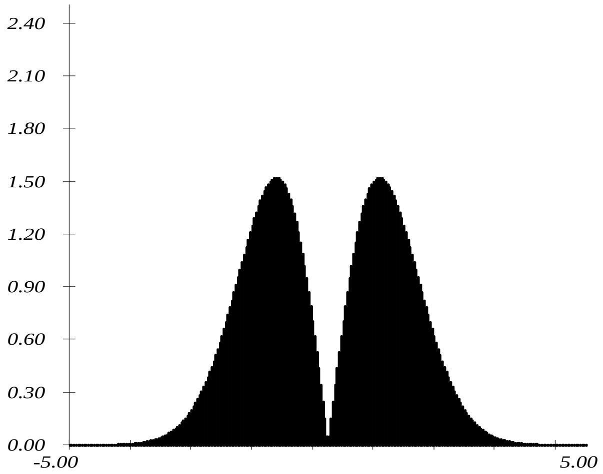

Figure 7: Ideal distribution of deviations D for the gaussian noise model (see text).

of D, is shown in figure 6. If the flow field errors were exactly modeled by the simple additive gaussian noise terms of equation (7), 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 7. 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 8 and the corresponding frame of the recovered flow field is shown in figure 9.

Figure 8: A frame from the original "Nap-of-the-earth" (NOE) fly-through sequence. The sequence was was provided by the NASA Ames Research Center.

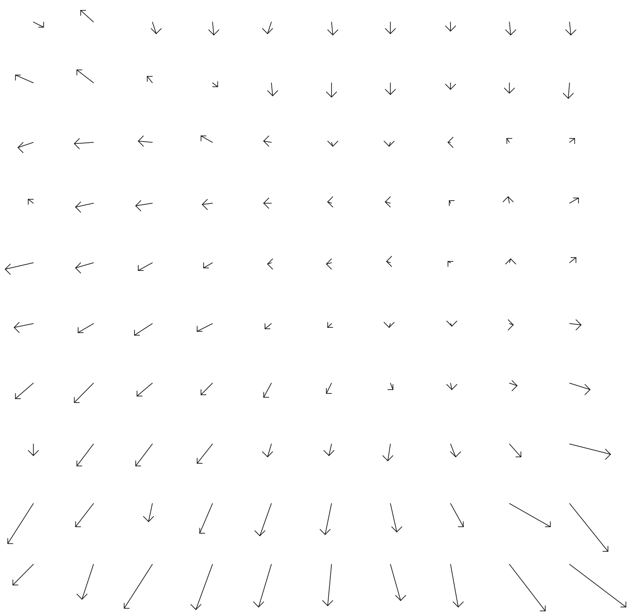

Figure 9: A frame of the recovered optical flow for the NOE sequence.

5 Conclusions

We have introduced probabilistic extensions of gradient techniques which compute two-dimensional optical flow distributions. Viewing the problem probabilistically has three advantages: (1) It produces useful extensions of the standard quadratic gradient techniques for computing optical flow, including an automatic gain control mechanism, and the incorporation of a prior bias on the flow, (2) It provides (two-dimensional) flow vector confidence information, allowing later stages of processing to weight their use of the vectors accordingly, and (3) It provides a framework for properly combining flow information with probabilistic information from other sources. Future work will include quantitative analysis of the errors in the recovered flow, and extensions of the distributed approach to handle the multi-modal flow distributions that arise near motion boundaries and in situations of motion transparency.

References

[1] B K P Horn and B G Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[2] B D Lucas and T Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, 1981.

[3] P Anandan. A computational framework and an algorithm for the measurement of visual motion. *Intl. J. Comp. Vis.*, 2:283–310, 1989.

[4] David J. Heeger. Optical flow using spatiotemporal filters. *Intl. J. Comp. Vis.*, pages 279–302, 1988.

[5] Richard Szeliski. Bayesian modeling of uncertainty in low-level vision. *International Journal of Computer Vision*, 5(3):271–301, December 1990.

[6] David J. Heeger. Personal Communication, July 1990.

[7] Lyn Quam. Hierarchical warp stereo. In *Proceedings of the DARPA Image Understanding Workshop*, September 1984.

[8] Peter J. Burt. Fast filter transforms for image processing. *Computer Graphics and Image Processing*, 16:20–51, 1981.