On Optical Flow

Article · September 1999

Source: CiteSeer

CITATIONS
19

READS
232

3 authors, including:

John L. Barron
The University of Western Ontario
103 PUBLICATIONS **8,452** CITATIONS

SEE PROFILE

David J. Fleet
University of Toronto
250 PUBLICATIONS **33,986** CITATIONS

SEE PROFILE

On Optical Flow

J. L. Barron
Dept. of Computer Science, University of Western Ontario
London, Ontario, N6A 5B7, Email: barron@csd.uwo.ca

S. S. Beauchemin
Dept. of Computer Science, University of Western Ontario
London, Ontario, N6A 5B7, Email: beau@csd.uwo.ca

D. J. Fleet
Dept. of Computing Science, Queen's University
Kingston, Ontario, K7L 3N6, Email: fleet@qucis.queensu.ca

Abstract

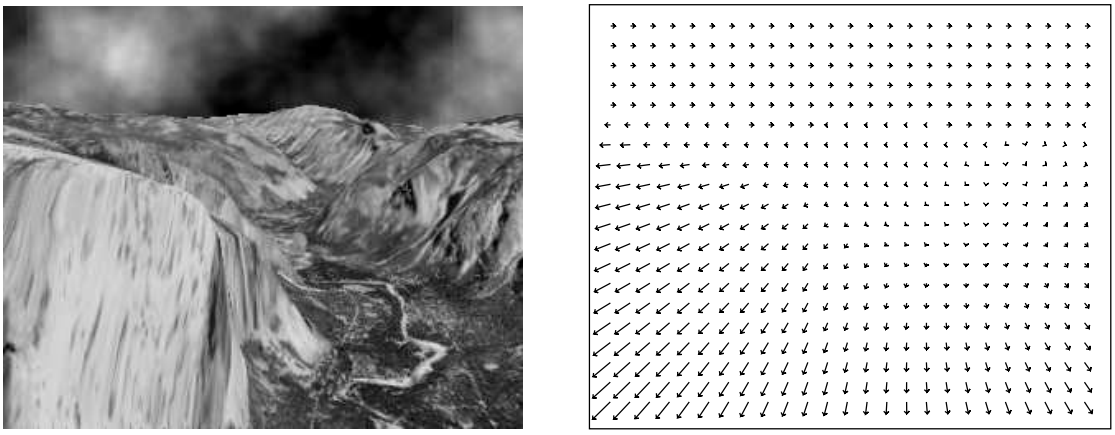
The measurement of optical flow is a fundamental problem in Computer Vision. Many techniques have been presented in the literature and many more continue to appear. How would one select an appropriate method for a given task? We overview the various classes of optical flow methods and describe examples of each, emphasizing both the advantages and drawbacks of each class of methods. We conclude the paper with an example of optical flow as a tool for measuring the growth of corn seedlings.

1 Introduction

A fundamental problem in the processing of image sequences is the measurement of **optical flow** or **image velocity**, an approximation to the 2-d motion field of an image sequence. The 2-d image motion is the projection of the 3-d velocities of 3-d object points onto the image plane that results from the relative motion of the sensor and/or the environment (see Figure 1.1). Once computed, optical flow can be used to to infer the 3-d motion and structure parameters of the camera and the scene (including time to collision), to perform motion detection and object segmentation, to perform motion-compensated encoding, and to compute stereo disparity (two papers[5,27] provide an extensive list of references).

Optical flow methods can be classified as belonging to one of three groups: (1) **Differential**: these methods compute image velocity from spatio-temporal intensity derivatives, (2) **Frequency-based**: these methods use energy/phase information in the output of velocity tuned filters and (3) **Matching**: these methods compute image displacements by matching various images features over a small number (usually 2 or 3) images.

All these approaches are thought to be broadly equivalent in many respects, yet differences in implementation can lead to substantial differences in performance. In comparing different optical flow methods, it is useful to distinguish three conceptual stages of processing: (a) prefiltering or smoothing with low-pass/band-pass filters in order to extract signal structure of interest and to enhance the signal-to-noise ratio, (b) the extraction of basic measurements, such as spatiotemporal derivatives (to measure normal components of velocity) or local correlation surfaces and (c) the integration of these measurements to produce a 2-d flow field, which often involves assumptions about the smoothness of the underlying flow field. Some algorithms produce normal velocity measurements (see below), as there is growing interest in the direct use of normal velocity, thereby side-stepping some of the assumptions inherent in current methods for integrating measurements to find 2-d velocity[6].



(a) (b)

Figure 1.1: (a) A frame from the **Yosemite Fly-Through** sequence created by Lynn Quam at SRI and (b) its correct optical flow field.

2 Optical Flow Methods

We begin with a brief description of the three main classes of optical flow methods and give examples of each.

2.1 Differential Techniques

Differential techniques compute velocity from spatiotemporal derivatives of image intensity or filtered versions of the image (using low-pass or band-pass filters). The first instances used first-order derivatives and were derived from image translation[15,23], i.e.

$$I(\mathbf{x}, t) \quad = \quad I(\mathbf{x} - \mathbf{v}\, t, 0) \; , \tag{2.1}$$

where $\mathbf{v} = (u, v)^T$ and $\mathbf{x} = (x, y)$. From a 1^{st} order Taylor expansion of (2.1)[23] or more generally from an assumption that intensity is conserved, $dI(\mathbf{x}, t)/dt = 0$, the *gradient constraint equation* is easily derived:

$$\nabla I(\mathbf{x}, t) \cdot \mathbf{v} \; + \; I_t(\mathbf{x}, t) \quad = \quad 0 \; , \tag{2.2}$$

where $I_t(\mathbf{x}, t)$ denotes the partial time derivative of $I(\mathbf{x}, t)$, $\nabla I(\mathbf{x}, t) = (I_x(\mathbf{x}, t), I_y(\mathbf{x}, t))^T$ is the spatial intensity gradient and $\nabla I \cdot \mathbf{v}$ denotes the usual dot product. Equation (2.2) yields only the normal component of image velocity relative to spatial contours of constant intensity. For example, Figure 2.1a shows a local contour of constant intensity that moves up and to the right with velocity $\mathbf{v}$. The gradient constraint projects $\mathbf{v}$ onto the direction of the spatial gradient (normal to the contour). This annihilates the tangential component, $\mathbf{v}_t$, thereby constraining only the normal component $\mathbf{v}_n = s\mathbf{n}$. The normal speed s and the normal direction $\mathbf{n}$ are given by

$$s(\mathbf{x}, t) \quad = \quad \frac{-\, I_t(\mathbf{x}, t)}{\| \nabla I(\mathbf{x}, t) \|_2} \text{ and } \mathbf{n}(\mathbf{x}, t) \quad = \quad \frac{\nabla I(\mathbf{x}, t)}{\| \nabla I(\mathbf{x}, t) \|_2} \; . \tag{2.3}$$

Another way to see this is to note that there are two unknown components of velocity $\mathbf{v}$ in (2.2), constrained by only one linear equation. This equation describes a line in velocity space as shown in Figure 2.1b. Any velocity on this line satisfies (2.2). The velocity on the line with the smallest magnitude is the normal velocity $\mathbf{v}_n$. Further constraints in addition to (2.2) are necessary to solve for both components of $\mathbf{v}$, uniquely specifying a point on the line. When the local structure of the image is one dimensional, then further constraints on flow are not available in that immediate neighbourhood and velocity is underconstrained; this is commonly called the *aperture problem*[31].

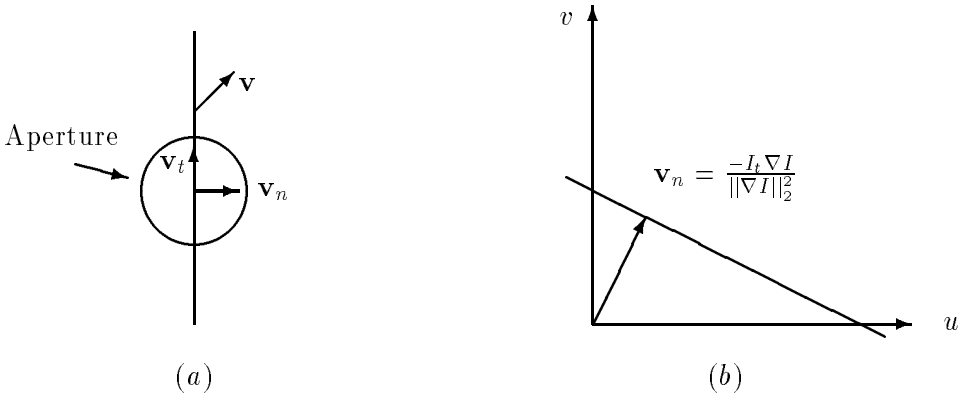


Figure 2.1: (a) The aperture problem: only the component of velocity normal to the line's orientation, $\mathbf{v}_n$, can be recovered. The tangential component of velocity, $\mathbf{v}_t$, cannot be recovered. (b) The Motion Constraint Equation (2.2) yields a line in $\mathbf{v} = (u, v)$ space, the velocity with the smallest magnitude on that line is $\mathbf{v}_n$. Another velocity on the line is the correct full velocity $\mathbf{v}$.

Second-order differential methods use second-order derivatives (the Hessian of I) to constrain 2-d velocity[33,37,38]:

$$\begin{bmatrix} I_{xx}(\mathbf{x}, t) & I_{yx}(\mathbf{x}, t) \\ I_{xy}(\mathbf{x}, t) & I_{yy}(\mathbf{x}, t) \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} I_{tx}(\mathbf{x}, t) \\ I_{ty}(\mathbf{x}, t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.4}$$

Equation (2.4) can be derived from (2.1) or from the conservation of $\nabla I(\mathbf{x}, t)$, $d\nabla I(\mathbf{x}, t)/dt = \mathbf{0}$. Strictly speaking, the conservation of $\nabla I(\mathbf{x}, t)$ implies that first-order deformations of intensity (e.g. rotation or dilation) should not be present. This is therefore a stronger restriction than (2.2) on permissible motion fields. To measure image velocity, assuming $d\nabla I(\mathbf{x}, t)/dt = \mathbf{0}$, the constraints in (2.4) may be used in isolation or together with (2.2) to yield an over-determined system of linear equations[20]. However, if the aperture problem prevails in a local neighbourhood, then because of the sensitivity of numerical differentiation, 2^{nd}-order derivatives cannot usually be measured accurately enough to determine the tangential component of $\mathbf{v}$. As a consequence, velocity estimates from 2^{nd}-order methods are often assumed to be sparser and less accurate than estimates from 1^{st}-order methods.

Another way to constrain $\mathbf{v}(\mathbf{x})$ is to combine local estimates of normal velocity and/or 2-d velocity through space and time, thereby producing more robust estimates of $\mathbf{v}(\mathbf{x})$[35].

There are two common approaches for accomplishing this. The first approach fits the normal velocity measurements in each neighbourhood to a local model for 2-d velocity (e.g. a low-order polynomial model), using least-squares minimization or a Hough transform[15,25,30,35,40]. Usually $\mathbf{v}(\mathbf{x})$ is taken to be constant, although linear models for $\mathbf{v}(\mathbf{x})$ have been used successfully[40,16]. For example, Lucas and Kanade[30] and others[2,25,36], implemented a weighted least-squares (LS) fit of local first-order constraints (2.2) to a constant model for $\mathbf{v}$ in each small spatial neighbourhood Ω by minimizing

$$\sum_{\mathbf{x}\in\Omega} W^2(\mathbf{x})\left[\nabla I(\mathbf{x},\, t)\cdot\mathbf{v} + I_t(\mathbf{x},\, t)\right]^2, \tag{2.5}$$

where $W(\mathbf{x})$ denotes a window function that gives more influence to constraints at the centre of the neighbourhood than those at the periphery. The solution to (2.5) is given by

$$\mathbf{v} \;=\; (A^T W^2 A)^{-1} A^T W^2 \mathbf{b}\;, \tag{2.6}$$

where, for n points $\mathbf{x}_i \in \Omega$ at a single time t,

$$A = [\nabla I(\mathbf{x}_1),\, ...,\, \nabla I(\mathbf{x}_n)]^T, W = \mathrm{diag}[W(\mathbf{x}_1),\, ...,\, W(\mathbf{x}_n)]\,, \text{ and } \mathbf{b} = -(I_t(\mathbf{x}_1),\, ...,\, I_t(\mathbf{x}_n))^T.$$

Here, $A^T W^2 A$ is a 2×2 matrix:

$$A^T W^2 A \;=\; \begin{bmatrix} \sum W^2(\mathbf{x}) I_x^2(\mathbf{x}) & \sum W^2(\mathbf{x}) I_x(\mathbf{x}) I_y(\mathbf{x}) \\ \sum W^2(\mathbf{x}) I_y(\mathbf{x}) I_x(\mathbf{x}) & \sum W^2(\mathbf{x}) I_y^2(\mathbf{x}) \end{bmatrix}, \tag{2.7}$$

where all sums are taken over points in the neighbourhood Ω.

Waxman and Wohn's velocity functional method[40] assumes $\mathbf{v}(\mathbf{x},\, t) = (u(\mathbf{x},\, t),\, v(\mathbf{x},\, t))$ at any image point on a curved surface can be expressed as a 2^{nd} order Taylor series expansion about that point. Using the motion constraint equation, they obtain 1 linear equation relating a normal velocity to $\mathbf{v}(\mathbf{x})$ and its 1^{st} and 2^{nd} order spatial derivatives. Given 12 or more normal velocities in a local neighbourhood one can than solve for $\mathbf{v}$ and its derivatives in the least squares sense. This reduces to 8 equations in the planar case. Murray and Buxton[32] developed a similar approach for planar surfaces.

The second approach uses global smoothness constraints (regularization) in which the velocity field is defined implicitly in terms of the minimum of a functional defined over the image[23,33]. Horn and Schunck[23] combined the gradient constraint (2.2) with a global smoothness term to constrain the estimated velocity field $\mathbf{v}$, minimizing

$$\int_D (\nabla I \cdot \mathbf{v} + I_t)^2 + \lambda^2(\parallel \nabla u \parallel_2^2 + \parallel \nabla v \parallel_2^2)\, d\mathbf{x} \tag{2.8}$$

defined over a domain D, where the magnitude of λ reflects the influence of the smoothness term.

Nagel was one of the first to use second-order derivatives to measure optical flow[33]. Like Horn and Schunck, the basic measurements are integrated using a global smoothness constraint. As an alternative to the constraint in (2.8), Nagel suggested an *oriented-smoothness* constraint in which smoothness is not imposed across steep intensity gradients (edges) in an attempt to handle occlusion. The problem is formulated as the minimization of

$$\int_D (\nabla I^T \mathbf{v} + I_t)^2 + \frac{\alpha^2}{||\nabla I||_2^2 + 2\delta}\left[(u_x I_y - u_y I_x)^2 + (v_x I_y - v_y I_x)^2 + \delta(u_x^2 + u_y^2 + v_x^2 + v_y^2)\right] d\mathbf{x}.$$

(2.9)

Minimizing (2.9) with respect to $\mathbf{v}$ attenuates the variation of the flow $\nabla\mathbf{v}$ in the direction perpendicular to the gradient.

Hildreth[22] proposed a regularization method that finds the velocity of a contour subject to a smoothness constraint and the minimization of the motion constraint equation for each normal velocities measured along the contour.

Of course, one requirement of differential techniques is that $I(\mathbf{x}, t)$ must be differentiable. This implies that temporal smoothing at the sensors is needed to avoid aliasing and that numerical differentiation must be done carefully. We propose that spatio-temporal Gaussian smoothing, typically with a standard deviation of between 1.5 and 3.0 be used and that 4-point central differences with kernels $\frac{1}{12}(-1, 8, 0, -8, 1)$ be used for differentiation[5]. The often stated restrictions that gradient-based techniques require image intensity to be nearly linear, with velocities less than 1 pixel/frame, arise from the use of 2 frames, poor numerical differentiation or input signals corrupted by temporal aliasing. For example, with 2 frames, derivatives are estimated using 1^{st}-order backward differences, which are accurate only when *1)* the input is highly over-sampled or *2)* intensity structure is nearly linear. When aliasing cannot be avoided in image acquisition, one way to circumvent the problem is to apply differential techniques in a coarse-fine manner, for which estimates are first produced at coarse scales where aliasing is assumed to be less severe, with velocities less than 1 pixel/frame. These estimates are then used as initial guesses to warp finer scales to compensate for larger displacements[8,26]. Such methods are discussed at greater length below.

The need for confidence measures on computed velocities cannot be understated, they can be used to threshold flow fields or as a weight term in a subsequent calculation. Most techniques do not provide confidence measures. The smallest eigenvalue of (2.7) was used successfully as a confidence measure[5,36]. Other proposals include the determinant of a Hessian matrix (Gaussian curvature)[41], the condition number of a solution matrix[38] and the magnitude of $||\nabla I||_2$[5].

2.2 Frequency-Based Methods

A second class of optical flow techniques is based on the use of velocity-tuned filters[1,9,21,16]. They are called frequency-based methods owing to the design of velocity-tuned filters in the Fourier domain[1,17]. The Fourier transform of a translating 2-d pattern (2.1) is

$$\hat{I}(\mathbf{k}, \omega) \quad = \quad \hat{I}_0(\mathbf{k})\ \delta(\omega + \mathbf{v}^T\mathbf{k})\ , \tag{2.10}$$

where $\hat{I}_0(\mathbf{k})$ is the Fourier transform of $I(\mathbf{x}, 0)$, $\delta(k)$ is a Dirac delta function, ω denotes temporal frequency and $\mathbf{k} = (k_x, k_y)$ denotes spatial frequency. This shows that all nonzero power associated with a translating 2-d pattern lies on a plane through the origin in frequency space.

Interestingly, it has been shown that certain energy-based methods are equivalent to correlation-based methods[1] and to the gradient-based approach of Lucas and Kanade[2]. Indeed, the motion constraint equation becomes $\mathbf{k} \cdot \mathbf{v} + \omega = 0$ in frequency space.

Heeger's method[21] formulated the optical flow problem as a least-squares fit of energy output of a family of space-time Gabor filters to a plane in frequency space. Ideally, for a single translational motion, the responses of these filters are concentrated about a plane in

frequency space. Heeger derives an equation for the expected response $R(u,\, v)$ of a Gabor-energy filter tuned to frequency $(k_x,\, k_y,\, \omega)$ for translating white noise as a function of velocity. If M_i, $1 \leq i \leq 12$, denote the set of filters with different orientation tunings and $\bar{m}_i$ and $\bar{R}_i$ be the sum of measured and predicted energies, m_j and R_j, from filters j in the set M_i then a least-squares estimate for (u, v) that minimizes the difference between the predicted and measured motion energies is given by the minimum of

$$f(u,v) \;=\; \sum_{i=1}^{12} \left[m_i - \bar{m}_i \frac{R_i(u,v)}{\bar{R}_i(u,v)} \right]^2 . \tag{2.11}$$

The fitting of filter energies to a plane in frequency has become a somewhat common theme, providing the basis for several optical flow methods[9].

The other main optical flow methods that exploit velocity-tuned filters are the class of phase-based approaches. For example, Fleet and Jepson[16] define component velocity in terms of the instantaneous motion normal to level phase contours in the output of band-pass filters. (Note: normal velocity is normal to local intensity structure while component velocity is normal to local phase structure.) The filters are used to decompose the input signal according to scale, speed and orientation. Each filter output is complex-valued and may be written as

$$R(\mathbf{x},\, t) \;=\; \rho(\mathbf{x},\, t)\, \exp[i\phi(\mathbf{x},\, t)] \;, \tag{2.12}$$

where $\rho(\mathbf{x}, t)$ and $\phi(\mathbf{x}, t)$ are the amplitude and phase parts of R. The component of 2-d velocity in the direction normal to level phase contours is then given by $\mathbf{v}_n = s\mathbf{n}$, where the normal speed and direction are given by

$$s \;=\; \frac{-\,\phi_t(\mathbf{x},\, t)}{\parallel \nabla\phi(\mathbf{x},\, t) \parallel_2} \text{ and } \mathbf{n} \;=\; \frac{\nabla\phi(\mathbf{x},\, t)}{\parallel \nabla\phi(\mathbf{x},\, t) \parallel_2} \;, \tag{2.13}$$

where $\nabla\phi(\mathbf{x},\, t) = (\phi_x(\mathbf{x},\, t),\, \phi_y(\mathbf{x},\, t))^T$. In effect, this is a differential technique applied to phase rather than intensity. The phase derivatives are computed using the identity

$$\phi_x(\mathbf{x},\, t) \;=\; \frac{\mathrm{Im}[R^*(\mathbf{x},\, t)\; R_x(\mathbf{x},\, t)\,]}{|R(\mathbf{x},\, t)|^2} \;, \tag{2.14}$$

where R^* is the complex conjugate of R. The use of phase is motivated by their claim that the phase component of band-pass filter outputs is more stable than the amplitude component when small deviations from image translations that regularly occur in 3-d scenes are considered[18]. However, phase can also be unstable, with instabilities occurring in the neighbourhoods about phase singularities. Such instabilities can be detected with a straightforward constraint on the instantaneous frequency of the filter output and its amplitude variation in space-time[17,18]:

$$\parallel\; \nabla \log R(\mathbf{x},\, t) \;-\; i(\mathbf{k},\, \omega)\; \parallel_2 \;\;\leq\;\; \sigma_k\, \tau \;, \tag{2.15}$$

where $(\mathbf{k}, \omega)$ denotes the spatiotemporal frequency to which the filter is tuned, σ_k is the standard deviation of the isotropic amplitude spectra they use and τ denotes a threshold that can be used to reject unreliable component velocity measurements. A second constraint on the amplitude of response is also used to ensure a reasonable signal-to-noise ratio. As currently

formulated Fleet and Jepson's method does not provide confidence measures. Finally, given the stable component velocity estimates from the different filter channels, a linear velocity model is fit to each local region. To ensure that there is sufficient local information for reliable velocity estimates, they introduce further constraints on the conditioning of the linear system and on the residual LS error.

The class of phase-based approaches can also be generalized to include zero-crossing methods[12,13,22,41], since zero-crossings can be viewed as level phase-crossings. For example, Waxman, Wu and Bergholm[41] apply spatiotemporal filters to binary edge maps to track edges in real-time. They apply a number of filters based on Gaussian envelopes and their derivatives to edge maps of an image sequence and derive velocity for ratios of the filters' responses.

2.3 Matching Techniques

Accurate numerical differentiation may be impractical because of noise, because a small number of frames exist or because of aliasing in the image acquisition process. In these cases differential/frequency approaches may be inappropriate and it is natural to turn to matching techniques. Region-based matching[3,28,29] and feature-based matching[4] are two approaches. Region-based matching methods define velocity $\mathbf{v}$ as the shift $\mathbf{d} = (d_x, d_y)$ that yields the best fit between image regions at different times. Finding the best match amounts to maximizing a similarity measure (over $\mathbf{d}$), such as the normalized cross-correlation or minimizing a distance measure, such as the sum-of-squared difference (SSD):

$$\begin{aligned} SSD_{1,2}(\mathbf{x};\mathbf{d}) &= \sum_{j=-n}^{n}\sum_{i=-n}^{n} W(i,j)\,[I_1(\mathbf{x}+(i,j)) - I_2(\mathbf{x}+\mathbf{d}+(i,j))]^2 \\ &= W(\mathbf{x}) * [I_1(\mathbf{x}) - I_2(\mathbf{x}+\mathbf{d})]^2 \ , \end{aligned} \tag{2.16}$$

where W denotes a discrete 2-d window function, and $\mathbf{d} = (d_x, d_y)$ take on integer values.

There is a close relationship between the SSD distance measure, the cross-correlation similarity measure, and differential techniques. Minimizing the SSD distance amounts to maximizing the integral of product term $I_1(\mathbf{x})I_2(\mathbf{x}+\mathbf{d})$. Also, the difference in (2.16) can be viewed as a window-weighted average of a first-order approximation to the temporal derivative of $I(\mathbf{x}, t)$.

Matching techniques reported by Anandan[3] and Singh[35] are based on a Laplacian pyramid and a coarse-to-fine SSD-based matching strategy. The Laplacian pyramid[11] allows the computation of large displacements between frames and helps to enhance image structure. The pyramid image structure also yields considerable computational savings. Measures of principle curvature[3] or eigenvalues of a covariance matrix[35] served as confidence measures.

3 Performance Analysis

We have conducted a performance analysis of nine different optical flow techniques[5] that includes instances of the three classes of optical flow presented above. For a number of synthetic image sequences (for which the correct flows are known) and real image sequences (for which the correct flows are unknown) we analyzed these techniques with respect to the accuracy, density of flow and computational efficiency. In table 1.1 we show a set of error results for the Yosemite Fly Through sequence for a number of techniques described in[5]. Since the image sequences were all appropriately sampled with small motions (typically 1-4

pixels/frame) conditions were favourable for differential approaches. For example, we did not consider *stop-and-shoot* sequences[14]; in fact, only matching techniques would be appropriate here for this type of data due to their robustness with respect to aliasing. We found Fleet and Jepson's phase-gradient based method[16] produced the best quantitative results overall with Lucas and Kanade's intensity-gradient based method[30] being second. Lucas and Kanade's method was significantly more efficient than Fleet and Jepson's method and in addition provided good confidence measures[36]. Matching (correlation-based) techniques were not as good, the error depended on how close the motion was to an integer number of pixels per frame, that is, these techniques had difficulty in measuring subpixel motions. Matching techniques are obviously more suitable for larger motions and aliased (under sampled) image sequences (the coarse-to-fine processing in a Gaussian or Laplacian pyramid yield these properties), which we did not test. The optical flow programs and data used in this study are available via ftp (`ftp` to `ftp.csd.uwo.ca` and `cd` to `pub/vision`) and comparative results are being reported in current conferences.

Technique	Average Error	Standard Deviation	Density
Horn and Schunck (original)	31.69°	31.18°	100%
Horn and Schunck (modified)	9.78°	16.19°	100%
Horn and Schunck (modified) $\|\|\nabla I\|\| \geq 5.0$	5.59°	11.52°	32.9%
Lucas and Kanade ($\lambda_2 \geq 1.0$)	4.28°	11.41°	35.1%
Uras et al. (unthresholded)	8.94°	15.61°	100%
Nagel	10.22°	16.51°	100%
Nagel $\|\|\nabla I\|\|_2 \geq 5.0$	6.06°	12.02°	32.9%
Anandan	13.36°	15.64°	100%
Singh (Step 1, $n = 2$, $w = 2$)	15.28°	19.61°	100%
Singh (Step 2, $n = 2$, $w = 2$)	10.44°	13.94°	100%
Heeger	11.93°	23.16°	44.8%
Waxman et al. $\sigma_f = 2.0$	20.05°	23.23°	7.4%
Fleet and Jepson ($\tau = 1.25$)	5.28°	14.34°	30.6%
Fleet and Jepson ($\tau = 2.5$)	4.63°	13.42°	34.1%

Table 3.1: *Summary of* **Yosemite** *2-d Velocity Results*

More recently, we examined the use of reconstruction (warping) error as a quantitative error metric for those flows where the correct flow fields are unknown[27]. Given a computed optical flow field for some image we use it to reconstruct the next image in the sequence and then compute the RMS difference between the actual next image and its reconstructed version. We found that several backward and forward reconstruction methods using bicubic interpolation on Gaussian smoothed or confidence measure weighted velocity fields provided the best correlation between quantitative and RMS errors for synthetic images sequences (the reader is directed to a paper[27] for full details). Reconstruction error was then used to provide a quantitative error metric for real image sequences with unknown correct flow.

4 Problems and Current Research Directions

Still, several major problems remain:

- Acquisition of image sequences using conventional video cameras typically exhibit temporal aliasing. The velocities are often larger than 10 pixels/frame, and there is only a minimal amount of temporal blurring. In addition, as there is usually little mechanical stabilization, the motion often varies radically from frame to frame.

- Lighting effects can also be a problem in many image sequences. Constant scene illumination and Lambertian reflectance are assumed for most current optical flow methods (inherent in the use of (2.2)). Highlights, shadows, transparency, and varying illumination, have only been studied to a limited degree.

- Finally, occlusion remains a very difficult problem, especially for small fragmented objects such as bushes or trees.

Some reflectance phenomena, for example transparency, can lead to multiple motions. This can usually be dealt with using some of the approaches to occlusion (segmentation) described below.

4.1 Warping and Coarse-to-Fine Strategies

One way to deal with large erratic displacements is to dispense with larger amounts of temporal filtering, and instead use coarse-to-fine control strategies and image warping in order to keep the image sufficiently well registered at the scale of interest where simple forms of numerical differentiation remain reasonably accurate[3,8,35]. Bergen et al.[8] proposed a hierarchical framework that unifies several different model-based flow methods. Using an affine flow model in an image region allows one both to judge the quality of that model's fit to the data (perhaps splitting the region if necessary) and to fill in sparse flow fields using the computed affine parameters. Multiple motions can also be thought of as a set of velocity layers, each describing a particular motion[39].

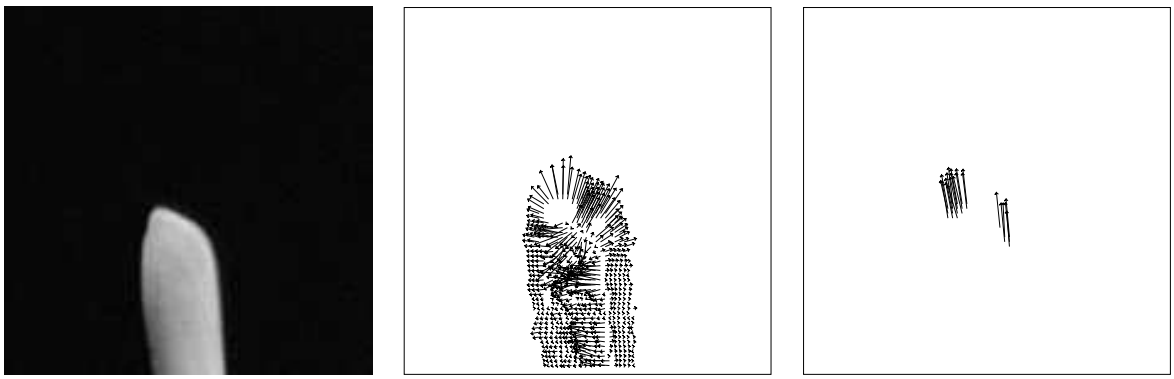
Such modifications to the basic framework appear to be suitable for large regions of the image. But the tracking of small target moving over longer distances within this framework is still often problematic.

4.2 Motion Segmentation

Another current theme in current research on flow concerns the notion of segmentation in the face of occlusion and transparency. Attempts to estimate discontinuous motion can be classified into two groups. The first group consists of binary line processes[19] that explicitly represent intensity discontinuities, for example by *breaking* the smoothness constraint at pixels having a large spatial gradient[10,26]. Another class of methods iteratively decomposes the scene into different objects, by first fitting a single model, then finding regions inconsistent with the origin fit. These inconsistent regions then becomes starting points for another single motion model, from which regions of large deviation can be extracted, etc. A variation on this theme involves the use of mixed velocity distribution models[24,34] to find several velocities simultaneously in a given region of the image. Here the less dominant motions are initially treated as outliers. Mixed velocity models are also appropriate for measuring transparent motions. Fleet and Jepson's method is capable of measuring component velocities but no means is provided for segmenting the component velocities into the two or more full velocities each of the component velocities belong to.

5 Plant Growth: An Application of Optical Flow

We used the optical flow method attributed to Lucas and Kanade[30] with some of the modifications proposed by Simoncelli et al.[36] plus some additional thresholding criteria stated in Barron and Liptay[7] to measure the growth rate of a corn seeding using optical flow. We used a Sony colour camera with non-square pixels: each pixel was 9.8×10^{-6} meters high and 8.4×10^{-6} meters wide. We scaled all horizontal components of velocity by $\frac{8.4}{9.8}$ to take this into account. We used a YIQ colour to grayvalue transformation (each pixel of a grayvalue image was computed as 0.299, 0.587 and 0.114 respectively of the corresponding red, green and blue pixels of a colour image) to convert the images into grayscale images. Constant spatio-temporal scene illumination was maintained during the growth sequence by using fluorescent lighting and blocking the windows of the experimental room to ensure the room was devoid of natural light. Constant scene illumination ensures there are no non-zero intensity derivatives not due to motion and hence no detection of false motion. Surface reflectivity was minimized by using a black rayon velvet cloth as the scene's background. The camera was placed in close proximity to the front of the seedling to ensure sufficient textural detail was captured by image acquisition. The temporal sampling was 2 minutes per frame and was empirically determined. We employed both eigenvalue thresholding[5] and similarity thresholds (local velocities should be similar and upwards)[7] to remove outlier velocities. Figure 5.1 shows one image of the sequence plus the computed normal and thresholded full image velocity fields. The average rate of growth for this image was $1.982 \times 10^{-7} \pm 8.9393 \times 10^{-9}$ meters per second. An experiment using optical flow to measure plant growth while the plant's root temperature was varied showed that there was good correlation between root temperature and growth rate.



(a) Corn Seedling **(b) Normal Flow** **(c) Full Flow**

Figure 5.1: *An image of the corn seedling sequence and its normal and full optical flow fields. The corn seedling's root temperature was 24° C, the normal and full flow fields were sampled by 2 and scaled by 50.0. The flows were computed using balanced smoothing and differentiation. The full flow (c) was computed from neighbourhoods of local normal flow (b) and then thresholded as described in Barron and Liptay[7].*

Acknowledgements This work has been supported in part by NSERC Canada, the Gov-

ernment of Canada (through IRIS) and the Government of Ontario (through ITRC).

References

1. E. H. Adelson and J. R. Bergen, *Spatiotemporal energy models for the perception of motion*, J. Opt. Soc. Am. **A 2(2)** (1985), pp284-299.
2. E.H. Adelson and J.R. Bergen ,*The extraction of spatiotemporal energy in human and machine vision*. IEEE Proc. of Workshop on Visual Motion (1986), pp151-156.
3. P. Anandan, *A computational framework and an algorithm for the measurement of visual motion*, IJCV **2** (1989), pp283-310.
4. S. T. Barnard and W. B. Thompson, *Disparity analysis of images*, IEEE PAMI **2(4)** (1980), pp333-340.
5. J. L. Barron, D. J. Fleet, and S. S. Beauchemin, *Performance of optical flow techniques*, IJCV **12(1)** (1994), pp43-77.
6. J. L. Barron, A. D. Jepson, and J. K. Tsotsos, *The feasibility of motion and structure from noisy time-varying image velocity information*, IJCV **5(3)** (1990), pp239-269.
7. J.L. Barron and A. Liptay, *Optical flow to measure minute increments in plant growth*, BioImaging (to appear), 1994.
8. J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg, *Three-frame algorithm for estimating two-component image motion*, IEEE PAMI **14(9)** (1992),pp886-896.
9. J. Bigun, G. Granlund, and J. Wiklund, *Multidimensional orientation estimation with applications to texture analysis and optical flow*, IEEE PAMI **13** (1991), pp775-790.
10. M. J. Black, *Robust incremental optical flow*, PhD thesis, Yale University, (1992).
11. P.J. Burt and E.H. Adelson, *The laplacian pyramid as a compact image code*, IEEE Trans. on Communications **31** (1983), pp532-540.
12. B. F. Buxton and H. Buxton, *Computation of optic flow from the motion of edge features in image sequences*, Image and Vision Computing **2(2)** (1994), pp59-75.
13. J.H. Duncan and T.C. Chou, *On the detection of motion and the computation of optical flow*, IEEE PAMI **14(3)** (1992), pp346-352.
14. R. Dutta, R. Manmatha, L. Williams, and E.M. Riseman, *A data set for quantitative motion analysis*, IEEE Proc. of CVPR, San Diego (1989), pp159-164.
15. C. Fennema and W. Thompson, *Velocity determination in scenes containing several moving objects*, CGIP **9** (1979), pp301-315.
16. D. J. Fleet and A. D. Jepson, *Computation of component image velocity from local phase information*, IJCV **5(1)** (1990), pp77-104.
17. D.J. Fleet, *Measurement of Image Velocity*, Kluwer Academic Publishers, Norwell, (1992).
18. D.J. Fleet and A.D. Jepson, *Stability of phase information*, IEEE PAMI (in press), 1993.
19. S. Geman and D. Geman. *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images.* IEEE PAMI **6(6)** (1984), pp721-741.
20. F. Girosi, A. Verri, and V. Torre, *Constraints for the computation of optical flow*, IEEE Proc. of Visual Motion Workshop (1989), pp116-124.

21. D. J. Heeger. *Optical flow using spatiotemporal filters*, IJCV, **1** (1988), pp279-302.
22. E. C. Hildreth, *The computation of the velocity field*, Proc. R. Soc. Lond., **B 221** (1984), pp189-220.
23. B. K. P. Horn and B. G. Schunck, *Determining optical flow*, Artificial Intelligence **17** (1981), pp185-204.
24. A. D. Jepson and M. Black, *Mixture models for optical flow computation*, Technical Report RBCV-TR-93-44 University of Toronto, (1993).
25. J.K. Kearney, W.B. Thompson, and D.L. Boley, *Optical flow estimation: An error analysis of gradient-based methods with local optimization*, IEEE PAMI **9** (1987), pp229-244.
26. C. Koch, H. T. Wang, R. Battiti, B. Mathur, and C. Ziomkowski, *An adaptive multiscale approach for estimating optical flow: Computational theory and physiological implementation*, IEEE Proc. of Workshop on Visual Motion (1991), pp111-122.
27. T. Lin and J.L. Barron, *Image reconstruction error for optical flow*, (to appear) Vision Interface, May (1994).
28. J. J. Little and A. Verri. *Analysis of differential and matching methods for optical flow*, IEEE Proc. of Visual Motion Workshop (1989), pp173-179.
29. J.J. Little, H.H. Bulthoff, and T.A. Poggio, *Parallel optical flow using local voting*, IEEE Proc. of ICCV (1988), pp454-459.
30. B. D. Lucas and T. Kanade, *An iterative image-registration technique with an application to stereo vision*, DARPA Image Understanding Workshop (1981), pp121-130.
31. D. Marr and E. C. Hildreth, *Theory of edge detection*, Proc. R. Soc. Lond. **B 290** (1980), pp187-212.
32. D.W. Murray and B.F. Buxton, *Reconstructing the optic flow from edge motion: An examination of two different approaches* , 1st Conf. on AI Applications (1984), pp382-388.
33. H. H. Nagel, *On the estimation of optical flow: Relations between different approaches and some new results*, Artificial Intelligence **33**, pp299-324.
34. B. G. Schunck, *Image flow segmentation and estimation by constraint line clustering*, IEEE PAMI **11(10)** (1989), pp1010-1027.
35. A. Singh, *Optic flow computation: A unified perspective*, IEEE Computer Society Press, (1992).
36. E.P. Simoncelli, E.H. Adelson and D.J. Heeger, *Probability distributions of optical flow*, IEEE Proc. of CVPR (1991), pp310-315.
37. O. Tretiak and L. Pastor, *Velocity estimation from image sequences with second-order differential operators*, IEEE Proc. of ICPR (1984), pp20-22.
38. S. Uras, F. Girosi, A. Verri, and V. Torre *A computational approach to motion perception*, Biological Cybernetics **60** (1988), pp79-87.
39. J. Wang and E.H.. Adelson, *Layered representation for motion analysis*, IEEE Proc. of CVPR (1993).
40. A. M. Waxman and K. Wohn, *Contour evolution, neighborhood deformation, and global image flow: planar surfaces in motion*, Int. J. of Rob. Vis. **4(3)** (1985).
41. A. M. Waxman, J. Wu, and F. Bergholm, *Convected activation profiles and the measurement of visual motion*, IEEE Proc. of CVPR (1988), pp717-723.