International Journal of Computer Vision, 6:2, 133–145 (1991)

Computing Optical Flow Across Multiple Scales: An Adaptive Coarse-to-Fine Strategy

ROBERTO BATTITI*, EDOARDO AMALDI**, AND CHRISTOF KOCH
Computation and Neural Systems Program, Division of Biology, 216-76, California Institute of Technology, Pasadena, CA

Received April 9, 1990. Revised October 15, 1990

Abstract
Single-scale approaches to the determination of the optical flow field from the time-varying brightness pattern assume that spatio-temporal discretization is adequate for representing the patterns and motions in a scene. However, the choice of an appropriate spatial resolution is subject to conflicting, scene-dependent, constraints. In intensity-based methods for recovering optical flow, derivative estimation is more accurate for long wavelengths and slow velocities (with respect to the spatial and temporal discretization steps). On the contrary, short wavelengths and fast motions are required in order to reduce the errors caused by noise in the image acquisition and quantization process.

Estimating motion across different spatial scales should ameliorate this problem. However, *homogeneous* multiscale approaches, such as the standard multigrid algorithm, do not improve this situation, because an optimal velocity estimate at a given spatial scale is likely to be corrupted at a finer scale. We propose an *adaptive* multiscale method, where the discretization scale is chosen locally according to an estimate of the relative error in the velocity estimation, based on image properties.

Results for synthetic and video-acquired images show that our coarse-to-fine method, fully parallel at each scale, provides substantially better estimates of optical flow than do conventional algorithms, while adding little computational cost.

1 Reliable Estimation of the Optical Flow

During the last decade there has been an increasing interest in analyzing sequences of time-varying images and in particular in determining the 2-D motion or velocity field, which is the projection of the 3-D velocity field onto the image plane—see (Nagel 1978) and (Verri & Poggio 1989) for reviews. The two main approaches that have been proposed for determining the optical flow (the apparent motion of the brightness pattern approximating the underlying motion field) are *intensity-based methods* and methods based on the *matching of tokens*, such as zero-crossings or other high-level features—for a review see (Ullman 1981) and (Hildreth & Koch 1987). The intensity-based methods, in turn, can be subdivided into two subclasses. *Correlation, second-order* or *spatio-temporal energy* models essentially multiply the linearly filtered intensity value at a given point with the linearly filtered intensity value arriving, delayed in time, from a neighboring receptor (Hassenstein & Reichardt 1956; Adelson & Bergen 1985; van Santen & Sperling 1984; Poggio & Reichardt 1973; Watson & Ahumada 1985; Reichardt et al. 1988). *Differential methods*, on the other hand, exploit the relationship between velocity and spatial and temporal gradients in the image brightness (Fennema & Thompson 1979; Horn & Schunck 1981; Hildreth 1984; Bülthoff et al. 1989; Nagel 1978; Yuille & Grzywacz 1988; Wang et al. 1989; Uras et al. 1988).

*Now at Dipartimento di Fisica, Univ. de Genova, Via Dodecanesco 33, 16146, Genoa, Italy.
**Now at Department of Mathematics, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland.

Both the correlation as well as the gradient approach make a basic assumption about the *scale* of the velocity relative to the spatial neighborhood and to the temporal discretization step or delay. For instance, if the velocity of the pattern is much larger than the ratio of the spatial to the temporal sampling step an incorrect velocity value will be obtained.

Direction-selective cells in the primate visual system exhibit a range of spatial sizes, in particular if receptive field size is compared between different cortical areas, such as between the primary visual cortex (V1) and the middle temporal area (MT;—see, for instance, (Maunsell & Van Essen 1983). We were thus motivated to study how integrating motion information across different spatial scales could help improving the estimate of the optical flow—see also (Koch et al. 1989).

The multigrid algorithm with the "full approximation storage" scheme has been suggested as a way to solve the differential equation in Horn and Schunck's method (Brandt 1977; Terzopoulos 1986). This algorithm is computationally more efficient than single-scale methods (it converges in a time proportional to the number of pixels in the image) and leads to a consistent result at different spatial scales. Unfortunately, both the multigrid method and simpler coarse-to-fine continuation schemes tend to suffer from their homogeneous computational structure. In some cases this causes the optical-flow detection process to oscillate between different estimates at different scales or even to converge to a wrong solution (Enkelmann 1988; Glazer 1984). Indeed, if no explicit direction is given in order to select locally the appropriate scale, different scales will, in general, provide conflicting information.

We propose a method for tuning the discretization grid to a measure of the reliability of the information derived from a given scale. This measure will be based on a local estimate of the *relative error* in the flow field due to noise and discretization. The flow of control is from coarse to fine scale. We present some relevant experimental results obtained with synthetic and real-world image sequences.

The main qualities of our approach are its algorithmic simplicity and its parallel nature at any given scale, making it a valid candidate for real-time vision systems as well as for the mammalian visual system. Alternative approaches, based on local optimization and iterative registration, have been studied by Kearney, Thompson, and Boley (1984).

In this article we do not discuss the introduction of binary motion discontinuities á la Geman and Geman, which are necessary in order to prevent smoothing over different physical objects and preserve object boundaries (Battiti 1989; Geman & Geman 1984; Harris et al. 1990; Hutchinson et al. 1988; Poggio et al. 1988; Marroquin 1984). We discuss in another publication both psychophysical and electrophysiological evidence favoring the existence of a multiscale, coarse-to-fine strategy for computing optical flow in the primate visual system (Wang et al. 1991).

2 Shortcomings of Homogeneous Differential Methods

Homogeneous differential methods estimate the optical flow using a hierarchy of resolution grids and a solution process that is the same for all points in the image. In the following, we first summarize Horn and Schunck's approach (Horn & Schunck 1981) showing some well-known limitations related to quantization of intensity values and estimation of derivatives with discretized formulas. We will then motivate our alternative *adaptive* strategy.

Horn and Schunck start with a definition of the optical flow given by the following *brightness constancy equation*—see also (Fennema & Thompson 1979; Nagel 1978)

$$\frac{dE}{dt} = E_x u + E_y v + E_t = 0 \qquad (1)$$

with the optical flow given by $(u, v) = (dx/dt, dy/dt)$ and E_x, E_y, and E_t denoting the spatial and temporal brightness derivatives. This, of course, yields only one linear equation in two unknowns, giving rise to the well-known aperture problem (Marr & Ullman 1981). In this way of formulating the optical flow, the problem is ill-posed (Poggio et al. 1985). Horn and Schunck used a membrane-type of smoothness constraint to regularize the problem, leading to the minimization of the following functional:

$$\Phi = \int\int_{\text{Image}} (E_x u + E_y v + E_t)^2 + \alpha^2(u_x^2 + u_y^2 + v_x^2 + v_y^2)dxdy \qquad (2)$$

with the regularization parameter α controlling the amount of smoothness. This functional embodies the conflicting demands of faithfulness to the data and smoothness. The appropriate Euler-Lagrange equations

$$(E_x u + E_y v + E_t)E_x = \alpha^2 \Delta u \tag{3}$$

$$(E_x u + E_y v + E_t)E_y = \alpha^2 \Delta v \tag{4}$$

give a necessary condition for an extremum of Φ. The algebraic system obtained after discretization can be solved using local and iterative "relaxation" methods. The solution method used throughout this work is Gauss-Seidel lexicographic relaxation. During an updating cycle the new approximation (u^{n+1}, v^{n+1}) of the flow field can be determined from the estimated brightness derivatives $(\tilde{E}_x, \tilde{E}_y$, and $\tilde{E}_t)$ and from the local average $(\bar{u}^n, \bar{u}^n)$ of the previous flow estimate by

$$u^{n+1} = \bar{u}^n - \frac{\tilde{E}_x(\tilde{E}_x\bar{u}^n + \tilde{E}_y\bar{v}^n + \tilde{E}_t)}{(\alpha^2 + \tilde{E}_x^2 + \tilde{E}_y^2)} \tag{5}$$

$$v^{n+1} = \bar{v}^n - \frac{\tilde{E}_x(\tilde{E}_x\bar{u}^n + \tilde{E}_y\bar{v}^n + \tilde{E}_t)}{(\alpha^2 + \tilde{E}_x^2 + \tilde{E}_y^2)} \tag{6}$$

Free boundary conditions are given by zero normal derivative. In the present scheme, computation starts from a field equal to zero on the coarsest scale, while in a real-time continuous scheme it should start from the previously determined field.

The basic assumption made in solving equations (3) and (4) using discretized versions (5) and (6) is that the spatial and temporal sampling steps are small with respect to the given image features and motion amplitudes. If the brightness changes rapidly on the scale given by the discretization step, the accuracy of the formulas for derivative estimation decreases, because in this case the step cannot be considered infinitesimal.

In the one-dimensional case, the derivative estimation problem can be illustrated by considering a one-dimensional sinusoidal intensity profile $\sin(2\pi/L)(x - vt)$ of wavelength L moving with velocity v. The brightness constancy equation determines the optical flow uniquely and the measured velocity v is given by

$$\tilde{v} = -\frac{\tilde{E}_t}{\tilde{E}_x}$$

$$= \frac{\dfrac{\sin(2\pi/L)(x + vt + v\Delta t) - \sin(2\pi/L)(x + vt - v\Delta t)}{2\Delta t}}{\dfrac{\sin(2\pi/L)(x + vt + \Delta x) - \sin(2\pi/L)(x + vt - \Delta x)}{2\Delta x}}$$

$$= \frac{\sin\,[(2\pi/L)v\Delta t]\;\Delta x}{\sin\,[(2\pi/L)\Delta x]\;\Delta t} \tag{7}$$

where $\tilde{E}_x$ and $\tilde{E}_t$ are the three-point approximations of the spatial and temporal brightness derivatives obtained using the spatial and temporal discretization steps Δx and Δt. Three-point derivatives provide a better estimate than the standard two-point forward difference formula ($O(\Delta x)^2$ versus $O(\Delta x)$). Moreover, the temporal and spatial derivatives are estimated at the same point—no phase shift is present, as explained by Little and Verri (1989). Figure 1 shows some characteristic graphs of the relative error in the velocity as a function of the true velocity v for different values of the dimensionless ratios $v\Delta t/L$ and $\Delta x/L$.

While in the limit of L converging to infinity, equation (7) converges to the correct velocity v, the relative error in the computed velocity becomes of the order of 100% even for small velocities when the wavelength is less than approximately three spatial sampling steps ($L < 3\Delta x$). When the wavelength is between one and two sampling steps ($\Delta x < L < 2\Delta x$), *motion reversal* may occur (depending on the magnitude of $v\Delta t$), that is, sign $(\tilde{v} \cdot \text{sign}\,(v)) = -1$. If the grid size Δx goes to zero, the velocity estimate in equation (7) converges to

$$\frac{\sin\left(\dfrac{2\pi}{L}\,v\Delta t\right)}{\dfrac{2\pi\Delta t}{L}}$$

Thus, reducing the grid size to very small values while leaving the temporal sampling rate fixed, will not necessarily lead to a better velocity estimate. In fact, the estimated velocity $\tilde{v}$ is equal to the true velocity v only if $\Delta x = v\Delta t$, that is if the grid size is identical to the interframe motion!

To deal with this problem one can consider a resolution pyramid—see Burt (1984) and the contained references. Because the high spatial frequencies are attenuated at lower spatial resolution, the spatial and temporal derivatives of the brightness are smoothed and their estimate is more accurate, provided that discretization errors do not become dominant. There has been some previous work on multiscale determination of the optical flow (Enkelmann 1988; Glazer 1984; Terzopoulos 1986). Terzopoulos applied the multigrid algorithm to the Euler-Lagrange equations (3) and (4). The idea of the multigrid method consists of starting from an approximation with smoothed error obtained by relaxation on the fine grid and in determining a correction of this approximation on the coarser grid. This

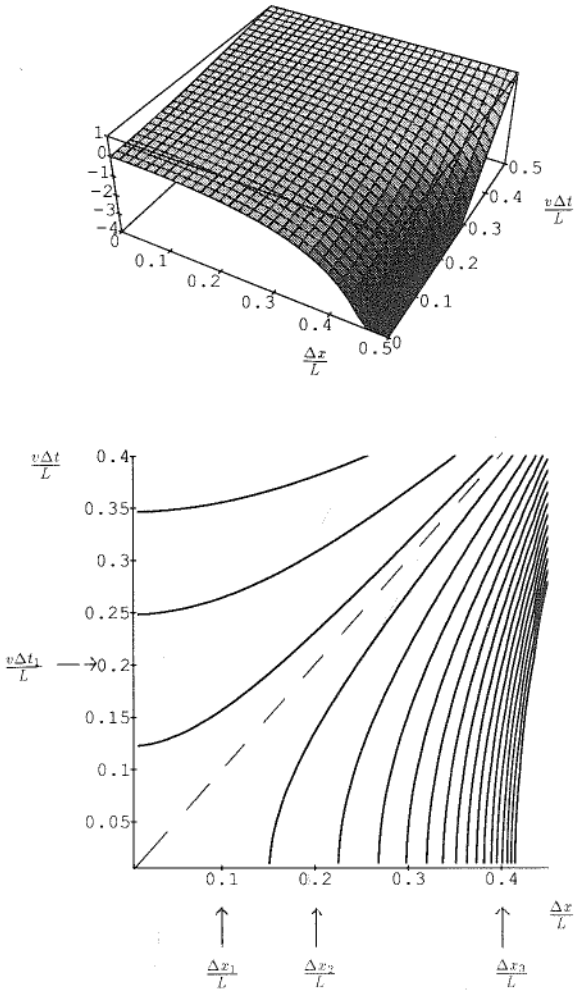

Fig. 1. Estimated relative error, $(v - \tilde{v})/v$ in the case of a single moving sinusoidal pattern $f(x, t) = \sin(2\pi/L)(x - vt)$ (with $v = 1$), as a function of the relative spatial $\Delta x/L$ and temporal sampling steps $v\Delta t/L$. The error diverges for $\Delta x/L = 0.5$. The estimated velocity $\tilde{v}$ was computed following equation (7), based on the three-point approximation for the spatial and temporal derivatives. The 3D plot illustrates that the error is zero along the diagonal $\Delta x = v\Delta t$. In other words, the velocity estimate is optimal if the spatial sampling step is identical to the interframe motion. The relative error becomes negative (motion reversal) for all points $\Delta x > v\Delta t$, is positive for $\Delta x < v\Delta t$ (for $\Delta x < 0.5$), and diverges for $\Delta x = 0.5$. The second plot shows lines of constant relative error spaced 0.2 apart. For points falling on the diagonal $\Delta x = v\Delta t$ (dotted line), the error is zero. We indicated schematically the situation occurring with the use of three spatial and one temporal sampling intervals.

is computationally less expensive and can be done recursively by relaxation on the coarse grid with correction on the next coarser grid. The fine-to-coarse and coarse-to-fine intergrid transfers are realized using, respectively, restriction and interpolation operators with local averaging properties. Note that the starting approximation itself can be obtained in a coarse-to-fine fashion, using nested iterations. Terzopoulos reported, for the case of an expanding Lambertian sphere, a substantial speed-up with respect to the single-scale relaxation (Terzopoulos 1986). It is important to point out that this result applies to an image that contains a unique dominant spatial frequency (related to the sphere diameter). Because in this special case the velocity is parallel to the brightness gradient, the first iteration is already sufficient (in the absence of noise) to recover the correct optical flow. However, the multigrid method turns out to be much less effective for more complex images with superposed frequencies, or even for single frequencies if, as will be shown, a grid coarser than the finest one provides a better estimate. This difficulty has also been encountered by Glazer (1984) and Enkelmann (1988) and is relevant to any homogeneous multiscale scheme, when conflicting information is present at different scales.

An example is given in one dimension by considering two scales with a 2 : 1 resolution and an intensity profile which is the sum of two sine waves of different wavelengths L_1 and L_2. Suppose that, in terms of the fine grid spatial step, $L_1 = 3$, $L_2 = 6$, and the intensity profile velocity is equal to 2 (in the following, for simplicity, Δt is equal to 1). On the coarse grid the higher frequency is almost completely suppressed by the smoothing operation preceding the subsampling process and the measured velocity is equal, according to equation (7), to the true velocity $v = 2$. Figure 2 shows that on the fine scale there is at least a 50% error in the velocity for any combination of the two frequencies. In particular, the measured velocity is equal to 1 for an intensity profile with only the low frequency and it has an opposite sign when the ratio between the high and low frequencies is greater than 0.5. It is worth noting that if $v = 1$, the correct velocity would be recovered at the fine scale.

Typically, the image brightness is a superposition of different frequencies corresponding to the different objects and textures in the scene. Thus, a multiscale scheme *à la multigrid*, involving a bidirectional information flow from high-to-low and low-to-high resolution, is not appropriate because it is likely to mix incoherent information from the different scales. The scheme may not converge or it may converge to an incorrect result.[1]

The previous examples and considerations suggest a new strategy. It starts by estimating the flow field at

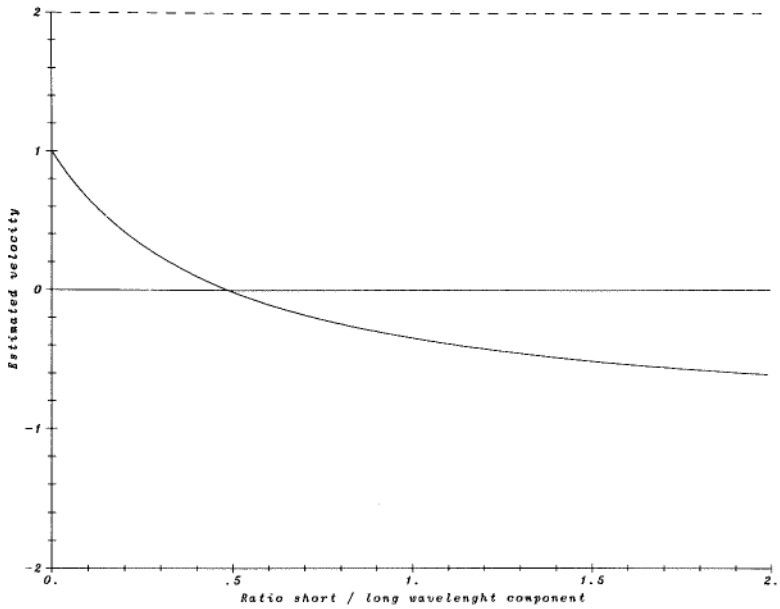

Fig. 2. Measured velocity at the fine scale ($\Delta x = 1$) for two superimposed sinusoidal patterns as a function of the ratio of the amplitude of the short to the amplitude of the long wavelength components. The correct velocity, equal to 2, is recovered at the coarser scale ($\Delta x = 2$); dashed line).

a reasonably coarse scale. This approximation is then improved on successive finer scales only in regions of the image *where its estimated error is greater than a predefined threshold*. We therefore obtain a *local inhomogeneous* approach, where areas of the images characterized by different spatial frequencies or by different motion amplitudes are processed at the appropriate resolutions, avoiding corruption of good estimates by inconsistent information from a different scale.

3 Estimation of the Flow Field Error

Let us now derive an *estimate for the relative error* in the flow field. The effects caused by spatial and temporal quantization (finite-grid step in the derivative estimation formulas) and intensity quantization are discussed separately. The error estimate will be derived in the one-dimensional case and then extended to two dimensions using rotational invariance.

3.1 Error in Derivative Estimation

We first consider the contribution to the flow-field error due to the approximation of the brightness derivatives.

Let $f(x - vt)$ be a one-dimensional translating brightness profile. Taylor's expansion yields the three-point approximation to the first derivative

$$\frac{f(y+h) - f(y-h)}{2h} = f'(y) + \frac{f'''(y)h^2}{6} + O(h^4) \quad (8)$$

In 1-D, the *brightness constancy equation* reduces to $E_x v + E_t = 0$, where $E(x, t) = f(x - vt)$ (see equation (1)). It is easy to show (see appendix) that, neglecting higher-order terms, the three-point approximations of the temporal and spatial derivatives are given by

$$\begin{aligned} \tilde{E}_t &\approx f_t - \frac{vf'''}{6}(v\Delta t)^2 \\ \tilde{E}_x &\approx f_x + \frac{f'''}{6}(\Delta x)^2 \end{aligned} \quad (9)$$

where Δx and Δt are the spatial and temporal sampling steps. By substitution in the brightness constancy equation, we obtain an approximate expression for the measured velocity as a function of the correct velocity

$$\tilde{v} = -\frac{\tilde{E}_t}{\tilde{E}_x} \approx -\frac{f_t - vf'''(v\Delta t)^2/6}{f_x + f'''(\Delta x)^2/6} \quad (10)$$

which leads (by second-order Taylor's expansion) to the relative error

$$\frac{\delta v}{|v|} = \frac{|\tilde{v} - v|}{|v|} \approx \left| \frac{f'''}{6f'} \left((v\Delta t)^2 - (\Delta x)^2 \right) \right| \qquad (11)$$

Thus, provided that higher-order terms can be neglected, the relative error in the flow field due to the three-point approximation of the brightness derivatives is close to zero when the interframe motion $v\Delta t$ is of the order of the spatial sampling step Δx. In particular, for a sinusoidal brightness profile $\sin(2\pi/L)(x - vt)$ of wavelength L, we have

$$\frac{\delta v}{|v|} \approx \frac{2\pi^2}{3L^2} \left| \left((\Delta x)^2 - (v\Delta t)^2 \right) \right| \qquad (12)$$

Notice that the approximation inherent in equation (9) breaks down for large values of Δx and $v\Delta t$.

3.2 Quantization Error

We shall now estimate the flow field relative error due to the quantization of the intensity levels. This provides an upper bound on the relative error due to the noise in the image brightness. Under the assumption that the errors have a Gaussian distribution and therefore

$$\delta v \approx \sqrt{\left(\frac{\partial v}{\partial E_x} \right)^2 \delta E_x^2 + \left(\frac{\partial v}{\partial E_t} \right)^2 \delta E_t^2}$$

and assuming that the one-dimensional constancy equation $v \approx -E_t/E_x$ holds (where v is not null) we arrive at the following expression for the relative error:

$$\frac{\delta v}{|v|} \approx \sqrt{(\delta E_x/E_x)^2 + (\delta E_t/E_t)^2}\,, \qquad (13)$$

where δE_x and δE_t are the errors on the temporal and spatial derivatives respectively. We here assume that the image intensity is an integer going from 0 to a maximum value n, so that the maximum quantization error in the intensity is less than 1. If we consider the errors induced by the quantization process using the three-point estimate of the derivatives, we have

$$\delta E_x \approx \frac{1}{2\Delta x} \quad \text{and} \quad \delta E_t \approx \frac{1}{2\Delta t} \qquad (14)$$

Because we are interested in an upper bound on the error, we have used the maximum error instead of the average discretization error or the standard deviation. These quantities can be calculated using a statistical model for the considered images, as explained by Kamgar-Parsi and Kamgar-Parsi (1989). Because $|v| \approx |E_t/E_x|$, we can rewrite

$$\frac{\delta v}{|v|} \approx \frac{\sqrt{(\delta E_x)^2 + (\delta E_t/v)^2}}{|E_x|} \approx \frac{\sqrt{1/(2\Delta x)^2 + 1/(2v\Delta t)^2}}{|E_x|} \qquad (15)$$

In the following we shall denote the spatial and temporal differences with $\Delta_x E$ and $\Delta_t E$, respectively, that is,

$$\Delta_x E = E(x + \Delta x, t) - E(x - \Delta x, t)$$
$$\Delta_t E = E(x, t + \Delta t) - E(x, t - \Delta t)$$

Now $|E_x| \approx |\Delta_x E/2\Delta x|$ and therefore

$$\frac{\delta v}{|v|} \approx \sqrt{\frac{1}{(\Delta_x E)^2} + \frac{1}{(2vE_x\Delta t)^2}} \qquad (16)$$

Using the constancy equation, we arrive at

$$\frac{\delta v}{|v|} \approx \sqrt{\frac{1}{(\Delta_x E)^2} + \frac{1}{(2E_t\Delta t)^2}} \qquad (17)$$

and because $E_t \approx \Delta_t E/2\Delta t$,

$$\frac{\delta v}{|v|} \approx \sqrt{\frac{1}{(\Delta_x E)^2} + \frac{1}{\Delta_t E)^2}} \qquad (18)$$

3.3 Overall Relative Error

To quantify the overall relative error, we add the error term due to the three-point approximation of the derivatives to the error term caused by the quantization of the image intensity. Thus,

$$\frac{\delta v}{|v|} \approx C(x)|(\Delta_t E)^2 - (\Delta_x E)^2| + \sqrt{\frac{1}{(\Delta_x E)^2} + \frac{1}{(\Delta_t E)^2}} \qquad (19)$$

where the function $C(x)$ depends on the first and third brightness derivatives at the image point x under consideration.

The first term refers to the approximation of the derivatives and can be obtained from equation (11) using the constancy equation and the two basic expressions $E_t \approx \Delta_t E/2\Delta_t$ and $E_x \approx \Delta_x E/2\Delta x$. Because this term does not depend on the number n of brightness quantization levels and since $\Delta_t E$ as well as $\Delta_x E$ are proportional to n, the function $C(x)$ must be proportional to $1/n^2$. This proportionality can be easily shown for a sinusoidal intensity profile. In this case, the first term of equation (19) can be rewritten, according to equation (12), as

$$\frac{2\pi^2(\Delta x)^2}{3L^2}\left[\left(\frac{\Delta_t E}{\Delta_x E}\right)^2 - 1\right] \tag{20}$$

Let us introduce the parameter ρ (*fractional range* of intensity values in a given image), defined by ρ = (maximum-intensity–minimum-intensity)/n. The typical scale for the value of the brightness derivative is given by the range of intensity values of the sinusoid ρn, divided by the wavelength L (i.e., $\Delta_x E/\Delta x \approx \rho n/L$). This relation implies that $(\Delta x/L)^2 \approx (\Delta_x E/\rho n)^2$, which leads—by substitution in equation (20)—to the inverse relation between $C(x)$ and n^2. After completing this substitution, we obtain the following relative error estimate:

$$\frac{\delta v}{|v|} \approx \frac{C}{\rho^2 n^2}|(\Delta_t E)^2 - (\Delta_x E)^2| + \sqrt{\frac{1}{(\Delta_x E)^2} + \frac{1}{(\Delta_t E)^2}} \tag{21}$$

where the value for C is $2\pi^2/3$ as suggested by the above argument. For a general image, the fractional range of the image ρ was estimated using the standard deviation σ in the distribution of intensity values ($\rho = \sigma/n$). Other values of C are possible, for instance derived by considering the average magnitude of the derivative of a sinusoidal pattern, but change little the overall result.

Paradoxically, the first term in equation (21), which explicitly includes n, *does not* depend on n, while the second term, in which n does not appear, *does* depend on n. The "difference" terms (like $\Delta_x E$) grow linearly with the number of discretization levels n, whereas ρ remains constant. Therefore, because C is a constant, the first term in expression (21) will not depend on n, while the second term, which expresses the contribution due to the quantization process, has a $1/n^2$ dependency. Thus, the amplitude of quantization errors can be reduced by increasing the number of quantization levels. It is clearly difficult to determine—and even to estimate—the third derivative of the intensity at every point in the image; but our tests show that, as a working hypothesis, we can consider it as a constant independent of the image position. In practice, we shall use the constant estimated for sinusoidal gratings given in equation (21). Note that approximations are necessary since it is not possible to evaluate the error in the optical flow accurately without knowing precisely the optical flow itself. It is important to point out that the final result in equation (21) presents in a concise way the trade-off between the two kinds of errors introduced.

According to equation (12), there is an optimal scale at which the error contribution from the derivative estimation will be zero. The spatial discretization step should be equal to the interframe motion, that is, $\Delta x = v\Delta t$. The error caused by an incorrect estimate of the image brightness derivatives will increase both above and below this scale. The second error term due to discretization, on the other hand, can be made smaller by going to finer and finer scales. Because the overall error estimate *itself* may become erroneous if very high spatial frequencies are present, the optimal scale for a given pixel is defined as the *coarsest* scale where the relative error is less than a selected threshold. We therefore arrive at a coarse-to-fine approach.

The two-dimensional estimate of the overall relative error is obtained from equation (21) by rotational invariance, substituting $(\Delta_x E)^2$ with the sum of the squared differences in the two dimensions $(\Delta_x E)^2 + (\Delta_y E)^2$. This amounts to measuring the field unreliability according to the error on the component of the velocity that is parallel to the brightness gradient.

4 The Error-Based Adaptive Multiscale Scheme

Preliminary processing consists in building the Gaussian pyramid associated with successive images (Burt 1984). This is a 2 : 1 resolution pyramid using three or four different spatial resolution levels computed from a sequence of three images. The coarser versions of these images are obtained by local averaging using the 5-point mask proposed by Burt (1984). The procedure is then repeated iteratively to construct the low-resolution versions of the three images. For an appropriate choice of the mask, this result closely approximates the convolution of the original images with Gaussian filters (Burt 1984). In the tests that we carried out, the finest scales contained 129 by 129 pixels. The number of layers depends on the image size; in our experiments we usually used three resolution layers. The spatial and temporal derivatives of the brightness are then calculated independently at each level of the pyramid using the three-point approximations (Little & Verri 1989).

Our strategy is based on a coarse-to-fine continuation scheme, and the locally adaptive discretization is implemented using an *inhibition flag* associated with each point in the pyramid. This flag is set whenever the estimated optical flow at the corresponding pixel

is considered sufficiently accurate (with respect to the selected threshold). Once the preliminary processing is terminated, the Horn and Schunck relaxation algorithm described in equations (5) and (6) is applied at the lowest resolution for a selected number of cycles. After the relaxation cycles are completed, the relative error in the flow field—according to equation (21)—is calculated for every pixel at the lowest resolution. This quantity is then used to decide about the local reliability of the optical flow. For every pixel a test is done to see whether the error is below a defined threshold T_{err}. If the test is satisfied, the grid point corresponding to this pixel at the finer resolution in the pyramid and its immediate four neighbors (in the east, west, north, south directions) are inhibited. The optical flow values are then interpolated (with bilinear interpolation) to the next finer scale where they are used as initial approximation for further local relaxations. *Inhibited* pixels will not participate in the relaxation process and will maintain the optical flow values interpolated from coarser resolutions, thereby preventing the corruption of a reliable estimate due to poor derivative approximations at the new scale. This procedure is then repeated iteratively, such that the optical flow is only updated at those locations where no flag is set. If a flag is set, the associated point at the next finest grid and its neighbors will be inhibited. The optimal grid structure for a given image is translated into a pattern of active and *inhibited* grid points in the pyramid, as illustrated in figure 3. The final result of the computation is a reconstruction of the optical flow at the different spatial resolutions, together with the estimate of the relative error.

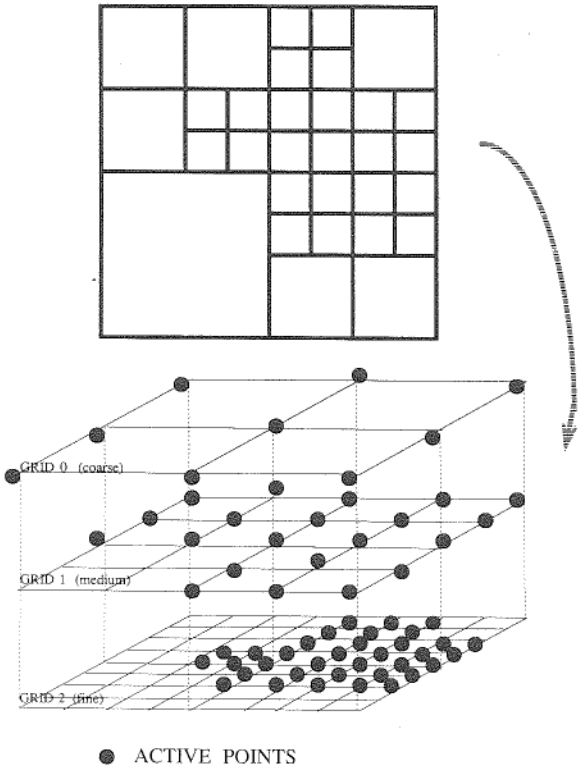

Fig. 3. Adaptive grid and activity pattern in our adaptive multiresolution pyramid. Absence of an active point at any particular location signifies that the value of the optical flow at that particular location is simply interpolated from the corresponding point in the next coarsest level.

5 Experimental Results

To measure in a quantitative way the correctness of the derived optical flows, a series of synthetic images with known velocity fields were generated. We also used natural images, acquired via a video camera, with a measured spatial displacement. Both cases allow us to compare our estimated optical flow field against the correct velocity field.

5.1 Two-Dimensional Sinusoidal Patterns

The generated images show a "plaid" pattern, a superposition of sine waves of different wavelengths in the vertical and horizontal directions. The intensity of a pixel with coordinates (i, j) obeys

$$I(i,j) = \frac{255}{4(1+R)^2} \times \left[1 + R + \sin\left(\frac{2\pi}{L} i\right) + R \sin\left(\frac{2\pi}{l} i\right)\right] \times \left[1 + R + \sin\left(\frac{2\pi}{L} j\right) + R \sin\left(\frac{2\pi}{l} j\right)\right] \tag{22}$$

The relative amount of short versus long wavelength component is determined by the parameter R, the intensity is normalized to obtain values in the range (0–255). The first example illustrates the basic difficulty arising in a multiscale strategy. The parameter R is 1.0, the long and short wavelengths are 7.5 and 3.2.[2] The resulting image is displayed in figure 4. This pattern was moved in the plane in the north-east direction, with

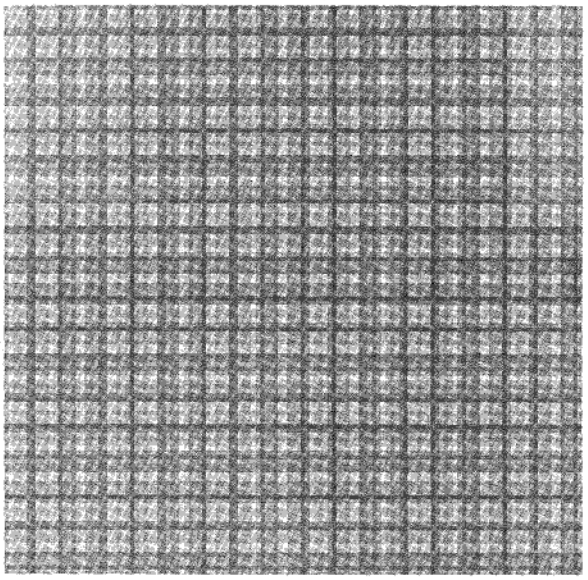

Fig. 4. The two-dimensional plaid pattern with long ($L = 7.5$) and short ($L = 3.2$) wavelengths (see equation 22).

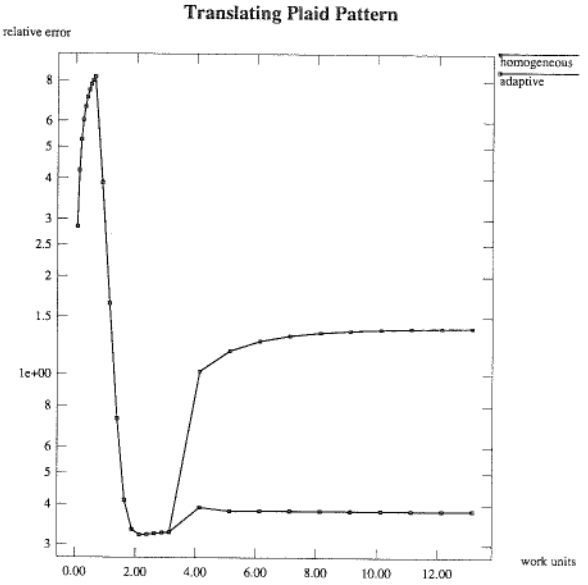

Fig. 5. Comparison of homogeneous versus adaptive multiscale strategy using the translating "plaid" pattern of figure 4. Graphs show the relative error in the optical flow as a function of *work units* (a *work unit* is defined as the amount of computation used for a complete relaxation at the finest scale). Line (a): multiscale homogeneous strategy (with no adaptation; top curve). Line (b): multiscale adaptive strategy (curve at bottom). The adaptive algorithm "freezes" the result at the intermediate grid because the error measure is below the threshold T_{err} and interpolates to the finest grid. Notice the logarithmic ordinate.

a velocity equal to (1, 2). Comparison of the results of the homogeneous versus the adaptive coarse-to-fine strategy are shown in figure 5. Ten iterations are carried out on every discretization grid, bilinear interpolation of results is applied before relaxation is initiated on a finer grid.

Relaxation on the coarsest grid produces an optical field whose difference with the correct motion flow *increases* as a function of the iteration number. This is caused by large quantization errors on this grid (the intensity is almost constant and discretization errors are large). The situation is better on the intermediate grid. In spite of incorrect initial values obtained from the coarser grid, the error is rapidly reduced after the first relaxations. Error in derivative estimation reaches in this case the minimum value (motion on this scale is less than the dominant wavelength).

After interpolation to the finest grid, the homogeneous scheme continues the relaxation process, driving the result to a *worse* solution. This is caused by bad derivative estimation (motion on this scale is not small in comparison with the shorter wavelength). On the contrary, the adaptive scheme recognizes that the error on the intermediate scale is less than the given threshold T_{err} (0.4 in this case) in most of the corresponding image pixels, so that no further updating of the optical flow at this particular location at finer scales is necessary. Thus, relaxation at the next finer scale is inhibited and the error in the final optical flow is similar to that on the middle scale.

The difference in the qualitative structure of the derived optical flow can be appreciated in figure 6. Finally, figure 7 shows a display of the *estimated* error (according to equation (19)) on the different scales. The quantization error is largest at the coarser scale, while the derivative estimation error is largest at the finest scale. The total error reaches the minimum on the middle scale. Furthermore, for the range of images we have considered, our adaptive coarse-to-fine continuation method shares with the standard multigrid algorithm its speed. Using our coarse-to-fine strategy reduced the computational time for our images by a factor of 50–100, when the latter converged to the correct solution.

5.2 *Tuning Curve for Natural Images*

The images used for this test show a pine cone moving in the upward direction (figure 8). They were acquired

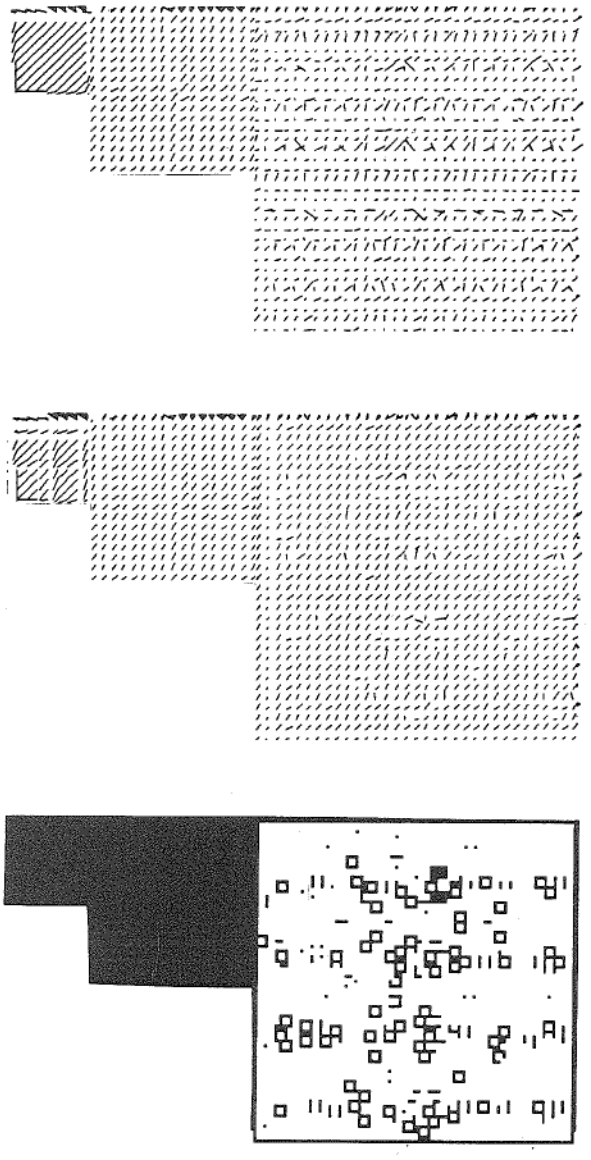

Fig. 6. Reconstructed optical flow for the translating "plaid" pattern of figure 4. (a) homogeneous multiscale strategy; (b) adaptive multiscale strategy; (c) active (black) and inhibited (white) points. Only locations at the finest grid are inhibited.

with a S-VHS video camera and a Targa frame grabber. Movement was executed by adjusting a tripod sustaining the object by 0.25 cm every frame. Measured velocity in pixels is 1.6 pixel / frame. Tests were carried out for sets of three images taken every one, two, and three frames. Thus, the effective velocity was 1.6, 3.2, and 4.8 pixels per frame. The average velocity (on a window centered on the pine cone) obtained with the homogeneous multiscale algorithm is compared with

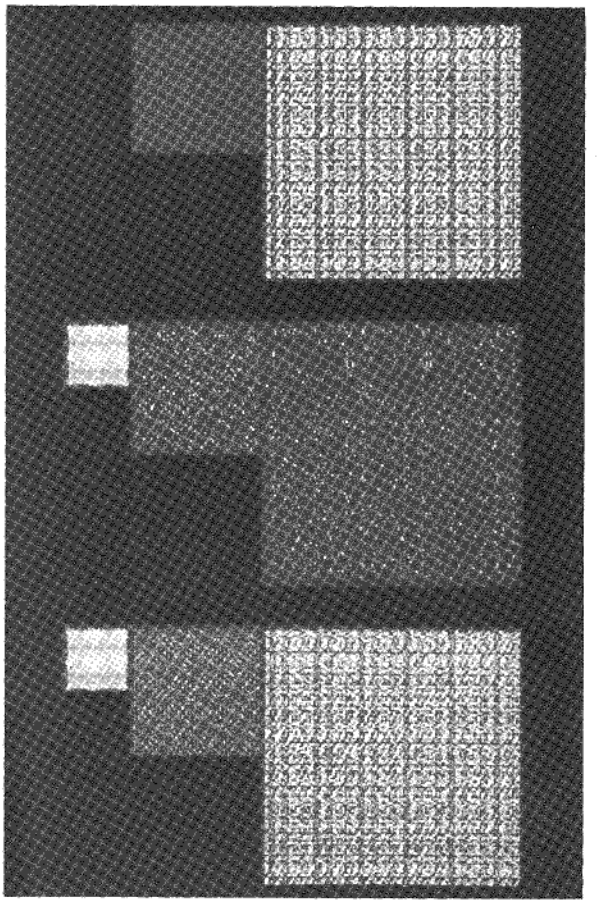

Fig. 7. Estimated error on different scales for the "plaid" pattern of figure 4. The intensity value is proportional to the error. Errors in the derivative estimation (top panel), quantization errors (middle panel), and the total error (sum of the two; bottom panel) are shown. Errors in estimating the spatial and temporal derivatives are minimized at the coarsest scale by low-pass filtering. However, the large grid spacing at this scale causes relatively large quantization errors. The total error is minimized at the intermediate grid.

that obtained with the adaptive version. While this second version always produces a better estimate, the difference is particularly significant for large-motion amplitudes, as shown in figure 8. In this case the fine-scale derivative information is completely erroneous. For the large-amplitude motion this, in fact, leads to motion reversal where the algorithm signals motion in the direction opposite to the direction of true motion. This is recognized by the adaptive scheme that *freezes* the solution obtained at coarser grids, producing a much better final estimate. We used our adaptive algorithms with both 3 and 4 grids; thus, in the latter case, the image was being analyzed on grids ranging from 129 × 129 pixels to 17 × 17 pixels. It is obvious that

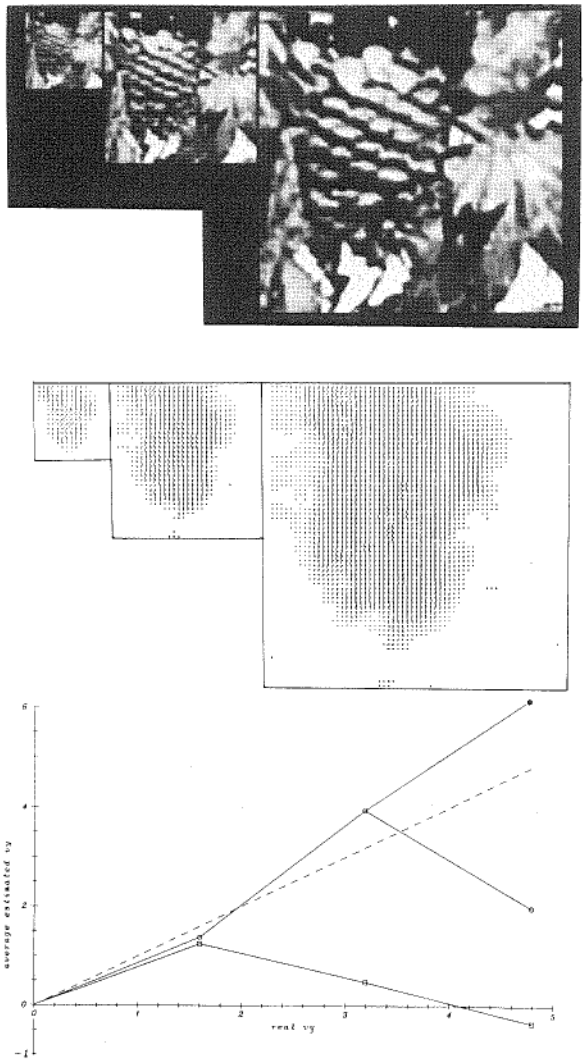

Fig. 8. Optical flow obtained using a mechanical setup. The pine cone (top panel) was translated upward by 1.6, 3.2, and 4.8 pixels per frame. The resultant optical flow field at the three different scales is shown in the middle panel. The average estimated velocity in the central area of the pine cone, obtained using the homogeneous method (open squares) and using our adaptive multiscale method with three (open circles) or four grids (open circle in upper branch), is plotted against the true velocity. It is clear that the adaptive multiscale method does substantially better than the homogeneous method, effectively employing only a single scale.

larger-motion amplitudes require corresponding coarser and coarser grids for optimal performance. The correct velocity corresponds to the dotted line.

6 Discussion

Our motivation for this research was to study how the velocity range over which motion algorithms give a reliable estimate of the optical flow can be expanded. Both intensity and token-based motion algorithms only estimate the magnitude of the optical flow correctly within a range of velocities, a range dictated by the spatial and temporal discretization steps used. If, for instance, the interframe motion, $v\Delta t$, is much larger than the spatial discretization step, the estimate of the spatial and temporal derivatives will be incorrect and the computed velocity will be very different from its correct value (see equation (7)). While the standard multigrid algorithm—see (Enkelmann 1988; Glazer 1984; Terzopoulos 1986; Brandt 1977)—speeds up the convergence times of gradient-based motion algorithms by orders of magnitude, it does not improve the situation from the point of view of velocity range. While inspection of figure 5 illustrates that the relative error in the magnitude of the velocity is minimal at some intermediate grid size, the error increases for the standard, homogeneous multigrid algorithm at the finer grid size. In effect, the algorithm is limited by the spatial step size of the finest grid used.

We here provided a substantial improvement by locally computing an estimate of the relative error in the optical flow. This error is simply the sum of an error term due to the estimation of the velocity using the ratio of the temporal to the spatial intensity gradient and an error term due to the quantization of the image. The error estimate is simple to compute, being a function of the square of the temporal and spatial differences of the image intensity—see equation (21)—and does not depend on any particular assumption about texture or patterns. It contains one free parameter, C, governing the relative importance of the two error terms. We adopt a value of C on the basis of the heuristic sinusoidal-wave assumption. The flow field is computed at the coarse scale using Horn and Schunck's method and is improved at successive finer scales in those areas of the image where the relative error estimate is greater than a predefined threshold. We did not experiment with mixed coarse-to-fine-to-coarse strategies, having found that a single coarse-to-fine sweep leads to excellent results while still preserving the order of magnitude speedup experienced with multigrid algorithms (Enkelmann 1988; Glazer 1984; Terzopoulos 1986). Our adaptive scheme is parallel at any given scale; and deals with multiple motions and/or multiple patterns and textures across the visual scene. It leads to a substantially improved estimate of the optical flow and results in a substantial speedup with respect to the standard Horn and Schunck single-scale algorithm (almost two orders of magnitude for our images).

An important issue is the optimal number and size of the spatial grids used. Specifically, the coarsest spatial grid used should be matched to the maximum expected motion amplitude. Furthermore, the factor α by which the spatial sampling size Δx is reduced when going from one scale to the next coarsest, influences how close the algorithm approximates the optimal grid which minimizes the error expression equation (21). We here always use $\alpha = 2$; however, a factor of $\sqrt{\alpha}$ (obtained by rotating each grid by 45°) will most likely lead to better results.

We do not discuss here the use of analog or binary line discontinuities, which greatly improve the final optical flow, particularly in the presence of multiple, independently moving objects (Hutchinson et al. 1988; Poggio et al. 1988; Harris et al. 1990). It is straightforward to implement motion discontinuities within our multiscale framework—see, in particular, (Battiti 1989 and 1990).

Kearney et al. (1984) previously proposed an "iterative registration technique," for optimizing gradient-based optical-flow algorithms using up to four different resolution grids. Their method proceeds by computing a first estimate of the optical flow at any particular location. This flow estimate is then used to register the frame pair of each successive iteration of the estimation procedure. As the optical flow usually differs across any image, this procedure must be repeated at every image point. Our method, on the other hand, is fully parallel and involves a much simpler, one-step, operation; that is, computing a simple scalar function of the square of the temporal and spatial intensity differences, $\Delta_x E$ and $\Delta_t E$.

We have limited the discussion here to the gradient method of estimating optical flow. Recently proposed methods, such as (Uras et al. 1988 and Yuille & Grzywacz 1988) share the same derivative estimation and noise problems we have described. However, as the relative error in equation (21) does not depend on the actual computation of the flow estimate, but only on directly observable image properties, correlation-based optical-flow methods as well as the mathematically equivalent spatial-temporal energy methods will also profit from a similar adaptive multiscale approach.

Acknowledgments

This work was carried out while R. Battiti and E. Amaldi were in the laboratory of Prof. Geoffrey Fox at Caltech. We gratefully acknowledge his kind support. John Harris provided detailed comments, in particular for figure 1. G. Fox is partly funded by DOE grant DE-FG-03-85ER25009, NSF grant SIT-8700064 and by IBM. C. Koch is partly supported by the National Science Foundation, the Office of Naval Research, and the James S. McDonnell Foundation.

References

Adelson, E.H., and Bergen, J.R. 1985. Spatio-temporal energy models for the perception of motion. *J. Opt. Soc. Amer.* A 2: 284–299.

Battiti, R. 1989. Surface reconstruction and discontinuity detection: a fast hierarchical approach on a two-dimensional mesh. *Proc. 4th Conf. Hypercube Concurrent Computers and Applications*, Monterey, CA.

Battiti, R. 1990. Multiscale methods, parallel computation and neural networks for real-time computer vision. Ph.D. Dissertation, California Institute of Technology.

Brandt, A. 1977. Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computations* 31: 333–390.

Bülthoff, H.H., Little, J.J., and Poggio, T. 1989. Parallel computation of motion: computation, psychophysics and physiology. *Nature* 337: 549–553.

Burt, P.J. 1984. The pyramid as a structure for efficient computation. In *Multiresolution Image Processing and Analysis*, Rosenfeld, A. (ed.), Springer-Verlag, pp. 6–35.

Enkelmann, W. 1988. Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences. *Comput. Vision, Graph. Image Process.* 43: 150–177.

Fennema, C.L. and Thompson, W.B. 1979. Velocity determination in scenes containing several moving objects. *Comput. Graph. Image Process.* 9: 301–315.

Geman, S., and Geman D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI-6: 721–741.

Girosi, F., Verri, A., and Torre, V. 1989. Constraints for the computation of the optical flow. *Proc. IEEE Workshop on Visual Motion*, Irvine, CA, March, pp. 116–124.

Glazer, F. 1984. Multilevel relaxation in low-level computer vision. In *Multiresolution Image Processing and Analysis*, A. Rosenfeld (ed.). Springer-Verlag, pp. 312–330.

Harris, J., Koch, C., Staats, E., and Luo, J. 1990. Analog hardware for detecting discontinuities in early vision. *Intern. J. Comput. Vision* 4: 211–223.

Hassenstein, B., and Reichardt, W. 1956. Systemtheoretische Analyse der Zeit, Reihenfolgen, und Vorzeichenauswertung bei der Bewegungsperzepion des Rüsselkäfers *Chlorophanus*. *Z. Naturforsch.* 11b: 513–524.

Hildreth, E.C. 1984. Computations underlying the measurement of visual motion. *Artificial Intelligence* 23: 309–354.

Hildreth, E., and Koch, C. 1987. The analysis of visual motion: from computational theory to neuronal mechanisms. *Annu. Rev. Neurosci.* 10: 477–533.

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

Hutchinson, J., Koch, C., Luo, J., and Mead, C. 1988. Computing motion using analog and binary resistive networks. *IEEE Computer* 21: 52–61.

Kamgar-Parsi, B., and Kamgar-Parsi, B. 1989. Evaluation of quantization error in computer vision. *IEEE Trans Patt. Anal. Mach. Intell.* PAMI-11: 929–940.

Kearney, J.K., Thompson, W.B., and Boley, D.L. 1984. Optical flow estimation: an error analysis of gradient-based methods with local optimization. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI-9: 229–244.

Koch, C., Wang, H.T., and Mathur, B. 1989. Computing motion in the primate visual system. *J. Exper. Biol.* 146: 115–139.

Little, J., and Verri, A. 1989. Analysis of differential and matching methods for optical flow. *Proc. IEEE Workshop Visual Motion*, Irvine CA, March, pp. 173–180.

Marr, D., and Ullman, S. 1981. Directional selectivity and its use in early visual processing. *Proc. Roy. Soc. London* B 211: 151–180.

Marroquin, J. 1984. Surface reconstruction preserving discontinuities. *M.I.T. Artif. Intell. Lab. Memo* **792**, MIT: Cambridge, MA.

Maunsell, J.H.R., and Van Essen, D.C. 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed and orientation. *J. Neurophysiol.* 49: 1127–1147.

Nagel, H.H. 1978. Analysis techniques for image sequences. *Proc. 4th Intern. Joint Conf. Patt. Recog.*, Kyoto, Japan, November.

Nagel, H.H., and Enkelmann, W. 1986. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI-8 (5): 565–593.

Poggio, T., and Reichardt, W., 1973. Considerations on models of movement detection. *Kybernetik* 13: 223–227.

Poggio, T., Gamble, E.B., and Little, J.J. 1988. Parallel integration of vision modules. *Science* 242: 436–440.

Poggio, T., Torre, V., and Koch, C. 1985. Computational vision and regularization theory. *Nature* 317: 314–419.

Reichardt, W., Schlögel, R.W., and Egelhaaf, M. 1988. Movement detectors of the correlation type provide sufficient information for local computation of 2-D velocity field. *Naturwissenschaften* 75: 313–315.

Terzopoulos, D. 1986. Image analysis using multigrid relaxation methods. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI-8: 129–139.

Ullman, S. 1981. Analysis of visual motion by biological and computer systems. *IEEE Computer*, 14: 57–69.

Uras, S., Girosi, F., Verri, A., and Torre, V. 1988. A computational approach to motion perception. *Biological Cybernetics* 60: 79–87.

van Santen, J.P.H., and Sperling, G. 1984. A temporal covariance model of motion perception. *J. Opt. Soc. Amer.* A 1:451–473.

Verri, A., and Poggio, T. 1989. Motion field and optical flow: qualitative properties. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI-11: 490–498.

Wang, H.T., Mathur, B., and Koch, C. 1989. Computing optical flow in the primate visual system. *Neural Computation*, 1: 92–103.

Wang, H.T., Mathur, B., and Koch, C. 1991. A multiscale adaptive network model of motion computation in primates. In: *Advances in Neural Information Processing Systems*, Touretzky, D.S. and Lippman, R., eds., Morgan Kaufmann, San Mateo, in press.

Watson, A.B., and Ahumada, A.J. 1985. Model of human visual-motion sensing. *J. Opt. Soc. Amer.* A 2:322–341.

Yuille, A.L., and Grzywacz, N.M. 1988. A computational theory for the perception of coherent visual motion. *Nature* 333: 71–73.

Notes

1. It may oscillate between two different grids with conflicting information, for example.
2. These represent "generic" wavelengths (not multiples of the grid step to avoid particular effects), chosen to give different "dominant" components at different scales.

Appendix: Three Point Approximation of Derivatives

We shall derive the third-order expressions for the three-point approximations of the temporal and spatial brightness derivatives. Let $f(x - vt)$ be a one-dimensional translating brightness profile. Taylor's expansion provide the three-point formula for the first-order brightness derivatives:

$$\frac{f(y + h) - f(y - h)}{2h} = f'(y) + \frac{f'''(y)h^2}{6} + O(h^4)$$

The approximation of the temporal derivative is given by

$$\tilde{E}_t = \frac{f(x - v(t + \Delta t)) - f(x - v(t - \Delta t))}{2\Delta t}$$

which becomes, by setting $y = x - vt$,

$$\tilde{E}_t = \frac{f(y - v\Delta t) - f(y + v\Delta t)}{2\Delta t} n$$

which can be transformed, using the above Taylor expansion, to

$$-vf' - \frac{vf'''(v\Delta t)^2}{6} + O((v\Delta t)^4)$$

Since $f_t = f'(x - vt)(-v) = -vf'$, we arrive at

$$\tilde{E}_t = f_t - \frac{vf'''}{6}(v\Delta t)^2$$

where the higher-order terms are neglected. A similar expression holds for the approximation of the spatial derivative

$$\tilde{E}_x = f_x + \frac{f'''}{6}(\Delta x)^2$$

where Δx is the spatial sampling step.