The Computation of Optical Flow

S.S. Beauchemin and J.L. Barron
beau@csd.uwo.ca barron@csd.uwo.ca

Dept. of Computer Science
University of Western Ontario
London, Ontario, Canada, N6A 5B7

Abstract

Two-dimensional image motion is the projection of the three-dimensional motion of objects, relative to a visual sensor, onto its image plane. Sequences of time-ordered images allow the estimation of projected two-dimensional image motion as either instantaneous image velocities or discrete image displacements. These are usually called the *optical flow field* or the *image velocity field*. Provided that optical flow is a reliable approximation to two-dimensional image motion, it may then be used to recover the three-dimensional motion of the visual sensor (to within a scale factor) and the three-dimensional surface structure (shape or relative depth) through assumptions concerning the structure of the optical flow field, the three-dimensional environment and the motion of the sensor. Optical flow may also be used to perform motion detection, object segmentation, time-to-collision and focus of expansion calculations, motion compensated encoding and stereo disparity measurement. We investigate the computation of optical flow in this survey: widely known methods for estimating optical flow are classified and examined by scrutinizing the hypotheses and assumptions they use. The survey concludes with a discussion of current research issues.

Contents

List of Figures

1 Introduction

A fundamental problem in processing sequences of images is the computation of *optical flow*, an approximation to image motion defined as the projection of velocities of 3-d surface points onto the imaging plane of a visual sensor. Optical flow is often a convenient and useful image motion representation. However, there exist other motion descriptors, sometimes more general than optical flow, such as parametric models of motion, or descriptors adapted to restricted contexts, such as when elements of the geometry of the scene or the motion of the visual sensor are partially or completely pre-determined.

The importance of motion in visual processing cannot be understated: approximations to image motion may be used to estimate 3-d scene properties and motion parameters from a moving visual sensor [64, 91, 93, 121, 150, 149, 3, 16, 113, 66, 163, 165, 99, 52, 46, 73], to perform motion segmentation [24, 116, 125, 104, 138, 42, 76, 29, 9, 126, 45], to compute the focus of expansion and time-to-collision [124, 118, 141, 75, 143, 30], to perform motion-compensated image encoding [35, 41, 101, 105, 115, 164], to compute stereo disparity [12, 39, 77, 85], to measure blood flow and heart-wall motion in medical imagery [123] and, recently, to measure minute amounts of growth in corn seedlings [17, 87].

1.1 Motion and Structure Paradigms

Traditionally, approximations to image motion have been used to infer egomotion and scene structure. Towards this end, different motion and structure paradigms have been developed, sometimes using optical flow as an intermediate representation of motion, correspondences between image features, correlations or properties of intensity structures. These paradigms are generally classified into three main groups:

Velocity Three-dimensional motion and scene structure may be inferred from two-dimensional velocity fields [64, 93, 121] by relating the motion and structure parameters to optical flow. These parameters include instantaneous translation and rotation rates and possibly surface parameters or relative depth[1].

[1]Figure 1.1 shows one frame from the synthetic *Yosemite fly-through* sequence, produced by Lynn Quam at SRI.

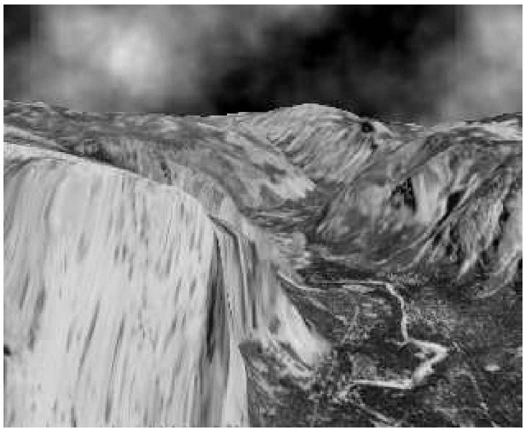
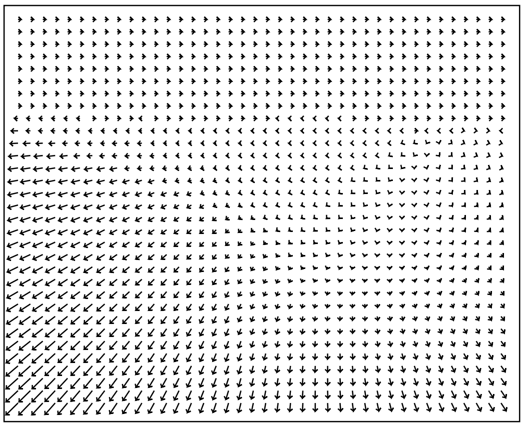

Figure 1.1: *a) (right): A frame of the Yosemite fly-through image sequence generated by Lynn Quam at SRI and b) (left): its optical flow.*

Disparity Image disparities, either established as image feature correspondences or local correlations, may be used to compute three-dimensional translation vectors, rotation matrices and surface attributes [91, 150].

Intensity Image intensities and their derivatives are sometimes used directly to obtain motion and structure parameters [6, 7, 71, 112, 67, 166], thus avoiding an explicit intermediate representation of image motion such as optical flow or disparity fields.

Usually, relating image motion estimates or intensity derivatives to three-dimensional motion and structure parameters results in sets of non-linear equations. In addition, each of these paradigms has its merits and detractions, depending on the intended use and the characteristics of the imagery. However, their evaluation is beyond the scope of this survey.

1.2 Optical Flow

The initial hypothesis in measuring image motion is that the intensity structures of local time-varying image regions are approximately constant under motion for at least a short duration

[70]. Formally, if $I(\mathbf{x},t)$ is the image intensity function, then

$$I(\mathbf{x},t) \approx I(\mathbf{x}+\delta\mathbf{x}, t+\delta t), \tag{1.1}$$

where $\delta\mathbf{x}$ is the displacement of the local image region at $(\mathbf{x},t)$ after time δt. Expanding the left-hand side of this equation in a Taylor series yields

$$I(\mathbf{x},t) = I(\mathbf{x},t) + \nabla I \cdot \delta\mathbf{x} + \delta t I_t + O^2, \tag{1.2}$$

where $\nabla I = (I_x, I_y)$ and I_t are the 1^{st}-order partial derivatives of $I(\mathbf{x},t)$, and O^2, the 2^{nd} and higher order terms, which are assumed negligible. Subtracting $I(\mathbf{x},t)$ on both sides, ignoring O^2 and dividing by δt yields

$$\nabla I \cdot \mathbf{v} + I_t = 0, \tag{1.3}$$

where $\nabla I = (I_x, I_y)$ is the spatial intensity gradient and $\mathbf{v} = (u,v)$ is the image velocity[2]. Equation (1.3) is known as the *optical flow constraint equation*, and defines a single local constraint on image motion (see Figure 1.2). However, this constraint is not sufficient to compute both components of $\mathbf{v}$ as the optical flow constraint equation is ill-posed[3]. That is to say, only $\mathbf{v}_\perp$, the motion component in the direction of the local gradient of the image intensity function, may be estimated. This phenomenon is known as the *aperture problem* [151] and only at image locations where there is sufficient intensity structure (or Gaussian curvature) can the motion be fully estimated with the use of the optical flow constraint equation (see Figure 1.3). For example, the velocity of a surface that is homogeneous or containing texture with a single orientation cannot be recovered optically. Since the normal velocity is in the direction of the spatial gradient ∇I, (1.3) allows one to write

$$\mathbf{v}_\perp = \frac{-I_t \nabla I}{||\nabla I||_2^2}. \tag{1.4}$$

Thus, the measurement of spatiotemporal derivatives allows the recovery of normal image velocity.

From this definition, it becomes clear that for optical flow to be exactly image motion, a number of conditions have to be satisfied. These are: *a) uniform illumination; b) lambertian surface reflectance* and *c) pure translation parallel to the image plane.* Realistically, these conditions are never entirely satisfied in scenery. Instead, it is assumed that these conditions hold

[2]The row convention for vectors is used, thus $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}\mathbf{y}^T$ represents inner product.

[3]The optical flow constraint equation is one linear equation in the two unknowns $\mathbf{v} = (u,v)$.

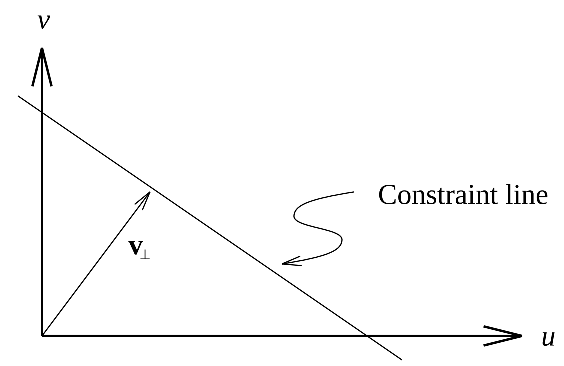

Figure 1.2: *The optical flow constraint equation defines a line in velocity space. The normal velocity* $\mathbf{v}_\perp$ *is defined as the vector perpendicular to the constraint line, i.e. the velocity with the smallest magnitude on the optical flow constraint line.*

locally in the scene and therefore locally on the image plane. The degree to which these conditions are satisfied partly determines the accuracy with which optical flow approximates image motion. Alternatively, one can measure the displacement of small image patches, for example by correlation, in short image sequences (usually 2 or 3 frames). Such *image displacements* constitute a valuable approximation to image velocity when certain conditions are met. In particular, the ratio of sensor translational speed to absolute environmental depth, the 3-d vertical and horizontal sensor rotations and the time interval between frames must be small quantities [3]. Optical flow may also be computed as the *disparity field* where, given two stereo images or two adjacent images in some sequence, features of interest in the images are extracted and matched via a correspondence process.

Essentially, performing 2-d motion detection involves the processing of scenes where the sensor is moving within an environment containing both stationary and nonstationary objects. Furthermore, visual events such as occlusion, transparent motions and non-rigid objects increase the inherent complexity of the measurement of optical flow.

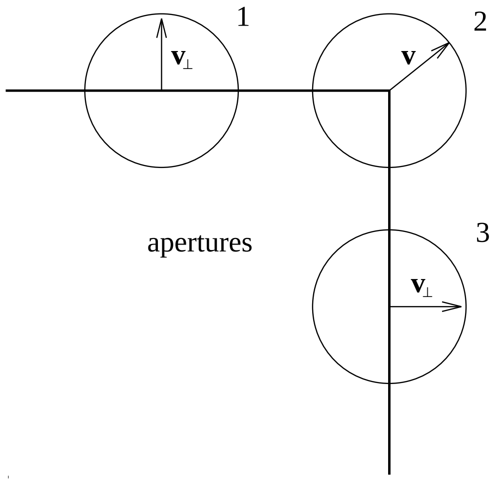

Figure 1.3: *Through apertures 1 and 3, only normal motions of the edges forming the square can be estimated, due to a lack of local structure. Inside aperture 2, at the corner point, the motion can be fully measured since there is sufficient local structure; both normal motions are visible.*

1.3 Hierarchical Processing

Traditionally, optical flow was computed using only one scale of resolution, usually defined by the visual sensor [70], leading to the problem of measuring large image motions. In this case, because of low sampling rates and aliasing effects, (1.3) becomes inappropriate. A general way of circumventing this problem is to apply optical flow techniques in a hierarchical, coarse-to-fine framework. Hierarchical frameworks allow the images to be decomposed in different scales of resolution in the form of Gaussian or Laplacian pyramids [8, 18, 44, 53]. Because of a low frequency representation at coarser resolutions, the optical flow constraint equation becomes applicable in the case of large image motions at the finest resolution [80]. In addition to handling fast motions, hierarchical processing also offers increased computational efficiency. In such frameworks, velocity or displacement estimates are cascaded through each resolution level as initial estimates subject to refinement. At the coarsest level, initial estimates are computed and then projected onto a finer level of resolution and refined once again. The final estimates are obtained when the refinement reaches the finest level of resolution (see Figure 1.4). Hierarchical processing is applicable to most optical flow techniques. For example, Glazer [54] adapted Horn and Schunck's differential technique to such a framework, Anandan [8] used a hierarchical area-based correlation method, Heeger [65] proposed a hierarchical energy-based filtering technique in a Gaussian pyramid and Bergen *et al.* [20, 21] proposed hierarchical parametric models for optical flow.

1.4 Problems and Issues

Much progress has been made in optical flow computation and yet, its accurate estimation remains difficult because of numerous theoretical and practical reasons. Theoretically, we believe that optical flow, as an approximation to image motion, largely determines the lower bound on accuracy. In addition, scene properties such as surface reflectance and informative image events such as transparency and occlusion were, until recently, not adequately dealt with in most models of image motion.

Optical flow and image motion The interpretation of intensity variation as pure relative motion is restrictive because velocity is a geometric quantity independent from illumination conditions. Hence, estimating optical flow from intensity variation only approximates image motion. Conditions which make optical flow different from image motion include

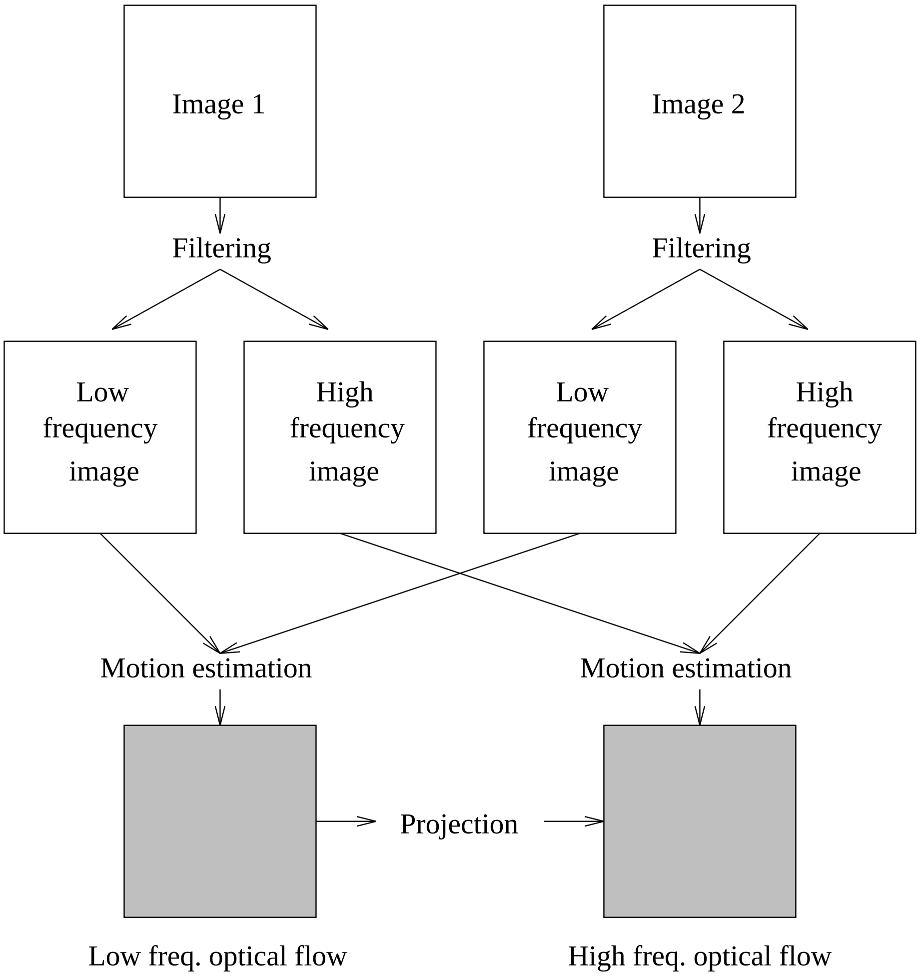

Figure 1.4: *The hierarchical computational model. Coarse motion estimates are projected to finer levels where they serve as initial estimates subject to refinement [8].*

the absence of texture, in which case optical flow is null, and when the true motion field violates the brightness consistency model used for its approximation [69]. Uniform scene illumination and Lambertian surface reflectance are either explicitly or implicitly assumed in most current optical flow methods which use some form of the brightness consistency assumption. Highlights, shadows, variable illumination and surface translucency are phenomena violating the assumption and have only been studied to a limited extent [102, 21, 48, 78].

Occluding surfaces and independently moving objects The problem posed by occluding surfaces is currently being addressed by the research community. Occlusion is difficult to analyze, despite the fact that occlusion constitutes an important source of visual information: optical flow at occlusion boundaries may be used to determine the direction of translation [93] and segment the scene into independently moving surfaces [3, 146, 153]. Until recently, most optical flow techniques relied on a single-surface hypothesis [70], which is a rare visual event. The difficulty of handling occlusion lies in the fact that image surfaces may appear or disappear in time, misleading tracking processes and causing numerical artifacts in intensity derivatives.

Transparency Transparent motions created by physically translucent surfaces is also found in imagery. The problem posed by transparent motions is mainly one of handling multiple motion distributions. Classical approaches to optical flow measurement which use single motion models are clearly inadequate [14]. Recently, mixed distributions and superposition principles have been applied to transparent motions [21, 78, 132].

Practical issues in computing optical flow were addressed in a recent study [14, 15] which analyzed nine techniques dating from 1981 to 1990 [70, 106, 108, 152, 95, 48, 47, 65, 8, 134, 156] for accuracy, density and reliability of measurements. To test the implementations of these algorithms, both synthetic and real data were used. The observed performance of these algorithms lead to the following conclusions:

Prefiltering and differentiation Temporal smoothing is required in order to avoid aliasing and numerical differentiation must be done carefully. The often stated requirement that differential methods require image intensity be nearly linear with motions less than one spatial unit per frame arises from the use of only two frames, poor numerical differentiation or input imagery corrupted by temporal aliasing. With two frames, derivatives are estimated using simple backward differences which are accurate only when the input is highly oversampled and the intensity structure is nearly linear. When temporal aliasing cannot be avoided, hierarchical methods, operating in a coarse-to-fine manner, provide better results.

Reliability measures The need for confidence measures to indicate the reliability of computed velocities cannot be understated. These confidence measures can be used to threshold optical flow fields or to weight velocities is post-measurement processing (in a motion and structure calculation, for example). Most current differential methods do not provide confidence measures. However, in Barron *et al.*'s study [14], the smallest eigenvalue of a least-squares matrix [133] was used successfully. Other possibilities, including the determinant of a Hessian matrix (Gaussian curvature) [156], the condition number of a solution matrix [48, 152] the magnitude of local image gradients, the principle curvature values [8] and the eigenvalues of a covariance matrix [136] were examined [14].

Accuracy Hierarchical correlation methods constitute robust motion measurement schemes for image sequences with significant contrast changes or large displacements and severe aliasing[4]. The test image sequences used by Barron *et al.* [14] are all appropriately sampled with small motions (typically between one and four pixels per frame) and were favourable to differential approaches. In spite of this, and as opposed to differential-based test results, their experiments demonstrate that correlation methods experience difficulty with subpixel motions as their error depends on the closeness of image motion to an integer number of pixels. Hierarchical differential-based methods (using image warping or registration) may provide an alternative to correlation methods.

One of the purposes of this study [14] was to analyze the performance of different optical flow methods and to and to encourage others to compare numerical results with theirs. Towards this end, several authors now compare the performance of their techniques with those of this study for the same image sequences [28, 59, 157, 90, 27, 162, 58, 50]. In addition, some experimental work evaluating differential techniques has recently appeared [60]. Unfortunately, a quantitative analysis is often impossible for real image data (to obtain the correct optical flow, one needs the 3-d motion parameters as well as the 3-d depth values everywhere). In this

[4]Dutta *et al.*'s *stop-and-shoot* sequences constitute interesting image sequence examples [43].

case, only a qualitative analysis may be performed, but it was observed that some optical flow fields, while being less accurate quantitatively, may appear better qualitatively, such as those obtained with methods incorporating global smoothing constraints. An obvious way to evaluate optical flow computations and yet avoid a quantitative analysis is to use the computed optical flow field in a motion and structure calculation and examine the accuracy of the 3-d motion parameters. DeMicheli *et al.* [99] used optical flow fields obtained with the method of Uras *et al.* [152] to estimate time-to-collision and angular velocity in a Kalman filter framework with good accuracy. More recently, Barron and Eagleson [13] have proposed a motion and structure algorithm to compute general 1^{st} and 2^{nd}-order 3-d motion and structure parameters from time varying optical flow, also in a Kalman filter framework.

1.5 Scope and Purpose

There exist numerous computational models for estimating image velocity, which we classify into the following main groups: *intensity-based differential methods* [93, 70, 95, 148, 44, 54, 53, 107, 108, 109, 152, 5, 147, 127, 128, 133, 137, 23, 21, 50], *frequency-based filtering methods* [48, 56, 65, 154] and *correlation-based methods* [8, 12, 79, 84, 131, 134, 144]. In addition, there exist methods for the computation of discontinuous or multiple-valued optical flow and techniques for performing temporal refinements of motion estimates as more information becomes available through the image-acquisition process. These methods are classified into the following groups: *multiple motion methods* and *temporal refinement methods*.

Most of these approaches can be understood as being comprised of three conceptual stages of processing: *prefiltering* (low-pass or band-pass) in order to extract signal structures of interest and to enhance the signal-to-noise ratio, *measurement extraction* of the basic image structures, such as spatiotemporal derivatives or local correlation surfaces and *measurement integration* either by regularization, correlation, or a least-squares computation. These approaches are thought to be broadly equivalent [1, 2] although differences in implementation can lead to significant differences in performance. Given this particular classification, this survey covers the optical flow techniques that do not require solving the correspondence problem. Hence, areas which are not covered by this survey are feature-based matching methods involving the correspondence problem and stereo approaches to image motion.

One of the most fundamental uses for optical flow is the computation of 3-d motion and structure. Typically, these reconstruction algorithms are *ill-conditioned*[5] and the accuracy of optical flow becomes of extreme importance. Achieving more accurate optical flow calculations requires not only careful attention to details, but also that realistic imaging properties be taken into account. In this survey, we examine both older and newer approaches to optical flow, with particular attention devoted to how the newer approaches address the accuracy, density and reliability issues raised by Barron *et al.* [14]. A recent survey [4] shows the current state-of-the-art up to 1988 not only for optical flow algorithms, but also for feature-based motion algorithms that require a solution to the correspondence problem and for motion and structure algorithms based on these two paradigms.

2 Optical Flow Techniques

We survey the following classes of optical flow techniques: *a) intensity-based differential methods*, *b) multiconstraint methods*, *c) frequency-based methods*, *d) correlation-based methods*, *e) multiple motion methods* and *f) temporal refinement methods.* The boundaries between each class of methods are not always clear: Weng's method [158] incorporates both phase-based and feature-based matching while Waxman *et al.*'s [156] applies a differential scheme on time-varying edge maps. We classify the former as a phase-based method and the latter as a differential method. In addition, multiple motion and temporal refinement methods for optical flow overlap with other classes. However, their importance dictates that they be covered separately. Following this classification, we describe representative examples of the current state-of-the-art work in optical flow measurement.

2.1 Differential Methods

Differential techniques compute image velocity from spatiotemporal derivatives of image intensities. The image domain is therefore assumed to be continuous (or differentiable) in space and time. Global and local 1^{st} and 2^{nd}-order methods based on (1.3) can be used to compute optical flow. Global methods use (1.3) and an additional global constraint, usually a smoothness regularization term, to compute dense optical flows over large image regions[6]. Local methods

[5] As opposed to the computation of optical flow, which is *ill-posed.*

[6] Often, the entire image is considered.

use normal velocity information in local neighbourhoods to perform a least-squares minimization to find the best fit for $\mathbf{v}$. The size of the neighbourhood for obtaining a velocity estimate determines whether each individual technique is local or global. A surface or contour model may also be used to intergrate normal velocities into full velocity. Occlusion, manifested by discontinuous optical flow, can be analyzed by line processes, mixed velocity distribution or parametric models. These techniques perform the segmentation of optical flow into regions corresponding to various independently moving objects or surfaces. Large 2-d motions may be analyzed in a hierarchical framework, possibly in conjunction with warping methods.

2.1.1 Global Methods

Often, an explicit use of (1.3) is made [5, 21, 44, 54, 53, 70, 106, 107, 108, 127, 128] in conjunction with a regularization term (usually a smoothness constraint). Combined, they form a functional which is minimized over the image domain. Regularization by requiring a slowly varying optical flow field was first introduced by Horn and Schunck [70] to disambiguate normal measurements and justified by the claim that neighbouring velocities, if corresponding to the same object surface, should be nearly identical. These constraints were used to define an error functional:

$$\int_D ((\nabla I \cdot \mathbf{v} + I_t)^2 + \lambda^2 \ \mathrm{tr}((\nabla \mathbf{v})^T (\nabla \mathbf{v})))d\mathbf{x} \tag{2.5}$$

over a domain of interest D, where $\mathbf{v} = (u, v)$. The solution for $\mathbf{v}$ is given as a set of Gauss-Seidel equations which are solved iteratively. Uniform illumination (at least locally) in the image domain of interest, orthographic projection, and pure translational motion parallel to the scene are conditions that must be met for the brightness constancy assumption ($\frac{dI}{dt} = 0$) to be satisfied. These hypotheses reduce the set of admissible visual events to restricted cases of realistic time-varying imagery and motivated the investigation of constraints generating more applicable equations. For instance, motion may cause a change in the density of features in a local image neighbourhood. Schunck [129] accounts for this by modifying (1.3), using the continuity equation from fluid dynamics and transports theory to obtain:

$$\nabla I \cdot \mathbf{v} + I(u_x + v_y) = -I_t. \tag{2.6}$$

This equation is equivalent to (1.3) but for an additional term containing flow divergence which expresses expansion or compression of an image neighbourhood as it undergoes an *affine* transformation[7]. Nagel [109] suggests that the optical flow constraint equation should be explicitly based on the geometric properties of the 3-d scene and derives

$$\nabla I \cdot \mathbf{v} + I_t = 4I\left(\frac{\hat{\mathbf{z}}\dot{\mathbf{P}}^T}{\hat{\mathbf{z}}\mathbf{P}^T} - \frac{\mathbf{P}\dot{\mathbf{P}}^T}{\mathbf{P}\dot{\mathbf{P}}^T}\right) \tag{2.7}$$

where $\mathbf{P}$ is a 3-d environmental point, $\dot{\mathbf{P}}$ is its 3-d velocity and $\hat{\mathbf{z}}$ is a unit vector along the line-of-sight axis. This equation assumes a known scene geometry. An experimental evaluation of these constraints (equations (1.3), (2.6) and (2.7)) in [160] demonstrates that (1.3) has slightly better accuracy when applied to ray-traced synthetic data. Negahdaripour and Yu [114] propose replacing (1.3) with a more general constraint that models a linear transformation over time of the image intensity values. Prince and McVeigh [123] derive the *variable brightness optical flow* equation for an application using MR (magnetic resonance) image sequences. This equation accounts for the fact that $\frac{dI}{dt} \neq 0$ in these images by modelling intensity changes over time as a function of MR parameters, motion and an initial magnetically induced tag pattern. Luettgen *et al.* [96] present a multiscale stochastic algorithm to regularize Horn and Schunck's smoothness constraint. The algorithm is non-iterative and provides confidence measures (multiscale error covariance statistics) to determine the optimal resolution level of optical flow fields. The framework is generalizable to other regularization problems. Mukawa [102] proposed a regularization method to compute optical flow with a global smoothness constraint and other constraints that model both diffuse and specular lighting effects (via Phong shading) for a moving object in a scene with one light source. The term being regularized is

$$\begin{aligned} E &= \sum_R \left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right) + \lambda \sum_R \left(q + I_x u + I_y v + I_t\right)^2 \\ &+ \mu \sum_R \left((q_x - cI_x)^2 + (q_y - cI_y)^2\right) + \nu \sum_R \left(c_x^2 + c_y^2\right) = 0. \end{aligned} \tag{2.8}$$

The first term incorporates Horn and Schunck's smoothness constraint [70]. The second term is the optical flow constraint equation with an additional term q that is the difference of diffuse and specular luminance over time, as first suggested by Cornelius and Kanade [38]. The third and fourth terms exploit a relationship between the spatial derivatives of luminance (q_x and q_y) and the original image intensity derivatives: $q_x = cI_x$ and $q_y = cI_y$. c is a function that involves computing the ratios of diffuse luminance at different times. The fourth term ensures

[7] An affine transformation includes translation, rotation and foreshortening and may be expressed with six parameters: the image velocity and its 1^{st}-order derivatives.

that c varies smoothly. λ, μ and ν are simply constants weighing the relative importance of each term in the minimization.

2.1.2 Local Models

Local models of velocity assuming single motion patterns are also common. For example, Lucas and Kanade use a local constant model for $\mathbf{v}$ [94, 95] which is solved as a weighted least-squares solution to (1.3). Velocity estimates are computed by minimizing

$$\sum_{\mathbf{x} \in R} W^2(\mathbf{x})(\nabla I(\mathbf{x}, t) \cdot \mathbf{v} + I_t(\mathbf{x}, t))^2, \tag{2.9}$$

where $W(\mathbf{x})$ denotes a window function and R is a spatial neighbourhood. Solutions for $\mathbf{v}$ are obtained in closed form. Modifications suggested by Simoncelli *et al.* [133] allow the use of the eigenvalue of the least-squares matrix involved in solving (2.9) as a confidence measure in subsequent processing. Chu and Delp also use a local least squares approach [37] and solved (1.3) with a total least-squares calculation that accounts for errors in I_t and also for independent errors in I_x and I_y[8]. Weber and Malik [157] also use total least-squares in their multiconstraint approach. Campani and Verri [34] use a local model which assumes 1^{st}-order variation in local motion measurements: an overconstrained system of equations is solved with least-squares to recover velocity and its spatial derivatives. However, these local models tend to react poorly in the presence of multiple motions within the neighbourhoods over which they operate.

The aperture problem may be analytically resolved by differentiating the optical flow constraint equations to obtain equations involving 2^{nd}-order intensity derivatives [63, 106, 108, 125, 148, 152]. These constraints generally provide two or more equations in the two components of $\mathbf{v}$ and, when nonsingular, can be used to obtain full motion estimates. For instance, Uras *et al.* [152] use the constraint

$$(\nabla\nabla I)\mathbf{v}^T = -\nabla I_t \tag{2.10}$$

which results in an analytical expression for both components of $\mathbf{v}$ at a single image point. Nagel [108] shows that image points with high Gaussian curvature, such as grayvalue corners,

[8]Consider fitting a line $y = mx + b$ to noisy y values. In standard least-squares, this is accomplished by minimizing the vertical distance of each y value from the fitted line assuming noiseless x values. In total least-squares, it is assumed that there is noise in both the x and y values and the technique minimizes the perpendicular distance of y values from the fitted line [167].

allows the recovery of full velocity in closed form[9]. Both Haralick and Lee [63] and Tretiak and Pastor [148] use a combination of (1.3) and (2.10) to overcome the aperture problem at individual image points and to estimate full velocity.

Another local approach, which avoids the need to estimate intensity derivatives altogether, uses the Gauss divergence theorem to convert the optical flow constraint equation into

$$\int_V (I_x + I_y + I_t) dV = \int_S uI dy dt + \int_S vI dx dt - \int_V I(u_x + v_y) dx dy dt + \int_S I dx dy = 0, \quad (2.11)$$

where S and V denote local integration over surfaces and volumes of intensity data [57]. The size of the surface and volume neighbourhoods must be sufficient to overcome the aperture problem.

2.1.3 Surface Models

Pioneering work by Longuet-Higgins and Prazdny [93] examines the form of the optical flow field for a moving monocular observer in a rigid scene. They derive the well known *image velocity* equation, relating 3-d motion and depth parameters to 2-d image motion (approximated as optical flow). They show that these parameters could be recovered from optical flow and its 1^{st} and 2^{nd}-order derivatives. Longuet Higgins [92] derive the conditions necessary to recover motion and structure from planar surface motion. It is shown that two planes with distinct surface orientations that are engaged in different 3-d motions may have the same optical flow field. Waxman and Wohn [155] describe the *dual* nature of these motion and structure parameters: one set of parameters can be derived from the other. Subbarao and Waxman [142] demonstrate that these solutions are unique over time. Horn [69] proves that multiple interpretations of a single optical flow field generated by arbitrarily shaped surfaces occur only rarely.

A number of planar motion techniques rely on normal velocity being available[10]. Waxman and Wohn's Velocity Functional method [155] assumes that a velocity at a point on a curved surface can be approximated by a 2^{nd}-order Taylor series expansion about that point. For velocity $\mathbf{v} = (u, v)$, one obtains

$$\mathbf{v}(x, y) = \mathbf{v}(0,0) + \frac{\partial \mathbf{v}}{\partial x} + \frac{\partial \mathbf{v}}{\partial y} + \frac{1}{2}\frac{\partial^2 \mathbf{v}}{\partial x^2} + \frac{\partial^2 \mathbf{v}}{\partial x \partial y} + \frac{1}{2}\frac{\partial^2 \mathbf{v}}{\partial y^2}. \quad (2.12)$$

[9] Gaussian curvature may be expressed as $\det(\nabla\nabla I)$.

[10] The measurement of 1^{st}-order spatiotemporal intensity derivatives allows the computation of normal velocity.

Using the normal velocity constraint equation, $v_n = \mathbf{v} \cdot \mathbf{n}$, a linear equation in twelve unknowns, the two components of $\mathbf{v}$ and their 1^{st} and 2^{nd}-order derivatives, is obtained. Given twelve or more normal velocities in a local neighbourhood on the curved surface, these parameters can be recovered. In the event that the local surface is planar, then

$$\frac{\partial^2 \mathbf{v}}{\partial x^2} = (2\frac{\partial^2 v}{\partial x \partial y}, 0) \quad \text{and} \quad \frac{\partial^2 \mathbf{v}}{\partial y^2} = (0, 2\frac{\partial^2 u}{\partial x \partial y}) \ , \tag{2.13}$$

allowing (2.12) to be written as one equation in eight unknowns. Hence, only eight normal velocities are required to recover the velocity of a planar surface.

Murray and Buxton [103] derive a relation between normal velocity parameters of a planar surface, leading to a linear equation of the form

$$||\mathbf{v}_\perp||_2^2 = \mathbf{c} \cdot \mathbf{p}, \tag{2.14}$$

where vectors $\mathbf{c}$ and $\mathbf{p}$ contain expressions for image coordinates of normal velocity, the 3-d motion parameters and the planar surface normal. Given eight normal velocities, the components of $\mathbf{p}$ can be determined and full velocity is obtained as:

$$\mathbf{v} \ = \ (p_1 x + p_2 y - p_3 f + p_7 \frac{x^2}{f} + p_8 \frac{xy}{f}, p_4 x + p_5 y - p_6 f + p_7 \frac{xy}{f} + p_8 \frac{y^2}{f}). \tag{2.15}$$

Further mathematical manipulation of $\mathbf{p}$ yields equations for the recovery of 3-d motion parameters and surface orientation. If the aperture problem cannot be overcome for some image locations in a neighbourhood, but that surface parameters can be estimated, then, because of the use of a surface model, an image velocity for each of these locations may be inferred.

2.1.4 Contour Models

Many differential approaches to image motion estimation rely on the presence of contours or edges in image sequences [33, 42, 68, 119, 156]. The computational stages of these methods consist of the extraction of relevant image contours with prefiltering techniques followed by a differential estimation of image motion. Essentially, contours or edges exhibit strong signal-to-noise ratios which facilitate their extraction. In addition, it is a common belief that they correspond to significant image structures, although this claim cannot be supported in any rigorous sense [47]. In addition, computing optical flow at edges often leads to sparse flow fields (usually, 10% of the field or less, depending on the density of edges).

Hildreth [68] proposes a smoothness constraint to be applied to normal velocity estimates along contours extracted from time-varying images. For a contour S, the normal velocity estimates should minimize

$$\int \frac{\partial \mathbf{v}}{\partial S} dS. \tag{2.16}$$

If at least two normal velocity estimates along S are different, then the minimization of the above integral yields a unique velocity field at contour S (see Figure 2.5). In practice, the

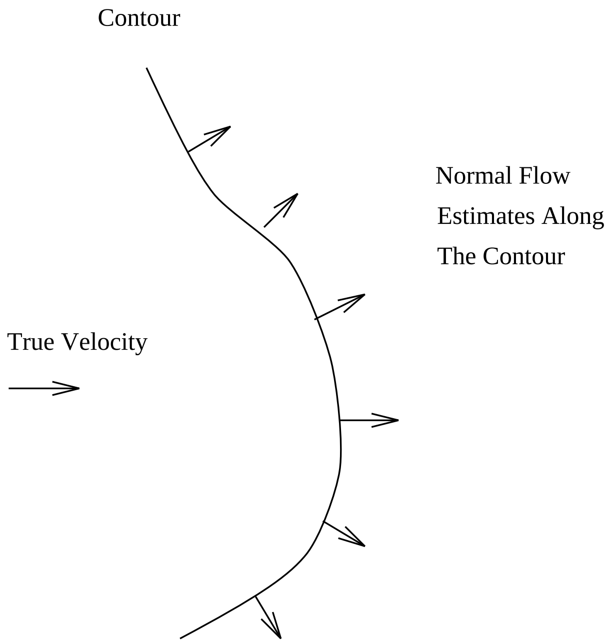

Figure 2.5: *If at least two normal flow estimates along a contour are different, then the full velocity of the contour can be uniquely determined.*

functional

$$\int \left(\frac{\partial \mathbf{v}}{\partial S}\right)^2 + \beta(\mathbf{v}\cdot\hat{\mathbf{n}} - \|\mathbf{v}_\perp\|_2)^2 dS, \tag{2.17}$$

where $\hat{\mathbf{n}}$ is a unit vector in the direction of $\mathbf{v}_\perp$, is minimized along contours. β is a weighting factor and $(\mathbf{v}\cdot\hat{\mathbf{n}} - \|\mathbf{v}_\perp\|_2)^2$ expresses the squared difference between estimated normal velocity and that predicted by the solution. Gong and Brady formulate a similar constraint to be minimized which includes a least-squares difference term for tangential velocity [55]:

$$\int \left(\frac{\partial \mathbf{v}}{\partial S}\right)^2 + \beta(\mathbf{v}\cdot\hat{\mathbf{n}} - \|\mathbf{v}_\perp\|_2)^2 dS + \alpha(\mathbf{v}\cdot\hat{\mathbf{t}} - \|\mathbf{v}_\|\|_2)^2 dS, \tag{2.18}$$

where $\mathbf{t}$ is a unit vector perpendicular to $\mathbf{n}$ and α is a scalar expressing the confidence associated with the tangential component of $\mathbf{v}$, proportional to the determinant of the Hessian matrix of the underlying intensity structure. They demonstrate that the tangential component of $\mathbf{v}$ can be reliably estimated wherever the determinant of the Hessian is non-zero.

Buxton and Buxton's method [33] for estimating optical flow is based on a model of the motion of edges in image sequences. They note that the signal-to-noise ratio is enhanced at locations of significant image features such as edges. Their approach, guided by psychophysical evidence of spatiotemporal filtering in human early visual processes, is a direct extension of the Marr and Hildreth centre-surround edge detection operator [98]. They compute spatiotemporal zero-crossings by convolving

$$S(\mathbf{x},t) = -\left(\left(\nabla^2 + \frac{1}{u^2}\frac{\partial^2}{\partial t^2}\right) G(\mathbf{x},t)\right) \tag{2.19}$$

with an image sequence, where the first term is the d'Alembert operator and the second term is a Gaussian function, given by

$$G(\mathbf{x},t) = u\left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} e^{-\alpha(\|\mathbf{x}\|_2^2 + u^2t^2)}, \tag{2.20}$$

where the parameters α and u control the spread of the envelope. Normal velocities can then be estimated at zero-crossings by computing the partial 1^{st}-order derivatives of $S(\mathbf{x},t)$ as

$$\mathbf{v}_\perp = \frac{S_t \nabla S}{\|\nabla S\|_2^2} \tag{2.21}$$

with a least-squares calculation. The choice for the values of the parameters α and u is made *a priori* since the distribution of edges in the image is unknown.

Duncan and Chou define a temporal edge detector which minimizes the effects of temporal variations in illumination [42]. The edge detector is the 2^{nd}-order temporal derivative of a Gaussian function:

$$S_t(\mathbf{x},t) = \frac{\partial^2 G(t)}{\partial t^2} = \frac{-2s^3}{\sqrt{\pi}}(1 - 2s^2t^2)e^{-s^2t^2} \tag{2.22}$$

which is convolved in time with the image sequence to produce a set of zero-crossings induced by moving edges. The authors theoretically and experimentally show that variance in illumination does not create zero-crossings in $S_t(\mathbf{x}, t)$. Central differences are used to estimate the 1^{st}-order derivatives of S_t in local neighbourhoods. Normal velocities are computed as

$$\nabla S_t \cdot \mathbf{v}_\perp - \frac{\partial S_t}{\partial t} = 0 \tag{2.23}$$

which is equivalent to the usual optical flow constraint equation. Lines defined by $\mathbf{v}_\perp$ are then intersected in local image regions to obtain full velocity estimates. Successful experiments are presented with synthetic images containing significant illumination variations.

Waxman *et al.* [156] apply spatiotemporal filters to edge maps in order to measure velocity at edges extracted with DOG zero-crossings [98]. Given a binary edge map $E(\mathbf{x}, t)$, an *activation profile* $A(\mathbf{x}, t)$ is created by smoothing the edge map with a spatiotemporal Gaussian filter:

$$A(\mathbf{x}, t) = G(\mathbf{x}, t, \sigma_x, \sigma_y, \sigma_t) * E(\mathbf{x}, t) \tag{2.24}$$

to which a differential method is applied. At edge locations, where the Gaussian profiles are centered, the spatial gradient of A should be zero, and therefore a 2^{nd}-order approach is adopted. The velocity estimates are given by

$$\mathbf{v} = \left(\frac{A_{xt}A_{yy} - A_{yt}A_{xy}}{A_{xx}A_{yy} - A_{xy}^2}, \frac{A_{yt}A_{xx} - A_{xt}A_{xy}}{A_{xx}A_{yy} - A_{xy}^2} \right), \tag{2.25}$$

where derivatives of A are computed by convolving the appropriate Gaussian derivative kernels with the edge maps.

A different method for tracking edges via zero-crossings is to use a radial configuration of DOG operators. Perrone [119] presents such a model to measure normal velocities of moving edges. In the radial configuration, temporal differences in the responses of the operators provide sufficient information towards determining normal velocity. A constant model can then be applied to the normal velocity distributions to obtain full velocity.

2.1.5 Multiconstraint Methods

Multiconstraint methods use multiple instances of (1.3) or (2.10) to provide unambiguous expressions for image motion at single image points [161, 90, 140, 137, 157, 100, 147]. Liu *et al.* [90] use equations (1.3) and (2.10): they expand the spatiotemporal image with Hermite polynomials and solve for $\mathbf{v}$ using standard least-squares. Residual of fit, condition number and determinant of the least-squares matrix act as confidence measures on the final optical flow field. Overconstrained systems of equations can also be obtained with multiple light sources [161] or by using images acquired with visual sensors tuned to different regions of the light spectrum. Spectral images include those in the visible (3 colour planes) and infrared spectrum. Markandey and Flinchbaugh show that their multispectral approach produced similar accuracy to Horn and Schunck's algorithm [70] for synthetic image data and an outdoor scene [97].

Functions other than (or coupled with) intensity may be substituted in the optical flow constraint equation to obtain overconstrained systems. These functions can be thought of as the output of operations applied to the image intensities. For example, responses of pairs of linearly independent filters can be used jointly with the optical flow constraint equation for this purpose [140, 137, 157]. Mitiche *et al.* [100] use an overconstrained system of equations constructed from the optical flow constraint equation for the same point in a number of different images, derived from the original one, generated by applying functions to compute local values for contrast, average, variance, entropy, median and power content. Srinivasan proposed a similar approach, using an overconstrained system of equations derived from images that are generated by applying six specialized spatiotemporal filters on the original images [140]. However, the aperture problem still cannot be resolved when facing singularities in overconstrained system of equations: these occur for particular intensity structures, including uniform intensity regions, highly structured or periodic textures, *etc.*

2.1.6 Hierarchical Approaches

Differential optical flow methods also exhibit problems with large 2-d motions, due to low sampling rates, thus violating the Shannon sampling theorem. Applying differential methods in a coarse-to-fine manner alleviates such problems. Image warping may be used to keep the images sufficiently well registered at the scale of interest so that numerical differentiation can be performed. Bergen *et al.*'s hierarchical framework [21] unifies several different model-based optical flow methods. Using a parametric model based on affine transformations, scene rigidity, surface planarity or general motion in an image region allows one to both judge the quality of the fit to the data (perhaps splitting the region if necessary) and to fill in sparse optical flow fields with the computed parameters.

Differential-based hierarchical frameworks are proposed by Glazer, Enkelmann and Battiti *et al.* [18, 44, 53, 54]. In addition, Bergen *et al.* have shown that many models of motion could be expressed within a hierarchical framework [20]. Adaptive hierarchical methods with respect to scale have also appeared [18, 82, 159, 96].

2.2 Frequency-Based Methods

A second class of optical flow techniques are based on the use of velocity-tuned filters. These techniques use orientation sensitive filters in the Fourier domain of time-varying images [1, 48, 56, 65, 154]. Among advantages brought by these methods, it is found that motion-sensitive mechanisms operating on spatiotemporally oriented energy in Fourier space can estimate motion in image signals for which matching approaches would fail. For example, the motion of random dot patterns may be difficult to capture with feature-based or correlation-based methods whereas, in Fourier space, the resulting oriented energy may be more readily extracted to compute motion [1].

The Fourier transform of a translating 2-d intensity signal specified in (1.1) is

$$\hat{I}(\mathbf{k}, \omega) \;=\; \hat{I}_0(\mathbf{k})\delta(\mathbf{v}^T\mathbf{k} + \omega), \tag{2.26}$$

where $\hat{I}_0(\mathbf{k})$ is the Fourier transform of $I(\mathbf{x}, 0)$ and $\mathbf{x}$ denotes spatial position. δ is a Dirac delta function and $\mathbf{k}, \omega$ denote spatiotemporal frequency. This yields the optical flow constraint equation in frequency space:

$$\mathbf{v}^T\mathbf{k} + \omega = 0, \tag{2.27}$$

which shows that the velocity of a translating 2-d pattern is a function of its spatiotemporal frequency and forms a plane through the origin of the Fourier space.

2.2.1 Orientation Selective Filtering

Adelson and Bergen [1] propose a class of computational schemes which exploits the fact that detecting image motion is equivalent to extracting spatiotemporal orientation. Gabor filtering is presented as a technique for extracting spatiotemporal energy. A Gabor filter is a Gaussian function multiplied by a sine or cosine wave. For example, the function

$$G(\mathbf{x}, t) = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\sigma_t} e^{-\left(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}+\frac{t^2}{2\sigma_t^2}\right)} \sin(2\pi(\mathbf{x}\cdot\mathbf{k} + \omega t)) \tag{2.28}$$

is a 3-d sine (odd) Gabor filter, where $(\mathbf{k}, \omega)$ is the central frequency at which response amplitude peaks. Adelson and Bergen note that the response pattern of such filters is affected by the contrast of the signal: stimuli with low contrast generate low response amplitudes and *vice versa*. Since the measurement of velocity is independent of contrast amplitudes, it is suggested that one uses ratios of responses from different filters for the extraction of velocity estimates.

Jähne [74] demonstrates that detection of spatiotemporal orientation is analogous to eigenvalue analysis of inertia tensor. For instance, a point at the origin of Fourier space corresponds to a region of constant intensity and corresponds to null eigenvalues in tensor space. Also, a line in frequency space is a spatially oriented pattern moving with constant velocity and its normal velocity can be obtained with the eigenvector corresponding to the null eigenvalue. Finally, a plane through the origin expresses a spatially distributed pattern moving with constant velocity and the eigenvector associated with the maximum eigenvalue of the inertia tensor gives the full velocity of the pattern. Hence, Jähne [74] suggests the use of inertia tensor analysis for detecting spatiotemporal orientation and to avoid the computational expense arising from the large number of filters involved in both Heeger's and Fleet and Jepson's methods [48, 65]. However, this algorithm was only tested on simulated image data. Barman *et al.*'s work [11] is similar to Jähne's, as they use spatiotemporal filtering to recover both velocity (and depth-scaled disparity in the case of motion stereo) and acceleration. A tensor is computed from the response of six or more quadrature filters, evenly spread in one half of the Fourier space. Eigenvalues and eigenvectors of the tensor allow one to determine which of several formulae can be used to compute full or normal velocity, if possible. Acceleration may also be recovered as curvature of the spatiotemporal surface in frequency space [10].

Often, these frequency-based velocity techniques are presented as biological models of human motion sensing: Watson and Ahumada [154] have defined an orientation-selective mechanism which agrees with psychophysical measurements of human motion sensing. Their mechanism uses a combination of 2-d spatial Gabor functions and 1-d temporal filters tuned to several orientations for the estimation of local image velocity. The integration of the responses of these sensors discriminates local measurements as each sensor within a directional group provides a linearly independent component of the velocity vector. Grzywacz and Yuille also propose a frequency-based model of visual motion sensing [56]. Their model uses 3-d orientation-selective Gabor filters to measure motion energy in frequency space. Estimates of velocity are also obtained by integrating the responses of populations of filters on a local basis.

2.2.2 Phase-Based Filtering

The method developed by Fleet and Jepson [48] defines component velocity in terms of the instantaneous motion of level phase contours in the output of band-pass velocity-tuned Gabor filters[11]. These filters are used to decompose the input signal according to scale, speed and orientation. Each filter output is complex-valued and can be expressed as

$$R(\mathbf{x},t) = \rho(\mathbf{x},t)e^{i\phi(\mathbf{x},t)}, \tag{2.29}$$

where $\rho(\mathbf{x},t)$ and $\phi(\mathbf{x},t)$ are the amplitude and phase part of the output signal. The component 2-d velocity in the direction normal to level phase contours is given by

$$\mathbf{v}_\perp = \frac{-\phi_t(\mathbf{x},t)\nabla\phi(\mathbf{x},t)}{\|\nabla\phi(\mathbf{x},t)\|_2^2}. \tag{2.30}$$

$\phi_t(\mathbf{x},t)$ is the temporal derivative of the phase and $\nabla\phi(\mathbf{x},t)$ is its spatial gradient. Phase derivatives are computed using the identity

$$\nabla\phi(\mathbf{x},t) = \frac{\mathrm{Im}[R^*(\mathbf{x},t)\nabla R(\mathbf{x},t)]}{|R(\mathbf{x},t)|^2}, \tag{2.31}$$

where R^* is the complex conjugate of $R(\mathbf{x},t)$, $\nabla R(\mathbf{x},t)$ is the gradient of $R(\mathbf{x},t)$ and Im denotes the imaginary part of a complex number. Fleet and Jepson relate velocity to local phase information because of the relative insensitivity of the phase signal to amplitude variations due to changes in scene illumination. These distortions are often a consequence of the geometry of perspective projection. Component velocity can be obtained from the output of each velocity-tuned channel on the condition that the phase signal is stable. Usually, instabilities are associated with neighbourhoods about phase singularities which may be detected with a constraint stating that the distance between the instantaneous frequency and the peak tuning frequency of the filter should be minimal. Such a constraint, when met, is sufficient to avoid velocity estimation at phase singularities. Because each channel is considered independently, there may be multiple measurements at a single image location. Then, if there is a sufficient number of estimates, full velocity is recovered at a single point or in a small neighbourhood by solving a linear system of equations relating the measurements to an affine model of optical

[11]The term *component* velocity is used to denote the velocity normal to local phase structure while *normal* velocity denotes velocity normal to local intensity structure.

flow. As currently formulated, Fleet and Jepson's method does not provide confidence measures. There is also a requirement for a large number of filters to cover frequency space. It is also important to note that local phase information is also used in stereopsis for the measurement of image disparities by Jenkin *et al.* [77] and Langley *et al.* [85]. Weng [158] demonstrates that the phase part of a signal in a particular frequency channel provides sufficient information to reconstruct the signal within a multiplicative constant. Windowed Fourier phase is used as the correlation primitive in Weng's matching algorithm.

Phase is more robust than either intensity derivatives or energy-based filter responses for varying scene illumination. This can be seen intuitively by considering a signal of the form $A\cos(\omega t + \phi)$. Changing the amplitude, A, of the signal will change its derivative values or the response of an energy based (amplitude-squared) filter but will have little effect on the phase of the signal or its derivatives. Phase-based methods will also respond to the component velocities of multiple motions due to occlusion or transparency, although a suitable measurement integration method, such as a mixed velocity model, is required to cluster them into the correct full velocities. Typically, frequency-based methods do not provide a means of assigning confidence to the computed velocities. Thresholds may be provided [48, 65] but their use is binary: either a velocity is found at an image location or one is not.

2.2.3 Hierarchical Approaches

Heeger presents a computational model for the estimation of image velocity which uses quadrature pairs of spatiotemporal Gabor filters [65]. A family of Gabor-energy filters, tuned to the same spatial frequency band, but to different spatial orientations is defined. The magnitude of the spatial frequency tuning is therefore invariant and other sets of filters can be designed for different frequency channels. However, a single set of such filters can be cascaded through a Gaussian pyramid in order to cover different channels. For a translating 2-d pattern, the responses of these filters are concentrated about a plane in frequency space. Parseval's theorem is used to derive the expected responses $R_i(\mathbf{v})$ of the Gabor-energy filters for a translating stimulus:

$$R_i(\mathbf{v}) = e^{-4\pi^2\sigma_x^2\sigma_y^2\sigma_t^2 H_i(\mathbf{v},\mathbf{k}_i,\omega_i)} \tag{2.32}$$

where

$$H_i(\mathbf{v},\mathbf{k}_i,\omega_i) = \frac{\mathbf{v}\cdot\mathbf{k}_i + \omega_i}{(u\sigma_x\sigma_t)^2 + (v\sigma_y\sigma_t)^2 + (\sigma_x\sigma_y)^2}.$$

The variables σ_x, σ_y and σ_t are the standard deviations of the Gabor filters. If m_i is the measured energy for filter i and $\bar{m}_i$ and $\bar{R}_i$ are the sums of m_i and R_i which belong to the set of filters M_i having the same spatial orientation as the i^{th} filter:

$$\bar{m}_i = \sum_{j \in M_i} m_j$$

$$\bar{R}_i = \sum_{j \in M_i} R_j(\mathbf{v}),$$

then a non-linear least-squares solution for $\mathbf{v}$ which minimizes the difference between the predicted and measured energies,

$$f(\mathbf{v}) = \sum_{i=1}^{n} \left(m_i - \bar{m}_i \frac{R_i(\mathbf{v})}{\bar{R}_i(\mathbf{v})} \right)^2, \tag{2.33}$$

should yield the correct estimate for $\mathbf{v}$. Certainty measures for velocity estimates are expressed as conditional probability densities. A computationally efficient method of convolving Gabor filters, exploiting their separability properties, is also presented by Heeger.

2.3 Correlation-Based Methods

Numerical differentiation is sometimes impractical because of small temporal support (only a few frames) or poor signal-to-noise ratio [14]. In these cases, differential or frequency approaches may not be appropriate and it is natural to consider matching methods.

2.3.1 Correlation-Based Matching

Typically, good, matchable features, such as corner points, are sparse while poor, easily mismatched features, such as edges, are denser. Even when reasonably unique features are available, establishing the correct correspondences can be problematic. Also further complicating matters is occlusion of features that may lead to matching errors.

Correlation-based matching approaches are less sensitive to these problems: they do not rely on the presence of significant image features and variable correlation window sizes can be used near occlusion boundaries to handle multiple motions [88]. These approaches define displacement (which is an approximation to velocity) as a shift that yields the best fit between contiguous time-varying image regions. Most of these approaches originate from computational stereopsis, where the task is to correlate image regions of a pair of images taken from different viewing positions, under perspective projection. It is assumed that, at least locally, distortions caused by the shift in the viewing angle are negligible [77]. Matching image regions often amounts to maximizing a similarity measure. In particular, a correlation coefficient between two functions f and g is defined as the integral of their product:

$$\int_D f(x+\delta x)g(x)dx.$$

Finding δx which maximizes this integral amounts to finding the shift between f and g, if $f(x+\delta x) = g(x)$.

Kories and Zimmerman use a monotonicity operator for matching image regions in adjacent images [84]. The operator is a 3×3 window in which the central grayvalue is compared with its neighbours and classified according to the number of grayvalues that have a lower value than the central one. The matching process proceeds by first merging adjacent grayvalues sharing the same class. The centroids of grayvalue regions sharing the same classification are then tracked with a simple correlation algorithm to establish disparity estimates.

Sutton *et al.* [144] proposed a correlation method which allows linear deformations of small image regions. A bilinear reconstruction of the intensity surface is computed for local image regions Ω_i. The shift of such an intensity surface under linear deformation is expressed as

$$\mathbf{x'}_{\Omega_i} = \mathbf{x}_{\Omega_i} + \mathbf{d}_{\Omega_i} + \nabla \mathbf{d}_{\Omega_i} \tag{2.34}$$

where $\mathbf{x'}_{\Omega_i}$ is the position of the deformed neighbourhood Ω_i, $\nabla \mathbf{d}_{\Omega_i}$ are the deformation parameters and $\mathbf{d}$ is the shift of Ω_i in time. A correlation coefficient

$$C(\mathbf{d}_{\Omega i}, \nabla \mathbf{d}_{\Omega i}) = \int_D (\Omega_i - \Omega_i')^2 d\mathbf{x} \tag{2.35}$$

which describes the squared differences of neighbourhoods Ω_i and Ω'_i, is minimized by a search in the motion domain for values of $\mathbf{d}_{\Omega i}$ and $\nabla \mathbf{d}_{\Omega i}$. The estimates for these six parameters must be obtained iteratively, as no closed form solution for C is presented.

A region matching method that allows affine deformations of intensity is presented by Kalivas and Sawchuk [79]. An objective function that is defined in terms of a displacement field undergoing an affine transformation is minimized over the entire image.

In order to avoid the computational expense of iterative minimizations of functionals, Little *et al.* [89] suggest that the use of partially overlapping regions for matching is sufficient to approximate an isotropic smoothness term imposed on disparity estimates. Their approximation of the functional

$$\int_D (\phi(I(\mathbf{x},t), I(\mathbf{x}+\delta t\mathbf{v}, t+\delta t)) + \lambda(\nabla\mathbf{v}\nabla\mathbf{v}^T))d\mathbf{x} \tag{2.36}$$

where ϕ is a correlation operator, is shown to be correct as the extent of the matching region becomes larger.

2.3.2 Hierarchical Approaches

In the presence of large disparities, non-hierarchical correlation algorithms become sensitive to false matches, due to the increase in search spaces required to handle the faster motion. In addition, the correlation of image areas is, in general, computationally intensive. In order to reduce the amount of computations and the potential for mismatches, one may use coarse estimates of motion to direct the matching process. Ogata and Sato's algorithm [116] provides the correlation computation with coarse estimates of motion obtained from velocity-tuned Gabor filters. These estimates can then be used to restrict the sizes of search areas and thereby reduce the number of computations usually necessary to obtain disparities.

The size of correlation windows is an important parameter for region matching. For instance, within a correlation window, there must be enough variation in the signal to reliably determine disparity. However, the variation of disparity within the same window must remain negligible as local matches operate under the hypothesis of a constant velocity model. The optimal window size then depends on the structure of the underlying signal. Okutomi and Kanade [117] propose a statistical model of disparity within correlation windows which assumes that disparity values are constant but exhibit increasing uncertainty as they are farther from the central point of the window. This model establishes a relation between the window size and the uncertainty of disparity. This relation allows one to minimize the uncertainty of the measurements by adjusting the size of the correlation window.

Hierarchical matching techniques can improve the accuracy of the disparities by operating on several frequency channels extracted from the images to be processed. Low frequency channels are used to estimate large disparities which can be refined by adding higher frequency channels into the matching process. Anandan's method [8] is based on a Laplacian pyramid and a coarse-to-fine SSD-based matching strategy. The Laplacian pyramid allows for the estimation of large inter-frame disparities and helps to enhance image structure, such as edges, that is thought to be important for matching (see Figures 1.4 and 2.6). The SSD (sum of squared difference)

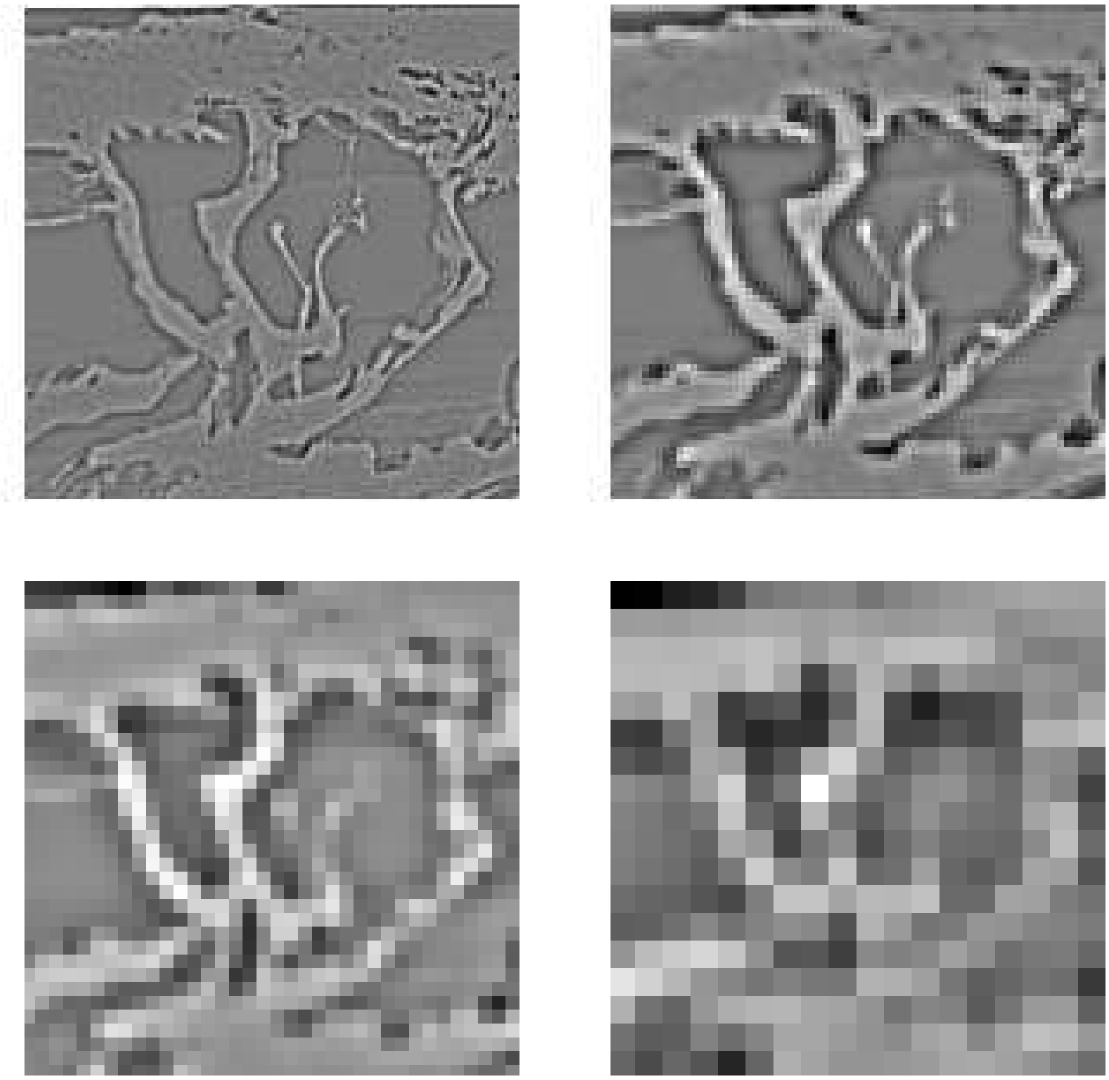

Figure 2.6: *Hierarchical image decomposition: the original images are decomposed in a hierarchical set of frequency channels prior to optic flow estimation.*

measure is defined as

$$S(\mathbf{x},\mathbf{d}) = \sum_{j=-n}^{n}\sum_{i=-n}^{n} W(i,j)(I(\mathbf{x}+(i,j),t) - I(\mathbf{x}+(i,j),t+1))^2 \tag{2.37}$$

where $W(i,j)$ denotes a weighting function and $\mathbf{d}$ is restricted to the square neighbourhood of size $(2n+1)^2$ centered at $\mathbf{x}$. At the coarsest level, the correct displacements are assumed to be one spatial unit per frame or less. SSD minima are first located to integer accuracy within small image regions. Subpixel dispacements are then computed by finding the minimum of a quadratic approximation to the SSD surface about the integer location which best minimizes $S(\mathbf{x},\mathbf{d})$. Confidence measures are derived from the principal curvatures of the SSD surface and used as weights in the functional

$$\int_D \mathrm{tr}((\nabla\mathbf{d})^T(\nabla\mathbf{d})) + c_{max}(\mathbf{d}\cdot\hat{e}_{max} - \mathbf{d}_0\cdot\hat{e}_{max})^2 + c_{min}(\mathbf{d}\cdot\hat{e}_{min} - \mathbf{d}_0\cdot\hat{e}_{min})^2 d\mathbf{x} \tag{2.38}$$

which is to be minimized over the entire image velocity domain D. $\hat{e}_{max}$ and $\hat{e}_{min}$ are the normalized principal directions of maximum and minimum curvature. $\mathbf{d}_0$ denotes the displacements obtained by the minimization of $S(\mathbf{x},\mathbf{d})$ in the laplacian pyramid (usually, three levels are used). At the coarsest level, where the largest motion is less than one spatial unit per frame, SSD minima are located to subpixel accuracy by finding the minimum of the SSD correlation surface and smoothed using iterative equations based on (2.38). Then, using an overlapped projection scheme, these displacements are projected to the next level in the pyramid. Matching and smoothing are performed in this manner at each level in the pyramid from the coarsest level down to the finest level, the original image, yielding the final optical flow field.

Singh's approach [136] is similar to Anandan's method [8] as it also uses SSD minimizations and a two-stage computation. The first stage consists of the computation of SSD values with three adjacent high-pass filtered images I_{-1}, I_0 and I_1. The three-frame SSD surface is computed as

$$S(\mathbf{x},\mathbf{d}) = S_{0,1}(\mathbf{x},\mathbf{d}) + S_{0,-1}(\mathbf{x},-\mathbf{d}). \tag{2.39}$$

The surface S is then converted into a probability distribution, defined as

$$R_c(\mathbf{d}) = e^{-kS(\mathbf{x},\mathbf{d})}, \tag{2.40}$$

where

$$k = \frac{\ln(0.95)}{\min(S(\mathbf{x},\mathbf{d}))}.$$

Velocity estimates $\mathbf{d}_c$ are obtained by computing the weighted average using the $R_c(\mathbf{d})$ values for a given image area. Covariance matrices S_c associated with disparities $\mathbf{d}_c$ are also computed.

The second step of Singh's algorithm propagates velocity using a neighbourhood smoothness constraint. Again, a weighted average approach is used in computing $\mathbf{d}_a$, an average of $\mathbf{d}_c$ over small image regions. $\mathbf{d}_c$ and $\mathbf{d}_a$ are then used to create a covariance matrix S_n. The correct disparity field minimizes the functional

$$\int_D (\mathbf{d}-\mathbf{d}_a)S_n^{-1}(\mathbf{d}-\mathbf{d}_a)^T + (\mathbf{d}-\mathbf{d}_c)S_c^{-1}(\mathbf{d}-\mathbf{d}_c)^T d\mathbf{x}, \tag{2.41}$$

which expresses the requirement of a smoothly varying disparity field across D. The eigenvalues λ_1 and λ_2 of matrix $[S_n + S_c]^{-1}$ act as confidence measures for the estimates. Singh recommends the use of a Laplacian pyramid with a coarse-to-fine strategy as in [8] to estimate larger velocities. In addition, Sing's framework was extended with a Kalman filter approach in order to record motion estimates along with their confidence measures and to integrate new measurements with existing estimates [135].

2.4 Multiple Motion Methods

Many phenomena can cause multiple image motions. Among them, occlusion and transparency are important in terms of their occurrence and significance in realistic imagery. In addition, their information content is useful to later stages of processing, such as motion segmentation and 3-d surface reconstruction. Occlusion boundaries are described by the partial occlusion of a surface by another, while transparency is defined as occlusion of a surface by translucent material. In realistic imagery, one finds occlusion to be the most frequent cause of discontinuous motion[12].

Among the limitations inherent to gradient-based methods, the requirement of differentiable intensity structures throughout the image domain is perhaps the most restrictive. At motion discontinuities where most of the information resides, the use of (1.3) becomes problematic, since the intensity derivatives theoretically do not exist. In addition, typical correlation-based techniques are sensitive to occlusion as image structures near occlusion boundaries may appear or disappear in time, possibly leading to mismatches. Furthermore, local optical flow constraints such as (1.3) and local correlation methods are often coupled with global requirements which

[12] Prazdny [122] was one of the first to explicitly allow for disparity discontinuities locally in two stereo images.

impose a spatial continuity on optical flow. It is obvious that such isotropic requirements cannot be satisfied in general, as imagery often contains motion discontinuities.

2.4.1 Line Processes

Other functionals, that attempt to estimate discontinuous motion, have been developed. One strategy to handle occlusion involves using binary line processes [51] that explicitly model intensity discontinuities. Koch *et al.* [83] relax the imposition of a smoothness constraint at those pixels having a large spatial gradient. This prevents smoothing over discontinuities and assumes that motion discontinuities occur in the same location as intensity discontinuities [120]. Black [23] also shows how line processes could be used in a robust (Kalman filter-like) framework.

The non-binary inhibition of smoothness across intensity contours, was proposed by Nagel [106, 107, 108, 111]. This approach is based on the minimization of the functional

$$\int_D (\nabla I \cdot \mathbf{v} + I_t)^2 + \lambda^2 \ \mathrm{tr}((\nabla \mathbf{v})^T W (\nabla \mathbf{v})) d\mathbf{x} \tag{2.42}$$

where

$$W = \frac{1}{\|\nabla I\|_2^2 + 2\delta} \begin{pmatrix} {I_y}^2 + \delta & -I_x I_y \\ -I_x I_y & {I_x}^2 + \delta \end{pmatrix},$$

in which the quantity $(u_x I_y - u_y I_x)^2 + (v_x I_y - v_y I_x)^2$, representing the spatial variation of $\mathbf{v}$ in the direction perpendicular to the image gradient, is minimized across intensity contours. This functional is known as the *oriented smoothness* constraint. The minimization procedure can be implemented using finite differences [14] or finite elements [81, 127]. This oriented smoothness approach has been recently extended into the temporal domain [110]. A similar method using an intensity-weighted smoothing procedure is presented by Aisbett [5]. This approach is characterized by the inhibition of an isotropic smoothness constraint for image regions containing significant intensity variations. Contrary to Nagel's method [108], the inhibition of the smoothness constraint is not directional. In addition, the image intensities are assumed differentiable and the domain of application of the algorithm is explicitly restricted to images which satisfy this requirement.

Closed curves may also be used to separate image regions exhibiting different velocities. Schnörr [128] proposes a method which consists of defining such a curve, delimiting an arbitrary area around the region where the existence of an independently moving object is assumed, thus creating two domains Ω_i and Ω_o. Given two velocity fields $\mathbf{v}_i$ and $\mathbf{v}_o$, the closed curve defining the two domains is iteratively refined by minimizing the functional

$$\int_{\Omega_i} f(\hat{\mathbf{v}}_i, \mathbf{v}_i) d\mathbf{x} + \int_{\Omega_o} g(\hat{\mathbf{v}}_o, \mathbf{v}_i) d\mathbf{x} \tag{2.43}$$

where $f = (\hat{\mathbf{v}}_i - \mathbf{v}_i)^2$, $g = (\hat{\mathbf{v}}_o - \mathbf{v}_o)^2$ and $\hat{\mathbf{v}}_i$ and $\hat{\mathbf{v}}_o$ are measured velocities within Ω_i and Ω_o respectively. However, assuming *a priori* knowledge about the position of the independently moving object limits the generality of this method.

2.4.2 Mixed Velocity Distributions

Another strategy for estimating discontinuous optical flow is to make explicit a model for mixed velocity distributions (usually two) at each image point. A method of estimating discontinuous motion on a local basis, presented by Schunck [130], uses the optical flow constraint equation (1.3) to compute several constraint lines in velocity space for small spatial neighbourhoods. Clusters of intersections of these lines with the constraint line of the central point of the neighbourhood are analyzed to determine the smallest cluster containing at least half of the intersection points. The middle point of this cluster thus defines the motion estimate $\mathbf{v}$. If two motion patterns are present within the neighbourhood, then $\mathbf{v}$ is considered as the dominant one. Hence, velocity can be correctly estimated across motion boundaries (see Figure 2.7). However, neighbourhood sizes must include significant constraint line variations, as finding intersections of constraint lines may become ill-conditioned otherwise (this is simply another manifestation of the aperture problem). Jepson and Black's *mixture models* [78] also follow this approach, but use a robust estimation framework.

When multiple motions arise within a single image region, a least-squares solution to the optical flow constraint line clustering problem leads to an average estimate of these multiple motions. Noting that difficulty, Black [22, 23, 24] reformulates the problem of estimating optical flow by using *robust* estimators. This framework consists of the minimization of a functional that expresses the various assumptions made about image motion:

$$E(\mathbf{v}, \bar{\mathbf{v}}) = \lambda_D E_D(\mathbf{v}) + \lambda_S E_S(\mathbf{v}) + \lambda_T E_T(\mathbf{v}, \bar{\mathbf{v}}), \tag{2.44}$$

where $E_D(\mathbf{v})$ is the optical flow constraint equation, $E_S(\mathbf{v})$ is a spatial coherence constraint (a spatial smoothness term) and $E_T(\mathbf{v}, \bar{\mathbf{v}})$ is a temporal coherence constraint:

$$E_T(\mathbf{v}, \bar{\mathbf{v}}) \;\; = \;\; \rho_T(\mathbf{v} - \bar{\mathbf{v}}, \sigma_t) \tag{2.45}$$

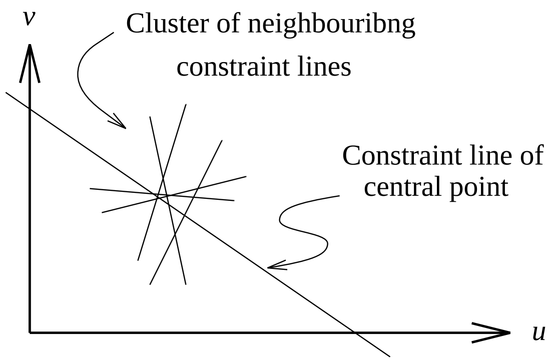

Figure 2.7: *The constraint line from the central point of the image region being considered intersects with the constraint lines of neighbouring image points. Clusters of intersections allow the determination of the dominant velocity pattern [130].*

$$E_D(\mathbf{v}) = \rho_D(\nabla I \cdot \mathbf{v} + I_t, \sigma_d) \quad (2.46)$$

$$E_S(\mathbf{v}) = \sum_{n \in \Omega} \rho_S(\mathbf{v} - \mathbf{v}_\perp, \sigma_s). \quad (2.47)$$

$\bar{\mathbf{v}}$ is a prediction about $\mathbf{v}$ at time $t+1$. ρ_D, ρ_S and ρ_T are *robust*, Lorentzian M-estimators. Their use is motivated by the fact that the distribution of multiple motions within a single image region is not Gaussian and to account for events unmodelled by the brightness constancy assumption. The robustness of these statistical estimators is characterized by their relative insensitivity to deviations from the assumed statistical model in the set of measurements, allowing the estimation of discontinuous optical flows (see Figure 2.8).

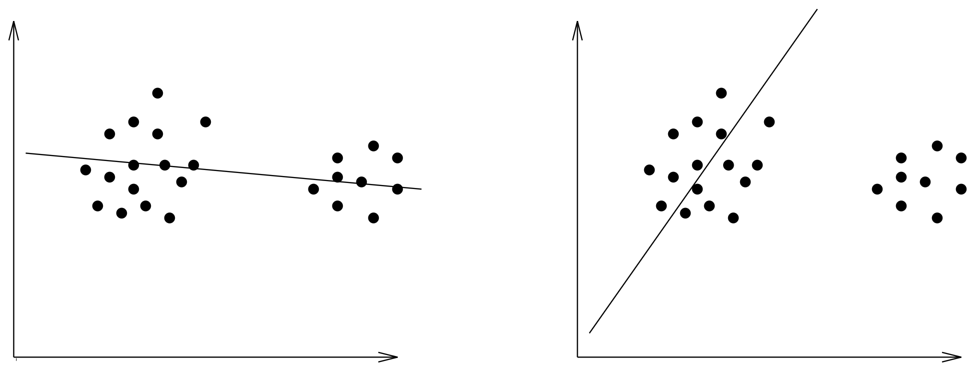

Figure 2.8: *a) (left): A least-squares fit through a cloud of points. b) (right): A robust fit through the same cloud of points. The influence function of the Lorentzian probability distribution tends to zero rapidly for deviations from the mean. These are considered outliers [23].*

Multiple patterns of motion within a single image region also arise from partial transparency of occluding surfaces. Bergen *et al.* [21] present an algorithm for estimating up to two different motions within a single intensity neighbourhood. The algorithm uses the following steps: let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two distinct velocities within an arbitrary image region. An iterative process is applied for estimating $\mathbf{v}_1$ and warping the corresponding image region on the next two frames to compute two difference images, D_1 and D_2, used in turn to estimate $\mathbf{v}_2$. If $\mathbf{v}_1$ is a reasonable estimate of one velocity pattern then the residual intensity structure in D_1 and D_2 reflects the velocity $\mathbf{v}_2$. The algorithm is iterated until the estimates $\mathbf{v}_1$ and $\mathbf{v}_2$ stabilize. It is generally sufficient to assume $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}$ initially, if no *a priori* knowledge is available. A least-squares method is employed for solving $\mathbf{v}_1$ and $\mathbf{v}_2$, using (1.3).

Other approaches for the measurement of multiple motions exist: the distribution of motion patterns may be regarded as a superposition of data distributions. Shizawa and Mase [132] apply a superposition principle to multiple motions and show that existing algorithms for optical flow, 3-d motion and structure, *etc.* can be generalized to handle many motion distributions. Similarly, multiple motions can be thought of as a set of layers, each describing a particular motion, over a particular domain. Techniques to separate these layers have been proposed by many authors: Darrell and Pentland explicitly take the support of homogeneous regions into account by using a multi-layer, cooperative robust estimation framework. Jespon and Black use an expectation-maximization (EM) algorithm to group a wide variety of component velocities into a fixed number of layers. Irani *et al.* [72] determine a dominant motion in an image using a least-squares approach and then group and segment the outlying motions. Their approach assumes that there is only one dominant motion and many outlying motions, each of which is assumed to correspond to independently moving objects. Adiv [3] uses a Hough transform on a pre-computed optical flow field to group regions having velocities consistent with roughly planar surfaces. This grouping is based on finding neighbouring velocities sharing the same affine transformations. Negahdaripour and Lee [113] present a segmentation process based on a hierarchical clustering method that does not assume a pre-computed optical flow field. They fit an affine model to small regions of the image and then repeatedly merge neighbouring regions based on similarity of their affine parameters. Then, given two sufficiently large planar regions, a motion and structure calculation can be performed. Wang and Adelson [153] use a clustering algorithm to group velocities into layers, each consistent with an affine motion. Bober and Kittler [28] use a block-based Hough transform in a robust estimation framework (redescending kernels) to obtain robust velocity estimates, including multiple motions, by clustering coherent motions at the same time the motion estimation is performed. Two confidence measures based on support functions are also proposed. In addition, hierarchical frameworks are known to separate motion components with respect to spatiotemporal frequencies. Burt *et al.* [32] suggest that multiple motions could be handled separately using different spatiotemporal frequency channels.

2.4.3 Parametric Models

Parametric models generally describe image motion with bivariate polynomials of varying order in the image coordinates and provide strong constraints on motion, which usually results in the accurate inference of optical flow [27]. These models possess desirable qualities: the motion of large image regions may be described with a single set of parameters, due to the increased flexibility of representation. In addition, parametric models are adequate for the description of discontinuous optical flow as each segmented region may be described with a particular set of motion parameters.

Bergen *et al.* [21, 20] consider the computation of optical flow from the viewpoint of image registration: given an image sequence, the parameters that best align an image with the next in the sequence are to be computed. This framework unifies many of the approaches already surveyed. In all cases, a function is to be minimized with respect to different parameters modelling velocity. Their algorithm has four basic components: pyramid construction, motion estimation, image warping and coarse-to-fine refinement. A Laplacian pyramid is used to hierarchically represent the image data [31] and motion estimation is performed by SSD minimization with respect to a particular model of motion. Image warping uses the current parameter values to compute an optical flow field at time t and then reconstructs another image at time t from the image at time $t-1$. Image reconstruction is performed using bilinear interpolation. The warped image is then compared with the original image and an error measure, based on image difference, is minimized with the Gauss-Newton method. The last component is a propagation of motion estimates from one level in the pyramid to the next lower level where they are used as initial guesses for the iterative refinement.

Using Bergen *et al.*'s classification, the optical flow methods presented above can be thought of as parametric, quasi-parametric or non-parametric [20, 21]. Parametric models fully describe the individual motion with a bivariate equation. For example, affine models approximate image velocity as:

$$\begin{aligned} u(x,y) &= a_1(x,y) + a_2x + a_3y \\ v(x,y) &= a_4(x,y) + a_5x + a_6y \end{aligned} \tag{2.48}$$

which is reasonably valid when surfaces are far from the observer or the image region under analysis is small. This models optical flow as a superposition of uniform motion and rotation, dilation and shear. This is the model used by Fleet and Jepson [48, 47] to integrate component velocities in local neighbourhoods. Adiv [3] also uses this model to segment and fit velocity measurements to local planar patches in the first stage of his algorithm. Spetsakis [139] uses an affine flow model and a hierarchy of Gabor filters. A second model assumes planarity of local surfaces:

$$\begin{aligned} u(x,y) &= a_1(x,y) + a_2 x + a_3 y + a_7 x^2 + a_8 xy \\ v(x,y) &= a_4(x,y) + a_5 x + a_6 y + a_7 xy + a_8 y^2, \end{aligned} \tag{2.49}$$

which is the *velocity functional method* of Waxman and Wohn [155] which they extended to include 2^{nd}-order curved surfaces. In both (2.48) and (2.49) the a_i's are neighbourhood center velocities or 1^{st} and 2^{nd}-order velocity derivatives. These parameters completely describe a planar surface velocity field. Any constant velocity model, such as Lucas and Kanade's [95] is also an example of a parametric model, although very simple. [21, 20]. Usually, the order of a parametric model describes its applicability for large image regions.

Non-parametric models are those typically used in global optical flow recovery. Horn and Schunck's global smoothness [70] or Nagel's oriented smoothness constraint [108] are examples of non-parametric models. Quasi-parametric models use a combination of parametric and non-parametric models. Bergen *et al.* place rigid motion models in this class. Rigid motion arises from rigidly moving scene objects under perspective projection. Direct motion and structure methods are examples of this model: Hanna [61, 62] shows that the rigidity assumption can be used to overcome the aperture problem in most cases. These parametric models are presented in a unified hierarchical framework [21]. The hierarchy yields increased computational efficiency and also allows for increased accuracy and robustness via coarse-to-fine refinements and image warping.

Black and Jepson [27] determine coarse optical flow via a correlation method [26] and fit parametric models to segmented regions of the image by hypothesizing local planarity and using coarse velocities to perform segmentation. Standard area-based regression techniques further refine these motion estimates. Deviations from planarity are modelled by allowing local deformations in the motion estimates. Hence, their approach does not try to fit a single parametric model to the whole image, but many parametric models to individual segmented regions. Also, Haddadi and Kuo [58] propose a parametric smoothness model that decomposes optical flow into irrotational and solenoidal fields and imposes a smoothness constraint on each field separately. Parameters are iteratively improved and smoothing across motion boundaries is avoided.

2.5 Temporal Refinement Methods

Most methods presented above for computing optical flow do not incorporate motion estimates from previous calculations within an image sequence being acquired: given two or more images, optical flow is computed only for one of the images. Recently, there has been some interest in incremental computation of optical flow [23, 25, 135, 50, 36]. The advantages include instantaneous access to optimal velocity estimates, accuracy improvement as the integration of optical flow over time is performed, computational efficiency gained by updating the estimates with the current frame and the ability to adapt to discontinuous optical flow as the observer or scene objects abruptly vary their motions.

Black's [23] algorithm may also be viewed as an incremental model because it minimizes an objective function that incorporates conservation of image intensity and spatiotemporal coherence in a robust estimation framework. Temporal continuity allows prediction of the next image velocity, assuming uniform acceleration. Warping with bilinear interpolation is used to estimate the acceleration. Black and Anandan [25] use a Markov Random field (MRF) method in a reformulation of Black's approach. The MRF algorithm is parallel, local and detects occlusion boundaries in an incremental fashion but, as formulated, can only handle integer motions. Singh [135] uses a Kalman filter to integrate velocity estimates computed by a hierarchical correlation method [134]. The Kalman filter reduces the uncertainty of the estimates over time. This framework also detects occlusion boundaries. Fleet and Langley [49] use low-pass recursive filters to produce and update gradient-based velocity estimates from a sequence of images. Chin *et al.* [36] present an extension of Horn and Schunck's dense optical flow algorithm [70] that uses a temporal coherence constraint to produce near optimal, recursive flow estimates from multiple frames.

3 Discussion

Although many methods and strategies have appeared, the estimation of image motion remains a challenging task: to date, except in limited circumstances, no technique is able to generate sufficiently accurate and dense optical flow fields to allow the general recovery of motion and scene parameters in a realistic environment. In fact, motion and structure algorithms need very accurate optical flow to carry useful 3-d motion and structure computations [16]. Also needed are accurate means of determining the reliability of computed image velocities. Such reliability measures have been proposed: covariance matrix eigenvalues [133, 136], Gaussian curvature [152], principle curvature [8, 65], spatial gradient [14], eigenvalues of a least-squares matrix [133] and support function values [28]. These confidence measures allow for thresholding, yielding more accurate but sparser optical flow fields. They may also be integrated in subsequent processing, such as weights in a least-squares motion and structure calculation.

Of importance in an accurate estimation of image motion in the surveyed methods is the requirement for appropriate spatiotemporal sampling rates, in order to compute accurate spatiotemporal derivatives for differential-based methods, to reduce the search areas for matching-based methods or to limit the amount of aliasing when estimating optical flow with frequency-based filtering methods. Too often, the assumption that imagery is free of aliasing effects is made. Conventional cameras usually produce imagery with severe temporal aliasing, especially for significant image motions. Reducing aliasing effects may be accomplished by increasing temporal sampling rates, image prefiltering or by using hierarchical processing. Of course, increased temporal sampling rates lead to more accurate optical flow computations. However, for a number of reasons, including small temporal support (only a few images) or fast image motion, such appropriately sampled imagery is not always available. In such cases, accurate temporal derivatives may be difficult to obtain and hierarchical matching-based methods seem to be a natural choice. It has also been observed that some prefiltering of the image sequence prior to the extraction of basic image motion measurements, such as intensity derivatives or correlation surfaces, significantly increases the accuracy of results [14]. For instance, a spatiotemporal Gaussian smoothing of the image sequence results in more accurate derivatives for the methods of Lucas and Kanade [94, 95] and Horn and Schunck [70]. Anandan's and Singh's computational schemes [8, 136] also use prefiltering of the images by computing hierarchical Laplacian images. It is believed that this high-pass filtering emphasizes image structures that are desirable for correlation.

Often, very restrictive assumptions about image motion are posed. For example, one of these assumptions requires neighbouring velocities arising from the relative motion of a single surface to be similar [70]. This requirement is usually imposed by applying isotropic smoothness constraints onto velocity estimates. However, only a few simple cases of realistic imagery exhibit continuous motion fields: realistic imagery, such as outdoor scenes, possess complex structures for which global single surface assumptions are inadequate. Attempts at estimating discontinuous image motion were proposed by Corneluis and Kanade [38] and Nagel [106, 108, 110] in the form of an inhibition of the smoothness requirement across intensity discontinuities. However, it is obvious that intensity discontinuities may not necessarily represent motion discontinuities [145]. The problem posed by occluding surfaces needs further investigation. Occlusion is a important source of visual information: optical flow at occlusion boundaries can be used to determine the direction of translation (the focus of expansion) [93] and segment the scene into independently moving objects [146], yet optical flow estimation at occlusions is problematic. Adequate approaches to handling occlusion include line approaches which explicitly model intensity discontinuities and prevent smoothing over them, layered or superposed parametric models, and mixed velocity distribution models assume the presence of usually two velocities and discriminate them according to some criteria. In addition, occlusion has been recently analytically described in Fourier space [19, 49].

Alternatively, optical flow may be estimated with local constraints only [21, 34, 48, 95, 152]. In these schemes, no smoothness requirements are imposed and motion discontinuities may be preserved. Of course, the accuracy at motion boundaries or at regions of transparency highly depends on the model of motion being used. For instance, single motion models are inadequate for handling occlusion and transparency properly. Nonetheless, the use of local constraints or parametric models may be more appropriate in general [14], as no arbitrary smoothness requirement is imposed on the structure of optical flow.

Lighting effects also constitute a problem in many image sequences. Constant scene illumination and Lambertian surface reflectance are either implicitly or explicitly assumed for most current optical flow methods that use the brightness constancy assumption. Although the effects of highlights, shadows and illumination conditions on the estimation of optical flow have only been studied to a limited degree, it is possible to partially compensate for these effects and estimate image motion as a geometric quantity if the characteristics of the light sources are known. Towards this, the use of multiple light sources [161] and sets of multispectral constraints on image motion [97] have been used. Shading effects have also been modelled [102].

Aside from lighting conditions, some surface reflectance phenomena also pose difficulties. For example, transparent surfaces usually lead to multiple motions while highlights may create false motions. Perhaps due to its difficulty and infrequency of occurrence, transparent motions have been mainly ignored and attempts at estimating multiple motions have just begun to appear [21, 78]. Methods using orientation and velocity sensitive filters may contribute to the solving of this particular problem, as they provide multiple measurements for each location [48]. Alternatively, superposition principles and layered motions [40, 132, 153] are promising frameworks. However, segmenting multiple motion distributions remains difficult if no *a priori* assumption is made on the number of distributions present within a support region.

Lastly, we would like to emphasize that much of the image motion literature presents flow field examples for a few image sequences, which can only be judged qualitatively. While the *theory* of optical flow computation is being addressed, the *practise* of optical flow is often neglected: far too little of the published work provide quantitative error analysis. Usually, only a qualitative comparison is possible. Even then, it is often difficult to assess which techniques are quantitatively better as authors typically use their favorite image sequences, which are not usually available to the community and for which the correct image motion is unknown. A widely available set of images for comparative testing is needed. These images should have known optical flow fields and allow a quantitative error analysis. This is especially the case for the newer work, as with the layered approaches to optical flow, where little or no quantitative analysis exists.

There are various means of performing quantitative error analysis when correct optical flow information is available: error can be expressed as absolute error, relative error, angle error [48], RMS or SNR ratios, allowing one to compare optical flows for the same image sequence. Furthermore, quantitative analysis is possible without motion information: RMS image reconstruction error has been used to measure error for real image sequences when the correct motion information is unavailable [41, 86, 96, 164, 105, 115, 164].

Acknowledgements This work has been supported in part by NSERC Canada, the Government 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. *Journ. Opt. Soc. Am.*, A 2(2):284–299, 1985.

[2] E. H. Adelson and J. R. Bergen. The early detection of motion boundaries. In *IEEE Proc. of Workshop on Visual Motion, Charleston*, pages 151–156, 1986.

[3] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE PAMI*, 7(4):384–401, 1985.

[4] J. K. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images - a review. *Proceedings of the IEEE*, 76(8):917–935, 1988.

[5] J. Aisbett. Optical flow with intensity-weighted smoothing. *IEEE PAMI*, 11(5):512–522, 1989.

[6] J. Aloimonos and C. M. Brown. Direct processing of curvilinear sensor motion from a sequence of perspective images. In *IEEE Proc. of 2nd Workshop on Computer Vision: Representation and Control*, pages 167–174, 1986.

[7] J. Aloimonos and I. Ristigous. Determining the 3d motion of a rigid planar patch without correspondence, under perspective projection. In *Workshop on Motion, Representation and Analysis*, pages 167–174. IEEE, 1986.

[8] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *IJCV*, 2:283–310, 1989.

[9] N. Ancona. A fast obstacle detection method based on optical flow. In *Proc. of ECCV*, pages 267–271, 1992.

[10] H. Barman. *Hierarchical Curvature Estimation in Computer Vision.* PhD thesis, Dept. Electrical Engineering, Univ. of Linkoping, 1991.

[11] H. Barman, L. Haglund, H. Knutsson, and G. H. Granlund. Estimation of velocity, acceleration and disparity in time sequences. In *IEEE Proc. of Workshop on Visual Motion*, pages 44–51, 1991.

[12] S. T. Barnard and W. B. Thompson. Disparity analysis of images. *IEEE PAMI*, 2(4):333–340, 1980.

[13] J. L. Barron and R. Eagleson. Motion and structure from time-varying optical flow. In *Vision Interface*, pages 104–111, May 1995.

[14] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.

[15] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt. Performance of optical flow techniques. In *IEEE Proc. of CVPR*, pages 236–242, Champaign, Illinois, June 1992.

[16] 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):239–269, 1990.

[17] J. L. Barron and A. Liptay. Optic flow to measure minute increments in plant growth (in press). *BioImaging*, 1994.

[18] R. Battiti, E. Amaldi, and C. Koch. Computing optical flow across multiple scales: An adaptive coarse-to-fine strategy. *IJCV*, 6(2):133–145, 1991.

[19] S. S. Beauchemin and J. L. Barron. The structure of occlusion in fourier space. In *Vision Interface*, pages 112–119, May 1995.

[20] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proc. of ECCV, Santa Margherita, Italy*, pages 237–252. Springer-Verlag, 1992.

[21] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg. Three-frame algorithm for estimating two-component image motion. *IEEE PAMI*, 14(9):886–896, 1992.

[22] M. J. Black. A robust gradient-method for determining optical flow. Technical Report YALEU/DCS/RR-891, Yale University, New-Haven, CT, 1991.

[23] M. J. Black. *Robust incremental optical flow.* PhD thesis, Yale University, 1992.

[24] M. J. Black and P. Anandan. A model for the detection of motion over time. In *Proc. of ICCV*, pages 33–37, 1990.

[25] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *IEEE CVPR*, pages 296–302, 1991.

[26] M. J. Black and P. Anandan. A framework for robust estimation of optical flow. In *Proc. of ICCV*, pages 231–236, 1993.

[27] M. J. Black and A. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. Technical Report SPL-94-053, Xerox Systems and Practices Laboratory, 1994.

[28] M. Bober and J. Kittler. Robust motion analysis. In *IEEE CVPR*, pages 947–952, 1994.

[29] P. Bouthemy and E. Francois. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *IJCV*, 10(2):159–182, 1993.

[30] P. Burlina and R. Chellappa. Time-to-x: Analysis of motion through temporal parameters. In *CVPR*, pages 461–468, 1994.

[31] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Trans. on Communications*, 31:532–540, 1983.

[32] P. J. Burt, R. Hingorani, and R. J. Kolczynski. Mechanisms for isolating component patterns in the sequential analysis of multiple motion. In *IEEE Proc. of Workshop on Visual Motion*, pages 187–194, 1991.

[33] 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):59–75, 1984.

[34] M. Campani and A. Verri. Computing optical flow from an overconstrained system of linear equations. In *Proc. of ICCV*, pages 22–26, 1990.

[35] B. Carpentieri and J. A. Storer. A split-merge parallel block-matching algorithm for video displacement estimation. In *Data Compression Conference*, pages 239–248, 1992.

[36] T. M. Chin, W. Karl, and A. Willsky. Probabilistic and sequential computation of optical flow using temporal coherence. *IEEE Trans. Image Processing*, pages 773–788, 1994.

[37] C. H. Chu and E. J. Delp. Estimating displacement vectors from an image sequence. *J. Opt. Soc. Am.*, A 6(6):871–878, 1989.

[38] N. Cornelius and T. Kanade. Adapting optical flow to measure object motion in reflectance and x-ray image sequences. In *Proc. of ACM Siggraph/Sigart Interdisciplinary Workshop on Motion*, pages 50–58, 1983.

[39] V. Cornilleau-Peres and J. Droulez. Stereo correspondence from optical flow. In *Proc. of ECCV*, pages 326–330, 1990.

[40] T. Darrell and A. Pentland. Robust estimation of a multi-layered motion representation. In *IEEE Proc. of Workshop on Visual Motion*, pages 173–178, 1991.

[41] E. Dubois. The sampling and reconstruction of time-varying imagery with application in video systems. *Proc. of IEEE*, 73(4):502–522, 1985.

[42] J. H. Duncan and T. Chou. On the detection of motion and the computation of optical flow. *IEEE PAMI*, 14(3):346–352, 1992.

[43] R. Dutta, R. Manmatha, L. Williams, and E. M. Riseman. A data set for quantitative motion analysis. In *IEEE Proc. of CVPR*, pages 159–164, San Diego, California, 1989.

[44] W. Enkelmann. Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences. In *IEEE Proc. of Workshop on Motion: Representation and Analysis*, pages 81–87, 1986.

[45] W. Enkelmann. Obstacle detection by evaluation of optical flow fields from image sequences. In *Proc. of ECCV*, pages 134–138, 1990.

[46] I. Fermin and A. Imiya. Two-dimensional motion computation by randomized method. Technical Report TR ICS-4-6-1994, Dept. of Information and Computer Sciences, Chiba University, 1994.

[47] D. J. Fleet. *Measurement of Image Velocity.* Kluwer Academic Publishers, Norwell, 1992.

[48] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *IJCV*, 5(1):77–104, 1990.

[49] D. J. Fleet and K. Langley. Computational analysis of non-fourier motion. *Vision Research*, 34(22):3057–3079, 1995.

[50] D. J. Fleet and K. Langley. Recursive filters for optical flow. *IEEE PAMI*, 17(1), 1995.

[51] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *PAMI*, 6(6):721–741, November 1984.

[52] A. Giachetti, M. Campani, and V. Torre. The use of optical flow for autonomous navigation. In *ECCV*, pages 146–151, 1994.

[53] F. C. Glazer. Computation of optical flow by multilevel relaxation. Technical Report COINS-TR-87-64, Univ. of Mass., 1987.

[54] F. C. Glazer. Hierarchical gradient-based motion detection. In *DARPA Proc. of Image Understanding Workshop*, pages 733–748, Los Angeles, California, 1987.

[55] S. Gong and M. Brady. Parallel computation of optic flow. In *Proc. of ECCV*, pages 124–133, 1990.

[56] N. M. Grzywacz and A. L. Yuille. A model for the estimate of local velocity by cells in the visual cortex. *Proc. R. Soc. Lond.*, B 239:129–161, 1990.

[57] N. Gupta, S. Raghavan, and L. Kanal. Robust recovery of image motion. In *ACCV*, 1993.

[58] N. Haddadi and C.-C. J. Kuo. Computation of dense optical flow with a parametric smoothness model. Technical Report USC-SIPI Report 223, Dept. of Electrical Engineering-Systems, University of Southern California, 1992.

[59] L. Haglund. *Adaptive Multidimensional Filtering.* PhD thesis, Dept. Electrical Engineering, Univ. of Linkoping, 1992.

[60] P. Handschack and R. Klette. Evaluation of differential methods for image velocity measurement. *Computer and Artificial Intelligence (to appear)*, 1995.

[61] K. J. Hanna. Direct multi-resolution estimation of ego-motion and structure from motion. In *IEEE Motion Workshop*, pages 156–162, 1991.

[62] K. J. Hanna and N. E. Okamoto. Combining stereo and motion for direct estimation of scene structure. In *ICCV*, pages 357–365, 1993.

[63] R. M. Haralick and J. S. Lee. The facet approach to optic flow. In *DARPA proc. of Image Understanding Workshop*, pages 84–93, Arlington, Virginia, 1983.

[64] J. C. Hay. Optical motions and space perception: An extension of gibson's analysis. *Psychological Review*, 73(6):550–565, 1966.

[65] D. J. Heeger. Optical flow using spatiotemporal filters. *IJCV*, 1:279–302, 1988.

[66] D. J. Heeger and A. D. Jepson. Subspace methods for recovering rigid motion 2: Algorithm and implementation. *IJCV*, 7(2):95–117, 1992.

[67] J. Heel. Direct estimation of structure and motion from multiple frames. Technical Report 1190, MIT AI Meno, 1990.

[68] E. C. Hildreth. The computation of the velocity field. *Proc. R. Soc. Lond.*, B 221:189–220, 1984.

[69] B. K. P. Horn. Motion fields are hardly ever ambiguous. *IJCV*, 1:259–274, 1987.

[70] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–204, 1981.

[71] B. K. P. Horn and E. J. Weldon. Computationally efficient methods for recovering translational motion. In *Proceedings of the 1st International Conference on Computer Vision*, pages 2–11, London, England, 1987.

[72] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *IJCV*, 12(1):5–16, 1994.

[73] M. Irani, B. Rousso, and S. Peleg. Recovery of egomotion using image stabilization. In *CVPR*, pages 454–460, 1994.

[74] B. Jahne. Motion determination in space-time images. In *Proc. of ECCV*, pages 161–173, 1990.

[75] R. C. Jain. Direct computation of the focus of expansion. *IEEE PAMI*, 5(1):58–63, 1983.

[76] R. C. Jain. Segmentation of frame sequences obtained by a moving observer. *IEEE PAMI*, 6(5):624–629, 1984.

[77] M. R. M. Jenkin, A. D. Jepson, and J. K. Tsotsos. Techniques for disparity measurement. *CVGIP*, 53(1):14–30, 1991.

[78] A. D. Jepson and M. Black. Mixture models for optical flow computation. In *IEEE CVPR*, pages 760–761, 1993.

[79] D. S. Kalivas and A. A. Sawchuk. A region matching motion estimation algorithm. *CVGIP*, 54(2):275–288, 1991.

[80] 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(2):229–244, 1987.

[81] H. Kirchner and H. Niemann. Finite element method for determination of optical flow. *Pattern Recognition*, 13(2):131–141, 1992.

[82] C. Koch, H. T. Wang, R. Battiti, B. Mathur, and C. Ziomkowski. An adaptive multi-scale approach for estimating optical flow: Computational theory and physiological implementation. In *IEEE Proc. of Workshop on Visual Motion*, pages 111–122, 1991.

[83] C. Koch, H. T. Wang, B. Mather, A. Hsu, and H. Suarez. Computing optical flow in resistive networks and in the primate visual system. In *IEEE Proc. of Workshop on Visual Motion*, 1989.

[84] R. Kories and G. Zimmerman. A versatile method for the estimation of displacement vector fields from image sequences. In *IEEE Proc. of Workshop on Motion: Representation and Analysis*, pages 101–106, 1986.

[85] K. Langley, T. J. Atherton, R. G. Wilson, and M. H. E. Lacombe. Vertical and horizontal disparities from phase. *Image and Vision Computing*, 9(4):296–302, 1991.

[86] T. Lin and J. L. Barron. Image reconstruction error for optical flow. In *Vision Interface, Banff National Park (to appear)*, May 1994.

[87] A. Liptay, J. L. Barron, T. Jewett, and I. Van Wesenbeeck. Optic flow as an ultra-sensitive technique for measuring seedling growth in long image sequences. *submitted*, 1994.

[88] J. J. Little. Accurate early detection od discontinuities. In *Vision Interface*, pages 2–7, 1992.

[89] J. J. Little, H. H. Butlhoff, and T. Poggio. Parallel optical flow using local voting. In *Proc. of ICCV*, pages 454–459, 1988.

[90] H. Liu, M. Herman T.-H. Hong, and R. Chellappa. A reliable optical flow algorithm using 3-d hermite polynomials. Technical report, National Institute of Standards and Technology (NIST), 1993.

[91] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 223:133–135, 1981.

[92] H. C. Longuet-Higgins. The visual ambiguity of a moving plane. *Proc. R. Soc. Lond.*, B223:165–175, 1984.

[93] H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proc. R. Soc. Lond.*, B 208:385–397, 1980.

[94] B. D. Lucas. *Generalized image matching by the method of differences.* PhD thesis, Carnegie-Mellon Univ., 1984.

[95] B. D. Lucas and T. Kanade. An iterative image-registration technique with an application to stereo vision. In *DARPA Proc. of Image Understanding Workshop*, pages 121–130, 1981.

[96] M. R. Luettgen, W. C. Karl, and A. S. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Trans. on Image Processing*, 3(1):41–64, January 1994.

[97] V. Markandey and B. E. Flinchbaugh. Multispectral constraints for optical flow computation. In *Proc. of ICCV*, pages 38–41, 1990.

[98] D. Marr and E. C. Hildreth. Theory of edge detection. *Proc. R. Soc. Lond.*, B 290:187–212, 1980.

[99] E. De Micheli, V. Torre, and S. Uras. The accuracy of the computation of optical flow and of the recovery of motion parameters. *IEEE PAMI*, 15(5):434–447, 1993.

[100] A. Mitiche, Y. F. Wang, and J. K. Aggarwal. Experiments in computing optical flow with the gradient-based, multiconstraint method. *Pattern Recognition*, 20(2):173–179, 1987.

[101] F. W. Mounts. A video encoding system using conditional picture-element replenishment. *Bell System Technical Journal*, 48:2545–2554, 1969.

[102] N. Mukawa. Estimation of shape, reflection coefficients and illuminant direction from image sequences. In *ICCV*, 1993.

[103] D. W. Murray and B. F. Buxton. Reconstructing the optic flow from edge motion: An examination of two different approaches. *1st Conference on AI Applications*, pages 382–388, 1984.

[104] D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE PAMI*, 9(2):220–228, 1987.

[105] H. G. Musmann, P. Pirsch, and H. J. Grallert. Advances in picture coding. *Proc of IEEE*, 73(4):523–548, 1985.

[106] H.-H. Nagel. Constraints for the estimation of vector fields from image sequences. In *IJCAI*, pages 945–951, Karlsruhe, Federal Republic of Germany, August 1983.

[107] H.-H. Nagel. Displacement vectors derived from second-order intensity variations in image sequences. *CVGIP*, 21:85–117, 1983.

[108] H.-H. Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33:299–324, 1987.

[109] H.-H. Nagel. On a constraint equation for the estimation of displacement rates in image sequences. *IEEE PAMI*, 11(1):13–30, 1989.

[110] H.-H. Nagel. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optic flow. In *Proc. of ECCV*, pages 139–148, 1990.

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

[112] S. Negahdaripour and B. K. P. Horn. Direct passive navigation. *IEEE PAMI*, 9(1):168–176, 1987.

[113] S. Negahdaripour and S. Lee. Motion recovery from image sequences using only first order optical flow information. *IJCV*, 9(3):163–184, 1992.

[114] S. Negahdaripour and C. H. Yu. A generalized brightness change model for computing optical flow. In *ICCV*, pages 2–11, 1993.

[115] A. N. Netravali and J. D. Robbins. Motion compensated television coding: Part 1. *Bell System Technical Journal*, 58:631–670, 1979.

[116] M. Ogata and T. Sato. Motion-detection model with two stages: Spatiotemporal filtering and feature matching. *J. Opt. Soc. Am.*, A 9(3):377–387, 1992.

[117] M. Okutomi and T. Kanade. A locally adaptive window for signal matching. In *Proc. of ICCV*, pages 190–199, 1990.

[118] I. Overington. Gradient-based flow segmentation and location of the focus of expansion. *Alvey Vision Conference*, pages 860–870, 1987.

[119] J. A. Perrone. Simple technique for optical flow estimation. *J. Opt. Soc. Am*, A 7(2):264–277, 1990.

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

[121] K. Prazdny. Motion and structure from optical flow. In *IJCAI*, pages 702–704, 1979.

[122] K. Prazdny. Detection of binocular disparities. *Biological Cybernetics*, 52:93–99, 1985.

[123] J. L. Prince and E. R. McVeigh. Motion estimation from tagged mr image sequences. *IEEE Trans. on Medical Images*, 11(2):238–249, 1992.

[124] D. Regan and K. I. Beverley. How do we avoid confounding the direction we are looking and the direction we are moving. *Science*, 215:194–196, 1982.

[125] W. Reichardt, R. W. Schlogl, and M. Egelhoaf. Movement detectors of the correlation type provide sufficient information for local computation of 2d velocity fields. *Naturwissenschaften*, 75:313–315, 1988.

[126] A. Rognone, M. Campani, and A. Verri. Identifying multiple motions from optical flow. In *Proc. of ECCV*, pages 258–266, 1992.

[127] C. Schnorr. Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *IJCV*, 6(1):25–38, 1991.

[128] C. Schnorr. Computation of discontinuous optical flow by domain decomposition. *IEEE PAMI*, 8(2):153–165, 1992.

[129] B. G. Schunck. Image flow: Fundamentals and future research. In *IEEE Proc. of CVPR*, pages 560–571, San Francisco, California, 1985.

[130] B. G. Schunck. Image flow segmentation and estimation by constraint line clustering. *IEEE PAMI*, 11(10):1010–1027, 1989.

[131] G. L. Scott. Four-line method of locally estimating optic flow. *Image and Vision Computing*, 5(2):67–72, 1987.

[132] M. Shizawa and K. Mase. Principle of superposition: A common computational framework for analysis of multiple motion. In *IEEE Proc. of Workshop on Visual Motion*, pages 164–172, 1991.

[133] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *IEEE Proc. of CVPR*, pages 310–315, 1991.

[134] A. Singh. An estimation-theoretic framework for image flow computation. In *Proc. of ICCV*, pages 168–177, Osaka, Japan, 1990.

[135] A. Singh. Incremental estimation of image flow using a kalman filter. In *IEEE Proc. of Workshop on Visual Motion*, pages 36–43, 1991.

[136] A. Singh. *Optic flow computation: A unified perspective.* IEEE Computer Society Press, 1992.

[137] P. Sobey and M. V. Srinivasan. Measurement of optical flow by a generalized gradient scheme. *J. Opt. Soc. Am.*, A 8(9):1488–1498, 1991.

[138] L. A. Spacek. Edge detection and motion detection. *Image and Vision Computing*, 4(1):43–56, 1986.

[139] M. E. Spetsakis. An optical flow estimation algorithm that uses gabor filters and affine model for flow. Technical report, Dept. of Computer Science, York University, 1994.

[140] M. V. Srinivasan. Generalized gradient schemes for the measurement of two-dimensional image motion. *Biological Cybernetics*, 63:421–431, 1990.

[141] M. Subbarao. Bounds on time-to-collision and rotational component from first-order derivatives of image flow. *CVGIP*, 50:329–341, 1990.

[142] M. Subbarao and A. M. Waxman. On the uniqueness of image flow solutions for planar surfaces in motion. In *3rd Workshop on Computer Vision: Representation and Control*, pages 129–140, 1985.

[143] V. Sundareswaran. A fast method to estimate sensor translation. In *ECCV*, pages 263–257, 1992.

[144] M. A. Sutton, W. J. Walters, W. H. Peters, W. F. Ranson, and S. R. McNeil. Determination of displacement using an improved digital correlation method. *Image and Vision Computing*, 1(3):133–139, 1983.

[145] W. B. Thompson, K. M. Mutch, and V. A. Berzins. Dynamic occlusion analysis in optical flow fields. *IEEE PAMI*, 7(4):374–383, 1985.

[146] W. B. Thompson and T. Pong. Detecting moving objects. *IJCV*, 4(1):39–57, 1990.

[147] M. Tistarelli and G. Sandini. Estimation of depth from motion using anthropomorphic visual sensor. *Image and Vision Computing*, 8:271–278, 1990.

[148] O. Tretiak and L. Pastor. Velocity estimation from image sequences with second-order differential operators. In *IEEE Proc. of ICPR*, pages 20–22, Montreal, Quebec, 1984.

[149] R. Y. Tsai and T. S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE PAMI*, 6(1):13–27, 1984.

[150] R. Y. Tsai, T. S. Huang, and W. Zhu. Estimating three-dimensional motion parameters of a rigid planar patch 2: Singular value decomposition. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 30(4):525–534, 1982.

[151] S. Ullman. *The interpretation of visual motion.* MIT Press, Cambridge, London, 1979.

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

[153] J. Y. A. Wang and E. H. Adelson. Representing moving images with layers. *IEEE Trans. Image Processing*, 3(5):625–638, 1994.

[154] A. B. Watson and A. J. Ahumada Jr. Model of human visual motion sensing. *J. Opt. Soc. Am.*, A 2(2):322–341, 1985.

[155] A. M. Waxman and K. Wohn. Contour evolution, neighborhood deformation, and global image flow: planar surfaces in motion. *Int. J. of Robotics Res.*, 4(3):95–108, 1985.

[156] A. M. Waxman, J. Wu, and F. Bergholm. Convected activation profiles and the measurement of visual motion. In *IEEE Proc. of CVPR*, pages 717–723, 1988.

[157] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. In *Proc. of ICCV*, pages 12–20, 1993.

[158] J. Weng. A theory of image matching. In *Proc. of ICCV*, pages 200–209, 1990.

[159] G. Whitten. A framework for adaptive scale space tracking solutions to problems in computer vision. In *Proc. of ICCV*, pages 210–220, 1990.

[160] D. Willick and Y. Yang. Experimental evaluation of motion constraint equations. Technical Report 89-4, Dept. of Computational Science, Univ. of Saskatchewan, 1989.

[161] R. J. Woodham. Multiple light source optical flow. In *Proc. of ICCV*, pages 42–46, 1990.

[162] Y. Xiong and S. A. Shafer. Moment and hypergeometric filters for high precision computation of focus, stereo and optical flow. Technical Report CMU-RI-TR-94-28, Dept. of Computer Science, Carnegie-Mellon Univ., 1994.

[163] Z. Zhang and O. D. Faugeras. Three-dimensional motion computation and object segmentation in a long sequence in stereo frames. *IJCV*, 7(3), 1992.

[164] H. Zheng and S. D. Blostein. An error-weighted regularization algorithm for image motion-field estimation. *IEEE Trans. on Image Processing*, 2(2):246–252, 1993.

[165] Q. Zheng and R. Chellappa. Automatic feature point extraction and tracking in image sequences for unknown image motion. In *ICCV*, pages 335–339, 1993.

[166] H. Zinner. Determining the kinematic parameters of a moving imaging sensor by processing spatial and temporal intensity changes. *J. Opt. Soc. Am*, A 3(9):1512–1517, 1986.

[167] M. D. Zoltowski. Signal processing applications of the method of total least squares. In *IEEE 21st Annual Asilomar Conf. on Signals, Systems and Computers*, pages 290–296, 1987.