



Motion estimation based on the direction of intensity gradient

Pierre-Yves Burgi*

Swiss Center for Electronics and Microtechnology, 2007 Neuchâtel, Switzerland

Received 2 January 2002; received in revised form 7 January 2004; accepted 15 January 2004

Abstract

The optical flow constraint (OFC) equation has been extensively studied in computer vision to estimate motion from image sequences. Traditionally the OFC relies on spatio-temporal brightness variations caused by motion. However, many other features are in principle appropriate, including gradient directions of the image intensity. Given gradient directions are more tolerant to changes in lighting, they seem to be an adequate choice for computing optical flows. Their applicability in the OFC is, however, not straightforward as the gradient direction is independent of the gradient magnitude, and is not defined in homogeneous areas. To palliate to these difficulties, a form of OFC equation based on probability distributions of gradient directions is proposed. The performance of the approach is assessed by experiments realized on synthetic and real world images.

© 2004 Elsevier B.V. All rights reserved.

PACS: 07.05.Pj; 42.30.Tz; 02.50.Cw

Keywords: Motion field; Optical flow constraint; Histogramming; Intensity gradient direction; Parametric estimation

1. Introduction

Motion estimation aims at recovering 3-D velocity fields of moving objects. One possible approach is to base this estimation on the spatio-temporal variations in magnitude of some particular image features. However, because those image features are generally obtained from the projection of moving objects onto 2-D optical sensors, motion estimation is often an under-constrained problem. Thus, in most cases motion estimation is instead concerned with the determination of 2-D motion fields, which represent the projection of 3-D velocity fields onto a plane. Because of the way light is reflected by the surface of the objects, most often than not the reconstruction of such motion fields cannot be done exactly. Consequently, motion estimation is generally limited to compute the so-called optical flow, which provides a set of motion vectors resembling as closely as possible the motion field.

A large amount of literature is devoted to the reconstruction of optical flows. While a wide variety of methods has been proposed to perform such a reconstruction

(as reviewed for instance in Ref. [1], or more recently in Ref. [2]), this paper focuses specifically on differential techniques. Such techniques have originally been based on the spatio-temporal variations in image brightness, which led to the well-known optical flow constraint (OFC) equation [3,4]:

$$\vec{\nabla} E \cdot \vec{v} + \frac{\partial E}{\partial t} = 0, \quad (1)$$

where E stands for the image brightness, \vec{v} is the optical flow, and $\vec{\nabla}$ is the two-dimensional spatial gradient operator. This equation expresses the invariance of E over time, which is strictly valid only for constant lightings illuminating the scene, and for translating Lambertian surfaces (a rotating homogeneous Lambertian sphere might as well yield no variations in E). Any deviations from these conditions will undermine this invariance and thus make the optical flow computed by the OFC equation disagree with the true motion field [5].

Unless one is only interested in recovering the velocity perpendicular to the edges, the determination of the optical flow is further impaired by the fact that Eq. (1) poses a single local constraint on the two-dimensional optical flow (in short, there is one equation for two unknowns).

* Corresponding author. Present address: IT Division, University of Geneva, rue du Général-Dufour 24, CH-1211 Genève 4, Switzerland.

E-mail address: pierre-yves.burgi@adm.unige.ch (P.-Y. Burgi).

This problem is commonly referred to as the aperture problem [6]. Two main techniques have been devised to solve this problem (for a review of these various techniques, see, e.g. [1,7]). The first technique, known as regularization, consists in imposing a smoothness constraint on the global optical flow [4,8,9]. The second technique resorts to a least-squares fit of the local (first-order) OFC equation to a constant model for \tilde{v} in a small spatial neighborhood [10–12]. This latter technique is known as the multi-constraint method. When combined to local regularization, the multi-constraint method has been shown to be particularly efficient in estimating motion when several objects are moving simultaneously in the scene [13].

The objective of the present work is to reformulate motion estimation using gradient directions as the invariant. This choice is motivated by their invariance to global light changes, which typically occur in artificial fluorescent lightings and by their relative insensitivity to changes in illumination direction, a property recently demonstrated on real scenes [14]. The capacity of gradient direction in estimating image motion is assessed quantitatively and qualitatively on a wide range of image sequences. The optical flow is determined according to a multiconstraint-based approach. Results are promising and provide for the first time the evidence that gradient directions can indeed be used to estimate image motion with good accuracy.

The remainder of the paper is organized as follows: Section 2 presents the theory of the motion estimation problem based on gradient directions, while the details of the computer implementation of the method are given in Section 3. Experimental results are then presented in Section 4, followed by a discussion in Section 5. Finally, the main conclusions are summarized in Section 6.

2. Theory

2.1. Motion equation for gradient directions

The OFC equation has originally been proposed on the assumption that intensity is conserved. Taking instead the gradient direction as the conserved quantity is possible on the ground that such a feature is also invariant to motion. The gradient direction (θ) is obtained through the application of the operator that maps E (the luminance) to $\arctan(E_y/E_x)$, where E_x and E_y denote the spatial derivatives in the horizontal and vertical directions, respectively. It can readily be shown that such an operator pertains to the class of the so-called ‘planar rigid motion invariant’ (PRMI) operators. These operators are such that the result of the operator on a rigidly moved image is the same as that obtained by moving the result of the operator on the original image, a requisite for using the OFC equation [15]. Substituting E in Eq. (1) with the gradient directions,

θ , results into the motion equation

$$\frac{\partial}{\partial t} \theta(x, y, t) = -\tilde{v} \theta(x, y, t) \cdot \tilde{v}(x, y, t), \quad (2)$$

where $\theta = \arctan(E_y/E_x)$. This equation involves second-order derivatives with respect to the image intensity. While second-order derivatives have been considered in previous works [8,16,17], it must nevertheless be acknowledged that Eq. (2) leads to intriguing difficulties. Such difficulties arise from the fact that gradient directions specify angle values in range $[0, 2\pi[$, and apart from the extra precaution one must apply when computing these derivatives (difference between two angles cannot exceed π), it must also be recognized they are not defined in regions with homogeneous image intensity. In practice, image regions are rarely perfectly homogeneous. Small variations in image intensity occurring over the object’s surface is rather the rule, and therefore the status of the gradient directions goes from undefined to noisy, which does not necessarily improve the situation for estimating motion, particularly if large angle variations occur between two successive frames. Furthermore, considering straight edges (previously smoothed out for regularizing the computation of E_x and E_y), there will be no variations in the edges’ orientation not only in a direction parallel to them (which is to be expected), but also perpendicular to them. This latter invariance comes from the fact that gradient direction is independent of gradient magnitude, and thus even though the luminance is varying perpendicularly to the edge, the direction of these variations remains constant. Consequently, there seems to be no possibility to recover the optical flow, or altogether its component in the direction of the spatial gradient, in the vicinity of such straight edges based on Eq. (2). As it is shown in the following sections, a solution for computing the optical flow based on gradient directions can be found by embedding the motion estimation problem within a statistical framework.

2.2. Statistical framework for motion estimation

In this section, a statistical framework is introduced as a means to circumvent the above-mentioned difficulties appearing when computing the gradient directions in homogeneous regions and at proximity of straight edges. As previously indicated, these difficulties stem from the fact that gradient directions are not defined in homogeneous image regions, and do not vary in the vicinity of straight edges (in the parallel and perpendicular directions). One possible solution to these two problems is to represent gradient directions as statistical distributions rather than as scalar values. In homogeneous regions such distributions are assumed to be flat, indicating that all angles are equally probable, whereas near edge structures they are assumed to peak at the corresponding gradient directions with, ideally, the amplitude of the peak increasing as the distance to the edge diminishes.

There are at least two possible ways of interpreting gradient direction within a statistical framework. The first, strictly local, is based on the intrinsic measurement errors occurring when gradient direction is calculated from image intensity sensed by photo-captors. Due to variations in lightness, spatial quantification, intrinsic noise in the photo-captors, and other factors, those directions are necessarily probabilistic. A direction being defined only for two and higher dimensions, a strict local interpretation cannot, however, be given much credence. At corners, T -junctions, and other spatial structures involving several edge elements, the notion of gradient direction is ambiguous, and becomes entangled with the regularization process needed for determining the spatial derivatives. The second statistical interpretation of gradient directions is therefore based on several local measurements performed within a small image region domain \mathbb{D} , and which are then lumped together. Denoting by $\theta_i(\vec{x}_i)$ the measured direction at time i and spatial position $\vec{x}_i = (x, y)$, and $p(\psi_i(\vec{x}_i) | \theta_i(\vec{x}_i))$ the estimated local statistical distribution of the gradient direction given the measured direction, then the probability density function (PDF) of the gradient directions at position \vec{x} can be expressed as

$$p(\psi_i(\vec{x}) | \{\theta_i(\vec{x})\}) = \frac{1}{K_i(\vec{x})} \sum_{i \in \mathbb{D}} p(\psi_i(\vec{x}_i) | \theta_i(\vec{x}_i)), \quad (3)$$

where ψ is the angular argument of the estimated distribution, the brackets $\{\cdot\}$ on the left-hand side denote variations around \vec{x} within the spatial domain \mathbb{D} , and $K_i(\vec{x})$ is a normalizing factor so that

$$\int_0^{2\pi} p(\psi_i(\vec{x}) | \{\theta_i(\vec{x})\}) d\psi = 1. \quad (4)$$

In practice $p(\psi_i(\vec{x}_i) | \theta_i(\vec{x}_i))$ is not known; even if we assume that $p(\psi_i(\vec{x}_i) | \theta_i(\vec{x}_i))$ is Gaussian (a common procedure to take account of several contributing noise factors), the resulting global PDF $p(\psi_i(\vec{x}) | \{\theta_i(\vec{x})\})$ is not necessarily so. For instance, at T -junctions these PDFs will be characterized by two peaks. This case is illustrated in Fig. 1.

The global PDFs can be estimated empirically through non-parametric estimator such as histogramming. In this case, to avoid over-fitting or undersampling the PDFs, the number of bins of the histograms has to be scaled in proportion to the number of direction samples, which depends on the size of domain \mathbb{D} . Only the image points with defined gradient directions contribute to increase the value contained in the corresponding bin. Such histograms will be flat in homogeneous regions, and will deviate from a uniform distribution as soon as gradient directions are defined within \mathbb{D} , that is, whenever the gradient norm is different from zero. This is illustrated in Fig. 2 for a square moving on a uniform background. In this example, as long as the edge of the square is within domain \mathbb{D} , the PDF peaks at the corresponding gradient direction; afterwards the PDF becomes flat.

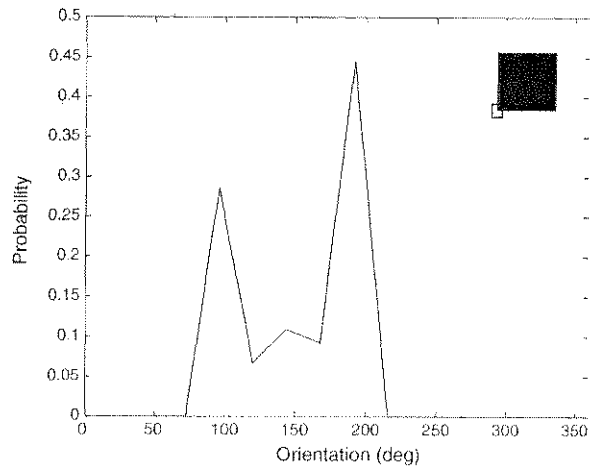


Fig. 1. Histogram of the probability distribution function (PDF) at a corner (bin's width 22.5°). The two main peaks at 90° and 180° correspond to the gradient directions of two orthogonal edges. The remaining non-zero bins are due to the low-pass spatial filtering, which by smoothing the straight corner introduces other gradient directions. The span of the region domain \mathbb{D} used for computing the histogram is indicated around the lower-left corner with a small square. The fact that this region is not exactly centered on the corner explains why the two peaks have differing amplitudes.

It will be shown in Section 4 that the histogramming method described above has the capacity to retrieve with relatively good accuracy optical flows. However, higher accuracy can be obtained by using a more elaborate parametric estimator, based on a priori knowledge of the edge's location. The main idea, when estimating the PDF, is to weigh differently the contribution of a gradient

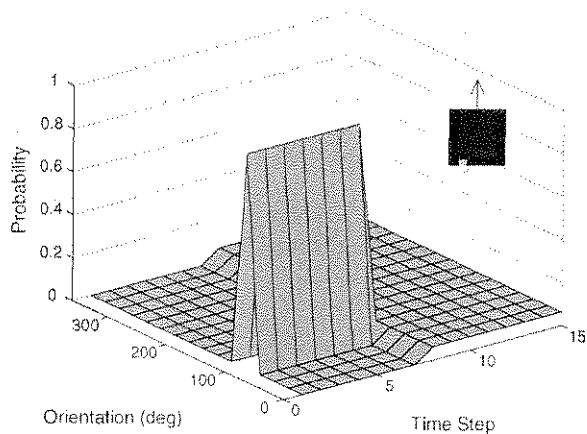


Fig. 2. Histogram of the probability distribution function (PDF) for a square moving on a uniform background (bin's width 22.5°). The motion direction is upward (see pointing arrow) with respect to the region domain \mathbb{D} , which is visible as a small square on the lower edge. The temporal axis has arbitrary units corresponding to the time interval between two successive frames. This moving edge yields a step-like PDF, that is, a high probability profile during the first time steps, until it moves out from the domain \mathbb{D} and the distribution becomes flat.

direction located on an edge from one located off an edge, where gradient directions are supposedly less reliable. The shape of the resulting PDFs will then depend on the distance to an edge, and ideally will be flat in homogeneous image regions and will peak on the edge.

Statistical criteria for deciding whether an image point is on or off an edge have been proposed by Konishi et al. [18]. In their work, these authors have investigated on several hundred natural images various edge cues and their statistical relevance for deciding that image pixels are on, near, or off object boundaries. They showed in particular that the magnitude of the image gradient has a fairly good discriminative power in such a classification. To formalize the dependence of $\psi(\vec{x})$ on the measured gradient magnitude, denoted by $\|\vec{\nabla}E_t(\vec{x})\|$, the conditional probability $p(\psi_t(\vec{x})|\theta_t(\vec{x}), \|\vec{\nabla}E_t(\vec{x})\|)$ is introduced. Given $\theta_t(\vec{x})$ is independent of $\|\vec{\nabla}E_t(\vec{x})\|$ (from the gradient magnitude one cannot infer the gradient direction), and assuming ψ are equiprobable in range $0 \cdots 2\pi$, this conditional probability can further be developed as

$$p(\psi_t(\vec{x})|\theta_t(\vec{x}), \|\vec{\nabla}E_t(\vec{x})\|) \propto p(\psi_t(\vec{x})|\theta_t(\vec{x}))p(\psi_t(\vec{x})|\|\vec{\nabla}E_t(\vec{x})\|). \quad (5)$$

Based on a large set of natural images, Balboa and Grzywacz [19] have determined the probability distribution that a point pertains to an occluding border. They found that such a probability increases supra-linearly with contrast, to reach a plateau for large contrasts. Taking inspiration from their experimental work, herein, it is assumed as a first approximation that the conditional probability $p(\psi_t(\vec{x})|\|\vec{\nabla}E_t(\vec{x})\|)$ is simply described by a sigmoid, that is,

$$p(\psi_t(\vec{x})|\|\vec{\nabla}E_t(\vec{x})\|) = \frac{\|\vec{\nabla}E_t(\vec{x})\|^4}{\alpha + \|\vec{\nabla}E_t(\vec{x})\|^4}, \quad (6)$$

where α specifies the inflexion point of the sigmoid. This probability is small for small gradient magnitudes, increases supra-linearly around the inflexion point, to saturate for very large gradient magnitudes. Taking this conditional probability back into Eq. (5), and considering a spatial domain \mathbb{D} around \vec{x} , the estimator for the PDF takes the form

$$p(\psi_t(\vec{x})|\{\theta_t(\vec{x}), \|\vec{\nabla}E_t(\vec{x})\|\}) = \frac{1}{K'_t(\vec{x})} \sum_{i \in \mathbb{D}} p_0 + p(\psi_t(\vec{x}_i)|\theta_t(\vec{x}_i)) \frac{\|\vec{\nabla}E_t(\vec{x}_i)\|^4}{\alpha + \|\vec{\nabla}E_t(\vec{x}_i)\|^4}, \quad (7)$$

where p_0 is a constant contributing to making the estimation robust to small values (typically found near homogeneous regions), and the normalizing constant $K'_t(\vec{x})$ guarantees that

$$\int_0^{2\pi} p(\psi_t(\vec{x})|\{\theta_t(\vec{x}), \|\vec{\nabla}E_t(\vec{x})\|\}) d\psi_t = 1. \quad (8)$$

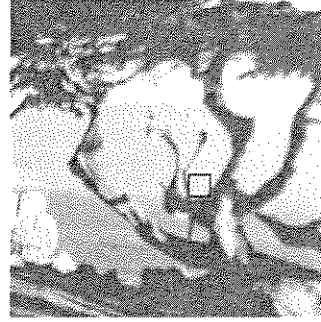


Fig. 3. Location of domain \mathbb{D} used for estimating the PDFs shown in next figure. This domain, marked by a square near the top of the tree's trunk, is 11 by 11 pixels.

If $p(\psi_t(\vec{x}_i)|\theta_t(\vec{x}_i))$ is assumed to be Gaussian, that is $p(\psi_t(\vec{x}_i)|\theta_t(\vec{x}_i)) = N(\psi_t(\vec{x}_i), \sigma^2)$, then $K'_t(\vec{x})$ reduces to

$$K'_t(\vec{x}) = np_0 + \sum_{i \in \mathbb{D}} \frac{\|\vec{\nabla}E_t(\vec{x}_i)\|^4}{\alpha + \|\vec{\nabla}E_t(\vec{x}_i)\|^4}, \quad (9)$$

where n is the number of pixels comprised in domain \mathbb{D} . The three parameters of this estimator are the location of the inflexion point, specified by α , constant p_0 , and the variance of $N(\psi_t(\vec{x}_i), \sigma^2)$. An example of estimated PDF for a natural image (Fig. 3) with $\alpha = 100$, $p_0 = 0.1$, and $\sigma^2 = 0.5$ is shown in the right part of Fig. 4. The left panel of this figure shows the PDF estimated with the non-parametric method (see also Fig. 2 for another example based on the non-parametric method). The smoothing effect of the parametric estimator, compared to the non-parametric estimator, is readily apparent in this figure, particularly in the homogeneous regions. The location of the spatial domain \mathbb{D} used for these estimations is visible as a small square near the center of Fig. 3.

2.3. Motion equation

In Section 2.2 the PDF for the estimated directions was expressed in non-parametric and parametric forms. Whatever the form, these PDFs are spatio-temporal functions, more simply written as $F(\psi; \vec{x}, t)$. This function is 2π periodic, and must verify

$$\int_0^{2\pi} F(\psi; \vec{x}, t) d\psi = 1. \quad (10)$$

Assuming the statistics within the domains \mathbb{D} remain constant in the time interval δt , in spite of the objects contained in the image being shifted and possibly rotated, then

$$F(\psi; \vec{x}, t) = F(\psi + \delta\psi; \vec{x} + \delta\vec{x}, t + \delta t) \quad (11)$$

should hold the exceptions being at the boundaries of the images, or when occlusions occur. In this equation,

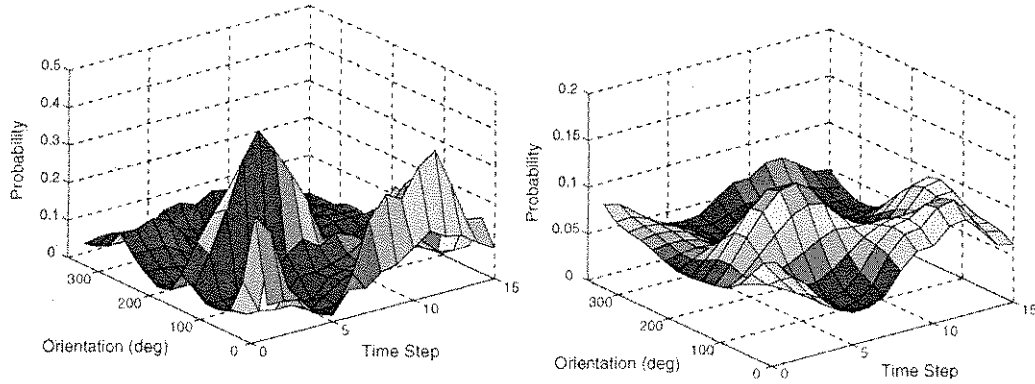


Fig. 4. Time variation of the PDF of gradient direction estimated using non-parametric (left) and parametric (right) methods for a tree translating leftwards (corresponding to a rightwards motion of the camera normal to its line of sight along the X-axis). For the position of the spatial domain \mathbb{D} , see previous figure. The temporal axis has arbitrary units corresponding to the time interval between two successive frames.

$\delta\vec{x}$ and $\delta\psi$ account for the linear and rotational motion components, respectively. Expanding the right-hand side in Taylor series, taking the limit $\delta t \rightarrow 0$, and keeping only the first order terms, one gets the OFC motion equation

$$\frac{\partial F}{\partial t} = -\vec{\nabla}F \cdot \vec{v} - \frac{\partial F}{\partial \psi} \omega, \quad (12)$$

where \vec{v} and ω represent the displacement velocity and angular speed, respectively. How \vec{v} and ω can be estimated from this motion equation is addressed in the next section. It must be noted that in the case $F(\psi; \vec{x}, t)$ takes the form of a multivariate Gaussian distribution (near an elongated edge for instance), this equation is under-determined, which is an indication that the image region is subject to the aperture problem. However, in practice analyzing the shape of $F(\psi; \vec{x}, t)$ is not straightforward, and thus another criterion must be defined to identify the aperture problem, as discussed in Section 2.4.

Eq. (12) describes a motion equation based on a two frame scheme, during which it is assumed that the velocity is constant, or very slightly changing in the case ω is non-zero (as rotations necessarily imply an acceleration). Given $F(\psi; \vec{x}, t)$ stands for a PDF, a more general form of this equation, known as Kolmogorov's equation [20], comprises diffusion terms. These terms would express in this case random variations in the velocity and rotation between two frames, and are only appropriate if the goal is to track image features, as shown elsewhere [21]. The presence of the diffusion terms makes motion estimation more robust to sharp changes in the trajectory, or to the case where the moving objects become temporarily occluded. Also, velocity estimation improves over several time frames, but only if the image motion is temporally coherent with the motion model. The aim of this present paper, however, is not to exploit temporal coherency for improving motion estimation, but rather to propose a form of the OFC equation that uses as image features the gradient directions.

2.4. Motion estimation

A method for determining the optical flow from Eq. (12) is now proposed. This method relies on the fact that at a given spatial position, there is potentially enough information about motion contained in the spatio-temporal variation of the estimated probability distribution of the gradient directions. In the situation where the spatial domain \mathbb{D} encompasses more than just one extended edges or the edges happen to be curved, evidence gathered over a range of gradient directions can be used to determine the different motion components without being subject to the aperture problem. One possibility resides in finding \vec{v} and ω , at a given spatial position, best verifying the motion equation over a range of gradient directions. This can be stated more formally by using the least-squares method, where the global error measure, given by

$$e(\vec{x}, t) = \int_0^{2\pi} \left[\frac{\partial F}{\partial t} + \vec{\nabla}F \cdot \vec{v} + \omega \frac{\partial F}{\partial \psi} \right]^2 d\psi, \quad (13)$$

is sought to be minimized. Such a minimization is expressed by

$$\frac{\partial}{\partial \vec{a}} e(\vec{x}, t) = \vec{0}, \quad (14)$$

where $\vec{a} = (v_x, v_y, \omega)^T$. This derivative provides three linear equations, one for ω and two for velocity \vec{v} . In matrix notation, these equations are expressed as

$$\begin{pmatrix} v_x \\ v_y \\ \omega \end{pmatrix} = \begin{pmatrix} \int F_x^2 d\psi & \int F_x F_y d\psi & \int F_x F_\psi d\psi \\ \int F_y F_x d\psi & \int F_y^2 d\psi & \int F_y F_\psi d\psi \\ \int F_\psi F_x d\psi & \int F_\psi F_y d\psi & \int F_\psi^2 d\psi \end{pmatrix}^{-1} \begin{pmatrix} -\int F_t F_x d\psi \\ -\int F_t F_y d\psi \\ -\int F_t F_\psi d\psi \end{pmatrix}, \quad (15)$$

where F is short notation for $F(\psi, \vec{x}, t)$, and F_x, F_y, F_ψ , and F_t denote the derivatives with respect to space, gradient direction, and time, and the integrations are performed over $[0 \dots 2\pi]$. In the computer implementation the integrals are approximated by discrete sums (corresponding to the bins of the histogram) and the four derivatives (F_x, F_y, F_ψ , and F_t) are evaluated for each bin.

According to the proposed method, the recovered velocities \vec{v} are those pointing towards the 'true' motion direction, if not subject to the aperture problem. The aperture problem appears over extended linear edges. In this case, only the velocity perpendicular to the edges can be computed. This aperture problem can usually be diagnosed by inspecting the eigenvalues of the first-right-hand-side matrix in Eq. (15). If the three eigenvalues are sorted in descending order, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$, and λ_3 is greater than a threshold λ , then \vec{v} and ω can be computed from Eq. (15). If $\lambda_3 < \lambda$, and $\int F_\psi^2 d\psi$ is greater than a threshold τ , then there is indication that the image point is subject to the aperture problem. In this case, only a normal velocity estimate can be computed, and the rotational component ω can be determined using a reduced form of Eq. (15):

$$\omega = - \frac{\int F_t F_\psi d\psi}{\int F_\psi^2 d\psi}. \quad (16)$$

The denominator of this equation must be greater than τ to ensure consistent solutions; it will be zero (or close to) in the case of a rotating disk for instance. Compared to affine models, or other second-order methods, the proposed method has thus the advantage of potentially providing rotational information locally over extended elements, which are in principle subject to the aperture problem. Finally, when the three eigenvalues are smaller than λ and $\int F_\psi^2 d\psi$ is smaller than τ , no velocity and no rotational component are possibly computed.

To give a more intuitive apprehension of the method for estimating \vec{v} and ω , a schematic representation of the spatio-temporal variation of function $F(\psi, \vec{x}, t)$ for a translating and rotating surface is illustrated in Fig. 5. Translations

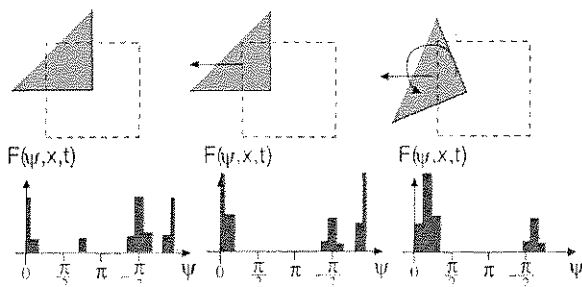


Fig. 5. Illustration of motion estimation based on probability density functions. Left: initial probability distribution, estimated at a specific spatial position within the spatial domain delimited with dashed lines. Middle: Left-ward motion modifies the shape of the PDF. Right: rotation modifies slightly the shape and shifts the whole PDF.

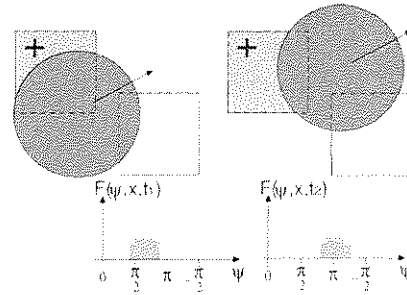


Fig. 6. Illustration of apparent rotations, as induced by translating curved boundaries. Two time frames are shown. Depending on the location of the spatial domain (squares), the sign of the induced rotational component ω is either positive (dark grey) or negative (light grey).

and rotations have two different effects on the PDF. In principle translations affect the shape of the PDF, while rotations shift the whole function while preserving its shape. However, because the region domains \mathbb{D} are finite, the situation is more intricate. For one, a purely translating shape with curved edges induces shifts in the PDF mimicking rotations, as illustrated in Fig. 6. Similarly, when an object is not fully contained in a region domain, pure rotations may also affect the shape of the PDF, as for translations (see right panel in Fig. 5). This is because rotations modify the spatial position of the rotating elements (exception to this rule is when the whole shape is contained within \mathbb{D}). Consequently, there is some ambiguity in the variations of the PDFs relatively to the object's motion. Also, part of this ambiguity is resolved by the least-squares method, which provides the translational and rotational components that best satisfy the OFC equations.

3. Algorithm

3.1. Description of the algorithm

This section is concerned with the practical calculation of Eq. (15). To begin, the spatial derivatives E_x and E_y , where E stands for the luminance, must be determined as the gradient directions depend on these values. The determination of E_x and E_y is done using a forward differencing scheme. To make these derivatives well-posed, the image sequences are smoothed spatially with a Gaussian filter with a standard deviation of 1 pixel. From the gradient directions $\theta = \arctan(E_y/E_x)$, the PDFs are then established, either through simple histogramming, or according to Eq. (7). For the sake of symmetry, the first bin spans $-\pi/n \dots \pi/n$, the second $\pi/2 \dots 3\pi/n$, and so on, where n is the number of bins.

Critical to motion estimation is the number of bins. For one, it should be great enough so as to have enough equations for the least-squares estimation. However, the larger the number of bins, the more information must be gathered to fill them, which in turn requires more extent

domains \mathbb{D} . However, too large domains \mathbb{D} can be detrimental for motion estimation as (1) the assumption of constantness of the optical flow within \mathbb{D} might break down, and (2) the position of the velocity vector associated to the changes in the PDF becomes inaccurate. As a trade-off, the number of bins was fixed to 8 (each bin spans 45°) and \mathbb{D} was set to 3×3 pixels, unless stated otherwise.

Another requirement for computing Eq. (15) is that the derivatives F_x , F_y , and F_ψ are well-posed. This implies that spatial and temporal smoothing of the PDFs $F(\psi; \vec{x}, t)$ are done carefully. Indeed, any methods based on differential techniques crucially depend on this regularization process, which has a large impact on the accuracy of the optical flow [22]. Also, to allow for comparisons of the resulting optical flows obtained with the proposed methods and those obtained with other methods described in Barron et al. [7], the spatio-temporal smoothing applied on $F(\psi; \vec{x}, t)$ was the same as the one described by these authors: they used a spatio-temporal Gaussian kernel with a standard deviation of 1.5 pixels in space and 1.5 frames in time, sampled out to three standard deviations. The temporal support for the entire filtering process was 15 frames. Such a filter was applied individually to each bin of the PDFs. To compute F_ψ , a similar filter was applied on the bins, with the additional constraint that F is 2π periodic with respect to ψ . As in Barron et al. [7], derivatives were computed with four-point central differences, with mask coefficients $(-1/12, 8/12, 0, -8/12, 1/12)$.

The final step of the algorithm concerns the inversion of the 3×3 matrix in Eq. (15). This inversion can lead to instabilities in the case the system of equation is under-determined (due to the aperture problem) or is ill-conditioned because $F(\psi; \vec{x}, t)$ does not vary sufficiently. Some confidence measure is thus required to separate reliable from unreliable optical flows. Confidence measures based on the matrix's determinant value or on the matrix's minimum eigenvalue have been shown to be particularly good indicators [22]. Both criteria have been tried and the minimum eigenvalue tended to give slightly better results than the one based on the determinant. Furthermore, the eigenvalue criterion was also applied in Barron et al. [7], and thus choosing it makes the quantitative comparisons more convenient. The way this criterion was applied on the three eigenvalues was previously described in Section 2.4.

3.2. Computational complexity

The proposed motion estimation methods (either non-parametric or parametric) involve at each pixel, in addition to low-pass filtering, three main steps: (1) computation of the gradient direction, (2) establishment of the PDF, and (3) resolution of Eq. (15). The overall computational complexity of these methods is linear with respect to the number of pixels contained in the image, the number of operations being proportional to the kernel size of the low-pass filters, and to the number of bins used to approximate the PDFs.

Step 1 requires the calculations of the horizontal and vertical spatial gradients of the image brightness. The gradient directions are then obtained by taking the inverse tangent of the ratios of the vertical and horizontal spatial gradients. All the operators involved in these calculations (forward differencing scheme and inverse tangent) are straightforward. Also, with new advances in microtechnology, the computation of gradient directions, including the low-pass filtering stage, is hard-wired in some imagers (see for instance [23]). This kind of imager makes the gradient direction and gradient amplitude 'freely' available at no extra computational cost. Recently, an imager for extracting hard-wired gradient directions based on normalized spatial gradients (contrasts) has even been described [24].

Step 2 consists either in simply establishing the histogram of the gradient directions in the non-parametric case, or in computing Eq. (5) in the parametric case. Eq. (5) is relatively computationally costly as it involves the product of Gaussian and sigmoidal functions, and requires the calculation of the norm of the spatial gradient. In both cases, the histograms are built by collecting the gradient directions within a pixel neighborhood, usually taken to be 3×3 pixels. To obtain PDFs these histograms must be normalized, a straightforward operation that involves the summation of all bin values followed by the multiplication of each bin value with the inverse of the result of the summation.

Step 3 is the core of motion estimation. It asks for the resolution of Eq. (15), which involves for each pixel and in the most complete case (that is, when including the rotational component), the determination of 12 components followed by a 3×3 matrix inversion. The components of the matrices are made of sums of pair products of a combination of the temporal, spatial (horizontal and vertical), and rotational PDF derivatives. The number of sums is proportional to the number of bins used to sample the PDFs. In most cases, the number of bins is conveniently set when chosen to be 8. This last number should not be confused with the number of gradient directions lumped together to build the PDFs, and which is specified by the spatial neighborhood (3×3 pixels if not otherwise indicated). Consequently, the total number of derivatives one has to compute at each time step is given by the number of pixels \times 8 (number of bins) \times 4 (two spatial, one temporal and one rotational derivatives).

In addition to these three steps, spatio-temporal and rotational smoothing is necessary to guarantee stability of the derivative operators. As indicated above, smoothing is obtained by convolving the image with a spatio-temporal Gaussian kernel, as described in Barron et al. [7]. The difference, however, is that this convolution is done for each bin of the histogram function (4D smoothing kernel), and is thus typically eight times more computationally intensive.

The overall complexity of the proposed algorithms is similar to that found in the Lucas and Kanade's (LK) method [10], which is also based on least-squares fit to

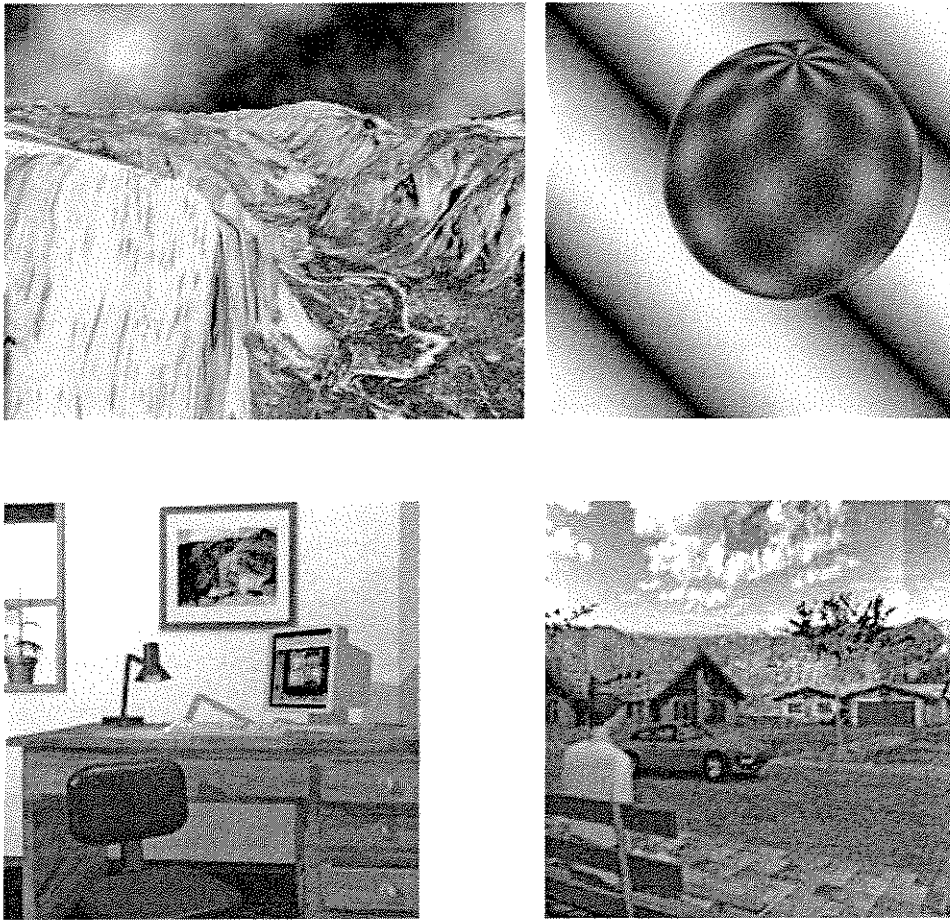


Fig. 7. From top left to bottom right are shown one sample from the Yosemite, Sphere, Office, and Street synthetic sequences.

determine the velocity vectors. Typically, this fit involves 25 equations, a number that corresponds to the spatial domain considered (5×5 pixels). In their approach the least-squares fit is dependent on a separable isotropic window function (whose 1D weights are set to $[0.0625, 0.25, 0.375, 0.25, 0.0625]$). The number of 25 equations is larger than the 8 involved in the proposed approaches based on gradient directions (corresponding to the number of bins used to quantify the PDFs). However, the LK method does not need a normalization step, nor the evaluation of Eq. (5) and only needs a 3D smoothing kernel (not 4D). Taken together, these differences make the computation time of the gradient direction based methods (with the parameters used) roughly 10 times that of the LK method.

4. Results

The accuracy of the proposed algorithms has been quantitatively assessed on eight test images proposed by

Barron et al. [7] and Galvin et al. [25], four of which are visible in Fig. 7. For these sequences, the results obtained with the non-parametric and parametric models are both presented. The first sequence consists of the superposition of two sinusoidal plane waves with wavelength of 16 pixels/cycle (SinusB). The second test case involves a translating dark square over a bright background which has uniform velocity $\vec{v} = (4/3, 4/3)^T$ (Square2). The two next sequences show a tree in a landscape. Motion is simulated by translational camera motions, either normal to its line of sight along the X-axis (*Translating Tree*), or along its line of sight (*Diverging Tree*). The fifth sequence is an outdoor scene (*Yosemite*) where motion is more complex as both divergent and translational motions are present. The three last sequences were obtained by ray tracing, and thus were more realistic for the light conditions. The first represents a rotating sphere (*Sphere*), the second a view of an office seen through a camera translating along its line of sight (*Office*), and the third a street with a combination of a car moving right while the camera is translating to the left (*Street*).

For all these image sequences the true optical flow is available. Consequently, it is possible to calculate at a given time the sum of all errors (aggregate error) involved in the estimation of each motion vector. Error measurement consists in calculating the angle [7]

$$\psi_E = \arccos(\vec{v}_c \cdot \vec{v}_e), \quad (17)$$

where \vec{v}_c is the correct velocity and \vec{v}_e is the estimate, and both vectors are represented as 3D direction vectors

$$\vec{v} = \frac{1}{\sqrt{v_1^2 + v_2^2 + dt^2}}(v_1, v_2, dt)^T \quad (18)$$

with time step $dt = 1$. Using this formulation, ψ_E measures the angular deviations from the correct space-time orientation, that is, takes into account the direction and the magnitude of the motion vectors.

The aggregate error for the eight sequences, along with the standard deviation and the density of the flow fields, are provided in Table 1. The flow fields corresponding to the images shown in Fig. 7 are illustrated in Figs. 8–11 using small arrows (for the sake of clarity, only one arrow over two is displayed in these representations). The capability of the method in estimating the normal velocity component when the full velocity cannot be computed (due to

the aperture problem) is shown for two sequences in Table 2. All these error measurements were obtained by assuming the rotational component ω to be zero, which reduces by one the number of unknowns in the system of equations. When not, the errors tended to be larger (about 10%), which is understandable in view that no rotational components perpendicular to the plan are present in these image sequences.

In addition to the results for the non-parametric and parametric methods, those obtained for the LK method [10] are also provided in these tables for comparison. The motivation for having chosen the LK method comes from the comparative study performed by Barron et al. [7] where it was shown that the LK method is the most accurate differential technique, a conclusion also supported recently by Galvin et al. [25] based on quantitative optical flow evaluations performed on the ray tracing image sequences (Fig. 7). Furthermore, as pointed out in Section 3.2, the complexity of the LK algorithm is comparable to the non-parametric and parametric methods (to the exception of the low-pass filtering stage). Therefore, the LK performance represents a convenient reference against which methods based on differential techniques of similar complexity can be confronted.

Table 1
Summary of velocity results for non-parametric (I), parametric (II), and Lucas and Kanade (LK) methods

Image	Method	Average error (°)	Standard deviation (°)	Density (%)
SinusB ($\lambda \geq 0$)	I	17	13.8	100
SinusB ($\lambda \geq 0$)	II	4.2	3.6	100
SinusB ($\lambda \geq 0$)†	II	2.05	0.86	100
SinusB ($\lambda \geq 0$)	LK	2.5	0.15	100
Square2 ($\lambda \geq 10^{-3}$)	I	3.5	2.5	10
Square2 ($\lambda \geq 10^{-4}$)	II	0.3	0.17	9
Square2 ($\lambda \geq 1.0$)	LK	0.2	0.18	9
Translating tree ($\lambda \geq 3 \times 10^{-3}$)	I	2.3	2.5	36
Translating tree ($\lambda \geq 10^{-4}$)	II	0.72	0.69	45
Translating tree ($\lambda \geq 5 \times 10^{-5}$)‡	II	0.59	0.53	55
Translating tree ($\lambda \geq 0.5$)	LK	0.70	0.70	57
Diverging tree ($\lambda \geq 3 \times 10^{-3}$)	I	4.0	3.4	51
Diverging tree ($\lambda \geq 10^{-4}$)	II	2.7	2.2	54
Diverging tree ($\lambda \geq 1.0$)	LK	2.0	2.0	54
Yosemite ($\lambda \geq 5 \times 10^{-3}$)	I	7.0	9.3	19
Yosemite ($\lambda \geq 2 \times 10^{-4}$)	II	5.2	9.7	30
Yosemite ($\lambda \geq 0.5$)	LK	4.8	10.8	50
Sphere ($\lambda \geq 8 \times 10^{-3}$)	I	6.8	10.4	5
Sphere ($\lambda \geq 5 \times 10^{-4}$)	II	5.9	10.1	5
Sphere ($\lambda \geq 2.0$)	LK	10.0	11.9	6
Office ($\lambda \geq 2.5 \times 10^{-3}$)	I	6.8	6.4	25
Office ($\lambda \geq 10^{-4}$)	II	5.3	5.6	24
Office ($\lambda \geq 0.5$)	LK	4.9	6.1	27
Street ($\lambda \geq 3 \times 10^{-3}$)	I	9.9	14.2	45
Street ($\lambda \geq 10^{-4}$)	II	7.2	14.7	49
Street ($\lambda \geq 0.5$)	LK	6.2	15.0	54

Parameters were the same for all experiments, that is, domain size $10 \times 3 \times 3$, $\sigma^2 = 0.5$, $p_0 = 0.1$, $\alpha = 100$, to the exceptions of SinusB marked with † for which $\alpha = 10000$ (to give more weight to higher gradients in the motion estimation), and Translating tree indicated with ‡ for which the neighborhood size was set to 11×11 pixels.

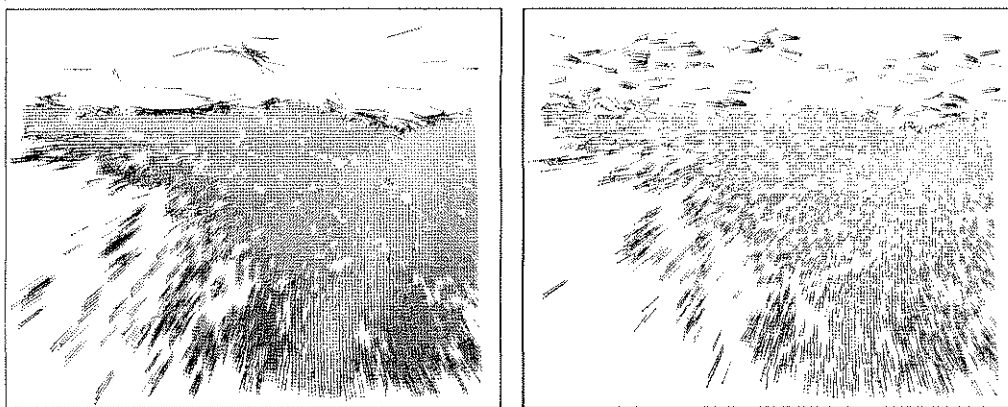


Fig. 8. Flow fields for the LK method (left, density 50%) and parametric method (right, density 30%) obtained for the Yosemite sequence.

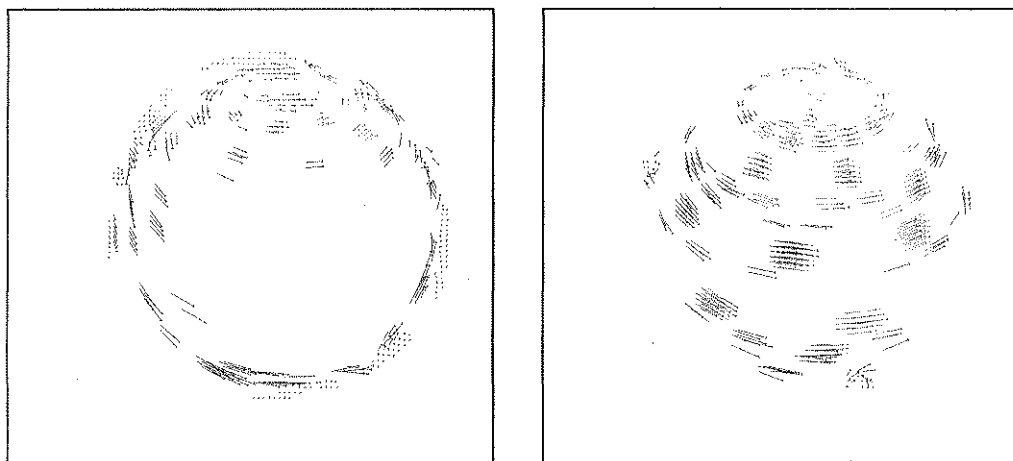


Fig. 9. Flow fields for the LK method (left, density 6%) and parametric method (right, density 5%), obtained for the Sphere sequence.



Fig. 10. Flow fields for the LK method (left, density 27%) and parametric method (right, density 24%) obtained for the Office sequence.

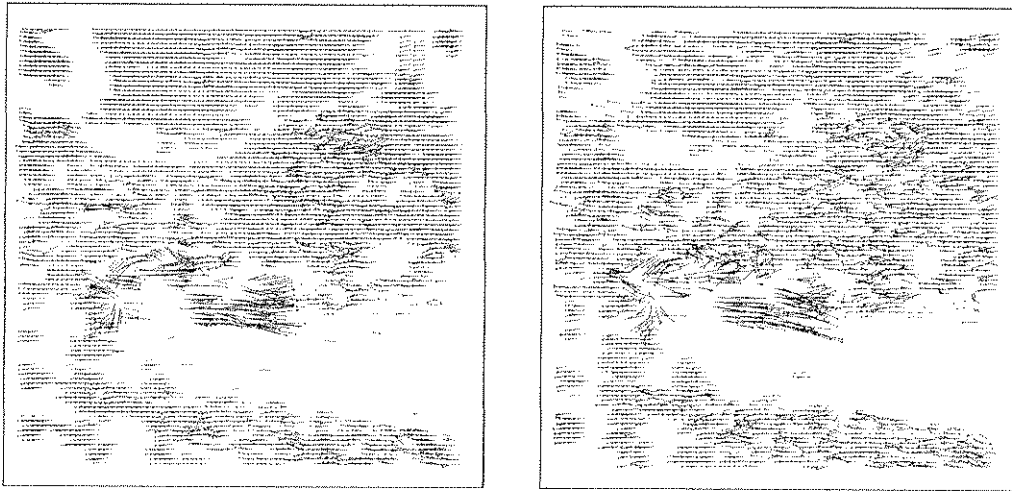


Fig. 11. Flow fields for the LK method (left, density 54%) and parametric method (right, density 49%) obtained for the Street sequence.

The quantitative results summarized in Tables 1 and 2 indicate that the parametric estimation (method II) is much more satisfying than the non-parametric one (method I). Furthermore, the aggregate errors corresponding to the parametric method are similar, and in a few instances smaller, than those obtained by the LK method. An instance of smaller error is the translating tree. The velocity being the same over the whole image, a larger domain size (11×11) could be used to estimate the PDFs, which improved the accuracy of the motion estimation while being unaffected by the increased positional inaccuracy due to the larger spatial domain. The aggregate error is, however, not always an ideal criterion for comparing the performances. For instance, for the Yosemite sequence, the LK method indicates an error of 4.8° for a density of 50%, which is less than the error of 5.2° for a density of 30% obtained with method II. However, if one looks at the optical field, shown in Fig. 8, one can see that the field computed by the LK method (left panel) is very sparse in the upper part, which is the region where the optical flow is the most difficult to obtain. In spite of the lower density (30%), method II (Fig. 8, right panel) has a fairly good representation of the optical flow in this region, which, although it contributes to increase

the aggregate errors, looks more stable than the sparser flow recovered by the LK method. The same can be said for the Sphere where for equivalent densities, the proposed method is more accurate and has a much better representation of the flow field than that provided by the LK method (Fig. 9). This observation is further confirmed with the Office sequence, shown in Fig. 10, where for equivalent densities the optical flow for the region near the chair, a region with low contrasts and with lighting variations between two frames, is again better represented using method II. The results of another complex sequence (Street) are shown in Fig. 11. In the regions where the rightward-moving car is on cluttered background, and where the tree in the background is subject to lighting changes, the flow field is difficult to obtain with accuracy. In those regions, the proposed methods (I and II) tend to provide more velocity vectors than the LK method. However, because these vectors are more subject to errors, the neat aggregate errors demonstrate an increase of at least one more degree than the aggregate error otherwise obtained with the LK method.

Compared to other second-order methods, such as those proposed by Nagel [16] and Uras et al. [17] for instance, the proposed parametric method (method II) is more accurate,

Table 2
Summary of normal component velocity for non-parametric (I), parametric (II), and Lucas and Kanade (LK) methods

Image	Method	Average error ($^\circ$)	Standard deviation ($^\circ$)	Density
Square2 ($\lambda = 5 \times 10^{-3}$)	I	1.2	0.99	30
Square2 ($\lambda = 5 \times 10^{-4}$)	II	0.14	0.10	14
Square2 ($\lambda = 10^{-4}$)	LK	0.25	0.60	15
Diverging tree ($\lambda = 5 \times 10^{-3}$)	I	2.2	2.3	58
Diverging tree ($\lambda = 5 \times 10^{-4}$)	II	1.3	1.2	55
Diverging tree ($\lambda = 5.0$)	LK	0.94	0.76	45

Parameters as specified in Table 1.

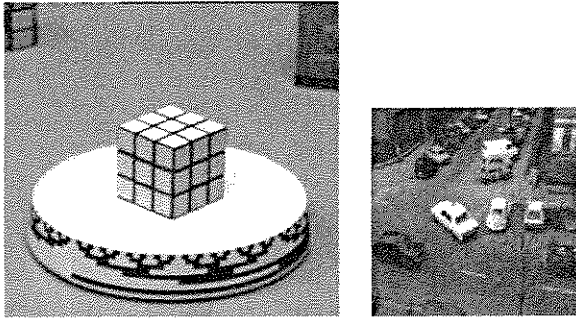


Fig. 12. Natural image sequences, a rotating Rubik cube (left image) and moving cars (right image).

except for the translating square and tree where Uras et al. obtained an error of 0.15° (for a density of 26%) and 0.46° (for a density of 42%), respectively. It must be remarked that these authors obtained their results through

the application of a selection: they looked for the eight estimates that best satisfy a constraint, and of these, they further chose the estimate with the largest confidence. Similarly it must be mentioned that Nagel [16] introduced a regularization constraint imposed on the motion field (in the direction perpendicular to the spatial gradients). In spite of these more sophisticated schemes for improving their results, their aggregate errors are larger than those obtained with the parametric method, but slightly smaller than those obtained with the non-parametric method (method I).

The new algorithms have been tested on two real image sequences, visible in Fig. 12. The first consisted in a rotating Rubik cube and the second in moving cars (the 'Hamburg taxi sequence'). As mentioned in [7], it is difficult to assess the errors in the computed optical flow for natural image sequences. Background noise is generally evident and its occurrence can be reduced by increasing the confidence criterion λ at the detriment of the optical flow density. Parameters of the simulations were those described above

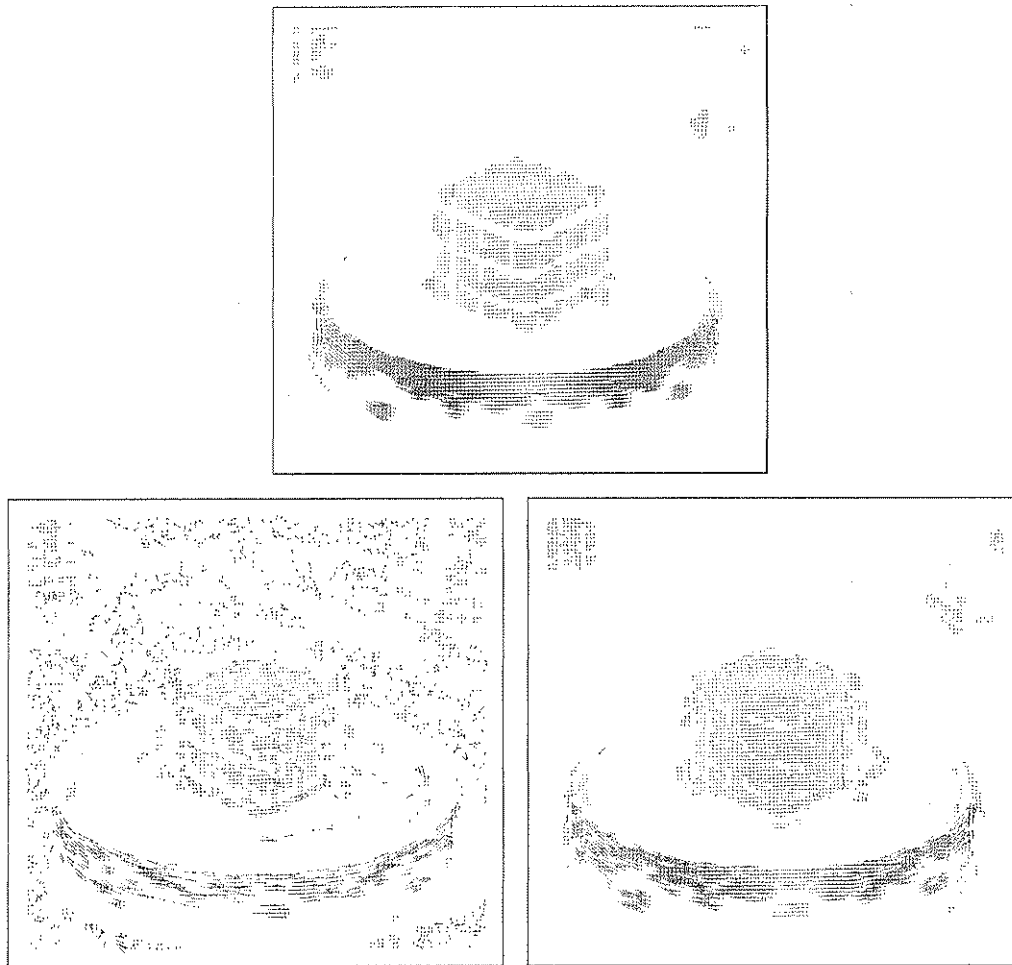


Fig. 13. Optical flows for the Rubik sequence. Upper panel shows the results for the LK method ($\lambda_{\min} = 1$, density 19%), while the left and right lower panels show the results for the non-parametric ($\lambda_{\min} = 5 \cdot 10^{-5}$, density 21%) and parametric methods ($\lambda_{\min} = 10^{-4}$, density 20%), respectively.

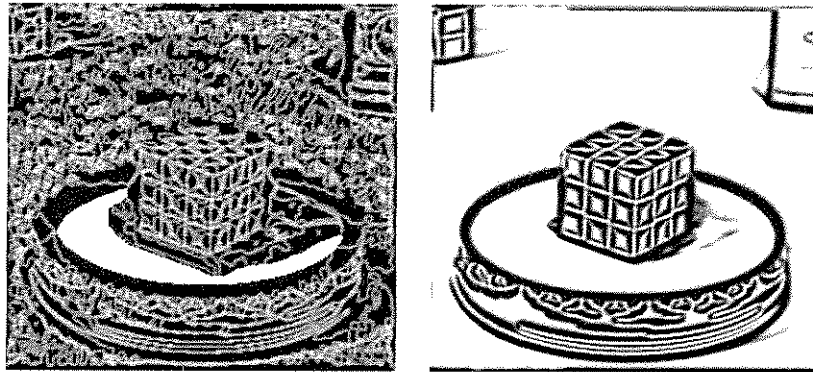


Fig. 14. Entropy of the PDFs for the non-parametric (left) and parametric (right) methods. High entropy, shown in white, is characteristic of uniform distributions, whereas low entropy, shown in black, is typical of peaked distributions. With the parametric method, PDFs distant from edge structures are characterized by higher and less erratic entropy than that obtained with the non-parametric method.

for the synthetic image sequences. The optical flows have been determined for the LK, the non-parametric and parametric methods. The results for the Rubik cube are shown in Fig. 13 for the three methods. The effect of the a priori knowledge on the location of the edges on the recovered optical flow is easily perceptible: the parametric method provides much less noisy velocity vectors in homogeneous regions. The reason for these differences in the amount of noise can be better grasped by looking at the entropy of the PDFs computed at each pixel, and shown for the Rubik cube in Fig. 14. With the parametric method (right panel), the transitions in the entropy values when passing from an edge, which typically has low entropy (black areas in Fig. 14), to homogeneous regions, which have high entropy (white areas in Fig. 14) are much more pronounced than those obtained with the non-parametric method (left panel). Furthermore, in those homogeneous

regions the entropy tends to be less variable between two neighbor positions in the parametric case than in the non-parametric one. These two observations thus clarify why motion estimation is more robust when performed through the parametric model.

Finally, the results for the taxi sequence, for the LK and parametric methods, are shown in Fig. 15. In the resulting optical flows, it can be seen that in image regions where the brightness constraint is weak, for instance at proximity of the dark car in the left part of the image, the proposed methods based on gradient directions provide denser flow fields than those obtained with the LK method, an observation already pointed out for the synthetic image sequences.

To demonstrate the ability of the proposed methods to estimate the rotational component, rotating surfaces (octagons—see Fig. 16) translating down-left-ward were

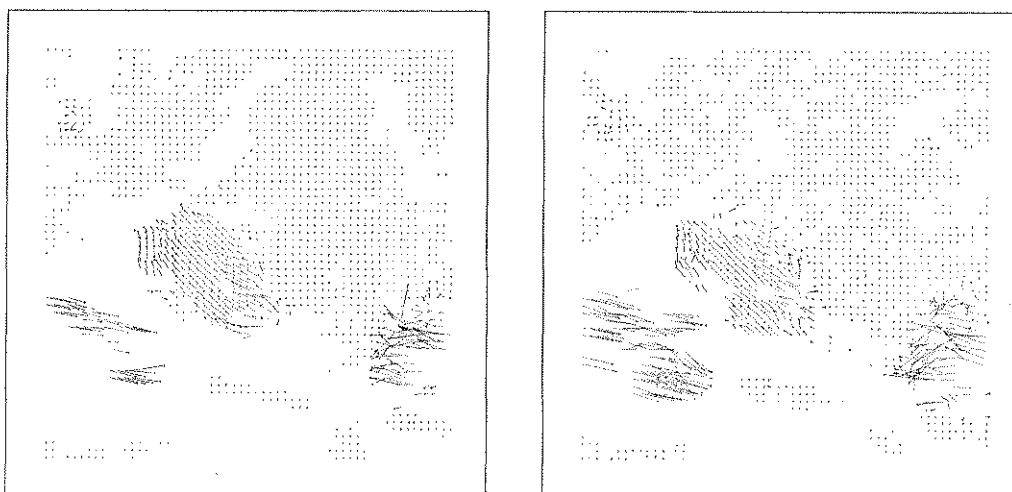


Fig. 15. Optical flows for the Taxi sequence. Left panel shows the results for the LK method ($\lambda_{\min} = 0.5$, density 52%), and right panel those for the parametric method ($\lambda_{\min} = 10^{-4}$, density 48%). In the region where a car with low luminance contrast is moving rightward (lower-left part of the image), this latter method provides a denser flow field than that provided by the LK method.

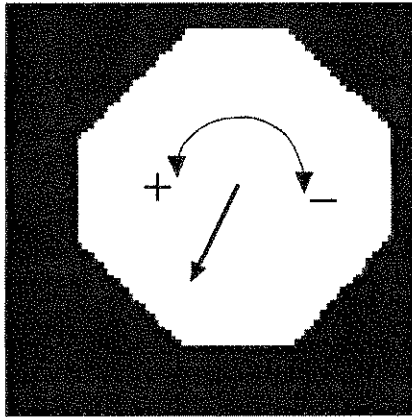


Fig. 16. Octagon translating down-left-ward, while rotating clockwise (–) or counterclockwise (+).

considered. For such affine motions, the application of an affine model can be used to compute the translational and rotational components from the optical flow \vec{v} , determined for instance through the LK method, and shown in Fig. 17. For this type of image sequence with most of the image points subject to the aperture problem, such an affine method works with limited accuracy. Second-order differential techniques are also subject to this aperture problem as the angular velocity is estimated from the optical flow [17]. The alternative is to apply Eq. (12) to estimate the rotational components based on shifts of the PDFs. However, for reasons already evoked in Fig. 5, the combination of rotations and translations can lead to ambiguous temporal variations of the PDFs. In spite of these difficulties, the sign of the rotational component and the direction of the local motion were correctly estimated (lower panels in Fig. 18). Also, as visible in Fig. 18, the rotational component has been recovered over the entire length of the contours, even in regions subject to the aperture problem. In those regions, the matrix to be inverted in Eq. (15) reduces to a scalar,

and the other confidence level τ , defined in Section 2.4, was set to 0.05 to prevent the computation of ω from being based on ill-conditioned data (the other parameters of the model were set to the same values as for the other experiments). By averaging the amplitude of all rotational components of the dominant sign the curl of the affine motion could be estimated. It was found to be 0.028 and -0.035 for the counterclockwise and clockwise rotations, respectively, while the true curl was 0.03 and -0.03 . The application of the affine model based on the flow fields obtained with the LK method (Fig. 17) yielded a curl of 0.04 and -0.037 , respectively.

5. Discussion

The OFC equation has originally been based on the image brightness to extract motion [3,4]. However, any other visual features obtained through motion invariant operators are in principle applicable to estimate motion based on the OFC equation [15]. Although these visual features should also include gradient directions, as pointed out a long time ago [26], no evidence of their applicability for estimating motion in image sequences has yet been reported in the literature. This lack of evidence is certainly accountable on several problems related to the fact that gradient directions represent angles, taking values in range $[0, 2\pi]$ independently of the gradient magnitude. In image regions with (nearly) homogeneous brightness, this assumption makes the estimation of motion ill-posed. Furthermore, there are no variations in the edges' orientation not only in a direction parallel to them (which is to be expected), but also perpendicular to them. These problems are peculiar to the gradient direction as other visual features such as the edgeness (magnitude of the gradient of intensity), the curvature (or higher-order derivatives), the cornerness (as defined for instance in [15]), and so on, have graded values directly interpretable by the motion equation. Another

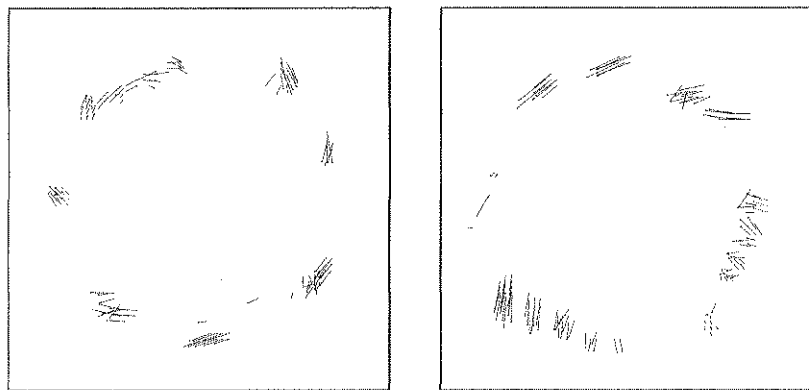


Fig. 17. Optical flow recovered with the LK method for the octagon rotating clockwise (left panel) or counterclockwise (right panel), while translating down-left ward.

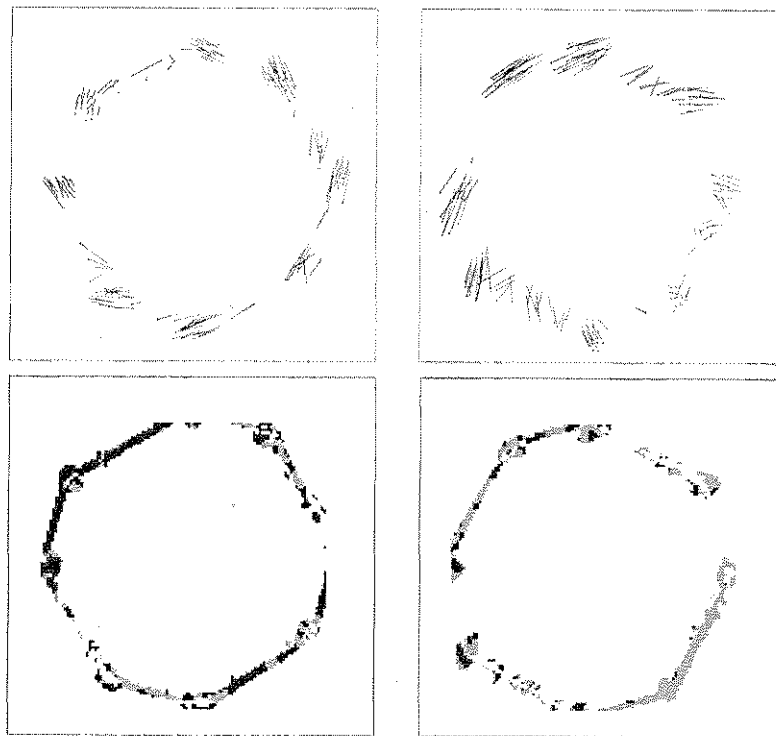


Fig. 18. Optical flow and rotational component recovered with the parametric method for the octagon rotating clockwise (left panels) or counterclockwise (right panels), while translating down-left ward. The optical flows \vec{v} and the rotational component ω are shown in the upper and lower panels, respectively. In lower panel, black stands for negative (clockwise) rotation, and grey stands for positive (counterclockwise) rotation.

difficulty is that the difference between two angles cannot be larger than π , which implies some extra precautions when evaluating the time and space derivatives of the angle function. To circumvent all these problems, motion estimation was proposed to be based on the spatio-temporal variation of the PDF of the gradient directions, evaluated at each spatial position. The soundness of this approach has been checked on synthetic and real image sequences and compared to other differential methods of similar complexity.

To establish the distribution of the gradient directions, two methods have been explored. The first method consisted in histogramming the gradient directions without making any assumptions on the shape of the distribution (non-parametric method). The idea of using histograms to solve computer vision problems is not new and has previously been proposed for recognition tasks [27–29]. However, in the case of motion estimation, such a crude non-parametric estimation did not perform well on most of the image sequences considered; the reason for this bad performance was the important sensitivity of the histograms to noise. The second method, classified as parametric, embedded a priori knowledge on the statistics of the noise (taken as Gaussian), and weighted the contribution of an image pixel accordingly to its distance to an edge. Previous works have shown that, statistically, gradient magnitude and contrast provide

a reliable cue for deciding whether a pixel is on or off an edge [18,19]. Based on these studies, the PDF was made dependent on the gradient magnitude, becoming sharper on and at proximity of edges, and flatter in homogeneous regions. The effect of this a priori knowledge can be sensed in Fig. 14, where the entropy of the PDFs (measured at each image point) is shown for the parametric and non-parametric methods. Compared to the non-parametric approach, the parametric method has sharper transitions in the entropy function when passing from an edge to more uniform regions and is less subject to small brightness variations. Because of these characteristics, the parametric method fared better than the non-parametric one. It also yielded quantitative results mostly similar, but overall of better quality (particularly in regions with brightness variations) than those provided by LK method [10], which is recognized as one of the best differential method [7,25]. However, the proposed methods do not compete, at least on synthetic scenes, with more sophisticated (and thus computationally more costly) methods, such as phase-based techniques [30] and with methods based on improved motion models [31], particularly with those relying on a coarse to fine analysis [32].

There is no apparent reason why a coarse-to-fine approach would not apply in the case of the motion equation based on PDFs. Larger spatial domains increase

the accuracy of the velocity vectors, and makes the method more tolerant to large motions. To support this assertion, the Translating tree sequence has been analyzed while skipping one frame over two. Using a domain size of 7×7 pixels, the parametric method yielded an aggregate error of 1.0° (density 24%), which is better than the error of 2.9° (density 22%) obtained in these conditions with the LK method. As stressed in the previous section, larger domain sizes tend to make more fuzzy the position of the velocity vectors. However, within the framework of a coarse-to-fine approach, such a limitation is to be expected. Furthermore, the proposed methods based on PDFs could benefit from regularization schemes such as proposed for instance by Ju et al. [13]. However, the main goal of this work was before all to demonstrate the potential of gradient direction in estimating the optical flow, based on the OFC equation, without resorting to warping or regularization techniques to improve the results.

An interesting property of the proposed parametric method is that for a given flow field density, the resulting velocity vectors can be computed in regions where the LK method would fail due to the weakness of the brightness constraint. This is especially evident for the Yosemite sequence where in spite of the higher density of the LK's flow field (50 versus 30%) the velocity vectors corresponding to the translating clouds in the upper part of the image are not well represented. The fact that the LK method tends to over-represent motion vectors in image regions where brightness is particularly stable explains the slightly better quantitative performances obtained on some synthetic sequences. Also, the capacity of the parametric method in estimating velocity vectors in image regions where lighting conditions vary between successive frames reflects a larger insensitivity of the image gradients to changes in illumination direction, a property recently pointed out for real world images using probabilistic measures [14].

Finally, compared to second-order differential techniques, such as those proposed by Nagel [16] and Uras et al. [17], which have also been included in the comparison study of Barron et al. [7] and Galvin et al. [25], the proposed parametric method yielded better results (to the exception of the translating square and tree), in spite of the techniques used by these authors to improve their results. For instance, Uras et al. applied a specific scheme consisting in dividing the image into 8×8 pixel regions and looking for the eight estimates from each region that best satisfy a constraint. Of these, they further chose the estimate with the largest confidence, and assigned the recovered velocity to the entire region. Similarly it must be mentioned that Nagel [16] introduced a regularization constraint imposed on the motion field (in the direction perpendicular to the spatial gradients) to improve the results, particularly at occlusions. As for the angular velocity, and contrary to Uras et al.'s approach, the proposed method based on gradient directions makes possible to retrieve this component without having to

invert a matrix, which authorizes its evaluation on extended linear edges, subject to the aperture problem (where the matrix would be singular). This is feasible because from one frame to the next rotating image structures make the PDFs shift in ψ , and from the time rate of these shifts the angular velocity can be estimated.

6. Conclusions

A variety of statistical methods have been proposed in the literature to improve motion estimation. Markov random field models, some based on a Bayesian formulation, have been used to regularize the solutions, while taking into account occlusions and flow field discontinuities at the boundaries [9,33,34]. The optical flow being only an approximation of the motion flow, several authors proposed to view it as a probabilistic quantity [13,21,35,36]. At a lower level, the uncertainty in the gradient constraint has also been considered in order to take into account image noise, filter inaccuracies, quantization, etc. in the estimation of the derivatives. This aspect has previously been explored for brightness [35,37,38], and its generalization to gradient direction represents a cornerstone of this study. Extension of this work to other visual features than gradient direction has not been considered but is a straightforward possibility as the proposed formalism based on PDFs, to the exception of the rotational component, is general. Moreover, the possibility of recovering the angular velocity based on shifts of the PDFs represents a novel approach in motion estimation, which has also been successfully validated in this study. Peculiar to this approach is the fact that the rotational components can be estimated even in regions subject to the aperture problem.

Within this statistical framework it has been shown on the basis of quantitative and qualitative results performed on simple and on complex image sequences that gradient direction is a reliable cue for estimating motion. Compared to the LK method, described in the literature as the most accurate differential technique [7,25], the proposed parametric method based on gradient directions is shown to be more robust in regions with lighting variations between two frames. Such increased robustness is obtained at the price of a higher computational cost, mainly stemming from the low-pass filtering stage, which asks for a 4D smoothing kernel. Overall, the success of the methods based on gradient direction to estimate motion is accountable on the larger insensitivity of this visual cue to global light changes, and to changes in illumination direction [14]. Furthermore, the performance of these methods can further be improved if the a priori knowledge concerned with the position of the edges [18,19] is also taken into account to lessen the effects of small brightness variations on motion estimation.

Acknowledgements

The author is grateful to Alan Yuille and to Pierre-François Ruedi for their fruitful comments on an earlier version of this paper. This work profited from discussions with Friedrich Heitger, Pascal Nussbaum, and Florent Glück.

References

- [1] A. Mitiche, P. Bouthemy, Computation and analysis of visual motion: a synopsis of current problems and methods, *International Journal Computer Vision* 19 (1996) 29–55.
- [2] H. Haussecker, Motion, in: B. Jähne, H. Haussecker, P. Geissler (Eds.), *Handbook of Computer Vision and Applications*, Academic Press, New York, 1999, pp. 310–396, Chapter 13.
- [3] G. Fennema, W. Thompson, Velocity determination in scene containing several moving objects, *Computer Graphics and Image Processing* 9 (1979) 301–315.
- [4] B. Horn, B. Schunck, Determining optical flow, *Artificial Intelligence* 17 (1981) 185–203.
- [5] A. Verri, T. Poggio, Motion field and optical flow: qualitative properties, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (1989) 490–498.
- [6] D. Marr, S. Ullman, Directional selectivity and its use in early visual processing, *Proceedings of the Royal Society of London* 211(B) (1981) 151–180.
- [7] J. Barron, D. Fleet, S. Beauchemin, Performance of optical flow techniques, *International Journal Computer Vision* 12 (1994) 43–77.
- [8] H.-H. Nagel, Displacement vectors derived from second-order intensity variations in image sequences, *Computer Vision, Graphics and Image Processing* 21 (1983) 85–117.
- [9] J. Konrad, E. Dubois, Bayesian estimation of motion vector fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (1992) 910–927.
- [10] B. Lucas, T. Kanade, An iterative, image registration technique with an application to stereo vision, *Proceedings of the DARPA Image Understanding Workshop*, 1981, pp. 121–130.
- [11] A. Mitiche, Y. Wang, J. Aggarwal, Experiments in computing optical flow with the gradient-based multiconstraint method, *Pattern Recognition* 20 (1987) 173–179.
- [12] A. DelBimbo, P. Nesi, J. Sanz, Optical flow computation using extended constraints, *IEEE Transactions on Image Processing* 5 (1996) 720–739.
- [13] S. Ju, M. Black, A. Jepson, Skin and Bones: Multi-layer, Locally Affine, Optical Flow and Regularization with Transparency, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1996, pp. 307–314.
- [14] H. Chen, P. Belhumeur, D. Jacobs, In search of illumination invariants, *Proceedings of the International Conference on Computer Vision and Pattern*, 2000, pp. 254–261.
- [15] J. Weng, T.S. Huang, N. Ahuja, *Motion and Structure from Image Sequences*, Springer, Berlin, 1993.
- [16] H.-H. Nagel, On the estimation of optical flow: relations between different approaches and some new results, *Artificial Intelligence* 33 (1987) 299–324.
- [17] S. Uras, F. Girosi, A. Verri, V. Torre, A computational approach to motion perception, *Biological Cybernetics* 60 (1988) 79–87.
- [18] S. Konishi, A. Yuille, J. Coughlan, S. Zhu, Fundamental bounds on edge detection: an information theoretic evaluation of different edge cue, *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 1999, pp. 573–579.
- [19] R. Balboa, N. Grzywacz, The minimal local-asperity hypothesis of early retinal lateral inhibition, *Neural Computation* 12 (2000) 1485–1517.
- [20] A. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
- [21] P.-Y. Burgi, A. Yuille, N. Grzywacz, Stochastic motion estimation based on temporal coherence, *Neural Computation* 12 (2000) 1839–1867.
- [22] J. Brandt, Improved accuracy in gradient-based optical flow estimation, *International Journal Computer Vision* 25 (1997) 5–22.
- [23] M. Barbaro, P.-Y. Burgi, A. Mortara, P. Nussbaum, F. Heitger, A 100 × 100 pixel silicon retina for gradient extraction with steering filter capabilities and temporal output coding, *IEEE Journal of Solid-State Circuits* 37 (2002) 160–172.
- [24] P.-F. Ruedi, P. Heim, F. Kaess, E. Grenet, F. Heitger, P.-Y. Burgi, S. Gyger, P. Nussbaum, A 128 × 128 pixel 120-dB dynamic-range vision-sensor chip for image contrast and orientation extraction, *IEEE Journal of Solid-State Circuits* 38 (2003) 2325–2333.
- [25] B. Galvin, B. McCane, K. Novins, D. Mason, S. Mills, Recovering Motion Fields: an Evaluation of Eight Optical Flow Algorithms, *Proceedings of the British Machine Vision Conference*, 1998, pp. 195–204.
- [26] K. Wahn, L. Davis, P. Thrift, Motion estimation based on multiple local constraints and nonlinear smoothing, *Pattern Recognition* 16 (1983) 563–570.
- [27] M. Swain, D. Ballard, Color indexing, *International Journal of Computer Vision* 7 (1991) 11–32.
- [28] Schiele, B., Crowley, J., Recognition without correspondence using multi-dimensional receptive field histograms, *Tech. Rep. 453*, M.I.T. Media Laboratory, 1997.
- [29] B. Mel, Seemore: combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition, *Neural Computation* 9 (1997) 777–804.
- [30] D. Fleet, A.D. Jepson, Computation of component image velocity from local phase information, *International Journal of Computer Vision* 5 (1990) 77–104.
- [31] H. Liu, T.H. Hong, M. Herman, R. Chellapa, A general motion model and spatio-temporal filters for computing optical flow, *International Journal Computer Vision* 22 (1997) 141–172.
- [32] M. Black, A. Jepson, Estimating optical flow in segmented images using variable-order parametric models with local deformations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18 (1996) 972–986.
- [33] F. Heitz, P. Bouthemy, Multimodal estimation of discontinuous optical flow using Markov random fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15 (1993) 1217–1232.
- [34] S. Li, *Markov Random Field Modeling in Computer Vision*, Springer, Berlin, 1995.
- [35] E. Simoncelli, Bayesian multiscale differential optical flow, in: B. Jähne, H. Haussecker, P. Geissler (Eds.), *Handbook of Computer Vision and Applications*, Academic Press, New York, 1999, pp. 397–422, Chapter 14.
- [36] M. Chang, A. Tekalp, M. Sezan, Simultaneous motion estimation and segmentation, *IEEE Transactions on Image Processing* 6 (1997) 1326–1333.
- [37] M. Luetgen, W. Karl, A. Willsky, Efficient multiscale regularization with applications to the computation of optical flow, *IEEE Transactions on Image Processing* 3 (1994) 41–64.
- [38] M. Black, P. Anandan, The robust estimation of multiple motions: parametric and piecewise-smooth flow fields, *Computer Vision and Image Understanding* 63 (1996) 75–104.