

# A New Approach to Motion Pattern Recognition and Its Application to Optical Flow Estimation

J. Chamorro-Martínez and J. Fernández-Valdivia

**Abstract**—In this paper, a new methodology for extracting motion patterns is applied to optical flow estimation in the presence of multiple motions. The proposed approximation deals with the problem in two stages. In the first one, the most important motions are segmented; in the second one, the optical flow is estimated on the basis of the motions detected in the previous stage. To extract relevant motions, a new approach based on a spatio-temporal filtering is presented. The approach groups together parts of a moving object that have been separated into various filter responses because of the object's spatial structure, thereby avoiding the spatial dependency problem associated with a representation based on spatio-temporal filters. The proposed model, therefore, generates one "motion pattern" for each motion detected in the sequence. To obtain an optical flow estimation, which is able to represent multiple velocities, the gradient constraint is applied to the output of each filter so that multiple estimations of the velocity at the same location may be obtained. For each "motion pattern" detected in the previous stage, the velocities at a given point corresponding to the same motion are then combined using a probabilistic approach. In the application to optical flow estimation, the use of "motion patterns" allows multiple velocities to be represented, while the combination of estimations from different filters helps reduce the aperture problem. This technique is illustrated on real and simulated data sets, including sequences with occlusion and transparencies.

**Index Terms**—Motion segmentation, multiple motions, optical flow, spatio-temporal.

## I. INTRODUCTION

MOTION analysis is a fundamental task in video processing. A good detection and an adequate representation of motions are basic steps in video applications, such as coding, object tracking, or robot navigation. In this framework, motion segmentation, i.e., the process of dividing the scene into regions representing moving objects, is an open problem.

The most common proposals in the literature to motion segmentation rely on a frame-by-frame analysis, where feature correspondences between consecutive frames are investigated [1], [2]. Although this kind of approach works well in many cases, it is common for there to be problems in the presence of noise, occlusions, or transparencies. To overcome such problems, some authors propose the use of extended features to find correspondences between frames. Nevertheless, the success of these models depends on the stability of detection of such features over multiple frames, and the way of solving the

correspondence problem [3]. In this sense, some interesting approaches rely on geometrical approximations [4]–[8], which face the problem of segmenting multiple (rigid) motions on the basis of feature correspondences in multiple (affine) views of the scene [9]–[11]. Apart from the problems caused by occlusions or transparencies, one of the main inconveniences of these approaches is the need for two or more views of a scene, something that may not always be available.

Unlike frame-by-frame analysis (or analysis over a small number of frames), some approaches represent the sequence as a spatio-temporal volume. From this point of view, a moving object may be observed as a three-dimensional (3-D) object, where the  $x$  and  $y$  axes correspond to the spatial dimensions, and the third axis corresponds to the temporal dimension [12]. Therefore, the goal of this kind of method is to find those spatio-temporal volumes corresponding to significant motions. In this framework, some important proposals are based on a bandpass filtering operation with a set of spatio-temporal filters [13]–[17]. These approaches are derived by considering the motion problem in the Fourier domain, where the spectrum of a spatio-temporal translation lies in a plane, the orientation of which depends on the direction and velocity of the motion [18]–[21]. This kind of technique is especially interesting to face cases of occlusion and transparencies, since two objects with different motions will be separated into different areas in the frequential spectrum, despite if one of them occludes the other or presents a certain degree of transparency.

Although spatio-temporal filters provide a powerful tool for separating the motions presented in a sequence, it is nevertheless true that one of the main problems of these schemes is that components of the same motion with different spatial characteristics are separated into different responses. In addition, a filter response will be modified if the spatial orientation or the scale changes (this depends on the spatial orientation of the underlying signal). This spatial orientation dependence can be clearly seen in the example in Fig. 1(a). Next to the original sequence (corresponding to a square moving), the output of two filters is shown. As the figure shows, the perpendicular sides of the square are separated in two different responses. Although all the parts of the object have the same movement, these will, therefore, be separated into different responses. This case becomes worse with more complex movements (for example, rotation) and with more spatially complex objects. For example, in the case of the rotating bar as shown in Fig. 1(b), the difficulty stems from the fact that the movement of the object generates a 3-D shape with a large number of spatio-temporal orientations. Since the filters are selective to orientations, this movement, as the figure shows, is dispersed among various filters.

Manuscript received March 3, 2004; revised July 14, 2004. This work was supported by the MCYT (Spain) under Grant TIC2003-01504. This paper was recommended by Editor V. Marik.

The authors are with the Department of Computer Science and Artificial Intelligence, University of Granada, Spain (e-mail: jesus@decsai.ugr.es; jfv@decsai.ugr.es).

Digital Object Identifier 10.1109/TSMCC.2006.876044

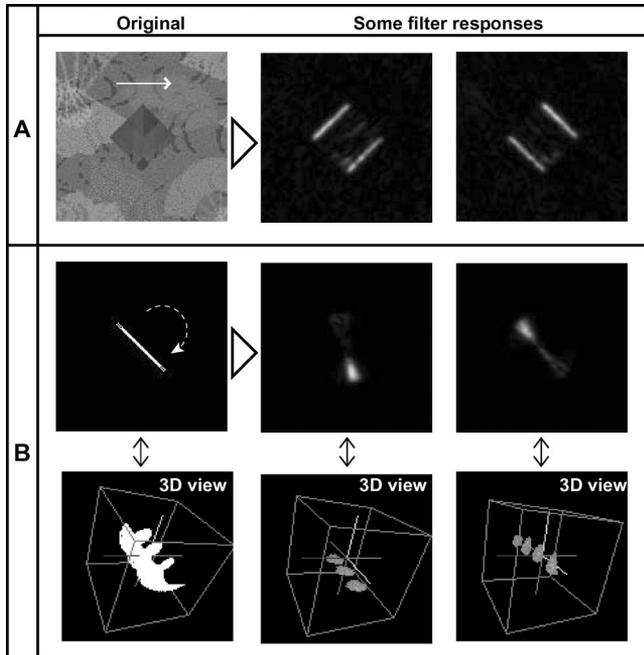


Fig. 1. Examples of spatio-temporal filter responses.

This problem, inherent to all the models based on spatio-temporal filtering, is carefully examined in this paper. Concretely, we develop a methodology for motion segmentation on the basis of a spatio-temporal filtering approach, in such a way that a motion will be detected as a spatio-temporal volume in the 3-D space. To solve the problems described above concerning to the use of filters, we propose a novel approach that groups the separated responses obtained by the filters to extract coherent and independent motions. Using a new distribution of 3-D logGabor filters over the spatio-temporal spectrum, a *motion* is detected as an invariance in statistical structure across a range of spatio-temporal frequency bands. This new scheme recomposes responses that, even with different spatial characteristics, have continuity in its motion. Preliminary results were obtained in [22] for translation motions; in this paper, we improve the different stages of the model, allowing us to face more complex motions like rotations or articulated object movements.

Dealing with multiple motion is also an important problem when estimating optical flow [23], [24]. With the presence of occlusions and transparencies, more than one velocity may be presented at the same point (for example, let us consider a sheet of glass crossing over an opaque object). In such cases, the techniques that do not consider the presence of multiple motions will generate erroneous estimations, which will combine the different velocities present at one point into a single vector. These problems are currently being addressed by the research community with models such as those based on the use of mixed velocity distributions (usually two) at each point [25], the models based on line processes [26], the parametric models [27], or the frequency-based techniques (which use the aforementioned spatio-temporal filters to separate the motions [28], [29]). Nevertheless, although they do consider the presence of occlusions and transparencies in their calculations, the majority of these

techniques do not generate a representation as an output, which allows more than one velocity per point. This means that, in cases like that of transparency, the result does not really reflect the movement associated with the point [30], [31].

To confront this problem, on the basis of the motion patterns detected in the previous stage, in this paper we develop an application for optical flow estimation in presence of multiple motions. The main novelties of this technique, introduced in [32] for simple motions and extended here for more complex movements, are its capability to represent multiples velocities at the same point and its adaptability to the cases where there is motion with occlusions and transparencies.

The rest of the paper is organized as follows. Section II describes the new method to motion pattern recognition and Section III shows its application to optical flow estimation. The performance of the model is evaluated in Section IV, where results with real and synthetic sequences are shown. Finally, the main conclusions are summarized in Section V.

## II. MOTION PATTERN RECOGNITION

In this section, the methodology used to detect multiple motions is presented. There are three main stages to this analysis: a spatio-temporal filtering, the computation of the distance between filter responses, and a clustering process. A diagram illustrating the analysis on a sequence corresponding to a handclap is shown in Fig. 2 (in this example, the objective is to separate the two hand motions).

In the first stage, the original sequence is represented as a spatio-temporal volume, where a moving object corresponds to a 3-D pattern. Its Fourier transform is then calculated to perform the analysis in the frequential domain. Given a bank of spatio-temporal logGabor filters [33], a subset of these is selected so that significant spectral information may be extracted. These selected filters are applied over the original spatio-temporal image so that a set of active responses may be obtained (it should be noted that only one subset of filters is used).

In the second stage, the filter responses are compared on the basis of the distance between statistical structures. These distances are computed over relevant points that are calculated as local energy peaks on the filter response. As a result, the distances between active filters are obtained.

In the third stage, a clustering over the set of active filters is performed to highlight response invariance. Each cluster obtained in this step defines a motion pattern. In the output box of Fig. 2, two collections of filters corresponding to the two hand motions are shown.

In Sections II-A–II-C, we shall explain these three stages in greater detail.

### A. Spatio-Temporal Filtering

The aim of this first stage is to obtain a set of filters that responds to the most significant motions (or motion parts) in the sequence (see the “Spatio-temporal filtering” box in Fig. 2). To decompose the input sequence, a bank of logGabor filters is used. These quadrature filters minimize the uncertainty relation (like Gabor functions) and have no dc component by definition.

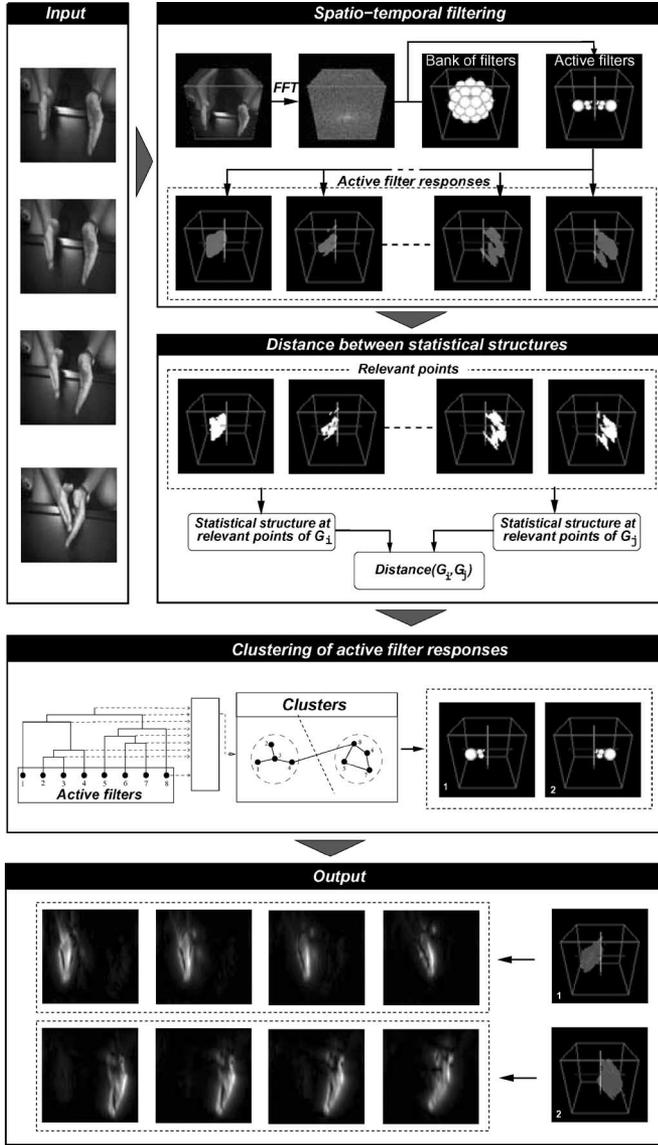


Fig. 2. General diagram describing the model.

Furthermore, several publications claim that logGabor filters provide a better description of the cortical cell response of the visual system in mammals [34].

A logGabor function can be represented in the frequential domain as

$$\phi(\rho, \theta, \varphi) = e^{-\frac{(\log(\frac{\rho}{\rho_o}))^2}{2(\log(\frac{\sigma_\rho}{\rho_o}))^2}} e^{-\frac{(\theta - \theta_o)^2}{2\sigma_\theta^2}} e^{-\frac{(\varphi - \varphi_o)^2}{2\sigma_\varphi^2}} \quad (1)$$

where  $\sigma_\theta$ ,  $\sigma_\varphi$ , and  $\sigma_\rho$  are the angular and radial standard deviations,  $(\theta_o, \varphi_o)$  is the orientation of the filter, and  $\rho_o$  is the central radial frequency. The separability of (1) reduces the computational cost.

The convolution of a real sequence with a logGabor function (whose real and imaginary parts are in quadrature) results in a complex spatio-temporal image. The local energy can be expressed as [35]

$$E(x, y, z) = \sqrt{O_{\text{even}}^2(x, y, z) + O_{\text{odd}}^2(x, y, z)} \quad (2)$$

 TABLE I  
 ANGULAR COORDINATES ASSOCIATED WITH THE BANK OF FILTERS (OVER AN SPHERE OF RATIO 1)

$(\theta, \varphi)$		
(0.52, 0.00)	(-0.62, -0.53)	(-1.29, 0.45)
(0.62, 0.53)	(-1.08, 0.97)	(-1.05, 0.00)
(1.08, 0.97)	(1.08, -0.97)	(-1.29, -0.45)
(-1.08, 0.97)	(0.62, -0.53)	(1.29, -0.45)
(-0.62, 0.53)	(1.05, 0.00)	(1.57, 0.00)
(-0.52, 0.00)	(1.29, 0.45)	-

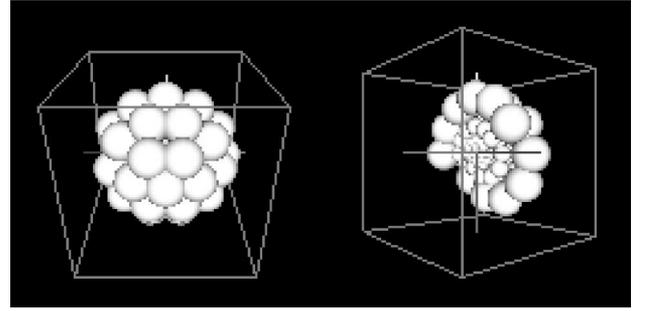


Fig. 3. Distribution of spatio-temporal filters.

where  $O_{\text{even}}(x, y, z)$  and  $O_{\text{odd}}(x, y, z)$  are the result of convolving the original sequence with the even-symmetric and odd-symmetric logGabor filter, respectively.

In Section II-A1, we shall explain how to distribute the filters, how to select a subset of filters corresponding to significant motions, and how to select relevant points in a given filter.

1) *Distribution of Filters*: The bank of filters should be designed so that it tiles the frequential space uniformly. In our model, we shall employ a spherical scheme with spatio-temporal multiresolution that distributes the center of the filters on the surface of spheres with different radii (one for each considered resolution level). In the Appendix, we propose an algorithm to uniformly distribute the filter centers in the frequential space. By using that algorithm, we can vary the number and distribution of the filters. The higher the number of filters employed, the higher the selectivity of the filters, but the higher the computational cost also. On the contrary, if we diminish the number of filters, the execution time will be reduced, but the filters will be less selective. In this paper, different distributions have been analyzed, and finally, a bank consisting of 51 filters (whose spherical orientations have been obtained by applying algorithm 1 with  $N = 6$ ) has been chosen that gives us a tradeoff between the computational cost and selectivity of the bank. Therefore, we propose a bank with the following features.

- 1) For each radial frequency, we deal with 17 spherical orientations over dynamic planes. The angular coordinates used in the proposed bank are shown in Table I.
- 2) The radial axis is divided into three equal octave bands. The wavelength in each orientation is set to 3, 6, and 12 pixels, respectively.
- 3) The radial bandwidth is set to 1.2 octaves.
- 4) The angular bandwidth is set to  $30^\circ$ .

The resultant collection of filters  $\phi_1, \phi_2, \dots, \phi_G$ , with  $G = 51$ , is illustrated in Fig. 3. Due to conjugate symmetry in the Fourier domain, the filter is only designed on half the 3-D frequential

space. The logGabor filters are displayed by spheres drawn in the 3-D frequential space at the point where the amplitude has decreased to  $e^{-1/2}$  of its maximum

2) *Active Filters*: To reduce the computational cost and to avoid noisy or less relevant filter responses, filters that isolate spectral information corresponding to significant motions are selected. Given a filter  $\phi_i$ , a measure of its relevance is defined as

$$w_i = \frac{1}{\text{Card}[V(i)]} \sum_{(\rho, \theta, \varphi) \in V(i)} |F(\rho, \theta, \varphi)| \quad (3)$$

where  $|F(\rho, \theta, \varphi)|$  is the amplitude of the Fourier spectrum at  $(\rho, \theta, \varphi)$ , and  $V(i)$  represents a spectral volume associated with the filter  $\phi_i$ . To calculate  $V(i)$ , we consider that a point  $(\rho, \theta, \varphi)$  in the spatio-temporal frequential domain will belong to  $V(i)$  if

$$|\rho - \rho_o| \leq \sigma_\rho \quad |\theta - \theta_o| \leq \sigma_\theta \quad |\varphi - \varphi_o| \leq \sigma_\varphi \quad (4)$$

where  $\sigma_\theta, \sigma_\varphi, \sigma_\rho$ , and  $(\theta_o, \varphi_o)$  are the logGabor filter parameters (we should mention that it is not necessary to calculate the responses of each filter for these weights to be obtained).

Using the filter relevance measure defined in (5), an unsupervised classification method is performed for each scale so that the filters may be grouped into two classes: active and non-active. The cluster whose filters have the highest weights will determine the set of active filters (henceforth, denoted *Actives*). In our implementation, a hierarchical clustering [36] is used with a dissimilarity function between classes defined as

$$\Gamma(C_i, C_j) = |\mu_i - \mu_j| \quad (5)$$

where

$$\mu_k = \frac{1}{\text{Card}[C_k]} \sum_{r \in C_k} w_r. \quad (6)$$

3) *Selection of Relevant Points*: For each active filter, a set of “relevant points” is computed, and we shall focus on these at a later stage. These points are calculated via local energy peaks on the filter responses [37]: given the local energy map  $E_i$  of a filter  $\phi_i$  (2), the maximal of  $E_i$  in the direction of the filter will determine the set of relevant points.

The use of peaks on local energy is based on Morrone *et al.*'s work [38], [37]. They postulate that features are perceived at points where the Fourier components are maximally in phase, and for these points of phase congruency to be detected, they use an energy function extracted by using the standard method of squaring the outputs of two filters that are in a quadrature. Features are then signaled by peaks in local energy functions. Hence, the detection of peaks on the  $E_i$  map acts as a detector of significant features on the filter response.

In [22], we introduce the use of local energy peaks for selecting “relevant points,” but in that case the number of peaks is very high and the execution time increases significantly. To extract less peaks (just the most significant ones), we introduce in this paper a selection based on a hysteresis procedure [39]. We begin by selecting a set of “sure” points (as in [39], we select those points with an energy exceeding 95% in the accumulated histogram as peaks) and, for the remaining peaks, we select

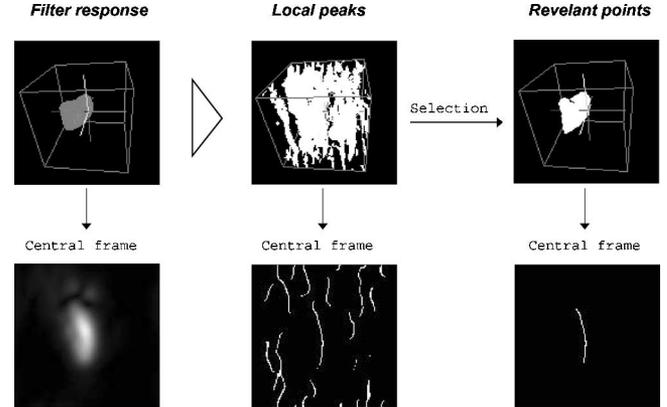


Fig. 4. Example of relevant points selection.

those for which there is a path to a “sure” point in which all the points in the path have an energy that is equal to or higher than a threshold (we set this threshold as the energy associated with 50% in the accumulated histogram). For more details about this selection method, see [39].

Fig. 4 shows an example of relevant point selection. An active filter response and its relevant points are shown for a sequence corresponding to a handclap (see Fig. 2). The top row shows the point selection procedure: Given the spatio-temporal response, the peaks of the local energy are obtained and then a subset of these is selected. The bottom row shows the central frame associated with each 3-D image.

To conclude this first stage (see the “Spatio-temporal filtering” box in Fig. 2), it is worth mentioning that we have obtained a set of filters (which we have called “Actives”) for the results of this step and these correspond to the most significant motions (or motion parts) in the sequence. These filters will be the starting point for future stages when we will attempt to segment motions on the basis of their responses.

## B. Distance Between Filter Responses

In this second stage, given a decomposition of the original sequence into its most significant components (that is to say, the subset of active filters), a new element is needed: a distance between the statistical structures of two filter responses. To represent a statistical structure, we use the concepts of *separable feature* and *integral feature*, which were introduced in [40]. A separable feature is defined as any relevant characteristic that may be obtained at one point (phase, local contrast, energy, etc.). The combination of any subset of separable features will define an integral feature at a given point  $(x, y, z)$ . In this paper, the following five separable features proposed in [22] and [40] will be used: phase, local energy, local standard deviation, local contrast of the local energy, and local entropy. In addition, three new separable features, more related with apparent motion information, have been tested: the first and second derivatives at each point, and the estimation of the apparent motion given by (23).

Let  $T^i(x, y, z) = [T_k^i(x, y, z)]_{k=1,2,\dots,L}$  be an integral feature at  $(x, y, z)$ , which combines  $L$  separable features, denoted  $T_k^i$ , computed on the response of the filter  $\phi_i$ . Let  $\hat{d}(T^i, T^j)$

be the distance between two integral features  $T^i(x, y, z)$  and  $T^j(x, y, z)$  given by the equation

$$\hat{d}(T^i, T^j) = \sum_{k=1}^L \frac{1}{\text{Max}_k} d(T_k^i, T_k^j) \quad (7)$$

with  $\text{Max}_k$  being a normalization factor [40], and  $d(\cdot)$  a distance between separable features [this measure  $d(\cdot)$  is defined for each separable feature in [40]].

Based on the previous equation, a distance between the responses of two filters  $\phi_i$  y  $\phi_j$  is defined as

$$\hat{D}(\phi_i, \phi_j) = D[i, j]^2 + D[j, i]^2 \quad (8)$$

where

$$D[r, s] = \frac{1}{\text{Card}[P(r)]} \left( \sum_{P(r)} |\hat{d}[T^r, T^s]|^\beta \right)^{\frac{1}{\beta}} \quad (9)$$

with  $\hat{d}[T^r, T^s]$  being the distance between integral features given by (7), and  $P(r)$  the set of relevant points for the filter  $\phi_r$ .

The default value of the exponent  $\beta$  in (9) is 3. Graham [41] discussed several interpretations of the Quick pooling formula and the selection of the pooling exponent at some length.

To summarize this second stage (see the ‘‘Distance between statistical structures’’ box in Fig. 2), the distance defined in this section enables us to calculate the difference between any pair of active filters (obtained in the previous stage). This distance will be necessary in later stages for a clustering process to be carried out. For this distance to be calculated, two important aspects have been incorporated: First, the selection of points of interest on which to focus the analysis of the filter; and second, the use of a feature vector (which we have called ‘‘integral feature’’) for characterizing each relevant point.

### C. Clustering of Active Filters

The aim of this final stage is to group active filters (obtained in the first stage), which have motion continuity (see the ‘‘Clustering of active filter responses’’ box in Fig. 2). For this, we shall use the distance defined in Section II-B and we shall search for filter groupings (which we shall call ‘‘motion patterns’’), which will characterize motions of the original sequence.

To obtain a partition  $S_1, S_2, \dots, S_N$  of active filters, with  $S_i$  representing a motion pattern, a clustering of the dataset  $X = \{\phi_i \in \text{Actives}\}$  into an unknown number  $N$  of clusters is performed. For this purpose, a hierarchical clustering is used [36] with a dissimilarity function between classes defined on the basis of distances between statistical structures as

$$\delta(S_n, S_m) = \min\{\hat{D}(\phi_i, \phi_j), \phi_i \in S_n, \phi_j \in S_m\} \quad (10)$$

where  $\hat{D}(\phi_i, \phi_j)$  is given by the (8).

1) *Selection of the Best Partition:* In the clustering process, we compute all levels of the hierarchy (as the number of active filters is usually small, the computational cost is not increased). To select level  $l$  of the hierarchy, which will define the best partition  $P^l = S_1, S, \dots, S_N$ , we shall propose the following

goodness function:

$$f(P^l) = \frac{\gamma_{P^l}^*}{\varepsilon_{P^l}^*} \quad (11)$$

where  $\varepsilon_{P^l}^*$  and  $\gamma_{P^l}^*$  are two measures of the congruence and separation of the partition  $P^l$ , respectively, given by the equations

$$\varepsilon_{P^l}^* = \max\{\varepsilon_n | S_n \in P^l\} \quad (12)$$

$$\gamma_{P^l}^* = \min\{\gamma_n | S_n \in P^l\}. \quad (13)$$

The congruence degree  $\varepsilon_n$  and the separation degree  $\gamma_n$  of a cluster  $S_n$  are defined as

$$\varepsilon_n = \max\{\text{cost}(\pi_{i,j}^*) | \phi_i, \phi_j \in S_n\} \quad (14)$$

$$\gamma_n = \min\{\delta(S_n, S_m) | m = 1, \dots, N \text{ with } m \neq n\} \quad (15)$$

where  $\delta(S_n, S_m)$  is defined in (10), and  $\text{cost}(\pi_{i,j}^*)$  is the cost of the optimal path between two elements  $\phi_i$  and  $\phi_j$  in  $S_n$ , which is calculated in the following way: Let us consider a complete graph, denoted  $G_{S_n}$ , where each node corresponds to a filter in  $S_n$  and where each arc has a weight calculated by using the distance between filters given by (8). Let  $\prod_{ij}$  be the set of possible paths linking  $\phi_i$  and  $\phi_j$  in  $G_{S_n}$ ; given a path  $\pi_{ij} \in \prod_{ij}$ , its cost is defined as the greatest distance between two consecutive points on the path

$$\text{cost}(\pi_{ij}) = \max\{\hat{D}(\phi_r, \phi_{r+1}) / \phi_r, \phi_{r+1} \in \pi_{ij}\} \quad (16)$$

where  $\phi_r$  and  $\phi_{r+1}$  are two consecutive elements of  $\pi_{ij}$ , and  $\hat{D}(\phi_r, \phi_s)$  is defined in (8). The optimum path  $\pi_{ij}^* \in \prod_{ij}$  between  $\phi_i$  and  $\phi_j$  is then defined as the path that links both filters with minimum cost

$$\pi_{ij}^* = \underset{\pi_{i,j} \in \prod_{i,j}}{\text{argmin}} \{\text{cost}(\pi_{i,j})\}. \quad (17)$$

Due to the merging process of the hierarchical clustering and the distance between classes used in this case (10), the congruence degree  $\varepsilon_n$  is equal to the distance between the two clusters, which were merged together to obtain  $S_n$  [40]. The calculation of  $\varepsilon_n$  does not, therefore, increase the computational cost of the clustering.

To summarize this last stage (see the ‘‘Clustering of active filter responses’’ box in Fig. 2), we should point out that by way of result, we obtain a set of clusters (each of which groups a set of active filters) with the special characteristic that each cluster matches a motion of the sequence. The motion associated with each cluster is what we have called the ‘‘motion pattern.’’

## III. OPTICAL FLOW ESTIMATION

In this section, the frequency-based model described in Section II will be used to obtain an optical flow estimation that is capable of representing multiple motions. First, the differential formulation to optical flow estimation will be introduced in Section III-A. This formulation will be used in Section III-B to calculate velocity estimations from each active filter. As the motion patterns detected by our model correspond to clusters

of active filters, in Section III-C, we shall propose a methodology to integrate the estimations of the same cluster to represent multiple velocities at the same location (Section III-D).

### A. Differential Formulation

While many methods for computing optical flow have been proposed in literature, a study by Barron *et al.* [42] suggest that the gradient-based approaches are the most accurate. In addition, they also offer an efficient solution that produces subpixel displacement estimations. By writing image intensity as a function of position and time  $f(x, y, t)$ , the gradient formulation of optical flow is based on the following differential brightness constancy constraint equation:

$$\mathbf{f}_e \cdot \mathbf{v} + f_t = 0 \quad (18)$$

where  $\mathbf{f}_e = (f_x, f_y)$  and  $f_t$  are the spatial and temporal derivatives, respectively, of image  $f$  at a given spatial location and time  $(x, y, t)$ , and  $\mathbf{v} = (v_x, v_y)^T$  is the two-dimensional (2-D) velocity vector at location  $(x, y, t)$  (for notational simplicity, spatial and temporal location parameters are omitted). This formulation assumes that changes in image intensity are only due to the translation of local image intensity and not due to changes in lighting or reflectance. As (18) places a single linear constraint, it cannot be solved unless some additional constraints are imposed. Typically, this problem is overcome by imposing various global smoothness constraints [43] or alternatively, a locally constant velocity may be assumed and linear constraints combined over local regions [44].

Within the gradient-based approaches, a probabilistic framework to optical flow estimation was proposed by Simoncelli *et al.* [45]. In this approach, the authors propose the computation of an expression for the probability of the image velocity conditional to the image gradient  $P(\mathbf{v} | \mathbf{f}_e, f_t)$ . Two independent additive Gaussian noise terms  $\mathbf{n}_1$  and  $n_2$  are introduced in (18) to describe first, the uncertainty in the assumptions of the constraint equation, and second, the errors in the temporal derivative computations. Thus, (18) is rewritten as

$$\mathbf{f}_e \cdot (\mathbf{v} - \mathbf{n}_1) + f_t = n_2, \quad n_i \sim N(0, \Delta_i) \quad (19)$$

which describes the conditional probability  $P(f_t | \mathbf{v}, \mathbf{f}_e)$ . Choosing a zero-mean Gaussian with covariance  $\Delta_p$  for the prior distribution  $P(\mathbf{v})$ , we can use Bayes' rule to compute an expression for the probability  $P(\mathbf{v} | \mathbf{f}_e, f_t)$ . The resulting distribution is Gaussian and if we choose  $\Delta_1$  to be a diagonal matrix, with diagonal entry  $\kappa_1$ , and write the scalar variance of  $n_2$  as  $\kappa_2 \equiv \Delta_2$ , the mean and covariance of the Gaussian distribution will be given by [45]

$$\mu_{\mathbf{v}} = -\Delta_{\mathbf{v}} \cdot \frac{\mathbf{d}}{\kappa_1 \|\mathbf{f}_e\|^2 + \kappa_2} \quad (20)$$

$$\Delta_{\mathbf{v}} = \left[ \frac{\mathbf{M}}{\kappa_1 \|\mathbf{f}_e\|^2 + \kappa_2} + \Delta_p^{-1} \right]^{-1} \quad (21)$$

with

$$\mathbf{M} = \begin{bmatrix} f_x^2 & f_x f_y \\ f_y f_x & f_y^2 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} f_x f_t \\ f_y f_t \end{bmatrix}. \quad (22)$$

This solution computes velocity for one point in isolation, but the constraint at a single location is insufficient to uniquely specify a solution. To solve this, we can combine information in small neighborhoods (as in [44]). Assuming that the noise at each point in the neighborhood is independent, the resulting mean and variance are given by [45]

$$\mu_{\mathbf{v}} = -\Delta_{\mathbf{v}} \cdot \sum_r \frac{w_r \mathbf{d}_r}{\kappa_1 \|\mathbf{f}_e(x_r, y_r, t)\|^2 + \kappa_2} \quad (23)$$

$$\Delta_{\mathbf{v}} = \left[ \sum_r \frac{w_r \mathbf{M}_r}{\kappa_1 \|\mathbf{f}_e(x_r, y_r, t)\|^2 + \kappa_2} + \Delta_p^{-1} \right]^{-1} \quad (24)$$

where  $w_r$  is a weighting function that gives more influence to elements at the center of the neighborhood than to those at the periphery, with the points in the neighborhood indexed by  $r$ , and where  $\mathbf{M}_r = \mathbf{M}(x_r, y_r, t)$  and  $\mathbf{d}_r = \mathbf{d}(x_r, y_r, t)$  are defined as in (22).

### B. Estimation for a Spatio-Temporal Filter Response

To estimate the velocity  $\mathbf{v}_i$  at a given point  $(x, y, t)$  of the  $i$ th filter  $\phi_i$ , the probabilistic approach described in Section III-A is used. Using the odd response of the filter  $\phi_i$ , the velocity  $\mathbf{v}_i$  is, therefore, defined on the basis of a Gaussian random variable  $\mathbf{v}_i$  with mean  $\mu_{\mathbf{v}_i}$  and covariance  $\Delta_{\mathbf{v}_i}$

$$\mathbf{v}_i \sim N(\mu_{\mathbf{v}_i}, \Delta_{\mathbf{v}_i}), \quad i = 1, \dots, N \quad (25)$$

where  $\mu_{\mathbf{v}_i}$  and  $\Delta_{\mathbf{v}_i}$  are calculated using (23) and (24). Therefore, given a point  $(x, y, t)$ , we shall have an estimation  $\mathbf{v}_i$  for each active filter  $\phi_i$ ; in other words, we will have a vector of estimations  $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N]$  associated with each point, with  $N$  being the number of active filters.

*1) Confidence Measure:* Because the use of filters that are selective to spatio-temporal orientations aggravates the aperture problem, it is, therefore, necessary to use an adequate confidence measure to discard unreliable estimations. It is well known that the covariance matrix  $\Delta_{\mathbf{v}_i}$  can be used to define a confidence measure of the estimation  $\mathbf{v}_i$  [45]. In this paper, we shall use the smallest eigenvalue of  $\Delta_{\mathbf{v}_i}^{-1}$  as the confidence measure of  $\mathbf{v}_i$  [42] and this will be denoted  $\lambda_{\mathbf{v}_i}$

$$\lambda_{\mathbf{v}_i} = \min \{ \lambda_1^i, \lambda_2^i \} \quad (26)$$

where  $\lambda_1^i$  and  $\lambda_2^i$  are the two eigenvalues of  $\Delta_{\mathbf{v}_i}^{-1}$  [for the sake of simplicity, we have omitted the spatio-temporal parameters  $(x, y, t)$  in the notation  $\lambda_{\mathbf{v}_i}(x, y, t)$ ].

Therefore, an estimation  $\mathbf{v}_i$  at a given point  $(x, y, t)$  of the  $i$ th filter  $\phi_i$  will be accepted if  $\lambda_{\mathbf{v}_i} \geq \text{Threshold}_{\phi_i}$ , where  $\text{Threshold}_{\phi_i}$  is a confidence threshold associated with the filter  $\phi_i$ . Under the assumption that every relevant point of the filter will generate a reliable estimation, the following approximation is proposed to calculate  $\text{Threshold}_{\phi_i}$ :

$$\text{Threshold}_{\phi_i} = \min \{ \lambda_{\mathbf{v}_i}(x, y, t) / (x, y, t) \in P(\phi_i) \} \quad (27)$$

where  $P(\phi_i)$  represents the set of relevant points of the filter  $\phi_i$ . In this way, we accept as reliable any estimation that is the same as or better than the worst estimation obtained for the set of relevant points.

### C. Estimation for a Motion Pattern

This section shall describe the methodology for integrating the estimations corresponding to the set of filters, which comprise a motion pattern. Let  $S_k$  be the  $k$ th motion pattern detected in the sequence, and let  $\{\phi_i^k\}_{i=1,\dots,L_k}$  be the set of  $L_k$  grouped filters in  $S_k$ . Let  $\Omega_k$  be the set of estimations  $\mathbf{v}_i \sim N(\mu_{\mathbf{v}_i}, \Delta_{\mathbf{v}_i})$  obtained from  $\{\phi_i^k\}_{i=1,\dots,L_k}$ , which are above the confidence threshold. The integration will be performed on the basis of a linear combination

$$\hat{\mathbf{v}}_k = \sum_{\mathbf{v}_i \in \Omega_k} \alpha_i \mathbf{v}_i \quad (28)$$

with  $\hat{\mathbf{v}}_k$  representing the velocity at point  $(x, y, t)$  of the motion pattern  $P_k$ , and  $\alpha_i$  given by the equation

$$\alpha_i = \frac{\|\mu_{\mathbf{v}_i}\| \lambda_{\mathbf{v}_i}}{\sum_{\mathbf{v}_j \in \Omega_k} \|\mu_{\mathbf{v}_j}\| \lambda_{\mathbf{v}_j}}. \quad (29)$$

In this equation, the norm  $\|\mu_{\mathbf{v}_i}\|$  measures the ‘‘amount of motion’’ detected at this point by the filter  $\phi_i$ , while  $\lambda_{\mathbf{v}_i}$  measures the reliability of the estimation  $\mathbf{v}_i$  (26). The denominator in (29) guarantees that  $\sum_{\Omega_k} \alpha_i = 1$ .

If we assume that  $\mathbf{v}_i$  are independent variables,  $\hat{\mathbf{v}}_k$  will be a random variable with a Gaussian distribution with the mean  $\mu_{\hat{\mathbf{v}}_k} = \sum_{\Omega_k} \alpha_i \mu_{\mathbf{v}_i}$  and the covariance  $\Delta_{\hat{\mathbf{v}}_k} = \sum_{\Omega_k} \alpha_i^2 \Delta_{\mathbf{v}_i}$ .

### D. Representation of Multiple Velocities

The motion patterns allow the relevant motions presented in a given sequence to be separated; therefore, in the optical flow estimation problem, they can be used to decide whether there are multiple velocities at the same location or not. Based on this idea, our scheme will obtain the velocities at a given point  $(x, y, t)$  directly from the estimations calculated for each motion pattern as

$$\bar{\mathbf{v}} = \{\hat{\mathbf{v}}_k\}_{k=1\dots K} \quad (30)$$

where  $K$  is the number of motion patterns detected in the sequence and  $\hat{\mathbf{v}}_k$  is the optical flow estimation at point  $(x, y, t)$  of the  $k$ th motion pattern  $S_k$ . It should be noted that due to the use of confidence measures, we will not always have  $K$  estimations at each point.

## IV. RESULTS

In this section, the results obtained with real and synthetic sequences are shown. In both cases, we have tested sequences with simple motions, occlusions, and transparencies.

### A. Synthetic Sequences

To show how our model ‘‘see’’ a motion as a 3-D volume in the spatio-temporal domain, a synthetic case of pure translational motion with a constant speed is displayed in Fig. 5. More specifically, the example shows three bars with velocities of  $(1, 0)$ ,  $(-1, 0)$  and  $(0, -1)$  pixels/frame, respectively. Looking at the 3-D representation of the original sequence, three independent planes can be seen corresponding to the three bars in motion. Our model separates each of these planes into three

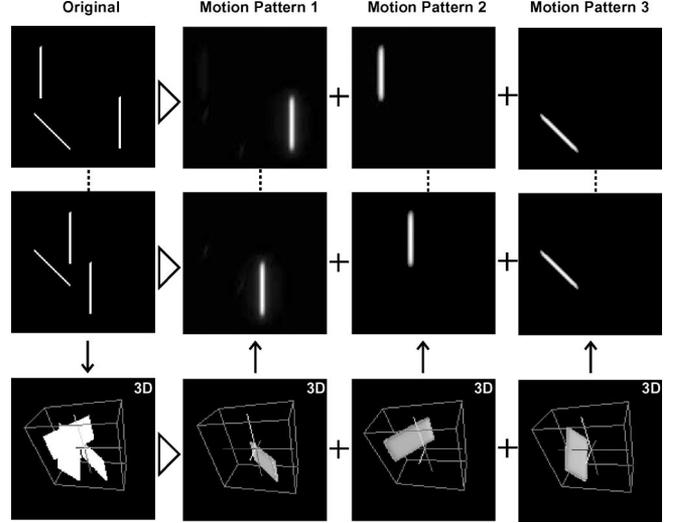


Fig. 5. Synthetic sequence with three translating bars.

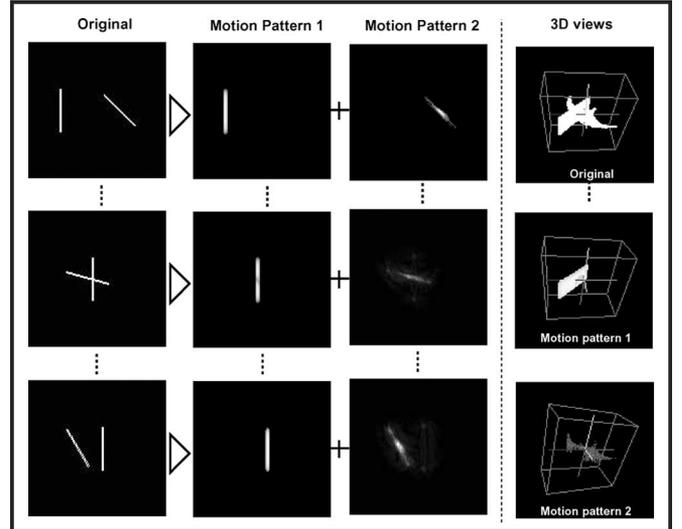


Fig. 6. Synthetic sequence with a translating and a rotating bar.

different spatio-temporal outputs corresponding to the three motions (each spatio-temporal image is obtained by the sum of the filter responses of the cluster). From this 3-D representation, the sequence associated with each motion pattern is extracted.

The example in Fig. 6 shows another more complex case with a bar rotating at an angular velocity of  $5^\circ$ /frame while moving with a velocity of  $(-1, 0)$  pixels/frame. The difficulty of this case lies in the rotation, since a motion of this type spreads across various planes in the Fourier spectrum, and therefore, activates various filters corresponding to different orientations (this is a similar effect to that which occurs with a circumference in the 2-D case). In this example, our model groups the different responses corresponding to the rotation movement and separates this from the translation movement of the first bar. This is precisely one of the more interesting points of the proposed model: The capacity to regroup responses corresponding to a same

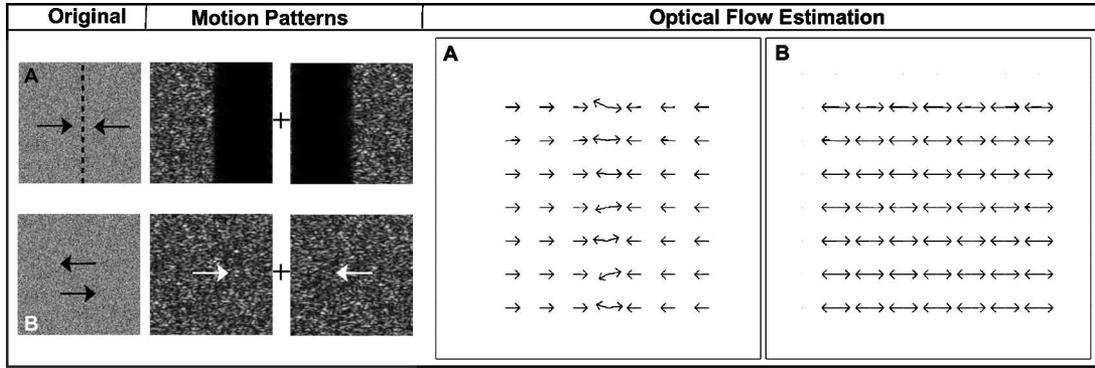


Fig. 7. Optical flow estimation with synthetic sequences.

object based on its continuity in the spatio-temporal domain. Nevertheless, the full rotation of an object is a complex case in which our model can encounter great difficulties since, as we have already mentioned, this type of motion is spread over a large number of filters. In this situation, what may occur is that, in the response of a filter, energy appears associated with two distinct rotations, in which case our model will tend to group the two movements into a single pattern (this will happen, for example, in the case of two bars rotating in opposite directions). This problem is inherent to the interpretation of motion in the frequential space, and to solve this, a postprocessing of the outputs of each filter searching for nonconnected components is necessary.

Focusing our study on the optical flow application, Fig. 7 shows two synthetic sequences, which have been generated with Gaussian noise of mean 1 and variance 0. The first example [Fig. 7(a)] shows a sequence where a background pattern with velocity  $(-1, 0)$  frames/image is occluded by a foreground pattern with velocity  $(1, 0)$ . The second example [Fig. 7(b)] shows two motions with transparency: an opaque background pattern with velocity  $(1, 0)$  and a transparent foreground pattern with velocity  $(-1, 0)$ . In both cases, the figure shows the central frame of the sequence, the motion patterns detected by the model (two in each case), and the optical flow estimated with our technique. In this example, we have used the values  $\kappa_1 = 0$ ,  $\kappa_2 = 1$ , and  $\kappa_p = 1e - 5$  (with  $\Delta_p^{-1} = \kappa_p I$  [45]) in (23) and (24), the spatial and temporal partial derivatives have been calculated using the kernel  $(1/12)(-1, 8, 0, -8, 1)$ , the gradient constraints have been applied in a local neighborhood of size  $5 \times 5$ , and the weight vector has been fixed to  $(0.0625, 0.25, 0.375, 0.25, 0.0625)$  [42]. We should point out that in the first example, our technique obtains two velocities at the occlusion points; in a similar way, in the second example, our methodology is able to estimate two velocities for each point of the frame.

Since we have access to the true motion field of the synthetic sequences, we can measure the performance of the proposed methodology. For this purpose, the following angular measure of error between the correct velocity  $\mathbf{v}_c$  and an estimated  $\mathbf{v}_e$  will be used [42]

$$e(\mathbf{v}_c, \mathbf{v}_e) = \arccos(\mathbf{v}'_c, \mathbf{v}'_e) \quad (31)$$

TABLE II  
MEAN ERROR OBTAINED WITH SEVERAL TECHNIQUES  
APPLIED TO THE SEQUENCES IN FIG. 7

		A (occlusion)	B (transparency)
Proposed technique	MV	0.84°	0.44°
Nestares	MV	3.93°	7.76°
Lucas&Kanade	SV	4.79°	50.89°
Horn&Schunk	SV	2.66°	52.77°
Nagel	SV	8.59°	45.81°
Anandan	SV	10.47°	47.78°
Singh	SV	2.97°	45.27°
Fleet&Jepson	SV	2.23°	43.19°
Uras	SV	3.96°	57.86°
Simoncelli	SV	5.97°	49.38°

MV: Multiple velocities; SV: Single velocity. Density is 100%.

where given a velocity  $\mathbf{v} = (v_x, v_y)$ , we calculate  $\mathbf{v}'$  as  $\mathbf{v}' = (v_x, v_y, 1)/\sqrt{v_x^2 + v_y^2 + 1}$ . Since our examples have points with two velocities, the error will be measured in relation to the nearest correct velocity at this point. If  $\Psi$  represents the set of correct velocities at the point  $(x, y, z)$ , the measure of error will, therefore, be given by the equation

$$E(\mathbf{v}_e) = \min\{e(\mathbf{v}_e, \mathbf{v}_r), \mathbf{v}_r \in \Psi\}. \quad (32)$$

For the two examples in Fig. 7, the mean error obtained with our model is 0.84° for the case with occlusion, and 0.44° for the sequence with transparency. It is important to point out that the proposed methodology generates two velocities at the occlusion points, which differentiates it from other classic techniques such as those studied by Barron *et al.* [42], in which only one velocity is obtained per point. If we apply any of these techniques to the examples in the figure, the estimations in points with double motion will tend to group the two velocities into one, and this generates erroneous estimations. The more points there are with two velocities, the greater this error will be, as shown in Table II. The method proposed by Nestares *et al.* [29] is, however, designed to represent multiple velocities. The fact that this technique also uses an approach based on spatio-temporal filtering makes it particularly interesting in this comparison. As can be seen, this technique reduces the mean error by generating double velocities in the occlusion points.

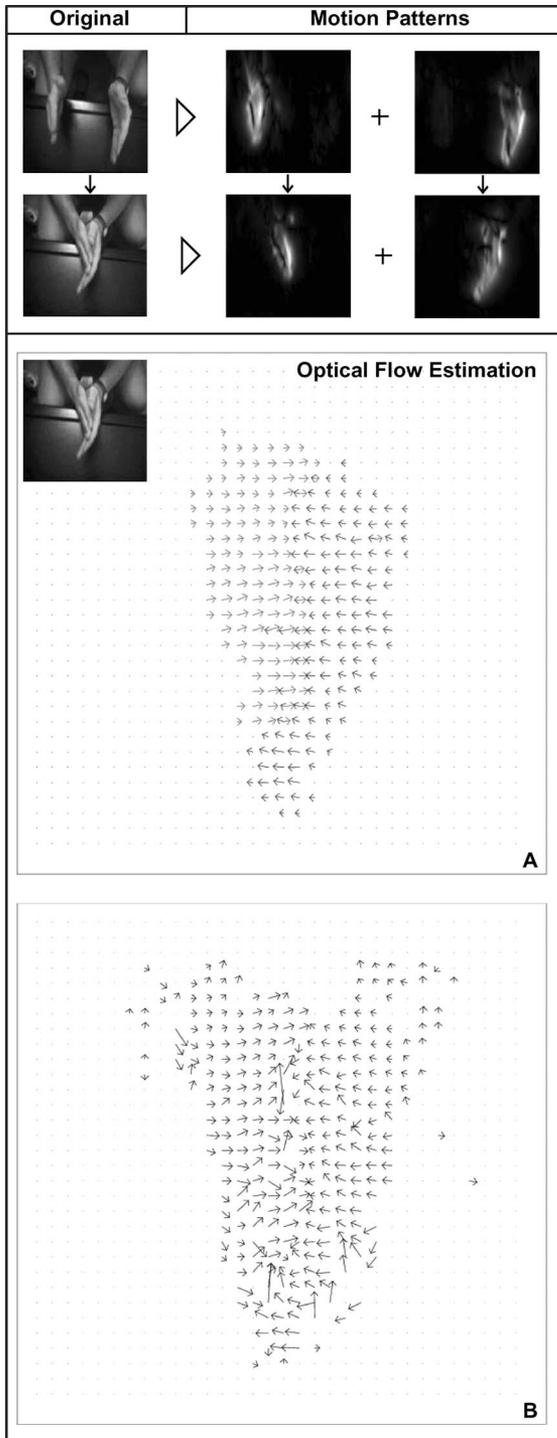


Fig. 8. Real sequence corresponding to a handclap. (a) Optical flow estimated with our technique. (b) Optical flow estimated employing Simoncelli's technique (a single velocity per point).

**B. Real Sequences**

Figs. 8–13 show some examples with real sequences. In this case, we have used the values  $\kappa_1 = 0$ ,  $\kappa_2 = 1$ , and  $\kappa_p = 0.5$  with the same partial derivatives and weight parameters used in the synthetic case. For each example, the figure shows the first and last frames of the original sequence, the motion patterns

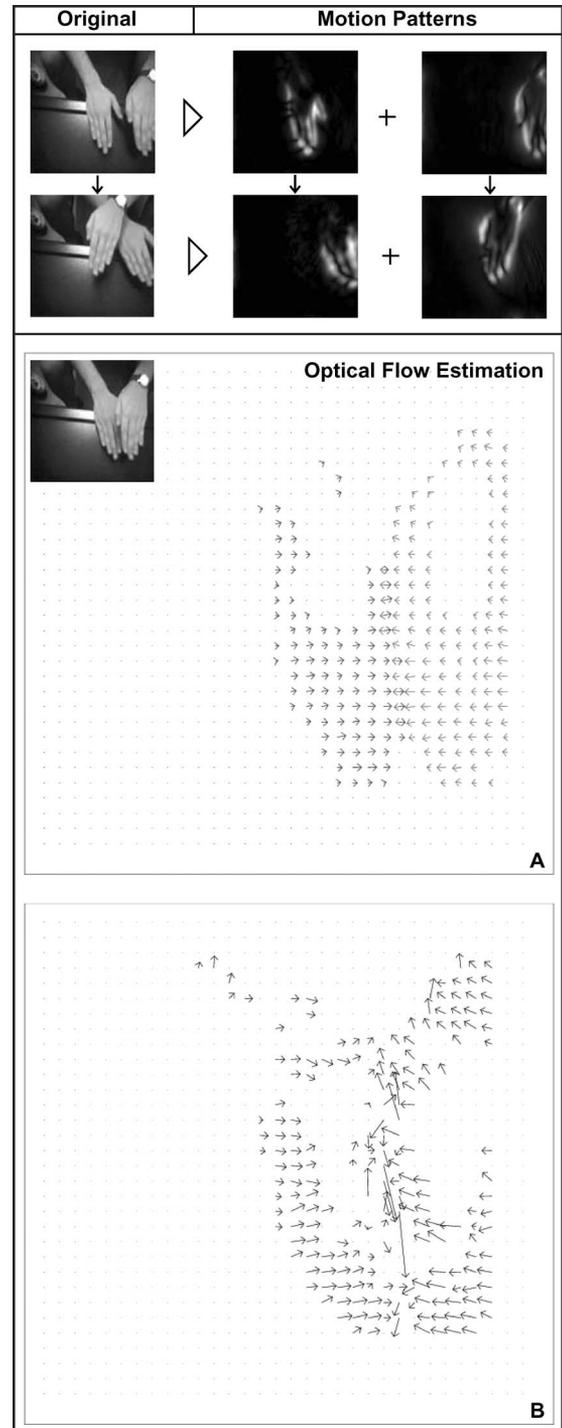


Fig. 9. Real sequence with occlusion where a hand is crossing over another one. (a) Optical flow estimated with our technique. (b) Optical flow estimated using Simoncelli's technique (a single velocity per point).

detected in each case, and the optical flow estimated with our technique. As we do not have the true motion field for real image sequences, we can only show the computed flow field.

1) *Translations With Occlusion and Transparencies:* The first example (Fig. 8) corresponds to a double motion without occlusions where two hands are clapping. The second example (Fig. 9) shows an example of occlusion where a hand is crossing

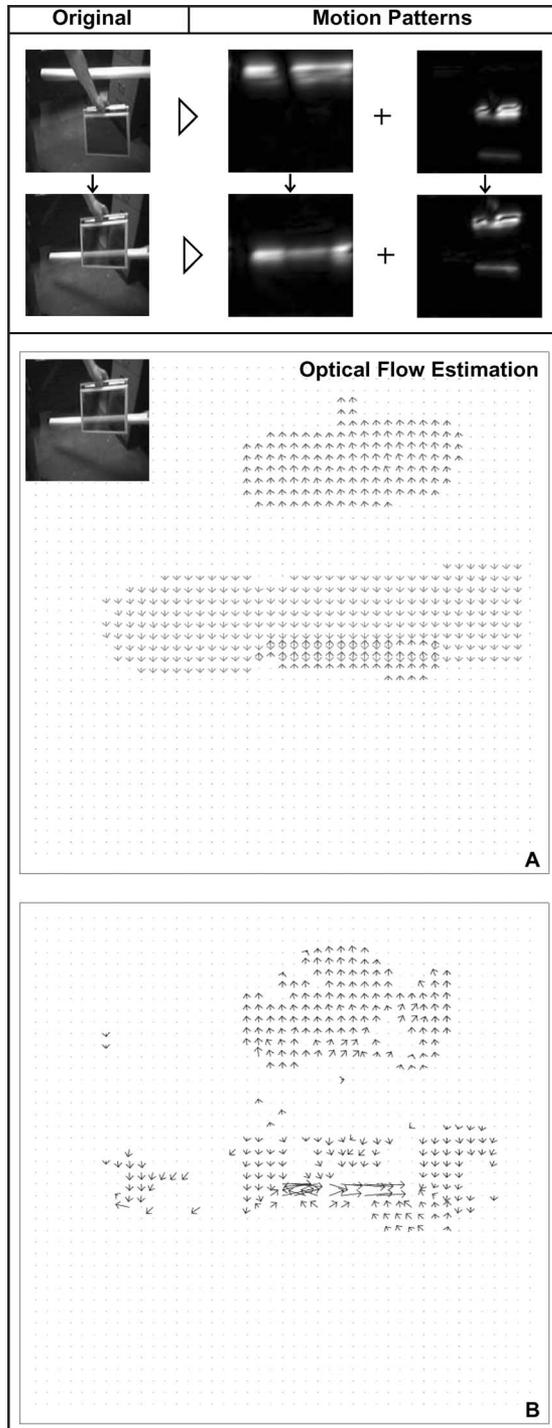


Fig. 10. Real sequence where a bar is occluded by a transparent object. (a) Optical flow estimated with our technique. (b) Optical flow estimated employing Simoncelli's technique (a single velocity per point).

over another one. In this case, where the occlusion is almost complete in some frames, the motion combines translation and rotation without a constant velocity. The third case shows an example of transparency where a bar is occluded by a transparent object (Fig. 10). In all cases, our methodology extracts two motion patterns that separate the two motions presented in the sequence and estimates two velocities in the occlusion

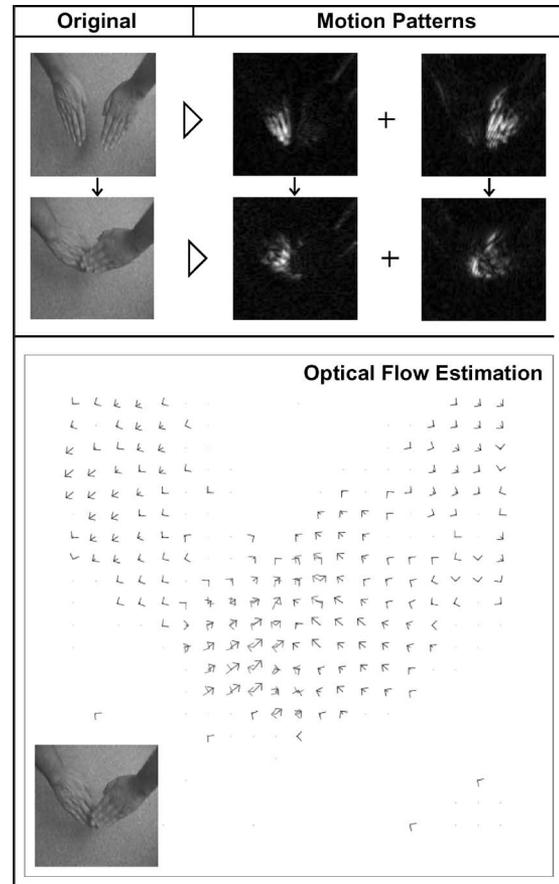


Fig. 11. Optical flow estimated with our technique for a real sequence with two rotating hands.

points. This is shown in Figs. 8–10, together with the velocities estimated employing the Simoncelli's technique [45] as described in Section III-A (which uses a similar approach, but without a multiple velocity representation). We should mention that our model solves the problem of occlusion by means of the spatio-temporal continuity of forms. Furthermore, this approach is capable of detecting motions even when different velocities and spatial orientations are present.

2) *Rotations*: The example in Fig. 11 shows a case with occlusion where two hands start in a vertical position and move closer together, ending with a rotation of approximately  $90^\circ$ . As the result shows, the proposed methodology generates two patterns corresponding to the two hands. It should be highlighted from this example that in the filtering phase, the rotation movements cause various filters to be activated (16, in particular), although the subsequent grouping phase unites the responses corresponding to the same movement. Once again, we should stress the importance of the grouping based on statistical continuity over a set of spatio-temporal responses. As in previous examples, the proposed methodology estimates two velocities at the occlusion points.

3) *Camera Movement*: Fig. 12 shows a sequence taken from a static scene with a translating and rotating camera. The scene has two objects: one close to the camera and the other one far away from it. As the figure shows, the model detects two

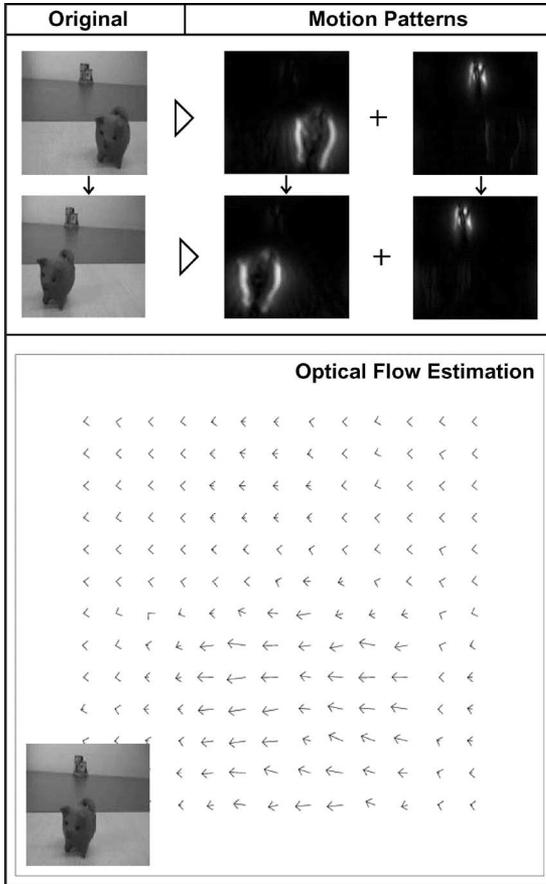


Fig. 12. Optical flow estimated with our technique for a real sequence taken from a static scene with a translating and a rotating camera.

movements associated with each object. We should bear in mind that although two objects are static in the scene, the camera movement makes the objects appear to move in the opposite direction to the camera. The “velocity” of this movement will be determined by its distance from the camera, so that a close object will appear to move with a greater velocity than the one far away. Our model detects these two movements, which, although moving in the same direction, have different velocities (and this will imply that different filters will be activated). Nevertheless, if in an example of these characteristics, the two objects are at the same distance from the camera (so that they appear to move in the same direction and at the same velocity), they will be identified as the same motion pattern. Fig. 12 also shows the optical flow obtained where the estimation for each object can be observed, as well as the movement of the background in the opposite direction to that of the camera.

4) *Articulated Objects*: Finally, Fig. 13 shows an example with an articulated object in which each component of the object moves independently. More specifically, the example shows an object with two components rotating and approaching. As the result shows, our model detects two patterns, one for each component. Due to the rotation movement, each component cause different filters to be activated, although all have a statistical congruence, which enables their subsequent regrouping into a

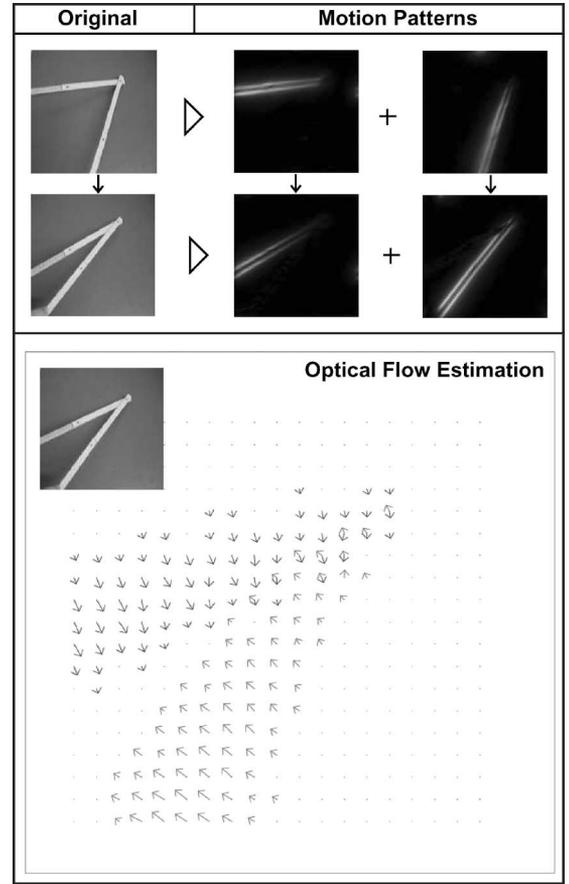


Fig. 13. Optical flow estimated with our technique for a real sequence with an articulated object.

single motion pattern. The optical flow, as we can see, gathers the velocities associated with each component, detecting occlusion points in the union areas of the two components.

## V. CONCLUSION AND FUTURE WORK

This paper presents a new methodology for motion pattern detection applied to optical flow estimation. The proposed technique is able to represent multiple motions on the basis of a new frequency-domain approach, dealing with cases of superposed motions, occlusions, and transparencies. The fact of working in the frequency domain allows objects in motion to be separated efficiently, since these objects will be located in different areas of the Fourier spectrum (this location will depend on the direction and velocity of the motion). To do so, it is necessary to design the spatio-temporal filters, which match these motions: In this approach, the logGabor functions have been adopted as an appropriate method for constructing filters of arbitrary bandwidth, and they have been spread uniformly in the frequency space on the basis of a spherical scheme. The problem of the spatio-temporal filters, although they are a very powerful tool when it comes to separating motions, is that they depend on the spatial orientation of the moving object: components of the same motion with different spatial characteristics are separated into different responses. To tackle this problem, a problem inherent

to all models based on spatio-temporal filtering, the proposed technique identifies a “motion pattern” on the basis of invariance in statistical structure across a range of spatio-temporal frequency bands.

Due to the spatial structure of an object in motion, an object is normally separated into several filter responses. Our new approach, however, enables the information about such motions to be collected, and consequently, the proposed model generates one output for each coherent, independent motion detected in the sequence, thereby avoiding the classic problem associated with a representation based on the spatio-temporal filters. Nevertheless, experiments show that this type of model based on spatio-temporal filtering is particularly sensitive in those cases where two different objects share the same motion (or part of it). In these cases, the filters associated with this type of motion gather both objects in their response, so that a model such as the one proposed would assign them to the same motion pattern. As a future line of work, techniques will be studied to allow two or more objects to be separated which, because of their motion, have been detected by the same filter. To do so, it will be necessary to analyze the responses of the filters in the search for nonconnected components.

Given a set of “motion patterns,” a methodology for obtaining the optical flow corresponding to a spatio-temporal filter response has been presented, using confidence measures to ensure only reliable estimations. One of the main features of the proposal is the possibility of representing more than one velocity at a point. This is extremely important in situations where there are occlusions or transparencies, since in these cases, the techniques that do not consider the presence of multiple motions will generate erroneous estimations. To carry out the estimation, a probabilistic combination of velocities corresponding to the clustering of the set of filters in a given motion pattern has been proposed. The use of “motion patterns” allows multiple motions to be represented, while the combination of estimations from different filters and the confidence measures have reduced the initial aperture problem.

The technique has been illustrated on several data sets. Real and synthetic sequences combining occlusions and transparencies have been tested. In all the cases, the final results enlighten the consistency of the proposed algorithm.

#### APPENDIX

In this Appendix, an algorithm to uniformly distribute the filter centers over the frequential space is presented. As pointed out in Section II-A1, our approach uses a spherical scheme with spatio-temporal multiresolution that distributes the center of the filters on the surface of spheres with different radii (one for each considered resolution level). Algorithm 1 distributes points on the surface of a sphere of radius 1 taking as parameter the number of orientations  $N$  in the static plane (this allows to particularize the algorithm for the case of static images). This way, we shall obtain the different spheric orientations of the bank used for each resolution level. Due to conjugate symmetry in the Fourier domain, the filter is only designed on half the 3-D frequential space.

---

**Algorithm 1** Algorithm for points distribution over an sphere of radius 1.

---

1) Input:

$N$ : Number of orientations in the static plane

2) Initialization

$$X = \emptyset$$

$$s = (\pi/N)$$

3) For  $i = 0, \dots, (N/2)$

$$\alpha = i \cdot (\pi/N)$$

$$r = \cos(\alpha)$$

$$\beta = (s/r)$$

$$M = \lceil (2\pi/\beta) \rceil$$

For  $j = 0, \dots, M - 1$

$$x = r \cdot \cos(j \cdot (2\pi/M))$$

$$y = r \cdot \sin(j \cdot (2\pi/M))$$

$$z = \sin(\alpha)$$

If  $i \neq 0$  or  $j < (M/2) - 1$

$$X = X \cup (x, y, z)$$

4) Output:  $X$

---

#### REFERENCES

- [1] F. Moscheni, S. Bhattacharjee, and M. Kunt, “Spatiotemporal segmentation based on region merging,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 9, pp. 897–915, Sep. 1998.
- [2] A. R. Mansouri and J. Konrad, “Multiple motion segmentation with level sets,” *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 201–220, Feb. 2003.
- [3] D. Tweed and A. Calway, “Integrated segmentation and depth ordering of motion layers in image sequences,” *Image Vision Comput.*, vol. 20, pp. 709–724, 2002.
- [4] P. H. S. Torr, “Geometric motion segmentation and model selection,” *Phil. Trans. R. Soc. A*, vol. 356, pp. 1321–1340, 1998.
- [5] R. Vidal and R. Hartley, “Motion segmentation with missing data using power factorization and GPCA,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2004, vol. 2, pp. 310–316.
- [6] K. Kanatani, “Motion segmentation by subspace separation and model selection,” in *Proc. IEEE Int. Conf. Comput. Vision*, 2001, vol. 2, pp. 586–591.
- [7] M. Han and T. Kanade, “Multiple motion scene reconstruction with uncalibrated cameras,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 884–894, Jul. 2003.
- [8] C. Y. Vincent and T. Tjahjedi, “Obstacle detection by direct estimation of multiple motion and scene structure from a moving stereo rig,” in *Proc. IEEE Int. Conf. Syst. Man, Cybern.*, 2003, vol. 3, pp. 2326–2331.
- [9] R. Vidal and S. Sastry, “Optimal segmentation of dynamic scenes from two perspective views,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2003, vol. 1, pp. 281–286.
- [10] R. Hartley and R. Vidal, “The multibody trifocal tensor: Motion segmentation from 3 perspective views,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vision. Pattern Recognit.*, 2004, vol. 1, pp. 769–775.
- [11] W.-S. Tong, C.-K. Tang, and G. Medioni, “Simultaneous two-view epipolar geometry estimation and motion segmentation by 4D tensor voting,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1167–1184, Sep. 2004.
- [12] K. Korimilli and S. Sarkar, “Motion segmentation based on perceptual organization of spatio-temporal volumes,” in *Proc. 15th Int. Conf. Pattern Recognit.*, 2000, vol. 3, pp. 844–849.
- [13] E. H. Adelson and J. R. Bergen, “Spatiotemporal energy models for the perception of motion,” *J. Opt. Soc. Amer. A.*, vol. 2, no. 2, pp. 284–299, Feb. 1985.
- [14] H. Liu, M. Hong, M. Herman, and A. Chellappa, “A general network model and spatio-temporal filters for computing optical flow,” *Int. J. Comput. Vision*, vol. 22, no. 2, pp. 141–172, 1997.

- [15] L. Wiskott, "Segmentation from motion: Combining Gabor and Mallat wavelets to overcome the aperture and correspondence problem," *Pattern Recognit.*, vol. 32, pp. 1751–1766, 2007.
- [16] O. Nestares, C. Miravet, J. Santamaria, and R. Navarro, "Automatic segmentation of low visibility moving objects through energy analysis of the local 3D spectrum," in *Proc. SPIE*, 1999, vol. 3642, pp. 13–22.
- [17] F. Dekeyser, P. Boutheymy, and P. Perez, "Spatio-temporal wiener filtering of image sequences using a parametric motion model," in *Proc. Int. Conf. Image Process.*, 2000, vol. 1, pp. 208–211.
- [18] A. B. Watson and A. J. Ahumada. (1983). A look at motion in the frequency domain. NASA Goddard Space Flight Center, Greenbelt, MD, Tech. Rep., NASA Tech. Memo. TM-84352. [Online]. Available: <http://vision.arc.nasa.gov/publications/LookAtMotion.pdf>
- [19] S. S. Beauchemin and J. L. Barron, "On the Fourier properties of discontinuous motion," *J. Math. Imag. Vision*, vol. 13, pp. 155–172, 2000.
- [20] —, "The frequency structure of 1D occluding image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 2, pp. 200–206, Feb. 2000.
- [21] W. Yu, S. Beauchemin, and K. Daniilidis, "Oriented structure of the occlusion distortion: Is it reliable?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1286–1290, Sep. 2002.
- [22] J. Chamorro-Martínez, J. Fdez-Valdivia, and J. Martínez-Baena, "A spatio-temporal filtering approach to motion segmentation," in *Pattern Recognition and Image Analysis*, F. J. Perales, A. J. C. Campilho, N. Perez de la Blanca, and A. Sanfeliu, Eds. Berlin, Germany: Springer-Verlag, Jun. 2003, pp. 193–203.
- [23] H.-Y. Wang and K.-K. Ma, "Accurate optical flow estimation using adaptive scale-space and 3D structure tensor," in *Proc. IEEE ICIP*, 2002, vol. 2, pp. 301–304.
- [24] M. Pingault, E. Bruno, and D. Pellerin, "A robust multiscale B-spline function decomposition for estimating motion transparency," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1416–1426, Nov. 2003.
- [25] B. G. Schunck, "Image flow segmentation and estimation by constraint line clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 10, pp. 1010–1027, Oct. 1989.
- [26] H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 5, pp. 565–593, Sep. 1986.
- [27] M. J. Black and P. Anandan, "The robust estimation of multiple motion: Parametric and piecewise smooth flow fields," *Comput. Vision Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.
- [28] D. J. Heeger, "Model for the extraction of image flow," *J. Opt. Soc. Amer. A.*, vol. 4, no. 8, pp. 1455–1571, 1987.
- [29] O. Nestares and R. Navarro, "Probabilistic estimation of optical flow in multiple band-pass directional channels," *Image Vision Comput.*, vol. 19, no. 6, pp. 339–351, 2001.
- [30] E. Memin and P. Perez, "Dense estimation and object-based segmentation of the optical flow with robust techniques," *IEEE Trans. Image Process.*, vol. 7, no. 5, pp. 703–719, May 1998.
- [31] T. Brox, A. Bruhn, N. Papenberger, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. ECCV 2004*, Lecture Notes in Computer Science, vol. 3024, T. Pajdla and J. Matas, Eds. Berlin, Germany: Springer-Verlag, 2004, pp. 25–36.
- [32] J. Chamorro-Martínez and J. Fdez-Valdivia, "Optical flow estimation based on the extraction of motion patterns," in *Proc IEEE ICIP*, Barcelona, Spain, Sep. 2003, vol. 1, pp. 925–928.
- [33] J. G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 36, no. 7, pp. 1169–1179, Jul. 1988.
- [34] M. Hawken and A. Parker, "Spatial properties of neurons in the monkey striate cortex," in *Proc. R. Soc. Lond., B*, 1987, vol. 231, pp. 251–288.
- [35] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A.*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [36] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice Hall, 1988.
- [37] M. C. Morrone and R. A. Owens, "Feature detection from local energy," *Pattern Recognit. Lett.*, vol. 6, pp. 303–313, 1987.
- [38] M. C. Morrone and D. C. Burr, "Feature detection in human vision: A phase-dependent energy model," in *Proc. R. Soc. Lond., B*, 1988, vol. 235, pp. 221–245.
- [39] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
- [40] R. Rodríguez-Sánchez, J. A. García, J. Fdez-Valdivia, and X. R. Fdez-Vidal, "The RGFF representation model: A system for the automatically learned partitioning of 'visual patterns' in digital images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 1044–1072, Oct. 1999.
- [41] N. Graham, *Visual Pattern Analyzers*. London, U.K.: Oxford Univ. Press, 1989.
- [42] J. L. Barron, D. J. Fleet, S. Beauchemin, and S., "Performance of optical flow techniques," *Int. J. Comput. Vision*, vol. 12, no. 1, pp. 43–77, 1994.
- [43] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, pp. 185–203, 1981.
- [44] B. Lucas and T. Kanade, "An interactive image registration technique with an application to stereo vision," in *DARPA Proc. Image Understanding Workshop*, 1981, pp. 121–130.
- [45] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger, "Probability distributions of optical flow," in *IEEE Proc. CVPR'91*, 1991, pp. 310–315.



**Jesús Chamorro-Martínez** was born in Granada, Spain. He received the M.S. and Ph.D. degrees from the University of Granada, Granada, in 1995 and 2001, respectively, both in computer science.

Since 1996, he has been with Computer Science Department, University of Granada, where he is now an Associate Professor. His research interests include motion analysis and optical flow estimation, information retrieval from image databases, soft computing image processing, and biomedical image analysis.



**Joaquín Fernández-Valdivia** was born in Granada, Spain. He received the M.S. and Ph.D. degrees from the University of Granada, Granada, in 1986 and 1991, respectively, both in mathematics.

Since 1988, he has been with the Computer Science Department (DECSAI) at University of Granada, where he is now an Associate Professor. His research work is summarized in over 100 papers in scientific journals and conference proceedings and three books in the field of computer vision.

His current interest includes computer vision, image representation, feature detection, visual target distinctness, image coding, and biomedical applications.

Dr. Fernández-Valdivia is a Member of the IAPR Association and the IEEE Computer Society.