Dealing with Multiple Motions in Optical Flow Estimation

Jesús Chamorro-Martínez, Javier Martínez-Baena,
Elena Galán-Perales, and Beén Prados-Suárez*

Department of Computer Science and Artificial Intelligence
University of Granada, Spain
{jesus,jbaena,elena,belenps}@decsai.ugr.es

Abstract. In this paper, a new approach to optical flow estimation in presence of multiple motions is presented. Firstly, motions are segmented on the basis of a frequency-based approach that groups spatio-temporal filter responses with continuity in its motion (each group will define a *motion pattern*). Then, 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", the velocities at a given point are then combined using a probabilistic approach. The use of "motion patterns" allows multiple velocities to be represented, while the combination of estimations from different filters helps reduce the aperture problem.

Keywords: Optical flow, multiple motions, spatio-temporal models

1 Introduction

Optical flow estimation, viewed as an approximation to image motion, is a very useful task in video processing [1]. In this framework, an open problem is how to deal with the presence of multiple motions at the same location [2]. 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 which do not consider the presence of multiple motions will generate erroneous estimations which will combine into a single vector the different velocities present at one point. 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 [3], the models based on line processes [4], the parametric models [5] or the frequency-based techniques (which use spatio-temporal filters to separate the motions [6, 7]). 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 work has been supported by the MCYT (Spain) under grant TIC2003-01504.

J.S. Marques et al. (Eds.): IbPRIA 2005, LNCS 3522, pp. 52–59, 2005.

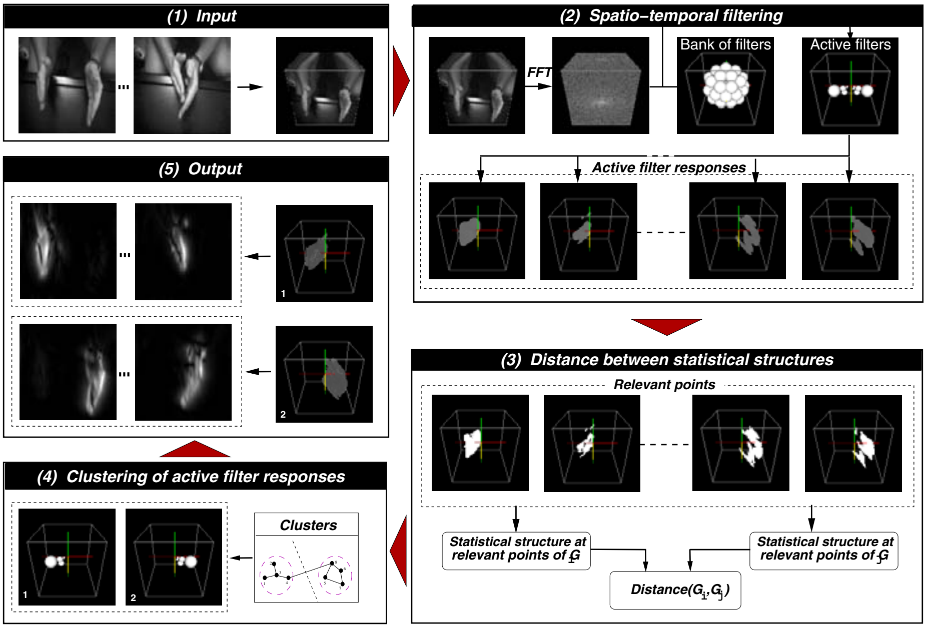

Fig. 1. A general diagram describing the motion segmentation model.

In order to confront this problem, in this paper we develop a methodology for optical flow estimation that is able to represent multiple velocities at the same point. To detect points with multiple motions, the model introduced in [8] is used. This model is a frequency-based approach that groups spatio-temporal filter responses with continuity in its motion (each group will define a *motion pattern*). Given a motion pattern (a group of filters), the proposed technique apply the gradient constraints to the output of each filter in order to obtain multiple estimates of the velocity at the same location. Then velocities at each point are combined using probability rules.

The rest of the paper is organized as follows. Section 2 introduces the spatio-temporal filtering approach to motion segmentation and Section 3 shows its application to optical flow estimation with of multiple motions. Results with real and synthetic sequences are shown in Section 4 and, finally, the main conclusions are summarized in Section 5.

2 Motion Patterns

To detect multiple motions at the same location, the frequency-domain approach introduced in [8] is used. This methodology is based on three main stages: 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 Figure 1 (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 three-dimensional pattern. Its Fourier transform is then calculated in order to perform the analysis in the frequential domain. Given a bank of spatio-temporal logGabor filters, 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 (only one subset of filters is used).

In the second stage, the distances between active filters are obtained. These distances are computed over relevant points which are calculated as local energy peaks on the filter response.

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 Figure 1, two collections of filters corresponding to the two hand motions are shown. For more details about these three stages, see [8].

3 Optical Flow Estimation

In this section, the frequency-based model described in Section 2 will be used to optical flow estimation in presence of multiple motions

3.1 Differential Formulation

Within the gradient-based approaches, based on the well known differential brightness constancy constraint equation, a probabilistic framework to optical flow estimation was proposed by Simoncelli et al. [9]. In this approach, two independent additive Gaussian noise terms $\mathbf{n_1}$ and n_2 are introduced in the constancy constraint equation [9], and the velocity at a given point is defined as a Gaussian random variable with mean and covariance:

$$\mu_{\mathbf{v}} = -\Delta_{\mathbf{v}} \cdot \sum_r \frac{w_r \mathbf{d}_r}{\kappa_1 \left\|\mathbf{f_e}(x_r, y_r, t)\right\|^2 + \kappa_2} \tag{1}$$

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

with $\mathbf{f_e} = (f_x, f_y)$ and f_t being, respectively, the spatial and temporal derivatives, where w_r is a weighting function that gives more influence to elements at the center of the neighborhood, with the points in the neighborhood indexed by r, Δ_p the covariance of the prior distribution $P(\mathbf{v})$, $\mathbf{M}_r$ and $\mathbf{d}_r$ defined as

$$\mathbf{M_r} = \begin{bmatrix} f_x^2(r) & f_x(r) f_y(r) \\ f_y(r) f_x(r) & f_y^2(r) \end{bmatrix} \qquad \mathbf{b_r} = \begin{bmatrix} f_x(r) f_t(r) \\ f_y(r) f_t(r) \end{bmatrix} \tag{3}$$

and $\kappa_1$and κ_2 two parameters associated to $\mathbf{n_1}$ and n_2 respectively (see [9] for more details)

3.2 Estimation for a Spatio-temporal Filter Response

In order to estimate the velocity $\mathbf{v}_i$ at a given point (x, y, t) of the i-th filter ϕ_i, the probabilistic approach described in Section 3.1 is used. Using the odd response of the filter ϕ_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 $\Delta_{\mathbf{v}_i}$:

$$\mathbf{v}_i \sim N(\mu_{\mathbf{v}_i}, \Delta_{\mathbf{v}_i}) \quad i = 1, \ldots N \tag{4}$$

where $\mu_{\mathbf{v}_i}$ and $\Delta_{\mathbf{v}_i}$ are calculated using Equations (1) and (2). Therefore, given a point (x, y, t), we shall have a vector of estimations $[\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_N]$, with N being the number of active filters

Confidence Measure It is well known that the covariance matrix $\Delta_{\mathbf{v}_i}$ can be used to define a confidence measure of the estimation $\mathbf{v}_i$ [9]. In this paper, we shall use the smallest eigenvalue of $\Delta_{\mathbf{v}_i}^{-1}$ as the confidence measure of $\mathbf{v}_i$ [10] and this will be denoted $\lambda_{\mathbf{v}_i}$:

$$\lambda_{\mathbf{v}_i} = min\left\{\lambda_1^i, \lambda_2^i\right\} \tag{5}$$

where λ_1^i and λ_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 ϕ_i will be accepted if $\lambda_{\mathbf{v}_i} \geq Threshold_{\phi_i}$, where $Threshold_{\phi_i}$ is a confidence threshold associated to the filter ϕ_i. Under the assumption that every relevant point of the filter will generate a reliable estimation, the following approximation is proposed to calculate $Threshold_{\phi_i}$:

$$Threshold_{\phi_i} = min\left\{\lambda_{\mathbf{v}_i}(x, y, t) \ / \ (x, y, t) \in P(\phi_i)\right\} \tag{6}$$

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

3.3 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 $\left\{\phi_i^k\right\}^{i=1,\ldots L_k}$ be the set of L_k grouped filters in S_k. Let Ω_k be the set of estimations $\mathbf{v}_i \sim N(\mu_{\mathbf{v}_i}, \Delta_{\mathbf{v}_i})$ obtained from $\left\{\phi_i^k\right\}^{i=1,\ldots L_k}$ which are above the confidence threshold. The integration will be performed on the basis of a linear combination

$$\widehat{\mathbf{v}}_k = \sum_{\mathbf{v}_i \in \Omega_k} \alpha_i \mathbf{v}_i \tag{7}$$

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

$$\alpha_i = \frac{\left\|\mu_{\mathbf{v}_i}\right\| \lambda_{\mathbf{v}_i}}{\sum_{\mathbf{v}_j \in \Omega_k} \left\|\mu_{\mathbf{v}_j}\right\| \lambda_{\mathbf{v}_j}} \tag{8}$$

In this equation, the norm $\|\mu_{\mathbf{v}_i}\|$ measures the "amount of motion" detected at this point by the filter ϕ_i, while $\lambda_{\mathbf{v}_i}$ measures the reliability of the estimation $\mathbf{v}_i$ (Equation (5)). The denominator in (8) guarantees that $\sum_{\Omega_k} \alpha_i = 1$.

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

3.4 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:

$$\overline{\mathbf{v}} = \{\widehat{\mathbf{v}}_k\}_{k=1\ldots K} \tag{9}$$

where K is the number of motion patterns detected in the sequence, and $\widehat{\mathbf{v}}_k$ 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.

4 Results

4.1 Synthetic Sequences

Figure 2 shows two synthetic sequences which have been generated with Gaussian noise of mean 1 and variance 0. The first example (Figure 2(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 (Figure 2(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 $\Delta_p^{-1} = \kappa_p I$ [9]) in Equations (1) and (2) as it is proposed in [9], the spatial and temporal partial derivatives have been calculated using the kernel $\frac{1}{12}(-1, 8, 0, -8, 1)$, the gradient constraints have been applied in a local neighborhood of size 5×5, and the weight vector has been fixed to $(0.0625, 0.25, 0.375, 0.25, 0.0625)$ [10].

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 angular error [10]. Table 1 shows a comparision between our methodology and other classic techniques such as those studied by Barron et al. [10].

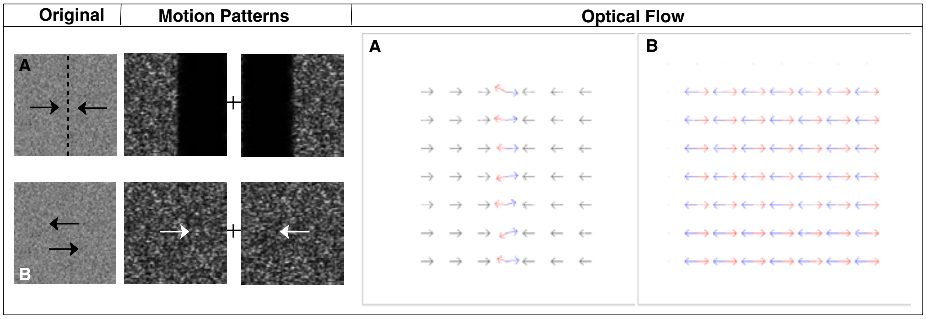

Fig. 2. Results with synthetic sequences.

Table 1. Mean error obtained with several techniques applied to the sequences in Figure 2. MV: Multiple velocities. SV: Single velocity. Density is 100%.

		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°
Uras	SV	3.96°	57.86°
Simoncelli	SV	5.97°	49.38°

4.2 Real Sequences

Figure 3 shows some examples with real sequences. In this case, we have used the values $\kappa_1 = 0$, $\kappa_2 = 1$ and $\kappa_p = 0.5$ (as it is proposed in [9]) with the same partial derivatives and weight parameters used in the synthetic case. For each example, the figure shows the first and last frame of the original sequence, the motion patterns detected in each case, the optical flow estimated with our technique and the optical flow estimated employing the Simoncelli's technique [9] as described in section 3.1 (which uses a similar approach, but without a multiple velocity representation). As we do not have the true motion field for real image sequences, we can only show the computed flow field.

The first example (Figure 3(A)) shows a case of occlusion where a hand is crossing over another one. The second case shows an example of transparency where a bar is occluded by a transparent object (Figure 3(B)). Finally, Figure 3(C) shows an example with an articulated object with two components rotating and approaching independently. In all the cases, our methodology extracts two motion patterns and estimates two velocities in the occlusion points.

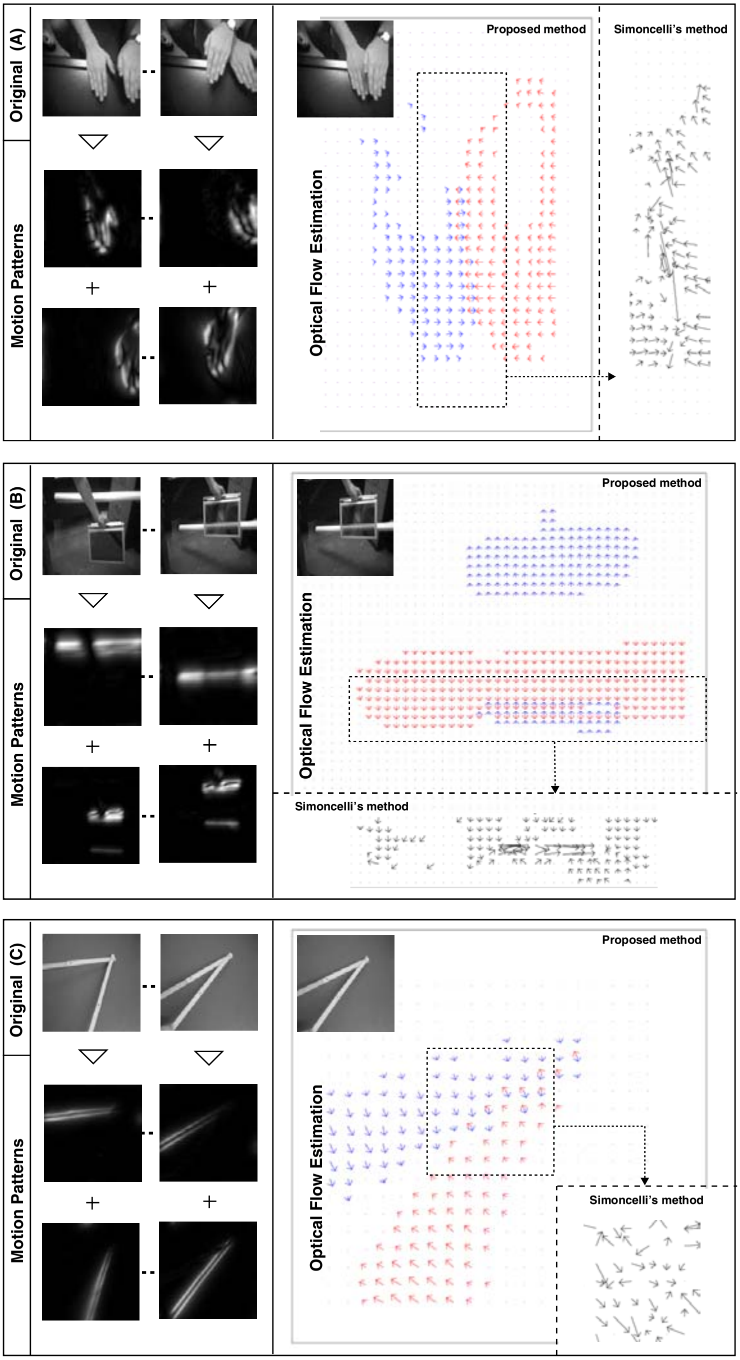

Fig. 3. Results with real sequences.

5 Conclusions

In this paper, a new methodology for optical flow estimation has been presented. The proposed technique is able to represent multiple velocities on the basis of a new frequency-domain approach capable to detect "motion patterns" (that is, a clustering of spatio-temporal filter responses with continuity in its motion). A methodology to obtain the optical flow corresponding to a spatio-temporal filter response has been proposed, using confidence measures to ensure only reliable estimations. A probabilistic combination of velocities corresponding to the set of filters clustering in a given motion pattern has been proposed. 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 with occlusions or transparencies.

References

1. B.K.P. Horn and B.G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.
2. M. Pingault, E. Bruno, and D. Pellerin, "A robust multiscale b-spline function decomposition for estimating motion transparency," *IEEE Transactions on Image Processing,*, vol. 12, pp. 1416–1426, 2003.
3. B.G. Schunck, "Image flow segmentation and estimation by constraint line clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 10, pp. 1010–1027, 1989.
4. H . Nagel and W. Enkelmann, "An investigation of smootness constraints for the estimation of displacement vector fields from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 565–593, 1986.
5. M.J. Black and P. Anandan, "The robust estimation of multiple motion: parametric and piecewisesmooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.
6. D.J Heeger, "Model for the extraction of image flow," *Journal of the Optical Society of America A.*, vol. 4, no. 8, pp. 1455–1571, 1987.
7. O . Nestares and R . Navarro, "Probabilistic estimation of optical flow in multiple band-pass directional channels," *Image and Vision Computing*, vol. 19, no. 6, pp. 339–351, 2001.
8. J. Chamorro-Martinez, J. Fdez-Valdivia, and J. Martinez-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., Lecture Notes in Computer Science LNCS 2652, pp. 193–203. Springer, June 2003.
9. E.P. Simoncelli, E.H. Adelson, and D.J. Heeger, "Probability distributions of optical flow," *IEEE Proceedings of CVPR'91*, pp. 310–315, 1991.
10. J.L. Barron, D.J. Fleet, and S. Beauchemin, S, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.