

OPTICAL FLOW ESTIMATION BASED ON THE EXTRACTION OF MOTION PATTERNS

J. Chamorro-Martínez and J. Fdez-Valdivia

Department of Computer Science and Artificial Intelligence
University of Granada, Spain

e-mail : {jesus,jfv}@decsai.ugr.es

ABSTRACT

In this paper, a new methodology for optical flow estimation that is able to represent multiple motions is presented. To separate motions at the same location, a new frequency-domain approach is used. This model, based on a band-pass filtering with a set of log-Gabor spatio-temporal filters, groups together filter responses with continuity in its motion (each group will define a *motion pattern*). Given a motion pattern, the gradient constraints is applied to the output of each filter in order to obtain multiple estimates of the velocity at the same location. Then, the velocities at each point of the motion pattern are combined using probabilistic rules. The use of “motion patterns” allows to represent multiple motions, while the combination of estimates from different filters helps to reduce the initial aperture problem. This technique is illustrated on real and simulated data sets, including sequences with occlusion and transparencies.

Keywords: Optical flow, multiple motions, spatio-temporal models, motion pattern.

1. INTRODUCTION

The estimation of the optical flow, an approximation to image motion, is an important problem in processing sequences of images. Many techniques have been proposed in the literature; for example, differential methods, which rely on the assumption that the intensity levels in the image remain constant over the time [1], matching techniques, which operate by matching small regions of intensity, and frequency-based methods, which are based on spatio-temporally oriented filters [2, 3].

An important point to take into account in the optical flow estimation is the presence of multiple motions at the same location. Occlusions and transparencies are two common examples of this phenomena, where traditional methods fail. These problems are currently being addressed by the research community; see, for example, the strategies based on the use of mixed velocity distributions (usually two) at each point [4], the line processes based models [5] or the parametric models [6]. Another important group of techniques are based on spatio-temporal filters [2]. These approaches are derived by considering the motion problem in the Fourier domain: the spectrum of a spatio-temporal translation lies in a plane whose orientation depends on the direction and velocity of the motion. Although the filters are a powerful tool to separate the motions presented in a sequence [7], the main problem of these schemes is that orientation selectivity tends to increase the aperture problem. Moreover, components of the same motion with different spatial characteristics are separated in different filters responses.

In this paper, we develop a methodology for optical flow estimation that is able to represent multiple motions. To separate mo-

tions at the same location, the model introduced in [7] is used. This model is a frequency-based approach that groups filter responses with continuity in its motion (each group will define a *motion pattern*). This grouping allows to eliminate the problems describe above relating to the spatial dependency. Given a motion pattern (a group of filters), we first apply the gradient constraints to the output of each filter in order to obtain multiple estimates of the velocity at the same location. Then we combine the velocities at each point of the motion patterns using probability rules. The use of “motion patterns” allows to represent multiple motions, while the combination of estimates from different filters helps to reduce the initial aperture problem.

2. MOTION PATTERNS

To separate motions at the same location, the frequency-domain approach introduced in [7] is used. The figure 1 shows a general diagram describing how the data flows through the model. This diagram illustrates the analysis on a given sequence showing a clap of hands. The endpoint of analyzing this sequence is to separate the two hand motions. In a first stage, a three-dimensional representation is performed from the original sequence and then its Fourier transform is calculated. Given a bank of spatio-temporal logGabor filters, a subset of them is selected in order to extract significant spectral information. These selected filters are applied over the original spatio-temporal image in order to obtain a set of active responses (note that we only use a subset of filters).

In the second stage, for each pair of active filters, their responses are compared based on the distance between their statistical structure, computed over those points which form relevant points of the filters (we calculate these points as local energy peaks on the filter response) As a result, a set of distances between active filters is obtained [8].

In a third stage, a clustering on the basis of the distance between the active filter responses is performed to highlight invariance of responses. Each of the cluster obtained in this stage defines a motion pattern. In figure 1, two collections of filters have been obtained for the input sequence.

3. OPTICAL FLOW ESTIMATION

In this section, the frequency-based model introduced in section 2 will be used to obtain an optical flow estimation able to represent multiple motions. In section 3.1 a technique based on the classic gradient constraint is proposed to obtain the optical flow estimation corresponding to each filter response. In section 3.2, a methodology to integrate the estimations corresponding to the

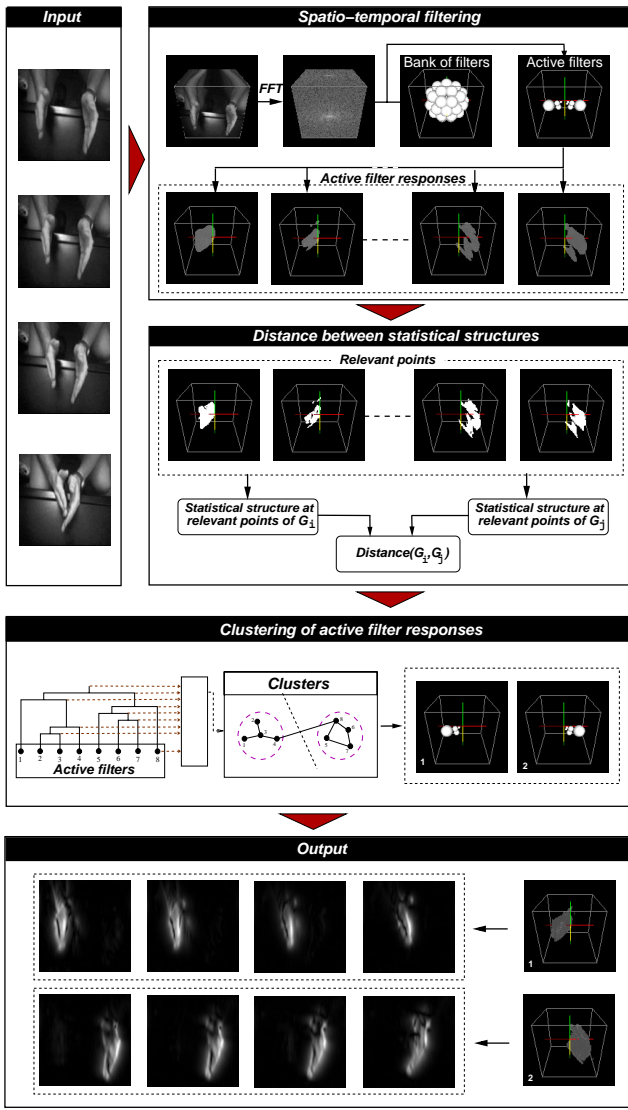


Fig. 1. A general diagram of the frequency-based model.

grouped filters in each motion pattern is described. Finally, in section 3.3 the proposed multiple motion representation is defined.

3.1. Estimation of a spatio-temporal filter response

To estimate the velocity \mathbf{v}_i at a given point (x, y, t) of the i -th filter ϕ_i , an analysis similar to the probabilistic approach proposed in [9] is used. Thus, and using the odd response of the filter, the velocity \mathbf{v}_i at a given point (x, y, t) is defined on the basis of a Gaussian random variable \mathbf{v}_i with mean $\mu_{\mathbf{v}_i}$ and covariance $\Delta_{\mathbf{v}_i}$:

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

where $\mu_{\mathbf{v}_i}$ and $\Delta_{\mathbf{v}_i}$ are calculated as

$$\mu_{\mathbf{v}_i} = -\Delta_{\mathbf{v}_i} \cdot \sum_r \frac{w_r \mathbf{d}_r}{\gamma_1 \|\mathbf{f}_e^r\|^2 + \gamma_2} \quad (2)$$

$$\Delta_{\mathbf{v}_i} = \left[\sum_r \frac{w_r \mathbf{M}_r}{\gamma_1 \|\mathbf{f}_e^r\|^2 + \gamma_2} + \Delta_p^{-1} \right]^{-1} \quad (3)$$

with R being the number of points in the neighborhood of (x, y, t) , w_r being a weight vector that gives more influence to elements at the center of the neighborhood than to those at the periphery, Δ_p the covariance of the prior distribution of \mathbf{v}_i [9], and \mathbf{M}_r and \mathbf{d}_r defined as

$$\mathbf{M}_r = \begin{bmatrix} f_x^2 & f_x f_y \\ f_y f_x & f_y^2 \end{bmatrix} \quad \mathbf{d}_r = \begin{bmatrix} f_x f_t \\ f_y f_t \end{bmatrix} \quad (4)$$

with $\mathbf{f}_e = (f_x, f_y)$ and f_t being the spatial and temporal partial derivatives [9] (for the sake of simplicity, we have removed the spatio-temporal parameters (x, y, t) in the notation). Thus, given a point (x, y, t) , we will have an estimation for each active filter.

3.1.1. Confidence measure

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 will use the smallest eigenvalue of $\Delta_{\mathbf{v}_i}^{-1}$ as confidence measure of \mathbf{v}_i [10] and it will be noted as $\lambda_{\mathbf{v}_i}$:

$$\lambda_{\mathbf{v}_i} = \min \{ \lambda_1^i, \lambda_2^i \} \quad (5)$$

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 left out 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 $\lambda_{\mathbf{v}_i} \geq T_{\phi_i}$, where T_{ϕ_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 T_{ϕ_i} :

$$T_{\phi_i} = \min \{ \lambda_{\mathbf{v}_i}(x, y, t) / (x, y, t) \in P(\phi_i) \} \quad (6)$$

where $P(\phi_i)$ represents the set of relevant points of the filter ϕ_i [7].

Note the importance of having an adequate confidence measure when working with filters which are selective to spatio-temporal orientations.

3.2. Estimation of a motion pattern

In this section, the methodology to integrate the estimations corresponding to the set of filters which compose a motion pattern is described. Let P_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 P_k . Let Ω_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

$$\hat{\mathbf{v}}_k = \sum_{\mathbf{v}_i \in \Omega_k} \alpha_i \mathbf{v}_i \quad (7)$$

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

$$\alpha_i = \frac{\|\mu_{\mathbf{v}_i}\| / \lambda_{\mathbf{v}_i}}{\sum_{\mathbf{v}_j \in \Omega_k} \|\mu_{\mathbf{v}_j}\| / \lambda_{\mathbf{v}_j}} \quad (8)$$

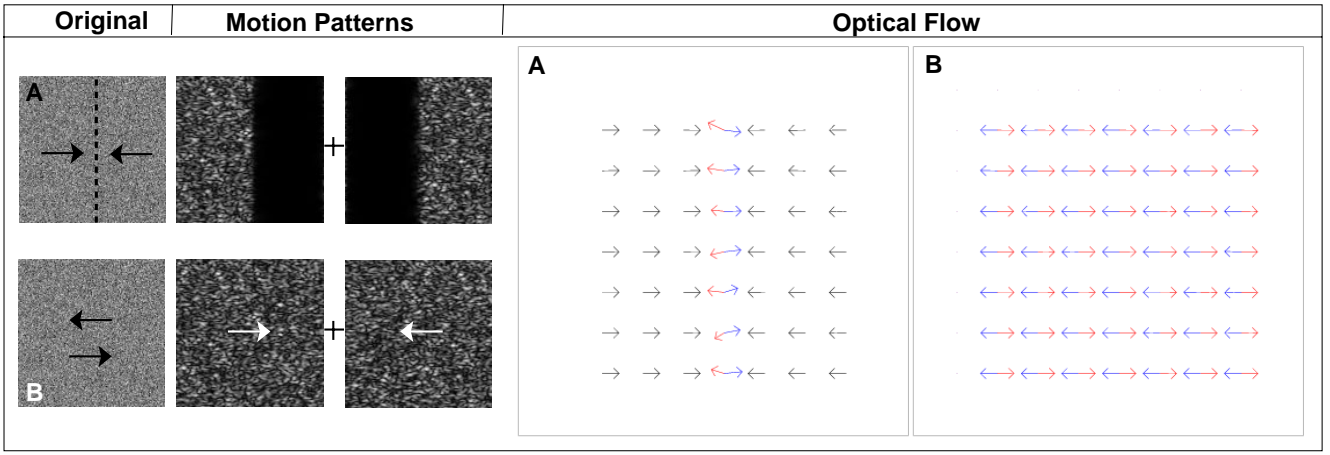


Fig. 2. Results with synthetic sequences.

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, $\hat{\mathbf{v}}_k$ will be a random variable with a Gaussian distribution with mean $\mu_{\hat{\mathbf{v}}_k} = \sum_{\Omega_k} \alpha_i \mu_{\mathbf{v}_i}$ and covariance $\Delta_{\hat{\mathbf{v}}_k} = \sum_{\Omega_k} \alpha_i^2 \Delta_{\mathbf{v}_i}$.

3.3. Multiple velocities representation.

The motion patterns allow to separate the relevant motions presented in a given sequence; therefore, they become an adequate tool to represent multiple velocities at the same location. Thus, our scheme will obtain the set of velocities $\bar{\mathbf{v}}$ at a given point (x, y, t) directly from the set of estimations calculated for each motion pattern:

$$\bar{\mathbf{v}} = \{\hat{\mathbf{v}}_k\}_{k=1\dots K} \quad (9)$$

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

4. RESULTS

In this section, the results obtained with real and synthetic sequences are showed to prove the performance of our model.

4.1. Synthetic sequences

The figure 2 shows two synthetic sequences which have been generated with Gaussian noise of mean 1 and variance 0. In this case, we have used the values $\gamma_1 = 0$, $\gamma_2 = 1$ y $\gamma_p = 1e - 5$ (with $\Delta_p^{-1} = \lambda_p I$ [9]) in equations (2) and (3). 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]. 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

	A (occlusion)	B (transparency)
Proposed technique	0.84°	0.44°
Nestares	3.93°	7.76°
Lucas&Kanade	4.79°	50.89°
Horn&Schunk	2.66°	52.77°
Nagel	8.59°	45.81°
Anandan	10.47°	47.78°
Singh	2.97°	45.27°
Uras	3.96°	57.86°

Table 1. Mean error comparison (techniques applied to the sequences in figure 2)

two motions with transparency: an opaque background pattern with velocity $(1,0)$, and a transparent foreground patterns 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 using multiple motions representation. Note 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 filed of the synthetic sequences, we can measure the performance of the proposed methodology. For this purpose, the following angular measure of error [10] between the correct velocity \mathbf{v}_c and an estimate \mathbf{v}_e will be used:

$$e(\mathbf{v}_c, \mathbf{v}_e) = \arccos(\mathbf{v}'_c, \mathbf{v}'_e) \quad (10)$$

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. Thus, if Ψ represents the set of correct velocities at the point (x, y, z) , the measure of error will be given by the equation:

$$E(\mathbf{v}_e) = \min \{e(\mathbf{v}_e, \mathbf{v}_r), \mathbf{v}_r \in \Psi\} \quad (11)$$

Table 1 shows a comparison between our methodology and the seven techniques discussed in [10] (the mean error for the two examples in figure 2 is reported in each case). As table 1 shows, the

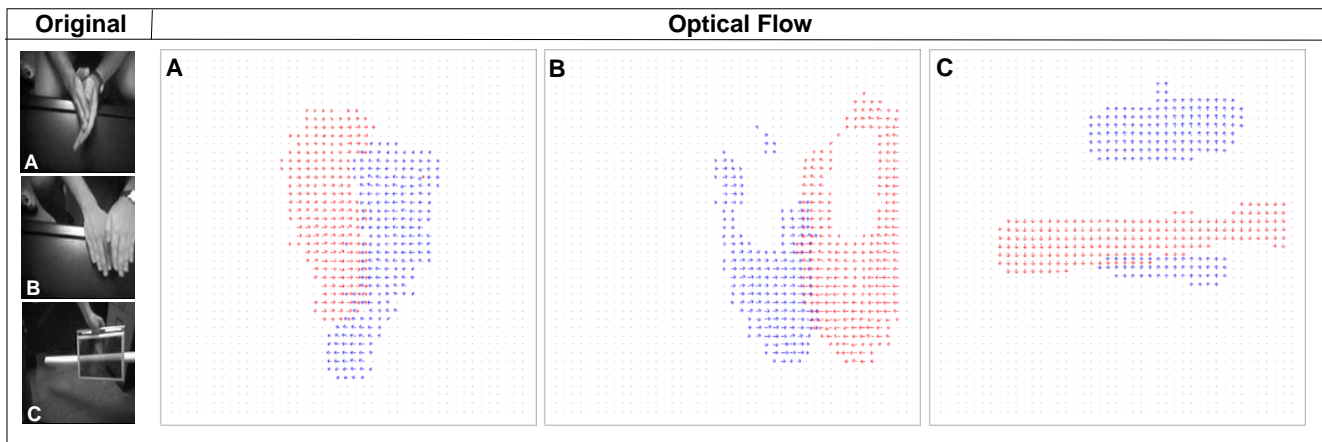


Fig. 3. Results with real sequences.

proposed method outperforms the other methods in all the cases (see in particular the example with transparency).

4.2. Real sequences

Figure 3 shows three examples with real sequences. In this case, we have used the values $\lambda_1 = 0$, $\lambda_2 = 1$ and $\lambda_p = 0.5$ with the same partial derivatives and weight parameters used in the synthetic case. For each example, the figure shows the central frame of the sequence and the optical flow estimated with our technique (for real images sequences, we do not have the true motion field, so we can only show the computed flow field). The first example (figure 3(A)) corresponds to a double motion without occlusions where two hands are clapping. The second one (figure 3(B)) shows an example of occlusion where a hand is crossing 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 (figure 3(C)). In all the cases, our methodology separates the two motions presented in the sequence and it estimates two velocities in the occlusion points.

5. CONCLUSIONS

In this paper, a new methodology for optical flow estimation has been presented. The proposed technique is able to represent multiple motions 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. The use of “motion patterns” has allowed to represent multiple motions, 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 transparency have been tested. In all the cases, the final results enlightens the consistency of the proposed algorithm.

6. REFERENCES

- [1] B.K.P. Horn and B.G. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] H. Nagel and W. Enkelmann, “An investigation of smoothness 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.
- [6] M.J. Black and P. Anandan, “The robust estimation of multiple motion: parametric and piecewise smooth flow fields,” *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.
- [7] J. Chamorro-Martinez, J. Fdez-Valdivia, J.A. Garcia, and J. Martinez-Baena, “A frequency-domain approach for the extraction of motion patterns,” *Proceedings of the ICASSP 2003, In press*, 2003.
- [8] R. Rodriguez-Sanchez, J.A. Garcia, 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1044–1072, 1999.
- [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.