A FREQUENCY-DOMAIN APPROACH FOR THE EXTRACTION OF MOTION PATTERNS

J. Chamorro-Martínez, J. Fdez-Valdivia, J.A. García, and J. Martínez-Baena

Department of Computer Science and Artificial Intelligence
University of Granada, Spain
e-mail: {jesus,jfv,jags,jbaena}@decsai.ugr.es

ABSTRACT

In this paper, a new frequency-domain approach to represent motions is proposed. The new scheme is based on a band-pass filtering with a set of logGabor spatio-temporal filters. It is well known that one of the main problems of these approaches is that a filter response varies with the spatial orientation of the underlying signal. To solve this spatial dependency, the proposed model allows to recombine information of motions that has been separated in several filter responses due to its spatial structure. For this purpose, motion patterns are detected as invariance in statistical structure across a range of spatio-temporal frequency bands. This technique is illustrated on real and simulated data sets, including sequences with occlusion and transparencies.

Keywords: Motion representation, spatio-temporal models, motion pattern, logGabor filters, .

1. INTRODUCTION

The motion is one of the most important information sources used in the analysis of an scene. Usually, the representation of motion allows to arrange the scene in order to make easier its analysis [1, 2, 3].

Many types of motion representation models have been proposed in the literature. In particular, some important approaches are those based on a band-pass filtering operation with a set of spatio-temporal filters selective to orientation and scale [4, 5, 6, 7]. These approaches are supported on the interpretation of motion in the Fourier domain, where the spectral information of a spatio-temporal translation goes to a plane whose orientation depends on the direction and velocity of the motion [8, 9, 10]. 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 in different responses. Moreover, a filter response will change if the spatial orientation or scale vary.

To solve the problems described above, we propose a new frequency based approach that groups the separated responses in order to obtain coherent and independent motion patterns. For this purpose, and using a new distribution of 3D logGabor filters over the spatio-temporal spectrum, *a motion pattern* is calculated as an invariance in statistical structure across a range of spatio-temporal frequency bands. This new scheme recombines responses that, even with different spatial characteristics, have continuity in its motion.

This work has been supported by the DGES (Spain) under grant PB98-1374

2. THE PROPOSED MODEL

The figure 1 shows a general diagram describing how the data flows through the proposed 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. 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. As a result, a set of distances between active filters is obtained.

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.

2.1. Bank of spatio-temporal filters

To decompose the sequence, a bank of logGabor filters is used. A logGabor function can be represented in the frequency domain as:

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

where σ_θ , σ_φ and σ_ρ are the angular and radial standard deviation, (θ_o, φ_o) is the orientation of the filter, and ρ_o is the central radial frequency. The bank of filters should be designed so that it tiles the frequency space uniformly. Hence we consider a bank with the following features:

1. For each radial frequency, 17 spherical orientations over dynamic planes are considered. Table 1 shows the angular coordinates used in the proposed bank.
2. The radial axis is divided into 3 equal octave bands. The wavelength in each orientation is set at 3, 6 and 12 pixels respectively.
3. The radial bandwidth is chosen as 1.2 octaves
4. The angular bandwidth is chosen as 30 degrees

The resultant filter bank is illustrated in the top of figure 1. Due to the conjugate symmetry in the Fourier domain, the filter design is only carried out on the half 3D frequency space.

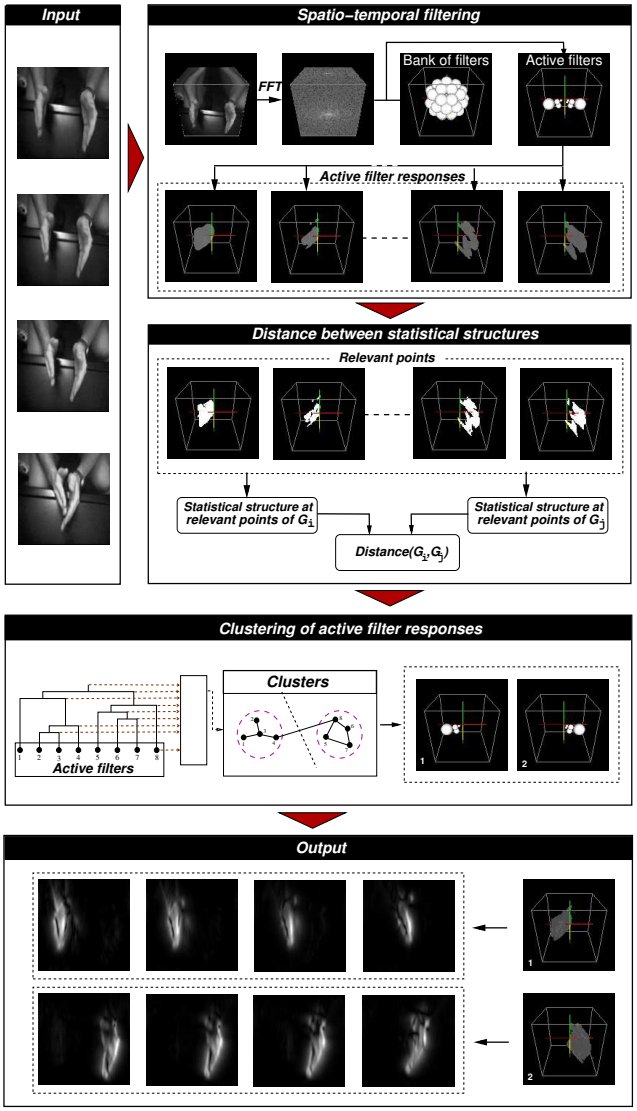

Fig. 1. A general diagram of the proposed model.

2.1.1. Active filters

In order to reduce the computational cost and to avoid the noisy or less relevant filter responses, a selection of filters that isolate spectral information corresponding to significant motions is performed. Given a filter ϕ_i, a measure of its relevance is defined as:

$$w_i = \frac{1}{Card[V(i)]} \sum_{(\rho,\theta,\varphi)\in V(i)} |F(\rho,\theta,\varphi)| \qquad (2)$$

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

$$|\rho-\rho_o| \leq \sigma_\rho \ , \ |\theta-\theta_o| \leq \sigma_\theta \ y \ |\varphi-\varphi_o| \leq \sigma_\varphi \qquad (3)$$

where σ_θ, σ_φ, σ_ρ and (θ_o,φ_o) are the logGabor filter parameters (let us remark that it is not necessary to calculate the responses of each filter to obtain these weights)

Using the filter relevance measure defined in (2), an unsupervised classification method is performed for each scale to group the filters into two classes: active ones and non-active ones. The cluster whose filters have the highest weights will determine the set of active filters (that will be noted $Active$). In our implementation, a hierarchical clustering [11] is used with a dissimilarity function between classes defined as

$$\delta(C_i,C_j) = |\mu_i - \mu_j| \qquad (4)$$

where

$$\mu_k = \frac{1}{Card\,[C_k]} \sum_{r\in C_k} w_r \qquad (5)$$

For each active filter, a set of 'relevant points' is computed. We calculate these points as local energy peaks on the filter responses [12]: given the response E_i of a filter ϕ_i, the maximal of E_i in the direction of the filter will determine the set of points which will focus our attention in the next stage.

2.2. Distance between filter responses.

In this section, a distance between the statistical structures of a given pair of filter responses is proposed. To represent a statistical structure, we use the notions of *separable feature* and *integral feature* introduced in [13]. A separable feature is defined as any relevant characteristic that may be obtained for a 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 five separable features proposed in [13] will be used.

Let $T^i(x,y,z) = \left[T^i_k(x,y,z)\right]_{k=1,2,\ldots L}$ be an integral feature at (x,y,z) which combines L separable features, noted as T^i_k, computed on the response of the filter ϕ_i. Let $\widehat{d}\left(T^i,T^j\right)$ be the distance between two integral features $T^i(x,y,z)$ and $T^j(x,y,z)$ given by the equation:

$$\widehat{d}\left(T^i,T^j\right) = \sum_{k=1}^{L} \frac{1}{Max_k} d(T^i_k,T^j_k) \qquad (6)$$

with Max_k being a normalization factor [13], and $d(\cdot)$ a distance between separable features (this measure $d(\cdot)$ is defined for each separable feature in [13])

Based on the previous equation, a distance between the responses of two filters ϕ_i y ϕ_j is defined as:

$$\widehat{D}(\phi_i,\phi_j) = D\,[i,j]^2 + D\,[j,i]^2 \qquad (7)$$

where

$$D\,[r,s] = \frac{1}{Card\,[P(r)]} \left(\sum_{P(r)} \left| \widehat{d}[T^r,T^s] \right|^\beta \right)^{\frac{1}{\beta}} \qquad (8)$$

with $\widehat{d}\,[T^r,T^s]$ being the distance between integral features given by (6), and $P(r)$ the set of relevant points for the filter ϕ_r .The default value of the exponent β in (8) has been fixed to 3.

(θ, φ)		
$(0.52, 0.00)$	$(-0.62, -0.53)$	$(-1.29, 0.45)$
$(0.62, 0.53)$	$(-1.08, 0.97)$	$(-1.05, 0.00)$
$(1.08, 0.97)$	$(1.08, -0.97)$	$(-1.29, -0.45)$
$(-1.08, 0.97)$	$(0.62, -0.53)$	$(1.29, -0.45)$
$(-0.62, 0.53)$	$(1.05, 0.00)$	$(1.57, 0.00)$
$(-0.52, 0.00)$	$(1.29, 0.45)$	

Table 1. Angular coordinates associated to the bank of filters (over an sphere of ratio 1).

2.3. Clustering of active filters

In order to obtain a partition $C_1, C_2, ..., C_N$ of active filters, with C_i representing a motion pattern, a clustering of the dataset $X = \{\phi_i \in Actives\}$ into an unknown number N of clusters is performed. For this purpose, a hierarchical clustering is used [11] with a dissimilarity function between classes defined on the basis of distances between statistical structures as

$$\delta(C_n, C_m) = min\left\{\widehat{D}(\phi_i, \phi_j)\ ,\ \phi_i \in C_n\ ,\ \phi_j \in C_m\right\} \quad (9)$$

where $\widehat{D}(\phi_i, \phi_j)$ is given by the equation (7).

2.3.1. Selection of the best partition

To select the level l of the hierarchy which will define the best partition $P^l = C_1, C_2, ..., C_N$, we propose the following function of goodness

$$f(P^l) = \frac{\gamma^*_{P^l}}{\varepsilon^*_{P^l}} \quad (10)$$

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:

$$\varepsilon^*_{P^l} = max\left\{\varepsilon_n \mid C_n \in P^l\right\} \quad (11)$$

$$\gamma^*_{P^l} = min\left\{\gamma_n \mid C_n \in P^l\right\} \quad (12)$$

The congruence degree ε_n and separation degree γ_n of a cluster C_n are defined as

$$\varepsilon_n = max\left\{cost(\mu^*_{i,j}) \mid \phi_i, \phi_j \in C_n\right\} \quad (13)$$

$$\gamma_n = min\left\{\delta(C_n, C_m) \mid m = 1, ..., N\ with\ m \neq n\right\} \quad (14)$$

where $\delta(C_n, C_m)$ is defined in (9), and $cost(\mu^*_{i,j})$ is the cost of the optimal path between two elements ϕ_i and ϕ_j in C_n calculated as follow: let $\prod_{ij}$ be the set of possibles paths linking ϕ_i and ϕ_j in C_n; given a path $\pi_{ij} \in \prod_{ij}$, its cost is defined as the greatest distance between two consecutive points on the path:

$$cost(\pi_{ij}) = max\left\{\widehat{D}(\phi_r, \phi_{r+1})\ /\ \phi_r, \phi_{r+1} \in \pi_{ij}\right\} \quad (15)$$

where ϕ_r and ϕ_{r+1} are two consecutive elements of π_{ij}, and $\widehat{D}(\phi_r, \phi_s)$ is defined in equation (7). The optimum path $\pi^*_{ij} \in \prod_{ij}$ between ϕ_i and ϕ_j is then defined as the path that links both filters with minimum cost:

$$\pi^*_{ij} = \underset{\pi_{i,j} \in \Pi_{i,j}}{argmin}\ \{cost(\pi_{i,j})\} \quad (16)$$

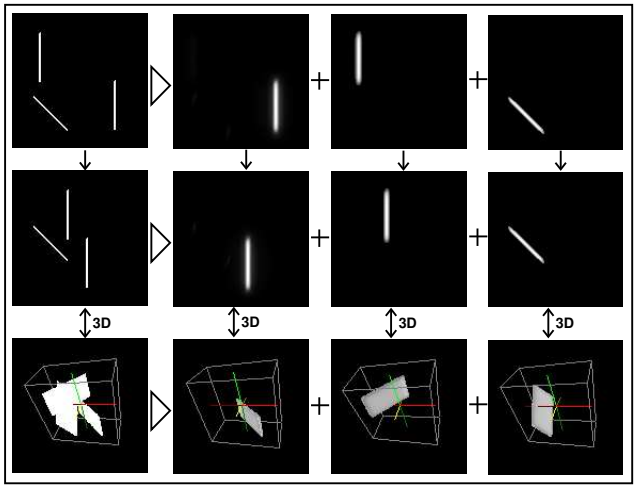

Fig. 2. Output of the model with a synthetic sequence.

Due to the merging process of the hierarchical clustering and the distance between classes used in this case (equation (9)), the congruence degree ε_n equals to the distance between the two cluster which were merged together to obtain C_n [13]. Thus, the calculus of ε_n do not increase the computational cost of the clustering.

3. RESULTS

In this section, the results obtained with real and synthetic sequences are showed to prove the performance of our model. Several cases have been tested, from simple motion to occlusions and transparencies. In all the cases, the figures show the first and the last frame of the original sequence and the corresponding motions patterns detected in each case

A synthetic case of pure translational motion with constant speed is showed in figure 2. Specifically, the example shows three bars with velocities of (1,0), (-1,0) and (0,-1) pixels/frame respectively. Looking at the 3D representation of the original sequence, three independent planes can be seen corresponding to the three bars in motion. Our model separates each one of these planes into three different spatio-temporal outputs corresponding to the three motions (each spatio-temporal image is obtained by the sum of the filters responses of the cluster). From this 3D representation, the sequence associated to each motion pattern is extracted.

Figure 3 shows three examples with real sequences. The first case corresponds to a double motion without occlusions where two hands are moving to clap. The second one 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 placed in the first plane. As figure 3 shows, in all the cases our model generates an output for each motion pattern present in the sequence. Let us remark that the problem of the occlusion is solved by our model by mean of the spatio-temporal continuity of forms. Furthermore, this approach is capable of detecting motions even when different velocities and spatial orientations are present.

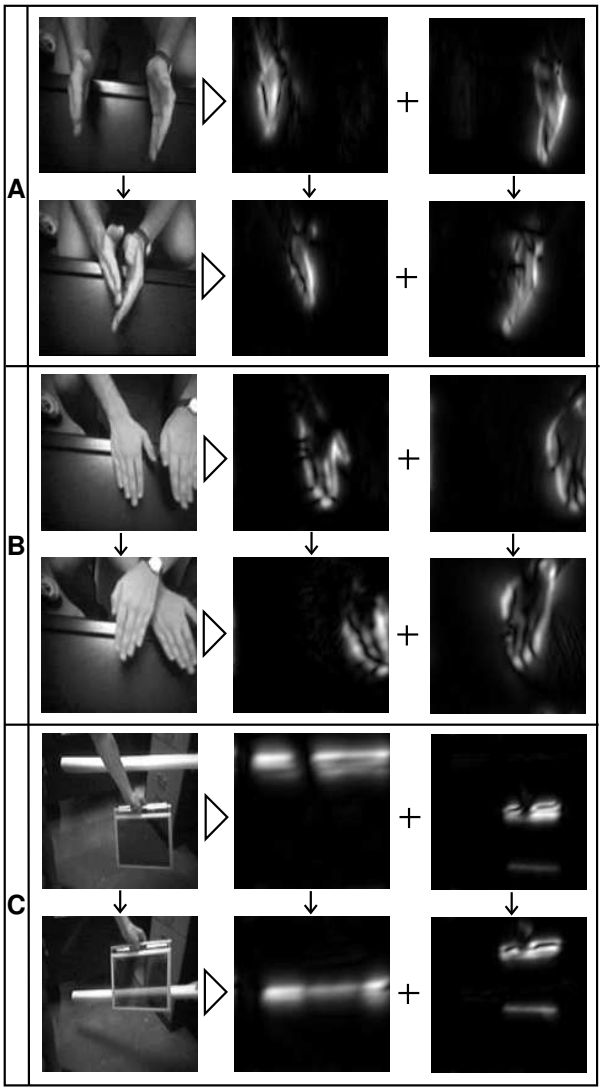

Fig. 3. Results with real sequences.

4. CONCLUSIONS

In this paper, a new approach to represent motions has been presented. For this purpose, a motion pattern has been identified on the basis of invariance in statistical structure across a range of spatio-temporal frequency bands. To span the spatio-temporal spectrum, logGabor functions have been adopted as an appropriate method to construct filters of arbitrary bandwidth. The new approach allows to recombine information of motions that has been separated in several filter responses due to its spatial structure; as a result, the proposed model generates an output for each coherent and independent motion detected in the sequence, avoiding the classic problem associated with a representation based on spatio-temporal filters.

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.

5. REFERENCES

[1] A.M. Dockstader, S.L.and Tekalp, "Adaptive motion estimation using local measures of texture and similarity," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2091–2094, 2000.

[2] L. Wiskott, "Segmentation from motion: combining gabor and mallat wavelets to overcom de aperture and correspondence problem," *Pattern Recognition*, vol. 32, pp. 1751–1766, 1999.

[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] E.H. Adelson and J.R. Bergen, "Spatiotemporal energy models for the perception of motion," *Journal of the Optical Society of America A.*, vol. 2, no. 2, pp. 284–299, Feb 1985.

[5] 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.

[6] F. Dekeyser, P. Bouthemy, and P Perez, "Spatio-temporal wiener filtering of image sequences using a parametric motion model," *Proceedings of the International Conference on Image Processing*, vol. 1, pp. 208–211, 2000.

[7] H. Liu, M. Hong, M. Herman, and A. Chellappa, "A general motion model and spatio-temporal filters for computing optical flow," *International Journal of Computer Vision*, vol. 22, no. 2, pp. 141–172, 1997.

[8] A.B. Watson and A.J. Ahumada, "A look at motion in the frequency domain," Tech. Rep., NASA Tech. Memo. TM-84352, 1983.

[9] S.S Beauchemin and J.L. Barron, "On the fourier properties of discontinuous motion," *Journal of Mathematical Imaging and Vision*, vol. 13, pp. 155–172, 2000.

[10] S.S Beauchemin and J.L. Barron, "The frequency structure of 1d occluding image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 2, pp. 200–206, 2000.

[11] Anil K. Jain and Richard C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, 1988.

[12] M.C. Morrone and R.A. Owens, "Feature detection from local energy," *Pattern Recognition Letters*, vol. 6, pp. 303–313, 1987.

[13] 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.