We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



1

Energy Feature Integration for Motion Segmentation

Raquel Dosil, Xosé R. Fdez-Vidal, Xosé M. Pardo & Antón García
Universidade de Santiago de Compostela
Spain

1. Introduction

This chapter deals with the problem of segmentation of apparent-motion. Apparent-motion segmentation can be stated as the identification and classification of regions undergoing the same motion pattern along a video sequence. Motion segmentation has a great importance in robotic applications such as autonomous navigation and active vision. In autonomous navigation, motion segmentation is used in identifying mobile obstacles and estimating their motion parameters to predict trajectories. In active vision, the system must identify its target and control the cameras to track it. Usually, segmentation is based on some low level feature describing the motion of each pixel in a video frame. So far, the variety of approaches to deal with the problems of motion feature extraction and motion segmentation that has been proposed in literature is huge. However, all of them suffer from different shortcomings and up to date there is no completely satisfactory solution.

Recent approaches to motion segmentation include, for example, that of Sato and Aggarwal (Sato & Aggarwal, 2004), where they define the Temporal Spatio-Velocity (TSV) transform as a Hough transform evaluated over windowed spatio-temporal images. Segmentation is accomplished by thresholding of the TSV image. Each resulting blob represents a motion pattern. This solution has proved to be very robust to occlusions, noise, low contrast, etc. Its main drawback is that it is limited to translational motion with constant velocity.

It is very common to use a Kalman filter to estimate velocity parameters from intensity observations (Boykov & Huttenlocher, 2000). Kalman filtering alone presents severe problems with occlusions and abrupt changes, like large inter-frame displacements or deformations of the object. If a prior model is available, the combined use of Kalman filtering and template matching is the typical approach to deal with occlusions. For instance, Kervrann and Heitz (1998) define an a priori model with global and local deformations. They apply matching with spatial features for initialization and reinitialization of global rigid transformation and local deformation parameters in case of abrupt changes and Kalman filtering for tracking otherwise. Nguyen and Smeulders (2004) perform template matching and updating by means of Kalman filtering.

Template matching can deal even with total occlusions during a period of several frames. Nevertheless, when no prior model is available, the most common approach is statistical region classification, like Bayesian clustering (Chang et al., 1997; Montoliu and Pla, 2005). These techniques are very sensitive to noise and aliasing. Furthermore, they do not provide

Open Access Database www.i-techonline.com

Source: Vision Systems: Segmentation and Pattern Recognition, ISBN 987-3-902613-05-9
Edited by: Goro Obinata and Ashish Dutta, pp.546, I-Tech, Vienna, Austria, June 2007

a method for correlating the segmentations obtained for different frames to deal with tracking. Tracking is straightforward when the identified regions keep constant motion parameters along the sequence and different objects undergo different motion patterns. Otherwise, it is difficult to know the correspondences between the regions extracted from different frames, especially when large displacements or occlusions take place.

An early approach by Wang and Adelson (1994) tackle this issue using a layered representation. Firstly, they perform motion segmentation by region clustering under affine motion constraints. Layers are then determined by accumulating information about different regions from different frames. This information is related to texture, depth and occlusion relationships. The main limitations of this model, that make it unpractical in most situations, are that it needs a large number of frames to compute layers and significant depth variations between layers.

A very appealing alternative for segmentation is the application of an active model at each frame guided by motion features or a combination of motion and static features (Paragios & Deriche, 2000). Deformable models are able to impose continuity and smoothness constraints while being flexible.

The performance of any segmentation technique is strongly dependent on the chosen low-level features to characterize motion. In segmentation using active models, low-level features are employed to define the image potential. The simplest approach uses temporal derivatives as motion features, as in the work of Paragios and Deriche (2000). They use de inter-frame difference to statistically classify image points into static or mobile. Actually, the inter-frame difference is not a motion estimation technique, since it only performs motion detection without modelling it. It can only distinguish between static and mobile regions. Therefore, this method is only valid for static background scenes and can not classify motion patterns according to their velocity and direction of motion.

Most motion segmentation models are based on the estimation of optical flow, i.e., the 2D velocity of image points or regions, based on the variation of their intensity values. Mansouri and Konrad (2003) have employed optical flow estimation to segmentation with an active model. They propose a competition approach based on a level set representation. Optimization is based on a maximum posterior probability criterion, leading to an energy minimization process, where energy is associated to the overall residuals of mobile objects and static background. Residuals are computed as the difference between measured intensities and those estimated under the constraint of affine transformation motion model. However, optical flow estimations present diverse kinds of problems depending on the estimation technique (Barron et al., 1994; Stiller & Konrad, 1999). In general, most optical flow estimation techniques assume brightness constancy along frames, which in real situations does not always hold, and restrict allowed motions to some specific model, such as translational or affine motion. Particularly, differential methods for estimating the velocity parameters consistent with the brightness constancy assumption are not very robust to noise, aliasing, occlusions and large inter-frame displacements.

Alternatively, energy filtering based algorithms (Heeger, 1987; Simoncelli & Adelson, 1991; Watson & Ahumada, 1985; Adelson & Bergen, 1985; Fleet, 1992) estimate motion from the responses of spatio-temporal filter pairs in quadrature, tuned to different scales and orientations. Spatio-temporal orientation sensitivity is translated into sensitivity to spatial orientation, speed and direction of motion. These techniques are known to be robust to noise and aliasing, to give confident measurements of velocity and to allow an easy treatment of

the aperture problem, i.e., the reliable estimation of the direction of motion. However, to the best of our knowledge there is not motion segmentation method based on energy filtering.
Another important subject in segmentation with active models is how to initialize the model at each frame. A common solution is to use the segmentation of each frame to initialize the model at the next frame. Paragios and Deriche (2000) use this approach. The first frame is automatically initialized based on the inter-frame difference between the first two frames. The main problem of the initialization with the previous segmentation arises with total occlusions, when the object disappears from the scene for a number of frames, since no initial state is available when the object reappears. The case of large inter-frame displacements is also problematic. The object can be very distant from its previous position, so that the initial state might not be able converge to the new position. Tsechpenakis et al. (2004) solve these problems by initializing each frame, not using the previous segmentation, but employing the motion information available for that frame. In that work, motion features are only employed for initialization and the image potential depends only on spatial information.

1.2 Our Approach

In this chapter we present a model for motion segmentation that combines an active model with a low-level representation of motion based on energy filtering. The model is based solely on the information extracted from the input data without the use of prior knowledge. Our low level motion representation is obtained from a multiresolution representation by clustering of band-pass versions of the sequence, according to a criterion that links bands associated to the same motion pattern. Multiresolution decomposition is accomplished by a bank of non-causal spatio-temporal energy filters that are tuned to different scales and spatio-temporal orientations. The complex-valued volume generated as the response of a spatio-temporal energy filter to a given video sequence is here called a *band-pass feature*, *subband feature*, *elementary energy feature* or simply *energy feature*. We will call *integral features, composite energy features* or simply *composite features* to motion patterns with multiple speed, direction and scale contents generated as a combination of elementary energy features in a cluster. The set of filters associated to an energy feature cluster are referred to as *composite-feature detector*. Segmentation is accomplished using composite features to define the image potential and initial state of a geodesic active model (Caselles, 1997) at each frame. The composite feature representation will be applied directly, without estimating motion parameters.
Composite energy features have proved to be a powerful tool for the representation of visually independent spatial patterns in 2D data (Rodriguez-Sánchez et al., 1999), volumetric data (Dosil, 2005; Dosil et al., 2005b) and video sequences (Chamorro-Martínez et al., 2003). To identify relevant composite features in a sequence, it is necessary to define an integration criterion able to relate elementary energy features contributing to the same motion pattern. In previous works (Dosil, 2005; Dosil et al., 2005a; Dosil et al., 2005b), we have introduced an integration criterion inspired in biological vision that improves the computational cost and performance of earlier approaches (Rodriguez-Sánchez et al., 1999; Chamorro-Martínez et al., 2003). It is based on the hypothesis of Morrone and Owens (1987) that the Human Visual System (HVS) perceives features at points of locally maximal Phase Congruence (PC). PC is the measure of the local degree of alignment of the local phase of Fourier components of a signal. The sensitivity of the HVS to PC has also been

studied by other authors (Fleet, 1992; Oppenheim & Lim, 1981; Ross et al., 1989; du Buf, 1994). As demonstrated by Venkatesh and Owens (1990), points whose PC is locally maximal coincide with the locations of energy maxima. Our working hypothesis is that local energy maxima of an image are associated to locations where a set of multiresolution components of the signal contribute constructively with alignment of their local energy maxima. Hence, we can identify composite features as groups of features that present a high degree of alignment in their energy maxima. For this reason, we employ a measure of the correlation between pairs of frequency features as a measure of similarity for cluster analysis (Dosil et al., 2005a).

Here, we extend the concept of PC for spatio-temporal signals to define our criterion for spatio-temporal energy feature clustering. We will show that composite features thus defined are robust to noise, occlusions and large inter-frame displacements and can be used to isolate visually independent motion patterns with different velocity, direction and scale content.

The outline of this chapter is as follows. Section 2 is dedicated to the composite feature representation model. Section 3 is devoted to the proposed method for segmentation with active models. In section 4 we illustrate the behaviour of the model in different problematic situations, including some standard video sequences. In 5 we expound some conclusions of the work.

2. Composite-Feature Detector Synthesis

The method for extraction of composite energy features consists of the decomposition of the image in a set of band-pass features and their subsequent grouping according to some dissimilarity measure (Dosil, 2005; Dosil et al., 2005a). The set of frequency features involved in the process is determined by selecting from a predefined spatio-temporal filter bank those bands that are more likely to be associated to relevant motion patterns, which we call *active* bands. Composite-feature detectors are clusters of these active filters. Each visual pattern is reconstructed as a combination of the responses of the filters in a given cluster. Filter grouping is accomplished by applying hierarchical cluster analysis to the set of band-pass versions of the video sequence. The dissimilarity measure between pairs of frequency features is related to the degree of phase congruence between a pair of features, through the quantification of the alignment among their local energy maxima. The following subsections detail the process.

2.1 Bank of Spatio-Temporal Filters

The bank of spatio-temporal filters applied here (Dosil, 2005; Dosil et al., 2005b) uses an extension to 3D of the log Gabor function (Field, 1994). The filter is designed in the frequency domain, since it has no analytical expression in the spatial domain. Filtering is realized as the inner product between the transfer function of the filter and the Fourier transform of the sequence. Filtering in the Fourier domain is very fast when using Fast Fourier Transform and Inverse Fast Fourier Transform algorithms.

The filters' transfer function T is designed in spherical frequency coordinates as the product of separable factors R and S in the radial and angular components respectively, such that $T = R \cdot S$. The radial term R is given by the log Gabor function (Field, 1993)

$$R(\rho;\rho_i)=\exp\left(-\frac{\left(\log\left(\rho/\rho_i\right)\right)^2}{2\left(\log\left(\sigma_{\rho i}/\rho_i\right)\right)^2}\right), \tag{1}$$

where $\sigma_{\rho i}$ is the standard deviation and ρ_i the central radial frequency of the filter.
The angular component is designed to achieve orientation selectivity in both the azimuthal component ϕ_i of the filter, which reflects the spatial orientation of the pattern in a frame and the direction of movement, and the elevation component θ_i, related to the velocity of the motion pattern. For static patterns $\theta_i=0$. To achieve rotational symmetry, S is defined as a Gaussian on the angular distance α between the position vector of a given point $\mathbf{f}$ in the spectral domain and the direction of the filter $\mathbf{v}=(\cos\phi_i\cdot\cos\theta_i,\cos\phi_i\cdot\sin\theta_i,\sin\phi_i)$ (Faas & van Vliet, 2003)

$$S(\phi,\theta;\phi_i,\theta_i)=S(\alpha)=\exp\left(-\alpha^2/2\sigma_{\alpha i}^2\right),\ \text{ with }\ \alpha(\phi_i,\theta_i)=\operatorname{acos}\left(\mathbf{f}\cdot\mathbf{v}/\|\mathbf{f}\|\right), \tag{2}$$

where $\mathbf{f}$ is expressed in Cartesian coordinates and $\sigma_{\alpha i}$ is the angular standard deviation.
Active filters are selected from a predefined band partition of the 3D frequency space. Frequency bands are determined by the central frequency (ρ_i,ϕ_i,θ_i) of the filters and their width parameters $(\sigma_{\rho i},\sigma_{\alpha i})$. In the predefined bank, frequency is sampled so that $\rho_i=\{1/2, 1/4, 1/8, 1/16\}$ in pixels^{-1}. Parameter $\sigma_{\rho i}$ is determined for each band in order to obtain 2 octave bandwidth. θ_i is sampled uniformly while the number of ϕ_i samples decreases with elevation in order to keep the "density" of filters constant, by maintaining equal arc-length between adjacent ϕ_i samples over the unit radius sphere. Following this criterion, the filter bank has been designed using 23 directions, i.e. (ϕ_i,θ_i) pairs, yielding 92 bands. $\sigma_{\alpha i}$ is set to 25º for all orientations. Hence, the bank involves 4×23 filters that yield a redundant decomposition and cover a wide range of the spectrum.

2.2 Selection of Active Bands

To achieve improved performance, it is convenient to reduce the number of bands involved in cluster analysis. The exclusion of frequency channels that are not likely to contribute to motion patterns facilitates the identification of clusters associated to composite motion features. Furthermore, it reduces computational cost. Here, we have introduced a channel selection stage based on a statistical analysis of the amplitude responses of the band-pass features. Selected channels are called *active*.
Our method for the selection of active channels is based on the works of Field (1994) and Nestares et al. (2004). Field has studied the statistics of the responses of a multiresolution log-Gabor wavelet representation scheme that resembles the coding in the visual system of mammalians. He has observed that the filter responses histograms are not Gaussian, but leptokurtic distributions –pointed distributions with long tails–, revealing the sparse nature of both the sensory coding and the features from natural images. According to Field, when the parameters of the wavelet codification fit those in the mammalian visual system, the histogram of the responses is highly leptokurtic. This is reflected in the fourth cumulant of the distribution. Namely, he uses the kurtosis to characterize the sparseness of the response.

Regarding spatio-temporal analysis, Nestares et al. (2000) applied channel selection to a bank of spatio-temporal filters, with third order Gaussian derivatives as basis functions, based on the statistics of filters responses. They have observed that features corresponding to mobile targets present sparser responses than those associated to background –weather static or moving. This fact is illustrated in Fig. 1**.** They measure different statistical magnitudes reflecting sparseness of the amplitude response, realize a ranking of the channels based on such measures and perform channel selection by taking the n first channels in the ranking, where n is a prefixed number.

Based on these two works, we have designed our filter selection method. The statistical measure employed to characterize each channel is the kurtosis excess γ_2

$$\gamma_2 = k_4 / k_2^2 - 3 \tag{3}$$

where k_4 and k_2 are respectively the fourth and second cumulants of a histogram. If the kurtosis excess takes a positive value, the distribution is called leptokurtic and presents a

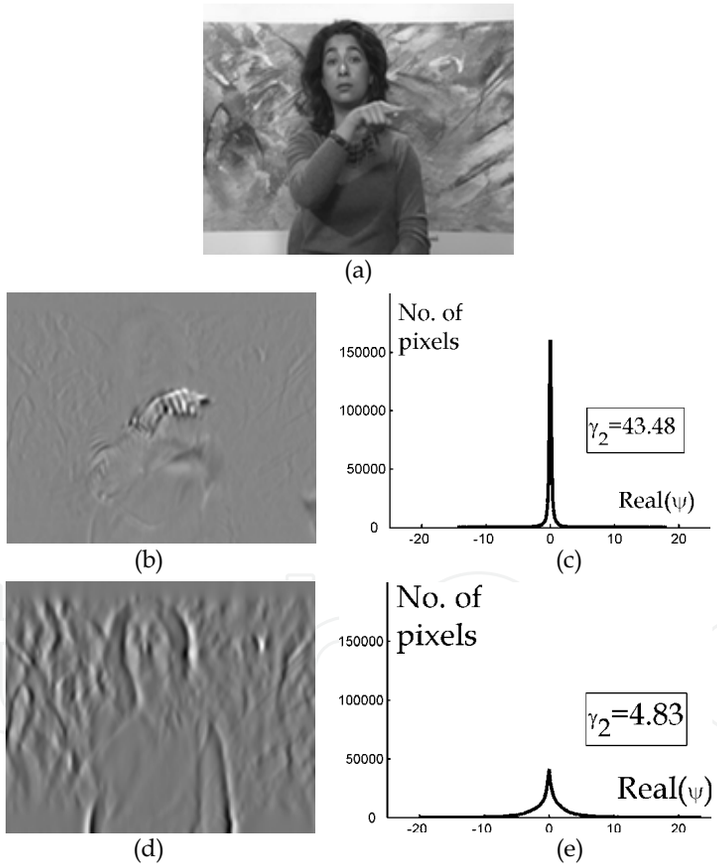


Fig. 1. (a) A frame of the standard sequence *Silent*, showing a moving hand. (b) and (d) A frame of the real component of two band-pass features of the *Silent* video sequence. (c) and (e) Histograms corresponding to band-pass features in (b) and (d)

narrow peak and long tails. If it is negative, the distribution is called platykurtic and presents a broad central lobe and short tails. Distributions with zero kurtosis excess, like the Gaussian distribution, are called mesokurtic.

We measure γ_2 for both the real and imaginary components of each feature ψ_i and then compose a single measure δ

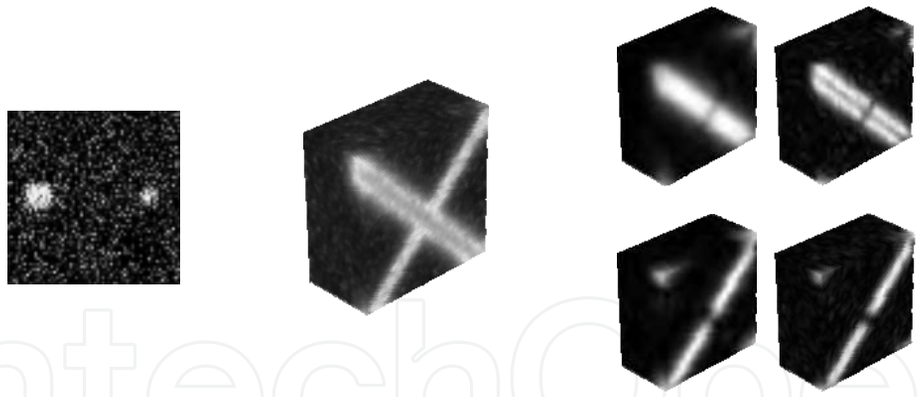
$$\delta_i = \gamma_2(\mathrm{Re}(\psi_i)) + \gamma_2(\mathrm{Im}(\psi_i)) \quad (4)$$

Instead of selecting the n first channels in the ranking of δ, we perform cluster analysis to identify two clusters, one for active channels with large values of δ and another for non active channels. Here, we have applied a k-means algorithm. The cluster of active channels is identified as the one with larger average δ.

2.3 Energy Feature Clustering

Integration of elementary features is tackled in a global fashion, not locally (point-wise). Besides computational efficiency, this provides robustness, since it intrinsically correlates same-pattern locations –in space and time–, avoiding grouping of disconnected regions.

As aforementioned, it seems plausible that the visual system of humans perceives features where Fourier components are locally in phase (Morrone & Owens, 1987). Our criterion for integration of frequency features is based on the assumption that a maximum in phase congruence implies the presence of maxima in the same location in a subset of subband versions of the data. Points of locally maximal phase congruence are also points of locally maximal energy density (Venkatesh & Owens, 1990). Hence, subband images contributing



(a) (b) (c)

Fig. 2. (a) A frame of a synthetic video sequence, where two light spots move from side to side with opposite direction. (b) A cut along the temporal axis of the total energy of the sequence. (c) Energy of some band-pass versions of the sequence. Those on top row correspond to one of the spots and present some degree of concurrence on their local energy maxima. Bottom row shows two band-pass features correspondent to the other motion pattern.

to the same visual pattern should present a large degree of alignment in their local energy maxima, i.e., their energy maxima present some degree of concurrence –see Fig. 2.
Here, the dissimilarity between two subband features is determined by estimating the degree of alignment between the local maxima of their local energy. Alignment is quantified using the correlation coefficient ρ of the energy maps of each pair $\{\psi_i, \psi_j\}$ of subband features. This measure has proved to produce good results in visual pattern extraction from volumetric data (Dosil, 2005; Dosil et al., 2005b). If $A(\psi) = ||\psi|| = (\mathrm{Im}(\psi)^2 + \mathrm{Re}(\psi)^2)^{1/2}$, the actual distance is calculated from $\rho(A_i, A_j)$ as follows

$$D_\rho(A_i, A_j) = \left(1 - \sqrt{(1 + \rho(A_i, A_j))/2}\right)^2 . \quad (5)$$

This distance function takes values in the range [0,1]. The minimum value corresponds to perfect match of maxima –linear dependence with positive slope– and the maximum corresponds to the case of perfect fit with negative slope, like, for example, an image and its inverse. This measure does not depend on the selection of any parameter and does not involve the discrete estimation of joint and/or marginal probabilities –histograms.
Our approach generates visual patterns by clustering of active bands. Dissimilarities between each pair of frequency features are computed to build a dissimilarity matrix. To determine the clusters from the dissimilarity matrix, a hierarchical clustering method has been chosen, using a Ward's algorithm to determine inter-cluster distance, which has proved to improve other metrics (Jain & Dubes, 1988). The number of clusters N_c that a hierarchical technique generates is an input parameter of the algorithm. The usual strategy to determine the N_c is to run the algorithm for each possible N_c and evaluate the quality of each resulting configuration according to a given validity index. A modification of the Davies-Boulding index proposed by Pal and Biswas (1996) has proved to produce good results for our application. It is a graph-theory based index that measures the compactness of the clusters in relation to their separation.
A stage of cluster merging follows cluster analysis. Clusters with average intercluster correlation values close to one –specifically, greater than 0.75– are merged to form a single cluster. This is made since we can not evaluate the quality of a single cluster containing all features. Besides, hierarchical algorithms can only analyse the magnitude of a distance in relation to others, not in an absolute fashion. This fact is often a cause of wrong classification, splitting clusters into smaller subgroups.

2.4 Composite Feature Reconstruction

The response ψ to an energy filter is a complex-valued sequence, where the real and imaginary components account for even and odd symmetric features respectively. In this section we describe how elementary complex features in a cluster are combined to obtain a composite-feature Ψ. We will use real, imaginary or amplitude representations depending on the application. For simple visualization we will employ only the real components. In the definition of the image potential of an active model, we are only interested on odd-symmetric components, which represent mobile contours, so only the imaginary parts of the elementary features will be involved. For initialization we are interested in the regions occupied by the moving objects, so the amplitude of the responses $||\psi||$ is the chosen representation.
Here we define the general rule for the reconstruction of Ψ based on a given representation

E of the responses of the filters, that can be either Re(ψ), Im(ψ) or the amplitude $A(\psi) = ||\psi|| = (\mathrm{Im}(\psi)^2 + \mathrm{Re}(\psi)^2)^{1/2}$. The easiest way of constructing the response Ψ of a set Ω_j of filters in a cluster j is by linear summation

$$\Psi^j(x,y,t) = \sum_{i \in \Omega_j} E_i(x,y,t). \tag{6}$$

However, simple summation presents one important problem. There might be features in the cluster that contribute, not only to the corresponding motion pattern, but also to other patterns or static structures in the sequence. Only points with contributions from all features in the cluster should have a non null response to the composite feature detector. To avoid this problem, we define the composite feature as the linear summation of elementary features weighted by a mask indicating locations with contribution of all features in the clustering. The mask is constructed as the summation of the thresholded responses $\widetilde{E}_i$ of the elementary features, normalized by the total number of features. Thresholding is accomplished by applying a sigmoid to the responses of elementary features, so that $\widetilde{E}_i \in [0,1]$. Therefore, the mask takes value 1 wherever all features contribute to the composite pattern and rapidly decreases otherwise, with a smooth transition

$$\Psi^j(x,y,t) = \frac{\sum_{i \in \Omega_j} \widetilde{E}_i}{Card(\Omega_j)} \sum_{i \in \Omega_j} E_i, \tag{7}$$

where Ω_j is the set of all bands in cluster j. The effect of masking is illustrated in Fig. 3. Reconstruction using different representations for E is illustrated in Fig. 4.

For visualization purposes, we will employ the real component $\Psi^j_{even} = \Psi^j(E_i = \mathrm{Re}(\psi_i))$. The odd-symmetric representation of Ψ is constructed by full-wave rectification of expression in equation (7), so that $\Psi^j_{odd} = \left|\Psi^j(E_i = \mathrm{Im}(\psi_i))\right|$ does not have into account the sign of the contour. The amplitude representation $\Psi^j_{amp} = \Psi^j(E_i = \|\psi_i\|)$ is used for initialization in general situations. The even-symmetric representation is used for initialization of objects with uniform contrast and is defined by applying a half-wave rectification $\max(\pm\Psi^j_{even}, 0)$, with sign depending on the specific contrast.



Fig. 3. A frame of the "silent" video sequence: *Left*: Input data. *Centre*: Even-symmetric representation of the response of one of the composite features detected, corresponding to the moving hand, calculated using equation (6) and, *Right*: using equation (7)

3. Motion Pattern Segmentation

The previously described method for feature clustering is able to isolate different static and dynamic patterns from a video sequence. Nevertheless, it is not suitable by itself to segment mobile objects for several reasons. To begin with, the mobile contours might present low contrast in some regions, giving place to disconnected contours. Furthermore, when the moving object is occluded by static objects, its contour presents static parts, so that the representation with motion patterns is incomplete. This happens also when a contour is oriented in the direction of motion; only the motion of the beginning and end of the segment is detected. For these reasons, we will produce a higher-level representation of the motion patterns from the proposed low-level motion representation.
In this work we have chosen an active model as a high level representation technique, namely, the geodesic active model. We will perform a segmentation process for each composite feature, which we will refer to as Ψ, omitting the superindex. From that pattern, we derive the initial state of the model and the image potential in each frame. After evolving a geodesic model in each frame, the segmented sequence is generated by stacking the segmented frames. A scheme of the segmentation method is presented in Fig. 5. Next subsections describe the technique in depth.

3.1 Geodesic Active Model

To accomplish segmentation, here we have chosen an implicit representation for object boundaries, where the contour is defined as the zero level set of an implicit function. Implicit active models present important advantages regarding parametric representations. The problem of contour re-sampling when stretching, shrinking, merging and splitting is avoided. They allow for the simultaneous detection of inner and outer contours of an object and naturally manage topological changes. Inner and outer regions are determined by the sign of the implicit function.

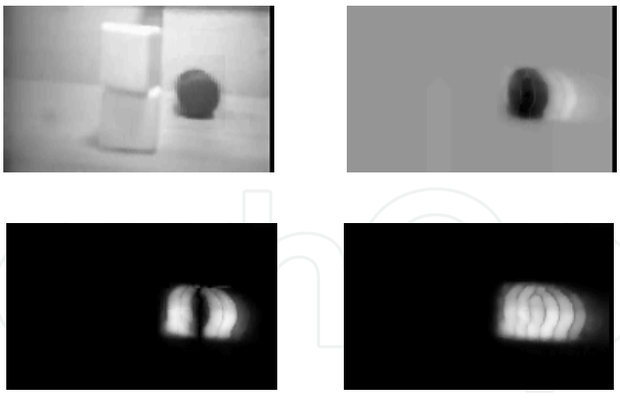


Fig. 4. Top left: One frame of an example sequence with a moving dark cylinder. The remainder images show different representations for one of the composite features identified by the presented representation method. Top right: Even representation. Bottom left: Odd representation. Bottom right: Amplitude representation.

The optimization model employed for segmentation is the geodesic active model (Caselles et al., 1997). The evolution of the contour is determined from the evolution of the zero-level set of an implicit function representing the distance u to the contour. Let $\Omega := [0, a_x] \times [0, a_y]$ be the frame domain and consider a scalar image $u_0(x, y)$ on Ω. We employ here symbol τ for time in the evolution equations of u to distinguish it from the frame index t. Then, the equations governing the evolution of the implicit function are the following:

$$\begin{aligned} u(x, y, t = t_k, \tau = 0) &= u_0(x, y, t = t_k) \quad \text{on } \Omega \\ \frac{\partial u}{\partial \tau} &= g(s)|\nabla u|(\kappa + c) + \nabla g(s) \nabla u \quad \text{on } \Omega \times (0, \infty) \end{aligned}' \tag{8}$$

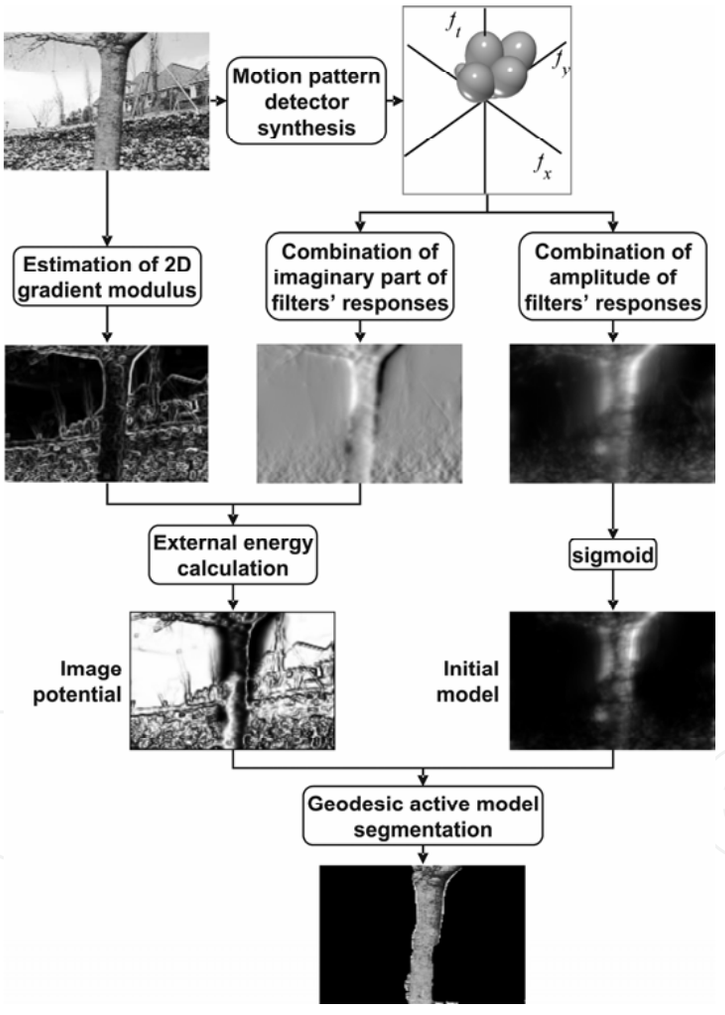


Fig. 5. Scheme of the segmentation technique

where c is real constant, g is a function with values in the interval [0, 1] that decreases in the presence of relevant image features, s is the selected image feature and κ is the curvature.
If the second term in the right side of the previous equation is not considered, what remains is the expression for the *geometric active model*, where $g \cdot (\kappa + c)$ represents the velocity of the evolving contour. The role of the curvature can be interpreted as a geometry dependent velocity. Its effect is also equivalent to the internal forces in a thin-plate-membrane spline model, also called *snake* (Kass et al., 1988). Constant c represents a constant velocity or *advection* velocity in the geometric active model and is equivalent to a balloon force in the snake model. Factor $g(s)$ has the effect of stopping the contour at the desired feature locations. The second term in the right side is the image dependent term, which pushes the level-set towards image features. It is analogous to external forces in the snake formulation. This term did not appear in the geometric active model, which made necessary the use of a constant velocity term to approach the level-set to the object boundary. With the use of a feature attraction term this is no longer necessary. However, if the model is initialized far away from the image features to be segmented, the attraction term may not have enough influence over the level-set. As a result, the constant velocity term is often used to compensate for the lack of an initialization stage.
The concrete implementation of the geodesic active model used here is the one described in (Weickert & Kühne, 2003). We do not employ balloon forces, since with the initialization, described in subsection 3.3, they are no longer needed, so then $c = 0$. In the following subsection we define the image potential as a function of the composite energy features.

3.2 Image Potential Definition

The expression for the image potential function is the same as in (Weickert & Kühne, 2003)

$$g(s) = \frac{1}{1 + (s/s_{\min})^p}, \tag{9}$$

with p and $s_{\min}$ being real constants.
The potential of the mobile contour depends on the odd-symmetric representation of the motion pattern, Ψ_{odd}, reconstructed as the rectified sum of the imaginary components of the responses to its constituent filters. This motion pattern may present artefacts, due to the diffusion of patterns from neighbouring frames produced when applying energy filtering. This situation is illustrated in Fig. 6.a and b. To minimize the influence of these artefacts, the motion pattern is modulated by a factor representing the localization of spatial contours. It is calculated from the 2D contour detector response by thresholding using a sigmoid function

$$C_m(x, y, t_k) = \frac{1}{1 + \exp(-K(C_s(x, y, t_k) - C_0))} \frac{\Psi_{odd}(x, y, t_k)}{\max(\Psi_{odd}(x, y, t_k))}, \tag{10}$$

where C_s is a spatial contour detector based on the frame gradient, C_0 is the gradient threshold and K is a positive real constant. The specific values taken here are $C_0 = 0.1$ and $K = 20$. The effect of this modulation can be observed in Fig. 6.c and d.
Although here we are interested in segmenting objects based on their motion features, it is convenient to include a spatial term in the potential. This is necessary to close the contour when part of the boundary of the moving object remains static –when there is a partial

occlusion by a static object or scene boundary or when part of the moving contour is parallel to the direction of motion. Therefore, the image feature s is the weighted sum of two terms, C_m and C_s, respectively related to spatio-temporal and pure spatial information.

$$s = w_s C_s + w_m C_m \text{ , with } w_s + w_m = 1 \text{ and } w_s, w_m > 0 \tag{11}$$

The weight of the spatial term w_s must be much smaller than the motion term weight w_m, so that the active model does not get "hooked" on a static contours not belonging to the target object. Here, the values of the weights have been set as follows: $w_s = 0.1$ and $w_m = 0.9$.
The spatial feature employed to define the spatial potential is the regularized image gradient. Regularization of a frame is accomplished here by feature-preserving 2D anisotropic diffusion, which brakes diffusion in the presence of contours and corners. The 3D version of the filter is described in (Dosil & Pardo, 2003). If $I^*(x, y, t_k)$ is the smoothed version of the k^{th} frame, then

$$C_s(x, y, t_k) = \left\| \nabla I^*(x, y, t_k) \right\| \Big/ \max\left(\left\| \nabla I^*(x, y, t_k) \right\| \right) \tag{12}$$

In the potential function g, $p = 2$ and $s_{\min}$ is calculated so that, on average, $g(s(x, y)) = 0.01$, $\forall x, y : C_m(x, y) > 0.1$. Considering the geodesic active model in a front propagation framework, $g = 0.01$ means a sufficiently slow speed of the propagating front to produce stopping in practical situations.

3.3 Initialization

The initial state of the geodesic active model is defined, in a general situation, from the amplitude representation of the selected motion pattern Ψ_{amp} unless other solution is specified. To enhance the response of the cluster we apply a sigmoid thresholding to Ψ_{amp}. The result is remapped to the interval [–1,1]. The zero-level of the resulting image is the initial state of the contour.

$$u_0(x, y, t_k) = \frac{2}{1 + \exp\left(-K\left(\Psi_{amp}(x, y, t_k) - \Psi_0 \right) \right)} - 1 \tag{13}$$

When the object remains static during a number of frames the visual pattern has a null response. For this reason, the initial model is defined as the weighted sum of two terms, respectively associated to the current and previous frames. The contribution from the previous frame must be very small.

$$u_0(x, y, t_k) = w_k \left(\frac{2}{1 + \exp\left(-K\left(\Psi_{amp}(x, y, t_k) - \Psi_0 \right) \right)} - 1 \right) + w_{k-1}\, u_{\tau = \tau_{\max}}(x, y, t_{k-1}) \tag{14}$$

with w_k and w_{k-1} being positive real constants that verify $w_k + w_{k-1} = 1$. In the experiments presented in next section, $w_k = 0.9$, $w_{k-1} = 0.1$, $K = 20$ and $\Psi_0 = 0.1$.

4 Results

In this section, some results are presented to show the behaviour of the method in problematic situations. The results are compared to an alternative implementation that

employs typical solutions for initialization and definition of image potential in a way similar to that of Paragios and Deriche (2000): the initial state is the segmentation of the previous frame and the image potential depends on the inter-frame difference. However, instead of defining the image potential from the temporal derivative using a Bayesian classification, the image potential is the same as with our method, except that the odd-symmetric representation of the motion pattern is replaced by the inter-frame difference $I_t(x,y,t_k)=I(x,y,t_k)-I(x,y,t_{k-1})$. This is to compare the performance of our low-level features with inter-frame difference under the equal conditions. The initial state for the first frame is defined by user interaction.

The complete video sequences with the original data and the segmentation results are available at http://www-gva.dec.usc.es/~rdosil/motion_segmentation_examples.htm.They are summarized in the next subsections.

4.1 Moving Background

In this example, we use part –27 frames– of the well-known sequence "flower garden". It is a static scene recorded by a moving camera –see Fig. 7. The estimation of the inter-frame difference along frames produces large values at every image contour. The temporal derivative can be thresholded, or more sophisticated techniques for classifying regions into mobile or static can be employed, as in (Paragios & Deriche, 2000). However, it is not possible to isolate independent motion patterns just from the information brought by I_t. In contrast, visual pattern decomposition allows isolation of motion patterns with different speeds, which in the 3D spatio-temporal domain is translated into patterns with different orientations. This is made clear visualizing a cut of the image and the motion patterns in the *x-t* plane, as in Fig. 8.

Consequently, the image potential estimated from the temporal derivative feature presents

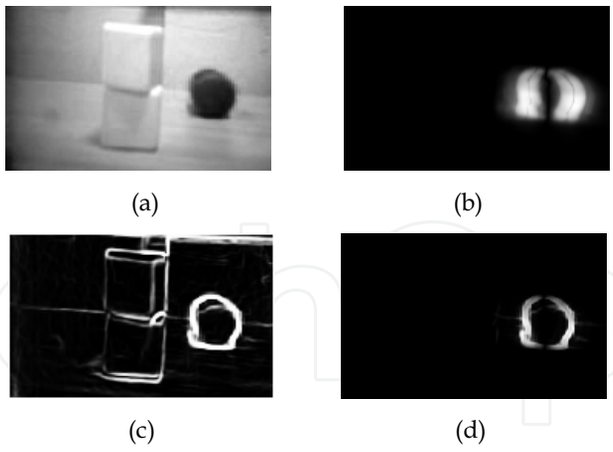


Fig. 6. (a) One frame of an example sequence where the dark cylinder is moving from left to right. For one of the composite features detected: (b) Ψ_{odd} representation. (c) Gradient after sigmoid thresholding. (d) Motion feature C_m from equation (10) as the product of images (b) and (c).

deep minima all over the image and the active model is not able to distinguish foreground objects from background, as can be seen in Fig. 9. The image potential in our implementation considers only the motion pattern corresponding to the foreground object, leading to a correct segmentation, as shown in Fig. 5.

4.2 Large Inter-Frame Displacements

When the sampling rate is too small in relation to the speed of the moving object, it is difficult to find the correspondence between the positions of the object in two consecutive frames. Most optical flow estimation techniques present strong limitations in the allowed displacements. Differential methods, based on the brightness constancy assumption, try to find the position of a pixel in the next frame imposing some motion model. Frequently, the search is restricted to a small neighborhood. This limitation can be overcome by coarse-to-fine analysis or by imposing smoothness constraints (Barron et al., 1994). Still, large displacements are usually problematic. The Kalman filter is not robust to abrupt changes when no template is available (Boykov & Hutterlocher, 2000).

When using the inter-frame difference in combination with an active model, the correspondence is accomplished through the evolution of the model from the previous state to the next one (Paragios & Deriche, 2000). However, when initializing with the previous segmentation, the model is not able to track the target if the previous segmentation does not intersect the object in the following frame. Fig. 10 shows an example case of this situation, taken from the standard sequence "table tennis". The alternative implementation of the active model fails to track the ball, as shown in the images of the second row of the figure. When using energy features, the composite motion patterns are isolated from each other. In this way, the correspondence of the motion estimations in different frames is naturally provided by the representation scheme, as shown in the third row of Fig. 10. This is also a property of techniques for motion estimation based on the Hough transform (Sato & Aggarwal, 2004). Nevertheless, this approach in not appropriate for this sequence, since the speed of the moving objects is variable in magnitude and direction –it is an oscillating movement –, so that it does not describe a straight line or a plane in the spatio-temporal domain –see left image in Fig. 10. Unlike the Hough transform, composite features combine elementary velocity-tuned features to deal with complex motion patterns, as can be seen in the image at the right of Fig. 10. We take advantage of both the isolation of the motion pattern and the integration of different velocity components associated to the moving objects, to initialize the model at each frame. Hence, the model arrives to a correct segmentation of the ball –see Fig. 10, bottom row– besides the large displacement produced and the changing direction or movement.

4.3 Total Occlusions

Occlusions give rise to the same problem as with fast objects. Again, initialization with composite frequency-features leads to a correct segmentation even when the object disappears from the scene during several frames. An example of this is presented in Fig. 12. In segmentation based on region classification (Chang et al., 1997; Montoliu & Pla, 2005) the statistical models extracted for each of the identified regions could be employed for tracking by finding the correspondence among them in different frames. However, occlusions carry

the additional problem of determining when the object has left the scene and when it reappears. The same problem applies for Kalman filter segmentation. Returning to the alternative implementation of the active model, when the object leaves the scene and no



Fig. 7. A frame of the "flower garden" video sequence.

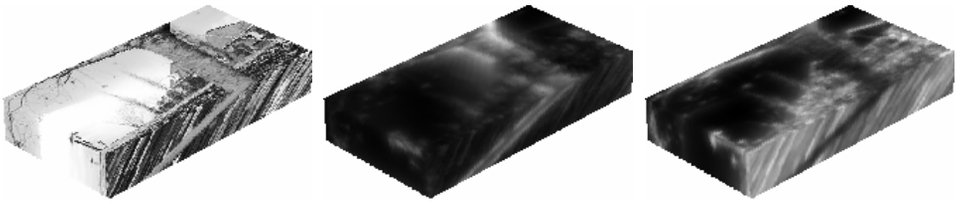


Fig. 8. A transversal cut of the original sequence: *Left*: Input data. *Centre and Right*: Ψ_{amp} of the two motion patterns isolated by the composite-feature representation model



Fig. 9. For the frame in Fig. 7, *Left*: Inter-frame difference, *Centre*: Image potential derived from I_t, *Right*: Segmentation obtained using image potential from image at the centre and initialization with the segmentation from previous frame.

other motion features are detected, the model collapses and the contour disappears from the scene in the remainder frames –see Fig. 12, second row. The solution of Paragios and Deriche could be employed to reinitialize the model by applying motion detection again, but it can not be ensured that the newly detected motion feature corresponds to the same pattern.

Again, due to the nature of our representation, the composite energy-features do not need a stage for finding correspondence between regions occupied by a motion pattern in different frames –see Fig. 12, third row. The model collapses when the cylinder disappears behind a static object and is reinitialized automatically when it reappears, without the need of a prior model. Initialization for this particular example has been achieved by using $\max(-\Psi_{even}, 0)$ of the selected composite feature, instead of the amplitude representation. This is because the target object does not present severe contrast changes in its surface, so half-wave rectification of the even-symmetric representation allows better localization of the object, facilitating convergence –the real component are inverted before rectification, since the object has negative contrast.

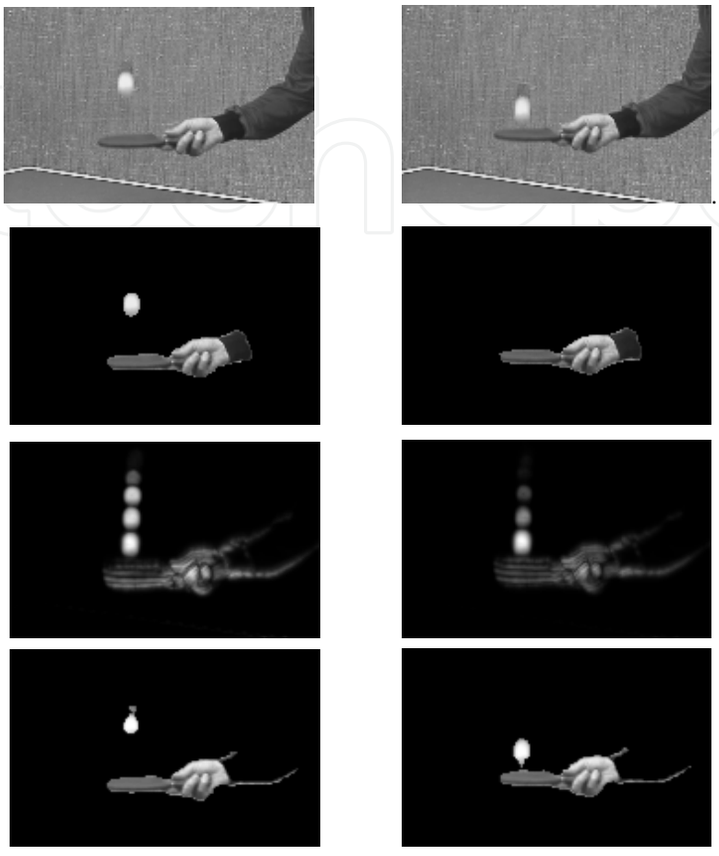


Fig. 10. *Top*: Two consecutive frames of the "table tennis" video sequence. For frames on top row, *2nd Row*: Segmentations produced by the alternative active model, *3rd Row*: Ψ_{amp} of the selected composite-feature, *Bottom*: Segmentation obtained with one of the detected composite-features

Fig. 13 shows another example presenting occlusions where the occluding object is also mobile. As can be seen, the alterative active model fails in segmenting both motion patterns, both due to initialization with previous segmentation and incapability of distinguishing both motion patterns, while our model properly segments both patterns using the composite-features provided by our representation scheme.

4.4 Complex Motion Patterns

The following example shows the ability of the method to deal with complex motion patterns and complex scenarios. In particular, the following sequence, a fragment of the standard movie know as "silent", presents different moving parts, each one with variable speed and direction and deformations as well, over a textured static background. As can be seen in the images from the top row of Fig. 14, the motion pattern of the hand can not be properly described by an affine transformation. Moreover, the brightness constancy assumption is not verified here.

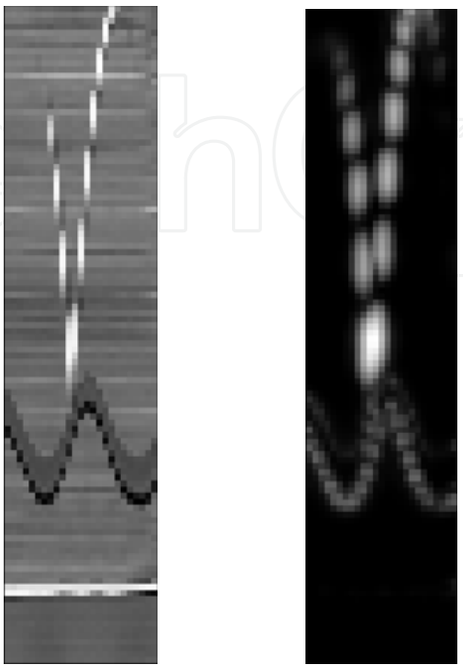


Fig. 11. Left: A cut of the "table tennis" sequence in the x-t plane. The white pattern corresponds to the ball and the gray/black sinusoidal pattern bellow corresponds to the bat. Right: A cut of the in the x-t plane of the Ψamp representation of the composite feature used in segmentation in bottom row of Fig. 14.

The active model based on the inter-frame difference is not able to properly converge to the contour of the hand, as seen in second row of Fig. 14. This is due to both the interference of other moving parts or shadows and wrong initialization. From the results, it can be seen that, despite the complexity of the image, the composite-feature representation model is able to isolate the hand and properly represent its changing shape in different frames – Fig. 14.

4.5 Discussion

In the examples presented, it can be observed that the proposed model for the representation of motion is able to group band-pass features associated to visually independent motion patterns without the use of prior knowledge. It must be said that the multiresolution scheme defined in section 2.1 has a great influence in the results, specially the selection of the number of filters and the angular bandwidth, which is related to the ability of the model to discriminate between different but proximal orientations, speeds and directions of motion.

In the comparison with the alternative implementation, which uses typical solutions for initialization and image potential definition, the proposed approach outperforms. Although there are other approaches that may present improved performance in solving some of the reported problems, it seems that none of them can successfully deal with all of them.

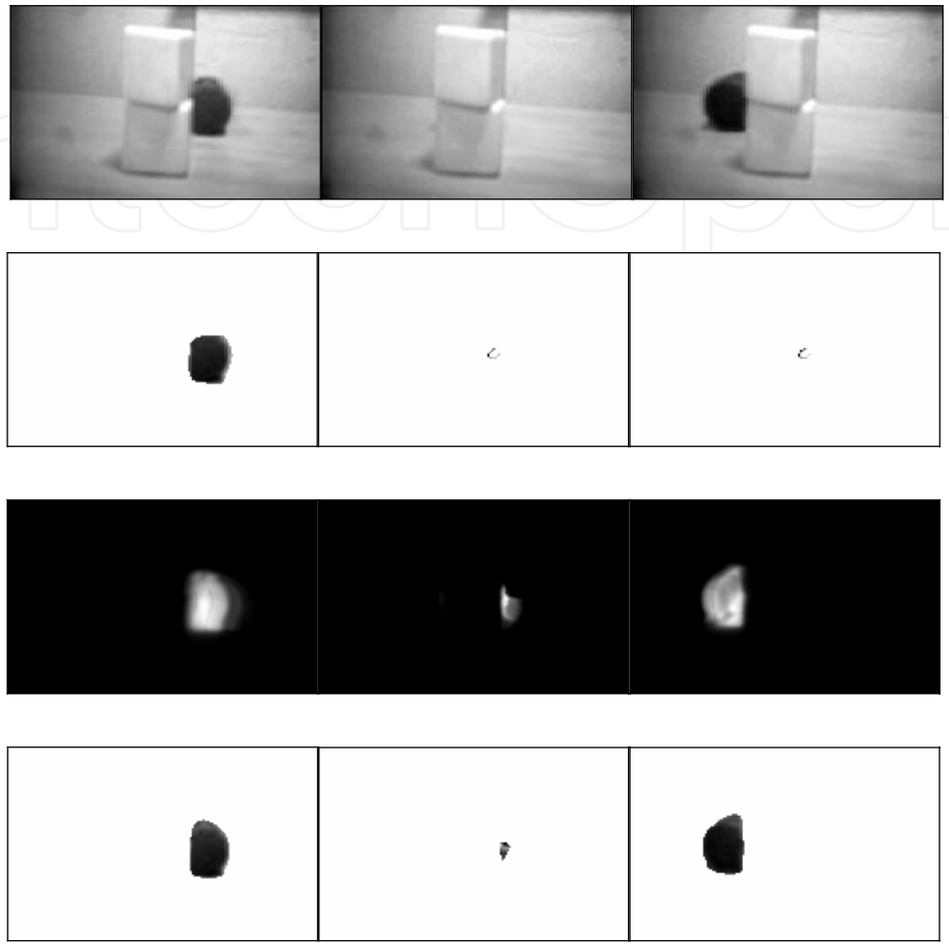


Fig. 12. *Top:* Three frames of a video sequence where a moving object is totally occluded during several frames. *2nd Row*: Segmentation using initialization with previous segmentation. *3rd Row*: Initialization of the frames using the Ψ_{amp} representation of one of the detected composite-feature. *Bottom*: Segmentation using initialization with the composite feature

The key characteristic of composite-feature representation scheme is that integration is accomplished by clustering on frequency bands, not by point-wise region clustering. This fact yields a representation that intrinsically correlates information from different frames, in a way similar to techniques based on the Hough transform –but not limited to constant speed and direction. This property is responsible for the robustness to partial and total occlusions and large inter-frame displacements or deformations. Furthermore, the proposed representation scheme does not limit the possible motion patterns to predefined models, like translational or affine motion, thanks to the composition of elementary motion features. This is evident in example from section 4.4 –"silent" video sequence– where also local deformations of the target appear. Besides, energy filtering provides robustness to noise and aliasing.

On the other hand, composite features present a larger temporal-diffusion effect than, for example, the inter-frame difference. However, this effect is suitably corrected by the gradient masking. Naturally, other typical shortcomings associated to velocity tuned filters can appear. For instance, there may be problems with low contrast regions, since the representation model is related to the contrast of features. This is observed in the example of

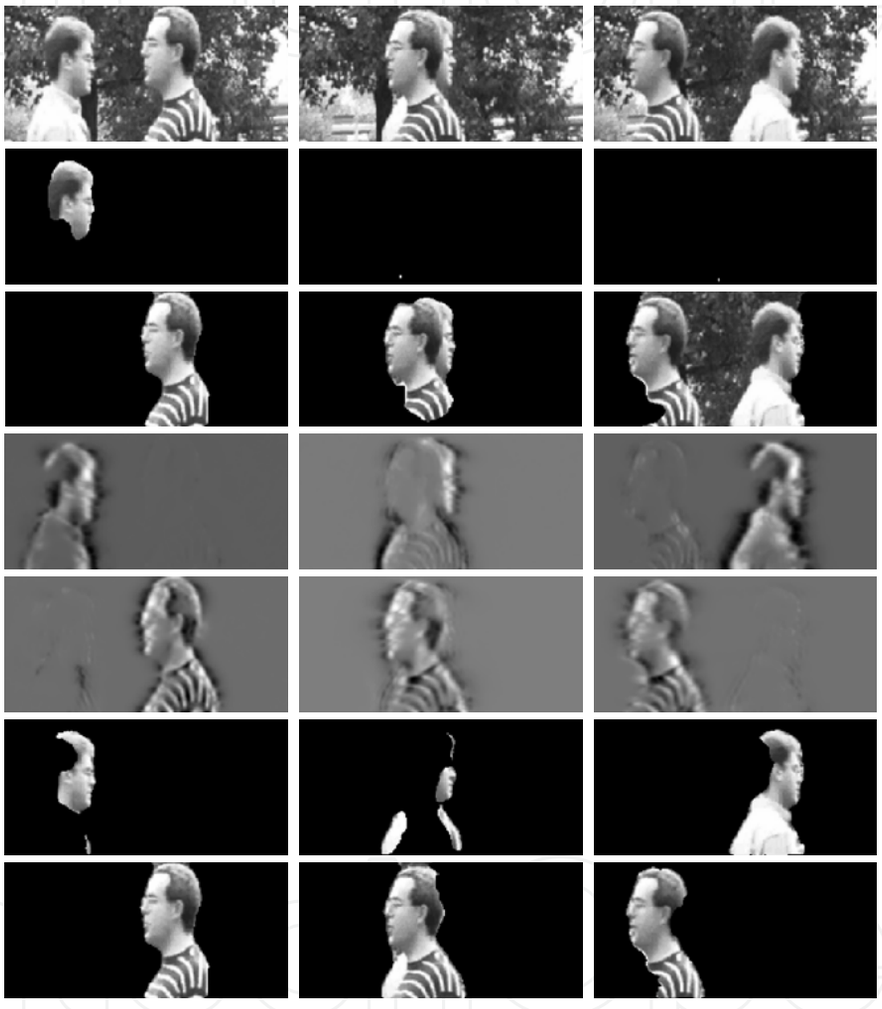


Fig. 13. Three frames of a sequence showing two occluding motion patterns. 1st row: Input data. 2nd and 3rd rows: Inter-frame difference based segmentation, using a different initialization for each of the motion patterns. 4th and 5th rows: Ψeven of two of the obtained composite-features, corresponding to the two motion patters. 6th and 7th rows: Segmentations produced using composite-features from rows 4th and 5th respectively.

section 4.1, where the contour between tree and flower bed is poorly defined –see Fig. 5.

5. Conclusions

In this chapter, a new active model for the segmentation of motion patterns from video sequences has been presented. It employs a motion representation based on composite energy features. It consists on the clustering of elementary band-pass features, which can be considered velocity tuned features. Integration is accomplished by extending the notion of phase congruence to spatio-temporal signals. The active model uses this motion information both for image potential definition and initialization of the model in each frame of the sequence.

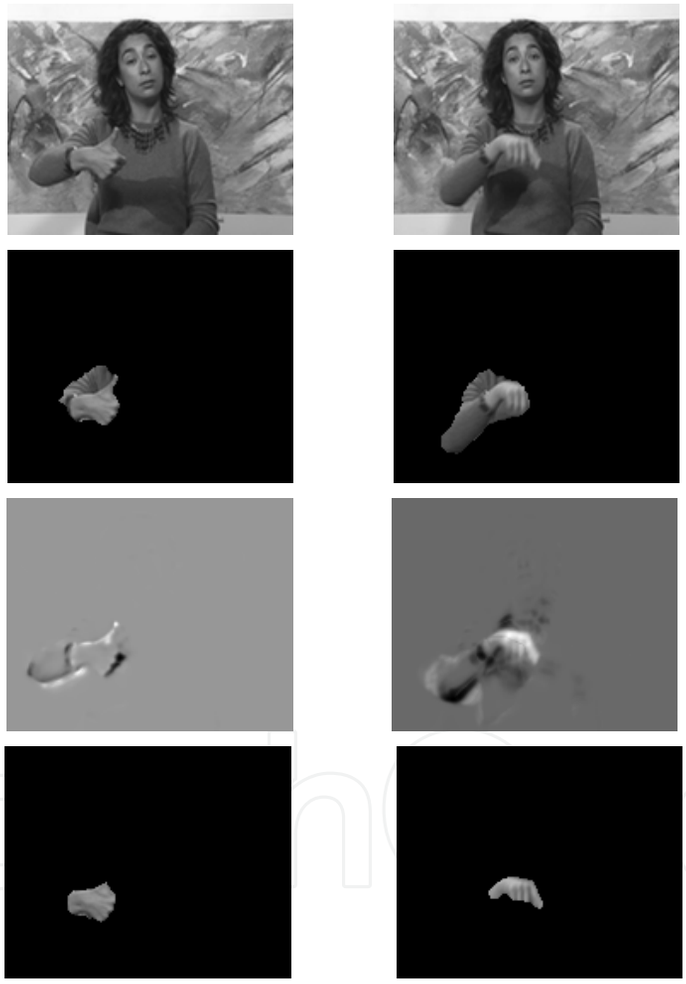


Fig. 14. Two frames of the "silent" video sequence: Top Row: Input data. 2nd Row: Segmentation using the active model based on the inter-frame difference. 3rd Row: Ψeven of the selected motion pattern. Bottom Row: Segmentation using the active model based on the composite-feature

The motion representation has proved to be able to isolate independent motion patterns from a video sequence. The integration criterion of spatio-temporal phase congruence gives place to a decomposition of the sequence into visually relevant motion patterns without the use of a priori knowledge.
The combination of geodesic active models and our motion representation yields a motion segmentation tool that presents good performance in many of the typical problematic situations, where previous approaches fail to properly segment and track, such as presence of noise and aliasing, partial and total occlusions, large inter-frame displacements or deformations, moving background and complex motion patterns. In the comparison with an alternative implementation, that employs typical solutions for initialization and definition of image potential, our method shows enhanced behavior.

6. Acknowledgements

This work has been financially supported by the Ministry of Education and Science of the Spanish Government through the research project TIN2006-08447.

7. References

Adelson, E.H. & Bergen, J.R. (1985). Spatiotemporal Energy Models for the Perception of Motion, J Opt Soc Am A, Vol. 2, No. 2, February 1985, pp. 284-299, ISSN: 1084-7529

Barron, J.L.; Fleet, D.J. & Beauchemin, S.S. (1994). Performance of Optical Flow Techniques, Int J Comput Vis, Vol. 12, No. 1, February 1994, pp. 43-77, ISSN: 0920-5691

Boykov, Y. & Huttenlocher, D.P. (2000). Adaptive Bayesian Recognition in Tracking Rigid Objects, Proceedings of the IEEE Comput Soc Conf Comput Vis Pattern Recogn (CVPR), Vol. II, pp. 697-704, Hilton Head Island (South Carolina), June 2000, IEEE Computer Society, Los Alamintos (CA), ISBN: 0-7695-0662-3

Caselles, V.; Kimmel, R. & Sapiro, G. (1997). Geodesic Active Contours, Int J Comput Vis, Vol. 22, No. 1, February 1997, pp. 61-79, ISSN: 0920-5691

Chamorro-Martínez, J.; Fdez-Valdivia, J.; García, J.A. & Martínez-Baena, J. (2003). A frequency Domain Approach for the Extraction of Motion Patterns, Proceedings of the IEEE Acoust Speech Signal Process, Vol. III, pp. 165-168, April 2003, IEEE Society, Los Alamitos (CA), ISBN: 0-7803-7663-3

Chang, M.M.; Tekalp, A.M. & Sezan, M.I. (1997). Simultaneous Motion Estimation and Segmentation, IEEE Trans Image Process, Vol. 6, No. 9, September 1997, pp. 1326-1333, ISSN: 1057-7149

Dosil, R. (2005) Data Driven Detection of Composite Feature Detectors for 3D Image Analysis. PhD Thesis, Universidade de Santiago de Compostela, ISBN: 84-9750-560-3, Santiago de Compostela (Spain) URL: http://www-gva.dec.usc.es/~rdosil/ficheiros/thesis_ dosil.pdf

Dosil, R.; Fdez-Vidal, X.R. & Pardo, X.M. (2005a). Dissimilarity Measures for Visual Pattern Partitioning, Proceedings of the 2nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA), Vol. II, pp. 287-294, Estoril (Portugal), June 2005, In: Lecture Notes in Computer Science, Vol. 3523, Marques, J. & Pérez de la Blanca, N. (Eds.), Springer-Verlag, Berlin Heidelberg, ISBN: 3-540-26154-0

Dosil, R.; Pardo, X.M. & Fdez-Vidal, X.R. (2005b). Decomposition of 3D Medical Images into Visual Patterns, IEEE Trans Biomed Eng, Vol. 52, No. 12, December 2005, pp. 2115-2118, ISSN: 0018-9294

Dosil, R. & Pardo, X.M. (2003). Generalized Ellipsoids and Anisotropic Filtering for Segmentation Improvement in 3D Medical Imaging, Image Vis Comput, Vol. 21, No. 4, April 2003, pp. 325-343, ISSN: 0262-8856

du Buf, J. (1994). Ramp Edges, Mach Bands and the Functional Significance of the Simple Cell Assembly, Biological Cybernetics, Vol. 70, No. 5, March 1994, pp. 449–461, ISSN: 0340-1200

Faas, F.G.A. & van Vliet, L.J. (2003). 3D-Orientation Space; Filters and Sampling, Proceedings of the 13th Scandinavian Conference in Image Analysis (SCIA), pp.36-42, Halmstad (Sweden), July 2003, In: Lecture Notes in Computer Science, Vol. 2749, Bigun, J. & Gustavsson, T. (Eds.), Springer-Verlag, Berlin Heidelberg, ISBN: 3-540-40601-8

Field, D.J. (1993). Scale–Invariance and self-similar "wavelet" Transforms: An Analysis of Natural Scenes and Mammalian Visual Systems, In: Wavelets, fractals and Fourier Transforms, pp. 151-193, Farge, M.; Hunt, J.C.R. & Vassilicos, J.C. (Eds.), Clarendon Press, Oxford, ISBN: 019853647X

Field, D.J. (1994). What is the Goal of Sensory Coding, Neural Computation, Vol. 6, No. 4, July 1994, pp. 559-601, ISSN: 0899-7667

Fleet, D. (1992). Measurement of Image Velocity, Kluwer Academic Publishers, ISBN: 0792391985, Massachusetts

Heeger, D.J. (1987). Model for the Extraction of Image Flow, J Opt Soc Am A, Vol. 4, No. 8, August 1987, pp. 1555-1471, ISSN: 1084-7529

Jain, A. & Dubes, R. Algorithms for Clustering Data, Prentice Hall, New Jersey, ISBN: 0-13-022278-X

Kass, M.; Witkin, A. & Terzopoulos, D. (1988). Snakes: Active Contour Models, Int J Comput Vis, Vol. 55, No. 4, January 1988, pp. 321-331, ISSN: 0920-5691

Kervrann, C. & Heitz, F. (1998). A Hierarchical Markov Modeling Approach for the Segmentation and Tracking of Deformable Shapes, Graph Model Image Process, Vol. 60, No. 3, May 1998, pp. 173-195, ISSN 1077-3169

Kovesi, P.D. (1996). Invariant Measures of Image Features from Phase Information, PhD. Thesis, The University or Western Australia, May 1996, URL: http://www.cs.uwa.edu.au/ pub/robvis/theses/PeterKovesi/

Mansouri, A.-J. & Konrad, J. (2003). Multiple Motion Segmentation with Level Sets, IEEE Trans Image Process, Vol. 12, No. 2, February 2003, pp. 201-220, ISSN: 1057-7149

Montoliu, R. & Pla, F. (2005). An Iterative Region-Growing Algorithm for Motion Segmentation and Estimation, Int J Intell Syst, Vol. 20, No. 5, May 2005, pp. 577-590, ISSN: 0884-8173

Morrone, M.C. & Owens, R.A. (1987). Feature Detection from Local Energy, Pattern Recognition Letters, Vol. 6, No. 5, December 1987, pp. 303-313, ISSN: 0167-8655

Nestares, O.; Miravet, C.; Santamaria, J. & Navarro, R. (2000). Automatic enhancement of noisy image sequences through localspatiotemporal spectrum analysis, Optical Engineering, Vol. 39, No. 6, June 2000, pp. 1457-1469, ISSN: 0091-3286

Nguyen, H.T. & Smeulders, A.W.M. (2004). Fast Occluded Object Tracking by a Robust Appearance Filter, IEEE Trans Pattern Anal Mach Intell, Vol. 26, No. 8, August 2004, pp. 1099-1104, ISSN: 0162-8828

Oppenheim, A. & Lim, J. (1981). The Importance of Phase in Signals, Proceedings of the IEEE, Vol. 69, No. 5, May 1981, pp. 529–541, ISSN: 0018-9219

Pal, N.R. & Biswas, J. (1996). Cluster Validation Using graph Theoretic Concepts, Pattern Recognition, Vol. 30, No. 6, June 1996, pp. 847-857, ISSN: 0031-3203

Paragios, N. & Deriche, R. (2000). Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects, IEEE Trans Pattern Anal Mach Intell, Vol. 22, No. 3, March 2000, pp. 266-279, ISSN: 1057-7149

Rodríguez-Sánchez, R.; García, J.A.; Fdez-Valdivia, J. & Fdez-Vidal, X.R. (1999). The RGFF Representational Model: A System for the Automatically Learned Partition of "Visual Patterns" in Digital Images, IEEE Trans Pattern Anal Mach Intell, Vol. 21, No. 10, October 1999, pp. 1044-1073, ISSN: 1057-7149

Ross, J.; Morrone, M.C. & Burr, D. (1989). The Conditions under which Mach Bands are Visible, Vision Research, Vol. 29, No. 6, 1989, pp. 699–715, ISSN: 0042-6989

Sato, K. & Aggarwal, J.K. (2004). Temporal Spatio-Temporal Transform and its Application to Tracking and Interaction, Comput Vis Image Understand, Vol. 96, No. 2, November 2004, pp. 100-128, ISSN: 1077-3142

Simoncelli, E.P. & Adelson, E.H. (1991). Computing Optical Flow Distributions using Spatio-Temporal Filters, MIT Media Lab. Vision and Modeling, Tech. Report No. 165, 1991,
http://web.mit.edu/persci/people/adelson/pub_pdfs/simoncelli_comput.pdf

Stiller, C. & Konrad, J. (1999). Estimating Motion in Image Sequences: A Tutorial on Modeling and Computation of 2D Motion, IEEE Signal Processing Magazine. Vol. 16, No. 6, July 1999, pp. 71-91, ISSN: 1053-5888

Tsechpenakis, G.; Rapantzikos, K.; Tsapatsoulis, N. & Kollias, S. (2004). A Snake Model for Object Tracking in Natural Sequences, Signal Process Image Comm, Vol. 19, No. 3, March 2004, pp. 219-238, ISSN: 0923-5965

Venkatesh, S. & Owens, R. (1990). On the Classification of Image Features, Pattern Recognition Letters, Vol. 11, No. 5, May 1990, pp. 339-349, ISSN: 0167-8655

Wang, J.Y.A. & Adelson, E.H. (1994). Representing Moving Images with Layers, IEEE Trans Pattern Anal Mach Intell, Vol. 3, No. 5, September 1994, pp. 325-638, ISSN: 1057-7149

Watson, A.B. & Ahumada Jr., A.J. (1985). Model for Human Visual-Motion Sensing, J Opt Soc Am A, Vol. 2, No. 2, February 1985, pp. 322-342, ISSN: 1084-7529

Weickert, J. & Kühne, G. (2003). Fast Methods for Implicit Active Contour Models, In: Geometric Level Set Methods in Imaging, Vision and Graphics, pp. 43-58, Osher, S. & Paragios, N. (Eds.), Springer, ISBN: 0-387-95488-0, New York



Vision Systems: Segmentation and Pattern Recognition
Edited by Goro Obinata and Ashish Dutta

ISBN 978-3-902613-05-9
Hard cover, 536 pages
Publisher I-Tech Education and Publishing
Published online 01, June, 2007
Published in print edition June, 2007

Research in computer vision has exponentially increased in the last two decades due to the availability of cheap cameras and fast processors. This increase has also been accompanied by a blurring of the boundaries between the different applications of vision, making it truly interdisciplinary. In this book we have attempted to put together state-of-the-art research and developments in segmentation and pattern recognition. The first nine chapters on segmentation deal with advanced algorithms and models, and various applications of segmentation in robot path planning, human face tracking, etc. The later chapters are devoted to pattern recognition and covers diverse topics ranging from biological image analysis, remote sensing, text recognition, advanced filter design for data analysis, etc.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Raquel Dosil, Xose R. Fdez-Vidal, Xose M. Pardo and Anton Garcia (2007). Energy Feature Integration for Motion Segmentation, Vision Systems: Segmentation and Pattern Recognition, Goro Obinata and Ashish Dutta (Ed.), ISBN: 978-3-902613-05-9, InTech, Available from: http://www.intechopen.com/books/vision_systems_segmentation_and_pattern_recognition/energy_feature_integration_for_motion_segmentation

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2007 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen