A 3-Step Algorithm Using Region-Based Active Contours for Video Objects Detection

Stéphanie Jehan-Besson

Laboratoire I3S, CNRS UNSA, 2000 route des Lucioles, 06903 Sophia Antipolis, France
Email: jehan@i3s.unice.fr

Michel Barlaud

Laboratoire I3S, CNRS UNSA, 2000 route des Lucioles, 06903 Sophia Antipolis, France
Email: barlaud@i3s.unice.fr

Gilles Aubert

Laboratoire J. A. Dieudonné, CNRS, Parc Valrose, 06108 Nice Cedex 2, France
Email: gaubert@math.unice.fr

We propose a 3-step algorithm for the automatic detection of moving objects in video sequences using region-based active contours. First, we introduce a very full general framework for region-based active contours with a new Eulerian method to compute the evolution equation of the active contour from a criterion including both region-based and boundary-based terms. This framework can be easily adapted to various applications, thanks to the introduction of functions named *descriptors* of the different regions. With this new Eulerian method based on shape optimization principles, we can easily take into account the case of descriptors depending upon features globally attached to the regions. Second, we propose a 3-step algorithm for detection of moving objects, with a static or a mobile camera, using region-based active contours. The basic idea is to hierarchically associate temporal and spatial information. The active contour evolves with successively three sets of descriptors: a temporal one, and then two spatial ones. The third spatial descriptor takes advantage of the segmentation of the image in intensity homogeneous regions. User interaction is reduced to the choice of a few parameters at the beginning of the process. Some experimental results are supplied.

Keywords and phrases: detection of moving objects, video sequences analysis, region-based active contours, level-set method.

1. INTRODUCTION

Automatic segmentation of video objects is still a matter of intensive research and is a crucial issue for the development of the new video coding standards MPEG-4 [1, 2] and MPEG-7 [3]. Indeed, object-based coding attracts now a considerable attention. Actually, it will open new areas of applications in the field of video production and multimedia interface.

Generally speaking, two classes of approaches may be considered for segmentation: region-based approaches such as Markov random fields [4], and boundary-based approaches. Originally, active contours were boundary-based methods. Indeed, snakes [5], balloons [6], or geodesic active contours [7] are driven by the minimization of energy towards the edges of an image. However, for video objects detection, it is interesting to incorporate region-based information in the evolution equation of the active contour.

In this context, we propose a generalized Eulerian framework for region-based active contours. This general framework is then applied to the detection of moving objects from video sequences acquired with either a moving or a static camera.

First, a general framework for region-based active contours is introduced. Starting from a criterion including both region-based and boundary-based terms, we build a new efficient method, based on shape optimization principles, to compute the evolution equation of an active contour. This method ensures the fastest decrease of the active contour towards a minimum of the criterion. In this criterion, each region is described by a function named *descriptor*. For a particular application, the user only has to choose well-adapted descriptors.

The second contribution of this paper is to propose a 3-step algorithm for detection of moving objects using the previous general framework. In this algorithm, the active contour evolves with successively three sets of descriptors. The first set is motion-based while the two others use spatial information, namely, edges and partition of the image into

intensity homogeneous regions. The first stage enables to detect moving objects while the two others refine the result. In our method, the user does not have to select the object. User interaction is thus reduced to the choice of a few parameters at the beginning of the segmentation process. As far as sequences with mobile cameras are concerned, camera motion compensation and segmentation are jointly performed. The camera motion model is directly included in the criterion to minimize. It is estimated from frame to frame using potential functions. We take advantage of the half quadratic theorem [8] for robust estimation with the strategy based on alternate minimizations [9].

The general framework for region-based active contours is detailed in Section 2, the 3-step algorithm for video objects detection is explained in Section 3. Section 4 illustrates the potential of our approach by applying it to real sequences.

2. REGION-BASED ACTIVE CONTOURS

The main idea of this part lies in the development of a general framework for the segmentation of an image in two different regions using region-based active contours. We want to find the image partition that minimizes a criterion including both region-based and boundary-based terms. In this framework, each region is described using a function that we name *descriptor of the region*. The introduction of such descriptors is interesting for two reasons. First, for a given application like detection of moving objects, various descriptors can be easily tested inside the same theoretical framework or hierarchically combined as in Section 3.

Some authors [10, 11, 12, 13, 14, 15, 16, 17] have proposed a way of adding region-based terms in the evolution equation of an active contour. These pioneer works are complementary and show the potential of region-based approaches. Some works suggest to choose the displacement of the active contour that will make the criterion decrease at each step. Some other works propose to compute the evolution equation of the active contour using Euler-Lagrange equations. With this last method of computation, the case of descriptors depending upon features globally attached to the region (region-dependent descriptors as the mean or the variance of a region) cannot readily be taken into account.

In this paper, we introduce a general framework and propose a new Eulerian method for the computation of the evolution equation of the active contour. The velocity vector of the active contour is computed by introducing a dynamical scheme directly in the criterion to minimize as in [18, 19]. However, we propose here a new formulation of the derivation using shape optimization tools. We compute the theoretical expression of the velocity vector that will make the active contour evolve as fast as possible towards a minimum of a criterion including both region-based and boundary-based terms. Moreover, this theoretical expression is computed for both region-dependent and region-independent descriptors. In the case of region-dependent descriptors, some additional terms

appear in the evolution equation that have not been computed so far.

2.1. Introduction of a general criterion

Let I_n be the intensity of image number n in the sequence and Ω_n the image domain of frame number n. The image domain is considered to be made up of two parts: the foreground part, containing the objects to segment, denoted by $\Omega_{n,\text{in}}$, and the background part denoted by $\Omega_{n,\text{out}}$. The discontinuities set is a curve denoted by Γ_n that defines the boundary between the two domains.

We search for the two domains $\Omega_{n,\text{in}}$ and $\Omega_{n,\text{out}}$ which minimize the following criterion J_n:

$$
\begin{aligned}
J_n&(\Omega_{n,\text{out}}, \Omega_{n,\text{in}}, \Gamma_n) \\
&= \iint_{\Omega_{n,\text{out}}} k^{(n,\text{out})}(\Omega_{n,\text{out}}, x, y)\, dx\, dy \\
&\quad + \iint_{\Omega_{n,\text{in}}} k^{(n,\text{in})}(\Omega_{n,\text{in}}, x, y)\, dx\, dy + \int_{\Gamma_n} k^{(n,b)}(x, y)\, ds.
\end{aligned}
\tag{1}
$$

The first two terms are region-based while the third term is boundary-based. The functions $k^{(n,\text{out})}(\cdot)$, $k^{(n,\text{in})}(\cdot)$, and $k^{(n,b)}(\cdot)$ are, respectively, called descriptor of the background region, descriptor of the objects, and descriptor of the boundary. The region descriptors may be globally attached to their respective regions, $\Omega_{n,\text{in}}$ or $\Omega_{n,\text{out}}$, and so depend on them. It happens when statistical features of a region like, for example, the mean or the variance, are selected as descriptors.

To compute an optimal solution, a dynamical scheme is introduced where the unknown regions become a function of an evolving parameter τ:

$$
\begin{aligned}
J_n&(\Omega_{n,\text{out}}(\tau), \Omega_{n,\text{in}}(\tau), \Gamma_n(\tau)) \\
&= \iint_{\Omega_{n,\text{out}}(\tau)} k^{(n,\text{out})}(\Omega_{n,\text{out}}(\tau), x, y)\, dx\, dy \\
&\quad + \iint_{\Omega_{n,\text{in}}(\tau)} k^{(n,\text{in})}(\Omega_{n,\text{in}}(\tau), x, y)\, dx\, dy \\
&\quad + \int_{\Gamma_n(\tau)} k^{(n,b)}(x, y)\, ds.
\end{aligned}
\tag{2}
$$

The functional $J_n(\Omega_{n,\text{out}}(\tau), \Omega_{n,\text{in}}(\tau), \Gamma(\tau))$ is thereafter denoted by $J_n(\tau)$, and the descriptors $k^{(n,\text{out})}(\Omega_{n,\text{out}}(\tau), x, y)$ and $k^{(n,\text{in})}(\Omega_{n,\text{in}}(\tau), x, y)$ are, respectively, denoted by $k^{(n,\text{out})}(\tau, x, y)$ and $k^{(n,\text{in})}(\tau, x, y)$, which gives

$$
\begin{aligned}
J_n(\tau) &= \iint_{\Omega_{n,\text{out}}(\tau)} k^{(n,\text{out})}(\tau, x, y)\, dx\, dy \\
&\quad + \iint_{\Omega_{n,\text{in}}(\tau)} k^{(n,\text{in})}(\tau, x, y)\, dx\, dy \\
&\quad + \int_{\Gamma_n(\tau)} k^{(n,b)}(x, y)\, ds.
\end{aligned}
\tag{3}
$$

Here $\Gamma_n(\tau)$ is modeled as an active contour that converges towards the final expected segmentation. Let Γ_{n_0} be the initial

curve, recall that we search for $\Gamma_n(\tau)$ as a curve evolving according to the following partial differential equation (PDE):

$$\frac{\partial \Gamma_n(\tau)}{\partial \tau} = \boldsymbol{v}_n, \tag{4}$$

where $\boldsymbol{v}_n$ is the velocity of the active contour for frame number n. The main problem lies in finding the velocity $\boldsymbol{v}_n$ from criterion (3) to get the fastest curve evolution towards the final segmentation.

2.2. Computation of the derivative

In order to obtain the evolution equation, the criterion $J_n(\tau)$ must be differentiated with respect to τ. Define the functional $k_n(x, y, \tau)$ such that

$$k_n(x, y, \tau) = \begin{cases} k^{(n,\mathrm{out})}(x, y, \tau), & \text{if } (x, y) \in \Omega_{n,\mathrm{out}}(\tau), \\ k^{(n,\mathrm{in})}(x, y, \tau), & \text{if } (x, y) \in \Omega_{n,\mathrm{in}}(\tau). \end{cases} \tag{5}$$

The criterion $J_n(\tau)$ writes as

$$\begin{aligned} J_n(\tau) &= \iint_{\Omega_n} k_n(x, y, \tau) dx\, dy + \int_{\Gamma_n(\tau)} k^{(n,b)}(x, y) ds \\ &= J_1(\tau) + J_2(\tau), \end{aligned} \tag{6}$$

where the image domain $\Omega_n = \Omega_{n,\mathrm{out}}(\tau) \cup \Omega_{n,\mathrm{in}}(\tau) \cup \Gamma_n(\tau)$ and

$$\begin{aligned} J_1(\tau) &= \iint_{\Omega_n} k_n(x, y, \tau) dx\, dy, \\ J_2(\tau) &= \int_{\Gamma_n(\tau)} k^{(n,b)}(x, y) ds. \end{aligned} \tag{7}$$

In order to compute the derivative of the criterion J_1, discontinuities must explicitly be taken into account. We first remind a general theorem concerning the Eulerian derivative of a time-dependent criterion. We define $\Omega(\tau)$ as a region included into Ω_n.

Theorem 1. *Let k be a smooth function on $\overline{\Omega} \times (0, T)$, and let $J(\tau) = \iint_{\Omega(\tau)} k(x, y, \tau) dx\, dy$, then*

$$\frac{dJ}{d\tau} = J'(\tau) = \iint_{\Omega(\tau)} \frac{\partial k}{\partial \tau} dx\, dy - \int_{\partial\Omega(\tau)} k(\boldsymbol{w} \cdot \boldsymbol{N}_{\partial\Omega}) ds, \tag{8}$$

where $\boldsymbol{w}$ is the velocity of $\partial\Omega(\tau)$ and $\boldsymbol{N}_{\partial\Omega}$ is the unit inward normal to $\partial\Omega(\tau)$.

For details of the proof see [20]. The variation of J is due to the variation of the functional $k(x, y, \tau)$ and to the motion of the integral domain $\Omega(\tau)$. As a corollary of Theorem 1, we get the following.

Corollary 2. *Suppose that the domain $\Omega(\tau)$ is made up of two parts, $\Omega_{\mathrm{in}}(\tau)$ and $\Omega_{\mathrm{out}}(\tau)$ separated by a moving interface $\Gamma(\tau)$ whose velocity is $\boldsymbol{v}$. The function $k(x, y, \tau)$ is supposed to be separately continuous in $\Omega_{\mathrm{in}}(\tau)$ and $\Omega_{\mathrm{out}}(\tau)$ but can be discontinuous across $\Gamma(\tau)$. Denote by $k^{(\mathrm{in})}$ and $k^{(\mathrm{out})}$ the value of k in*

$\Omega_{\mathrm{in}}(\tau)$ *and* $\Omega_{\mathrm{out}}(\tau)$, *respectively. Thus the derivative of $J(\tau)$ is written as*

$$\begin{aligned} J'(\tau) = &\iint_{\Omega(\tau)} \frac{\partial k}{\partial \tau} dx\, dy - \int_{\partial\Omega(\tau)} k(\boldsymbol{w} \cdot \boldsymbol{N}_{\partial\Omega}) ds \\ &+ \int_{\Gamma(\tau)} [[k]](\boldsymbol{v} \cdot \boldsymbol{N}) ds, \end{aligned} \tag{9}$$

where $[[k]]$ represents the jump of k across $\Gamma(\tau)$, $[[k]] = k^{(\mathrm{out})} - k^{(\mathrm{in})}$, and $\boldsymbol{N}$ the unit normal to $\Gamma(\tau)$ directed from $\Omega_{\mathrm{out}}(\tau)$ to $\Omega_{\mathrm{in}}(\tau)$.

It is now straightforward to get the derivative of the criterion $J_1(\tau)$: we take $\Omega(\tau) = \Omega_n$ the image domain, $k = k_n$, $\Omega_{\mathrm{out}}(\tau) = \Omega_{n,\mathrm{out}}(\tau)$, $\Omega_{\mathrm{in}}(\tau) = \Omega_{n,\mathrm{in}}(\tau)$, and $\Gamma(\tau) = \Gamma_n(\tau)$. The boundary of the image domain Ω_n is fixed and so, by explicitly taking the discontinuities into account, thanks to the corollary, we get the following expression:

$$\begin{aligned} J_1'(\tau) = &\iint_{\Omega_{n,\mathrm{in}}(\tau)} \frac{\partial k^{(n,\mathrm{in})}}{\partial \tau} dx\, dy + \iint_{\Omega_{n,\mathrm{out}}(\tau)} \frac{\partial k^{(n,\mathrm{out})}}{\partial \tau} dx\, dy \\ &+ \int_{\Gamma_n(\tau)} \left(k^{(n,\mathrm{out})} - k^{(n,\mathrm{in})}\right)(\boldsymbol{v}_n \cdot \boldsymbol{N}) ds. \end{aligned} \tag{10}$$

The derivative of J_2 is classical [7] and so, the derivative of the whole criterion is the following:

$$\begin{aligned} J_n'(\tau) = &\iint_{\Omega_{n,\mathrm{in}}(\tau)} \frac{\partial k^{(n,\mathrm{in})}}{\partial \tau} dx\, dy + \iint_{\Omega_{n,\mathrm{out}}(\tau)} \frac{\partial k^{(n,\mathrm{out})}}{\partial \tau} dx\, dy \\ &+ \int_{\Gamma_n(\tau)} \left(-k^{(n,b)} \cdot \kappa + \nabla k^{(n,b)} \cdot \boldsymbol{N}\right)(\boldsymbol{v}_n \cdot \boldsymbol{N}) ds \\ &+ \int_{\Gamma_n(\tau)} \left(k^{(n,\mathrm{out})} - k^{(n,\mathrm{in})}\right)(\boldsymbol{v}_n \cdot \boldsymbol{N}) ds, \end{aligned} \tag{11}$$

where $\kappa_n(x, y, \tau)$ is the curvature of $\Gamma_n(x, y, \tau)$.

For descriptors that do not depend on τ, the first two integrals are null. In this article, we express the derivative in the general case where descriptors may depend on τ. We take $k^{(n,\mathrm{in})}$ and $k^{(n,\mathrm{out})}$ as a combination of functions depending upon features globally attached to the regions. We have

$$\begin{aligned} k^{(n,\mathrm{in})}(x, y, \tau) &= g^{(\mathrm{in})}\left(x, y, G_1^{(\mathrm{in})}(\tau), \ldots, G_p^{(\mathrm{in})}(\tau)\right), \\ k^{(n,\mathrm{out})}(x, y, \tau) &= g^{(\mathrm{out})}\left(x, y, G_1^{(\mathrm{out})}(\tau), \ldots, G_m^{(\mathrm{out})}(\tau)\right), \end{aligned} \tag{12}$$

where

$$G_j^{(\cdot)}(\tau) = \iint_{\Omega_{n,\cdot}(\tau)} \psi_j^{(\cdot)}(x, y, \tau) dx\, dy \tag{13}$$

with $(\cdot) = (\mathrm{in})$ or (out). As an example, if the descriptor $k^{(n,\mathrm{in})}$ is chosen to be the mean of the region, we have

$$k^{(n,\mathrm{in})} = \frac{\iint_{\Omega_{n,\mathrm{in}}(\tau)} I_n(x, y) dx\, dy}{\iint_{\Omega_{n,\mathrm{in}}(\tau)} dx\, dy}. \tag{14}$$

And so it can be expressed as a combination of features depending on the region as

$$k^{(n,\text{in})}(x, y, \tau) = g^{(\text{in})}\left(x, y, G_1^{(\text{in})}(\tau), G_2^{(\text{in})}(\tau)\right), \qquad (15)$$

where

$$g^{(\text{in})} = \frac{G_1^{(\text{in})}(\tau)}{G_2^{(\text{in})}(\tau)},$$

$$G_1^{(\text{in})}(\tau) = \iint_{\Omega_{n,\text{in}}(\tau)} \psi_1^{(\text{in})}, \quad \psi_1^{(\text{in})}(x, y) = I_n(x, y), \qquad (16)$$

$$G_2^{(\text{in})}(\tau) = \iint_{\Omega_{n,\text{in}}(\tau)} \psi_2^{(\text{in})}, \quad \psi_2^{(\text{in})}(x, y) = 1.$$

Now we first compute the derivative of $k^{(n,\text{in})}$ in the general case using Theorem 1, we obtain that

$$\iint_{\Omega_{n,\text{in}}(\tau)} \frac{\partial k^{(n,\text{in})}}{\partial \tau} dx\, dy = \sum_{j=1}^{p} A_j^{(\text{in})} \iint_{\Omega_{n,\text{in}}(\tau)} \frac{\partial \psi_j^{(\text{in})}}{\partial \tau} dx\, dy$$
$$- \sum_{j=1}^{p} A_j^{(\text{in})} \int_{\Gamma_n(\tau)} \psi_j^{(\text{in})}(\boldsymbol{v}_n \cdot \boldsymbol{N}) ds, \qquad (17)$$

where

$$A_j^{(\text{in})} = \iint_{\Omega_{n,\text{in}}(\tau)} \frac{\partial g^{(\text{in})}}{\partial G_j^{(\text{in})}}\left(x, y, G_1^{(\text{in})}(\tau), \ldots, G_p^{(\text{in})}(\tau)\right) dx\, dy. \qquad (18)$$

The expression for $(\cdot) = (\text{out})$ is computed in the same manner while paying attention to the direction of the normal vector, and we thus obtain the general expression for the derivative of $J(\tau)$

$$J_n'(\tau) = \int_{\Gamma_n(\tau)} \left(k^{(n,\text{out})} - k^{(n,\text{in})}\right)(\boldsymbol{v}_n \cdot \boldsymbol{N}) ds$$
$$+ \int_{\Gamma_n(\tau)} \left(-k^{(n,b)} \cdot \kappa + \nabla k^{(n,b)} \cdot \boldsymbol{N}\right)(\boldsymbol{v}_n \cdot \boldsymbol{N}) ds$$
$$- \sum_{j=1}^{p} A_j^{(\text{in})} \int_{\Gamma_n(\tau)} \psi_j^{(\text{in})}(\boldsymbol{v}_n \cdot \boldsymbol{N}) ds$$
$$+ \sum_{j=1}^{m} A_j^{(\text{out})} \int_{\Gamma_n(\tau)} \psi_j^{(\text{out})}(\boldsymbol{v}_n \cdot \boldsymbol{N}) ds \qquad (19)$$
$$+ \sum_{j=1}^{p} A_j^{(\text{in})} \iint_{\Omega_{n,\text{in}}(\tau)} \frac{\partial \psi_j^{(\text{in})}}{\partial \tau} dx\, dy$$
$$+ \sum_{j=1}^{m} A_j^{(\text{out})} \iint_{\Omega_{n,\text{out}}(\tau)} \frac{\partial \psi_j^{(\text{out})}}{\partial \tau} dx\, dy.$$

2.3. From the derivative towards the evolution equation of the active contour

The goal of this part is to compute the expression of the velocity vector $\boldsymbol{v}_n$, which makes the curve evolve as fast as possible towards a minimum of the criterion J_n. In order to

deduce the velocity vector of the active contour from the derivative of the criterion, we have to make the velocity vector appear in the last two domain integrals by expressing them as boundary integrals. This can easily be done using Theorem 1, since domain integrals can always be expressed as boundary integrals. However, for simplicity, consider the cases where these the last two integrals are equal to zero, which happens for most region-dependent descriptors [20]. In that case, according to the inequality of Cauchy-Schwartz, the fastest decrease of $J_n(\tau)$ is obtained from (19) by choosing $\boldsymbol{v}_n = F_n \boldsymbol{N}$ which leads to the following evolution equation:

$$\frac{\partial \Gamma_n(\tau)}{\partial \tau} = \left[k^{(n,\text{in})} - k^{(n,\text{out})} + k^{(n,b)} \cdot \kappa - \nabla k^{(n,b)} \cdot \boldsymbol{N}\right. \\ \left. + \sum_{j=1}^{p} A_j^{(\text{in})} \psi_j^{(\text{in})} - \sum_{j=1}^{m} A_j^{(\text{out})} \psi_j^{(\text{out})}\right] \boldsymbol{N}. \qquad (20)$$

We may note that this expression is computed for different examples of region-dependent descriptors in [20, 21]. In that case, the additional terms have to be considered in order to get the fastest decrease of the active contour towards a minimum of the criterion.

If the descriptors do not depend on τ (region-independent descriptors), the PDE is then reduced to the following expression:

$$\frac{\partial \Gamma_n(\tau)}{\partial \tau} = \left[k^{(n,\text{in})} - k^{(n,\text{out})} + k^{(n,b)} \cdot \kappa - \nabla k^{(n,b)} \cdot \boldsymbol{N}\right] \boldsymbol{N}. \quad (21)$$

2.4. Implementation with the level-set method

In a video sequence, several objects may appear in a scene. So segmentation of video objects requires a method where topological changes are well handled in order to detect several objects from the same initial curve. On that account we use the level-set method, proposed by Osher and Sethian [22, 23], in order to implement the PDE (20). The key idea is to introduce an auxiliary image $U_n(x, y, \tau)$ such that $\Gamma_n(\tau)$ is the zero level set of U_n, $\Gamma_n(\tau) = \{(x, y)/U_n(x, y, \tau) = 0\}$. The function U_n is often chosen to be the signed distance function of Γ_n, and is then characterized by $|\nabla U_n| = 1$. This Eulerian formulation presents several advantages [23]. Firstly, the curve $\Gamma_n(\tau)$ may break or merge as the function U_n evolves, and topological changes are thus easily handled. Secondly, the function $U_n(x, y, \tau)$ always remains a function allowing efficient numerical schemes. Thirdly, the geometric properties of the curve, like the curvature and the normal vector field $\boldsymbol{N}$, can be estimated directly from the level-set function:

$$\kappa_n = \text{div}\left(\frac{\nabla U_n}{|\nabla U_n|}\right), \qquad \boldsymbol{N} = \frac{\nabla U_n}{|\nabla U_n|}. \qquad (22)$$

The evolution equation then becomes

$$\frac{\partial U_n(\tau)}{\partial \tau} = F_n |\nabla U_n|. \qquad (23)$$

The force F_n is computed only on the curve $\Gamma_n(\tau)$ but we can extend its expression to the whole image domain Ω_n. To

implement the level-set method, solutions must be found to circumvent problems coming from the fact that the signed distance function U_n is not a solution of the PDE (23) (see [24]). In this work, the function U_n is reinitialized so that it remains a distance function, see [25] for details.

3.   VIDEO OBJECT SEGMENTATION WITH A 3-STEP ALGORITHM

The main goal of this part is to describe an algorithm to segment moving objects in video sequences without any user interaction during the segmentation. The user only has to choose a set of parameters at the beginning of the segmentation process.

Our 3-step algorithm successively operates a motion-based segmentation and two spatial-based segmentations, all of them using region-based active contours. Three sets of time-independent descriptors $\{k^{(n,\mathrm{out})}(\cdot), k^{(n,\mathrm{in})}(\cdot), k^{(n,b)}(\cdot)\}$ are used for each segmentation step. Obviously, the first set is motion-based while the two others are spatial-based.

As far as the motion-based step is concerned, we propose two options. The first one (a) is dedicated to video sequences with a static camera whereas the second one (b) is dedicated to sequences with a mobile camera.

The algorithm does not require initial object selection by the user. Indeed the initial contour is choosen to be a rectangle near the borders of the first image (see Figure 1). The contour is then driven by the first set of motion-based descriptors, which allows us to detect moving objects more or less precisely. We then use the final contour of this first detection as an initial curve for the second step. The second stage drives the active contour towards the nearest edges of the image. The resulting contour is then taken as initial conditions for the third step which refines the detection by using segmentation of the image in intensity homogeneous regions. For each step, the active contour is driven by the PDE (21) by replacing the descriptors by their appropriate values.

Note that the color space used is (Y, C_b, C_r). The luminance Y of the image I_n is designated by $I_n(x, y, Y)$ while the two chrominances C_b and C_r are designated by $I_n(x, y, C_b)$ and $I_n(x, y, C_r)$, respectively. In this article $I_n(x, y)$ designates $I_n(x, y, Y)$.

3.1.   First step

Option a: motion-based descriptors for a static camera

For a static camera, motion may be detected by comparing the current frame with a background frame B_n.

(a1) Computation of the background frame

In this paper, B_n is computed with a robust estimation on a group of frames including the current frame. This frame is not necessarily the real background of the sequence. We then search for the frame B_n which minimizes,

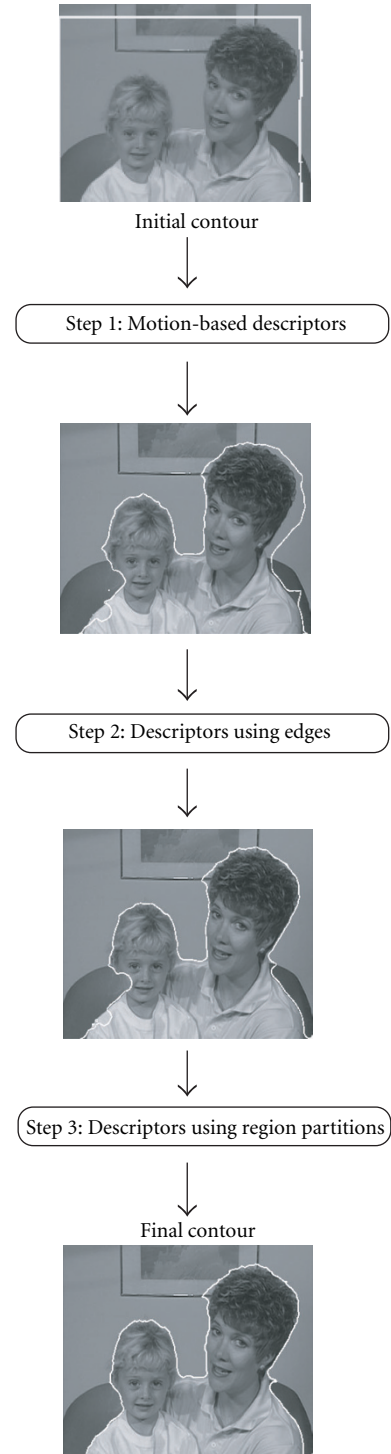$$\sum_{i\in[j,j+n_l]} \int_{\Omega_n} \varphi(|B_n - I_i|), \qquad (24)$$

FIGURE 1: *Mother and Daughter*, the 3-step algorithm.

where n_l is the number of frames chosen by the user to compute B_n and j is a number of frame chosen such that $n \in [j, j + n_l]$.

In the Bayesian framework, φ is known as the potential function and is introduced to eliminate outliers. Here, we

choose the Geman and McClure estimator [26]:

$$\varphi(t) = \frac{t^2}{1 + t^2}. \tag{25}$$

The minimization of (24) is done using the half quadratic theorem with the strategy based on alternate minimizations [9]. The background may also be computed as in [27].

(a2) Motion-based descriptors for a static camera

The motion-based descriptors are thus the following:

$$
\begin{aligned}
k^{(n,\text{out})} &= \sum_{\mathcal{V}} |B_n - I_n|, \\
k^{(n,\text{in})} &= \alpha_1, \\
k^{(n,b)} &= \lambda_1,
\end{aligned}
\tag{26}
$$

where α_1 and λ_1 are two positive constants. The term $\mathcal{V}$ designates a neighbourhood of (x, y).

Option b: motion-based descriptors for a mobile camera

For a mobile camera, the idea is to assume that the apparent motion of the background can be modeled by a 6-parameter affine motion model. These parameters are computed with a robust estimation using motion vectors evaluated by a classical block matching. The moving objects are supposed to be the pixels that are not compensated by the affine motion model. In this part, we first explain how the six parameters of the affine motion model are computed and then we detail the descriptors used.

(b1) Camera model

The camera motion can be modeled by a 6-parameter affine motion model, which is a good tradeoff between complexity and representativity [28]. So the apparent motion $w_n(p)$ of a point (x, y) of the static background, between frames I_{n-1} and I_n, is modeled by

$$w_n(p) = A_n p + t_n = \begin{bmatrix} a_{11}^n x + a_{12}^n y + t_1^n \\ a_{21}^n x + a_{22}^n y + t_2^n \end{bmatrix}, \tag{27}$$

with $p = (x, y)$.

We search for the six parameters of the camera model (27) which minimize the following criterion:

$$\sum_{p \in \Omega_n} \varphi(|u_n - A_n p - t_n|). \tag{28}$$

The function φ is introduced to eliminate outliers due to the motion of moving video objects. We choose the Geman and McClure estimator (25).

The motion field u_n is classically computed using a block matching algorithm between frames I_{n-1} and I_n.

In order to minimize criterion (28), we use the properties of half quadratic regularization with the strategy based on alternate minimizations [9]. The initial minimization problem

is in fact substituted for the equivalent problem:

$$(A_n, t_n) = \underset{(A_n, t_n)}{\operatorname{argmin}} \sum_{\Omega_n} b r^2, \tag{29}$$

where $r = |u_n - A_n p - t_n|$ and $b = \varphi'(r)/2r$.

The algorithm based on alternate minimizations is then the following, with k as the number of iteration:

Initialization(A_n^0, t_n^0)

Repeat

$$b^{k+1} = \frac{\varphi'(r^k)}{2r^k}$$

$$(A_n^{k+1}, t_n^{k+1}) = \underset{(A_n, t_n)}{\operatorname{argmin}} \sum_{\Omega_n} b^k (r^{k+1})^2$$

Until convergence

The minimization of $\sum_{\Omega} b^k (r^{k+1})^2$ is performed using a gradient descent method. We thus obtain the six parameters of the affine motion model.

At convergence of the algorithm, the image $b = \varphi'(r)/2r$ gives a representation of the outliers due to the motion of moving video objects.

(b2) Motion-based descriptors for a mobile camera

Descriptors are then the following:

$$
\begin{aligned}
k^{(n,\text{out})}(p) &= \sum_{p_i \in \mathcal{V}} |I_n(p_i) - \text{Proj}(I_{n-1})(p_i)|, \\
k^{(n,\text{in})}(p) &= \rho_1, \\
k^{(n,b)}(p) &= \lambda_1,
\end{aligned}
\tag{30}
$$

where λ_1 and ρ_1 are two positive constants.

The term $\text{Proj}(I_{n-1})$ designates the projection of the image I_{n-1} in the referential of image I_n:

$$\text{Proj}(I_{n-1})(p + w_n(p)) = I_{n-1}(p). \tag{31}$$

3.2. Second step: edge-based descriptors

Let Γ_n^1 be the final contour of step 1 (option a or b) and $\Omega_{n,\text{in}}^1$, $\Omega_{n,\text{out}}^1$ the two resulting domains. The region $\Omega_{n,\text{in}}^1$ contains pixels that are considered to belong to moving objects after step 1.

In this step, we want to make the active contour evolve towards the nearest edges in the image, and so, we choose the following descriptors:

$$k^{(n,\text{out})} = c(|\nabla I_n|), \qquad k^{(n,\text{in})} = \alpha_2, \qquad k^{(n,b)} = \lambda_2, \tag{32}$$

where λ_2 and α_2 are two positive constants.

The function c is defined as follows:

$$c(|\nabla I_n|) = \begin{cases} |\nabla I_n|, & \text{if } (x, y) \in \Omega_{n,\text{in}}^1, \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

This function allows to reach the nearest edge inside the first segmented region and not outside.

3.3. Third step: descriptors using regions partition

Let Γ_n^2 be the final contour of step 2, $\Omega_{n,\text{in}}^2$ contains the pixels of the image that are considered to belong to moving objects after step 2.

In this third step, we first make a partition of the image in intensity homogeneous regions using a region growing algorithm. We assume that a moving object is made with the union of several regions. We consider that if a region is mostly included in the final moving object of step 2, $\Omega_{n,\text{in}}^2$, then the active contour will be driven in order to include the whole region in the final moving object. On the contrary, if the region is mostly included in the background part, $\Omega_{n,\text{out}}^2$, then this region will be removed from the final segmentation.

3.3.1 Region growing method

Frames are segmented into intensity homogeneous regions. The pixel (x, y) belongs to the region R_i if it satisfies the following decision criterion:

$$
\begin{aligned}
\left| I_n(x, y, Y) - \mu_Y \right| &\leq \sigma, \\
\left| I_n(x, y, C_b) - \mu_{C_b} \right| &\leq \sigma, \\
\left| I_n(x, y, C_r) - \mu_{C_r} \right| &\leq \sigma,
\end{aligned}
\tag{34}
$$

where μ_Y, μ_{C_b}, μ_{C_r} are the average intensity values of the region R_i for, respectively, the luminance Y and the two chrominances C_b and C_r. The parameter σ is the variance, we take $\sigma = 8$.

We start from a pixel (x, y), and we check all the neighborhood points. Points that verify the criterion are inserted into the region. The region will expand until no more neighborhood points can be added. The algorithm may be improved by using much more efficient methods such as binary partition trees [29].

3.3.2 Descriptors

The decision criterion used is based on the percentage of pixels included in $\Omega_{n,\text{in}}^2$ as in [30] where it has been successfully applied. The new idea here is to apply this decision criterion directly in the criterion to minimize and so to embed the decision criterion in the velocity vector of the active contour. We call N_i the number of pixels of region R_i, and $N_{i,\text{seg}}$ the number of pixels of region $R_i \cap \Omega_{n,\text{in}}^2$, that is, the pixels of R_i that are inside the segmented region after step 2.

Descriptors are defined as follows:

$$
k^{(n,\text{out})} = d_{\text{out}}, \qquad k^{(n,\text{in})} = d_{\text{in}}, \qquad k^{(n,b)} = \lambda_3,
\tag{35}
$$

where λ_3 is a positive constant.

The functions d_{in} and d_{out} are chosen to be

$$
d_{\text{in}}(x, y) = \begin{cases} 0, & \text{if } \frac{\left| N_i - N_{i,\text{seg}} \right|}{N_i} \leq 0.02, \\ 1, & \text{otherwise}, \end{cases}
\tag{36}
$$

with $(x, y) \in R_i$,

$$
d_{\text{out}}(x, y) = \begin{cases} 1, & \text{if } \frac{\left| N_i - N_{i,\text{seg}} \right|}{N_i} \leq 0.02, \\ 0, & \text{otherwise}, \end{cases}
\tag{37}
$$

with $(x, y) \in R_i$.

3.4. Remarks

The parameters α_i and ρ_i interact as threshold parameters, see [31, 32] for detailed explanations. The parameter λ_i is a regularization parameter that smoothes the curve [33]. In order to reduce the number of parameters to choose, we take $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$.

The final contour of frame number n is used as an initial contour for frame number $n + 1$. The initial contour is reinitialized as a rectangle near the borders of every nb_{reinit} images, where nb_{reinit} is specified by the user (usually we take $4 \leq nb_{\text{reinit}} \leq 10$).

4. EXPERIMENTAL RESULTS

The algorithm has been tested on three real sequences from the research group COST 211 [34]: *Coastguard, Mother and Daughter,* and *Highway,* and on the sequence *Akiyo* from MPEG. For *Coastguard* we take $\rho_1 = 9, \lambda = 20$, and $\alpha_2 = 7000$; for *Mother and Daughter* we take $\alpha_1 = 2.5, \lambda = 15, n_l = 60$, and $\alpha_2 = 1000$; for *Highway* we take $\alpha_1 = 16, \lambda = 15, n_l = 30$, and $\alpha_2 = 1000$; and for *Akiyo* we take $\alpha_1 = 0.8, \lambda = 25$, $n_l = 20$, and $\alpha_2 = 500$.

In Figure 1, each step of the segmentation process is detailed. After the first stage, moving objects are roughly detected. The last two stages refine the result allowing an accurate detection of moving objects.

In Figures 2, 3, 4, and 5, the final active contour is shown with a white envelop. Moving objects are well detected in all sequences either with camera motion or not, which illustrates the potential of our approach.

4.1. Evaluation of segmentation results

In order to evaluate the quality of segmentation results, the European group COST 211 highlights some measurements based on the comparison to a reference segmentation. Three criteria are suggested [30], two for the evaluation of the spatial accuracy of the result and one for that of the temporal coherency of the shape along the sequence. The term VOP designates a video object plan as it has been defined in MPEG-4 [2]. The segmentation result for frame number n is denoted by A_n^{est} while the reference object is denoted by A_n^{ori}. We remind here these three criteria:

(1) Absolute VOP shape distorsion:

$$
d_{\text{abs}}\left(A_n^{\text{est}}, A_n^{\text{ori}} \right) = \frac{N_{\text{err}}}{N_{\text{frame}}}.
\tag{38}
$$

(2) Relative VOP shape distorsion:

$$
d_{\text{rel}}\left(A_n^{\text{est}}, A_n^{\text{ori}} \right) = \frac{N_{\text{err}}}{N_{\text{ori}}}.
\tag{39}
$$

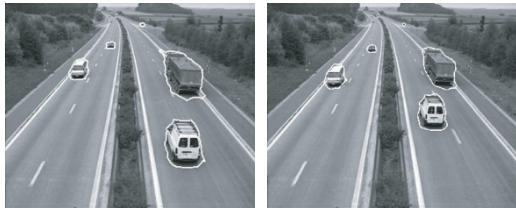FIGURE 2: *Mother and Daughter:* final contours in white and extracted video objects.

FIGURE 3: *Highway:* final contours in white.

(3) Temporal coherency:

$$T^{\text{est}}(n) = d_{\text{rel}}\left(A_n^{\text{est}}, A_{n-1}^{\text{est}}\right), \tag{40}$$

where N_{err} designates the number of deviating pixels of the segmented object compared to the reference mask, N_{frame} designates the number of pixels of a frame, and N_{ori} the number of pixels of the reference object. The temporal coherency $T^{\text{est}}(n)$ measures the average number of deviating pixels of the segmented object between two frames leading to an activity measure. This value may be compared to the activity measure of the reference object $T^{\text{ori}}(n)$.
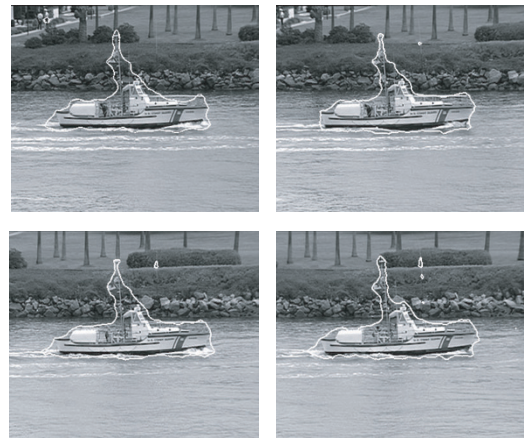

FIGURE 4: *Akiyo:* final contours in white.

FIGURE 5: *Coastguard:* final contours in white.

Reference segmentation results are available for sequences *Coastguard* and *Akiyo*. The three criteria have been computed for these two sequences. The results are given in Figures 6 and 7.

In Figure 6, we can remark that the distorsion results for *Akiyo* are very low, proving the accuracy of the segmentation results. The temporal coherency results show a low activity measure between two frames since the object does not move a lot from frame to frame. The object is well tracked over the sequence.

Coastguard has been acquired by a mobile camera and so segmentation is much more difficult since the moving object has to be extracted from a moving background. Moreover, the reference segmentation takes into account a large part of the moving water behind the boat while our algorithm detects only the water under the boat, as it can be seen in Figure 5. Despite all these difficulties, distorsion results in Figure 7 show a good accuracy of the segmentation and temporal coherency results show that the boat is well tracked over the sequence without being lost.
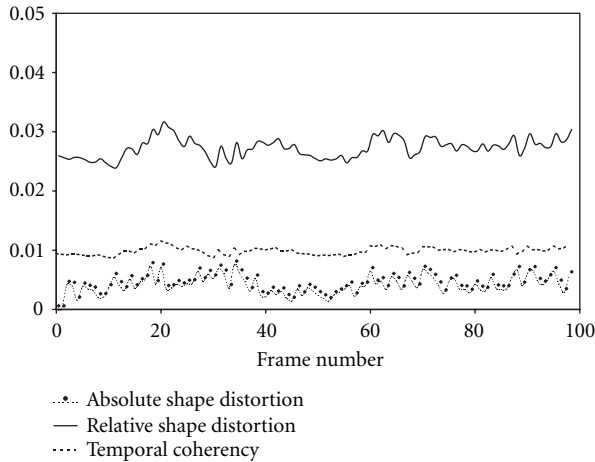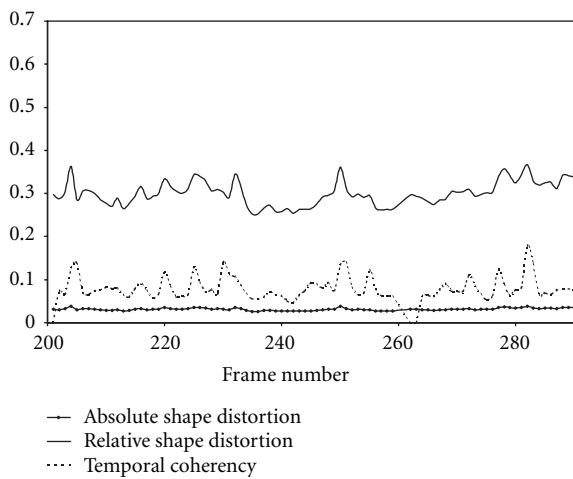
FIGURE 6: Evaluation results for *Akiyo*.

FIGURE 7: Evaluation results for *Coastguard*.

5. CONCLUSION

We propose a 3-step algorithm to segment moving objects using region-based active contours. Three hierarchical stages make the initial active contour evolve towards moving objects. The first step takes advantage of motion information, while the two others take advantage of spatial information, namely, edges and intensity homogeneous regions of the image. This algorithm can be easily extended by adding complementary steps and information as, for example, the temporal coherency of the video object.

REFERENCES

[1] R. Koenen, "MPEG-4: multimedia for our time," *IEEE Spectrum*, vol. 36, no. 2, pp. 26–33, 1999.

[2] International Organization for Standardization, "Overview of the MPEG-4 standard," ISO/IEC JTC1/SC29/WG11 N2725, March 1999.

[3] International Organization for Standardization, "MPEG-7: context and objectives," Atlantic City, NJ, USA, October 1998.

[4] J. Odobez and P. Bouthemy, "Detection of multiple moving objects using multiscale MRF with camera motion compensation," in *Proc. 1st IEEE International Conference on Image Processing*, pp. 257–261, Austin, Tex, USA, November 1994.

[5] M. Kass, A. Witkin, and D. Terzopoulus, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[6] L. Cohen, "On active contour models and balloons," *Computer Vision, Graphics, and Image Processing. Image Understanding*, vol. 53, no. 2, pp. 211–218, 1991.

[7] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.

[8] S. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, 1992.

[9] P. Charbonnier, G. Blanc-Féraud, L. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.

[10] L. Cohen, E. Bardinet, and N. Ayache, "Surface reconstruction using active contour models," in *Proc. SPIE '93 Conference on Geometric Methods in Computer Vision*, pp. 257–261, San Diego, Calif, USA, July 1993.

[11] R. Ronfard, "Region-based strategies for active contour models," *International Journal of Computer Vision*, vol. 13, no. 2, pp. 229–251, 1994.

[12] S. Zhu and A. L. Yuille, "Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, 1996.

[13] A. Chakraborty, L. Staib, and J. Duncan, "Deformable boundary finding in medical images by integrating gradient and region information," *IEEE Trans. on Medical Imaging*, vol. 15, no. 6, pp. 859–870, 1996.

[14] A.-R. Mansouri and J. Konrad, "Motion segmentation with level sets," in *Proc. IEEE International Conference on Image Processing*, vol. 2, pp. 126–130, Kobe, Japan, October 1999.

[15] N. Paragios and R. Deriche, "Geodesic active regions for motion estimation and tracking," in *Proc. International Conf. on Computer Vision*, vol. 1, pp. 688–694, Corfu, Greece, September 1999.

[16] C. Chesnaud, P. Réfrégier, and V. Boulet, "Statistical region snake-based segmentation adapted to different physical noise models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pp. 1145–1157, 1999.

[17] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[18] O. Amadieu, E. Debreuve, M. Barlaud, and G. Aubert, "Inward and outward curve evolution using level set method," in *Proc. IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.

[19] E. Debreuve, M. Barlaud, G. Aubert, I. Laurette, and J. Darcourt, "Space-time segmentation using level set active contours applied to myocardial gated SPECT," *IEEE Trans. on Medical Imaging*, vol. 20, no. 7, pp. 643–659, 2001.

[20] S. Jehan-Besson, M. Barlaud, and G. Aubert, "DREAM^{2}S: deformable regions driven by an Eulerian accurate minimization method for image and video segmentation," Technical Report RR 2001-14, Laboratoire I3S, 2001.

[21] S. Jehan-Besson, M. Barlaud, and G. Aubert, "Video object segmentation using Eulerian region-based active contours," in *Proc. International Conf. on Computer Vision*, Vancouver, British Columbia, Canada, July 2001.

[22] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-

Jacobi formulations," *J. Comput. Phys.*, vol. 79, no. 1, pp. 12–49, 1988.

[23] J. A. Sethian, *Level Set Methods*, Cambridge University Press, New York, NY, USA, 1996.

[24] J. Gomes and O. Faugeras, "Reconciling distance functions and level sets," *Journal of Visual Communication and Image Representation*, vol. 11, no. 2, pp. 209–223, 2000.

[25] E. Debreuve, *Segmentation par contours actifs en imagerie médicale dynamique: application en cardiologie nucléaire*, Ph.D. thesis, Univesité de Nice-Sophia Antipolis, 2000.

[26] S. Geman and D. E. McClure, "Bayesian image analysis: an application to single photon emission tomography," in *Proc. Statistical Computation Section*, Washington, Wash, USA, 1985.

[27] P. Kornprobst, R. Deriche, and G. Aubert, "Image sequence analysis via partial differential equations," *Journal of Mathematical Imaging and Vision*, vol. 11, no. 1, pp. 5–26, 1999.

[28] J. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, 1995.

[29] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 561–576, 2000.

[30] R. Mech, *Description of COST 211 Analysis Model (Version 4.0)*, COST 211, July 1998.

[31] S. Jehan-Besson, M. Barlaud, and G. Aubert, "Detection and tracking of moving objects using a new level set based method," in *Proc. Int. Conf. on Pattern Recognition*, pp. 1112–1117, Barcelona, Spain, September 2000.

[32] S. Jehan-Besson, M. Barlaud, and G. Aubert, "Region-based active contours for video object segmentation with camera compensation," in *Proc. IEEE International Conference on Image Processing*, Thessaloniki, Greece, October 2001.

[33] K. Siddiqi, Y. B. Lauzière, A. Tannenbaum, and S. W. Zucker, "Area and length minimizing flows for shape segmentation," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 433–443, 1998.

[34] *Working site for sequences and algorithms exchange*, http://www.tele.ucl.ac.be/EXCHANGE/.

Stéphanie Jehan-Besson received the engineering degree from École Centrale de Nantes, engineering school, in 1996 and a master of science in computer vision from the University of Nice-Sophia Antipolis in 1999. She is currently a Ph.D. student in the image processing group at Laboratory I3S, University of Nice-Sophia Antipolis, France. Her research interests include segmentation of video sequences, region-based active contours, and motion estimation.

Michel Barlaud received his Thèse d'État from the University of Paris XII. He is currently a professor of image processing at the University of Nice-Sophia Antipolis, and the leader of the image processing group of I3S. His research topics are image and video coding using scan-based wavelet transform, inverse problem using half quadratic regularization, and image and video segmentation using region-based active contours and PDEs. He is a regular reviewer for several journals, a member of the technical committees of several scientific conferences. He leads several national research, and development projects with French industries, and participates in several international academic collaborations (Universities of Maryland, Stanford, Louvain La Neuve and Naples). He is the author of a large number of publications in the area of image and video processing, and the editor of the book *Wavelets and Image Communication*, Elsevier, 1994.

Gilles Aubert received the Thèse d'État ès sciences mathématiques from the University of Paris VI, France, in 1986. He is currently professor of mathematics at the University of Nice-Sophia Antipolis and member of the J. A. Dieudonné Laboratory at Nice, France. His research interests are calculus of variations, nonlinear partial differential equations and numerical analysis; fields of applications including image processing and, in particular, restoration, segmentation, optical flow and reconstruction in medical imaging.