Rule-Driven Object Tracking in Clutter and Partial Occlusion with Model-Based Snakes

Gabriel Tsechpenakis

Center for Computational Biomedicine, Imaging and Modeling (CBIM), Division of Computer and Information Sciences,
Rutgers University, NJ 08854, USA
Email: gabtielt@cs.rutgers.edu

Konstantinos Rapantzikos

School of Electrical & Computer Engineering, National Technical University of Athens, Zografou, 15773 Athens, Greece
Email: rap@image.ntua.gr

Nicolas Tsapatsoulis

School of Electrical & Computer Engineering, National Technical University of Athens, Zografou, 15773 Athens, Greece
Email: ntsap@image.ntua.gr

Stefanos Kollias

School of Electrical & Computer Engineering, National Technical University of Athens, Zografou, 15773 Athens, Greece
Email: stefanos@cs.ntua.gr

In the last few years it has been made clear to the research community that further improvements in classic approaches for solving low-level computer vision and image/video understanding tasks are difficult to obtain. New approaches started evolving, employing knowledge-based processing, though transforming a priori knowledge to low-level models and rules are far from being straightforward. In this paper, we examine one of the most popular active contour models, snakes, and propose a snake model, modifying terms and introducing a model-based one that eliminates basic problems through the usage of prior shape knowledge in the model. A probabilistic rule-driven utilization of the proposed model follows, being able to handle (or cope with) objects of different shapes, contour complexities and motions; different environments, indoor and outdoor; cluttered sequences; and cases where background is complex (not smooth) and when moving objects get partially occluded. The proposed method has been tested in a variety of sequences and the experimental results verify its efficiency.

Keywords and phrases: model-based snakes, rule-driven tracking, object partial occlusion.

1. INTRODUCTION

In the last decade, snakes, a major category of active contours, have been given special attention in the fields of computer vision, image and video processing. They employ weak models, which deform in conformance with salient image features. The approaches proposed in the literature focus on either the highest accuracy of estimating moving silhouettes or the lowest computational complexity.

Active contours (snakes) were first introduced by Kass et al. [1]. A snake is actually a curve defined by energy terms, being able to deform itself in order to minimize its total energy. This total energy consists of an "internal" term, that enforces smoothness along the curve, and an "external" term, that makes the curve move towards the desired object boundaries. Many variations and extensions of snakes have been proposed and applied to certain applications [2, 3]. However, the majority of them faces three main limitations. The first one is the quality of the initialization that is crucial for the convergence of the algorithm. The second one is the need for parameter tuning that may lead to loss of generality, and the third one is the sensitivity to noise, clutter, and occlusions.

During the last decade, snakes and their variants were applied to motion segmentation [4, 5, 6, 7], object detection, localization, and tracking in video sequences [8, 9, 10, 11]. Most approaches require an initial shape approximation that is close to the objects' of interest boundaries [12]. The straightforward incorporation of prior knowledge in such

models is a very interesting property that makes them appropriate for capturing case-dependent constraints. Constraining the active contour representation to follow a global shape prior while preserving local deformations has drawn the interest of the research community. Cootes et al. [13] introduced the term "active shape models" to compensate for the extension of classical snakes with global constraints. They described a technique which allows an initial rough guess for the best shape, orientation, scale, and position to be refined by comparing a hypothesized model instance with image data, and using differences between model and image to deform the shape. The results demonstrate that their method can deal with clutter and limited occlusion. An efficient method towards the combination of low- and high-level information in a consistent probabilistic framework is proposed by Isard and Blake [14, 15]. The result is highly robust tracking of agile motion in clutter that runs in near real time. The condensation algorithm they introduced is a fusion of the statistical factor sampling algorithm for static, non-Gaussian problems with a stochastic differential equation model for object motion. Rouson and Paragios [16] proposed a two-stage approach using level-set representations. During the first stage, a shape model is built directly on the level-set space using a collection of samples. This model allows shape variabilities that can be seen as an "uncertainty region" around the initial shape. Then, this model is used as a basis to introduce the shape prior in an energetic form.

In the proposed approach, we consider a knowledge-based view of active contour models, which is appropriate for handling object tracking in partial occlusion, as well as tracking objects whose shape can be approximated by parameter-based models. We use shape priors and set them in a rather loose way to preserve the required deformations and introduce an uncertainty region around the contour to be extracted, which is based on motion history. In order to cope with partial occlusion, we use a rule-driven approach and provide several results. The algorithm seems to provide efficient solutions in terms of both accuracy and computational complexity. Head tracking has been selected as a test-bed application of the integrated model, where head is approximated by shape priors derived from an ellipsoid. This approach provides the constraint that the desired object is not strongly deformed in successive frames of video sequences, which is actually valid for most cases.

The paper is organized as follows. In Section 2 we review the classic snake model and provide information on the adopted model-based approach. Section 3 describes in detail the proposed tracking approach and Section 4 provides the experimental results. Future research directions are given in Section 5.

2. SNAKE MODEL

In general, snakes concern model and image data analysis through the definition of a linear energy function and a set of regularization parameters. Their energy function consists of two components, the internal or smoothness-driven one, which enforces smoothness along the snake, and the external or data-driven component, which depends on the image data according to a chosen criterion, forcing the snake towards the object boundaries. The goal is to minimize the total snake energy and this is achieved iteratively after considering an initial approximation of the object shape (prototype). Once such an appropriate initialization is specified, the snake can converge to the nearby energy minimum, using gradient descent techniques. According to that formulation, a snake is modeled as being able to deform elastically, but any deformation increases its internal energy causing a "restitution" force, which tries to bring it back to its original shape. At the same time, the snake is immersed in an energy field (created by the examined image), which causes a force acting on the snake. These two forces balance each other and the contour actively adjusts its shape and position until it reaches a local minimum of its total energy.

We consider a snake C_{snake} defined by a set $\mathbf{V}(\mathbf{s})$ of N ordered points (snaxels) $\{V_i(s) \mid i = 1, 2, \ldots, N\}$, corresponding to the positions $(x_i(s), y_i(s))$ in the image plane (s is a parameter denoting the normalized arc length in [0 1]. For simplicity, in the following the parameter s will be mentioned only when necessary). The total energy function E_{snake} is then defined by the weighted summation of the internal energy E_{int}, corresponding to the summation of the stretching and bending energies of the snake, and the external one which indicates how the snake evolves according to the features of the image:

$$E_{\text{snake}}(\mathbf{V}) = a_1 \cdot E_{\text{int}}(\mathbf{V}) + a_2 \cdot E_{\text{ext}}(\mathbf{V}), \qquad (1)$$

$$E_{\text{int}}(\mathbf{V}) = \sum_{i=1}^{N} e_{\text{int}}(V_i), \qquad (2)$$

$$E_{\text{ext}}(\mathbf{V}) = \sum_{i=1}^{N} e_{\text{ext}}(V_i), \qquad (3)$$

where $e_{\text{int}}(V_i)$ and $e_{\text{ext}}(V_i)$ are the internal and external energies corresponding to the point V_i, and the procedure of snake's convergence to the object boundary is given by the solution of its total energy minimization:

$$C_{\text{snake}} = \text{argmin} \left[a_1 \cdot E_{\text{int}}(\mathbf{V}) + a_2 \cdot E_{\text{ext}}(\mathbf{V}) \right], \qquad (4)$$

where a_1 and a_2 are the snake's regularization parameters.

2.1. Internal energy

The internal energy E_{int} has been given various definitions in the literature [17, 18, 19], depending on the application criteria. In our approach, we define the internal energy in terms of the snake curvature CU_{snake} and its point density distribution DV_{snake},

$$CU_{\text{snake}} = \frac{\dot{x} \cdot \ddot{y} - \ddot{x} \cdot \dot{y}}{\left(\dot{x}^2 + \dot{y}^2\right)^{3/2}},$$

$$DV_{\text{snake}} = \sqrt{\dot{x}^2 + \dot{y}^2}, \qquad (5)$$

where (x, y) parameterize the curve as $V_i = [x_i, y_i]$ and the first and second derivatives of (x, y) denote the velocity and the acceleration along the curve ($\dot{x} = dx/ds$, $\dot{y} = dy/ds$) and ($\ddot{x} = d^2x/ds^2$, $\ddot{y} = d^2y/ds^2$). Thus, the internal energy of the snake is defined as

$$e_{\text{int}}(V_i) = |CU_{\text{snake}}(V_i)| + DV_{\text{snake}}(V_i), \qquad (6)$$

where $|\cdot|$ denotes the magnitude of the corresponding quantities. In the discrete case, the value of the curvature at the kth point is calculated using the neighboring points to each side of it; the sign of the curvature is positive if the contour is locally convex, and negative if concave. Moreover, curvature distribution/function uniquely defines a propagating curve at different time instances although it is not affine invariant, and thus it is inappropriate in object recognition problems [18, 20]. In the proposed snake model, the points constituting a curve are not equally spaced and thus the distances between successive points represent the local elasticity of the snake. Finally, it should be noted that curvature and point density terms are often used in the literature [1, 19, 21], and in the present work they are used both as smoothness and curves similarity criteria, as described in the following sections. Figure 1 illustrates the curvature (curve smoothness) and point density (elasticity) distributions of a given snake.

2.1.1. Prior model constraints

The inclusion of a global shape model biases the snake contour towards a target shape, allowing some selectivity over image features. In several applications, the general shape, and possibly the location and orientation of objects, is known, and this knowledge may be incorporated into the deformable adaptive contour in the form of initial conditions, data constraints, constraints on the model shape parameters, or into the model fitting procedure. However, for efficient interpretation, it is essential to have a model that not only describes the size, shape, location, and orientation of the target object, but that also permits expected variations in these characteristics.

A number of researchers have incorporated knowledge of object shape into deformable models by using deformable shape templates. These models usually use global shape parameters to embody a priori knowledge of expected shape and shape variation of the structures and have been used successfully for many applications of automatic image interpretation. An excellent example in computer vision is the work of Yuille et al. [22], who constructed deformable templates for detecting and describing features of faces, such as the eye. Staib and Duncan [23] used probability distributions on the parameters of the representation and biased the model to a particular overall shape while allowing for deformations. Boundary finding is formulated as an optimization problem using a maximum a posteriori objective function. A model-based snake that is directly applicable in image space as opposed to parameter space is proposed in [24]. This method is simple and fast and therefore fits well to our intention to extend the previous formulation with a model prior constraint. We mention here that our goal is to illustrate the increased
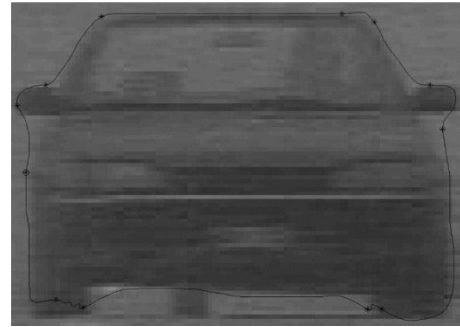
(a)
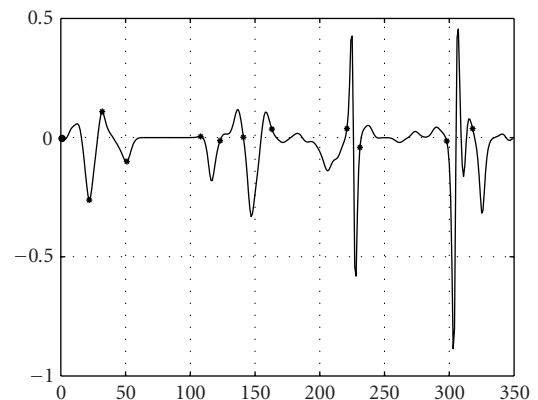
(b)
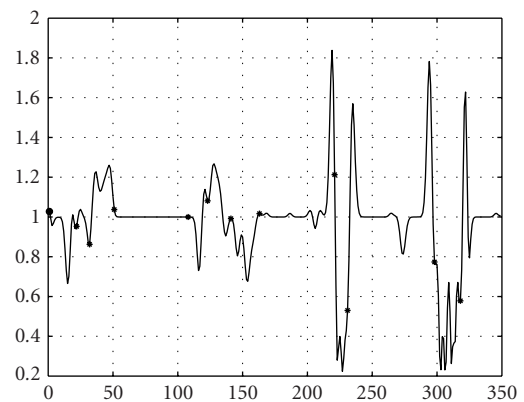
(c)

FIGURE 1: Curvature and point density distributions of a given contour. (a) The snake is locked at car boundaries whereas the circled areas denote parts of the curve of high curvature and point density: (b) curvature distribution and (c) point density distribution.

robustness of the proposed method provided by the inclusion of shape information rather than incorporating a novel shape prior constraint representation.

We formulate the model energy function by using a slightly different shape modeling than the one adopted in [24]. Therefore, we define the constraint energy term

$E_{\text{model}}(\mathbf{V}(s))$ as

$$
\begin{aligned}
E_{\text{model}}(\mathbf{V}(s)) &= \lambda \cdot \frac{1}{2} \cdot \sum_{i=1}^{N} e_{\text{model}}(V_i(s)) \\
&= \lambda \cdot \frac{1}{2} \cdot \sum_{i=1}^{N} \{V_i(s) - \text{model}(V_i(s))\}^T \\
&\quad \cdot \{V_i(s) - \text{model}(V_i(s))\},
\end{aligned} \tag{7}
$$

where λ is parameterized, since it can vary with position, and (6) is reformulated as

$$
e_{\text{int}}(V_i) = |CU_{\text{snake}}(V_i)| + DV_{\text{snake}}(V_i) + e_{\text{model}}(V_i). \tag{8}
$$

As an example, a generalized ellipse represented by (9) is used as a model ($\text{model}_{\text{ellipse}}$) here. Ellipse is a typical model for human faces and therefore is appropriate for head tracking, which is our test-bed application,

$$
\begin{aligned}
&\text{model}_{\text{ellipse}}(Vi(s)) \\
&= \begin{bmatrix} a \cdot \cos\vartheta \cdot \cos(2\pi s - \vartheta) - b \cdot \sin\vartheta \cdot \sin(2\pi s - \vartheta) \\ a \cdot \sin\vartheta \cdot \cos(2\pi s - \vartheta) + b \cdot \cos\vartheta \cdot \sin(2\pi s - \vartheta) \end{bmatrix},
\end{aligned} \tag{9}
$$

where a and b are the minor and major axes, respectively, and ϑ is the ellipsoid rotation. The model should take scaling, translation, and rotation under consideration. In order to meet the previous requirements, we base the minor and major axes and rotation calculation on a statistical representation of an ellipsoid as the covariance matrix $\mathbf{S}$ derived from the distribution of the last recovered (previous frame) solution points,

$$
S = \begin{bmatrix} e_1 & e_2 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \cdot \begin{bmatrix} e_1^\top \\ e_2^\top \end{bmatrix}. \tag{10}
$$

The eigenvalues λ_1 and λ_2 ($\lambda_1 \geq \lambda_2$) correspond to each of the principal directions e_1 and e_2, respectively. The eigenvalues determine the shape of the ellipsoid, while the eigenvectors determine the orientation as shown in Figure 2.

2.2. External energy

The external energy term, in most approaches, for each point V_i, is defined as

$$
e_{\text{ext}}(V_i) = 1 - |\nabla G_\sigma * I(x_i, y_i)| \cdot |\underline{g}(V_i) \cdot \underline{n}(V_i)|, \tag{11}
$$

where $|\nabla G_\sigma * I(x_i, y_i)|$ denotes the magnitude of the gradient of the image convolved with a Gaussian filter, of variance σ at point (x_i, y_i) corresponding to the snaxel V_i; $\underline{g}(V_i)$ is the respective gradient direction; and $\underline{n}(V_i)$ is the normal vector of the snake at the snaxel V_i.

The common problems in snake models are the presence of noise, background edges close to object boundaries, and edges in the interior of the desired object. These problems flow from the definition of the external energy and the Laplacian-of-Gaussian (LoG) term $\nabla G_\sigma * I$, especially in
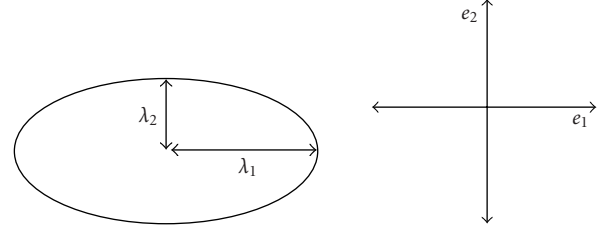

FIGURE 2: Proposed model constraining the obtained solutions to the application of the human head modeling and tracking.

cases where the initialization is not close enough to object boundaries. For that reason, snakes turn out to be efficient only in specific cases of images and video sequences. In the proposed model, another term is introduced instead, minimizing the local variance of the image gradient and preserving the most important image regions. This is achieved through morphological operations leading to a modified image gradient. In particular, the expression $|\nabla G_\sigma * I(x_i, y_i)|$ is replaced by a modified image gradient G_m and the image data criterion is strengthened through the square of G_m:

$$
e_{\text{ext}}(V_i) = 1 - G_m^2 \cdot |\underline{g}(V_i) \cdot \underline{n}(V_i)|. \tag{12}
$$

To obtain the modified image gradient, we first presmooth the image with a nonlinear morphological filter, called alternating sequential filter (ASF) [25] and we extract the morphological image gradient. The ASF used in our model is based on morphological area opening ($\circ$) and closing ($\bullet$) operations with structure elements of increasing scale. The main advantage of such filters is that they preserve line-type image structures, which is impossible to be achieved with, for example, median filtering. Figure 3 illustrates the performance of a frame's presmoothing with the proposed ASF; it can be clearly seen that noise is eliminated and the most important edges are preserved. More details can be found at [26]. Figure 4 illustrates the differences between the two image data criteria $|\nabla G_\sigma * I(x_i, y_i)|$ and G_m, presented in (11) and (12). It can be seen in Figures 4b and 4c that the proposed procedure clearly suppresses noise and retains the most important edges of the examined image, whereas Figures 4d and 4e illustrate the difference between image gradient and the proposed modified gradient, computed along a randomly selected image line.

Figure 4 clearly shows the advantages of the proposed external energy term for edge-based methods in terms of noise reduction and preservation of the most important edges. Comparing this external energy with related work found in the literature, except for the commonly used LoG-based definitions, a representative example is the respective term proposed in [27]. In this work, a Gaussian filter is used to obtain the image gradient, but an appropriate value of the Gaussian variance is required, which is done manually. Figure 5 illustrates the difference between the proposed external energy term and the one proposed in [27].

(a)

(b)

FIGURE 3: Frame presmoothing with the proposed ASF: (a)original frame and (b) filtered frame.
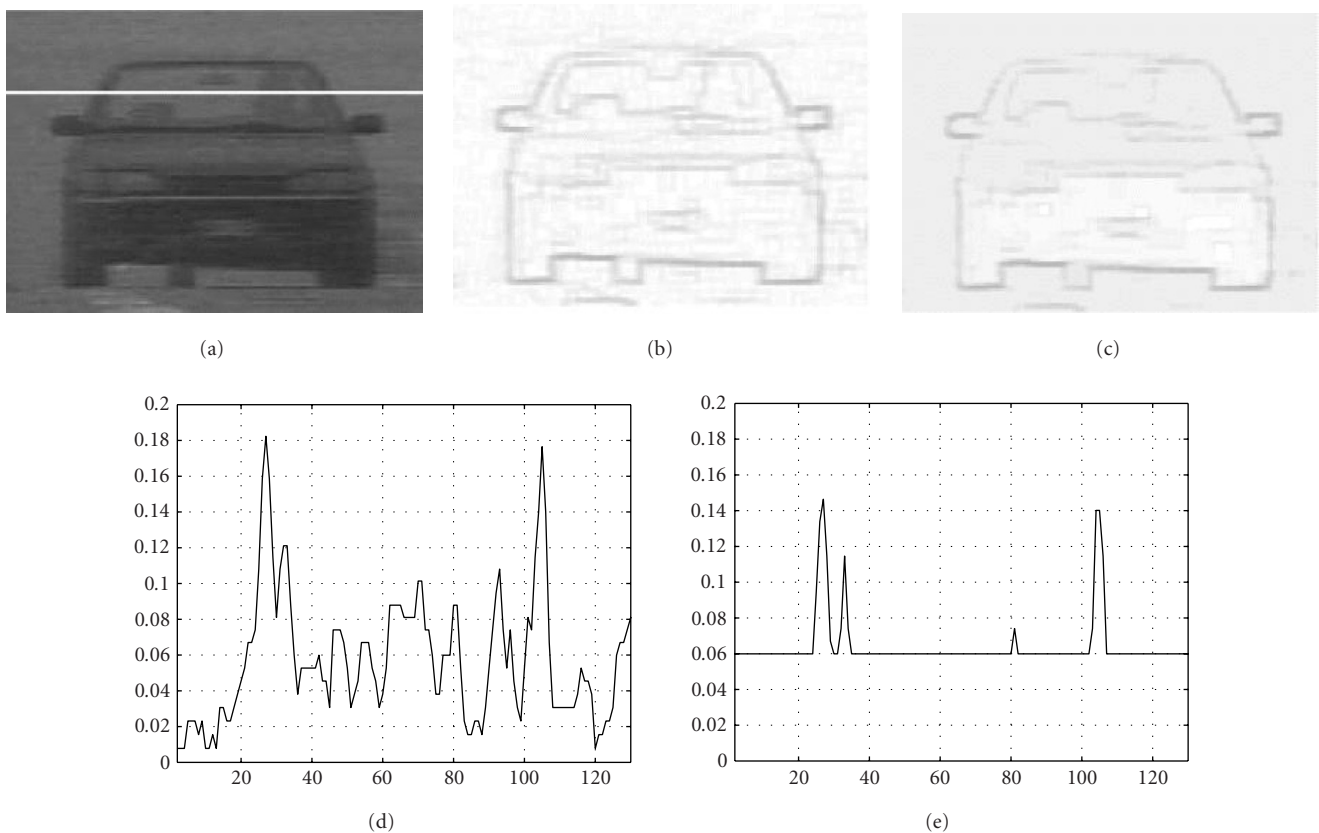

(a)

(b)

(c)

(d)

(e)

FIGURE 4: Differences between the image data criteria using the image gradient and the proposed one. (a) Original image, (b) image gradient, (c) modified image gradient, (d) image gradient computed along a randomly chosen row shown in (a), and (e) modified image gradient computed along the same row.

3. THE PROPOSED TRACKING APPROACH

Object tracking actually concerns the separation of moving objects from background [28], which is done so far in two different ways: (a) the motion-based approaches that rely on grouping motion information over time and (b) the model-based approaches that impose high-level semantic representation and knowledge. In these approaches, either geometrical properties or region-based features of the desired ob-

jects are extracted and utilized. Thus the methods proposed in the literature can be categorized in edge-based methods [14], which rely on the boundary information, and region-based ones [29], utilizing the information provided by the interior region of the tracked objects.

The main problems that tracking approaches are called upon to cope with are nonrigid (deformable) objects, objects with complicated (not smooth) contours, object movements that are not simple translations, and movement in

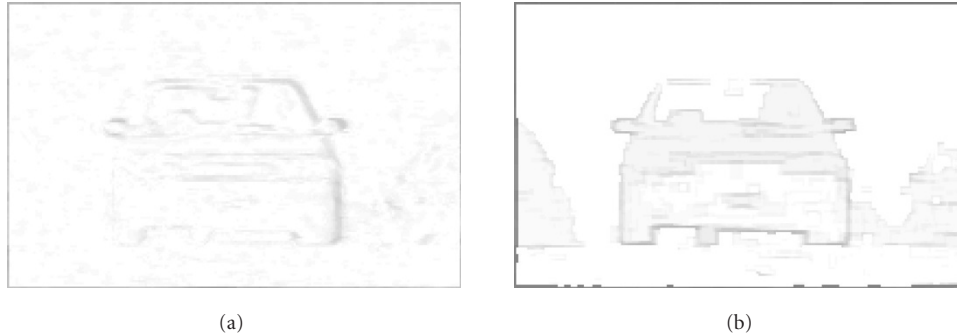(a)   (b)

FIGURE 5: Qualitative comparison between (a) a representative example of external energy term using Gaussian filtering and (b) the proposed external energy term.

natural sequences, where background is usually complicated and the amount of noise or external lighting changes are not known. The latter has been a motivation for many researchers, especially in the last years, to follow probabilistic approaches, for example, [30]. In addition, a more difficult problem emerges in many sequences, the occlusion, that is, when moving objects get occluded successively as time passes. This requires some assumptions about the shape, region, or motion of the tracked object in order to estimate its contour even in regions that are covered by other moving or static objects. In the following, we describe the proposed approach, which aims to cope with the above mentioned problems.

The proposed method consists of two main steps: the extraction of the "uncertainty regions" of each object in a sequence, and the estimation of the mobile object contours. The term "uncertainty regions" is used to describe the regions in a frame, where moving contours are possible to be located, whereas the estimation of the contours consists of an energy minimization procedure based on the proposed snake energy terms, described in Section 2. More specifically, the contour of a moving object is estimated first in a few successive frames of a sequence. This can be achieved with appropriate parameter initialization utilizing the proposed snake model. Then, for the next frames, a force-based approach is being followed to minimize the total snake energy inside the respective uncertainty regions, which are extracted using the displacement history of each point of the contour. The force-based approach is adopted as an alternative to direct energy minimization, while some rules are introduced to separate objects from background and to detect possible occlusions.

3.1. Uncertainty region estimation

The minimization procedure of snake's total energy is actually a problem of picking out the "correct" curve in the image, that is, the curve which corresponds to the object of interest among a set of candidate curves, given an initial estimate of the object's contour. In this section, we propose a way to determine a region around the snake initialization,

for each frame of a video sequence, in which the correct curve is located. This idea is not new, as stochastic models have been lately proposed in the literature, mostly as shape prior knowledge [8], to define possible positions of the curve points around an initialization. In the same direction, we introduce here the term "uncertainty region," which denotes that the minimization procedure (or the picking out of the correct curve) takes place inside that region, constraining the problem inside a narrow band around the snake initialization. Such regions are extracted by exploiting the motion history of the tracked contour (curve points' displacements in previous time instances), extracting statistical measurements of the motion. The previously estimated contour is deformed according to the previously calculated point displacements (initialization for the next frame), and the standard deviation of each point's mean motion is calculated; the uncertainty region around each point is then defined in terms of its corresponding standard deviation. The next step is to find the new position of each point of the curve, inside its corresponding uncertainty region, which corresponds to the minimum of a criterion, which is defined by the snake's energy terms described in Section 2.

We define the contour of an object, located in the Ith frame ($I > 1$), of a video sequence as a vector of complex numbers, that is,

$$
\begin{aligned}
\mathbf{C}^{(I)} &= \left[x_i^{(I)} + j \cdot y_i^{(I)} \mid i = 1, \dots, N \right] \\
&= \left[C_{(1)}^{(I)}, \dots, C_{(N)}^{(I)} \right],
\end{aligned}
\tag{13}
$$

where $C_{(k)}^{(I)} = x_k^{(I)} + j \cdot y_k^{(I)}$ is the location of the kth point of the contour. We define the instant motion of the kth point of the object contour, computed in the Ith frame, as

$$
m_{c,k}^{(I)} = MF^{(I-1,I)}(x_k, y_k),
\tag{14}
$$

where $MF^{(I-1,I)}(x_k, y_k)$ is the motion vector of the pixel (x_k, y_k) estimated with the use of a robust motion estimation technique proposed by Black and Anandan [31], between the successive frames $I - 1$ and I.

Based on the definition of the instant motion, we calculate the mean movement of the contour $\mathbf{C}$ up to frame I as

$$\bar{\mathbf{m}}_c^{(I)} = \left[\bar{m}_{c,1}^{(I)}, \ldots, \bar{m}_{c,N}^{(I)}\right], \tag{15}$$

where

$$\bar{m}_{c,k}^{(I)} = \bar{m}_{x,k}^{(I)} + j \cdot \bar{m}_{y,k}^{(I)} = \frac{1}{I-1} \sum_{i=1}^{I-1} m_{c,k}^{(i+1)} \tag{16}$$

is the corresponding mean movement of the kth point of the contour.

Similarly, the standard deviation of contour's mean movement is defined as

$$\bar{\mathbf{s}}_c^{(I)} = \left[\bar{s}_{c,1}^{(I)}, \ldots, \bar{s}_{c,N}^{(I)}\right], \tag{17}$$

where

$$\bar{s}_{c,k}^{(I)} = \left\{\frac{1}{I-1} \sum_{i=1}^{I-1} \left(\bar{m}_{x,k}^{(I)} - m_{x,k}^{(i+1)}\right)^2\right\}^{1/2}$$
$$+ j \cdot \left\{\frac{1}{I-1} \sum_{i=1}^{I-1} \left(\bar{m}_{y,k}^{(I)} - m_{y,k}^{(i+1)}\right)^2\right\}^{1/2} \tag{18}$$

is the standard deviation of kth point's mean movement.

In practice, (16) and (18) are computed based on the last L frames so as to take into account only the recent history of contour's movement, that is,

$$\bar{m}_{c,k}^{(I)} = \frac{1}{L} \sum_{i=I-L}^{I-1} m_{c,k}^{(i+1)}, \tag{19}$$

$$\bar{s}_{c,k}^{(I)} = \left\{\frac{1}{L} \sum_{i=I-L}^{I-1} \left(\bar{m}_{x,k}^{(I)} - m_{x,k}^{(i+1)}\right)^2\right\}^{1/2}$$
$$+ j \cdot \left\{\frac{1}{L} \sum_{i=I-L}^{I-1} \left(\bar{m}_{y,k}^{(I)} - m_{y,k}^{(i+1)}\right)^2\right\}^{1/2}. \tag{20}$$

The initial estimation of the object's contour $\mathbf{C}_{\text{init}}^{(I+1)}$ in the frame $I+1$ is computed based on the contour's current location and

(a) its mean motion when no abrupt movements are expected to occur, that is,

$$\mathbf{C}_{\text{init}}^{(I+1)} = \mathbf{C}^{(I)} + \bar{\mathbf{m}}_c^{(I)}, \tag{21}$$

or

(b) its instant motion when no knowledge about the motion of the desired object is available, that is,

$$\mathbf{C}_{\text{init}}^{(I+1)} = \mathbf{C}^{(I)} + \mathbf{m}_c^{(I+1)}, \tag{22}$$

where $\mathbf{m}_c^{(I+1)} = [MF^{(I,I+1)}(x_i, y_i) \mid i = 1, \ldots, N] = [m_{(c,1)}^{(I+1)}, \ldots, m_{(c,N)}^{(I+1)}]$.

The final solution, that is, the desired contour $\mathbf{C}^{(I+1)} = [C_{(k)}^{(I+1)} \mid k = 1, \ldots, N]$, is obtained by solving the following equations:

$$\mathbf{C}^{(I+1)} = \text{argmin}_{\mathbf{V} \in R} \Big\{ w_1 \cdot E_{\text{ext}}(\mathbf{V})$$
$$+ w_2 \cdot \Big[\mu_1 \cdot \left(D_{CU}^{(I)}(\mathbf{V}) + D_{DV}^{(I)}(\mathbf{V})\right)$$
$$+ \mu_2 \cdot E_{\text{model}}(\mathbf{V}) \Big] \Big\}, \tag{23}$$

$$D_{CU}^{(I)}(\mathbf{V}) = \sum_{k=1}^{N} \left\{ CU\left(\mathbf{C}_{(k)}^{(I)}\right) - CU(V_k) \right\}^2, \tag{24}$$

$$D_{DV}^{(I)}(\mathbf{V}) = \sum_{k=1}^{N} \left\{ DV\left(\mathbf{C}_{(k)}^{(I)}\right) - DV(V_k) \right\}^2, \tag{25}$$

where $E_{\text{ext}}(\mathbf{V})$ and $E_{\text{model}}(\mathbf{V})$ are given by (3) and (7), respectively, $CU(\mathbf{C}_{(k)}^{(I)})$ and $DV(\mathbf{C}_{(k)}^{(I)})$ are the curvature and the point density values of the contour $\mathbf{C}^{(I)}$ at the kth point. Parameters w_1 and w_2 represent the weights with which the energy-based terms of (23) participate in the minimization procedure, whereas μ_1 and μ_2 control the model's influence on the final solution; more about these weights is discussed in Section 3.3.

The set of all possible curves R, defining the uncertainty region, emerges by oscillating the points of the curve $\mathbf{C}_{\text{init}}^{(I+1)}$ according to the standard deviation of their mean movement, computed using (17) and (20). The Gaussian formulation for the point oscillations is mainly adopted to show that each point of the curve is likely to move in the same way (amplitude and direction) that it has been moving until the current frame. In this way, and for each contour point V_k, an uncertainty region is defined. If $C_{(k)}^{(I)}$ is the location of the kth point of the contour in frame I and this point was static during the previous L frames, then $\bar{s}_{c,k}^{(I)} = 0$ and its uncertainty region shrinks to a single point whose location coincides with $C_{\text{init},(k)}^{(I+1)}$. If point k was moving with invariable velocity, then the standard deviation of its movement is again $\bar{s}_{c,k}^{(I)} = 0$ and the previous case holds regarding its uncertainty region. On the other hand, if point k was oscillating in the previous L frames, the standard deviation of its movement is high and consequently its uncertainty region is large. Figure 6 illustrates the proposed approach in steps, in the case of face tracking. Figures 6a and 6b present two successive frames of a face sequence and the respective contours. Figure 6c presents the amplitude of the computed standard deviation (in pixels) of the contour mean motion, and based on this standard deviation, the uncertainty regions are then extracted (Figure 6d).

3.2. Force-based approach

The minimization of (23) is a procedure of high complexity: if N is the number of points determining the examined curve $\mathbf{C}$ and M is the number of all possible positions of each curve point $C_{\text{init},k}^{(I+1)}$ inside the extracted uncertainty region, assuming that M is the same for all points, then the number of all possible curves $\mathbf{r} \in R$ generated by points' oscillations is M^N. In order to avoid that problem, we propose a force-based
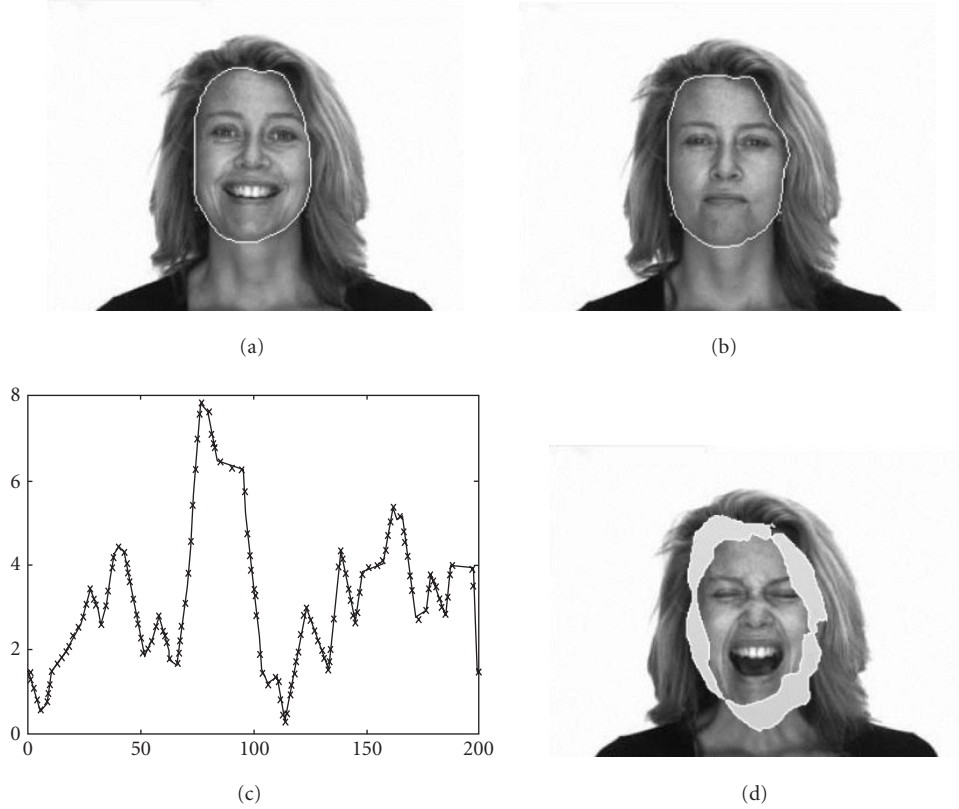
(a)

(b)

(c)

(d)

Figure 6: The proposed tracking approach in steps. (a), (b) Two successive frames of a face sequence and the respective contours. (c) Amplitude of the standard deviation of the contour mean motion leading to (d) the uncertainty regions of the curve.

approach (instead of using a dynamic programming algorithm) where the energy terms, participating in the snake energy function, are transformed into forces applied in each curve point so as to converge to the desired object boundaries.

We consider the curve $\mathbf{V}$ describing the object's contour. The object's contour at frame I is given by $\mathbf{C}^{(I)}$ and its initialization at frame $I + 1$ is given by $\mathbf{C}_{\text{init}}^{(I+1)}$. Also let $\underline{t}$ be the set of the tangential unit vectors and $\underline{n}$ the set of the normal vectors of curve $\mathbf{V}$, given by (28):

$$\mathbf{t} = [\underline{t}_k \mid k = 1, \ldots, N], \tag{26}$$

$$\mathbf{n} = [\underline{n}_k \mid k = 1, \ldots, N], \tag{27}$$

$$t_k = \frac{\nabla \mathbf{V}|_k}{||\nabla \mathbf{V}|_k||}, \qquad n_k = \frac{\nabla \underline{t}|_k}{||\nabla \underline{t}|_k||}. \tag{28}$$

We define the following forces acting at each contour point V_k:

$$
\begin{aligned}
\underline{F}_d(V_k) &= D_{DV}^{(I)}(V_k) \cdot \underline{t}_k \\
&= \left[DV\left(\mathbf{C}_{(k)}^{(I)}\right) - DV(V_k) \right] \cdot \underline{t}_k, \\
\underline{F}_c(V_k) &= D_{CU}^{(I)}(V_k) \cdot \underline{n}_k \\
&= \left[CU\left(\mathbf{C}_{(k)}^{(I)}\right) - CU(V_k) \right] \cdot \underline{n}_k.
\end{aligned}
\tag{29}
$$

$\mathbf{F_d} = [\underline{F}_d(V_k) \mid k = 1, \ldots, N]$ represents the stretching component that forces points to come closer or draw away from each other along the curve, and it is always tangential to it. Thus, if the distance between two curve points $\mathbf{C}_{(k)}^{(I)}$ and $\mathbf{C}_{(k+1)}^{(I)}$ is greater than the distance between V_k and V_{k+1}, then $\underline{F}_d(V_k) \cdot \underline{t}_k > 0$ and V_k is forced to draw away from V_{k+1}; otherwise, $\underline{F}_d(V_k) \cdot \underline{t}_k < 0$ and V_k is forced to come closer to V_{k+1}.

$\mathbf{F_c} = [\underline{F}_c(V_k) \mid k = 1, \ldots, N]$ represents the deformation of the curve along its normal direction. The property of the curvature distribution to take low values, where the curve is relatively smooth, and high values, where the curve has strong variations, makes $\mathbf{F_c}$ force curve to the initial shape (the one in the previous frame) and not to a smoother form. Moreover, we exploit the curvature's property to be positive where the curve is convex and negative where the curve is concave. Figure 7 illustrates the directions of $\underline{F}_c$ and $\underline{F}_d$ along a curve.

These forces represent the internal snake forces that deform the curve $\mathbf{V}$, initialized at $\mathbf{C}_{\text{init}}^{(I+1)}$, according to the shape of the contour $\mathbf{C}^{(I)}$ in the previous frame. The constraint of such a deformation is actually the first term of (23), that is, the external energy E_{ext}, which is transformed into force as described in the following.

We define $g_{m,k}(\overline{p})$, given by (30), to be the modified image gradient function of all pixels $\overline{p} = x_p + j \cdot y_p$, that (a)
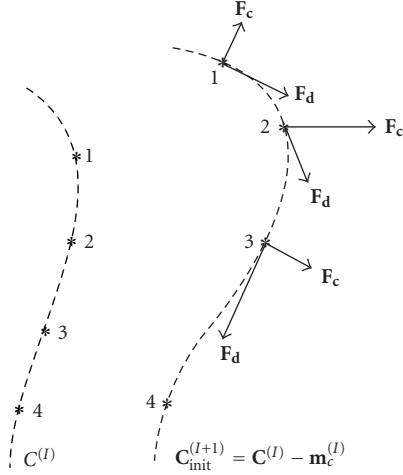
FIGURE 7: Curvature-based and point density-based forces $\underline{F}_c$ and $\underline{F}_d$, respectively, along the initialization of a curve $\mathbf{V}$ in the frame $I + 1$.

belongs to the uncertainty region $\mathbf{U}$ and (b) lies on the line segment that is defined by the normal direction of the curve $\mathbf{V}$ at point V_k,

$$g_{m,k}(\overline{p}) = \left[G_m(\overline{p}) \mid (V_k - \overline{p})^T \cdot \underline{n}_k = 1, \ \overline{p} \in \mathbf{U} \right]. \quad (30)$$

The maximum of this function determines the most salient edge pixel in the line segment defined above and thus defines the direction of the *external* snake force:

$$\overline{p}_k = \arg \max_{\overline{p}} \left[g_{m,k}(\overline{p}) \right], \quad (31)$$

$$\mathrm{sgn}_k = \begin{cases} +, & \text{if } \overline{p}_k \text{ inside the area is defined by } \mathbf{V}, \\ -, & \text{otherwise}, \end{cases} \quad (32)$$

where sgn_k denotes the sign/direction of the external force to be applied to V_k.

Then, the external snake force for each point V_k is given by

$$\underline{F}_e(V_k) = \mathrm{sgn}_k \cdot e_{\mathrm{ext}}(V_k) \cdot \underline{n}_k. \quad (33)$$

From the definition of the external energy term (12), it can be seen that it takes values close to zero in contour points corresponding to regions with high image gradient ($G_m^2(V_k) \simeq 1$) and values close to unity in regions with relatively constant intensity ($G_m^2(V_k) \simeq 0$). Thus, the term $\mathbf{F_e} = [\underline{F}_e(k) \mid k = 1, \dots, N]$ is proportional to G_m and forces the curve to the salient edges inside the extracted uncertainty region. In the definition of this force, we exploit the advantage of G_m against $|\nabla G_\sigma * I|$ to preserve the most important edges, as shown before, and thus the problem of the existence of many local maxima in (31) is eliminated.

In the force-based approach, the examined curve $\mathbf{V}$ marches towards the object's boundaries in the next frame, $I + 1$, according to the forces applied to it. Thus, the min-

imization of (23) can be approximated by using the internal and external snake forces defined above, in an iterative manner similar to the steepest descent approach [32], as it is summarized below. In particular, let $\mathbf{V}^{(\xi)}$ be the estimated contour in the ξ iteration, then the following equations hold:

$$\mathbf{V}^{(0)} = \mathbf{C}_{\mathrm{init}}^{(I+1)}, \quad (34)$$

$$\mathbf{V}^{(\xi)} = \mathbf{V}^{(\xi-1)} + \Delta \mathbf{V}^{(\xi)}, \quad (35)$$

$$\Delta \mathbf{V}^{(\xi)} = \left[\left\{ V_k^{(\xi-1)} \right\}^T \cdot \underline{F}_{\mathrm{tot}} \left(V_k^{(\xi-1)} \right) \mid k = 1, \dots, N \right], \quad (36)$$

$$\begin{aligned} \underline{F}_{\mathrm{tot}} \left(V_k^{(\xi-1)} \right) &= w_1 \cdot \underline{F}_e \left(V_k^{(\xi-1)} \right) \\ &+ w_2 \cdot \left[\mu_1 \cdot \left\{ \underline{F}_c \left(V_k^{(\xi-1)} \right) + \underline{F}_d \left(V_k^{(\xi-1)} \right) \right\} \right. \\ &\left. + \mu_2 \cdot \underline{F}_{\mathrm{model}} \left(V_k^{(\xi-1)} \right) \right], \end{aligned}$$

$$(37)$$

where $\underline{F}_d(V_k^{(\xi-1)})$, $\underline{F}_c(V_k^{(\xi-1)})$, and $\underline{F}_e(V_k^{(\xi-1)})$ are estimated according to (29) and (33), respectively, and $\underline{F}_{\mathrm{model}}(V_k^{(\xi-1)})$ is the regularization force, according to the specific model adopted, given by

$$\begin{aligned} \underline{F}_{\mathrm{model}} \left(V_k^{(\xi-1)} \right) \\ = \lambda \cdot \left\{ V_k^{(\xi-1)}(s) - \mathrm{model} \left(V_k^{(\xi-1)}(s) \right) \right\}. \end{aligned} \quad (38)$$

It is clear from the above definition that $\underline{F}_{\mathrm{model}}(V_k)$ forces contour point $V_k^{(\xi-1)}$ towards the model point $\mathrm{model}(V_k^{(\xi-1)}(s))$.

The final curve $\mathbf{V}$ corresponding to the contour $\mathbf{C}^{(I+1)}$ is obtained when one of the following criteria is satisfied.

(a) $F_\tau(\mathbf{V}^{(\xi)}) < a \cdot F_\tau(\mathbf{V}^{(\xi+1)})$, where

$$F_\tau \left(\mathbf{V}^{(\xi)} \right) = \sum_{k=1}^N \left| \underline{F}_{\mathrm{tot}}(V_k^{(\xi)}) \right|. \quad (39)$$

Parameter a is a positive constant in the range $0 < a < 1$. When a is selected to be close to 1, $\mathbf{C}^{(I+1)}$ is more likely to correspond to a local minimum solution; lower values of a increase the number of iterations and, therefore, the execution time. The statistical approach we follow to estimate the regions of uncertainty allows for the use of a close to 1.

(b) The maximum number of iterations is reached. In this case,

$$\begin{aligned} \mathbf{C}^{(I+1)} &= \mathbf{V}^{(\tilde{\xi})}, \\ \tilde{\xi} &= \mathrm{argmin}_\xi \left[F_\tau(\mathbf{V}^{(\xi)}) \right]. \end{aligned} \quad (40)$$

It must be noted that the use of the proposed steepest descent approach does not ensure that the final contour corresponds to the solution of (23). However, under the constraints we pose, even if $\mathbf{C}^{(I+1)}$ corresponds to a local minimum, it is close to the desired solution (global minimum).

(a)   (b)   (c)

(d)   (e)   (f)

(g)   (h)   (i)
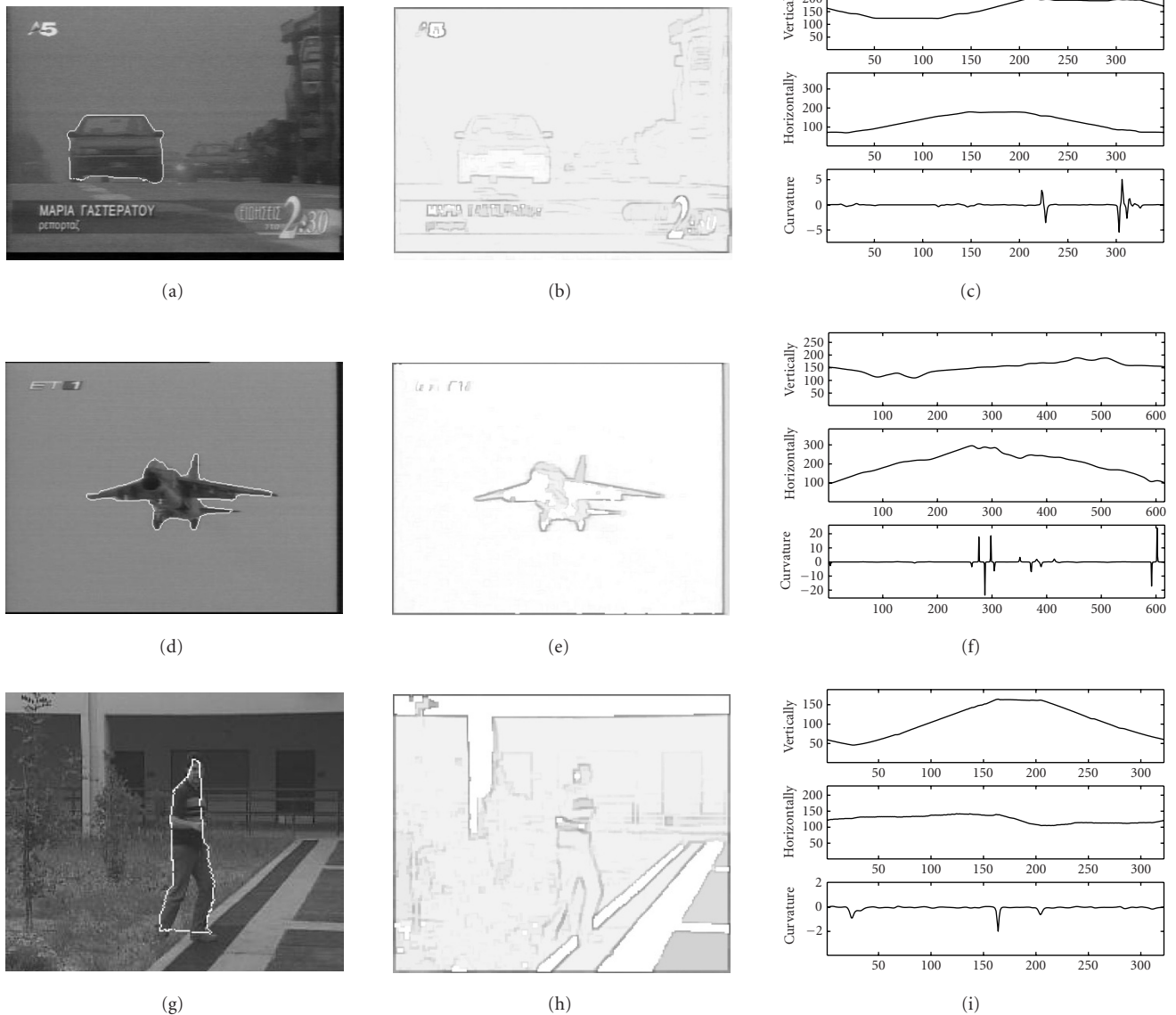
FIGURE 8: Curvature and external energy terms: (a), (d), (g) different cases of curves and background complexity, (b), (e), (h) respective external energies visualization, and (c), (f), (i) respective curvature distributions.

3.3. Weights estimation

In (23) and (37), four energy and force terms, respectively, participate in the minimization procedure with different weights w_1, w_2, μ_1, and μ_2. The choice of appropriate values for these weights is important for the method's performance. The values should be set depending on the amount of the background complexity and the smoothness of the object silhouette. For sequences with relatively smooth background (without any significant edges close to object boundaries, or edges far from object boundaries), the curve's external energy/force term is used as a reliable criterion and thus w_1 is set to higher value. Moreover, if the contour of the tracked object is complicated (not smooth) or noisy, the elasticity

and smoothness energy/force terms are not reliable and thus w_2 is set to lower values.

In order to automatically estimate the value of w_2, it suffices to count the curvature and point density distributions' zero crossings, which can give us the contour's local smoothness/elasticity. To estimate the value of w_1, it suffices to calculate the mean values of the external energy at all pixels $\overline{p}$ inside the extracted uncertainty region $\mathbf{U}$ (as verified by trial and error). Thus, smooth background inside the uncertainty region results in higher mean values and w_1 is set to a higher value, whereas low mean values correspond to cases of complex/noisy uncertainty regions (great number of edge pixels) and w_1 is set to a lower value.

Figure 8 illustrates three different sequences capturing

moving objects of different contour complexities. Figures 8a, 8d, and 8g represent the original images along with the moving object contours. At the first sequence, the background is relatively smooth and the object (car) has an uncomplicated contour. In the case of the aircraft, the background is also smooth but the contour is quite complicated, whereas in the third case, the walking man's contour is simple but the background is very cluttered. The respective external energies visualization is illustrated in Figures 8b, 8e, and 8h where the background complexity can be clearly seen; the modified image gradient preserves the most salient edges and eliminates noise. Finally, in Figures 8c, 8f, and 8i, the complexity of the respective object contours is presented in terms of the curvature. The first and the second subplots of each case illustrate the x and y coordinate distributions along the curves, whereas the third subplot represents the curvature distribution. As can be seen, the complexity of the contours is determined by the curvature variations, that is, the curvature's zero crossings: if $Z_{c,\mathbf{car}}$, $Z_{c,\mathbf{aircraft}}$, and $Z_{c,\mathbf{man}}$ are the numbers of the respective zero crossings and $N_{\mathbf{car}}$, $N_{\mathbf{aircraft}}$, and $N_{\mathbf{man}}$ are the number of points constituting the three contours, then $Z_{c,\mathbf{car}}/N_{\mathbf{car}} \simeq 0.03$, $Z_{c,\mathbf{aircraft}}/N_{\mathbf{aircraft}} \simeq 0.08$, and $Z_{c,\mathbf{man}}/N_{\mathbf{man}} \simeq 0.05$.

The parameters μ_1 and μ_2 related to the internal snake force can be set according to the application under consideration. If strict prior model knowledge is available (e.g., medical applications), then the model can strongly influence the solution. On the contrary, if there is no high certainty regarding the model prior, then the first term of the internal force should affect the solution more. This competitive relation of the two internal force terms can be easily represented by allowing one of them to change according to the other in a functional manner ($\mu_1 = f(\mu_2)$).

3.4. Rule-driven approach for complex background and partial occlusion cases

In order to separate background and object regions, especially when the background in not homogeneous (smooth), as well as to cope with moving object's partial occlusion that may occur, we introduce more constraints that $\overline{p}_k$ in (31) must obey, so that its estimation will be reasonable. The adopted motion estimation technique [31] ensures the distinction between moving background and foreground even in hard-to-detect cases (slightly different movements). Therefore, without loss of generality, we suppose that the background is static and possible occluding objects are also static. Let $\bar{m}_{c,k}^{(I)}$ be the mean estimated motion of the kth contour point at frame I, estimated through (16) or by any motion estimation algorithm as shown in Figure 14, and let $\overline{p}_l$ and $\overline{p}_m$ be the surrounding pixels of $\overline{p}_k$ on the line segment, along which the function $g_{m,k}(\overline{p})$ is computed. Then, $\overline{p}_k$ must fulfill the following two constraints/requirements:

(a) $\overline{p}_k$ must divide that line segment in two parts: an immiscibly moving and an immiscibly static one, that is,

$$u(\overline{p}_l) \simeq \bar{m}_{c,k}^{(I)}, \qquad u(\overline{p}_m) \simeq 0; \tag{41}$$

or

$$u(\overline{p}_l) \simeq 0, \qquad u(\overline{p}_m) \simeq \bar{m}_{c,k}^{(I)}; \tag{42}$$

(b) $\overline{p}_k$ must be a moving point with velocity close to $\bar{m}_{c,k}^{(I)}$, that is,

$$u(\overline{p}_k) \simeq \bar{m}_{c,k}^{(I)}, \tag{43}$$

where $u(\cdot)$ denotes the instant velocity.

Thus, taking the above constraints into consideration, we overcome cases such as (a) when the maximum is found in background: it is not a moving one and does not separate two immiscible (according to the motion) parts of the function g_m, (b) when the maximum is found inside the moving object region: although it is a moving one, it does not divide the function g_m in such two parts, (c) when occlusion occurs and the maximum is on the occluding object boundary: the maximum is not moving although it makes the region g_m separation, and (d) when occlusion occurs and the maximum is in the occluding object region: neither the maximum is moving nor it makes such a separation. In these cases, where these two constraints are not reached, we ignore the external force and evolve the curve according to its internal forces; in this way, we can obtain contours similar to the ones in the past frames. Figure 9 illustrates the detection of occlusion with the use of the above-defined rules that the local maximum $\overline{p}_k$ (shown as minimum), corresponding to a curve point k, must obey. It has to be mentioned that treating large occlusions is limited by the capabilities of the estimated motion field.

4. EXPERIMENTAL RESULTS

The performance of the proposed approach is tested over a large number of natural sequences, where specific tracking problems emerge. The results presented in this section concern cases of different object shape complexities, different motions, noisy video sequences, complicated backgrounds, as well as partial occlusion. Finally, a specific application of the proposed method is shown, where the desired objects are human heads. It has to be mentioned that the contour initialization for the first frame in a sequence is done manually. The adopted time-window parameter L (number of past successive frames) is set to 5 for all the sequences under consideration. Additionally, the motion at the very first frame of each sequence is supposed to be zero.

Figure 10 illustrates the case of tracking an object with complicated contour (low smoothness) moving in front of a relatively smooth background. In such a case, weight w_1 of (23) and (37) is significantly greater than w_2 ($[w_1, w_2] = [10, 1]$). In this case, the desired object (aircraft) is moving towards the shooting camera and even if the object is rigid, its projection on the image plane is deforming (its contour expands) along the time.

In Figure 11, the case of car tracking in six successive frames of a traffic sequence is presented. In this example, the
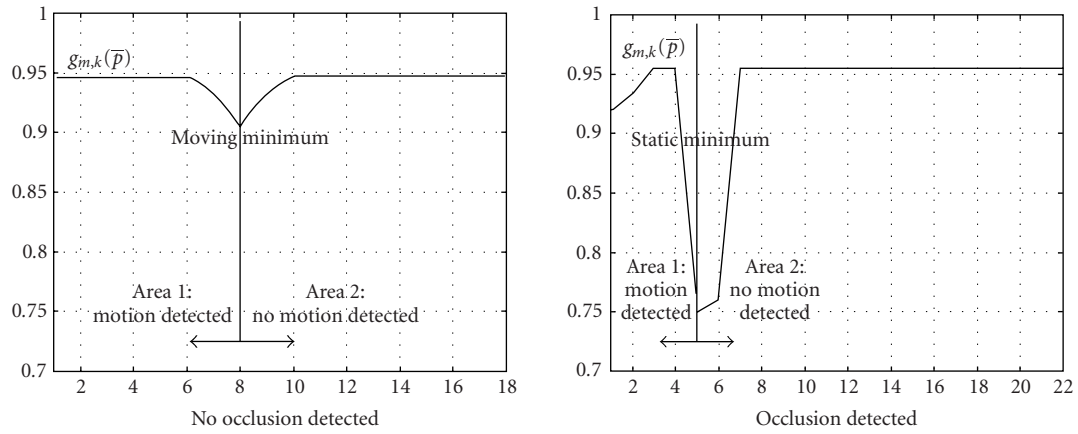
FIGURE 9: Detection of occlusion using the two rules of (41), (42), and (43) for the local minimum $\overline{p}_k$ of the function $1 - g_{m,k}(\overline{p})$.
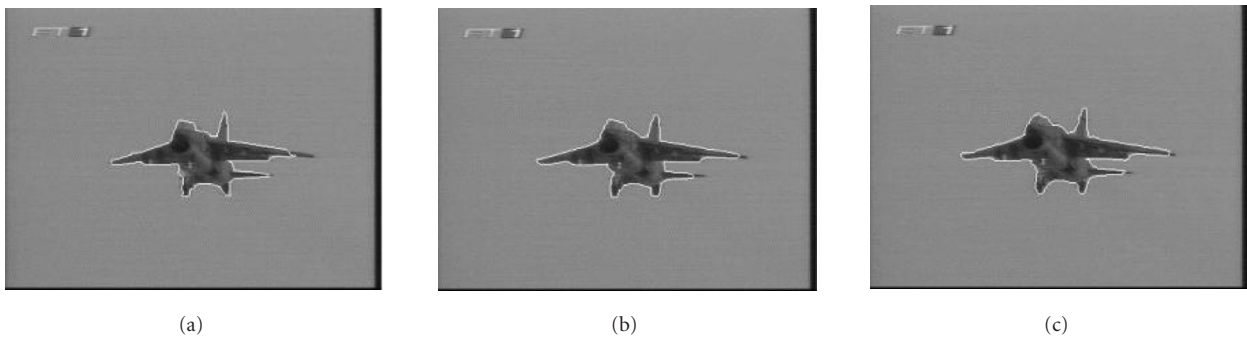

FIGURE 10: Example of tracking an aircraft approaching the airport: the case of complicated tracked contour with smooth background.
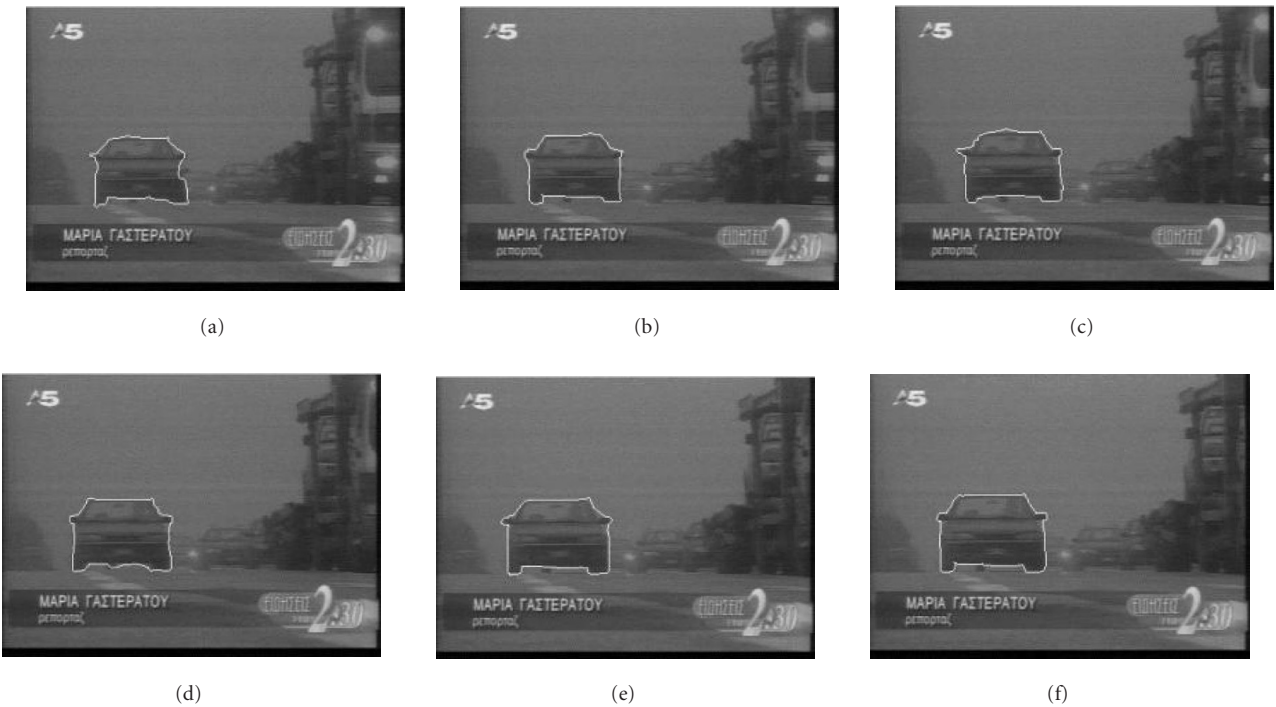

FIGURE 11: A moving car tracking in six, (a), (b), (c), (d), (e), and (f), successive frames of a traffic sequence. The background is relatively smooth close to car's boundary, while the car's contour is not very complicated.

(a)   (b)   (c)
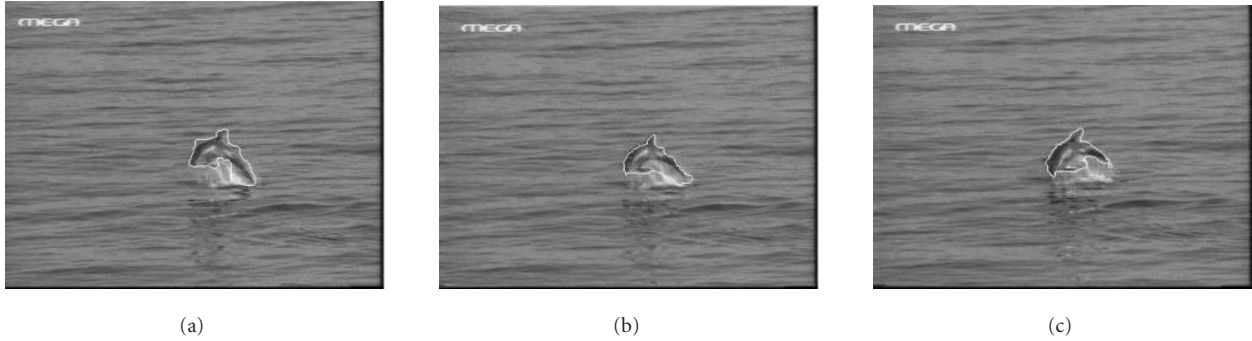
FIGURE 12: Object tracking in strong existence of clutter. For each of the transitions (a), (b) and (b), (c), the estimated weights are $[w_1, w_2] = [1, 10]$, since the background is not smooth enough.

(a)   (b)   (c)

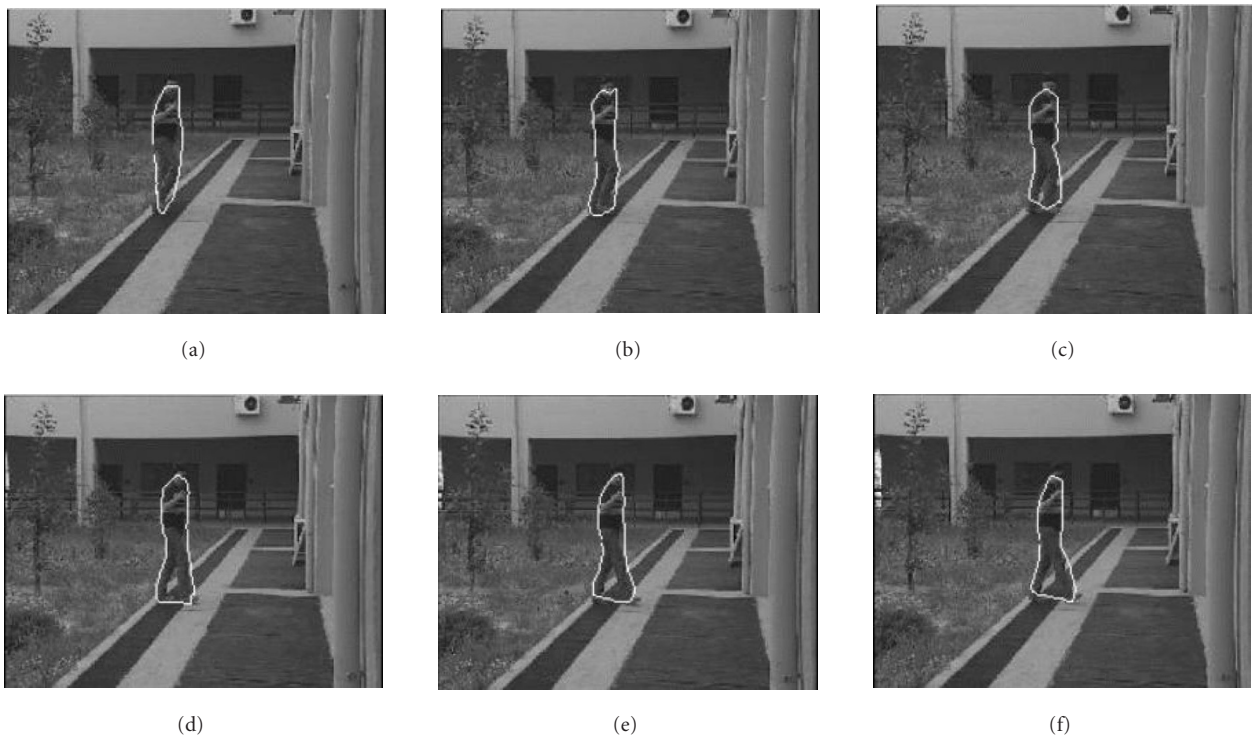(d)   (e)   (f)

FIGURE 13: Example of a man walking in a cluttered sequence. The main source of inaccuracies is the weak edges of the human body (especially close to the head). We choose $[w_1, w_2] = [1, 10]$ to reduce the effect of the background complexity.

desired object (car) is moving towards the camera and although it is rigid, its projection on the image plane is slowly deforming along time, as in the previous example. In this case, the sequence is more cluttered, although the car silhouette is smoother. The utilization of image presmoothing with the ASF and the modified image gradient, used for the external snake energy definition as described in Section 2, results in the estimation of such accurate contours.

Figure 12 illustrates the method's performance in a strongly cluttered sequence, where the object is nonrigid and its motion projection is both rotational and translational rather than a simple translation or expansion/shrink. The low accuracy of the method in this case is mainly due to the large uncertainty regions extracted. On the other hand, the object is well detected and localized in each frame of the sequence, due to the snake's external energy definition through the ASF prefiltering and the image modified gradient estimation.

In Figure 13, the proposed approach is applied to a strongly cluttered sequence, where the desired object is a man walking in one direction. The contour of the moving human body is strongly deforming along time, resulting in large uncertainty regions, whereas the weights w_1 and w_2 are estimated by values 1 and 10, respectively. The accuracy of the method is based on the snake's external energy definition and the rule-driven approach described in Section 3.4.
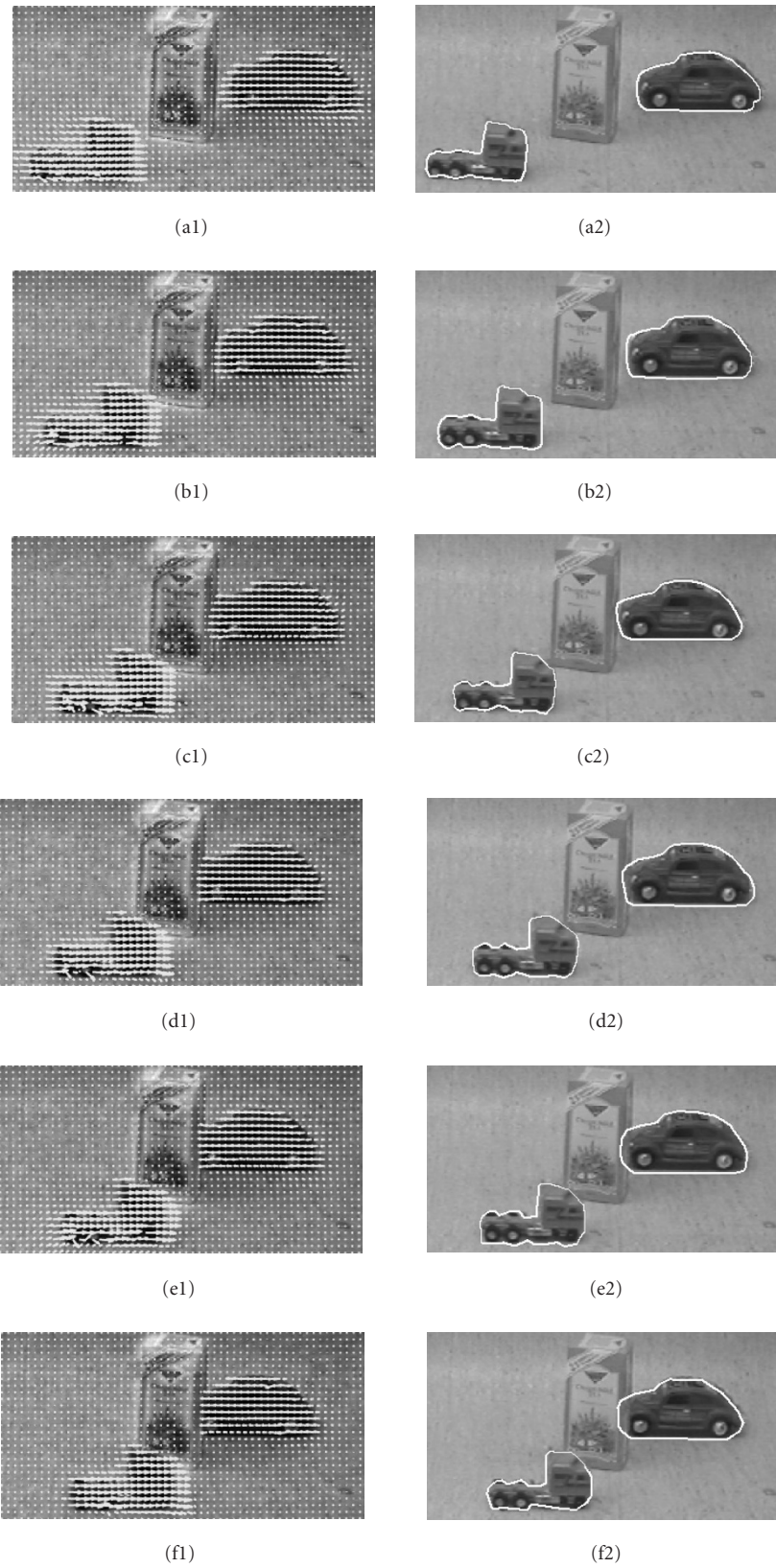
(a1)

(a2)

(b1)

(b2)

(c1)

(c2)

(d1)

(d2)

(e1)

(e2)

(f1)

(f2)

FIGURE 14: Occlusion case: motion vectors and obtained object contours utilizing the rule-driven tracking approach.

(a1)   (b1)   (c1)   (d1)

(a2)   (b2)   (c2)   (d2)

(a3)   (b3)   (c3)   (d3)

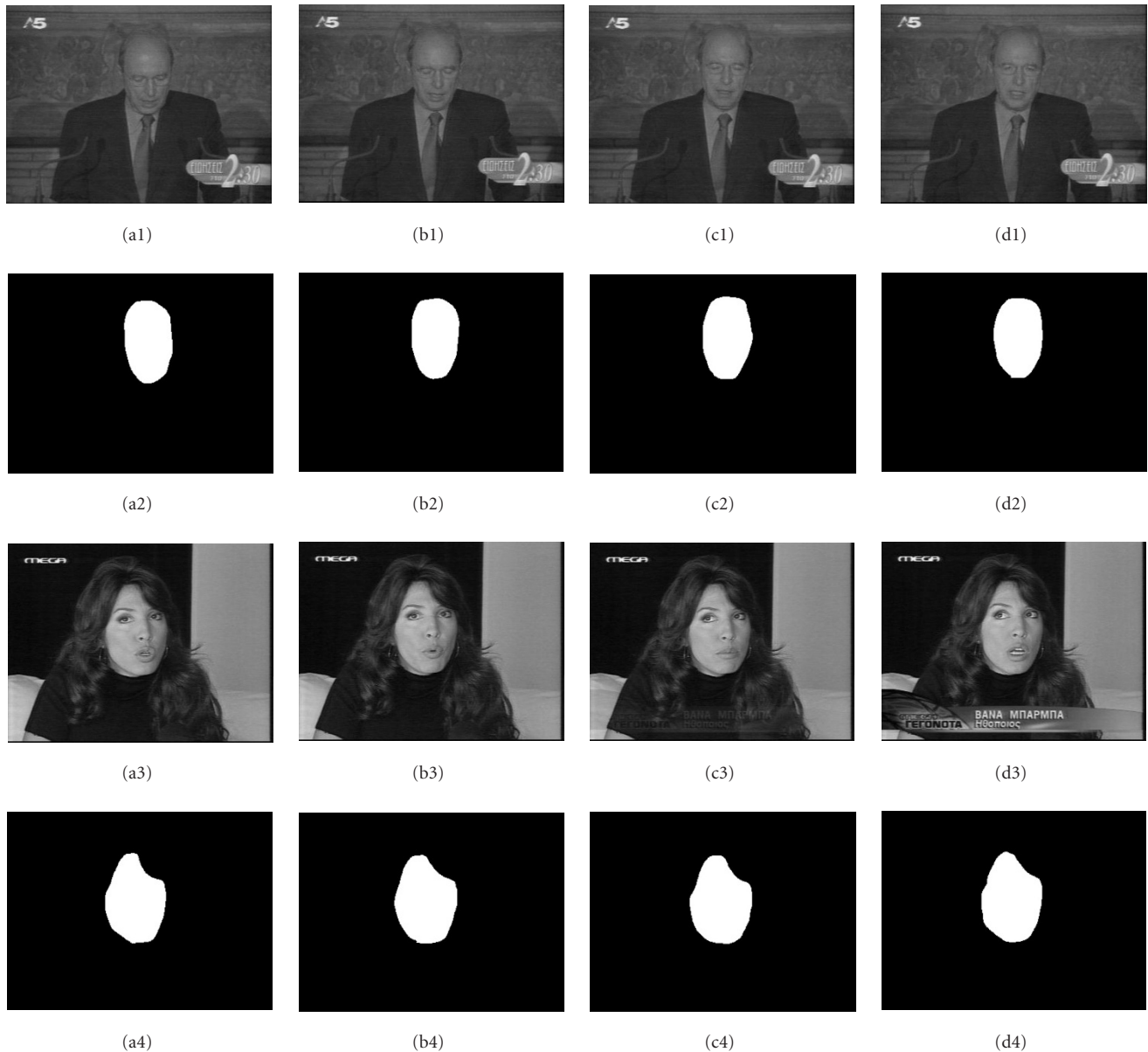(a4)   (b4)   (c4)   (d4)

FIGURE 15: Ground truth masks for selected frames of four TV sequences.

Figure 14 illustrates a case of two moving objects, where, as time goes by, the one is getting partially occluded by a static obstacle, while the other is moving in the front of the obstacle. In Figures 14a1, 14b1, 14c1, 14d1, 14e1, and 14f1, the motion estimates are illustrated, showing that the noise is effectively eliminated on the boundaries between the static and moving regions even when the occlusion occurs, whereas the respective Figures 14a2, 14b2, 14c2, 14d2, 14e2, and 14f2 show that both objects' contours are estimated with sufficient accuracy, due to the additional constraints (see (41), (42) and (43)) in which the maximum of (31) is imposed.

In order to demonstrate the efficiency of the proposed approach, we also evaluate the results of our technique

applied to head (face) extraction. Obtaining quantitative results for such an application area is hard since no extensive ground truth databases are available today. Therefore, in order to quantitatively assess the improvement of our algorithm, achieved by the addition of the geometrical model, we generated ground truth masks for available sequences.

Figure 15 shows the ground truth masks for selected frames. The presence of noise is strong since these sequences are extracted from TV clips. Consequently, the developed technique faces several difficulties not only due to occlusion/clutter cases but also due to vaguely defined object borders (weak gradient). The ground truth database consists of 100 images. Figures 16 and 17 show specific applications of

(a1)   (b1)   (c1)   (d1)

(a2)   (b2)   (c2)   (d2)

(a3)   (b3)   (c3)   (d3)

(a4)   (b4)   (c4)   (d4)

FIGURE 16: Face tracking results of the proposed rule-based method without applying the model-based term: the final results (contours) are illustrated in bold curves, whereas the shape prior (ellipse), to be further applied (Figure 17), is shown in soft curves.

the tracking method with and without the geometrical model for head contour tracking. We present representative example frames from the TV clip collection and comment on the various difficulties introduced. Even though the head is a rigid object, its contour is being deformed on the image plane along the time, due to the projection of its motion. The head contours produced by the rule-driven model-based approach (Figure 17) are obtained using an ellipsoid as shape prior.

In order to get a visual grip of the way that an ellipsoid (Section 2.1.1) affects the contour deformation, we present the results in two ways: we superimpose (a) the ellipse model (soft) and the nonmodel-based contour (Figure 16), and (b)

the model-based (bold) and the nonmodel-based contour on the original image (Figure 17). Obviously, as illustrated in Figure 16, an ellipsoid may affect the contour evolvement positively (Figures 16a1, 16b1, 16c1, and 16d1 case) since it fits well with the actual head contour, or negatively, due to strong fluctuations from the actual head shape (Figures 16a2, 16b2, 16c2, and 16d2 case—forehead area). We expect that the competitive nature of the different forces in the total energy formula will produce an acceptable result in terms of accuracy and total shape. Figure 17 shows the final results of the proposed technique that meet our expectations. The obtained contours are smooth and capture accurately the side

FIGURE 17: Face tracking results of the proposed method after applying the model-based (shape prior) term: the final contours are illustrated in bold curves, whereas the results obtained without the shape prior (Figure 16), are shown in soft curves.

regions of the head, as shown clearly in Figures 17a1, 17b1, 17c1, and 17d1; and 17a2, 17b2, 17c2, and 17d2. The actual head contour of Figures 17a4, 17b4, 17c4, and 17d4 case is fairly different than the corresponding "perfect" ellipsoid (Figures 17a4, 17b4, 17c4, and 17d4 case) since our ground truth generation is merely based on skin presence. However, the final model-based contour seems to be much closer to the shape of a head and therefore it can be treated as a better way to verify the presence of a human head in a sequence even in the case of partial occlusion (due to hair in the considered case).

Additionally, we provide a table with precision and recall measurements for the selected sequences in order to verify the improvement imposed by the model-based approach. As mentioned before, we extracted the ground truth masks based on skin activation. Consequently, we expect slightly lower precision than recall values, since areas covered, for example, by hair may be considered as "head," but they are not included in the ground truth masks. Table 1 compares the recall/precision values for the 16 selected frames used in Figures 16 and 17 and gives the overall values for the 100 images we tested.

TABLE 1: Recall and precision values, comparing the ground truth results with the ones obtained using the proposed method with and without the shape prior model.

Frame	Recall percentage (no model)	Precision percentage (no model)	Recall percentage (model)	Precision percentage (model)
a1	89.08	100.00	94.52	100.00
b1	87.01	100.00	100.00	97.44
c1	90.09	100.00	98.17	100.00
d1	90.09	100.00	98.17	100.00
a2	98.38	100.00	100.00	99.54
b2	98.63	100.00	95.14	100.00
c2	98.76	100.00	95.17	100.00
d2	100.00	96.68	100.00	99.52
a3	100.00	91.04	100.00	99.78
b3	100.00	94.18	95.55	100.00
c3	100.00	91.64	98.48	100.00
d3	100.00	91.26	99.62	100.00
a4	100.00	93.88	100.00	93.90
b4	100.00	97.29	100.00	91.60
c4	100.00	98.12	100.00	84.85
d4	100.00	97.09	100.00	97.83
Total	95.05	93.16	97.24	95.84

5. CONCLUSIONS AND FURTHER WORK

In this work, we have presented a probabilistic application of snakes for object tracking in clutter, partial occlusion, and complex backgrounds. Statistical measurements of the object contour motion history are extracted to obtain uncertainty regions, in which the estimated contours are to be localized. In this way, we constrain the solution in a narrow band around the next frames snake initialization. Moreover, utilizing various tools from image morphology, we eliminate noise. This approach is extended to cope with complex background and partial occlusion, introducing rule-based knowledge to separate objects from background and to detect occlusion. Finally, for specific applications where the desired object contours can be approximated by specific models, we use a shape prior knowledge in addition to the rule-driven approach so as to obtain more accurate contours. As indicated before, in this work our goal is to illustrate the increased robustness of the proposed method with the addition of a model rather than incorporating a novel prior constraint representation. Therefore, the future direction of our work is a more sophisticated representation and use of generalized geometric-based models, which will permit the method to deal even more efficiently with occlusions and perform tracking under various conditions (e.g., static or mobile camera). In this sense, a possible extension can be the incorporation of region-based tracking modules to the existing framework that will increase robustness. Additionally, covering large occlusions cases would require extensions of the method, for example, using higher-level representation of the moving regions, which is a topic of future research. We are currently examining such issues using "semantics" and ontological knowledge techniques in the framework of [33].

REFERENCES

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

[2] S. R. Gunn and M. S. Nixon, "A robust snake implementation: a dual active contour," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 1, pp. 63–68, 1997.

[3] L. D. Cohen and I. Cohen, "Finite-element methods for active contour models and balloons for 2-d and 3-d images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1131–1147, 1993.

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

[5] A.-R. Mansouri, T. Chomaud, and J. Konrad, "A comparative evaluation of algorithms for fast computation of level set PDEs with applications to motion segmentation," in *Proc. of International Conference on Image Processing (ICIP '01)*, vol. 3, pp. 636–639, Thessaloniki, Greece, October 2001.

[6] V. Caselles, R. Kimmel, G. Sapiro, and C. Sbert, "Minimal surfaces: a geometric three dimensional segmentation approach," *Numerische Mathematik*, vol. 77, no. 4, pp. 423–425, 1997.

[7] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.

[8] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, pp. 266–280, 2000.

[9] N. Paragios and R. Deriche, "A PDE-based level set approach for detection and tracking of moving objects," in *Proc. 6th International Conference on Computer Vision (ICCV '98)*, pp. 1139–1145, Bombay, India, January 1998.

[10] C. Vieren, F. Cabestaing, and J.-G. Postaire, "Catching moving objects with snakes for motion tracking," *Pattern Recognition Letters*, vol. 16, no. 7, pp. 679–685, 1995.

[11] M. Daoudi, F. Ghorbel, A. Mokadem, O. Avaro, and H. Sanson, "Shape distances for contour tracking and motion estimation," *Pattern Recognition*, vol. 32, no. 7, pp. 1297–1306, 1999.

[12] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 359–369, 1998.

[13] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[14] M. Israd and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *Proc. European Conf. on Computer Vision (ECCV '96)*, vol. 1, pp. 343–356, Cambridge, UK, 1996.

[15] M. Israd and A. Blake, "ICONDENSATION: unifying low-level and high-level tracking in a stochastic framework," in *Proc. 5th European Conf. Computer Vision*, vol. 1 of *Lecture Notes in Computer Science*, pp. 893–908, 1998.

[16] S. Rousson and N. Paragios, "Shape priors for level set representations," in *European Conference in Computer Vision*, pp. 78–93, Copenhagen, Denmark, June 2002.

[17] P. Delagnes, J. Benois, and D. Barba, "Active contours approach to object tracking in image sequences with complex background," *Pattern Recognition Letters*, vol. 16, no. 2, pp. 171–178, 1995.

[18] H. H. S. Ip and D. Shen, "An affine-invariant active contour model (AI-Snake) for model-based segmentation," *Image and Vision Computing*, vol. 16, no. 2, pp. 135–146, 1998.

[19] J. S. Park and J. H. Han, "Contour motion estimation from image sequences using curvature information," *Pattern Recognition Letters*, vol. 31, no. 1, pp. 31–39, 1998.

[20] Y. Avrithis, Y. Xirouhakis, and S. Kollias, "Affine-invariant curve normalization for object shape representation, classification and retrieval," *Machine Vision and Applications*, vol. 13, no. 2, pp. 80–94, 2001.

[21] F. Mohanna and F. Mokhtarian, "Improved curvature estimation for accurate localization of active contours," in *Proc. International Conference on Image Processing (ICIP '01)*, vol. 2, pp. 781–784, Thessaloniki, Greece, October 2001.

[22] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 133–144, 1992.

[23] L. H. Staib and J. S. Duncan, "Boundary finding with parametrically deformable models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1061–1075, 1992.

[24] S. R. Gunn and M. S. Nixon, "A model based dual active contour," in *Proc. British Machine Vision Conference*, E. Hancock, Ed., pp. 305–314, York, UK, 1994.

[25] P. Maragos, "Noise suppression," in *The Digital Signal Processing Handbook*, V. K. Madisetti and D. B. Williams, Eds., pp. 20–26, CRC Press, Boca Raton, Fla, USA, 1998.

[26] G. Tsechpenakis, N. Tsapatsoulis, and S. Kollias, "Snake modifications for object tracking in cluttered sequences: a probabilistic approach," submitted to Image Communication, http://www.image.ntua.gr/~gtsech/.

[27] C. L. Lam and S. Y. Yuen, "An unbiased active contour algorithm for object tracking," *Pattern Recognition Letters*, vol. 19, no. 5-6, pp. 491–498, 1998.

[28] J. M. Odobez and P. Bouthemy, "Separation of moving regions from background in an image sequence acquired with a mobile camera," in *Video Data Compression for Multimedia Computing*, H. H. Li, S. Sun, and H. Derin, Eds., pp. 283–311, Kluwer Academic Publishers, Boston, Mass, USA, 1997.

[29] F. G. Meyer and P. Bouthemy, "Region based tracking using affine motion models in long image sequences," *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 60, no. 2, pp. 119–140, 1994.

[30] N. Paragios and R. Deriche, "Geodesic active regions for motion estimation and tracking," in *Proc. 7th IEEE International Conference in Computer Vision (ICCV '99)*, Kerkyra, Greece, September 1999.

[31] M. J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[32] S. Haykin, "The steepest descent method," in *Neural Networks: A Comprehensive Foundation*, pp. 124–126, Macmillan College Publishing Company, New York, NY, USA, 1994.

[33] EU FP6 Network of Excellence, "Knowledge web," 2003–2007.

Gabriel Tsechpenakis was born in Athens, in 1975. He graduated from the School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), in 1999, and obtained his Ph.D. degree in 2003 from the same university. His Ph.D. was carried out in the Image, Video and Multimedia Systems Laboratory of NTUA. His current research interests focus on the fields of computer vision, machine learning, and human computer interaction. He is a Member of the Technical Chambers of Greece and the IEEE Signal Processing Society. He is currently a Postdoctoral Associate at the Center for Computational Biomedicine, Imaging and Modeling (CBIM) in Rutgers, The State University of New Jersey.

Konstantinos Rapantzikos received the Diploma and M.S. degree in electronics & computer engineering from the Technical University of Crete, Greece, in 2000 and 2002, respectively. He is currently working towards the Ph.D. degree in electrical engineering at the National Technical University of Athens, Greece. His thesis will be on computational modeling of human vision. His interests also include biomedical image processing and motion estimation in compressed/uncompressed video.

Nicolas Tsapatsoulis was born in Limassol, Cyprus, in 1969. He graduated from the Department of Electrical and Computer Engineering, National Technical University of Athens in 1994 and received his Ph.D. degree in 2000 from the same university. His current research interests lie in the areas of human computer interaction, machine vision, image and video processing, neural networks, and biomedical engineering. He is a Member of the Technical Chambers of Greece and Cyprus and a Member of IEEE Signal Processing and Computer Societies. Dr. Tsapatsoulis has published 13 papers in international journals and more than 35 in proceedings of international conferences. He served as Technical Program Cochair for the VLBV '01 workshop. He is a Reviewer of the IEEE Transactions on Neural Networks and IEEE Transactions on Circuits and Systems for Video Technology journals.
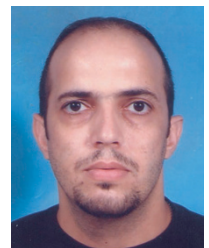
Stefanos Kollias was born in Athens in 1956. He obtained his Diploma from the National Technical University of Athens (NTUA) in 1979, his M.S. in communication engineering in 1980 from the University of Manchester, Institute of Science and Technology (UMIST), England, and his Ph.D. degree in signal processing from the Computer Science Division of NTUA. He is with the Electrical Engineering Department of NTUA since 1986, where he serves now as a Professor. Since 1990 he is the Director of the Image, Video and Multimedia Systems Laboratory of NTUA. He has published more than 120 papers in the above fields, 50 of which in international journals. He has been a Member of the technical or advisory committee or invited speaker in 40 international conferences. He is a Reviewer of 10 IEEE Transactions and of 10 other journals. Ten graduate students have completed their doctorate under his supervision, while another ten are currently performing their Ph.D. theses. He and his team have been participating in 38 European and national projects.