Utilization of a Snake Model for Object Tracking in Natural Sequences

G. Tsechpenakis, K. Rapantzikos, N. Tsapatsoulis and S. Kollias

National Technical University of Athens
School of Electrical and Computer Engineering
Tel: +30210 7723037, Fax: +30210 7722488
email: ntsap@image.ntua.gr

Abstract

Tracking moving objects in video sequences is a task that emerges in various fields of study: video analysis, computer vision, biomedical systems etc. In the last decade, special attention has been drawn to problems concerning tracking in real-world environments, where moving objects do not obey any afore-known constraints about their nature and motion or the scenes they are moving in. Apart from the existence of noise and environmental changes, many problems are also concerned, due to background texture, complicated object motion, and deformable and/or articulated objects, changing their shape while moving along time. Another phenomenon in natural sequences is the appearance of occlusions between different objects, whose handling requires motion information and, in some cases, additional constraints. In this work, we revisit one of the most known active contours, the Snakes, and we propose a motion-based utilization of it, aiming at successful handling of the previously mentioned problems. The use of the object motion history and first order statistical measurements of it, provide us with information for the extraction of uncertainty regions, a kind of shape prior knowledge w.r.t. the allowed object deformations. This constraining also makes the proposed method efficient, handling the trade-off between accuracy and computation complexity. The energy minimization is approximated by a force-based approach inside the extracted uncertainty regions, and the weights of the total snake energy function are automatically estimated as respective weights in the resulting evolution force. Finally, in order to handle background complexity and partial occlusion cases, we introduce two rules, according to which the moving object region is correctly separated from the background, whereas the occluded boundaries are estimated according to the object's expected shape. To verify the performance of the proposed method, some experimental results are included, concerning different cases of object tracking, indoors and outdoors, with rigid and deformable objects, noisy and textured backgrounds, as well as appearance of occlusions.

Key words: Active contours, Snakes, Tracking, Cluttered sequences, Occlusion

1 Introduction

This work addresses the problem of object contour modeling and tracking in natural video sequences. Specifically, given a video stream in which a known object of interest is in motion, the goal is to track the object's silhouette across time varying images. Applications abound in computer vision, video processing, including video analysis and understanding, object-based coding, content-based retrieval, remote surveillance, object recognition, as well as various biomedical problems, such as human organ motion analysis and tracking. According to the application examined, object tracking includes a number of problems that emerge, making its successful performance a very challenging task, especially regarding its generality and its independence from initial constraints. Moreover, during the last decade, object tracking has become one of the most important tasks, due to the information that new technologies require. A representative example of these technologies, are the MPEG-4 and MPEG-7 standards for video coding, according to which object localization along time is needed, in order to separate the front from the background, and handle them separately. Contrary to the MPEG-7 standard and many other applications, such as surveillance for security systems, MPEG-4 and a wide variety of applications require the exact contours of the moving objects, for shape description and recognition purposes. In this framework, and depending on the examined application, many approaches have been proposed in the literature, focusing either in the highest accuracy, or the lowest computational complexity. A great category of these approaches is the deformable templates [12] called active contours, that have come up and been drawn special attention, due to their performance in various problems, such as image and motion segmentation [4,14,3,20], object tracking [22,21,27,5] and, similarly to these approaches, audiovisual speech recognition.

Active contours were first introduced by Kass et al. in 1988, with a model called Snake [13], and since then many researchers have proposed various modifications of it, in order to apply it to edge detection, shape modeling and object tracking problems, under different constraints. However, many problems concerning these models arise, due to strong existence of noise, mostly in natural cluttered sequences [26], the requirement of an appropriate shape initialization [28] and parameter tuning [19]. Apart from these problems which are met due to the definition of the snake models, and thus can be considered as obvious and well known, there are also some cases that are not so far handled by Snakes: (a) the performance of snakes in sequences with complex backgrounds, i.e textured backgrounds containing strong edges close to the moving object boundaries, and (b) problems met when the desired moving objects get occluded by obstacles, which may be static or moving as well. These are mainly the reasons why many researchers have proposed other active contour models [22,11,20], utilizing region-based information, such as motion, color and texture, stochastic approaches and appropriate shape constraints [23].

In this work we focus on the above mentioned problems, i.e existence of noise, background complexity and occlusion handling, with a modified snake model which utilizes the object's motion history. Using the snake, which is a linear and thus computationally inexpensive model, and first order statistics, we try to balance between the computational cost of the proposed method, and its accuracy. The latter is needed, mainly due to some time-consuming morphological operations applied in each frame of the examined sequence. The proposed model is relatively robust to parameter tuning, while the required initialization of the snake in each frame is automatically extracted, along with a narrow band around it, in which the final solution (object contour) is located. Thus, the proposed method consists of three main steps: (a) the frame pre-processing, using a morphological filter, and the extraction of a morphological modified image gradient, (b) the extraction of the snake initialization and its "uncertainty regions", in which the final solution is located, and (c) the snake energy minimization procedure, which is approximated by a force-based curve evolution approach, inside the extracted uncertainty regions. Finally, a motion estimation scheme is used to help handling problems related to object occlusion.

Special attention has been given to the accuracy of the proposed approach, resulting in efficient solutions to a variety of applications: cases involving different classes of objects and object movements in natural sequences, where the amount of noise is not known, backgrounds that are highly textured, and moving objects that get successively occluded.

2 Snake Model

Snakes are edge-based models widely used for object shape modeling, in single images, and for tracking, in image sequences. So far, they have been successfully applied to problems where shape prototypes of the desired objects are given. In general, snakes concern model and image data analysis, through the definition of a linear energy function and a set of regularization parameters. This energy function consists of two parts: (a) the data-driven component (external energy), which depends on the image data according to a chosen criterion and (b) the smoothness-driven one (internal energy), which enforces smoothness along the snake. The goal is to minimize the total snake energy; this is achieved iteratively, after considering an initial estimate for 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 the above 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.

Following the definition of the snakes and the general idea they adopt, many problems arise, especially when examining cluttered natural sequences. The first concerns the snake energy function and the snake's convergence according to the energy minimization procedure: (a) there is a set of regularization parameters, used as weights in the total function, that need to be set in appropriate values, (b) an appropriate snake initialization around the desired object has to be defined manually, (c) the object boundaries have to be distinct, that is the respective edges have to be strong, and the background has to be relatively smooth, so that the energy minimization procedure is not trapped in undesirable local minima (edges) other than the object boundaries. These requirements are very difficult to be met in cluttered natural sequences, where weak object edges, the existence of noise, highly textured background or strong and persistent (along the time) edges close to the desirable object boundaries, are common and very frequent phenomena. Moreover, in real-world scenes, it often happens that a moving object of interest gets successively occluded by another static or moving object. In these cases, there is lack of necessary edge information, and thus snake models cannot be efficiently applied, without the exploitation of an appropriate shape prior knowledge.

Let us consider a snake representing a curve defined by a position vector $\underline{X}(p) = (x(p), y(p))$ on the image plane, and is generally parameterized by p, $0 \leq p \leq P$. The total energy function of the snake E_{snake}, is then a weighted summation of an internal energy factor E_{int}, corresponding to the summation of a bending and a stretching energy term, and an external energy factor E_{ext}, which denotes how the snake evolves according to the examined image features:

$$E_{snake} = a \cdot E_{int} + b \cdot E_{ext}, \quad E_{int} = \int_0^P e_{int}(p)dp, \quad E_{ext} = \int_0^P e_{ext}(p)dp, \qquad (1)$$

where $e_{int}(p)$ and $e_{ext}(p)$ are the respective internal and external energy terms defined for each point p of the curve, whereas a and b are normalization parameters.

For the internal energy E_{int}, a variety of definitions have been proposed, according to the examined application; indicatively three approaches are mentioned: the B-snake [6], where the curve is approximated by B-spline polynomials, the affine-invariant (AI-) snake [10] and the curvature-based snake [24] models. On the contrary, in most of the approaches, the external energy term at each point p is generally defined as [10]:

$$e_{ext}(p) = 1 - |\nabla G_\sigma * I(x(p), y(p))| \cdot |\underline{g}(p) \cdot \underline{n}(p)|, \qquad (2)$$

where $|\nabla G_\sigma * I(x(p), y(p))|$ (Gaussian-of-Laplacian) denotes the magnitude of the gradient of the image I ($I(x, y) \in [0, 1]$) convolved with a gaussian filter, of variance σ at point $(x(p), y(p))$, corresponding to the curve point p. The unit vectors $\underline{g}(p)$ and $\underline{n}(p)$ denote the image gradient direction and the normal vector of the snake at point p, respectively.

The definition of the external energy term, given in (2), is supposed to have the following effect: smooth regions of the examined image I, where its gradient has relatively low values, produce external energy values close to 1, whereas image edges produce external energy values close to 0. In this sense, the minimization procedure of the total energy function, forces the snake towards the edges closer to its initialization. This definition denotes that (a) the regions close to object boundaries must be smooth without significant edges and noise, that could create undesired minima, (b) if the background is noisy, then an appropriate value for the parameter σ of the gaussian filter must be manually chosen. These are mainly the reasons why it is not easy to use this definition for the external energy in natural cluttered sequences.

2.1 Internal Energy Definition

In this work, we adopt a snake model that utilizes the curvature K_{snake} [19,24] and the point density distribution DV_{snake}:

$$K_{snake} = \frac{\dot{x} \cdot \ddot{y} - \ddot{x} \cdot \dot{y}}{(\dot{x}^2 + \dot{y}^2)^{\frac{3}{2}}} \quad , \quad DV_{snake} = \sqrt{\dot{x}^2 + \dot{y}^2}, \tag{3}$$

assuming that that the snake points are not equally spaced along the curve. The first and second derivatives of (x, y) define the velocity and the acceleration, along the curve:

$$\underline{\dot{X}} = [\dot{x}, \dot{y}] = [\frac{dx}{dp}, \frac{dy}{dp}] \quad , \quad \underline{\ddot{X}} = [\ddot{x}, \ddot{y}] = [\frac{d^2x}{dp^2}, \frac{d^2y}{dp^2}], \tag{4}$$

and for each point p of the snake, they are calculated in terms of the neighboring points $p - 1$ and $p + 1$:

$$\dot{x}(p) = x(p + 1) - x(p - 1), \tag{5}$$

$$\dot{y}(p) = y(p + 1) - y(p - 1), \tag{6}$$

$$\ddot{x}(p) = x(p - 1) - 2x(p) + x(p + 1), \tag{7}$$

$$\ddot{y}(p) = y(p - 1) - 2x(p) + y(p + 1) \tag{8}$$

The curvature and the point density distribution functions are not affine-invariant, as can be easily proved, and can uniquely define a curve in specific time instances, whereas curvature denotes whether a part of a curve is convex ($K_{snake} > 0$) or concave ($K_{snake} < 0$). In our model, the curvature function is used to provide us with a local smoothness constraint, since values of K close to 0 denote smooth parts of the snake, but also a similarity criterion in the tracking approach as described in Section 3. The point density function is the snake local elasticity constraint, since it represents the distances between neighboring points along the snake. From eqs. (3) and (4) it is $DV_{snake} = |\underline{\dot{X}}|$. The internal energy of the proposed snake model is then,

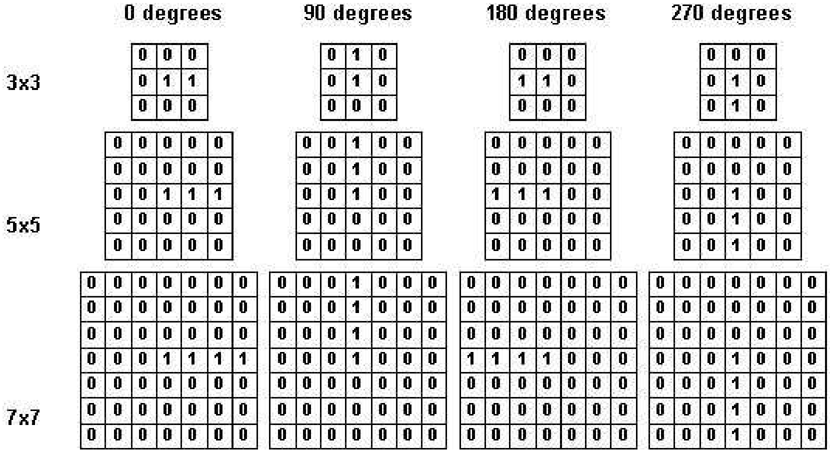
$$e_{int}(p) = |K_{snake}(p)| + DV_{snake}(p), \tag{9}$$

where $|\cdot|$ denotes the magnitude sign. Thus, the minimization of the internal energy forces the snake into a smoother form.

2.2 External Energy Definition

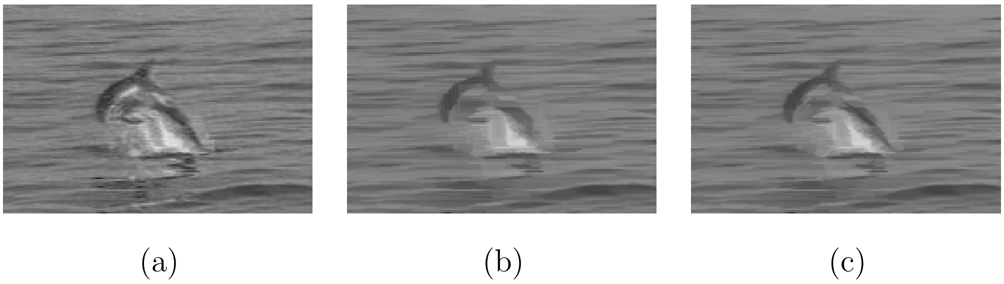
Regarding the external energy component, we use an alternative definition, in order to deal with noise, textured backgrounds and edges close to object boundaries. The proposed procedure extracts an image-based component, after appropriately pre-smoothing the examined image, in order to eliminate randomly distributed noise, and it provides a good alternative way for overcoming the problems that appear in definition (2); the proposed method uses various morphological operations that lead to a modified image gradient. The use of morphological operations may also eliminate object boundary edges, in cases that some of these edges are not well distinguishable, and thus force the snake into a more complicated form than the desired one. This problem is overcome with the use of the internal energy constraint, which does not allow great local deformations of the snake. Moreover, some background edges may remain even after the morphological procedures are applied, but this problem is overcome with the use of further constraints, as described in 3.2.3.
The first step is the image pre-smoothing with a non-linear morphological filter, a so-called ASF (Alternating Sequential Filter). This filter is based on morphological area opening ($\circ$) and closing ($\bullet$) operations with structure elements of increasing scale [15]. More specifically, if S_i, i=1,...,4, is the 2-pixel connected line segments oriented at 90(i-1) degrees, and nS_i is the corresponding (n+1)-pixel elements of size $n = 1, 2, 3, ...$, then the openings α_n and closings β_n that make up the filter, for an image I, are,

$$\alpha_n(I)(x,y) = \max_{i\in[1,4]} [I \circ nS_i(x,y)] \;\;,\;\; \beta_n(I)(x,y) = \min_{i\in[1,4]} [I \bullet nS_i(x,y)] \tag{10}$$



	0 degrees	90 degrees	180 degrees	270 degrees
3x3	$\begin{matrix}0&0&0\\0&1&1\\0&0&0\end{matrix}$	$\begin{matrix}0&1&0\\0&1&0\\0&0&0\end{matrix}$	$\begin{matrix}0&0&0\\1&1&0\\0&0&0\end{matrix}$	$\begin{matrix}0&0&0\\0&1&0\\0&1&0\end{matrix}$
5x5	$\begin{matrix}0&0&0&0&0\\0&0&0&0&0\\0&0&1&1&1\\0&0&0&0&0\\0&0&0&0&0\end{matrix}$	$\begin{matrix}0&0&1&0&0\\0&0&1&0&0\\0&0&1&0&0\\0&0&0&0&0\\0&0&0&0&0\end{matrix}$	$\begin{matrix}0&0&0&0&0\\0&0&0&0&0\\1&1&1&0&0\\0&0&0&0&0\\0&0&0&0&0\end{matrix}$	$\begin{matrix}0&0&0&0&0\\0&0&0&0&0\\0&0&1&0&0\\0&0&1&0&0\\0&0&1&0&0\end{matrix}$
7x7	$\begin{matrix}0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&1&1&1&1\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\end{matrix}$	$\begin{matrix}0&0&0&1&0&0&0\\0&0&0&1&0&0&0\\0&0&0&1&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\end{matrix}$	$\begin{matrix}0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\1&1&1&1&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\end{matrix}$	$\begin{matrix}0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&1&0&0&0\\0&0&0&1&0&0&0\\0&0&0&1&0&0&0\end{matrix}$

Fig. 1. *Structure elements of increasing scale ($n = 1, 2, 3$) for ASF construction*



(a) (b) (c)

Fig. 2. *Frame pre-smoothing with the proposed Alternating Sequential Filter: (a)original frame, (b) filtered frame with structure elements of 4 orientations, and (c) filtered frame with structure elements of 8 orientations.*

Then the filtered image I_{ASF} is obtained by the following cascade,

$$I_{ASF} = \beta_n \alpha_n ... \beta_2 \alpha_2 \beta_1 \alpha_1 \qquad (11)$$

In our method, we use that filter to preserve the line-type features (edges) of each frame of a sequence, without smoothing them, which cannot be achieved with, e.g., median filtering. Fig. 1 illustrates an example of structure elements of different orientations (0^o, 90^o, 180^o and 270^o) and increasing scale (3x3, 5x5 and 7x7), used for the construction of the ASF utilized in our implementation. Also, Fig. 2 illustrates the performance of image smoothing using the proposed ASF: (a) is the original image and (b) is the filtered image; it can be clearly seen that a great amount of clutter is eliminated. For a better implementation of the ASF, we propose the use of structure elements of 8 orientations, i.e 0^o, 45^o 90^o, 135^o, 180^o, 225^o, 270^o and 315^o, in order to preserve edges in a variety of orientations. The result of this implementation is illustrated in Fig. 2(c).

After filtering each frame of the examined sequence, we follow a procedure aiming at preserving the most important regions, and thus the most important

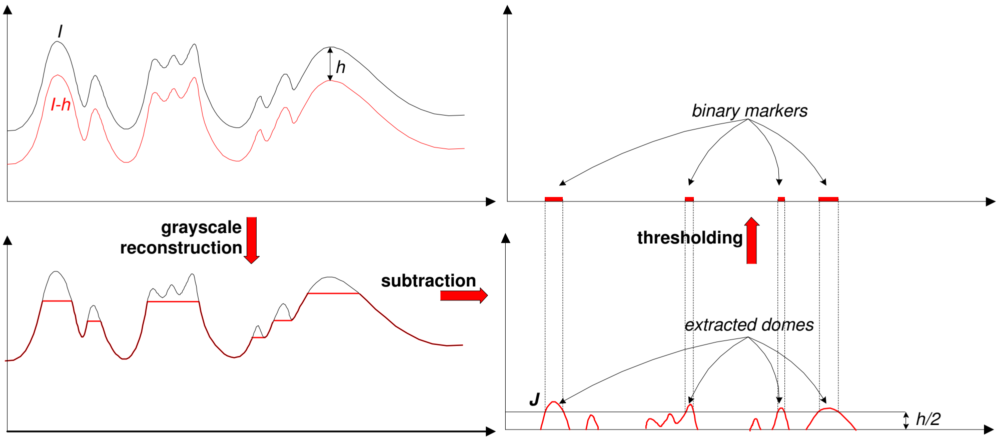


Fig. 3. *Binary markers extraction for modified gradient calculation: subtraction of a constant h from the initial image, grayscale reconstruction opening, local maxima of the image, and binary markers extraction using $h/2$ as a threshold.*

edges of the frame. This procedure is a part of the watershed transformation [16], and involves the extraction of binary markers of the most important regions. These markers are appropriately combined with the image gradient so that its variations, due to noise or weak edges, will be eliminated.

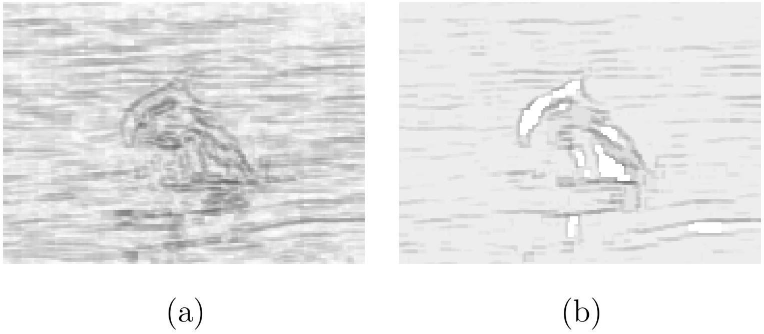
More specifically, as shown in Fig. 3, we subtract a constant h from the filtered image I_{ASF}, and we apply a morphological grayscale reconstruction opening between I_{ASF} and $I_{ASF} - h$. The result of this procedure is subtracted from I_{ASF}, and what we obtain is an image representing the local maxima of I_{ASF}; then, the most important maxima are obtained by thresholding this image with $h/2$. The same procedure is followed to extract the binary markers that represent the image most important minima; combining the binary marker images into one, a geodesic erosion reconstruction of image gradient G is computed,

$$G_m = \lim_{k \to \infty} [(m \ominus B) \bigvee G]^{(k)} \tag{12}$$

where m is the binary markers image, B is a symmetric structuring element of radius 1, k is the number of the successive operations, and "$\ominus$" and "$\bigvee$" denote the flat erosion and supremum operations, respectively [17].

In our implementation, we normalize the image intensity in the interval $[0, 1]$. Thus, a reasonable choice of the constant h is $0.1 \leq h \leq 0.3$; it is experimentally proved that the value of h is not crucial for the extraction of the image modified gradient. It must be also noted that instead of the gradient ∇I_{ASF}, we use the morphological image gradient, defined as $G = [s \oplus I_{ASF}] - [s \ominus I_{ASF}]$, where $\oplus$ and $\ominus$ are the dilation and erosion operations, and s is a 3×3 structure element.

After the modified image gradient G_m is extracted, the external snake energy



(a) (b)

Fig. 4. *Differences between (a) the image gradient* $\nabla G_{(\sigma=1)} * I$ *and (b) the proposed modified morphological image gradient* G_m.

is defined as,

$$e_{ext}(p) = 1 - G_m^2(p) \cdot |\underline{g}_m(p) \cdot \underline{n}(p)|, \tag{13}$$

where $\underline{g}_m(p)$ and $\underline{n}(p)$ denote the modified image gradient unit vector and the normal vector of the snake at the point p, respectively. Fig. 4 illustrates the differences between (a) the image gradient $\nabla G_\sigma * I$ ($\sigma = 1$) of eq. (2) and (b) the proposed modified gradient G_m of eq. (13), and it can be clearly seen that a great amount of noise and weak edges is eliminated.

3 Proposed Tracking Method

In object tracking, the goal is to track object's contour along time; this actually means that, for each frame of a video sequence, we aim at separating the object from the background. During the last decade, many approaches for contour tracking have been proposed in the literature, which can be roughly categorized in three main classes: (a) edge-based methods [11], utilizing the edge information of the examined images, (b) region-based methods [18], that rely on the information obtained by region features (color, texture etc.), and (c) combinations of edge and region-based methods [22], which exploit and combine the advantages of the previous two classes, formulating either active contours or geometric alternatives of them, such as the Level-Set theory. Amongst these methods, there are approaches relying on grouping motion information along time [18], introducing additional constraints in terms of high level semantic knowledge [7], or utilizing prior knowledge for the desired object's shape [23].
The issue of object tracking in natural and cluttered sequences, involves various problems, especially when its most general confrontation is needed. The generality of a method is one of the most important goals, which means that we have to deal with as few as possible initial constraints, or even make the tracking problem independent from initial conditions/constraints. The difficulties

that usually arise in real-world scenes are: (a) non-rigid (or articulated) moving objects, (b) moving objects with complicated contours, (c) object motions that are not simple translations, but also involve rotations, objects approaching or drawing away from the shooting camera (or camera zooming), (d) sequences with highly textured background, (e) existence of noise or abrupt/gradual environmental (external) changes, such as external lighting changes, (f) moving objects that get successively occluded by obstacles, that can be either static or moving.

In this Section, we describe the proposed tracking method that utilizes the snake model defined in the previous Section, and exploits the results of a motion estimation scheme proposed in [2]. Our method consists of three main steps: (a) estimation of the snake's initial position in frame $I+1$, given its position in the previous frame I, using a robust motion estimation technique, (b) definition of a narrow band around the initialization, in which the object contour is supposed to be located, and (c) estimation of the contour's final position in frame $I+1$. The narrow band around the object is called "uncertainty region", and this term is used to account for the uncertainty factor that is involved in estimating the shape and the position of a moving object in a scene. In this sense, the problem of finding the correct contour is constrained in a small image region, which allows us to reduce the computational cost.

3.1 Snake Initialization and Uncertainty Regions Extraction

Let us consider the contour $\mathbf{C}^{(I)}$ of the desired object in frame I of a video sequence, defined by P ordered points on the image plane, in terms of complex numbers, i.e

$$\mathbf{C}^{(I)} = [C^{(I)}(p) = x^{(I)}(p) + j \cdot y^{(I)}(p) \;\mid\; p \in P], \tag{14}$$

where $\underline{X}^{(I)}(p) = (x^{(I)}(p), y^{(I)}(p))$ is the position vector of each point $p \in P$ of the contour. We use the term "observed motion", $\mathbf{d}_c$, to define the motion of the contour, obtained by a motion estimation scheme, that is expected to give us the position of the contour in the next frame $I+1$:

$$\mathbf{d}_c^{(I+1)} = [d_c^{(I+1)}(p) \;\mid\; p \in P] = [MF^{(I,I+1)}(x(p), y(p)) \;\mid\; p \in P], \tag{15}$$

where $MF^{(I,I+1)}(x(p), y(p))$ is the motion vector between I and $I+1$, at point $(x(p), y(p))$, computed using the motion estimation scheme proposed by Black et al. [2].

In our work we utilize the specific motion estimation technique, considering it as an efficient way to overcome the common problems that arise in estimating the displacements or the optical flow in video streams. The most important problem that is handled by this technique is the trade-off between motion field over-smoothing and noise sensitivity. More specifically, the method proposed in [2] (a) eliminates noisy motion estimates, that cannot give us the motion information we need, (b) handles motion estimates' over-smoothing, which leads to smooth but not actual motion information, and (c) provides smooth motion fields on edges, which are the most informative features; this can lead to successful separation between the moving front and the background, or between two objects moving close to each other. The latter advantage is exploited in our work, in order to estimate the desired object contour in textured backgrounds and detect its possible occlusions.

The observed motion $d_c^{(I+1)}(p)$ of each contour point $C^{(I)}(p)$ is different from its actual motion, which is calculated after the contour new position is estimated in the next frame $I+1$. Thus, we define the "instant motion" of the contour between the frames I and $I+1$ as,

$$\mathbf{m}_c^{(I+1)} = [m_c^{(I+1)}(p) \ \mid \ p \in P] = [\underline{X}^{(I+1)}(p) - \underline{X}^{(I)}(p) \ \mid \ p \in P], \tag{16}$$

where $\underline{X}^{(\cdot)}(p)$ represents the position vector corresponding to the snake point p. This equation represents the displacement of snake final solution between the successive frames I and $I+1$, i.e the difference between the respective position vectors, after the object contour in the next frame is estimated.

The snake initialization $\mathbf{C}_{init}^{(I+1)}$ in the next frame $I+1$ is defined in terms of the observed motion,

$$\mathbf{C}_{init}^{(I+1)} = \mathbf{C}^{(I)} + \mathbf{d}_c^{(I+1)} \tag{17}$$

Fig. 5 illustrates the difference between the observed motion, according to the obtained motion field, of a contour point, and its instant motion. Snake initializations and the vectors of the observed motion are shown in dotted form, whereas the final positions of the contour and the instant motions are illustrated in solid line. The circled points represent the point p at its initial position in each frame, whereas the filled circles (dots) represent its final position, estimated according to the procedure described in the following subsections.

In order to form the uncertainty regions around the snake initialization in each frame of the examined sequence, which actually represents the area in which possible contour deformations can occur, we exploit the contour motion history, obtained by the contour instant motions in the L previous frames. In this way, we form the "expected" contour motion $\bar{\mathbf{m}}_c^{(I+1)}$, obtained as the contour's mean motion in the previous L frames,

$$\bar{\mathbf{m}}_c^{(I+1)} = [\bar{m}_c^{(I+1)}(p) \ \mid \ p \in P] = [\ \frac{1}{L} \sum_{i=I-L}^{I-1} m_c^{(i+1)}(p) \ \mid \ p \in P \ ] \tag{18}$$

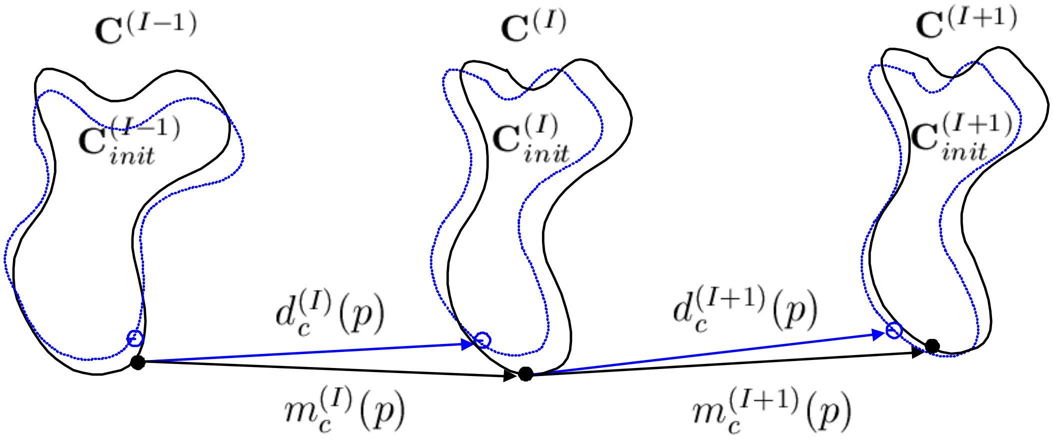


Fig. 5. *Differences between a contour observed motion and its instant motion in three successive time instances.*

The difference between the expected $\bar{\mathbf{m}}_c^{(I+1)}$ and the observed $\mathbf{d}_c^{(I+1)}$ motion, is used to formulate the uncertainty region, i.e the region where the acceptable deformations of the snake should be located. In particular, the expected solution (object boundaries), for the next frame $I+1$, is located in a narrow band around the snake initialization $\mathbf{C}_{init}^{(I+1)}$, whose width varies for each point p, and is equal to $2 \cdot sd^{(I+1)}(p)$ at each side of it (in the normal direction to the curve $\mathbf{C}_{init}^{(I+1)}$), where

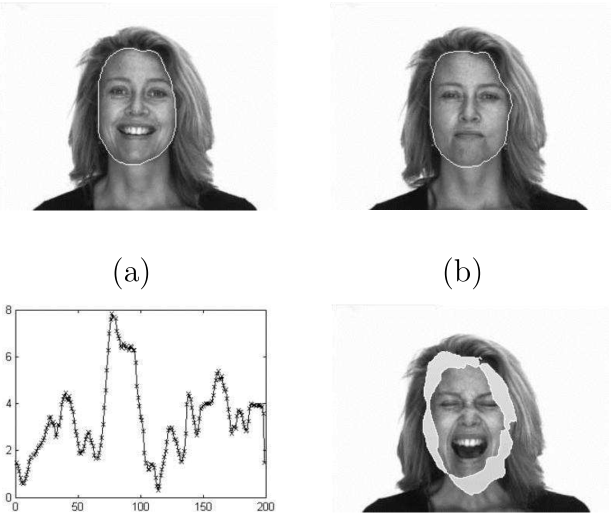
$$sd^{(I+1)}(p) = |\bar{m}_c^{(I+1)}(p) - d_c^{(I+1)}(p)| \tag{19}$$

In the above formulation we adopt the idea that the object contour is more likely to move in the same way, as it has been moving in the previous L frames of the sequence. In cases where the expected and observed contour positions coincide, there is clear evidence about the predicted contour position and the uncertainty region is narrowed. Thus, the final solution is taken as the initial one; however, exact match of observed and expected contour position is practically impossible, due to computational inaccuracies.

High values of $sd^{(I+1)}(\cdot)$ denote either a problem in the estimation of the observed contour position, or an abrupt change in the velocity of the tracked object, compared to its mean velocity in the previous frames. In both cases, the uncertainty about the predicted contour's initial position is high and the corresponding uncertainty regions are increased.

Fig. fig16 illustrates the proposed approach in steps, in the case of face tracking. Figs. 6(a) and (b) present two successive frames of a face sequence and the respective contours. Fig. 6(c) presents the amplitude of the computed deviation (in pixels) between the observed and the expected contour position. Based on this deviation, the uncertainty regions are then formed as shown in Fig. 6(d).

The exact location of the object contour (final solution) in the frame $I+1$ is



(a) (b)

(c) (d)

Fig. 6. *The proposed tracking approach in steps. (a)-(b) Two successive frames of a face sequence and the respective contours. (c) Amplitude of the deviation between the observed and the expected contour positions, leading to (d) the uncertainty regions of the curve.*

obtained by solving the following equations:

$$\mathbf{C}^{(I+1)} = \arg\min_{\mathbf{r}\in R}[w_1(D_K^{(\mathbf{r})} + D_{DV}^{(\mathbf{r})}) + w_2 \cdot E_{ext}^{(\mathbf{r})}], \tag{20}$$

$$D_K^{(\mathbf{r})} = \sum_{p\in P}[K_{(\mathbf{C}^{(I)})}(p) - K_{(\mathbf{r})}(p)]^2, \tag{21}$$

$$D_{DV}^{(\mathbf{r})} = \sum_{p\in P}[DV_{(\mathbf{C}^{(I)})}(p) - DV_{(\mathbf{r})}(p)]^2, \tag{22}$$

$$E_{ext}^{(\mathbf{r})} = \sum_{p\in P}[e_{ext}^{(\mathbf{r})}(p)] \tag{23}$$

where $K_{(\mathbf{C}^{(I)})}(p)$, $K_{(\mathbf{r})}(p)$ and $DV_{(\mathbf{C}^{(I)})}(p)$, $DV_{(\mathbf{r})}(p)$ are the curvature and the point density values of the contour $\mathbf{C}^{(I)}$ and the curve $\mathbf{r} \in R$ at the point p, respectively. R is the set of all possible curves r generated from the initialization $\mathbf{C}_{init}^{(I+1)}$ inside the extracted uncertainty region, and thus the minimization of eq. (20) is actually a problem of picking out the correct curve. The parameters w_1 and w_2 represent the weights with which the energy-based terms of eq. (20) participate in the minimization procedure (corresponding to a and b of eq. (1)).

The above equations impose the requirement of the similarity between the obtained solution and the contour estimated in the previous frame, in terms of the smoothness and the elasticity of the snake, which constrains the snake's evolution towards the object boundaries (external energy component).

3.2 Force-based Approach

The minimization of eq. (20) is a time-consuming procedure, since the number of curves generated from the snake initialization, inside the extracted uncertainty regions, is inhibitory for straightforward implementation. In this subsection we propose a force-based approximation of the minimization procedure inside the extracted uncertainty regions.

3.2.1 Internal Forces

If $\underline{\mathbf{n}}^{(I+1)} = [\underline{n}^{(I+1)}(p)|p \in P]$ and $\underline{\mathbf{t}}^{(I+1)} = [\underline{t}^{(I+1)}(p)|p \in P]$ are the sets of the normal and the tangential vectors of the snake initialization $\mathbf{C}_{init}^{(I+1)}$ in the frame $I+1$, respectively, we define the following forces:

$$\underline{\mathbf{F}}_d(p) = [DV_{(C^{(I)}(p))} - DV_{(C_{init}^{(I+1)}(p))}] \cdot \underline{\mathbf{t}}^{(I+1)}(p), \tag{24}$$

$$\underline{\mathbf{F}}_c(p) = [K_{(C^{(I)}(p))} - K_{(C_{init}^{(I+1)}(p))}] \cdot \underline{\mathbf{n}}^{(I+1)}(p) \tag{25}$$

The definition of these forces as being normal and tangential to the snake is motivated by the fact that $\underline{\mathbf{F}}_d$ corresponds to the distances between neighboring points along the snake, and thus it represents the elasticity force, whereas $\underline{\mathbf{F}}_c$ involves the curvature function, and thus represents the local deformation of the snake in its normal direction. Comparing these two forces with eqs. (21) and (22), we actually propose two force-based shape similarity constraints, denoting that the snake should not greatly deform its shape along time, corresponding to the snake internal energy components. In this sense, if two neighboring points increase their distance during the snake deformation, a tangential force $\underline{\mathbf{F}}_d$ is produced trying to bring them in their previous positions. Also, if the newly estimated position of a point produces higher or lower curvature value, i.e. the snake becomes less smooth locally, then a restitution force $\underline{\mathbf{F}}_c$, normal to the snake at this site, tries to bring the shape in its previous form.

3.2.2 External Force

Let us define $g_{m,p}(k)$, given by (26), be the modified image gradient function of all pixels $k = x_k + j \cdot y_k$ that belong to the uncertainty region $\mathbf{U}$, and lie on the line segment that is defined by the normal direction of the curve $\mathbf{C}_{init}^{(I+1)}$

at point p,

$$g_{m,p}(k) = [G_m(k) \mid (C_{init}^{(I+1)}(p) - k) \cdot \underline{n}^{(I+1)}(p) = 1, k \in \mathbf{U}] \tag{26}$$

The use of this function, for every point of the snake, is a way to determine the most salient edge pixel of the examined image $I+1$, inside the extracted uncertainty region, and thus indicate the possible object boundary pixels, in the normal direction of the snake. In this sense, the maximum of this function defines the direction of the snake external force, as follows,

$$\tilde{k}_p = \arg\max_k [g_{m,p}(k)], \tag{27}$$

$$sgn_p = \begin{cases} +, \text{ if } \tilde{k}_p \; inside\ the\ area\ defined\ by\ \mathbf{C}_{init}^{(I+1)} \\ -, \; otherwise \end{cases} \tag{28}$$

where sgn_p denotes the sign/direction of the external force to be applied to $C_{init}^{(I+1)}(p)$. The external snake force for each point $C_{init}^{(I+1)}(p)$ is then:

$$\underline{\mathbf{F}}_e(p) = sgn_p \cdot e_{ext}(p) \cdot \underline{n}^{(I+1)}(p) \tag{29}$$

This component is proportional to G_m and forces the snake to the salient edges inside the extracted uncertainty region. Thus when this force is applied, each point of the snake marches towards the object boundary edges, with a step defined by the value of the respective external energy term $e_{ext}(p)$. That is, when p is far from an edge, it is $e_{ext}(p) \simeq 1$ and the snake marches with constant "velocity"; otherwise, when p approaches an edge, $e_{ext}(p)$ reduces until it gets its minimum value ($e_{ext}(p) \simeq 0$), and the snake stops. If we used the image gradient $|\nabla G_\sigma * I|$, then the function (26) would contain a great number of maxima and thus the direction of the external force would probably lead to insufficient results. It must be also noted that the maximum of the function $g_{m,p}$ for each point p of the snake, corresponds to the minimum of the respective external energy term, according to eq. (13).
Although eq. (27) provides the most salient edge pixel in the normal direction of the snake at each point p, there are cases that this pixel is not the desired one, mainly due to four reasons:

- the ASF and the morphological operations, described in the previous Section, may eliminate noise but also some of the edges on the object boundaries, in case they are not strong enough,
- even the proposed modified image gradient may contain background edges close to the moving object, especially when these edges are strong and salient along the time (and thus cannot be eliminated with spatiotemporal filters),

- there are cases, not very usual in natural sequences, that both the moving object and the background are so smooth and with similar intensities, that the function $g_{m,p}$ in eq. (26) does not have any maxima, and
- there are cases that the moving object gets successively occluded by an obstacle, that may be a static or a different moving object, and thus there is loss of edge information for the occluded object region; in that case it is likely that the maximum of $g_{m,p}$ (if p is an occluded point) will lie on the boundary edge of the obstacle (which separates it from the moving object).

3.2.3 Rule-based Approach

In order to avoid the above mentioned problems, we use the information obtained from the motion field. In the following, we describe a method that separates the moving object from the background, detects possible occlusions and handles them, by estimating the occluded parts of the object contour.

Without any loss of generality, let us suppose that the background is highly textured and static, and the desired moving object gets successively hidden behind a static obstacle. Then, for each point p of the snake, the maximum $\tilde{k}_p = (\tilde{x_p}, \tilde{y_p})$ of $g_{m,p}$ does not probably lie on the desired edge. Let $k_l = (x_l, y_l)$ and $k_m = (x_m, y_m)$ be the surrounding points of $\tilde{k}_p$, on the line segment along which $g_{m,p}$ is computed. Then, in order to accept $\tilde{k}_p$ as the desired maximum (object boundary pixel), the three points $\tilde{k}_p$, k_l and k_m must obey the following two rules/constraints:

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

$$\left\{ \begin{array}{c} MF^{(I,I+1)}(k_l) \simeq d_c^{(I+1)}(p) \;\; and \;\; MF^{(I,I+1)}(k_m) \simeq 0 \\ or \\ MF^{(I,I+1)}(k_l) \simeq 0 \;\; and \;\; MF^{(I,I+1)}(k_m) \simeq d_c^{(I+1)}(p) \end{array} \right\}. \tag{30}$$

and

(b) $\tilde{k}_p$ must be a moving point with velocity close to $d_c^{(I+1)}(p)$, that is

$$MF^{(I,I+1)}(\tilde{k}_p) \simeq d_c^{(I+1)}(p) \tag{31}$$

Taking these constraints into consideration, we overcome cases that: (a) the maximum is found in background (Fig. 7(a)): it is not a moving one and does not separate two immiscible (according to motion) parts of the function $g_{m,p}$, (b) the maximum is found inside the moving object region (Fig. 7(b)): although it is a moving one, it does not divide the function $g_{m,p}$ in such two parts, (c) occlusion occurs and the maximum is on the occluding object boundary (Fig. 7(c)): the maximum is not moving, although it makes the

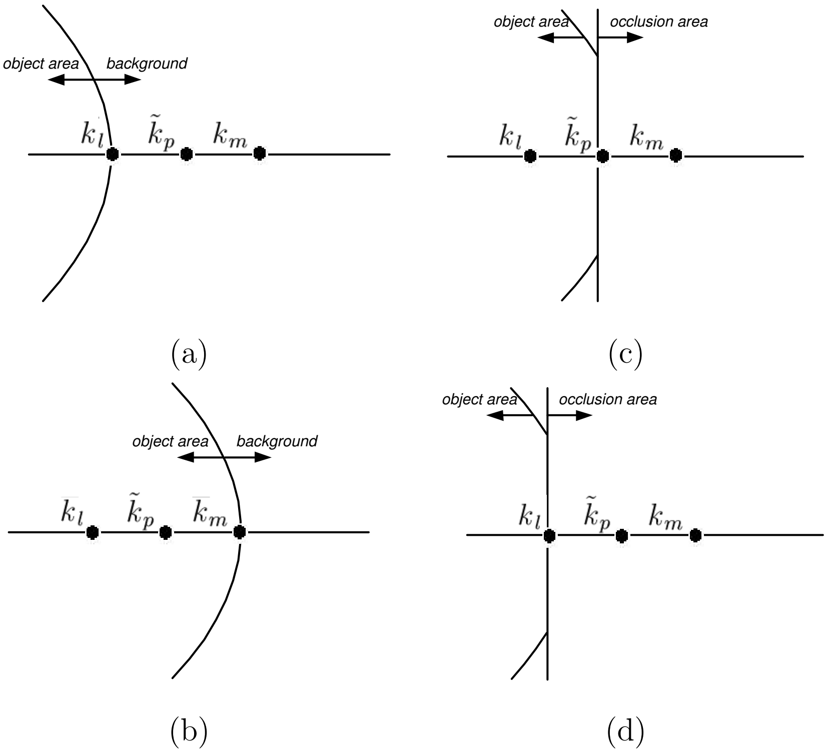


Fig. 7. *Four different cases in which the three points $\tilde{k}_p$, $\overline{k}_l$ and $\overline{k}_m$ do not obey both of the two rules for the separation of the moving object from the background or the occluding obstacle.*

region $g_{m,p}$ separation and (d) occlusion occurs and the maximum is in the occluding object region (Fig. 7(d)): neither the maximum is moving, nor it does such a separation.

If the above requirements are not met for the global maximum $\tilde{k}_p$ of the function $g_{m,p}$, at point p of the snake, we ignore it and continue searching for a local maximum that obeys these constraints. In occlusion cases, no such local maxima can be found and thus the external force is ignored, allowing the occluded parts of the snake to deform locally according to the internal forces. Fig. 8 illustrates detection of occlusion with the use of the above defined rules that the local maximum $\tilde{k}_p$, corresponding to a curve point p, must obey.

Finally, it should be noted that in occlusion cases, the more deformable a moving object is, or the more an object gets occluded, the less accurate the estimation of the occluded parts of its contour is. Moreover, when the desired object is already partially occluded, for the hidden parts of it, instead of the respective observed motion of eq. (15), we use the previously estimated instant motion $\mathbf{m}_c^{(I)}$ (eq. (16)) to initialize the snake in the next frame. Then we compute the deviation as follows,

$$sd^{(I+1)}(p) = \bar{m}_c^{(I+1)}(p) - m_c^{(I)}(p) \tag{32}$$

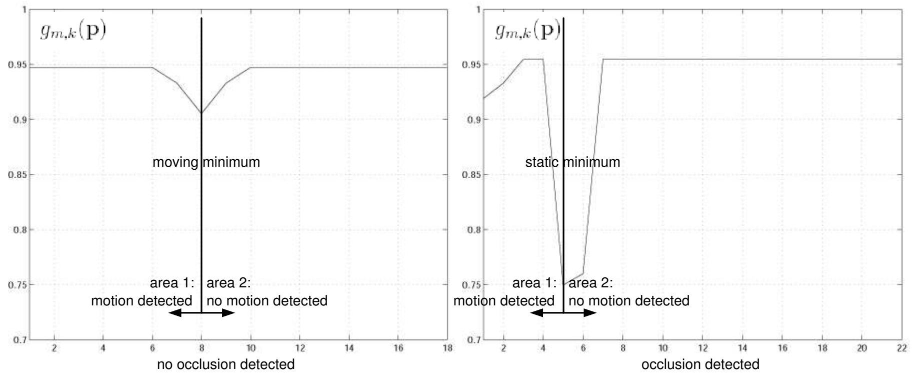


Fig. 8. *Detection of occlusion using the two rules of eqs. (30) and (31) for the local maximum $\tilde{k}_p$ of the function $g_{m,p}$ for a point p of the snake. The plots represent the function $1 - g_{m,p}$ and thus $\tilde{k}_p$ are presented as local minima of the $1 - g_{m,p}$.*

The latter is due to the obvious failure of the motion estimation scheme for the occluded parts, since no region information (intensity) is available for them.

3.2.4 Force-based Approximation of the Energy Minimization Procedure

In the force-based approach, the snake initialization $\mathbf{C}_{init}^{(I+I)}$ marches towards the object's boundaries in the next frame $I+1$, due to the internal and external forces applied to it. In this way, the minimization procedure of eq. (20) is approximated in an iterative manner similar to the steepest descent approach [8], as it is summarized below.
Let $\mathbf{C}^{(I+1),(\xi)}$ be the estimated contour in the ξ-est iteration, in the next frame $I+1$; then the following equations hold:

$$\mathbf{C}^{(I+1),(0)} = \mathbf{C}_{init}^{(I+1)} \tag{33}$$

$$\mathbf{C}^{(I+1),(\xi)} = \mathbf{C}^{(I+1),(\xi-1)} + \Delta\mathbf{C}^{(\xi)} \tag{34}$$

$$\Delta\mathbf{C}^{(\xi)} = [C^{(I+1),(\xi-1)}(p) + \underline{\mathbf{F}}_{tot}^{(\xi-1)}(p) \mid p \in P] \tag{35}$$

$$\underline{\mathbf{F}}_{tot}^{(\xi-1)}(p) = w_1 \cdot [\underline{\mathbf{F}}_{c}^{(\xi-1)}(p) + \underline{\mathbf{F}}_{d}^{(\xi-1)}(p)] + w_2 \cdot \underline{\mathbf{F}}_{e}^{(\xi-1)}(p), \tag{36}$$

where $\underline{\mathbf{F}}_{d}^{(\xi-1)}(p)$, $\underline{\mathbf{F}}_{c}^{(\xi-1)}(p)$ and $\underline{\mathbf{F}}_{e}^{(\xi-1)}(p)$ are estimated according to eqs. (24), (25) and (29) respectively, on the basis of $\mathbf{C}^{(I+1),(\xi-1)}$, instead of $\mathbf{C}_{init}^{(I+1)}$. The exponent $(\xi-1)$ in the forces' representations, denotes that they are calculated in the $\xi-1$ iteration, according to the form of the curve $\mathbf{C}^{(I+1),(\xi-1)}$ (for the internal forces), and its position on the image plane (for the external force).
The final solution $\mathbf{C}^{(I+1)}$ in the next frame $I+1$ is obtained when one of the following criteria is satisfied:
(a) the weighted summation of the corresponding forces' norms in the ξ-est

iteration has lower value than the respected summation in the next iteration, i.e.,

$$F_{\tau}^{(\xi)} < a \cdot F_{\tau}^{(\xi+1)}, \tag{37}$$

where

$$F_{\tau}^{(\xi)} = w_1 \cdot (\ \| \sum_{p \in P} \underline{\mathbf{F}}_c^{(\xi)}(p) \| + \| \sum_{p \in P} \underline{\mathbf{F}}_d^{(\xi)}(p) \| \ ) + w_2 \cdot \| \sum_{p \in P} \underline{\mathbf{F}}_e^{(\xi)}(p) \| \tag{38}$$

where a is a positive constant in the range $0 < a < 1$. When a is selected to be close to one, then $\mathbf{C}^{(I+1)}$ is more likely to correspond to a local minimum solution; lower values of α increase the number of iterations and, therefore, the execution time. The use of the uncertainty regions allows values of α get close to 1.
(b) the maximum number of iterations is reached; in this case,

$$\mathbf{C}^{(I+1)} = \mathbf{C}^{(I+1),(\tilde{\xi})}, \quad \tilde{\xi} = \arg\min_{\xi}[F_{\tau}^{(\xi)}] \tag{39}$$

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 eq. (20). However, under the constraints we pose, even if $\mathbf{C}^{(I+1)}$ corresponds to a local minimum, it is close to the desired solution.

3.3 Weight Estimation

As mentioned in Section 2, snake models generally involve a set of regularization parameters (parameters a and b in eq. (1)) taking part in the total energy function. In our approach, we represent these parameters as w_1 and w_2; they are the weights with which the internal and the external energy components of eq. (20) contribute in the minimization procedure. These weights are first calculated in the force-based implementation, in order to determine the importance of each force term, and thus determine the snake deformation according to (a) the snake elasticity and smoothness forces and (b) the force depending on the examined frame features (edges).
In order to make our method more straightforward, we propose an efficient and automatic estimation of these weights. Contrary to the energy minimization procedure of common snakes, where these weights are crucial for the accuracy of the obtained solutions, in our force-based method, w_1 and w_2 do not have to take specific values, but values that indicate the relativity between the "internal" and "external" information we utilize. This is mainly due the fact that we have constrained the problem (snake deformation) in a narrow band, i.e the

uncertainty region, and thus the final solution cannot be far from the actual object boundaries. Also, as mentioned in the force-based approach, although it is not easy to theoretically prove that the snake in its final position will correspond to the solution of the energy minimization, it is experimentally shown that at least it is very close to (or probably coincides with) it.

For the weight w_1 corresponding to internal force terms, it suffices to count the snake zero-crossings, that is the points at which the curvature function intersects the zero level. These points can determine how smooth the snake locally is, and thus denote how reliably we can apply the internal forces. The number of zero-crossings Z_c of a snake $\mathbf{C}^{(I)} = [C^{(I)}(p)|p \in P]$, in a frame I, with curvature function $K_{\mathbf{C}^{(I)}}$ (eq. (3)) is calculated according to,

$$Z_c = \sum_{p \in P} \left\{ \begin{array}{l} 1,\ if\ K_{\mathbf{C}^{(I)}}(p) \cdot K_{\mathbf{C}^{(I)}}(p+1) < 0 \\ 0,\ otherwise \end{array} \right\}. \tag{40}$$

Regarding weight w_2, corresponding to the external force component, we simply compute the summation $m_{(ext,\mathbf{U})}$ of the external energy values at all the pixels q inside the extracted uncertainty region $\mathbf{U}$:

$$m_{(ext,\mathbf{U})} = \sum_{q \in \mathbf{U}} e_{ext}(q), \tag{41}$$

In this sense, smooth uncertainty regions result to high values of $m_{(ext,\mathbf{U})}$, whereas uncertainty regions that are textured or contain strong edges, other than the object boundaries, result to significantly lower values of $m_{(ext,\mathbf{U})}$. Thus, if the extracted uncertainty region is textured, the external force component is assigned with lower weight, since it is considered as unreliable criterion for the snake deformation.

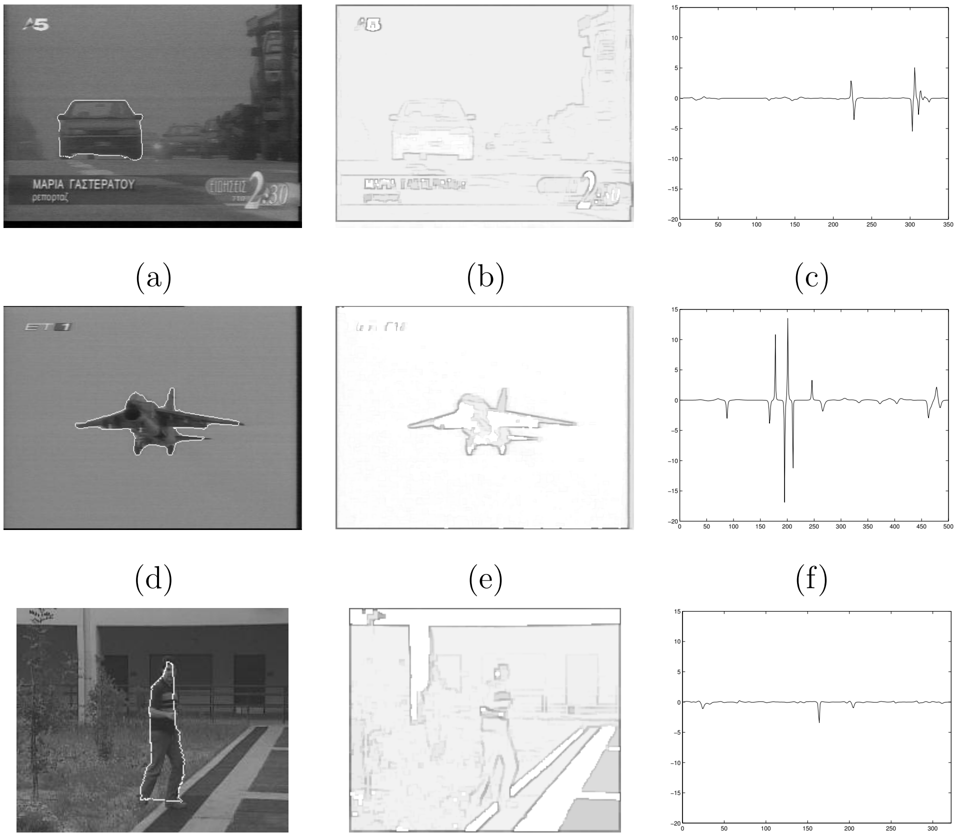
Practically, in order to set w_1 and w_2 values, according to the above and taking into account the number of snake points and the uncertainty region area, after extracting Z_c and $m_{(ext,\mathbf{U})}$, we compute the quantities,

$$\hat{Z}_c = \frac{Z_c}{P} \quad , \quad 0 \le \hat{Z}_c \le 1 \tag{42}$$

where P is the number of all snake points, and

$$\hat{m}_{(ext,\mathbf{U})} = \frac{m_{(ext,\mathbf{U})}}{Q} \quad , \quad 0 \le \hat{m}_{(ext,\mathbf{U})} \le 1 \tag{43}$$

where Q is the number of all pixels inside the uncertainty region ($q \in \mathbf{U}$), i.e. the uncertainty region area.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 9. *Curvature and external energy terms. (a,d,g) different cases of contours and background textures, (b,e,h) respective external energies visualization and (c,f,i) respective curvature distributions*

Thus, given the values of $\hat{Z}_c$ and $\hat{m}_{(ext,\mathbf{U})}$, the weights w_1 and w_2 are estimated by,

$$w_1 = \frac{100}{e^{5\hat{Z}_c}} \quad , \quad w_2 = \frac{100}{e^{5\hat{m}_{(ext,\mathbf{U})}}} \tag{44}$$

The factor 5, in the exponential parts of the denominators, is experimentally found to give the desired values of w_1 and w_2, as $\hat{Z}_c, \hat{m}_{(ext,\mathbf{U})} \in [0, 1]$. Let us now consider the examples of Fig. 9, where three objects of different shape complexity are in motion in either textured or smooth backgrounds. Figs. 9(a,d,g) represent the original images along with the moving object contours, Figs. 9(b,e,h) illustrate the proposed external energy components, whereas Figs. 9(c,f,i) are the respective curvature function plots. Indicatively, in the second example, the background is very smooth (Fig. 9(e)) and thus we estimate the external force with great reliability, but the aircraft's contour is complicated, which can be seen in the curvature plot, and thus the internal forces are not reliable.

3.4 Time-window L

In subsection 3.1 we described a method for the snake uncertainty region formation, utilizing the motion history of the tracked contour. In the respective eq. (18), it is denoted that these measurements involve the recent history of the tracked contour, i.e. the number L of the previous frames. We usually set $L = 5 \ldots 15$, according to the object motion, that is, the "smoother" the object motion is (almost constant velocity), the more informative a wide time-window is, and thus we use higher value for L; otherwise, we should use the very recent motion history, i.e a low value for L.

The question is what happens in the first frames of the examined sequence. In the first frame, we consider that the contour position has been manually set, whereas for the second frame we produce manually wide uncertainty regions. Then in the next frames, the object contours may be estimated not as accurately as we desire, but as time passes the motion history is getting "enriched" (updated) and thus the solutions get better.

4 Experimental Results

In the experimental results we present in this Section, we focus on natural sequences where very interesting cases can be encountered, due to the existence of noise, object deformations, complicated motions, and unpredictable external conditions (such as lighting changes, temporal clutter etc.). Moreover, cases with highly textured backgrounds and moving object partial occlusions are also handled. The performance of the proposed approach was tested over a large number of such sequences, and in this section we choose to illustrate the most representative examples.

Fig. 10 illustrates the case of tracking a car, moving towards the shooting camera. In this sequence, although the moving object is rigid, its silhouette is getting deformed slowly along time, due to its motion projection on the image plane. The background regions close to the object do not contain strong edges; however, some edges on the object boundaries are weak (low values of image gradient). As can be seen in (Fig. 10(a)), we assume that in the first frame the object contour is not estimated with the desired accuracy, but as long as motion history gets updated, the results are getting better.

Fig. 11 illustrates the tracking of the tail of a sperm whale. Object tracking in this case is considered to be difficult, due to the highly textured background and the strong clutter close to the desired object boundaries. The accuracy of the obtained results is enhanced due to the frame pre-filtering and the proposed modified image gradient. The results show that a very attractive application of the proposed method would be the sperm whale identification, based on its tail's contour local formations, which is a task in marine biology science.

In Figs. 12 and 13 the case of a small moving object that gets partially occluded is presented. Fig. 12 illustrates some indicative results of the motion estimation scheme [2] we utilize, and its ability to preserve smooth estimates on the boundary between two different objects. We exploit this advantage to track the finger nail in Fig. 13, efficiently separating the moving front from the background. Due to the slow motion of the finger, accurate nail contours are obtained, even in the frames where it is occluded by the obstacle (toy).

In the final example of Fig. 14, the case of a moving object that gets successively occluded is illustrated. The main problem in this case is the existence of weak edges on the object boundaries, due to the shadowing effects and the intensity similarity between the moving object region and the background, which leads to inaccurate motion estimates. On the other hand, the desired object is rigid, its motion is mainly translational (parallel to the camera's plane) and the obstacle's "crucial" edge is strong enough to indicate that occlusion occurs.

5 Conclusions and Further Work

In this paper we have presented a modified snake model assisted by rules, aiming at tracking objects in natural sequences, where the amount of noise is unknown, backgrounds may be highly textured, and partial object occlusions may occur. The internal energy of the proposed snake model is given in terms of two well known geometric characteristics, i.e the point density distribution (elasticity) and the curvature (smoothness) functions, whereas the proposed external energy is given in terms of a modified morphological image gradient. The object's motion history, along with a robust motion estimation scheme, provide the snake initializations for the next frames of the examined sequence, as well as uncertainty regions around these initializations, indicating the possible/allowable snake deformations. In this way, we constrain the contour estimation problem in a small frame region, and we follow a force-based approach to approximate the snake's energy minimization procedure, avoiding the point correspondence problem between two successive frames. Also, in order to handle object occlusions, as well as backgrounds with strong edges close to object boundaries, we propose a rule-based implementation of the proposed tracking model, utilizing again the motion estimation scheme [2]. The indicative experimental results we present illustrate the proposed method's success, when the afore mentioned problems occur.

Our future work focuses on tracking moving objects utilizing higher order statistics, in order to deal with articulated motion of objects. Also, we are currently investigating the estimation of force (energy) weights, in order to establish a reliable relation between them.

References

[1] Y. Avrithis, Y. Xirouhakis and S. Kollias, “Affine-Invariant Curve Normalization for Object Shape Representation, Classification and Retrieval,” *Machine Vision and Applications*, 13(2), pp. 80-94, 2001.

[2] M.J. Black and P. Anandan, “The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields,” *CVIU*, 63(1), pp. 75-104, 1996.

[3] V. Caselles, R. Kimmel, G. Sapiro, and C. Sbert, “Minimal Surfaces: A Geometric Three Dimensional Segmentation Approach,” *Numerische Mathematik*, 77(4), pp. 423-425, 1997.

[4] L. D. Cohen, “On Active Contour Models and Ballons,” *CVGIP: Image Understanding 53(2)*, pp. 211-218, 1991.

[5] M. Daoudi, F. Ghorbel, A. Mokadem, O. Avaro and H. Sanson, “Shape Distances for Contour Tracking and Motion Estimation,” *Pattern Recognition* 32, pp. 1297-1306, 1998.

[6] P. Delagnes, J. Benois and D. Barba, “Active Contours Approach to Object Tracking in Image Sequences with Complex Background,” *Pattern Recognition*, 16, pp. 171-178, 1995.

[7] D. Gravila and L. Davis, “3-D Model-based Tracking of Humans in Action: A Multi-view Approach,” *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR’96)*, San Fransisco, USA, pp. 73-80, 1996.

[8] S. Haykin, “Neural Networks” Macmillan College Publishing Company, *Section 5.3: The Steepest Descent Method*, pp. 124-126, 1994.

[9] M. Hoch and P. Litwinowicz, “A Practical Solution for Tracking Edges in Image Sequences with Snakes,” *The Visual Computer*, 12(2), pp. 75-83, 1996.

[10] H. S. Ip and S. Dinggang, “An Affine-Invariant Active Contour Model (AI-Snake) for Model-Based Segmentation,” *Image and Vision Computing*, 16(2), pp. 135-146, 1998.

[11] M. Israd and A. Blake, “Contour Tracking by Stochastic Propagation of Conditional Density,” *Proc. of European Conf. on Computer Vision (ECCV’96)*, 1, Cambridge, UK, pp. 343-356, 1996.

[12] A. K. Jain, Y. Zhong and M.-P. Dubuisson-Jolly, “Deformable Template Models: A review,” *Signal Processing* 71, pp. 109-129, 1998.

[13] M. Kass, A. Witkin and D. Terzopoulos, “Snakes: Active Contour Models,” *International Journal of Computer Vision*, 1(4), pp. 321-331, 1988.

[14] A. 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)*, pp. 636 -639, vol.3, Thessaloniki, Greece, 2001.

[15] P. Maragos, “Noise Suppression” The Digital Signal Processing Handbook, V.K Madisetti and D.B Williams Eds., CRC Press, 1998, Chapt. 74, pp. 20-21.

[16] P. Maragos, “Image Segmentation,” The Digital Signal Processing Handbook, V.K Madisetti and D.B Williams Eds., CRC Press, 1998, Chapt. 74, pp. 25-26.

[17] P. Maragos, “Computer Vision,” *Textbook from the post-graduate course “Computer Vision” in Electrical & Computer Eng. Dept. of National Technical University of Athens, instructor P. Maragos*, chapt. 5, 2002.

[18] F. Meyer and P. Bouthemy, “Region-Based Tracking Using Affine Motion Models in Long Image Sequences,” *CVGIP: Image Understanding* 60(2), pp. 119-140, 1994.

[19] F. Mohanna and F. Mokhtarian, “Improved curvature estimation for accurate localisation of active contours,” *In Proc. of International Conference on Image Processing (ICIP’01)*, pp. 781-784, vol.2, Thessaloniki, Greece, 2001.

[20] S. Osher and J. Sethian, “Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on the Hamilton-Jacobi Formulation,” *Journal of Computational Physics* 79, pp. 12- 49, 1988.

[21] N. Paragios and R. Deriche, “A PDE - Based Level Set Approach for Detection and Tracking of Moving Objects,” *In Proc. of the Sixth International Conference on Computer Vision (ICCV’98)*, pp. 1139-1145, Bombay, India, 1998.

[22] N. Paragios and R. Deriche, “Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects,” *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(3), pp. 266-280, 2000.

[23] N. Paragios and M. Rousson, “Shape Priors for Level Set Representations,” *European Conference in Computer Vision*, 22(3), Copenhagen, Denmark, 2002.

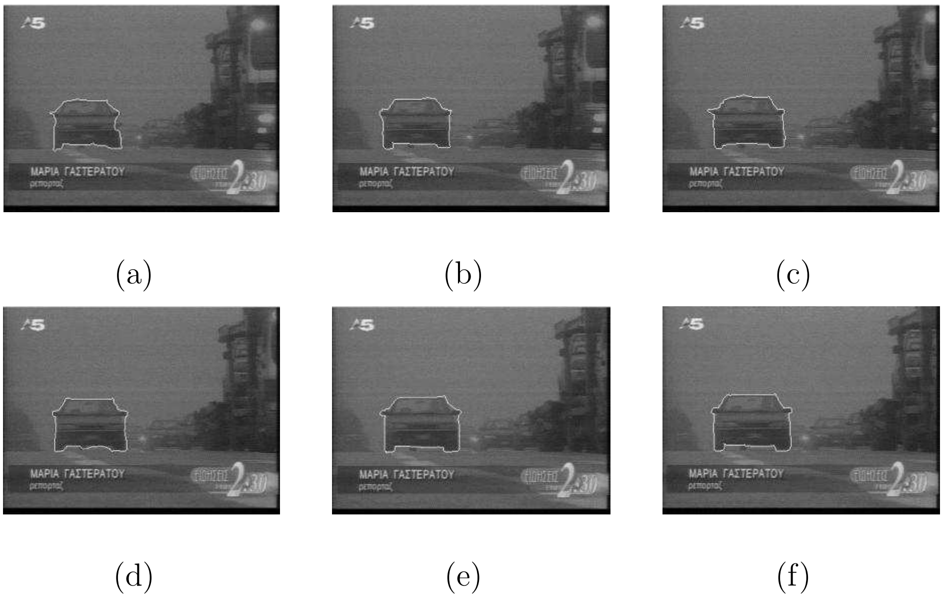
[24] J. S. Park and J.H. Han, “Contour Motion Estimation from Image Sequences Using Curvature Information,” *Pattern Recognition*, 31(1), pp. 31-39, 1998.

[25] J.S. Park and J.H. Han, “Contour Matching: A Curvature-based Approach,” *Image and Vision Computing*, 16, pp. 181-189, 1998.

[26] N. Peterfreund, “Robust Tracking of Position and Velocity with Kalman Filters,” *IEEE Trans. Pattern Analysis and Machine Intelligence* 21(6), pp. 564-569, 1999.

[27] C. Vieren, F. Cabestaing and J.G. Postaire, “Catching Moving Objects with Snakes for Motion Tracking” *Pattern Recognition*, 16, pp. 679-685, 1995.

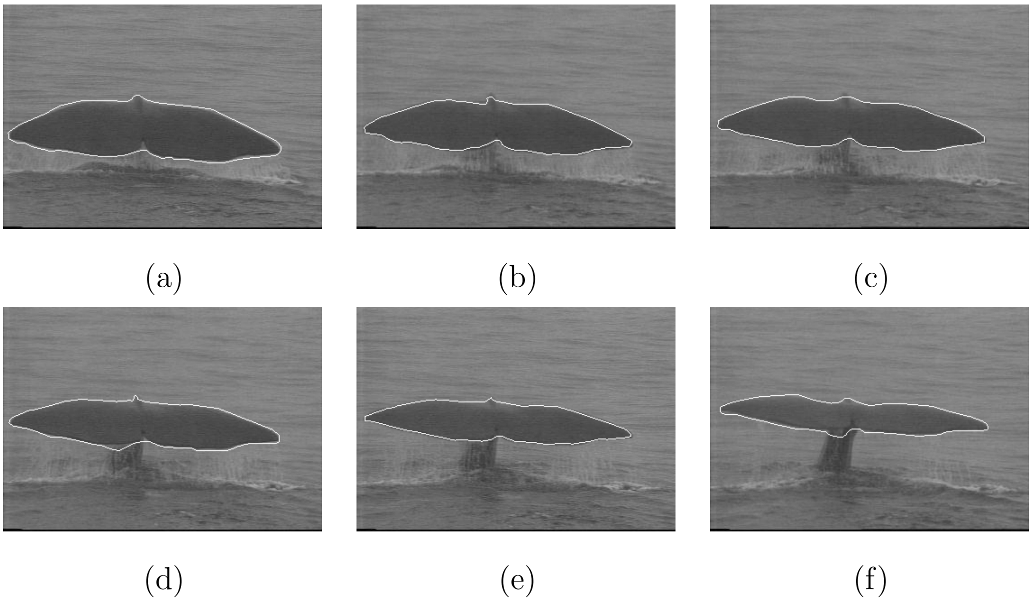
[28] C. Xu and J. L. Prince, “Snakes, shapes, and gradient vector flow,” *IEEE Trans. on Image Processing*, 7(3), pp. 359-369, 1998.



(a) (b) (c)

(d) (e) (f)

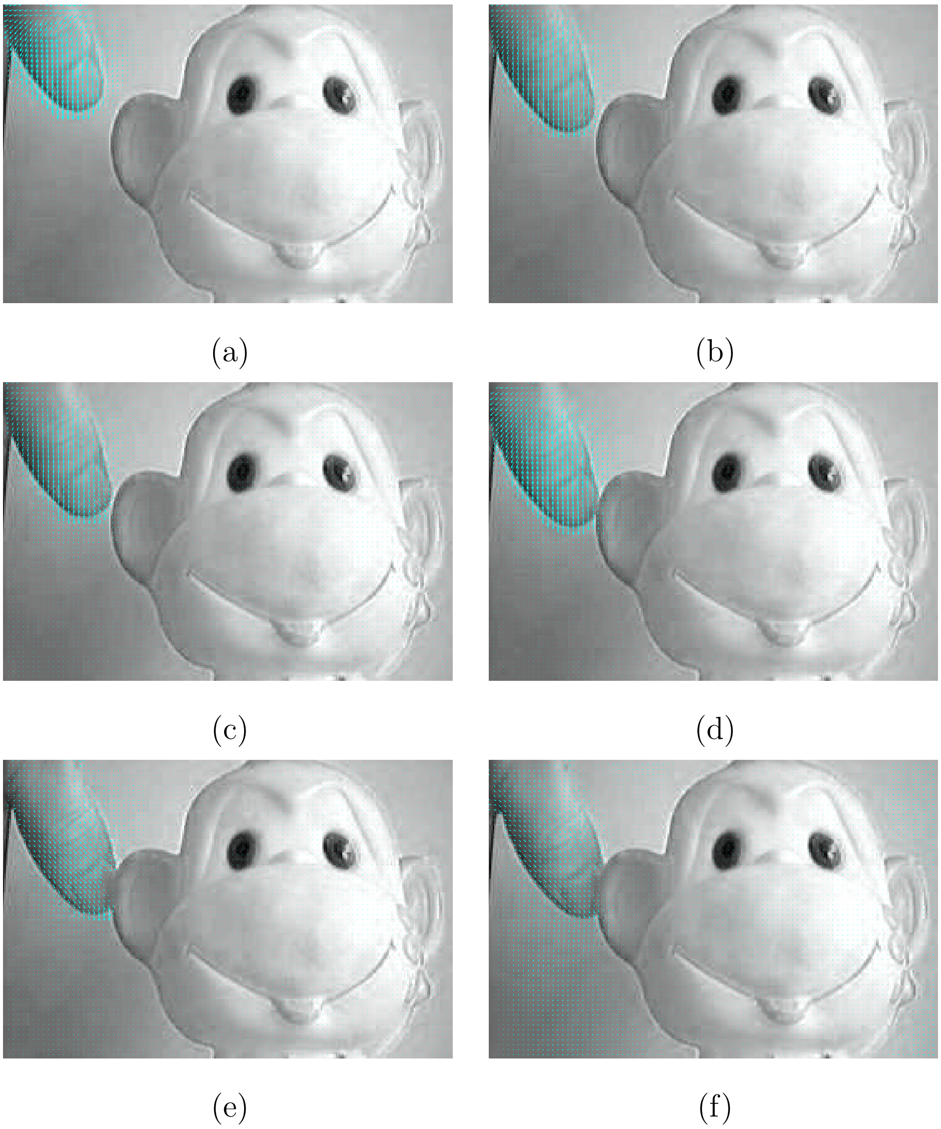
Fig. 10. *A moving car tracking in six (a,b,c,d,e,f) successive frames of a traffic sequence. The background is relatively smooth close to car's boundary, while the car's contour is not very complicated.*



(a) (b) (c)

(d) (e) (f)

Fig. 11. *A sperm whale tale tracking in six (a,b,c,d,e,f) successive frames of a traffic sequence: an example of a very cluttered sequence with highly textured background.*



(a)

(b)

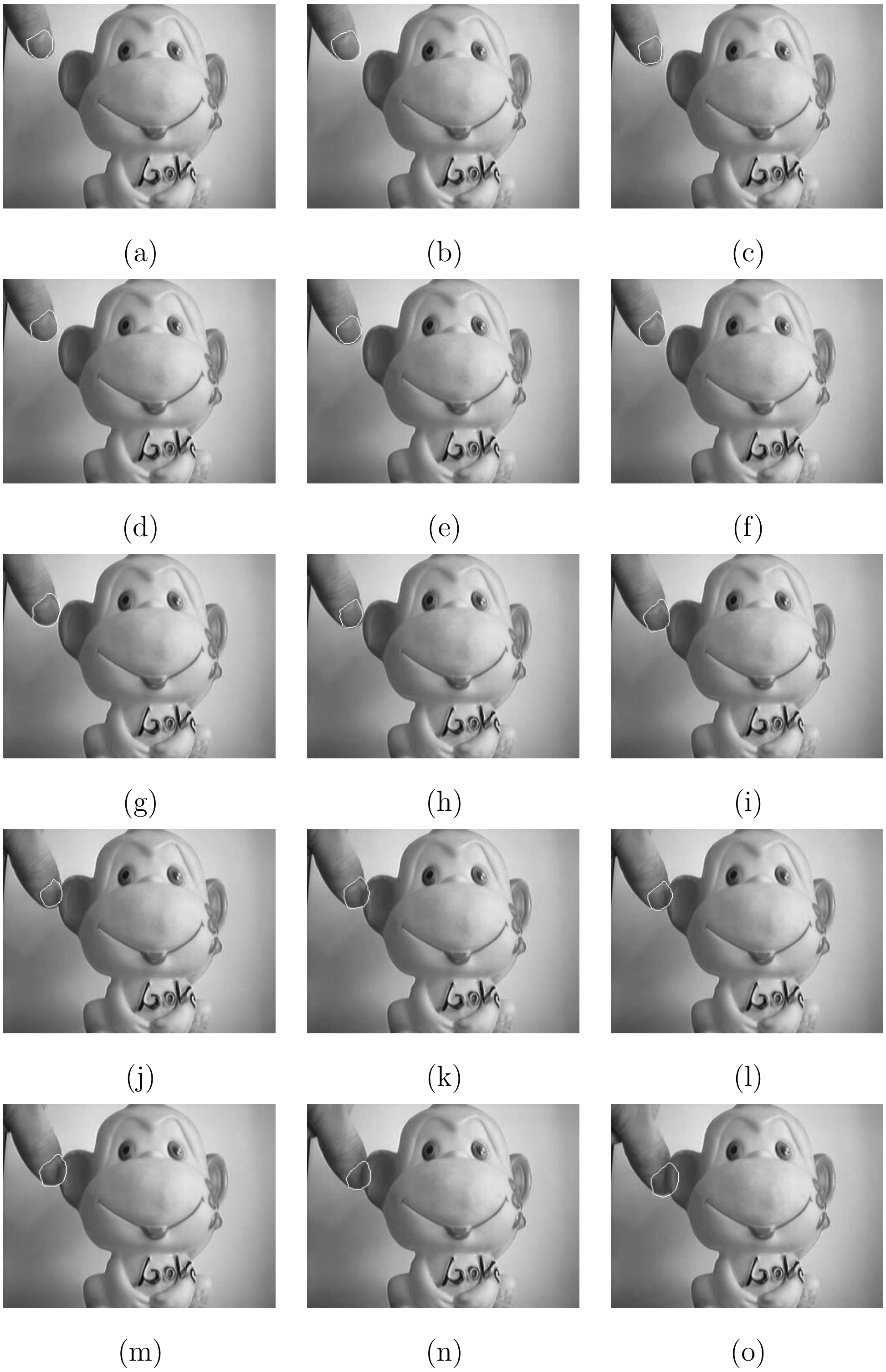
(c)

(d)

(e)

(f)

Fig. 12. *Indicative motion estimation results for the tracking example of Fig. 13.*



(a) (b) (c)

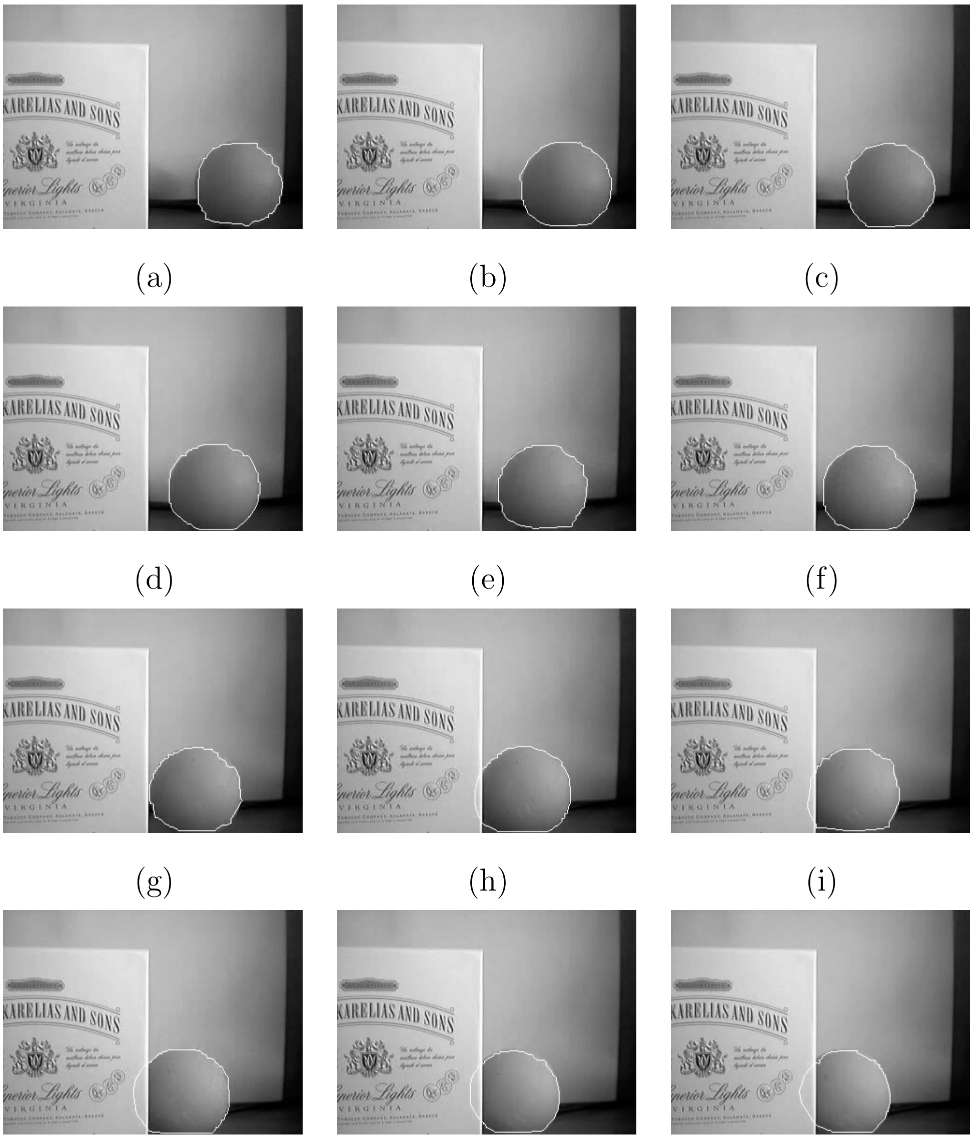
(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m) (n) (o)

Fig. 13. *Tracking a small rigid object in partial occlusion.*



(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

Fig. 14. *Object tracking with insufficient motion field and partial occlusion.*