Euclidean Invariant Snake for Joint Stereoscopic Segmentation

M.A. CHARMI(1), S. DERRODE(2) and F. GHORBEL(1)

(1) *Cristal laboratory, GRIFT,*
École Nationale des Sciences de l'Informatique,
Campus Universitaire de la Manouba,
Tn-2010 Manouba, Tunisia.
charmi.ma@free.fr, faouzi.ghorbel@ensi.rnu.tn

(2) *GSM, Institut Fresnel (CNRS UMR 6133),*
Faculté Paul Cézanne (Aix-Marseille III),
EGIM Nord, Domaine Universitaire de St Jérôme,
Fr-13013 Marseille Cedex 20, France.
stephane.derrode@fresnel.fr

Abstract

In this paper, we present the application of a recent method of snakes with geometric shape prior [6] for the segmentation of a pair of calibrated stereo images. Contours in one image of the pair are found using those in the other image as reference. The initialization of the snakes algorithm on the second image is performed automatically by matching only four particular points. The method is then adapted for object motion tracking in mono and stereo video sequences. The outlines of the algorithm are detailed and some experimental results are given and commented.

Keywords — *Snakes, shape prior, stereovision, invariant, Fourier descriptors, epilines, video tracking, stereo video tracking.*

1. Introduction

Snakes are used in many computer vision applications such as video [13], medical [8] and remote sensing [12] to segment and extract objects from the background. A snake is a curve $v(s)$ which move under an energy functional to converge to features of interest in images (lines or edges). The energy functional is the sum of an internal energy (the first and the second in (1)), which imposes elasticity and rigidity constraints on the curve, and an external energy, which attracts $v(s)$ to the boundary of the object:

$$E^*_{snake} = \int_0^1 \alpha \left|\frac{\partial v(s)}{\partial s}\right|^2 + \beta \left|\frac{\partial^2 v(s)}{\partial s^2}\right|^2 + E_{ext}(v(s))\, ds, \qquad (1)$$

where α et β are two constant weights.

These methods are characterized by their low computational cost which makes them tractable in nearly "real time", for applications such as object tracking in video sequences. However, problems related with initialization and convergence in concave boundaries sometimes limit their utility. From the initial work by Kass *et al.* [5], different solutions have been proposed in the literature, among them the Gradient Vector Flow (GVF) snakes [7] and balloons [8] are well-known efficient methods. In a recent work [6], we proposed to include a geometric-based shape prior information to help the snake to converge to the desired object, whatever its pose in the image. This technique is based on an Euclidean Fourier-based invariant description of the contour of reference (the *shape prior* or *template*) and the evolving snake at each iteration.

Only a few works dealing with active contours in stereo context have been presented during the last years. First, Kass *and al.* [5] minimize the disparity between the two evolving contours in the left and the right images. The formula for the proposed stereo coupled energy is:

$$E_{stereo} = \left(\frac{\partial v^R(s)}{\partial s} - \frac{\partial v^L(s)}{\partial s}\right)^2, \qquad (2)$$

where R and L denote the right and left views. Cham *and al.* [1] propose to enhance the B-splines active contours for tracking 3D curves in stereo images by considering epipolar constraints, when using calibrated affine cameras. In their work, a master frame called *canonical frame* is used to constrain the contour evolution. Many configurations are taken in consideration such as rigid and deformable curves with fixed and variable epipolar constraints. Also, depth and disparity maps derived from stereo images were used in [2] and [3] to deal with textured backgrounds and occluding objects in stereo images. In [2], an energy term depending on disparity is added to the snake energy, whereas a pre-processing step is done in [3] before the snake algorithm iterates. This step consists in removing the background texture, which is characterized by a homogenous surface in depth maps.

In this work, we present an extension of the method presented in [6] to segment an object into a pair of stereoscopic images acquired with calibrated cameras. It is done by using the first contour on the left image as a shape prior and well-known geometric constraints to help the second snake in the right image to converge. An automatic initialization of the right image snake is proposed using epipolar constraints. Finally, this strategy is applied to tracking in video stereo sequences.

The remaining of the paper is organized as follows: section 2 recalls the snakes algorithm with geometric shape prior. A detailed description of the application of snakes on a pair of stereo images are then presented in section 3. Extensions of snakes [6] for tracking in mono and stereo video sequences are then presented in section 4. Finally, we conclude on the method and present some perspectives to this work.

2. Snakes with shape prior

Since the first paper from Kass *et al.* [5], severall models of active contours with shape prior have been introduced. Cremers & al. [9] modified the Munford-Shah functional in order to add a statistical shape prior. They used an explicit parametrization of the contours using splines, in which each control point is associated to a Gaussian probability density function.

In [10], Zhong *et al.* present an affine-invariant deformable contour in a Bayesian framework. They introduce a new internal energy to define the global and local shape deformations of the contours between the shape domain and the image one. Interested readers may refer to [6] and [12] for a state-of-the-art on active contours with shape prior.

The remaining of this section is devoted to the presentation of the recent snake method with a geometric-based shape prior proposed in [6].

2.1. Presentation of the used Fourier invariant family

Let Γ be a discrete parametrization of a discrete closed curve with N points: $\Gamma(n) = (x(n); y(n))$; $n = 1, .., N$; where $x(n)$ and $y(n)$ are given according to the barycenter of Γ. Let $C_k(\Gamma)$ denotes the Discrete Fourier Transform (DFT) of Γ:

$$C_k(\Gamma) = \frac{1}{N} \sum_{n=0}^{N-1} \Gamma(n)\, e^{-j\frac{2\pi n k}{N}}, \qquad (3)$$

for $k = -\frac{N}{2}, \cdots, \frac{N}{2} - 1$.

The set of complex coefficients

$$I_k(\Gamma) = C_{k_0}^{-1}(\Gamma)\, C_k(\Gamma), \quad k_0 \neq 0,\ C_{k_0} \neq 0, \qquad (4)$$

forms a complete and stable [4] set of shape descriptors which are invariant to translation, rotation and scale factor, but not to the starting description point. The completeness property insures that the shape can be retrieved using the inverse DFT.

2.2. Embedding the shape invariants into snakes

To introduce shape prior information, we proposed to add a new force that guides the active contour in the image to a given template Γ_{ref}, independently of its pose, orientation and size. At each iteration t of the algorithm:

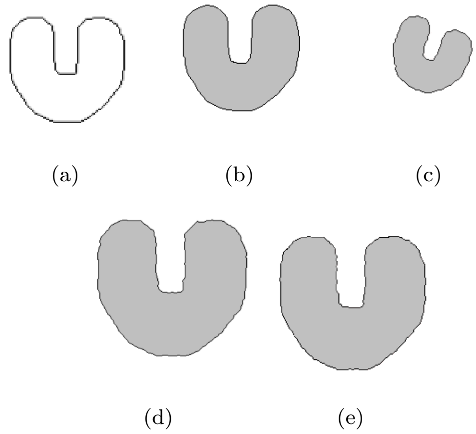


Figure 1. Results of the snake algorithm with shape prior on the "U" shape in (a) for several Euclidean transformations: (b) translation, (c) rotation, (d) scale factor, (e) scale factor and translation.

(1) First, we compute a linear mixture of the snake invariants at time t and the template invariants according to

$$I_k(\Gamma'_t) = (1 - c_{k,t})\, I_k(\Gamma_t) + c_{k,t}\, I_k(\Gamma_{ref}), \qquad (5)$$

where $c_{k,t} \in [0, 1]$ is a weight function that depends on the harmonic order k and time t. The function $c_{k,t}$ can be constant or represent a low-pass filter (*e.g. Hamming window*) to give more importance to low-order harmonics than to high-order ones. That way, $I_k(\Gamma'_t)$ can be considered as the invariant set of a curve Γ'_t influenced by the classical snake evolution Γ_t and by the template contour Γ_{ref}.

(2) Second, we reconstruct Γ'_t using the completeness property of the invariant set, according to

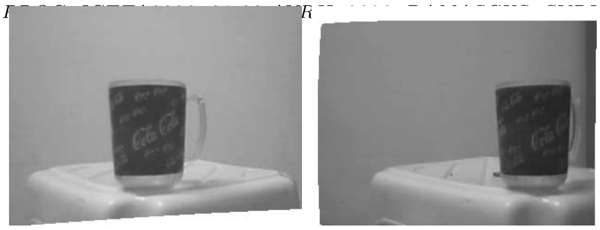
$$C_k(\Gamma'_t) = C_{k_0}(\Gamma'_t)\, I_k(\Gamma'_t), \qquad (6)$$

and the inverse DFT. Since we do not know the value for the harmonic C_{k_0} of curve Γ'_t, we put $C_{k_0}(\Gamma'_t) = C_{k_0}(\Gamma_t)$. Hence, Γ'_t is reconstructed with the same pose than Γ_t. The parameter k_0 is chosen, at each iteration, so that C_{k_0} is big compared to other harmonics.

We next define the new prior force of the snake as the difference between the snake Γ_t and the reconstructed shape Γ'_t after the invariants modification: $F^t_{prior} = \Gamma'_t - \Gamma_t$. The new forces of the snakes become $F = c_2\, F^t_{prior} - \nabla E_{ext}(v_t)$., where c_2 is a constant weight and E_{ext} is the classical external energy of the snake. Some results of the proposed method are shown in figure 1 (b→e) using (a) as template.

3. Application to stereo images

In this section, we describe the strategy adopted to extend the snakes with geometrical shape prior to the segmentation of a pair of stereo images acquired from calibrated cameras. In the remaining of the paper, we assume that we want to automatically find the boundaries of the right image I^R using the left image I^L as a reference.



(a) Left image I^L, (b) Right image I^R.

Figure 2. A pair of stereo images.

3.1. Method outline

The algorithm is divided into the three following steps which are illustrated with the "cup images" shown in Fig. 2.

Step 1– Segmentation of the left image

The first step of the algorithm is to find the edges of the left image using the GVF snake method which solves, to some extent, two well-known problems of snakes: the dependance of the method to the initial position of the contour and poor convergence into concave boundaries. The image in Fig. 3(a) shows the GVF segmentation result of Fig. 2(a).

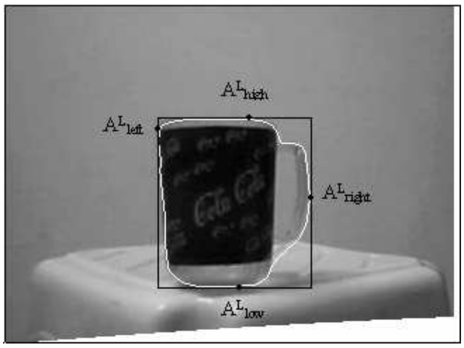
Step 2– Automatic initialization of the right image

The cameras are calibrated according to a simple procedure using mire. Then we rectify the pair of images, accordingly, two corresponding points A^L and A^R in the two images have the same X-coordinate: $X(A^R) = X(A^L)$. The rectification consists in rotating the PPM until focal planes become coplanar. This ensures that epipoles are at infinity, hence epipolar lines (or epilines) are parallel. In our case, epilines have the X-axis direction.

We then extract four particular points from the left image snake: the highest A^L_{high}, the lowest A^L_{low}, the most left A^L_{left} and the most right A^L_{right} as shown in Fig. 3(a). The epilines corresponding to A^L_{high} and A^L_{low} are given by: $X(A^R_{high}) = X(A^L_{high})$ and $X(A^R_{low}) = X(A^L_{low})$. Giving that A^L_{left} and A^L_{right} belong to their conjugate epipolar lines, their matching points can be found approximately by maximizing the correlation. The correlation product between a window centered at A^L_{left} (resp. A^L_{right}) and a moving window on the corresponding epipolar line is computed. We used the following normalized correlation:

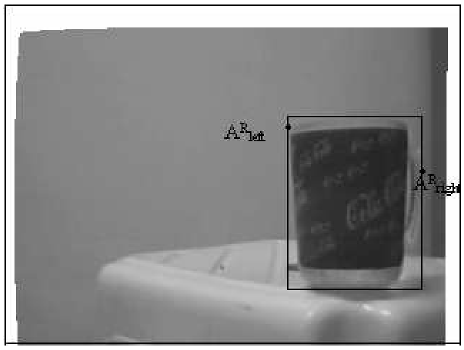
$$\mathrm{c}(x,y) = \frac{\sum_{i,j=-n}^{n} w(i,j) I(x+i,y+j)}{\sqrt{\sum_{i,j=-n}^{n} w(i,j)^2}\sqrt{\sum_{i,j=-n}^{n} I(x+i,y+j)^2}}, \tag{7}$$

where n is the half-size of the square window w and I is the image. Here x is fixed because we limit our query to epilines and only coordinate y varies. The maximum of $c(x,y)$ is reached at A^L_{high} and A^L_{low}.



(a) Results of GVF on I^L and the chosen points,

(b) Corresponding epilines in I^R of the highest and lowest points in I^L,



(c) Localization of the initialization of the snake in I^R.

Figure 3. Steps toward the automatic right image snake initialization from the left snake.

The rectangle limited by the epilines in A^R_{high} and A^R_{low} and the vertical lines passing by A^L_{high} and A^L_{low} is an approximation of the localization for the desired object in I^R (see obtained result in Fig. 3(c)).

The final initialization step consists in fitting the result found by the left snake in that rectangle. Since the two images are rectified, the transform between the found and the searched shape can be limited approximately to a translation and a stretching according to the Y-axis.

Step 3– Performing the snakes algorithm with shape prior

Finally, we perform the snakes algorithm with geometrical shape prior (as described in section 2) using the initialization obtained above and the left snake as reference.

(a) Result of the snake in I^L.

(b) Points of interest in I^L.

(c) Correspondance in I^R.

(d) Final result in I^R.

Figure 4. Results obtained at the end of each step of the algorithm.

3.2. Experimental results

We experimented our method on many calibrated stereo pair images. Figure 4 shows the results at the end of each step of the processing chain. The localization of the object in the left and right images are shown in (c) and (d). For correlation, we used a square window of 15 pixels. Another result is shown in 5. The values of α and β in eq. (1) are respectively 0.01 and 0.001. For the shape prior weights, we take $c_1 = 0.1$ and $c_2 = 0.05$. The contour is re-sampled to 128 nodes. The convergence is reached nearby the 150^{th} iteration.

4. Application to motion tracking

Active contours were already used in motion tracking [14], [13]. Indeed, these method have a low computational cost compared to other segmentation methods. In addition, we can make hypothesis on the transform between two instances of the tracked object in successive frames. This point was treated in [14], where authors assume that there is an affine transform between the two shapes.

4.1. Tracking in a single view

The first application consists of tracking an isolated object with a single camera. We propose to use the object contour at the i^{th} frame in the left sequence as a template for the $(i+1)^{th}$ frame. Indeed, the shape is approximately the same between two consecutive frames if we process an important number of frame per second.

In figure 6, we show the tracking results in a video sequence captured by a low-price webcam on the 33^{th}, 37^{th}, 72^{th}, 98^{th} and 119^{th} frames. The snake success to fit the object in all the frames. The number of iterations per frame is about 40. The complexity of the algorithm is $\Theta(Nln(N))$ per frame which correspond to the complexity of the DFT computation. It was performed under a P4 1.6 GHz, 256 Mo computer running windows. In addition, as we can see,

(a) Left image,

(b) Localization of the object in the right image,

(c) Right image.

Figure 5. Result of the snakes on a pair of stereo images.

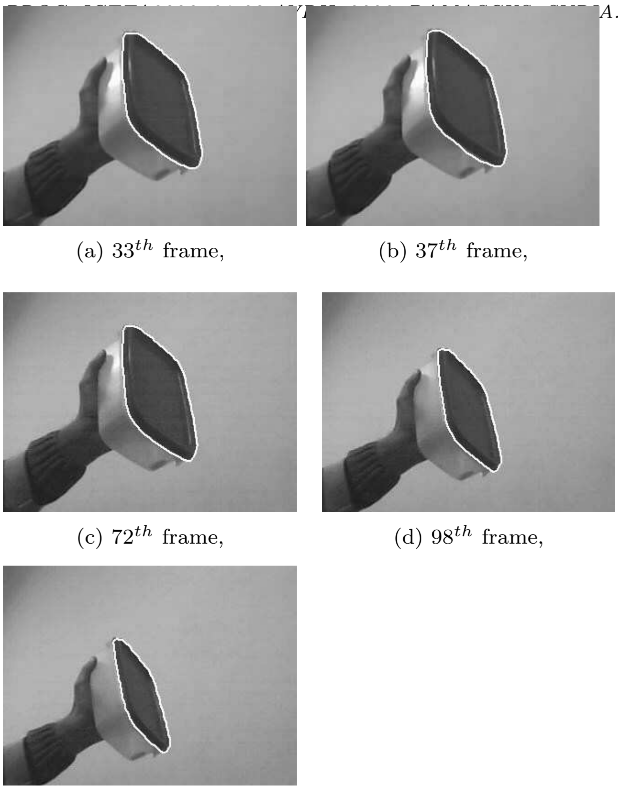
the method gives good results is more general than Euclidean transform (exemple: affine) which can be an experimental proof of the method robustness.

4.2. Tracking with two calibrated cameras

In the second application, the three-steps strategy explained above is used to track the object in the right view, according to the scheme described in Fig. 7. Two levels of shape prior are introduced: the communication between two consecutive frames in the left sequences and between the left and the right frames.

In the second result, presented in figure 8, we track a box using two calibrated cameras (cheap webcams with a low resolution). The obtained results are visually satisfying and we do never loose the object in the two views.

The used algorithm of snakes [6] enhances the tracking robustness in stereo sequences by adding shape constraints. However, the increase in computation time is relatively important compared to the clas-



(a) 33^{th} frame, (b) 37^{th} frame,

(c) 72^{th} frame, (d) 98^{th} frame,

(e) 119^{th} frame.

Figure 6. Application of the snake with geometric shape prior to motion tracking in mono video sequence.

sical snakes algorithm. In fact, the method requires additionally the computation of the DFT and the re-sampling of the curve many times at each frame. The cost of correlation computation and images rectification is also added.

5. Conclusion and further works

Some extensions of an already presented method of snakes with geometric shape prior to stereo images and tracking were presented in this paper. First, we propose a strategy for automatic initialization of one image of a pair using the edges of the other given some classical properties of stereovision. Then, edges in the second image are found using the first snake as template. The algorithm requires the matching of only four particular points. It is sensitive to the object retrieval by the snakes result applied to the first image and the quality of results obtained by the stereo correlation matching.

The second extension is the tracking of isolated object in mono and stereo sequences. The previous frame is used as template for the current one. The method enhances the tracking robustness, however, the complexity is increased due to the curve re-sampling, the DFT computation, rectification and stereo correlation when stereo sequences are considered.

As perspective, we plan to extend the snake prior to more general geometrical transform using suitable invariants especially those presented in [11] and further reduce the computational cost of the tracking in

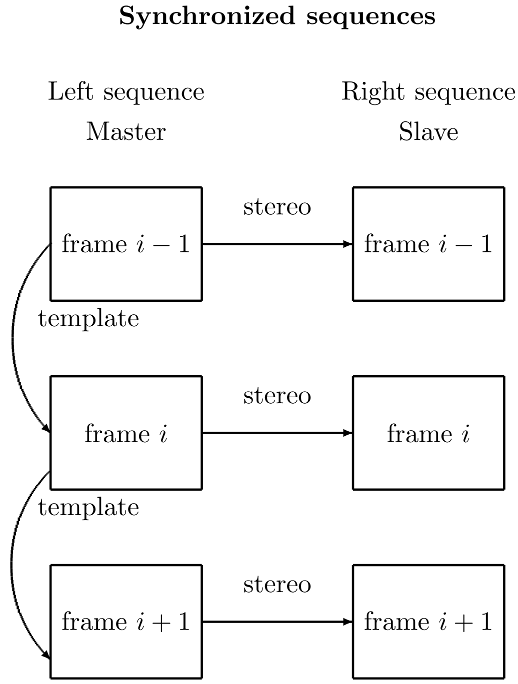


Figure 7. Using snakes with geometric shape prior and epipolar constraints for tracking objects with two calibrated cameras.

order to apply it for "real-time" applications.

acknowledgments

M.A. Charmi acknowledges the région PACA for their financial support through the "Med Accueil" Program. S. Derrode performed this work while at the Group GRIFT - ENSI, Tunisia.

(a) 51^{th} frame (b) 77^{th} frame (c) 90^{th} frame (d) 118^{th} frame

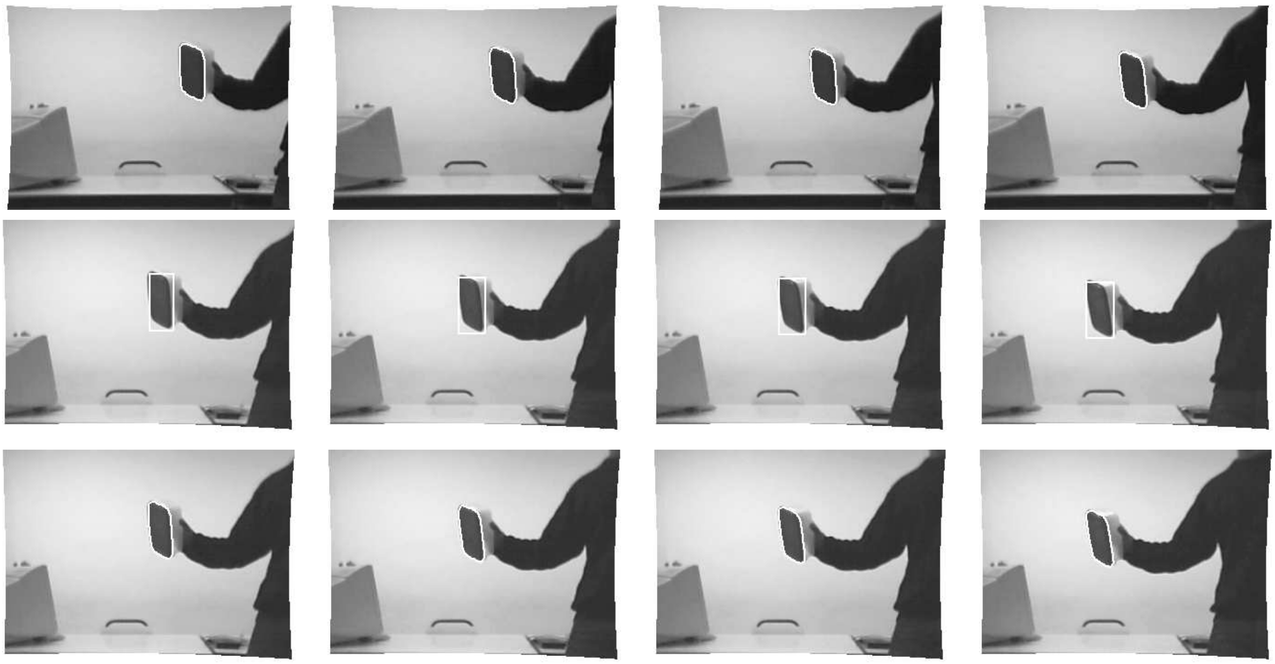


Figure 8. Tracking results on two stereo synchronized sequences: the first line represents the left view, the second line is the localisation using stereo geometry and the third is the snakes result in the right view.

6. References

[1] T.-J. Cham, and R. Cipolla, *Stereo coupled active contours, Proc. of the Computer Vision and Pattern Recognition Conf.(CVPR'97)*, pp. 1094-1097, 1997, Washington, DC, USA.

[2] M. Gelautz, and D. Markovic, *Recognition of object contours from stereo images: an edge combination approach*, 2nd Int. Symp. on 3D Data Processing, Visualization, and Transmission, Sept. 6-9, 2004, Thessaloniki, Greece.

[3] S.-H. Kim, J.-H. Choi, H.-B. Kim and J.-W. Jung, *A new snake algorithm for object segmentation in stereo images*, IEEE Int. Conf. on Multimedia and Expo (ICME'04), pp. 13-16, 27-30 June, 2004, Taipei, Taiwan.

[4] F. Ghorbel, *Stability of invariant Fourier descriptors and its inference in the shape classification*, 11th Int. Conf. in Pattern Recognition (ICPR'92), 30 Aug. - 3 Sept. 1992, The Hague, Netherlands.

[5] M. Kass, A.P. Witkin and D. Terzopoulos, *Snakes: active contour models*, Int. J. of Computer Vision, Vol. 1(4), pp. 321-331, January, 1988.

[6] S. Derrode, M.-A. Charmi and F. Ghorbel, *Fourier-based invariant shape prior for snakes*, IEEE Int. Conf. in Acoustics, Speech and Signal Processing (ICASSP'06), May 14-19 2006, Toulouse, France.

[7] C. Xu and J.L. Prince, *Gradient vector flow: a new external force for snakes*, IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR'97), pp. 66-71, June 1997, Los Alamitos, California, USA.

[8] L.D. Cohen, *On active contour models and balloons*, Graphical Models and Image Processing, Vol. 53(2), pp. 211-218, March, 1991.

[9] D. Cremers, C. Schnorr and J. Weickert *Diffusion-snakes: combining statistical shape knowledge and image information in a variational framework*, IEEE Workshop on Variational and Level Set Methods, pp. 137-144, July 6-9, 2001, Vancouver, Canada.

[10] X. Zhong, S.Z. Li and E.K. Teoh, *AI-Eigensnake: an affine-invariant deformable contour model for object matching*, Image and Vision Computing, Vol. 20(2), pp. 77-84, February 2002.

[11] F. Chaker, T. Bannour and F. Ghorbel, *A complete and stable set of affine-invariant Fourier descriptors*, 12th Int. Conf. in Image Analysis and Processing (ICIAP'03), pp. 578-581, 17-19 September, 2003, Mantova, Italy.

[12] M. Rochery, I. Jermyn and J. Zerubia, *Higher order active contours and their application to the detection of line networks in satellite imagery*, 2nd IEEE Workshop on Variational, Geometric and Level Set Methods, October 2003, Nice, France.

[13] H. Jiang, M.S. Drew, *A predictive contour inertia snake model for general video tracking*, Proc. of the Int. Conf. on Image Processing (ICIP'02), Vol. 3, pp 413-416, 24-28 June, 2002, Rochester, New York, USA.

[14] A. Black, R. Curwen and A. Zisserman, *Affine-invariant contour tracking with automatic control of spatial temporal scale*, Proc. of the Int. Conf. on Computer Vision (ICCV'93), pp. 66-75, 1993, Berlin, Germany.