Multiregion level set tracking with transformation invariant shape priors

Michael Fussenegger[1], Rachid Deriche[2], and Axel Pinz[1]

[1] Institute of Electrical Measurement and Measurement Signal Processing
Graz University of Technology
Schießstattgasse 14b, 8010 Graz, Austria
{Fussenegger, Axel.Pinz}@tugraz.at
http://www.emt.tugraz.at
[2] INRIA, 2004 route des Lucioles, BP 93
06902 Sophia Antipolis, France
Rachid.Deriche@sophia.inria.fr
http://www-sop.inria.fr/odyssee/

Abstract. *Tracking of regions and object boundaries in an image sequence is a well studied problem in image processing and computer vision. So far, numerous approaches tracking different features of the objects (contours, regions or points of interest) have been presented. Most of these approaches have problems with robustness. Typical reasons are noisy images, objects with identical features or partial occlusions of the tracked features. In this paper we propose a novel level set based tracking approach, that allows robust tracking on noisy images. Our framework is able to track multiple regions in an image sequence, where a level set function is assigned to every region. For already known or learned objects, transformation invariant shape priors can be added to ensure a robust tracking even under partial occlusions. Furthermore, we introduce a simple decision function to maintain the desired topology for multiple regions. Experimental results demonstrate the method for arbitrary numbers of shape priors. The approach can even handle full occlusions and objects which are temporarily hidden in containers.*

1 Introduction

Tracking of regions and object boundaries in image sequences is an important problem in computer vision (scene analysis and interpretation), video processing (video surveillance, object based video database search) and human-computer interaction.

Numerous tracking approaches have been developed, including early tracking algorithms to track feature points [1] and edge segments [2, 3], and several recent contributions to track parametric contours [4, 5]. Most of them had difficulties in handling topological changes such as the merging and splitting of overlapping object regions. For this, the level set method [6–8] is a more powerful technique. In the last few years various models have been proposed (see [9–15]). But there is always a problem with tracking multiple regions and none of these approaches uses the benefit of prior information to obtain a more robust tracking result. In this paper, we propose a novel approach that extends the well known level set model, such that it can simultaneously handle an arbitrary number of regions and competing shape priors.

An early work on region based tracking proposed by Bertalmio et al. [9] is based on morphing images. Paragios and Deriche [10] use a geodesic model that combines motion and edge information. Using the difference between the current image and the reference background, a region based model was proposed by Besson et al. [11]. In [12, 14] feature distributions of the object and the background were used for tracking. Freedman and Zhang [13] track a predefined distribution for the object region by minimizing a Kullback-Leibler or Bhattacharyya distance. All these approaches are restricted to one level set function and can only track one region. Shi and Karl [15] propose a new fast level set implementation that can handle multiple regions, but do not use prior information.

The integration of prior knowledge (in our case shape priors) into PDE based segmentation methods has delivered promising results (see [16–22]). Usually, the knowledge of one single shape prior is introduced into the contour evolution in a way that corrupted versions of a familiar object are reconstructed and all unfamiliar image structures are suppressed and often the localization of the shape must be known. Leventon et al. [18] use a Gaussian model to describe their shape priors. They assume a uniform distribution over pose parameters, that model translation and rotation. Rousson and Paragios [19] propose a transformation (scale, rotation and translation) for the shape prior that allows to segment familiar objects with an unknown position in the image scene. But like the approach of Leventon et al. they can handle only one shape prior and unfamiliar image structures are ignored. Cremers et al. ([23], [21]) presented an approach with dynamic labeling, that allows to use more than one shape prior and does not suppress unfamiliar image structures, but all shape priors are assigned to one level set function. Raviv et al. [22] present a novel approach that allows a projective transformation of the shape prior, but their approach is also limited to one region, furthermore the projective transformation needs too much calculation time for tracking applications. In [24] we present a segmentation level set framework that can handle an arbitrary number of regions with or without shape priors. Our segmentation algorithm in [24] is used in the initialization step of the multiple object tracking approach with shape priors, which we propose in this article.

The outline of the paper is as follows: Section 2 presents a level set formulation that can easily be extended with a single pose invariant shape prior (section 3). In section 4, we introduce our tracking algorithm. For the case of multiple object regions, we extend our tracking algorithm and introduce a logic function Ψ to incorporate topology in subsection 4.1. Results are presented in section 5. Finally conclusions are drawn in section 6.

2 A level set framework

In this section, we define a level set framework, that aims to maximize the color value homogeneity of the different regions. We assume each image of a video sequence is composed of a background region Ω_0 and N independent objects of interest $\Omega_1, ..., \Omega_N$.

Each of these $n = 1..N$ objects of interest is described with a level set function $\Phi_n : \Omega_n \rightarrow \mathbb{R}$, with $\Phi_n(\mathbf{x}) > 0$, if $\mathbf{x} \in \Omega_n$ and $\Phi_n(\mathbf{x}) < 0$, if $\mathbf{x} \in \Omega_0$.

There are different level set formulations, which could be possible choices [8, 25–27]. In this work, we use the level set formulation proposed by Paragios and Deriche [27, 28] to minimize the energy for each object region:

$$E_{D_n}(\Phi_n, p_n, p_0) = -\int_\Omega (H(\Phi_n)\log p_n + (1 - H(\Phi_n))\log p_0) d\mathbf{x} + \nu \int_\Omega |\nabla H(\Phi_n)| d\mathbf{x}. \quad (1)$$

H denotes the regularized Heaviside function and p_0 and p_n are the probability densities $p_i = p(\mathbf{x}|\Omega_i)$ of the background regions Ω_0 and the object region Ω_n, which cover the whole image domain Ω. For color images, we use the following multivariate Gaussian density:

$$p(\mathbf{x}|\Omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|} e^{-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i)}, \quad (2)$$

with the mean μ_i and the covariance matrix Σ_i of the multivariate color distribution of the region Ω_i. The last term of equation 1 takes into account the length of the contour weighted by the parameter ν. The minimization of the energy term in equation 1 can now be estimated according to the gradient descent equation

$$\frac{\partial \Phi_n}{\partial t} = \delta(\Phi_n)\Big[\nu \mathrm{div}\Big(\frac{\nabla \Phi_n}{|\nabla \Phi_n|}\Big) - log\frac{p_n}{p_0}\Big], \quad (3)$$

where $\delta(\Phi_n)$ is the derivative of $H(\Phi_n)$ with respect to its argument.

3 Adding a shape prior

To add a shape prior to the energy equation 1, we define a straight forward extension

$$E_n(\Phi_n, \Phi_0, p_n, p_0) = E_{D_n}(\Phi_n, p_n, p_0) + \lambda E_{S_n}(\Phi_n, \Phi_0), \quad (4)$$

with

$$E_{S_n}(\Phi_n, \Phi_0) = \int_\Omega \delta(\Phi_n)(\Phi_n - \Phi_{0n})^2 d\mathbf{x}, \quad (5)$$

where Φ_{0n} is the level set of the given training shape or the mean of a set of training shapes. $\lambda \geq 0$ indicates the weight of the prior. This formulation is simplified and does not consider invariance of the shape prior with respect to similarity transformations of the level set function. Nevertheless, equation 5 can be extended in this direction (cf. [18–20]), see section 3.1.

The minimization of the energy term can again be estimated according to the gradient descent equation:

$$\frac{\partial \Phi_n}{\partial t} = \delta(\Phi_n)\Big[\nu \mathrm{div}\Big(\frac{\nabla \Phi_n}{|\nabla \Phi_n|}\Big) - log\frac{p_n}{p_0} - 2\lambda(\Phi_n - \Phi_{0n})\Big] \quad (6)$$

3.1 A pose invariant formulation

During an image sequence the pose and the location of an object can change and the shape model has to be aligned. Possible solutions for simple or planar objects are presented in [19, 20], where a set of pose parameters are associated with the given prior Φ_{0n}. Rousson and Paragios [19] assume a global deformation A_n between Φ_n and Φ_{0n} that involves the parameters $[\mathcal{A} = (s; \theta; \mathbf{T})]$ with a scale factor s, a rotation angle θ and a translation vector **T**. The corresponding shape energy

$$E_{S_n}(\Phi_n, \Phi_0(\mathcal{A}_n)) = \int_\Omega \delta(\Phi_n)(s\Phi_n - \Phi_{0n}(\mathcal{A}_n))^2 d\mathbf{x} \tag{7}$$

is simultaneously optimized with respect to the segmentation level set function Φ_n and the pose parameters s, θ and **T**. The function is expanded with $\delta(\Phi_n)$, so that the shape prior is only estimated within the vicinity of the zero-crossing of the level set representation, which has a better performance than considering the whole image domain. Minimizing equation 7 leads to the following gradient descent for the level set function Φ_n:

$$\frac{\partial \Phi_n}{\partial t} = \delta(\Phi_n)\Big[\nu div\Big(\frac{\nabla \Phi_n}{|\nabla \Phi_n|}\Big) - log\frac{p_n}{p_0} - 2\lambda(s_n\Phi_n - \Phi_{0n}(\mathcal{A}_n))\Big] \tag{8}$$

The transformation $\mathcal{A}_\backslash$ is also dynamically updated to map Φ_n and Φ_{0n} in the best possible way. The calculus of variations for the parameter of $\mathcal{A}_n$ derives to the system:

$$\begin{aligned}
\frac{\partial s}{\partial t} &= 2\int_\Omega p(-\Phi_n + \nabla \Phi_{0n}(\mathcal{A}_n)\frac{\partial}{\partial s}\mathcal{A}_n)d\mathbf{x} \\
\frac{\partial \theta}{\partial t} &= 2\int_\Omega p(\nabla \Phi_{0n}(\mathcal{A}_n)\frac{\partial}{\partial \theta}\mathcal{A}_n)d\mathbf{x} \\
\frac{\partial \mathbf{T}}{\partial t} &= 2\int_\Omega p(\nabla \Phi_{0n}(\mathcal{A}_n)\frac{\partial}{\partial \mathbf{T}}\mathcal{A}_n)d\mathbf{x},
\end{aligned} \tag{9}$$

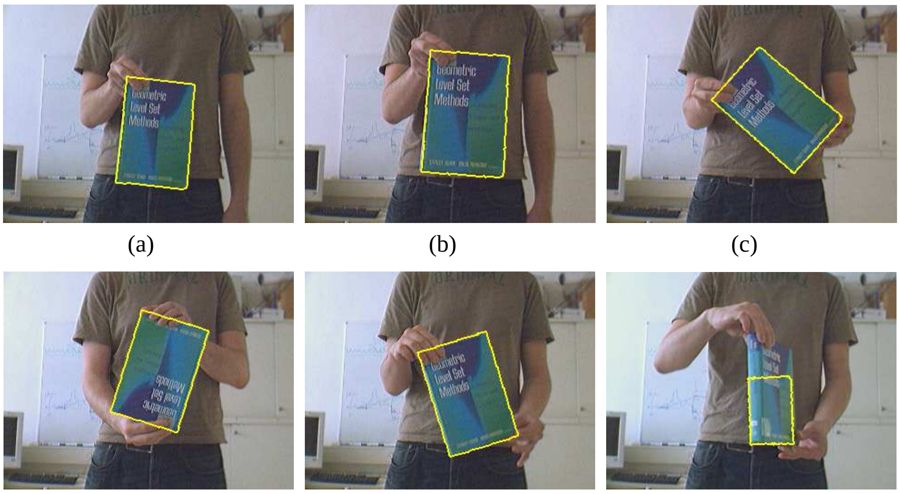
with

$$p = \delta(\Phi_n)(s\Phi_n - \Phi_{0n}(\mathcal{A}_n)). \tag{10}$$

Figure 1 shows the possible transformations of the shape prior: translation, rotation and scale. In figure 1(f) the limitation of our transformation model is shown at perspective distortions. Raviv et al. [22] presented an approach that allows a projective transformation of the shape prior, but their approach needs too much calculation time for tracking applications.

4 Tracking algorithm

In the initialization stage, we use the regions from the result of our multi region level set segmentation [24]. For every region of interest, we initialize a level set function Φ and calculate the means and covariance matrices for each region and the background. For



(a) (b) (c)

(d) (e) (f)

Fig. 1. Transformations of an object during a tracking sequence.

important or already learned objects, we assign a shape prior Φ_0 to the corresponding level set function Φ.

For the re-distancing step of the level set function Φ, we use a mixed approach: the PDE is used for reinitialization [29] in a small neighborhood of the zero level while the Fast Marching [30, 31] permits to extend the distance function to a larger band.

To track the object boundary, we compute the speed at all pixels in the band of the level set function Φ with equation 8 and calculate the new contour in the re-distancing step. In our implementation, the curve evolution stops when any of the following stopping conditions is satisfied:

1. Either, the transformation change of a pixel **x** is smaller than ε_1:

$$||\mathbf{x}_i - \mathbf{x}_{i-1}|| < \varepsilon_1, \text{ or} \tag{11}$$

2. a pre-specified maximum number of iterations is reached, or
3. the sum of the speed at each pixel is smaller than ε_2:

$$\sum_{\Omega_n} |\frac{\partial \Phi}{\partial t}| < \varepsilon_2. \tag{12}$$

4.1 Tracking of multiple objects

For the representation of multiple objects, we use one level set function Φ_n with or without an assigned shape prior Φ_{0n} and a decision function Ψ_n for each of N objects of interest. The decision function Ψ_n is defined as follows:

$$\Psi_n(x) = \begin{cases} 1 & \Phi_n(x) < 0, \Phi_l(x) > 0 \; (n \neq l) \\ 0 & \Phi_n(x) < 0, \Phi_l(x) < 0 \mid \exists \Phi_{0n} \; (n \neq l) \\ -1 & \Phi_n(x) < 0, \Phi_l(x) < 0 \mid \nexists \Phi_{0n} \; (n \neq l) \end{cases}$$

In the case of overlapping regions the different values for objects with and without shape priors allow an arbitrary topology, where objects with shape prior survive even with occlusions and they are handled as more important than regions without a shape prior. Our tracking algorithm consists of the steps shown in table 1.

- Step1:
  - Initialize all level set functions Φ_n.
  - Initialize all shape priors Φ_{0n} and assign them to the corresponding level set function Φ_n.
  - Calculate all means and covariance matrices for the regions and the background.
- Step2: (one cycle for region n)
  - Compute the parameters of transformation $\mathcal{A}_n$ for all shape priors Φ_{0n}.
  - Compute the gradient descent (equation 8) for all pixels in the band of Φ_n.
- Step3:
  - If the stopping condition is satisfied stop curve evolution for the actual region.
  - Else calculate the re-distancing for the actual region.
- Step4:
  - If the stopping condition for all regions is satisfied load new frame and start with step 2 with first region.
  - Else start with step 2 with next region.

Table 1. Tracking algorithm

Adding Ψ_n to equation 8 leads to following gradient descent:

$$\frac{\partial \Phi}{\partial t} = \begin{cases} u(x) - v(x) & \text{, for } \Psi_n = 1 \\ v(x) & \text{, for } \Psi_n = 0 \\ |u(x)| & \text{, for } \Psi_n = -1 \end{cases}$$

$u(x)$ and $v(x)$ are defined as follows:

$$u(x) = \delta(\Phi_n)\Big[\nu \text{div}\Big(\frac{\nabla \Phi_n}{|\nabla \Phi_n|}\Big) - log\frac{p_n}{p_0}\Big], \tag{13}$$

and

$$v(x) = 2\alpha_n \delta(\Phi_n)(s_n \Phi_n - \Phi_{Sn}(\mathcal{A}_n)). \tag{14}$$

the weighting parameter $\lambda = 0$ when there is no shape prior Φ_{0n} assigned to the level set function Φ_n. The scale s_n of the transformation $\mathcal{A}_n$ does not change for $\Psi_n \neq 1$.

Figure 2 demonstrates the value of Ψ. The first row shows three tracking results without the use of the decision function Ψ. In figure 2(b) the hand region grows into the ball because the color information is similar and in figure 2(c) the ball has vanished because it is totally occluded by the cup. In the second row, these problems do not occur because of the use of the decision function Ψ.

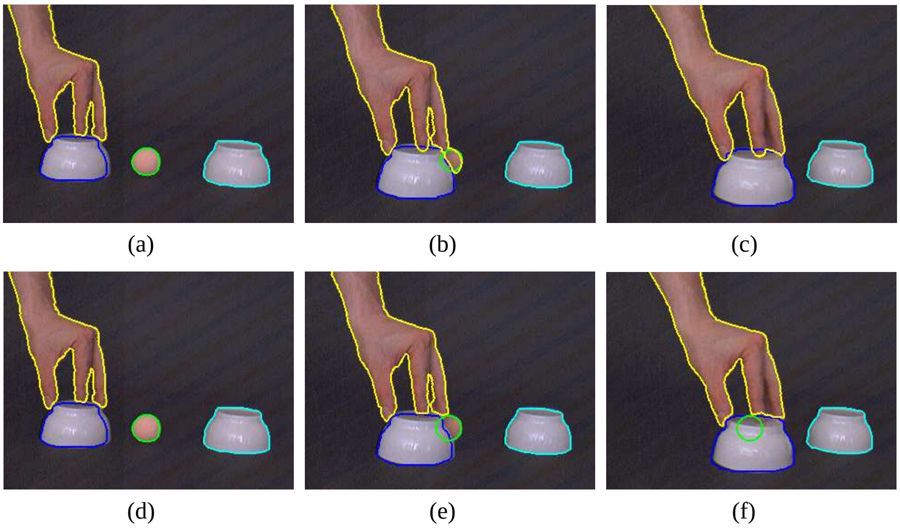


Fig. 2. Different tracking results of the *Game* sequence. First row without, second row with Ψ.

5 Tracking Results

We present two results of our tracking implementation, which run on a 2.0GHz PC under Linux. In all experiments, the segmentation result of frame 0 calculated with [24] is used for the initialization. Region tracking from frame to frame is performed via the algorithm described in table 1. We use the same parameters: $\nu = 1$, $\lambda = 1$ when a shape prior is assigned to the region and $\lambda = 0$ when not, for all experiments. The maximum number of iteration steps is set to 30, which assures a robust tracking even at fast movements. The first tracking experiment (figure 3) is performed on the *Game* sequence over 400 frames. We successfully track all objects in the image (hand, two cups, ball). The first image (figure 3(a)) shows the segmentation result from [24], which is used for the initialization. The hand region is tracked without prior information, because the changes of the hand shape can not be modeled with our shape priors. Shape priors are assigned to all other regions. The tracking speed strongly depends on the number of regions to track and on the length of the contour. For example we need only 0.1s per frame when we track only the ball with shape prior but 0.8s for the hand without shape prior. In this experiment we need an average time of 1.4s per frame for all objects. The second experiment (figure 4) is performed on the *Book 1* sequence over 250 frames. In this sequence, we track the book with shape prior and the orange object only with color information. Both objects are successfully tracked with 0.5s per frame. Our videos can be downloaded from `http://www.emt.tugraz.at/~pinz/data`.

6 Conclusion

We have introduced a novel level set based tracking framework that allows to track multiple object regions. Furthermore, we can add an arbitrary number of planar shape

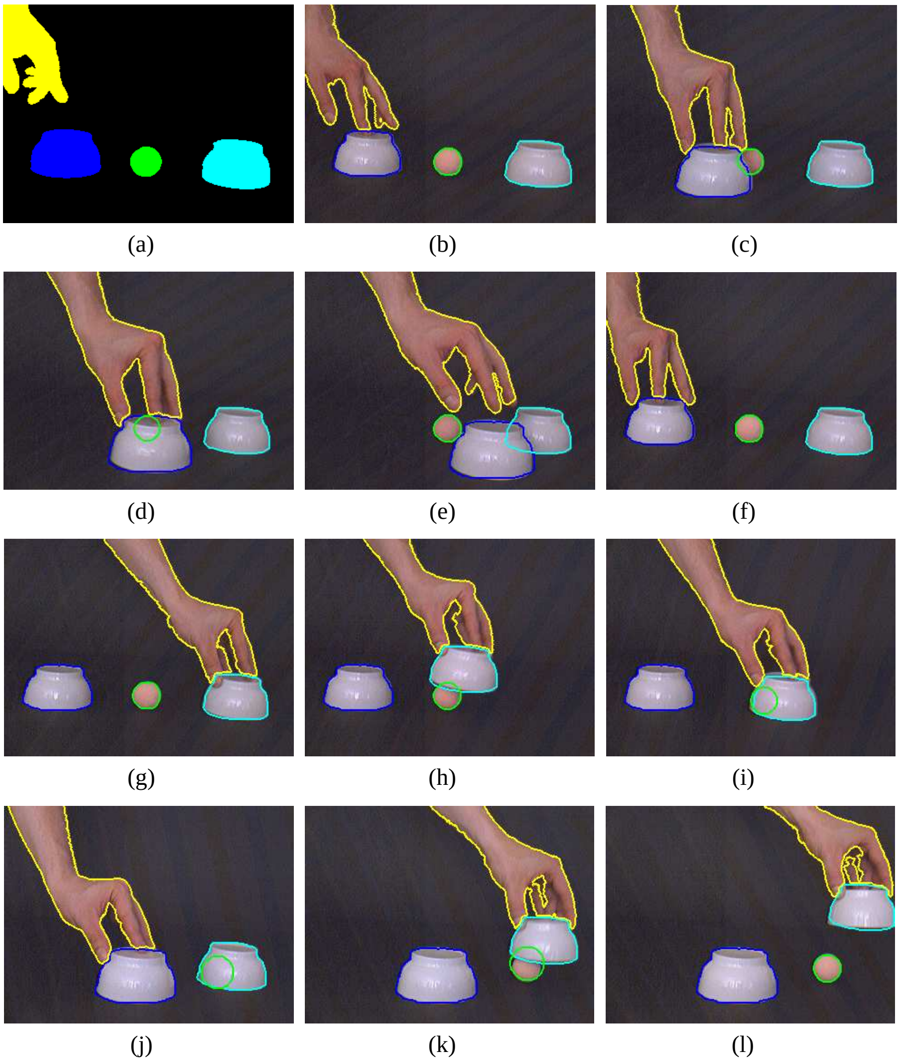
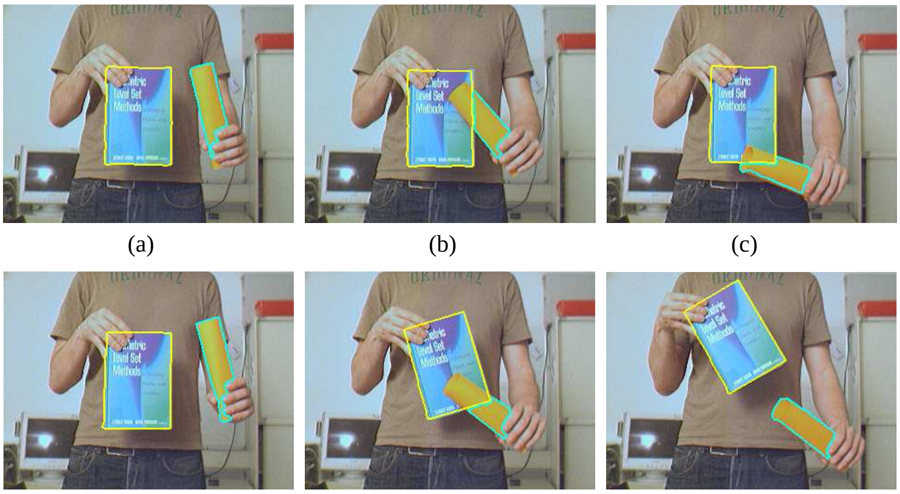


Fig. 3. Results of the *Game* (640x480) sequence. (a) initialization, (b-l) 11 out of 400 frames.

priors for already known or learned objects to get a more robust tracking result. Each shape prior is given by a fixed template (a given training shape or the mean of a set of training shapes) and respective pose parameters. A simple decision function Ψ ensures the desired topology for multiple regions tracking. Our approach can combine data-driven and recognition-driven information in the tracking process, and can for example be used to improve a cognitive vision system. Our approach has been successfully tested on a large number of real images, and it can even handle full occlusions which are temporally hidden in containers.



(a) (b) (c)

(d) (e) (f)

Fig. 4. Results of the *Book 1* (640x480) sequence. 6 frames of 250 are shown.

Acknowledgement

This research has been partly funded by the Austrian Science Foundation (FWF, project S9103-N04), the European project IMAVIS HPMT-CT-2000-00040 within the framework of the Marie Curie Fellowships Training Sites Programme and by the European project LAVA (IST-2001-34405). This is gratefully acknowledged.

References

1. Sethi, I.K., Jain, R.: Finding trajectories of feature points in an monocular image sequence. In: IEEE Trans. Pattern Anal. Machine Intell. Volume 9. (1987) 56–73
2. Crowley, J.L., Stelmaszyk, P., Discours, C.: Measuring image flow by tracking edge line. In: Proc. Second Int. Conf. of Computer Vision. (1988) 658–664
3. Deriche, R., Faugeras, O.D.: Tracking line segments. In: Proc. First European Conf. of Computer Vision. (1992) 259–268
4. Chen, Y., Rui, Y., Huang, T.S.: JPDAF based HMM for real-time contour tracking. In: Proc. of CVPR. (2001) 543–550
5. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. In: IEEE Trans. Pattern Anal. and Machine Intell. (2003) 564–577
6. Dervieux, A., Thomasset, F.: A finite element method for the simulation of Rayleigh-Taylor instability. In: Lecture Notes in Mathematics. (1979) 145–159
7. Dervieux, A., Thomasset, F.: Multifluid incompressible flows by a finite element method. In: International Conference on Numerical Methods in Fluid Dynamics. (1980) 158–163
8. Osher, S.J., Sethian, J.A.: Fronts propagation with curvature depend speed: Algorithms based on Hamilton-Jacobi formulations. In: Journal of Comp. Phys. Volume 79. (1988) 12–49
9. Bertalmio, M., Sapiro, G., Randall, G.: Morphing active contours: A geometric approach to topology-independent image segmentation and tracking. In: Proceedings of ICIP. (1998) 318–322

10. Paragios, N., Deriche, R.: Geodesic active contours and level sets for the detection and tracking of moving objects. In: IEEE Trans. Pattern Anal. Machine Intell. (2000) 266–280
11. Besson, S., Barlaud, M., Aubert, G.: Detection and tracking of moving objects using a new level set based method. In: Proceedings of ICPR. (2000) 1100–1105
12. Mansouri, A.: Region tracking via level set PDEs without motion computation. In: IEEE Trans. Pattern Anal. Machine Intell. (2002) 947–961
13. Freedman, D., Zhang, T.: Active contours for tracking distributions. In: IEEE Trans. Image Processing. (2004) 518–526
14. Yilmaz, A., Li, X., Shah, M.: Contour-based object tracking with occlusion handling in video acquired using mobile cameras. In: IEEE Trans. Pattern Anal. Machine Intell. (2004) 1531–1536
15. Shi, Y., Karl, W.C.: Real-time tracking using level sets. In: Proceedings of CVPR. (2005)
16. Yuille, A., Hallinan, P.: Deformable templates. In: A. Blake and A. Yuille, editors, Active Vison. (1992) 21–38
17. Cootes, T.F., A. Hill, C.J.T., Haslam, J.: Use of active shape models for locating structures in medical images. In: Image and Vison Computing. Volume 12:6. (1994) 355–365
18. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical shape influence in geodesic active contour. In: Proceedings of Conference Computer Vision and Pattern Recognition. Volume 1. (2000) 316–323
19. Rousson, M., Paragios, N.: Shape priors for level set representations. In: Proceedings of European Conference of Computer Vision. Volume 2351 of LNCS. (2002) 78–92
20. Chen, Y., Tagare, H.D., Thiruvenkadam, S., Huang, F., Wilson, D., Gophinath, K.S., Briggs, R.W., Geiser, E.A.: Using prior shapes in geometric active contours in a variational framework. In: International Journal of Computer Vison. Volume 50(3). (2002) 315–328
21. Cremers, D., Sochen, N., Schnoerr, C.: Multiphase dynamic labeling for variational recognition-driven image segmentation. In: Proceedings of European Conference of Computer Vision. (2004) 74–86
22. Riklin-Raviv, T., Kiryati, N., Sochen, N.A.: Unlevel-sets: Geometry and prior-based segmentation. In: Proccedings of ECCV. (2004) 50–61
23. Cremers, D., Sochen, N., Schnoerr, C.: Towards recognition-based variational segmentation using shape priors and dynamic labeling. In: Proceedings of Scale-Space 2003. (2003) 388–400
24. Fussenegger, M., Deriche, R., Pinz, A.: Multiregion level set tracking with transformation invariant shape priors. In: Proceedings of ACCV. (2006)
25. Chan, T., Vese, L.: Active contours without edges. In: IEEE Transaction on Image Processing. Volume 10(2). (2001) 266–277
26. Tsai, A., Yezzi, A.J., Willsky, A.S.: Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation and magnification. In: IEEE Transaction on Image Processing. Volume 10(8). (2001) 1169–1186
27. Paragios, N., Deriche, R.: Geodesic active regions : a new framework to deal with frame partition problems in computer vision. In: Journal of Visual Communication and Image Representation. Volume 13(1/2). (2002) 249–269
28. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for motion estimation and tracking. In: Computer Vision and Image Understanding. Volume 97(3). (2005) 259–282
29. Sussman, M., Smereka, P., Osher, S.: A level set approach for computing solutions to incompressible two phase flow. In: Journal of Computational Physics. (1994) 146–159
30. Adalsteinsson, D., Sethian, J.: A fast marching level set method for propagating interfaces. In: Journal of Computational Physics. (1995) 269–277
31. Sethian, J.A.: A fast marching level set method for monotonically advancing fonts. In: Proceedings of the National Academy of Sciences of USA. (1996) 1591–1595