PDE-BASED REGION TRACKING WITHOUT MOTION COMPUTATION BY JOINT SPACE-TIME SEGMENTATION

Abdol-Reza Mansouri[†], *Amar Mitiche, and Michael Aron*

INRS-Télécommunications, Institut National de la Recherche Scientifique
Place Bonaventure, Suite 6900, Montréal, Québec, Canada, H5A 1K6
[†] Division of Engineering and Applied Mathematics, Harvard University
Cambridge, MA 02138

1. PROBLEM STATEMENT

Tracking of regions in image sequences is one of the basic problems of image processing and computer vision, and plays an important role in numerous applications (search and retrieval in video databases, object based coding such as in MPEG-4, surveillance, automated image editing). Although numerous approaches to region tracking have been developed, most suffer from severe constraints and assumptions, typically by assuming particular motion models or constraining the range of interframe motion, or even assuming a fixed background. Furthermore, most of these algorithms perform tracking on a frame-to-frame basis as opposed to a multi-frame basis, leading to a possible lack of temporal coherence and a loss of tracking accuracy.

In this paper, we propose a novel algorithm for region tracking without motion computation that uses as its starting point the Bayesian framework for tracking previously developed [3], [4], extending it to a *multi-frame* tracking algorithm in which tracking is expressed as segmentation in the spatio-temporal domain [5]. Our proposed algorithm is expressed as the solution of level set partial differential equations, and the tracked region in a particular frame of the sequence is then obtained as the time slice of the level surface given by the level set equations.

The main benefit of our proposed algorithm is that contrary to numerous other tracking algorithms, it is a multi-frame tracking algorithm which does not assume the motion to be small [6], nor the background to be stationary [1], nor the region to be uniform in intensity on a uniform background [2], nor does it assume any motion models [5]. This leads to a tracking algorithm which combines the advantages of the Bayesian formulation developed for frame-to-frame tracking [3] with those of the formulation in the spatio-temporal domain [5]. We illustrate the performance of our algorithm on real image sequences with natural motion.

2. REGION TRACKING VIA JOINT SPACE-TIME SEGMENTATION

2.1. Basic Models and Level Set Evolution Equations

We first consider the Bayesian approach for frame-to-frame tracking developed in [3]. This will lead to a tracking functional which we will then generalize to multi-frame tracking. Let then $\{I^n, I^{n+1}\}$ be images at time instants t_n and t_{n+1}. Let the domain of both images be Ω. Let $\boldsymbol{R}_n \subset \Omega$ be a region in the image at time n (I^n) and let $\boldsymbol{R}_{n+1} \subset \Omega$ be the corresponding (unknown) region in the image at time $n+1$ (I^{n+1}). Assume there exists a given (finite or infinite) set Γ of bijective transformations $\psi : \Omega \rightarrow \Omega$, and that there exists a mapping $\phi_n \in \Gamma$ with $\phi_n(\boldsymbol{R}_n) = \boldsymbol{R}_{n+1}$ (and hence $\phi_n(\boldsymbol{R}_n^c) = \boldsymbol{R}_{n+1}^c$) such that

$$I^{n+1} \circ \phi_n(\boldsymbol{x}) = I^n(\boldsymbol{x}) + \mu_n(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \Omega, \tag{1}$$

where μ_n denotes a stationary zero-mean Gaussian white noise process with variance σ_n^2, with μ_i and μ_j independent for $i \neq j$. Most frame-to-frame tracking algorithms estimate $\boldsymbol{R}_{n+1}$ by first estimating ϕ_n; based on the estimate $\hat{\phi}_n$ of ϕ_n, the estimate $\hat{\boldsymbol{R}}_{n+1}$ of $\boldsymbol{R}_{n+1}$ is then computed as $\hat{\boldsymbol{R}}_{n+1} = \hat{\phi}_n(\boldsymbol{R}_n)$. To make the problem of estimating ϕ_n tractable, strong assumptions are then imposed on Γ; in particular, Γ is often assumed to be a small-dimensional space (e.g., the group of translations). Taking a different approach and, following [3], computing a Bayesian estimate of $\boldsymbol{R}_{n+1}$, it can be shown [3] that the maximum a posteriori estimator of $\boldsymbol{R}_{n+1}$ is, under some probabilistic assumptions, given by the solution of the following MAP estimation/energy minimization problem:

$$\begin{aligned}\hat{\boldsymbol{R}}_{n+1} &= \arg\max_{\boldsymbol{R} \subset \Omega} P(\boldsymbol{R}_{n+1} = \boldsymbol{R} | I^n, I^{n+1}, \boldsymbol{R}_n) \\ &= \arg\min_{\boldsymbol{R} \subset \Omega} \{-\log P(I^{n+1} | I^n, \boldsymbol{R}_n, \boldsymbol{R}_{n+1} = \boldsymbol{R}) \\ &\quad -\log P(\boldsymbol{R}_{n+1} = \boldsymbol{R} | I^n, \boldsymbol{R}_n)\}\end{aligned}$$

where the likelihood $-\log P(I^{n+1} | I^n, \boldsymbol{R}_n, \boldsymbol{R}_{n+1} = \boldsymbol{R})$ is, up to an additive constant, given by:

$$\begin{aligned}-\log P(I^{n+1} | I^n, \boldsymbol{R}_n, \boldsymbol{R}_{n+1} = \boldsymbol{R}) &= \int_{\boldsymbol{R}} \xi_n(\boldsymbol{x}) d\boldsymbol{x} \\ &+ \int_{\boldsymbol{R}^c} \eta_n(\boldsymbol{x}) d\boldsymbol{x},\end{aligned}$$

where $\boldsymbol{R}^c$ is the complement of $\boldsymbol{R}$ in Ω, and the functions ξ_n, η_n are given by:

$$\begin{aligned}\xi_n(\boldsymbol{x}) &= \inf_{\{\boldsymbol{z} : \|\boldsymbol{z}\| \leq \alpha, \boldsymbol{x}+\boldsymbol{z} \in \boldsymbol{R}_n\}} \frac{(I^{n+1}(\boldsymbol{x}) - I^n(\boldsymbol{x}+\boldsymbol{z}))^2}{2\sigma_n^2} \\ \eta_n(\boldsymbol{x}) &= \inf_{\{\boldsymbol{z} : \|\boldsymbol{z}\| \leq \alpha, \boldsymbol{x}+\boldsymbol{z} \in \boldsymbol{R}_n^c\}} \frac{(I^{n+1}(\boldsymbol{x}) - I^n(\boldsymbol{x}+\boldsymbol{z}))^2}{2\sigma_n^2},\end{aligned}$$

where α is the maximum range of frame-to-frame displacement of each image point. Choosing the prior $P(\boldsymbol{R}_{n+1} = \boldsymbol{R} | I^n, \boldsymbol{R}_n)$ to

This work was supported by the Natural Sciences and Engineering Research Council of Canada under Strategic Grant STR224122.

be a function of the boundary length of $\boldsymbol{R}$ as in [3] and letting the closed planar curve $\vec{\gamma}:[0,1]\to\mathbb{R}^2, s\mapsto\vec{\gamma}(s)$ be the estimator of the boundary $\partial\boldsymbol{R}_{n+1}$ of $\boldsymbol{R}_{n+1}$ (with $\boldsymbol{R}_{\vec{\gamma}}$ the subset of Ω bounded by $\vec{\gamma}$, $\boldsymbol{R}^c_{\vec{\gamma}}$ its complement, and ds arclength) finally leads to the following energy minimization problem:

$$\vec{\gamma}^\star = \arg\min_{\vec{\gamma}} E(\vec{\gamma}|I^n, I^{n+1}, \boldsymbol{R}_n)$$

where

$$\begin{aligned} E(\vec{\gamma}|I^n, I^{n+1}, \boldsymbol{R}_n) &= \int_{\boldsymbol{R}_{\vec{\gamma}}} \xi_n(\boldsymbol{x})d\boldsymbol{x} \\ &+ \int_{\boldsymbol{R}^c_{\vec{\gamma}}} \eta_n(\boldsymbol{x})d\boldsymbol{x} \\ &+ \lambda_L \oint_{\vec{\gamma}} ds. \end{aligned}$$

The minimization of the functional $\vec{\gamma}\mapsto E(\vec{\gamma}|I^n, I^{n+1}, \boldsymbol{R}_n)$ is performed by embedding the curve $\vec{\gamma}:[0,1]\to\mathbb{R}^2$ in a one-parameter family $\vec{\gamma}:[0,1]\times\mathbb{R}^+\to\mathbb{R}^2$ of plane curves such that $\vec{\gamma}(.,\infty)=\lim_{\tau\to\infty}\vec{\gamma}(.,\tau)$ be a minimum of E. Such a family is constructed by prescribing the evolution of $\vec{\gamma}$ according to the Euler-Lagrange descent equation of E, that is:

$$\frac{d\vec{\gamma}(s,\tau)}{d\tau} = -[\xi_n(\vec{\gamma}(s,\tau)) - \eta_n(\vec{\gamma}(s,\tau)) + \lambda_L\kappa_\gamma(s,\tau)]\vec{N}(s,\tau)$$

where $\vec{N}(s,\tau)$ is the unit normal to $\vec{\gamma}(.,\tau)$ at s pointing outward of $\boldsymbol{R}_{\vec{\gamma}}$, and κ_γ is the curvature of $\vec{\gamma}$. The level set representation of $\vec{\gamma}$ is given by defining a function $u:\Omega\times\mathbb{R}^+\to\mathbb{R}$ such that $\forall\tau\in\mathbb{R}^+$, $\vec{\gamma}([0,1],\tau)$ is the zero-level set $\{\boldsymbol{x}\in\Omega|u(\boldsymbol{x},\tau)=0\}$ of u. By convention, we take $u>0$ inside of $\vec{\gamma}$ and $u<0$ outside. It can be easily shown that for the zero level set of u to evolve according to the evolution equation of $\vec{\gamma}$, u itself must evolve according to the following partial differential equation:

$$\frac{\partial u(\boldsymbol{x},\tau)}{\partial\tau} = -[\xi_n(\boldsymbol{x}) - \eta_n(\boldsymbol{x}) + \lambda_L\kappa_u(\boldsymbol{x},\tau)]\|\vec{\nabla}u(\boldsymbol{x},\tau)\|,$$

where $\kappa_u = -\vec{\nabla}\cdot\frac{\vec{\nabla}u}{\|\vec{\nabla}u\|}$. The maximum a posteriori estimate $\hat{\boldsymbol{R}}_{n+1}$ of $\boldsymbol{R}_{n+1}$ is then given by the subset $\{\boldsymbol{x}\in\Omega|u(\boldsymbol{x},\infty)>0\}$ of Ω. This level set partial differential equation is the basic level set equation for tracking that we shall generalize to multi-frame tracking. As will be seen, the generalized equation is strikingly similar to this basic equation.

2.2. Tracking as joint space-time segmentation

Consider given a sequence $(I^n)_{n=0}^N$ of images, with t_n corresponding to time instant t_n, $t_N=T$, and $t_n<t_{n+1},\forall n$. Let $\boldsymbol{R}_0\in\Omega$ be a region in the image at time t_0 (I^0) which we wish to track for the rest of the sequence, i.e., for $n\geq1$. We thus wish to estimate the family $(\boldsymbol{R}_n)_{n=1}^N$ of subsets of Ω corresponding to $\boldsymbol{R}_0$, i.e., such that $\boldsymbol{R}_k$ be the region in the image at time t_k corresponding to $\boldsymbol{R}_0$. We assume that for each $n\geq1$, the regions in the pair $(\boldsymbol{R}_n,\boldsymbol{R}_0)$ are related through the basic model given in equation (1). Formulating this problem as a Bayesian estimation problem, we can write:

$$\begin{aligned} (\hat{\boldsymbol{R}}_n)_{n\geq1} &= \arg\max_{(\boldsymbol{R}_n)_{n\geq1}} P((\boldsymbol{R}_n)_{n\geq1}|(I^n)_n,\boldsymbol{R}_0) \\ &= \arg\min_{(\boldsymbol{R}_n)_{n\geq1}} \{-\log P((I^n)_n|I^0,\boldsymbol{R}_0,(\boldsymbol{R}_n)_{n\geq1}) \\ &- \log P((\boldsymbol{R}^n)_{n\geq1}|I^0,\boldsymbol{R}_0)\} \end{aligned}$$

At this point, we assume that the $(I^n)_{n\geq1}$ are conditionally independent given $I^0,\boldsymbol{R}_0$, and $(\boldsymbol{R}_n)_{n\geq1}$. Although this assumption does not hold in general, it is a reasonable assumption which makes the problem tractable. Furthermore, this assumption holds whenever the family Γ of transformations is small enough so that the knowledge of $\boldsymbol{R}_0$ and $(\boldsymbol{R}_n)_{n\geq1}$ allow the computation of the individual transformations ϕ_n, and is closely related to the conditional independence assumption formulated in [3]. This independence assumption allows us to write the negative log likelihood as:

$$\begin{aligned} &-\log P((I^n)_{n\geq1}|I^0,\boldsymbol{R}_0,(\boldsymbol{R}_n)_{n\geq1}) \\ &= \sum_{n\geq1} -\log P(I^n|I^0,\boldsymbol{R}_0,(\boldsymbol{R}_n)_{n\geq1}) \\ &= \sum_{n\geq1}\frac{1}{2\sigma_n^2}\int_{\boldsymbol{R}_n}\xi_n(\boldsymbol{x})d\boldsymbol{x} + \sum_{n\geq1}\frac{1}{2\sigma_n^2}\int_{\boldsymbol{R}_n^c}\eta_n(\boldsymbol{x})d\boldsymbol{x}, \end{aligned}$$

where ξ_n and η_n are now given by

$$\begin{aligned} \xi_n(\boldsymbol{x}) &= \inf_{\{\boldsymbol{z}:\|\boldsymbol{z}\|\leq n\alpha,\boldsymbol{x}+\boldsymbol{z}\in\boldsymbol{R}_0\}}(I^n(\boldsymbol{x})-I^0(\boldsymbol{x}+\boldsymbol{z}))^2 \\ \eta_n(\boldsymbol{x}) &= \inf_{\{\boldsymbol{z}:\|\boldsymbol{z}\|\leq n\alpha,\boldsymbol{x}+\boldsymbol{z}\in\boldsymbol{R}_0^c\}}(I^n(\boldsymbol{x})-I^0(\boldsymbol{x}+\boldsymbol{z}))^2. \end{aligned}$$

This definition of the functions ξ_n and η_n merely reflects the fact that tracking in frame I^n ($n\geq1$) is based on image I^0 and region $\boldsymbol{R}_0$. Note also that for technical reasons (justified below) the variance σ_n^2 has been excluded from the definition of ξ_n and η_n. We now let N tend to infinity, corresponding to a refinement of the discretization of the time interval $[0,T]$. We also assume, without any loss of generality, that all images in the sequence are equally spaced temporally, that is, for all $n=0,\ldots,N-1$, $t_{n+1}-t_n$ is a constant δ_N which is a function only of N. Clearly, as N tends to ∞, the temporal spacing δ_N between consecutive image frames goes to 0, and the sequence (I^n) can be viewed as a function $I:\Omega\times[0,T]\to\mathbb{R}$. By analogy to the construction of Brownian motion, we assume $\sigma=1/\sqrt{2\delta_N}$. As a result, as $N\to\infty$,

$$\begin{aligned} \sum_{n\geq1}\frac{1}{2\sigma_n^2}\int_{\boldsymbol{R}_n}\xi_n(\boldsymbol{x})d\boldsymbol{x} &\to \int_V \xi(\boldsymbol{x},t)d\boldsymbol{x}dt, \\ \sum_{n\geq1}\frac{1}{2\sigma_n^2}\int_{\boldsymbol{R}_n^c}\eta_n(\boldsymbol{x})d\boldsymbol{x} &\to \int_{V^c} \eta(\boldsymbol{x},t)d\boldsymbol{x}dt, \end{aligned}$$

where $V\subset\Omega\times[0,T]$ is the volume in the spatio-temporal domain spanned by the regions $\boldsymbol{R}_n$, and the functions ξ,η are given by:

$$\begin{aligned} \xi(\boldsymbol{x},t) &= \inf_{\{\boldsymbol{z}:\|\boldsymbol{z}\|\leq t\alpha,\boldsymbol{x}+\boldsymbol{z}\in\boldsymbol{R}_0\}}(I(\boldsymbol{x},t)-I(\boldsymbol{x}+\boldsymbol{z},0))^2 \quad (2) \\ \eta(\boldsymbol{x},t) &= \inf_{\{\boldsymbol{z}:\|\boldsymbol{z}\|\leq t\alpha,\boldsymbol{x}+\boldsymbol{z}\in\boldsymbol{R}_0^c\}}(I(\boldsymbol{x},t)-I(\boldsymbol{x}+\boldsymbol{z},0))^2. \end{aligned}$$

The preceding analysis suggests that if the image sequence is to be viewed as a mapping $I:\Omega\times[0,T]\to\mathbb{R}$, and the regions to be tracked as a *volume* V in the spatio-temporal domain $\Omega\times[0,T]$, then the negative log likelihood function $-\log P((I_t)_{t\geq0}|I^0,V)$ is given by the expression

$$-\log P((I_t)_{t>0}|I^0,V) = \int_V \xi(\boldsymbol{x},t)d\boldsymbol{x}dt + \int_{V^c}\eta(\boldsymbol{x},t)d\boldsymbol{x}dt$$

with η,ξ given in (2). For the negative log of the prior probability $-\log P(V|I^0,\boldsymbol{R}_0)$ we use a prior which favors volumes with minimal bounding area such as used in [5], yielding (up to an additive constant) the expression:

$$-\log P(V|I^0, \boldsymbol{R}_0) = \lambda \int_{\partial V} d\sigma$$

where $d\sigma$ is the element of surface area and ∂V the boundary of V. Therefore, given an image sequence $I : \Omega \times [0, T] \to \mathbb{R}$ and a region $\boldsymbol{R}_0 \subset \Omega$ in the image $I(., 0)$ corresponding to $t = 0$, the problem of multi-frame region tracking can be expressed as the problem of minimizing the energy functional

$$\begin{aligned} E[V|I, \boldsymbol{R}_0] &= \int_V \xi(\boldsymbol{x}, t) d\boldsymbol{x} dt \\ &+ \int_{V^c} \eta(\boldsymbol{x}, t) d\boldsymbol{x} dt \\ &+ \lambda \int_{\partial V} d\sigma \end{aligned} \quad (3)$$

over all spatio-temporal volumes $V \subset \Omega \times [0, T]$. The tracked region at time t is then given by the restriction to time t of the volume V which minimizes E. The computation of the spatio-temporal volume V which minimizes E is performed by embedding V in a one-parameter family $(V_\tau)_\tau$ of subsets of $\Omega \times [0, T]$, with the bounding surface S_τ of V_τ satisfying the Euler-Lagrange descent equation of the functional (3), given by [5]:

$$\frac{dS_\tau}{d\tau}(\boldsymbol{x}, t, \tau) = -[\xi(\boldsymbol{x}, t) - \eta(\boldsymbol{x}, t) + \lambda H(\boldsymbol{x}, t, \tau)]\vec{N}(\boldsymbol{x}, t, \tau),$$

where $H = H(\boldsymbol{x}, t, \tau)$ is the mean curvature of S_τ and $\vec{N}$ the outward unit normal to S_τ. The behavior of the proposed multi-frame tracking algorithm is evident in this evolution equation: A point $(\boldsymbol{x}, t)$ of the spatio-temporal domain which is closer in intensity to a point in $\boldsymbol{R}_0$ than to any point outside $\boldsymbol{R}_0$ will have $\xi(\boldsymbol{x}, t) \leq \eta(\boldsymbol{x}, t)$, by definition of the functions ξ and η; as a result, $\xi(\boldsymbol{x}, t) - \eta(\boldsymbol{x}, t)$ will be negative, which, omitting the curvature H, will encourage the bounding surface S of the volume V to grow, englobing $(\boldsymbol{x}, t)$. If, on the other hand, $(\boldsymbol{x}, t)$ is closer in intensity to a point in $\boldsymbol{R}_0^c$ than to any point inside $\boldsymbol{R}_0$, $\xi(\boldsymbol{x}, t) - \eta(\boldsymbol{x}, t)$ will be positive, encouraging the volume V to shrink at that point. The level set evolution equations for the multi-frame tracking algorithm are obtained by considering S as the zero level surface of a function $u : \Omega \times [0, T] \to \mathbb{R}$. The main advantages of solving active surface evolution equations via level set partial differential equations is the numerical stability of level set resolution schemes and the fact that the implicit surface representation provided by level sets allows for changes in surface topology. In the context of region tracking, this is of fundamental importance, as a particular region may split into numerous other regions during its evolution, or conversely a number of disjoint regions may merge.

The level set partial differential equation for the proposed multi-frame tracking algorithm is then given by:

$$\frac{\partial u}{\partial \tau}(\boldsymbol{x}, t, \tau) = -[\xi(\boldsymbol{x}, t) - \eta(\boldsymbol{x}, t) + \lambda H(\boldsymbol{x}, t, \tau)]\|\vec{\nabla} u(\boldsymbol{x}, t, \tau)\|$$

where $H = -\vec{\nabla} \cdot \frac{\vec{\nabla} u}{\|\vec{\nabla} u\|}$. Note the similarity of this equation with the level set evolution equation corresponding to frame-by-frame tracking.

3. EXPERIMENTAL RESULTS

We illustrate our tracking algorithm on two particularly challenging sequences of real images. The first sequence is the *Red Car sequence*. Four images are shown in Figure 1, the initial frame, the final frame (28), and intermediate frames 9, 25. The images are 400×300. The moving objects are two cars driving in opposite directions. Figure 2 shows the evolution of the zero-level surface, time being the vertical axis in the figure. The first image shows the initial surface, the second one after 3,000 iterations, and the last one at convergence after 10,000 iterations. Figure 3 shows the results of tracking, at convergence of the algorithm. The image of both cars is accurately outlined throughout the sequence.

The second sequence is the *Embryo sequence*. Figure 4 shows the initial frame, the final frame (61), and intermediate frames 12, 41. The images are 400×300. This is a challenging sequence for motion tracking algorithms, particularly because of the significant variation in shape of the moving object. Figure 5 shows the evolution of the zero-level surface. The first image is the initial surface, the second is the surface after 250 iterarions, and the last one is the surface at convergence after 500 iterarions. Finally, Figure 6 shows the results of tracking. The embryo has been properly tracked in spite of the significant change in shape over the sequence.

4. REFERENCES

[1] S. Jehan-Besson, M. Barlaud, G. Aubert, “Detection and tracking of moving objects using a new level set based method,” in *Proc. Int. Conf. Pattern Recognition*, Barcelona, 2000.

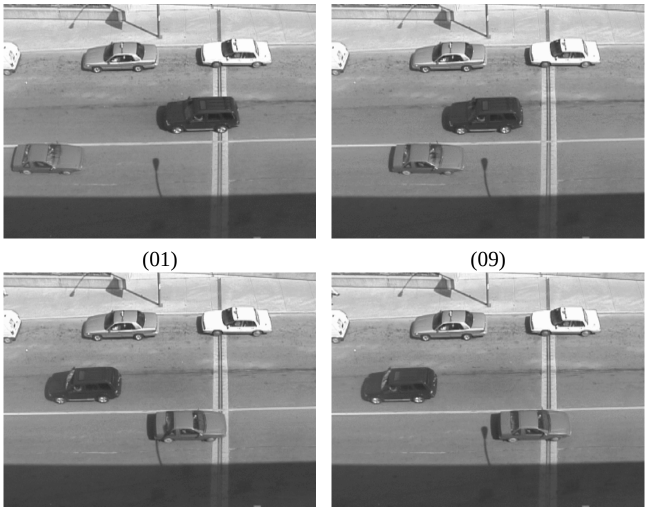
[2] N. Paragios and R. Deriche, “Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no.3 , pp. 266-280, 2000.

[3] A.-R. Mansouri, “Region tracking via level set PDEs without motion computation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, July 2002.

[4] A.R. Mansouri, A. Mitiche, “Region tracking via local statistics and level set PDEs”, *Proc. Int. Conf. Image Processing*, 2002.

[5] A. Mitiche, R. Feghali, A.-R. Mansouri, “Tracking Moving Objects As Spatio-Temporal Boundary Detection”, *Proceedings of the 5th IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI’02)*, Santa Fe, April 2002.

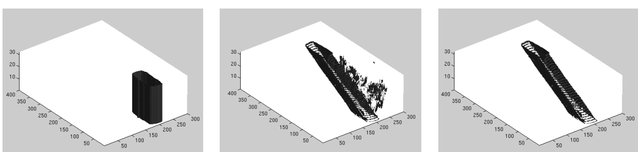
[6] M. Bertalmio, G. Sapiro, and G. Randall, “Morphing active contours: a geometric approach to topology-independent image segmentation and tracking,” in *Proc. Int. Conf. Image Processing*, vol. III, pp. 318-322, 1998.



(01) (09)

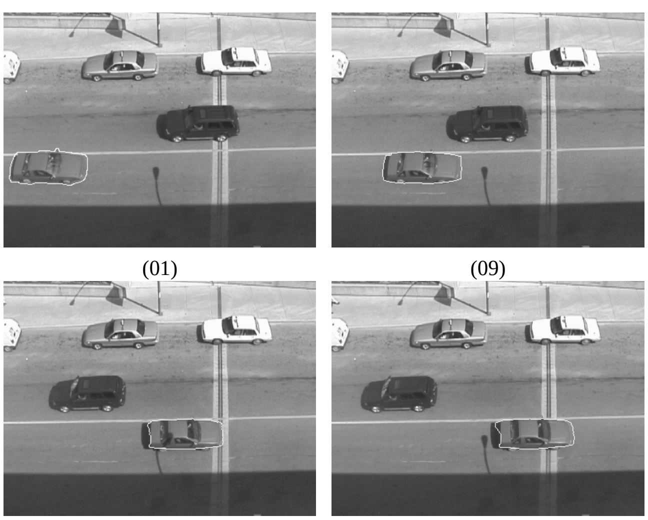
(25) (28)

Fig. 1. original red car sequence



(01) (03) (05)

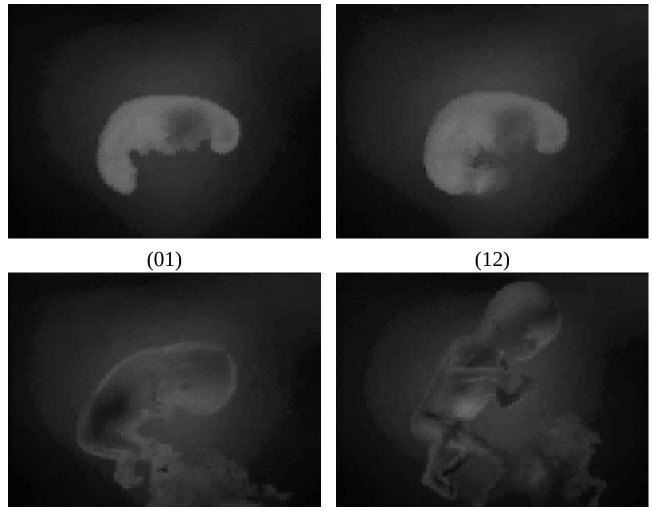
Fig. 2. zero level set evolution in 3D



(01) (09)

(25) (28)

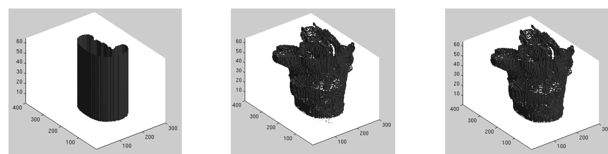
Fig. 3. red car tracking



(01) (12)

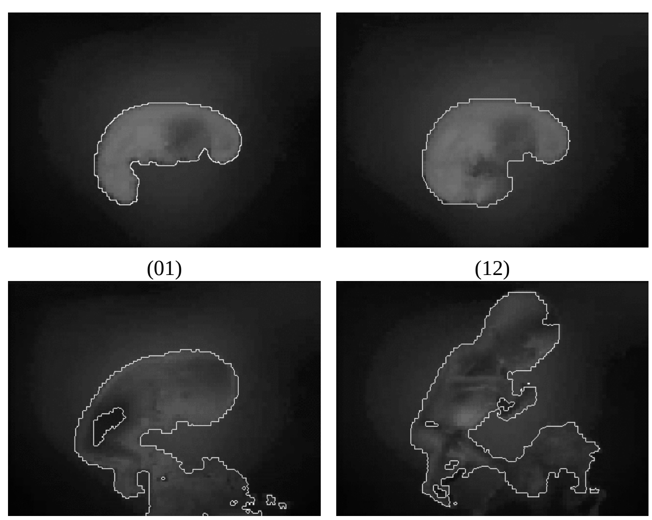
(41) (61)

Fig. 4. original embryo sequence



(0) (250) (500)

Fig. 5. zero level set evolution



(01) (12)

(41) (61)

Fig. 6. embryo tracking