TRACKING WITH SIMULTANEOUS CAMERA MOTION SUBSTRACTION BY LEVEL SET SPATIO-TEMPORAL SURFACE EVOLUTION

Rosario Feghali

Communications Research Center
3701 Carling Ave
Ottawa, Ontario, Canada, K2H 8S2
rosario.feghali@crc.ca

Amar Mitiche

INRS-Télécommunications
800, de La Gauchetière West, suite 6900
Montreal, Québec, Canada, H5A 1K6
mitiche@inrs-telecom.uquebec.ca

ABSTRACT

The purpose of this study is to investigate a method of tracking moving objects with a moving camera. This method interprets tracking as detection of the surface generated by motion boundaries in the spatio-temporal domain and estimates simultaneously the motion induced by camera movement. The problem is formulated as a Bayesian motion-based partitioning problem in the spatio-temporal domain of the image sequence. An energy functional is derived from the Bayesian formulation. The Euler Lagrange descent equations determine simultaneously an estimate of the image motion field induced by camera motion and an estimate of the spatio-temporal motion boundary surface. The Euler-Lagrange equation corresponding to the surface is expressed as a level set partial differential equation for topology independence and numerically stable implementation. The method has a simple initialization and allows the tracking of multiple objects with non simultaneous motions. Optical velocities on motion boundaries can be estimated from geometrical properties of the motion boundary surface. Several examples of experimental verification are given using synthetic and real image sequences.

1. INTRODUCTION

Tracking in the presence of a moving camera is a challenging problem in computer vision. When a camera moves, it generates motion over the entire image positional array and the tracking problem cannot be solved directly by simple motion detection as in the case of a static camera.

Methods of tracking with a moving camera fall in one of two categories. In one category, methods assumes that camera motion is given as an input or that the background scene has distinctive image properties. This leads to constraints that are valid only in background regions. Moving objects regions violate these constraints and can be thus detected [7]. Methods in the other category [4] perform tracking generally in three consecutive steps: (a) they assume that background motion is represented by a parametric model, and they compute an estimate of background motion parameters, (b) they detect moving objects based on image motion after compensation for camera motion, and, (c) they achieve tracking using a frame by frame temporal correspondence. The operations in sequential multi-step methods are not integrated by feedback and are therefore prone to cumulative errors from one step to another.

Tracking methods can be contour based or region based. Quite recently, both contour and region tracking have been formulated using level sets partial differential equations (PDE) [2, 3, 9, 1]. The level sets formalism is momentous for several reasons: (1) It follows a well posed variational statement of tracking where assumptions are transparent, (2) it accounts for topology changes during contour evolution and, (3) it can be implemented by stable numerical methods.

Existing methods of tracking formulated via level sets track moving image objects frame by frame. They require that these objects be segmented beforehand by some external process [9, 2, 3]. Some schemes assume that a good estimate of image motion is available at each instant of time, to be used as data [1], or that the background has known properties to be used for identification [2]. Under other assumptions, the background intensity pattern contrasts strongly with the pattern of moving objects, and inter-frame intensity difference statistics invariant in time can be computed [9]. Existing PDE-based methods also share the following shortcomings: Except for [3], they are not valid when there is camera motion, and only to those objects that are identified as moving at the start of tracking; objects that come into motion after tracking of objects is started cannot be tracked.

This study is along the vein of our previous one on tracking by explicit processing in the spatio-temporal domain [5, 6]. In that study there was no reference to the image motion induced by camera movement and, therefore, no estimation of this motion. As in [5, 6], we start by stating the problem as a Bayesian motion-based partitioning problem in the spatio-temporal domain. Representing background motion by a parametric model, the approach simultaneously estimates the model parameters while evolving a spatio-temporal surface so that at the end of its evolution it enfolds the volume generated by moving objects and thus partitions the spatio-temporal domain into the background region on one hand and the foreground region of the moving objects on the other hand. Surface evolution is implemented via level sets partial differential equations to afford topology free and numerically stable solutions. Furthermore, optical velocities can be estimated along motion boundaries from geometrical properties of the spatio-temporal surface. The approach has the following characteristics: (a) it allows tracking with a moving camera; (b) it allows tracking of several objects that have non-simultaneous motions; (c) it does not require prior estimation of camera motion; (d) it is implemented

This work was done while the main author was a PhD candidate at INRS-Télécommunications.

In Proceedings ICIP 2003, Sept. 14-17, 2003, Barcelona, Spain

via level set PDE's to allow topology free processing and numerically stable computation; finally, (e) it allows explicit recovery of motion along motion boundaries.

The remainder of this paper is organized as follows. Section 2 gives the formulation of the proposed tracking approach. Section 3 presents discussions and limitations of the approach. Section 4 gives experimental results and section 5 contains a conclusion.

2. FORMULATION

Let I be an image sequence defined over $D = \Omega \times]0, T[$ into $\mathbb{R}^+$, where $]0, T[$ is the time interval of the image sequence, and Ω an open subset of $\mathbb{R}^2$. Let S be a closed surface in D, R_S the region enclosed by S, $R_S^c = D \setminus R_S$ its complement in D, and the partition $\mathcal{P}_S = \{R_S, R_S^c\}$. We assume that background motion can be fully characterized by a set of parameters defined by the vector $\boldsymbol{\theta}$. The Maximum a Posteriori (MAP) estimate of $(S, \boldsymbol{\theta})$ is:

$$\begin{aligned}(\hat{S}, \hat{\boldsymbol{\theta}}) &= \arg\max_{S,\boldsymbol{\theta}} P((\mathcal{P}_S, \boldsymbol{\theta})/m) \\ &= \arg\max_{S,\boldsymbol{\theta}} \frac{P(m/(\mathcal{P}_S, \boldsymbol{\theta}))P(\mathcal{P}_S, \boldsymbol{\theta})}{P(m)}\end{aligned}$$

where m is a motion measurement. $P(m)$ is ignored because it is independent of $\boldsymbol{\theta}$ and S. $P(m/(\mathcal{P}_S, \boldsymbol{\theta})$ is the observation data term, and $P(\mathcal{P}_S, \boldsymbol{\theta})$ is the *a priori* term. Assuming conditional independence of the motion measurement, we find that maximizing this probability is equivalent to minimizing the following functional:

$$\begin{aligned}E(S, \boldsymbol{\theta}) = &-\int_{R_S} \log P(m(\boldsymbol{x})/(\mathcal{P}_S, \boldsymbol{\theta}))d\boldsymbol{x} \\ &-\int_{R_S^c} logP(m(\boldsymbol{x})/(\mathcal{P}_S, \boldsymbol{\theta}))d\boldsymbol{x} \\ &-\log P(\mathcal{P}_S, \boldsymbol{\theta})\end{aligned} \quad (1)$$

where $\boldsymbol{x} = (x, y, t)$. The first two terms on the right of (1) will be defined by the observation model. The last term will be defined by the model of prior.

2.1. Observation Model

Assuming small range motion so that motion is of small extent between consecutive instants of observation, let m be the normal component, $w_\perp$, of optical velocity, given by:

$$w_\perp = \begin{cases} \frac{-I_t}{\|\nabla I\|} & \text{for } \|\nabla I\| \neq 0 \\ 0 & \text{for } \|\nabla I\| = 0 \end{cases} \quad (2)$$

∇I and I_t being the spatial gradient and the temporal derivative of I, respectively. Define

$$w_\perp^* = w_\perp - w_{c\perp} \quad (3)$$

where $w_{c\perp}$ is the normal component of the image motion due to camera motion. $w_\perp^*$ is a function of $\boldsymbol{\theta}$, the parameters of the image motion due to camera motion. In a noiseless image sequence, $w_\perp^*$ is zero in the background, while in the regions of moving objects, it denotes motion activity due to moving objects intrinsic motions. We choose the following observation model:

$$P(m(\boldsymbol{x})/(\mathcal{P}_S, \boldsymbol{\theta})) \propto \begin{cases} e^{-\alpha e^{-(w_\perp^*(\boldsymbol{\theta}))^2}} & \text{for } \boldsymbol{x} \in R_S \\ e^{-\beta(w_\perp^*(\boldsymbol{\theta}))^2} & \text{for } \boldsymbol{x} \in R_S^c \end{cases} \quad (4)$$

where α and β are positive real constants and $\propto$ is the proportional to symbol. This choice will favor (in terms of higher probability values) partitions where high residual motion activity occurs in the region enclosed by the surface and low residual motion activity in the complement of this region.

2.2. Prior model

The model of prior is chosen to have a smoothing effect on the surface by favoring surface estimates that have small area.

$$P(\mathcal{P}_S, \boldsymbol{\theta})) \propto e^{-\lambda \int_S d\sigma} \quad (5)$$

2.3. Euler-Lagrange equations

Maximizing the *a posteriori* probability $P((\mathcal{P}_S, \boldsymbol{\theta})/m)$ is equivalent to minimizing the following energy functional:

$$\begin{aligned}E(S, \boldsymbol{\theta}) = \alpha \int_{R_S} e^{-(w_\perp^*(\boldsymbol{\theta}))^2} d\rho + \beta \int_{R_S^c} (w_\perp^*(\boldsymbol{\theta}))^2 d\rho \\ +\lambda \int_S d\sigma\end{aligned} \quad (6)$$

Background motion parameters represented by $\boldsymbol{\theta}$ intervene in the first two terms of equation (6). The third term depends only on geometrical properties of the spatio-temporal surface. Here following, we assume that background motion is translational. Therefore, $\boldsymbol{\theta} = (u, v)$, where u and v are the horizontal and vertical components of this motion, respectively. Assuming that these parameters are constant during the observation period, the minimization of $E(S, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and S yields the following set of descent equations:

$$\begin{cases} \frac{\partial u}{\partial \tau} = & -\alpha \int_{R_S} \frac{2I_x}{\|\nabla I\|} (\frac{I_t + uI_x + vI_y}{\|\nabla I\|}) e^{-(\frac{I_t + uI_x + vI_y}{\|\nabla I\|})^2} d\rho \\ & -\beta \int_{R_S^c} \frac{2I_x}{\|\nabla I\|} (\frac{I_t + uI_x + vI_y}{\|\nabla I\|}) d\rho \\ \frac{\partial v}{\partial \tau} = & -\alpha \int_{R_S} \frac{2I_y}{\|\nabla I\|} (\frac{I_t + uI_x + vI_y}{\|\nabla I\|}) e^{-(\frac{I_t + uI_x + vI_y}{\|\nabla I\|})^2} d\rho \\ & -\beta \int_{R_S^c} \frac{2I_y}{\|\nabla I\|} (\frac{I_t + uI_x + vI_y}{\|\nabla I\|}) d\rho \\ \frac{\delta S}{\delta \tau} = & -(2\lambda\kappa + \alpha e^{-(w_\perp^*(u,v))^2} - \beta(w_\perp^*(u,v))^2)\boldsymbol{n} \end{cases} \quad (7)$$

where $\boldsymbol{n}$ is the outward unit normal to S, κ is its mean curvature, and τ is the algorithmic time.

2.4. Level sets representation

Execution of the descent equation with respect to S by explicit representation of S as a set of points cannot allow changes in the topology of S. To obtain a topology-free evolution of this surface, and a numerically stable scheme, we adopt a level-set representation [10]. Surface S is represented implicitly as as the zero level of a one-parameter family of functions U, indexed by algorithmic time τ:

$$(\forall \tau) \quad U(x(\tau), y(\tau), t(\tau), \tau) = 0 \quad (8)$$

Using the third equation of (7), and defining U to be negative inside S and positive outside ($U = 0$ on S, by definition) so that ∇U is oriented as $\boldsymbol{n}$, i.e. $\boldsymbol{n} = \frac{\nabla U}{\|\nabla U\|}$, we get:

$$\frac{\partial U}{\partial \tau} = (2\lambda\kappa + \alpha e^{-(w_\perp^*(u,v))^2} - \beta(w_\perp^*(u,v))^2))\|\nabla u\| \quad (9)$$

According to (7) and (9), S evolves normal to itself at a speed:

$$\boldsymbol{s} = -(2\lambda\kappa + \alpha e^{-(w_\perp^*(u,v))^2} - \beta(w_\perp^*(u,v))^2)\boldsymbol{n} \quad (10)$$

Let the initial position of S be a surface S_0 that subsumes the volume generated by moving objects. With the proper choice of the constant coefficients, we will have the following behavior of S. While S is in the background, we will have $w^*_\perp \approx 0$ and $\mathbf{s} \approx (-2\lambda\kappa - \alpha)\boldsymbol{n}$: S will move inward, remain smooth because of the curvature term, and the speed of evolution will vary little because of the constant term α. Whenever and wherever it reaches the boundary of a moving object, $\mathbf{s} \approx (-2\lambda\kappa + \beta(w^*_\perp)^2)\mathbf{n}$. The term $\beta|w^*_\perp|$ acts to prevent the surface from penetrating into the region of motion activity beyond the motion boundary, while $-2\lambda\kappa$ has the same spatio-temporal smoothing effect.

The method is summarized as follows

1. Initialize S_0 and $\boldsymbol{\theta}_0 = (u_0, v_0)$.
2. Perform one iteration of the first two descent equations for u and v in (7).
3. Evolve S using one iteration of the level-set PDE descent equation (9)
4. Return to step 2 until convergence.

2.5. Estimation of velocities along motion boundaries

Within this scheme, motion boundaries at time t are obtained by intersecting the spatio-temporal surface by the plane $z = t$. Optical velocities at motion boundaries can be recovered from the spatio-temporal surface as follows. Let $C : t \longrightarrow (x(t), y(t), t)$ be a motion boundary point trajectory in spatio-temporal space. The tangent vector to C is $\boldsymbol{T} = (\frac{dx}{dt}, \frac{dy}{dt}, \frac{dt}{dt}) = (u, v, 1)$, where (u, v) is the optical velocity. If we assume no occlusion and that the trajectory of a motion boundary point is located on the spatio-temporal surface, then the velocity vector along this trajectory is tangent to the surface and, therefore, is orthogonal to the surface normal $\boldsymbol{n}$, yielding the following geometric constraint on optical velocity.

$$\boldsymbol{n} \cdot \boldsymbol{T} = 0 \tag{11}$$

If $\boldsymbol{n} = (n_x, n_y, n_t)$ then, equation (11) determines the component of optical velocity in the direction of (n_x, n_y). This constraint is valid at any point where the spatio-temporal surface is regular.

To estimate optical velocity from the component given by (11) we add a regularization constraint where optical velocities are considered to vary smoothly along motion boundaries. We proceed with the Hildreth iterations. These iterations minimize:

$$E(u, v) = \int_\Gamma (n_x u + n_y v + n_t)^2 + \|\nabla u\|^2 + \|\nabla v\|^2 \tag{12}$$

where Γ is the contour along which optical flow is estimated.

3. EXPERIMENTAL VERIFICATION

We validate the method on natural and synthetic sequences. For every sequence tested we first show the spatio-temporal surface at several stages during its evolution. Second, we provide tracking results obtained by taking temporal cuts across the spatio-temporal surface. Finally, we show optical velocities estimated along motion boundaries. The surface velocity used to evolve the spatio-temporal surface is:

$$\mathbf{s} = \left(-e^{-w^*_\perp} + 10w^*_\perp - 5\kappa\right)\mathbf{n} \tag{13}$$

The Rotating Fish sequence is a synthetically generated sequence where a fish is rotating against a background with a translational motion. Figure 1 shows the evolution of the spatio-temporal surface. The surface is simply initialized to a parallelepipedic shape. Throughout its evolution, we simultaneously refine the estimation of the camera-induced motion parameters. At the end of its evolution, the surface enfolds the volume generated by the rotating fish. Figure 2 shows the tracking results. We note that locations of motion boundaries are accurately delineated specifically at acute corners of the fish. Figure 3 shows velocities along motion boundaries for the Rotating Fish sequence at different instants of time. The estimated field correctly denote the rotation movement of the fish. The Walker sequence, is a natural sequence where a man is walking on a street. A moving camera results in an apparent translational background motion. Figures 4 and 5 show the spatio-temporal surface evolution and tracking results. Because tracking is solely based on a motion activity measurement, motion boundaries do not include portions of the pedestrian that are static during the person's movement. For instance, the bottom parts of the pedestrian's legs exhibit relatively low motion at certain instants in the sequence and are excluded. Optical velocities along motion boundaries are presented in Figure 6 and are consistent with the pedestrian's overall motion trajectory.

4. CONCLUSION

We proposed a new approach to perform tracking with a moving camera. Tracking as spatio-temporal motion boundary detection has many advantages. These include the detection of non-simultaneous motions and simple initialization. In contrast with existing tracking methods, the estimation of camera-induced motion parameters is not performed as a separate step. Here, we exploit the existing coupling between both problems by performing a simultaneous estimation of camera-induced motion parameters. Based on geometrical properties of the spatio-temporal surface, a new scheme to estimate optical velocities along motion boundaries is proposed. Experimental results show the validity of the proposed tracking approach and its potential.

5. REFERENCES

[1] M. Bertalmio, G. Sapiro, and G. Randall, "Morphing Active Contours," IEEE Trans. PAMI, vol. 22, no. 7, pp. 733-737, July 2000.

[2] S. Jehan-Besson, M. Barlaud, G. Aubert, "Detection and tracking of moving objects using a new level set based method," ICPR, Barcelona, septembre 2000.

[3] A. Mansouri, "Region Tracking via Level Set PDEs without Motion Computation," IEEE Trans. on PAMI, Vol. 24, No. 7, pp. 947-961, July 2002.

[4] R. Mech, M. Wollborn, "A Noise Robust Method for 2D Shape Estimation of Moving Objects in Video Sequences Considering a Moving Camera," Signal Processing, Vol. 66, No. 2, April 1998, pp. 203-217

[5] A. Mitiche, R. Feghali, and A. Mansouri, "Tracking Moving Objects as Spatio-Temporal Boundary Detection", Southwest Symposium on Image Analysis and Interpretation, Santa Fe, NM, pp. 106-110, 2002

[6] A. Mitiche, R. Feghali, and A. Mansouri, "Motion Tracking As Spatio-temporal Motion Boundary Detection," INRS-EMT Technical Report, EMT-004-1115.

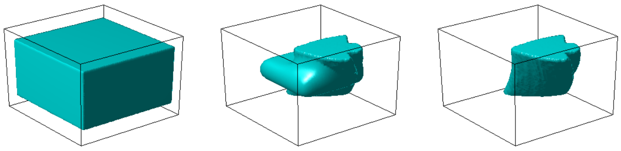

Fig. 1. The spatio-temporal surface evolution in the case of the Rotating Fish sequence (40 frames).

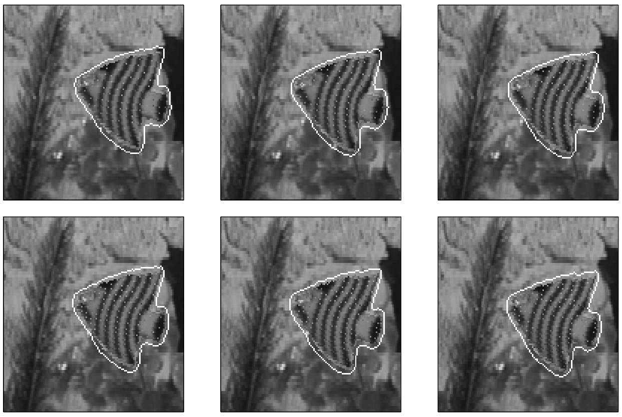

Fig. 2. Tracking results for Rotating Fish sequence, 1 frame interval, upper-left to lower-right.

[7] R.C. Nelson, "Qualitative detection of motion by a moving observer," Int. Journal of Computer Vision, vol. 7, no. 1, pp. 33-46, 1991.

[8] J.M. Odobez and P. Bouthemy, "Detection Of Multiple Moving Objects Using Multiscale MRF With Camera Motion Compensation," ICIP, Austin, Nov. 94.

[9] N. Paragios and R. Deriche, "Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects," IEEE Trans. PAMI, vol. 22, no. 3, pp. 266-280, March 2000.

[10] J. Sethian, "Level Set Methods," Cambridge University Press, 1996.

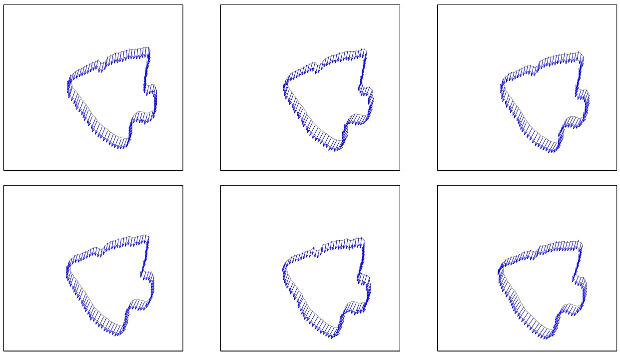

Fig. 3. Optical velocities along motion boundaries of Rotating Fish sequence, 1 frame interval, upper-left to lower-right.

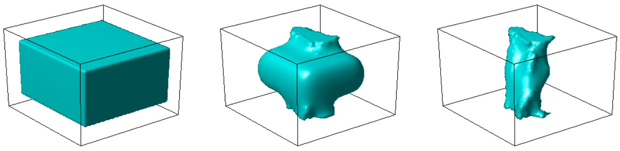

Fig. 4. The spatio-temporal surface evolution in the case of the Walker sequence. (34 frames)

Fig. 5. Tracking results for the Walker sequence, 1 frame interval, upper-left to lower-right.

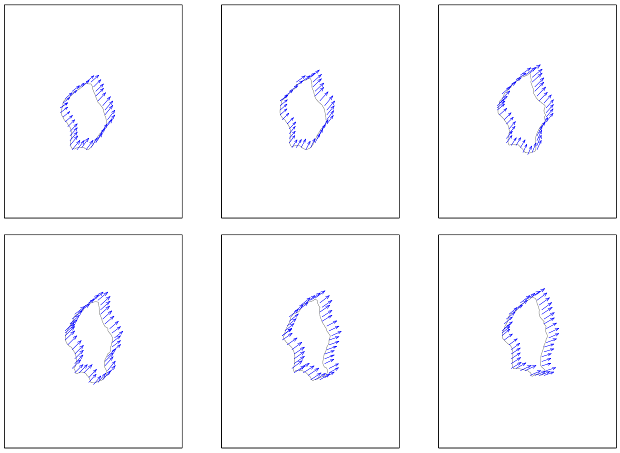

Fig. 6. Optical velocities along motion boundaries of Walker sequence, 1 frame interval, upper-left to lower-right.