Tracking Moving Objects as Spatio-Temporal Boundary Detection

A. Mitiche, R. Feghali, A. Mansouri

INRS-Télécommunications
Place Bonaventure, 900, de la Gauchetière Ouest, Niveau C
Montréal (Québec), Canada, H5A 1C6

Abstract.

The purpose of this study is to investigate tracking of moving objects in a sequence of images by detecting the surface generated by motion boundaries in the space-time domain. Estimation of this spatio-temporal surface is formulated as a Bayesian image partitioning problem. Minimization of the resulting energy functional seeks a solution biased toward smooth closed surfaces which coincide with motion boundaries, have small area, and partition the image into regions of contrasting motion activity. The Euler-Lagrange partial differential equations of minimization are expressed as level set evolution equations. The formulation does not require estimation of the image motion field and does not assume a known background. It allows multiple non-simultaneous independent motions to occur and, under some assumptions, can account for camera motion without prior estimation of this motion.

1. Introduction

In image motion analysis, tracking is the process of following image objects in their movement through an image sequence. Tracking plays a fundamental role in several applications: in robotics for visual servo-control; biomedecine to perform heart function measurements over a cardiac cycle; meteorology to follow the evolution of atmospheric disturbances; surveillance to secure sites of human activity; communications for televisual film editing or encoding of visual data; and man-machine communication for visual interfacing.

The level-set formalism [1][2] was introduced in tracking [3] [4] [5] to allow topology-independent formulations and numerically stable implementations. Current level-set formulations of tracking rest on strong assumptions. These assumptions include the availability of a good estimate of image motion, a known background and, objects having flat intensity patterns that contrast a flat background pattern. In addition, current methods share the following limitations: they are not valid if the image acquisition system moves and, once tracking of some objects is initiated, it is dedicated to these objects as no other object can be tracked that comes into motion subsequently.

The goal of this study is to investigate tracking of moving objects in an image sequence by detecting the surface generated by motion boundaries in the spatio-temporal domain. Estimation of this surface is stated as a MAP image partitioning problem, resulting in the minimization of an energy functional that seeks a spatio-temporal surface biased toward smooth closed surfaces which partition the image into regions of contrasting motion activity, coincide with motion boundaries, and have small area. The Euler-Lagrange descent equations are expressed as level set evolution equations to obtain a topology independent and numerically stable algorithm. The formulation does not assume a known background or require estimation of the image motion field. It allows multiple non-simultaneous independent motions to occur, and can account for camera motion without prior estimation of this motion. The analysis assumes short-range image motion. With moving cameras, it assumes that this short-range motion is locally approximately constant except at motion boundaries.

2. Formulation

Let I be an image sequence, $I : D = \Omega \times [0,T] \mapsto \mathbf{R}^+$, where Ω is an open subset of $\mathbf{R}^2$, and $[0,T]$ is an interval of time. Let m be a measurement related to image motion activity. Finally, let S be a closed surface in D, and R_S the region enclosed by S. Consider partitions $\mathcal{P}_S = \{R_S, R_S^c\}$ of D, where points in R_S and R_S^c are classified as corresponding to points on moving objects in space and to static objects in space, respectively. Using m as observation, the maximum a posteriori (MAP) estimate of S is:

$$\begin{aligned}\tilde{S} &= \arg\max_S P(\mathcal{P}_S|m) \\ &= \arg\max_S P(m|\mathcal{P}_S)P(\mathcal{P}_S)\end{aligned}$$

Assuming conditional independence of the motion measurement for $\mathbf{x} \neq \mathbf{y}$, gives:

$$P(m|\mathcal{P}_S) = \prod_{\mathbf{x}\in R_S} P(m(\mathbf{x})|\mathcal{P}_S) \prod_{\mathbf{x}\in R_S^c} P(m(\mathbf{x})|\mathcal{P}_S)$$

Therefore, the MAP estimation problem is equivalent to the following minimization problem:

$$\tilde{S} = \arg\min_S E(S)$$

where

$$\begin{aligned} E(S) = &- \int_{\mathbf{x}\in R_S} \log P(m(\mathbf{x})|\mathcal{P}_S) d\mathbf{x} \\ &- \int_{\mathbf{x}\in R_S^c} \log P(m(\mathbf{x})|\mathcal{P}_S) d\mathbf{x} - \log P(\mathcal{P}_S) \end{aligned} \tag{1}$$

The first two terms on the right-hand side of (1) are terms of conformity of the estimate to the measurements of motion activity. These terms are to be specified by an observation model. The last term is the prior, to be specified by a model of prior.

2.1 Observation model

Assuming small-range motion, let $m = W_\perp$, the normal component of optical velocity given by:

$$W_\perp = \begin{cases} \frac{-I_t}{\|\nabla I\|} & \text{for } \nabla I \neq 0 \\ 0 & \text{for } \nabla I = 0 \end{cases} \tag{2}$$

∇I and I_t being, respectively, the spatial gradient and the temporal derivative of I. We assume that $|W_\perp|$ is bounded and that its probability density function is an increasing function for the image of a moving object, a decreasing function for the complement. Under this assumption, the following observation model will bias the estimate toward surfaces that partition the spatio-temporal image into regions of contrasting motion activity:

$$P(W_\perp(\mathbf{x})|\mathcal{P}_S) \propto \begin{cases} e^{-\frac{\alpha}{1+|W_\perp|}} & \text{for } \mathbf{x} \in R_S \\ e^{-\beta|W_\perp|} & \text{for } \mathbf{x} \in R_S^c \end{cases} \tag{3}$$

where α and β are positive real constants, and $\propto$ is the proportional-to symbol. We are currently investigating parametric models that offer a more general definition of motion activity contrast.

2.2 Model of prior

Under the assumption that the image motion field is locally approximately constant everywhere except at motion boundaries, we differentiate the Horn and Schunck equation [6]

$$I_x u + I_y v + I_t = 0 \tag{4}$$

with respect to the spatial coordinates to obtain:

$$H \begin{bmatrix} u \\ v \end{bmatrix} + \nabla I_t \approx 0 \tag{5}$$

where H is the Hessian of I. Combining (4) and (5), we have:

$$\det(H) I_t - (\nabla I)^T H^* \nabla I_t \approx 0 \tag{6}$$

where T indicates transposition and H^* is the adjoint matrix of H. Let h be:

$$h = |\det(H) I_t - (\nabla I)^T H^* \nabla I_t| \tag{7}$$

For both static and moving cameras, h is expected to be approximately zero everywhere except at motion boundaries. Therefore, we following model of prior will bias the estimate toward surfaces that are motion boundaries and have small area:

$$P(\mathcal{P}_S) \propto e^{-(\lambda \int_S dS - \gamma \int_S h dS)} \tag{8}$$

where λ and γ are positive real constants.

With these models, the minimization problem can be written as:

$$\begin{aligned} \tilde{S} = Arg \min_S \{ &\alpha \int_{R_S} \frac{1}{1+|W_\perp|} dV + \beta \int_{R_S^c} |W_\perp| dV \\ &+ \lambda \int_S dS - \gamma \int_S h dS \} \end{aligned}$$

where the additive log of proportional-to constants have been removed. Since

$$\int_{R_S^c} |W_\perp| dV = \int_D |W_\perp| dV - \int_{R_S} |W_\perp| dV$$

and $\int_D |W_\perp|$ is independent of S, the minimization problem is:

$$\begin{aligned} \tilde{S} = Arg \min_S \{ &\alpha \int_{R_S} \frac{1}{1+|W_\perp|} dV - \beta \int_{R_S} |W_\perp| dV \\ &+ \lambda \int_S dS - \gamma \int_S h dS \} \end{aligned} \tag{9}$$

2.3 Euler-Lagrange equations and level set representation

The Euler-Lagrange partial differential equations that $\tilde{S}$ must satisfy are derived after transforming the volume integrals in the right-hand side of (9) into surface integrals over S using Gauss' divergence theorem. If we let

$$f = \frac{\alpha}{1+|W_\perp|} - \beta|W_\perp|$$
$$g = \lambda - \gamma h$$

the Euler-Lagrange descent equations to minimize (9) are:

$$\frac{d\Phi}{d\tau} = -(\nabla g \cdot \mathbf{n} + 2g\kappa + f)\mathbf{n} \qquad (10)$$

where, $\Phi : D \subset R^2 \mapsto S \subset R^3$, is a parametrization of S, $\mathbf{n}$ is the outward unit normal on S, and κ is the mean curvature. The right-hand side of (10) is independent of surface parametrization, as it should be, because we are interested in the image $S = \Phi(D)$ of Φ and not in Φ itself.

Execution of the descent equation by explicit representation of S as a set of points cannot implement changes in the topology of S. An alternative execution is via level sets where S is represented implicitely as the zero level set of a one-parameter family of functions u, indexed by τ:

$$(\forall\tau) \quad u \circ \Phi(\tau) = u(x(\tau), y(\tau), t(\tau), \tau) = 0 \qquad (11)$$

where x, y, and t are the spatio-temporal variables. Derivation of (11) with respect to τ yields the level set equation that dictates the evolution of u:

$$\nabla u \cdot \frac{d\Phi}{d\tau} + \frac{\partial u}{\partial \tau} = 0 \qquad (12)$$

or, using (10), and defining u to be positive inside S_R and negative outside ($u = 0$ on S, by definition) so that ∇u is oriented as $-\mathbf{n}$, i.e., $\mathbf{n} = -\nabla u/\|\nabla u\|$:

$$\frac{\partial u}{\partial \tau} = (\nabla g \cdot \frac{\nabla u}{\|\nabla u\|} - 2g\kappa - f)\|\nabla u\| \qquad (13)$$

In terms of the level set function u, mean curvature is expressed as [2]:

$$\kappa = \mathrm{div}(\frac{\nabla u}{\|\nabla u\|})$$

By construction, S can be recovered at any instant as the 0-level surface of function u regardless of its topology. According to (10) and (12), S evolves normal to itself at a speed

$$s = \nabla g \cdot \frac{\nabla u}{\|\nabla u\|} - 2g\kappa - f$$

Let the initial position of S be a surface S_0 that subsumes the volume generated by the moving objects and assume that S_0 is in the background.

With a static camera, we will have the following behaviour of S with proper choice of the constant coefficients. While S is in the background we will have $|W_\perp| \approx 0$ and $s \approx -2\lambda\kappa - \alpha$: S will move inward, remains 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, $s = -\gamma\nabla h \cdot \frac{\nabla u}{\|\nabla u\|} - 2(\lambda - \gamma h)\kappa + \beta|W_\perp|$. With proper choice of coefficients, the first two terms intervene to force S to conform to the motion boundary while remaining smooth, and the last term to prevent it from encroaching in the region of motion activity past the motion boundary.

With a moving camera, appropriate choice of the constant coefficients will produce the desired evolution of S. One choice is to set $\alpha = \beta = 0$. Therefore, S evolves to fit a smooth geodesic surface in spatio-temporal space. This generalizes geodesic contour methods [7][5] to three dimensions and to moving cameras. A problem with the choice $\alpha = \beta = 0$ is that the evolution S will be slow since dictated by curvature until it reaches the motion boundaries it is seeking. Another choice is to set bounds on $|W_\perp|$ to determine values of α and β for which $\frac{\alpha}{1+|W_\perp|} - \beta|W_\perp|$ remains negative to produce an inward velocity component, and set coefficient γ high enough to prevent continual movement of the surface inward and past motion boundaries. Because both choices are also applicable to static cameras, they are appropriate to cameras that function in both modes. However, neither choice is optimal for cameras which function only in the static mode.

3. Recovery of tracking information

Albeit implicitly, tracking has often been defined as follows: Given an agregate of motion boundaries (or regions) in the first image of a sequence of images, identify this agregate in every subsequent image.

With our method, motion boundaries at time t are obtained by intersecting the spatio-temporal surface by plane $z = t$. Optical velocity at motion boundary points can be recovered as well from the spatio-temporal surface. This can be done as follows: where parameter t is time, let $C : t \mapsto (x(t), y(t), t)$ be a motion boundary point trajectory in space-time. The tangent vector to C is $\mathbf{V} = (\frac{dx}{dt}, \frac{dy}{dt}, \frac{dt}{dt}) = (u, v, 1)$,

where (u, v) is optical velocity. Because the trajectory of a motion boundary point is on the spatio-temporal surface, assuming no occlusion, the velocity vector along this trajectory is tangent to the surface and, therefore, is orthogonal to the surface normal, yielding the following geometric constraint on optical velocity:

$$\mathbf{N} \cdot \mathbf{V} = 0 \tag{14}$$

where $\mathbf{N}$ is normal to the spatio-temporal surface. If $\mathbf{N} = (n_1, n_2, n_3)$, (14) determines the component of optical velocity in the direction of (n_1, n_2). This constraint, reminiscent of the Horn and Schunck equation which determines the component of optical velocity in the direction of the image gradient, is valid at any point where the spatio-temporal surface is regular.

To estimate optical velocity from the component given by (14), we can proceed with the Hildreth method [8] to compute a smooth estimate by regularization along the detected motion boundaries.

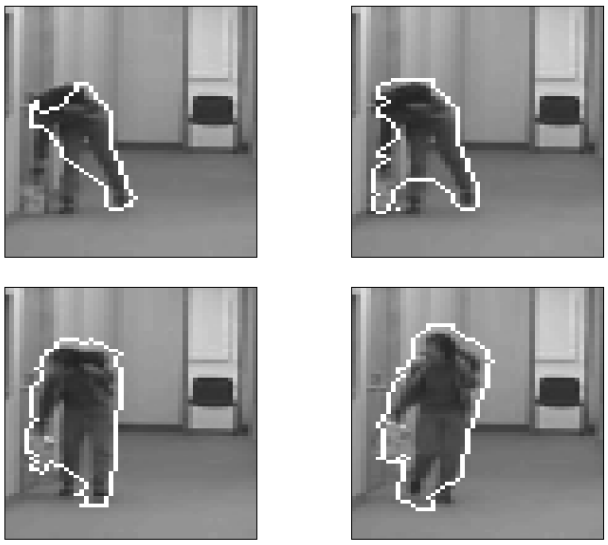

Fig. 2 Sample tracking results: person-bag sequence.

4. Examples

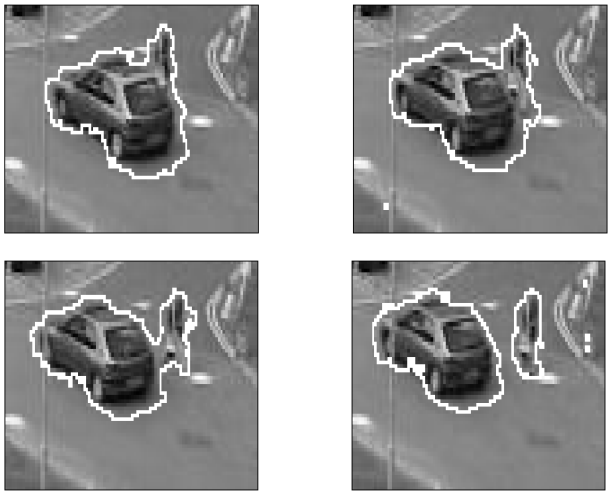

Fig. 1 Sample tracking results: car-pedestrian sequence.

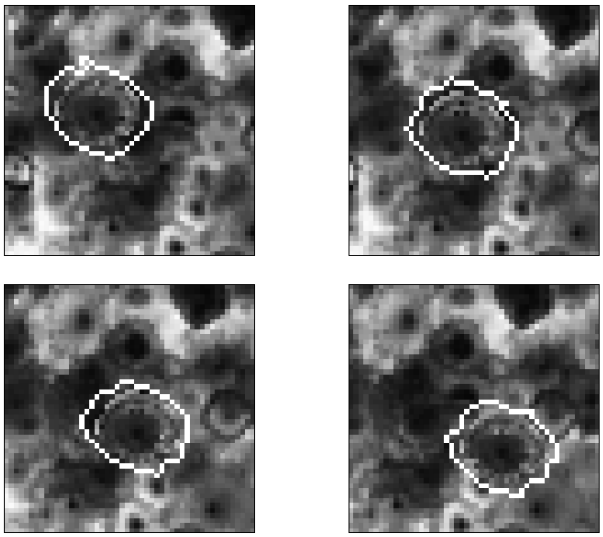

Fig. 3 Sample tracking results: Flower sequence.

Figures 1, 2, and 3 show samples of some of our results. The sequences processed are about 50 images long. The first image and the last image in each figure show the results at the start and end of sequence. The other two images show the results in-between. In the first sequence, Figure 1, a car drives by a walking pedestrian. This example illustrates how a motion boundary splits when the images of two moving objects (the pedestrian and the car in this instance) separate. Because tracking is based solely on a motion

activity measurement, the motion boundaries include the shadows at the rear of the car and front of the pedestrian. In the second sequence, Figure 2, a person stoops to pick up a bag. Here, the motion boundary adjusts to include the image of the bag when this bag is picked. In both sequences the background is fixed. In the third sequence, Figure 3, a flower moves against an also moving flower-patterned background.

References

[1] Osher, S., and J. Sethian, Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on the Hamilton-Jacobi Formulation, *Journal of Computational Physics*, vol. 79, 1988, pp. 1,139-1,145.

[2] Sethian, J., *Level set Methods*, Cambridge University Press (1996).

[3] Bertalmio, M., Sapiro, G., and G. Randall, Morphing Active Contours: A Geometric Approach to Topology-Independent Image Segmentation and Tracking, in: *International Conference on Image Processing*, pp. 318-322, Vol. 3 (1998).

[4] Mansouri, M., and J. Konrad, Minimum Description Length Region Tracking with Level Sets, in: *SPIE Image and Video Communications and Processing*, pp. 515-525, Vol. 3974 (2000).

[5] Paragios, N., 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, 2000, pp. 266-280.

[6] Horn, B. K. P., and Schunck, B. G., Determining Optical Flow, *Artificial Intelligence*, vol. 17, 1981, pp. 185-203.

[7] Caselles, V., Kimmel, R., and G. Sapiro, Geodesic Active Contours, *International Journal of Computer Vision*, vol. 22, No. 1, 1997, pp. 61-79.

[8] Hildreth, E. C., Computation Underlying the Measurement of Visual Motion, *Artificial Intelligence*, vol. 23, No. 3, 1984, pp. 309-355.