TRACKING OF MOVING OBJECTS UNDER SEVERE AND TOTAL OCCLUSIONS

By

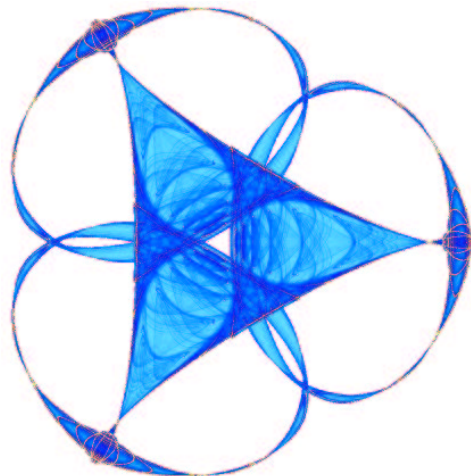**Alberto Bartesaghi**

and

Guillermo Sapiro

INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

TRACKING OF MOVING OBJECTS UNDER SEVERE AND TOTAL OCCLUSIONS

Alberto Bartesaghi and Guillermo Sapiro

University of Minnesota
Electrical and Computer Engineering Department
Minneapolis, MN 55455

ABSTRACT

We present an algorithm for tracking moving objects using intrinsic minimal surfaces which handles particularly well the presence of severe and total occlusions even in the presence of weak object boundaries. We adopt an edge based approach and find the segmentation as a minimal surface in 3D space-time, the metric being dictated by the image gradient. Object boundaries are represented implicitly as the level set of a higher dimensional function, and no particular object model is assumed. We also avoid explicit estimation of a dynamic model since the problem is regarded as one of static energy minimization. A set of interior points provided by the user is used to constrain the optimization, which basically corresponds to selecting the object of interest within the video sequence. The constraints are such that they restrict the resulting surface to be star-shaped in the 3D spatio-temporal space. We present some challenging examples that show the robustness of the technique.

1. INTRODUCTION

Despite numerous efforts by researchers, successfully tracking moving objects in surprisingly simple video sequences still remains a challenging problem. The goal is to track moving objects that possibly change shape and are subject to occlusions under stationary or moving camera conditions. A basic principle that is repeatedly used is to start with the segmentation on a single frame and then forward-track its evolution across subsequent frames. Due to noise, occlusions, poor motion models, etc., the tracking may get lost at some point and re-initialization procedures have to be devised. Another class of approach attempts to use information from both sides, that is, past and future frames are both taken into account so increased reliability and robustness can be achieved.[1] In this latter direction, we present an algorithm for tracking moving objects using intrinsic minimal surfaces which integrates information across all frames in a

---

[1] Although these approaches cannot be applied to real-time applications where future frames are not available, there is a large number of off-line video applications where these techniques can perform significantly better than forward-predictive techniques.

given sequence. We adopt an edge based approach and find the segmentation as a minimal surface in 3D space-time. Such techniques have been extensively used in the literature both for 2D and 3D object segmentation. The idea is to first design an energy functional that is minimized at the object of interest. Thereafter, the problem becomes one of non-convex optimization usually solved by following a gradient descent flow that gives a local minimizer of the energy. Initialization of the flow is a key step, that will allow recovery of the desired object provided the initial guess is close to the correct minima. Availability of robust optimization techniques is then very important, not only because they guarantee finding the correct minima but also because they allow us to concentrate on the design of the segmentation energy which will ultimately determine the performance of the algorithm. One such technique was introduced in [1] for 3D tomograms segmentation, and will be used and extended here to get the spatio-temporal minimal surface. Segmentation is achieved in a semi-automatic fashion: points inside the object are first specified by the user in a small number of keyframes. A basic interpolated trajectory is then obtained that allows us to compute a cylindrical transformed domain that provides a very convenient setting for carrying out the surface minimization. In a final step we convert the segmentation result back to the original cartesian domain.

2. BACKGROUND ON TRACKING AND VIDEO SEGMENTATION

Recently, sophisticated dynamical models have been introduced to forward-track the evolution of contours over time for implicitly represented contours in the presence of occlusions. In [2] significant occlusions have to be handled explicitly requiring the application of detection mechanisms that are then incorporated into the dynamic model of the system. The authors in [3] also assume an explicit model for the shape and motion of contours. In both these approaches the balance between inertial versus image related terms have to be carefully set for the contour to appropriately track the object of interest. Tracking of walking persons under severe occlusions has also been addressed in [4]. They require

a stationary camera and enforcement of motion periodicity constraints. Radial contours presented in [5] are closely related to our approach as it uses a similar 2D image segmentation technique. However, it only works on a frame-by-frame basis and a separate object model (and associated dynamics) need to be estimated in order to track the segmenting curve across frames. Video segmentation with 3D surfaces in space-time has also been addressed, see [6] for example. Here, the authors take a region based approach that cannot deal with severe or total occlusions because region properties are completely missing in the occluded frames. The authors in [7] propose an interactive system for the segmentation of video sequences. They represent contours parametrically and integrate information from past and future keyframes. An interactive environment is provided for the user to progressively refine the location of control points across frames.

3. ENERGY BASED SEGMENTATION

Let Γ be a contour embedded in $\Re^3$ that represents the boundary of interest. Its intrinsic area is

$$\int_\Gamma g(\Gamma)\,d\lambda \qquad (1)$$

where $g : \Re^3 \to (0,\infty]$ is the image derived metric. By minimizing the quantity in Equation (1), Γ is encouraged to go through areas of small cost (corresponding to boundaries) yielding the desired segmentation. The metric has typically the form: $g = f(I) + w$, where f depends on the 3D input image I, and w is a constant that can controls the smoothness of the minimizing contour. For the examples in this paper we use: $f(I) = \frac{1}{|\nabla I_\sigma|+\epsilon}$, where ϵ is a small constant to prevent the denominator from vanishing and I_σ is a smoothed version of I. Note that the image I here is the three dimensional cube in space-time obtained by stacking individual 2D frames across the time dimension, consequently the smoothing and gradient operations are computed intrinsically in 3D. Given an initial contour Γ_0, the solution to the segmentation problem is given by the steady state of the gradient descent flow. Although this technique has proven to be very successful, it strongly depends on the choice of the initial contour Γ_0 which has to be close to the desired minima. Only recently, the availability of global optimization techniques is allowing to address fundamental but challenging problems in image analysis as that of video segmentation under severe occlusions. In particular, we extend the minimization technique presented in [1], see also [8, 9, 10] for related techniques.

4. MINIMAL SURFACES FOR VIDEO

In video segmentation, the 3D surface in space-time generated by the movement of an object has usually a tubular ("cylindrical") shape. That is, if we disregard occlusions for the moment, integration of object boundaries across frames will render a surface with such shape. In this setting, let us analytically express the cylinder's axis as a function of time $P(t) : \Re \to \Re^2$ which will give the position at frame t of a point inside the cylinder. For each frame we then consider a polar coordinate system centered in $P(t)$ and assume that the lateral surface of the cylinder can be expressed as a single valued function $\rho(\theta)$, see Figure 1 (a). Extending to all frames, the complete surface can then be expressed in the cylindrical system (ρ, θ, t) as a function $\rho(\theta, t)$, see Figure 1 (b). Although this restricts the segmentation technique to the class of star shaped surfaces, it will provide a nice and general enough setting to solve the optimization problem as described below. Assuming that $P(t)$ is known, the segmentation problem is equivalent to the search of a function $\rho(\theta, t)$ representing the surface minimizing the energy in Equation (1). As similarly discussed in [1, 11], such a geometric construction introduces a scaling factor $\frac{1}{\rho}$ on the metric that avoids the global minima of zero energy. That is, for uniform g, concentric cylinders $\rho(\theta, t) = $ constant will all be minimizers of the energy.

Assuming the position of points inside the object changes slowly from frame to frame, the axis $P(t)$ can be specified as a sequence of discrete points at a few selected keyframes, see Figure 2. From this set of points, we can linearly interpolate in-between to get the coordinates of intermediate points. The number of points required is dependent on the complexity of the movement, the simplest ones will require a minimum of two points (*i.e.* in the first and last frames) but more complex motion paths may require additional intermediate points in order to guarantee that interpolated positions will fall inside the object in all intermediate frames.[2] Note that we do not require complete determination of the object's motion, but only an approximation so the correct object geometry is recovered.

4.1. Finding the minimal surface

The optimization is done with the technique presented in [1]. The basic idea is to solve the minimal surface problem by sectioning the 3D domain with 2D planes and find geodesics restricted to the cutting sections. If we assume

---

[2]Note that more sophisticated interpolation schemes can also be used, for example, polynomial or splines interpolants. Even dynamic models for the motion of single points can be incorporated so fewer points need to be specified. Tracking of a single axis point can easily be done with traditional tracking techniques or as the computation of minimal paths using the adequate image metric. That is, we can automatically compute the axis as a geodesic curve (on a metric space dictated by image features) between a pair of selected end points in the keyframes.
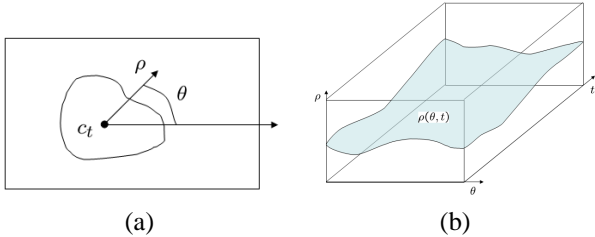
Fig. 1. (a) Polar coordinate transformation at each individual frame t. (b) Cylindrical transformed domain in ρ, θ, and t coordinates. To guarantee the tubular shape in the Cartesian domain, the transformed surface has to be periodic in the θ direction.
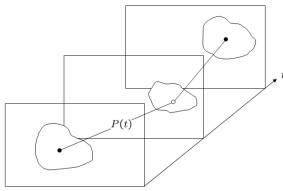

Fig. 2. Approximating a simple motion trajectory by linear interpolation. Selected points must be inside the object of interest at each frame and interpolated positions must also fall inside the object in all intermediate frames.

that each such 2D geodesic corresponds to the intersection of the 3D minimal surface with the slicing plane, we can recover the surface as the collection of curves. Planar geodesics between two points are computed by a non-iterative procedure [12]. This procedure can easily be modified to handle periodic geodesics, needed for the computation of geodesics in the θ direction in order to enforce the cylindric shape. The volume is traversed in both directions (θ and t, see Figure 1 (b)) in increasing order according to the intrinsic length of corresponding geodesics. By first processing slices with shorter intrinsic lengths we are relying on areas where the metric is strongly anisotropic. As we continue to process sections of increasing cost we start gradually enforcing the restriction that geodesics in both directions should be in agreement with each other as they are part of a single surface. For further details on the computation of the minimal surface we refer the reader to [1]. Once the segmentation in the transformed domain is obtained, we bring the result back to the Cartesian grid.

5. EXAMPLES ON REAL SEQUENCES

We present segmentation results for two different movies of real scenes with car objects. Figures 3 and 4 show results for a 280-frame movie segmented from only three points. Observe how shape changes are handled correctly and how the segmentation is not biased by the presence of occlu-

sions. The passing truck takes 30 frames which means that useful boundary information is completely missing in those frames. Figure 5 shows the recovered surface in 3D. Figure
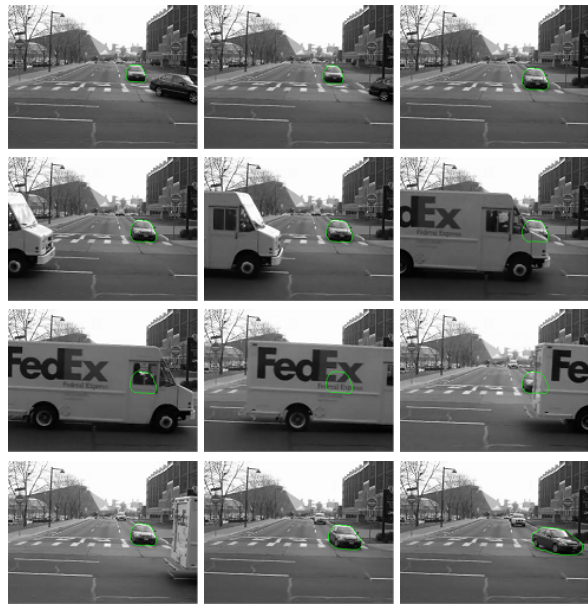

Fig. 3. Sequence of a car turning at a stop sign. From left to right, top to bottom: frames 1, 17, 117, 145, 150, 158, 162, 167, 177, 181, 250 and 280.

6 shows the segmentation of a passing car occluded by trees. This sequence has a very complex background including other moving objects (the truck on the back), very weak edges and some small camera movements. Not only the shape of the car changes across frames (because of the perspective), but also its area grows dramatically as it approaches the camera (from 347 to 2428 pixels), illustrating again the capacity of the algorithm in adapting to arbitrary shapes. The results were obtained from four points selected in the first, last, frames 20 and 50. Additional points were needed in this case because the speed of the car changes significantly and it cannot be linearly predicted from the first and last frames.

The running time of the algorithm is about a minute per hundred frames in a 1.2 Mhz laptop computer without careful optimization. Movie sequences are available for viewing at: http://mountains.ece.umn.edu/~abarte/sequences.

6. SUMMARY AND DISCUSSION

We presented an energy based video segmentation algorithm that finds objects boundaries as minimal surfaces in 3D space-time domain. Geometric constraints are enforced by specifying points inside the object to be segmented. The technique is both robust and accurate and can track objects that change shape and are subject to severe and total occlusions.

Fig. 4. Zoomed in view for the turning car sequence. The three selected points are shown in red. From left to right, top to bottom: frames 1, 17, 117, 145, 150, 158, 162, 167, 177, 181, 250 and 280.

7. REFERENCES

[1] A. Bartesaghi, G. Sapiro, and S. Subramaniam, "An energy-based three dimensional segmentation approach for the quantitative interpretation of electron tomograms," IMA Preprint Series 2009, University of Minnesota, Minneapolis, MN, USA, December 2004.

[2] M. Niethammer and A. Tannenbaum, "Dynamic geodesic snakes for visual tracking," in *CVPR*, Wash., D.C., USA, June 27 - July 02 2004, pp. 660–667.

[3] J. Jackson, A. Yezzi, and S. Soatto, "Tracking deformable moving objects under severe occlusions," in *IEEE Conference on Decision and Control*, Atlantis, Paradise Island, Bahamas, December 14-17 2004.

[4] Jiaya Jia, Tai-Pang Wu, Yu-Wing Tai, and Chi-Keung Tang, "Video repairing: Inference of foreground and background under severe occlusion," in *CVPR*, June-July 2004, vol. I, pp. 364–371.

[5] Y. Chen, T. Huang, and Y. Rui, "Optimal radial contour tracking by dynamic programming," in *ICIP*, Thessaloniki, Greece, October 2001.

[6] J. Konrad and M. Ristivojevic, "Joint space-time image sequence segmentation based on volume competition and level sets," in *ICIP*, Rochester, NY, USA, September 22-25 2002.

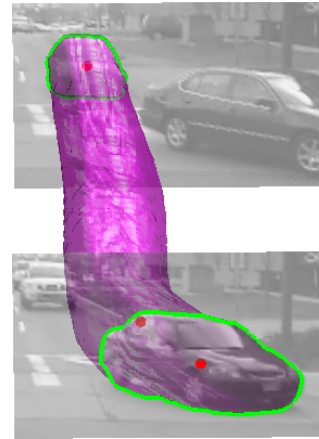[7] A. Agarwala, A. Hertzmann, D. H. Salesin, and S. M. Seitz, "Keyframe-based tracking for rotoscoping and animation,"

Fig. 5. Minimal surface in 3D space-time domain for the turning car sequence.

in *ACM SIGGRAPH'04*. 2004, vol. 23:3, pp. 584–591, ACM Press.

[8] Y. Boykov and V. Kolmogorov, "Computing geodesics and minimal surfaces via graph cuts," in *ICCV*, Nice, France, October 2003.

[9] B. Appleton and H. Talbot, "Globally optimal surfaces by continuous maximal flows," in *Digital Image Computing: Techniques and Applications, Proc. VIIth APRS*, November 2003, vol. 1, pp. 623–632.

[10] R. Ardon and L. D. Cohen, "Fast constrained surface extraction by minimal paths," in *2nd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, O. Faugeras and N. Paragios, Eds., Acropolis, Nice, France, October 2003, pp. 233–240.

[11] B. Appleton and H. Talbot, "Globally optimal geodesic active contours," *Journal of Mathematical Imaging and Vision, submitted*, 2002.

[12] L. D. Cohen and R. Kimmel, "Global minimum for active contour models: A minimal path approach," *IJCV*, vol. 24, no. 1, pp. 57–78, August 1997.
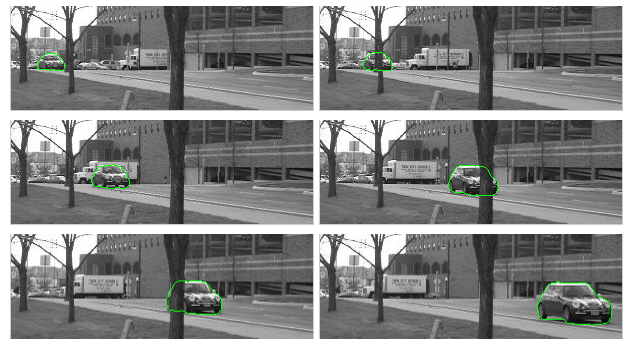

Fig. 6. Sequence of car partially occluded by trees (80 frames). We show frames 1, 13, 36, 55, 63 and 78.