Active Polygons for Object Tracking

Gozde Unal, Hamid Krim
North Carolina State University
ECE Dept
Raleigh, NC 27695

Anthony Yezzi
Georgia Institute of Technology
School of Electrical Engineering
Atlanta, GA 30332

Abstract

In this paper, we present a technique for object tracking in image sequences, which includes a novel velocity field estimation for vertex locations of a polygon. In this approach, after capturing boundaries of an object by a polygon from the first few frames of the image sequence, both the spatial segmentation and motion segmentation of a polygonal object can be achieved quickly by involving only the vertex locations and the adjacent edges in the computations. We carry out velocity field estimation at an active polygon vertex using the optical flow constraint on its two adjacent edges. A spatial segmentation phase follows to further refine object's vertex locations estimated by the optical flow. The advantage of our region-based active polygons over continuous active contours in object tracking in video applications is emphasized by provision of a compact representation of object features, particularly for simply connected target shapes, hence will be essential for their tracking.

1. Introduction

In a time-varying image sequence, $I(x, y, t) : [0, a] \times [0, b] \times [0, T] \rightarrow \mathbb{R}^+$, image motion may be described by a 2-D vector field $\boldsymbol{V}(x, y, t)$, which specifies the direction and speed of the moving target at each point (x, y). The measurement of visual motion is equivalent to computing $\boldsymbol{V}(x, y, t)$ from $I(x, y, t)$ [5]. Estimating the velocity field remains an important research topic in light of its ubiquitous presence in many applications and as reflected by wealth of previously proposed techniques. Approaches include differential techniques [6,8], energy-based techniques [4], phase-based techniques (see [1, 10] for more references). Differential techniques solve an optical flow equation which states that intensity or brightness of an image stays constant with time. Additional constraints on the velocity field are usually required to address the ill-posed nature of the problem. A translational motion of an object boundary monitored through a small aperture only allows for the detection and hence computation of the velocity component normal to the edge, with the tangential component remaining undetectable. Computing the velocity field hence involves regularizing constraints such as its smoothness and other variants.

Object tracking remains of great research interest in computer vision. Methods which exploit boundary-based information and/or region-based information have been proposed [9]. Continuous contours, and snake models [2], for example, are used for estimating the motion along the boundaries of an object that is to be tracked. If the information being utilized in the estimation is very local, however, the result is prone to errors. To overcome this problem, geodesic active regions, a framework which incorporates both boundary-based and region-based approaches have been proposed in [9]. This technique involves many different steps, in addition to a consistency check with respect to an affine motion following a motion detection. Our goal in this paper is to build on the previously achieved hindsight to develop a simple and efficient tracking algorithm nice adapted to polygonal objects. This is in effect an extension of recently developed evolution models which use region-based data distributions to capture polygonal object boundaries [11]. A fast numerical approximation of an introduced information measure which in its optimization yields a set of coupled ODEs determine the flow of polygon vertices to enclose an object.

The key idea of our proposed extension to video sequences and its resulting efficiency is centered around tracking of relatively few vertices and effectively the corresponding adjacent edges. A novel velocity estimation is proposed in tandem with a fast spatial and motion robust segmentation which follow the initial object delineation in the first few frames. We carry out a velocity field estimation at an active polygon vertex using the optical flow constraint on its two adjacent edges. A spatial segmentation phase follows to further refine object's vertex locations estimated by the optical flow. As argued below, this approach presents several advantages over active contours in video tracking; not the least of which is its feature parsimony which in turn greatly simplifies any tracking application.

The balance of the paper is as follows. In the next section, we briefly review our spatial segmentation step which involves our region-based active polygons. In Section 3, we explain the incorporation of velocity field estimation in case of an image sequence, and we provide simulation results in Section 4.

2. Active Polygons with an Information-Theoretic Criterion

In this section, we present gradient flows designed to move polygon vertices with the goal of parsing an image domain into meaningfully different regions. A general form of an energy functional over a region R which is to capture various objects may be written as a contour integral [11]

$$E(\boldsymbol{C}) = \iint_R f(x,y)\,dx\,dy = \oint_{\boldsymbol{C}=\partial R} \langle \boldsymbol{F}, \boldsymbol{N} \rangle\, ds, \quad (1)$$

where $\boldsymbol{N}$ denotes the outward unit normal to $\boldsymbol{C}$ (the boundary of R), ds the Euclidean arclength element, and where $\boldsymbol{F} = (F^1, F^2)$ is chosen so that $\nabla \cdot \boldsymbol{F} = f$. It can be shown [12] that a gradient flow for $\boldsymbol{C}$ with respect to E may be written as

$$\frac{\partial \boldsymbol{C}}{\partial t} = f\boldsymbol{N}.$$

The key to our approach is to consider a closed polygon $\boldsymbol{P}$ as the contour $\boldsymbol{C}$, with a fixed number of vertices, say $n \in \mathbb{N}$, $\{\boldsymbol{P}_1, .., \boldsymbol{P}_n\} = \{(x_i, y_i), i = 1, \ldots n\}$. We may parameterize $\boldsymbol{C}$ by $p \in [0, n]$ as

$$\boldsymbol{C}(p, \boldsymbol{P}) = L(p - \lfloor p \rfloor, \boldsymbol{P}_{\lfloor p \rfloor}, \boldsymbol{P}_{\lfloor p \rfloor + 1})$$

where $\lfloor p \rfloor$ denotes the largest integer which is not greater than p, and where $L(t, \boldsymbol{A}, \boldsymbol{B}) = (1-t)\boldsymbol{A} + t\boldsymbol{B}$ parameterizes between 0 to 1 the line from $\boldsymbol{A}$ to $\boldsymbol{B}$ with constant speed. Note that the indices of $\boldsymbol{P}$ should be interpreted as modulo n so that $\boldsymbol{P}_0$ and $\boldsymbol{P}_n$ denote the same vertex (recall $\boldsymbol{C}$ is a closed curve). Finally, note that $\boldsymbol{C}_p$ is defined almost everywhere (where $p \neq \lfloor p \rfloor$) by $\boldsymbol{C}_p(p, \boldsymbol{P}) = \boldsymbol{P}_{\lfloor p \rfloor + 1} - \boldsymbol{P}_{\lfloor p \rfloor}$. We may hence obtain the first variation of the energy functional E, and its minimization yields a gradient descent flow given by a set of ordinary differential equations (ODEs) for each vertex $\boldsymbol{P}_k$ as

$$\begin{aligned}\frac{\partial \boldsymbol{P}_k}{\partial t} &= \boldsymbol{u}_{1,k}^{\perp} \int_0^1 p f\big(L(p, \boldsymbol{P}_{k-1}, \boldsymbol{P}_k)\big)\, dp \\ &+ \boldsymbol{u}_{2,k}^{\perp} \int_0^1 (1-p) f\big(L(p, \boldsymbol{P}_k, \boldsymbol{P}_{k+1})\big)\, dp, \quad (2)\end{aligned}$$

where $\boldsymbol{u}_{1,k}^{\perp}$ (resp. $\boldsymbol{u}_{2,k}^{\perp}$) denotes the outward unit normal of edge $(\boldsymbol{P}_{k-1} - \boldsymbol{P}_k)$ (resp. $(\boldsymbol{P}_k - \boldsymbol{P}_{k+1})$), and which in the interest of space, we defer the proof of to [11]. These ODEs are solved simultaneously for each vertex of a polygon. Note how the information through the functional f is being integrated along adjacent edges for a vertex. In addition to a small number of vertices, their well-separated locations clearly distinguish our proposed approach from the snake-based methods.

Upon obtaining a polygonal evolution model for a generic energy functional, we next choose a "good" criterion that captures the underlying statistical properties of each region delineated by a contour (particularly a polygon) and defines a metric among them. This may be achieved by an information divergence measure that defines a distance among probability densities. On account of its properties [3], we pick as our energy functional the Jensen-Shannon criterion which, in its general form may be written for N probability densities (with prior probabilities, $\boldsymbol{a} = (a_1, \ldots, a_N)$ such that $\sum_{i=1}^{N} a_i = 1$) as

$$JS_a = H\left(\sum_{i=1}^{N} a_i p_i(\xi)\right) - \sum_{i=1}^{N} a_i H(p_i(\xi)) \quad (3)$$

where $p_i(\xi)$ denotes the probability density of pixel values in the i^{th} region, and H is the Shannon entropy. Estimation of a probability density and its entropy may be carried out in a variety of ways, we hence adopt a first order approximation of a density which achieves a maximum entropy solution, which is in turn used in approximating the entropy expression as proposed in [7]. Equipped with this numerical approximation of entropy H, and with a single polygon laid on the image domain, we obtain an approximate Jensen-Shannon criterion

$$\widehat{JS}_{a,2} = \frac{1}{2}\, a_1\, a_2 \sum_{j=1}^{m} (u_j - v_j)^2. \quad (4)$$

Measurements u_j (resp. v_j), $j = 1, .., m$, for region R_u inside the polygon (resp. R_v outside the polygon) are given by

$$u_j = \frac{\int_{R_u} G_j(I(\boldsymbol{x}))d\boldsymbol{x}}{|R_u|}, \quad v_j = \frac{\int_{\Omega \setminus R_u} G_j(I(\boldsymbol{x}))d\boldsymbol{x}}{|R_v|}, \quad (5)$$

with $|R_u| = \int_{R_u} d\boldsymbol{x}$, ($|R_v| = \int_{\Omega \setminus R_u} d\boldsymbol{x}$), $\boldsymbol{x} = (x, y)$. In order to capture statistical characteristics of regions with non-Gaussian densities, higher than 2$^{\text{nd}}$ order moments are used by choosing $G_1(\xi) = \xi e^{-\xi^2/2}$ as an odd function to measure asymmetry, and $G_2(\xi) = |\xi|$, or $e^{-\xi^2/2}$ as choices of even functions [7].

One specific choice of priors in Eq.(4), which are proportional to the ratio of area of each region to the total area of the image domain leads the image-based integrand f in Eq.(1) to take the form

$$f = \sum_{j=1}^{m} (u_j - v_j)\left((G_j(I) - u_j) + (G_j(I) - v_j)\right). \quad (6)$$

3. Velocity Estimation At Vertex Points

Our polygonal flows essentially integrate spatial image information along two adjacent edges of a vertex $\boldsymbol{P}_k$ to determine its speed and direction. This is exactly the same idea we will use in the estimation of a velocity field at a vertex of the active polygon laid on a time-varying image sequence. Our goal is to estimate velocity vector at each vertex $\boldsymbol{P}_k$ using the two adjacent edges as shown in Fig. 1. The velocity field $\boldsymbol{V}(x,y)$ at each point on an edge may

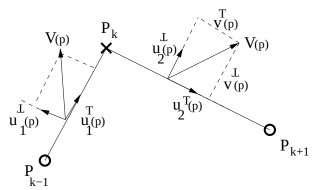

Figure 1. 2-D velocity field along two neighbor edges of a polygon vertex

be represented as $\boldsymbol{V}(p) = v^{\perp}(p)\boldsymbol{u}^{\perp}(p) + v^{T}(p)\boldsymbol{u}^{T}(p)$, where $\boldsymbol{u}^{T}(p)$ and $\boldsymbol{u}^{\perp}(p)$ are unit vectors in the directions tangent and normal to an edge. The optical flow constraint given by

$$\frac{\partial I}{\partial t} + \nabla I \cdot \boldsymbol{V} = 0, \quad (7)$$

provides a way to estimate the component $v^{\perp}$ of the velocity field directly from the time-varying image $I(x,y,t)$. Once an active polygon locks onto a target polygonal object, the unit direction vectors $\boldsymbol{u}^{\perp}$ and $\boldsymbol{u}^{T}$ are also known immediately. Intuitively, the set of measurements obtained from the image brightness constraint

$$v^{\perp}(x,y) = -I_t(x,y)/|\nabla I(x,y)| \quad (8)$$

from two edges of a vertex with a combined contribution provide sufficient information to infer the resultant motion at the vertex. Expecting a sensitivity of such instantaneous normal velocity measurements to noise, we integrate them in a weighted manner along two neighboring edges of a vertex to yield a more robust estimation. This leads us to propose a velocity estimation scheme at each vertex of an active polygon as

$$\begin{aligned}\boldsymbol{V}_k &= \boldsymbol{u}^{\perp}_{1,k}\frac{\int_0^1 p\, v^{\perp}\big(L(p,\boldsymbol{P}_{k-1},\boldsymbol{P}_k)\big)dp}{\int_0^1 p\, dp} \\ &+ \boldsymbol{u}^{\perp}_{2,k}\frac{\int_0^1 (1-p)\, v^{\perp}\big(L(p,\boldsymbol{P}_k,\boldsymbol{P}_{k+1})\big)dp}{\int_0^1 (1-p)\, dp} \quad (9)\end{aligned}$$

for $k = 1, \ldots n$. To introduce further robustness and achieve more reliable estimates, we make use of smoother spatial derivatives (larger neighborhoods) in the coarse of computing $v^{\perp}$.

To proceed with tracking of a polygonal object, we first involve the previously described segmentation technique to initially delineate the moving object boundaries. Upon initialization, and using Eq. (9), the velocity vector is estimated from two images $I(x,y,t)$ and $I(x,y,t+1)$, and is subsequently utilized to move the active polygon's handful of vertices to new locations on the next image $I(x,y,t+1)$ in the sequence. The ODE in Eq. (2) is then run for a short time for further refinement of the polygon delineation of the moving object. We substantiate our approach by the examples given in the next section.

4. Experimental Results

We show in Fig.2 snapshots from an image sequence where a rectangular object rotates in clockwise direction. The estimated velocity field vectors at each of the 4 vertices are also shown next to each snapshot image. It can be observed that directions of the velocity field are fairly well tracked.

Two IR image sequences are given in Fig.3 and Fig. 4. The targets appear as bright spots, usually in the form of polygonal shapes, and their tracking is successfully achieved by our technique. Similarly, tracking by our active polygons is demonstrated in Fig.5 where a model rocket is about to be launched.

References

[1] S. S. Beauchemin and J. L. Barron. The computation of optical flow. *ACM Computing Surveys*, 27(3):433–467, 1995.

[2] V. Caselles and B. Coll. Snakes in movement. *SIAM Journal on Numerical Analysis*, 33(12):2445–2456, 1996.

[3] A. B. Hamza, Y. He, and H. Krim. An information divergence measure for ISAR image registration. In *IEEE Workshop on Statistical Signal Processing*, pages 130–133, 2001.

[4] D. J. Heeger. Optical flow using spatiotemporal filters. *IJCV*, 1:279–302, 1988.

[5] E. C. Hildreth. Computations underlying the measurement of visual motion. *AI*, 23:309–354, 1984.

[6] B. K. P. Horn and B. G. Schunck. Determining optical flow. *AI*, 17:185–203, 1981.

[7] A. Hyvarinen. New approximations of differential entropy for independent component analysis and projection pursuit. Technical report, Helsinki Univ., 1997.

[8] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *IJCAI*, pages 674–679, 1981.

[9] N. Paragios and R. Deriche. Geodesic active regions for motion estimation and tracking. Tech. Report INRIA 1999.

[10] A. M. Tekalp. *Digital Video Processing*. Prentice Hall, 1995.

[11] G. B. Unal. *Curve and Polygon Evolution Techniques for Image Processing*. PhD thesis, NCSU, 2002.

[12] S. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE PAMI*, 18(9):884–900, 1996.

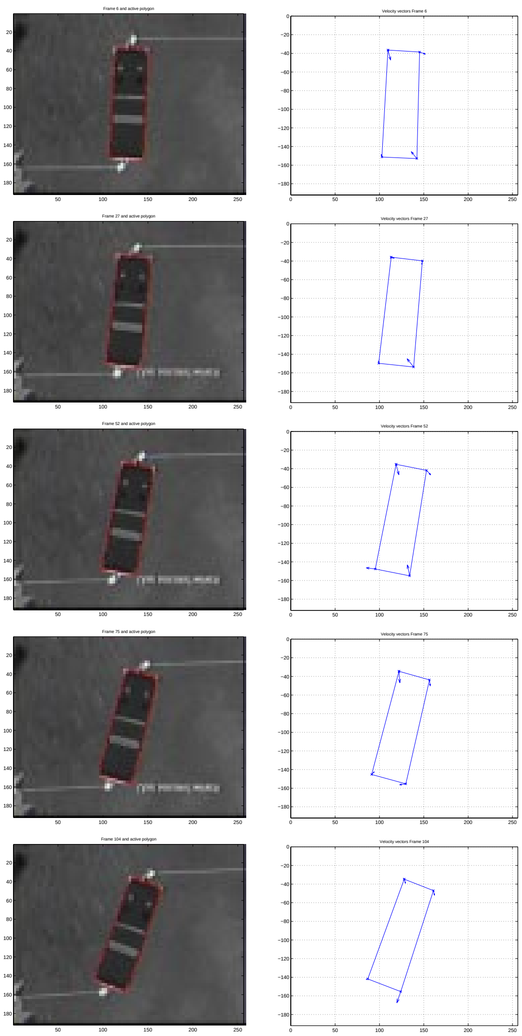

Figure 2. A rotating object (left), and its velocity field at 4 vertices (right) are shown.

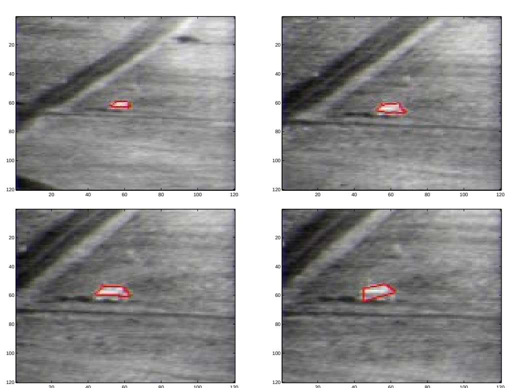

Figure 3. Tracking of an IR image sequence (left-right, top-bottom).

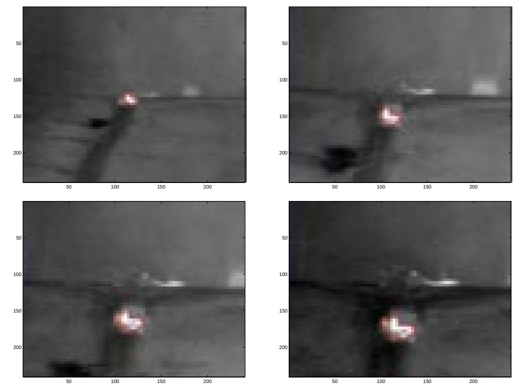

Figure 4. Tracking of another IR image sequence (left-right, top-bottom).

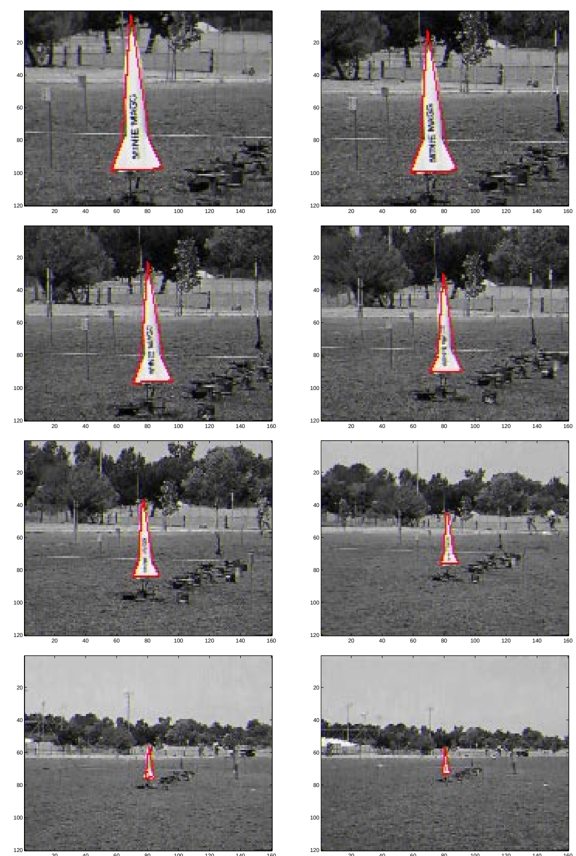

Figure 5. Tracking of a model rocket object (left-right, top-bottom).