Object Tracking with Ratio Cycles Using Shape and Appearance Cues*

M.E. Sargin, P. Ghosh, B.S. Manjunath and K. Rose
Center for Bio-Image Informatics
Department of Electrical and Computer Engineering, UC Santa Barbara
{msargin,pratim,manj,rose}@ece.ucsb.edu

Abstract

We present a method for object tracking over time sequence imagery. The image plane is represented with a 4-connected planar graph where vertices are associated with pixels. On each image, the outer contour of the object is localized by finding the optimal cycle in the graph such that a cost function based on temporal, appearance and shape priors is minimized. Our contribution is the particle filtering-based framework to integrate the shape cue with the temporal and appearance cues. We demonstrate that incorporating the shape prior yields promising performance improvement over temporal and appearance priors on various object tracking scenarios.

1. Introduction

Object tracking in image sequences is an active area of research in computer vision [10]. A majority of current tracking methods can be categorized into three groups: active contours methods, kernel methods and graph theory methods. In this work, we are particularly interested in the graph-based approaches since, under certain assumptions, they are able to provide globally optimal solutions in polynomial time.

In graph-based tracking methods, the image plane is represented with a graph where the vertices are associated with the pixels. Accordingly, the edges between each vertex and its immediate 4 (or 8) neighbors are designed to capture spatial dependencies between pixels. In this representation, tracking problem can be re-stated as finding the optimal labeling configuration where each vertex is labeled with either 0 (background) or 1 (object). Graph-cuts [2] provide an efficient solution to this label assignment problem with the energy function adopted from pairwise Markov random fields (MRF). In typical object tracking applications, the object can be represented with a single connected blob. However, graph-cuts do not make use of this important domain specific knowledge. In fact, including this constraint in the graph-cuts is NP-hard [9].

Alternatively, one can represent the object contour with a cycle in the graph. In this representation, non-self-intersecting cycles always yield a single connected blob. Furthermore, polynomial time algorithms exist to find optimal cycles in the graph [3]. In these algorithms, each edge has two weights and the target energy function is defined based on ratio of these weights on the cycle. The appearance cues are transformed to edge weights with the use of Green's theorem [4, 5]. The shape cues are also integrated in this framework using auxiliary vertices [8]. However, to the best of our knowledge there is no prior work which integrates both shape and the region-based appearance cues within the ratio cycle framework.

In this paper, we present a ratio cycle-based tracking framework which integrates both shape and the region-based appearance. The appearance is represented with color and edge cues. Shape cue is introduced based on a template shape which captures rigid transformation including translation, scale, rotation and shear. The temporal information is probabilistically integrated within the proposed framework through state-space filtering.

The organization of this paper is as follows. In Section 2 we provide technical explanation of the proposed method. The experimental results are given in Section 3. Finally, we conclude and discuss future directions in Section 4.

2. Method

Consider an image $I(x, y)$ with dimensions $X \times Y$ where X is the width and the Y is the height. We represent this image with a planar graph $G = \{V, E\}$, where $V = \{v_{x,y}\}$ is the set of vertices with $(X+1)\times(Y+1)$ elements. The vertices are located at the corners of

*This work is supported by grants NSF-ITR 0331697 and NSF-III 0808772.

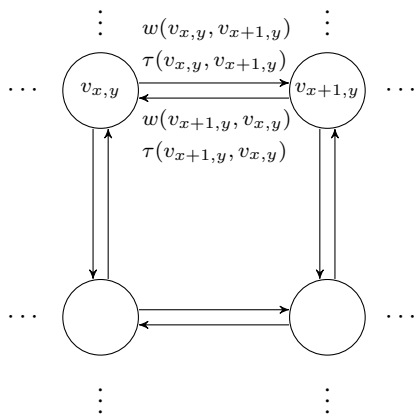

Figure 1. Representation of the image with the graph G.

the pixels such that the node at the upper left corner of the pixel $I(x, y)$ is $v_{x,y}$. The set of edges $E = \{e(v_{x,y}, v_{x',y'})\}$ are defined between 4 neighbors of each vertex.

In this graph, we also define a cycle C as a closed directed path, with no repeated vertices. The set of all cycles in G is denoted with $\mathcal{C}$. Given two weights, $w(e)$ and $\tau(e) \geq 0$, associated with each edge e, the task is to find the cycle C^* that maximizes the following:

$$C^* = \underset{C \in \mathcal{C}}{\operatorname{argmax}} \frac{\sum_{e \in C} w(e)}{\sum_{e \in C} \tau(e)}. \tag{1}$$

This maximization problem can be solved using Howard's iteration policy algorithm [3] with empirical time complexity $O(|E|)$.

2.1 Cost Function

Let $P_{I,S}(x, y)$ be the posterior probability of labeling the pixel at (x, y) with *object* label O:

$$P_{I,S}(x, y) = P(L(x, y) = O | I(x, y), S) \tag{2}$$

Here, $I(x, y)$ is the image intensity/color value at (x, y) and S is the shape prior. We assume that $P_{I,S}(x, y)$ can be approximated with the weighted sum rule [6]:

$$\begin{aligned} P_{I,S}(x, y) \approx \alpha \overbrace{P(L(x, y) = O | I(x, y))}^{P_I(x,y)} \\ + (1 - \alpha) \underbrace{P(L(x, y) = O | S)}_{P_S(x,y)}. \end{aligned} \tag{3}$$

The tracking is performed by finding the 8-connected region that maximizes a function of these probabilities for the pixels within the region. This function is designed in the form of (1) where $w(v_{x,y}, v_{x',y'})$ and $\tau(v_{x,y}, v_{x',y'})$ weights are set based on (3). By doing this, the idea is to relate the optimum blob in the image and the optimum cycle in G.

The numerator is designed such that it reflects the sum of $P_{I,S}(x, y)$ within the area surrounded by that cycle. Accordingly, the cycles favor to capture the regions with high $P_{I,S}(x, y)$. The area integration within the cycle is converted to line integration on the cycle using Green's theorem. In this way, integration of the $P_{I,S}(x, y)$ values over a region is performed efficiently by summing the edge weights around that region. Accordingly, one can derive the values of $w(v_{x,y}, v_{x',y'})$ as:

$$w(v_{x,y}, v_{x',y'}) = \begin{cases} 0 & y = y' \\ \sum_{\delta=1}^{x} P_{I,S}(\delta, y') & y = y' + 1 \\ -\sum_{\delta=1}^{x} P_{I,S}(\delta, y) & y = y' - 1. \end{cases} \tag{4}$$

The denominator weights $\tau(v_{x,y}, v_{x',y'})$ are set such that the cycles favor discontinues in $P_{I,S}$:

$$\tau(v_{x,y}, v_{x',y'}) = \begin{cases} \frac{1}{\alpha_\tau + (P_{I,S}(x_m, y-1) - P_{I,S}(x_m, y))^2} & y = y' \\ \frac{1}{\alpha_\tau + (P_{I,S}(x-1, y_m) - P_{I,S}(x, y_m))^2} & x = x' \end{cases} \tag{5}$$

where, x_m and y_m are $\min(x, x')$ and $\min(y, y')$, respectively. In this equation, α_τ adjusts the *edge* sensitivity. Larger α_τ values make the system less sensitive to the edges. In our experiments we empirically set it to 0.01.

2.2 Shape and Temporal Priors

The shape prior is represented with a single binary image, $S(x, y)$, which contains the silhouette of the object. The temporal prior is represented with the position (x, y), scale (s_x, s_y) and rotation θ of the object. The temporal prior, $\mathbf{t} = [x, y, s_x, s_y, \theta]^T$, is propagated in time using particle filtering [1] technique. In the particle filtering, the conditional density of $\mathbf{t}$ is modeled with a set of particles $\mathbf{T}_p = \{\mathbf{t}_{p,i}\}$ in $\mathbf{t}_p \in \mathbb{R}^5$ and propagated over time with re-sampling. The shape and the temporal priors are coupled to obtain $P_S(x, y)$:

$$\begin{aligned} P_S(x, y) =& P(L(x, y) = O | S, \mathbf{t}) \\ =& S(x', y') \end{aligned} \tag{6}$$

where, (x', y') is obtained by warping (x, y) from the I back to S using the transformation $\mathbf{t}$. Accordingly, for a

Figure 2. Sample tracking output. The maximum ratio-cycle is shown in green. The shape priors associated with 50 particles are shown in blue.

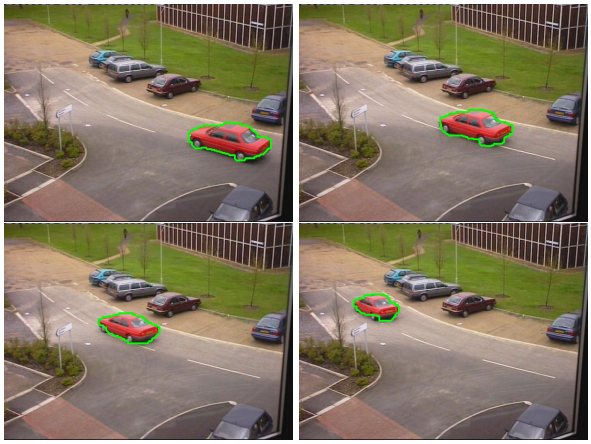

Figure 3. Tracking output on *car* sequence.

given S, each $\mathbf{t}$ deforms S to the image domain and results in a hypothetical $P_S(x, y)$. Given the set of all hypotheses, the task is to find the one that maximizes the cycle ratio (1). However, it is computationally infeasible to try all possible $\mathbf{t} \in \mathbb{R}^5$. Alternatively, one can obtain an approximate solution by properly sampling the set of all hypotheses based on the temporal information. For this purpose, the particle filtering provides an effective way to generate these random samples $\mathbf{t}_p$ from the estimated distribution of $\mathbf{t}$. In this way, only the relevant subspace of $\mathbb{R}^5$ is searched with remarkably less number of hypotheses. In our experiments, we observe that $50 - 500$ samples are sufficient for object tracking.

Accordingly, for a given S, the optimal cycle is obtained by maximizing (1) over the samples $\{\mathbf{t}_p\}$:

$$C^* = \max_{\mathbf{t}_p \in \mathbf{T}_p} \operatorname*{argmax}_{C \in \mathcal{C}} \frac{\sum_{e \in C} w(e|S, \mathbf{t}_p)}{\sum_{e \in C} \tau(e|S, \mathbf{t}_p)}. \qquad (7)$$

Note that, each S and $\mathbf{t}_p$ results in a new set of edge weights. Therefore, we represent this dependency with $w(e|S, \mathbf{t}_p)$ and $\tau(e|S, \mathbf{t}_p)$.

3. Experimental Results

In this section, we demonstrate the performance of the proposed framework on three different object tracking scenarios. The tracking performance is measured based on the F-measure. Let $\mathcal{R}_o$ and $\mathcal{R}_g$ be the output and the ground truth regions, respectively. The precision, $\mathcal{P}$ and recall $\mathcal{R}$ is measured as:

$$\mathcal{P} = \frac{|\mathcal{R}_o \cap \mathcal{R}_g|}{|\mathcal{R}_o|}, \; \mathcal{R} = \frac{|\mathcal{R}_o \cap \mathcal{R}_g|}{|\mathcal{R}_g|}. \qquad (8)$$

Finally, the F-measure is obtained by $2\mathcal{P}\mathcal{R}/(\mathcal{P} + \mathcal{R})$.

In all scenarios, the foreground and background models are obtained using manual segmentation on few frames. In the first scenario, we consider car tracking [1]. In this image sequence, the viewpoint changes with respect to the car which makes it hard to model with rigid deformations. However, in our framework, the optimal cycle can still deviate from the rigid transformation to match the image observations. Therefore, slight non-rigid deformations are still captured properly as illustrated in Fig. 3. In the second scenario, the object of interest is the mushroom toy (See Fig 4) and the main challenge is the additive color noise with $(\mu = 0, \sigma = 25)$. Although the color noise affects the $P_I(x, y)$ drastically, shape prior helps to locate the object properly. The third scenario includes artificially occluded lamp object. In this case, the advantage of shape prior completes the missing parts perfectly (See Fig 5). The results on these sequences are summarized in Table 1 where we also present the results without using the shape prior. The advantage of using the shape prior is clearly visible in all scenarios. The tracking videos can be found in [7].

4. Conclusions and Future Work

We presented a method for integrating both shape and region-based appearance cues within the ratio-cycle optimization framework. The experiments showed promising results on tracking an object of known shape. Although the optimal cycles can handle limited non-

[1] The image sequence is taken from PETS 2000 dataset (http://www.cvg.rdg.ac.uk/slides/pets.html).

Sequence	Shape Prior	Frames							Average
		1	5	15	25	35	45	50	
car	w/o Shape Prior	0.93	0.93	0.92	0.85	0.81	0.87	0.89	0.89
	Shape Prior	0.96	0.94	0.91	0.91	0.90	0.90	0.91	0.92
mushroom	w/o Shape Prior	0.92	0.91	0.90	0.95	0.74	0.73	0.73	0.84
	Shape Prior	0.95	0.94	0.96	0.96	0.94	0.97	0.96	0.95
lamp	w/o Shape Prior	0.69	0.69	0.43	0.27	0.54	0.85	0.91	0.63
	Shape Prior	0.98	0.95	0.95	0.94	0.97	0.96	0.95	0.96

Table 1. F-measures on three sequences.

Figure 4. Tracking output on *mushroom* sequence.

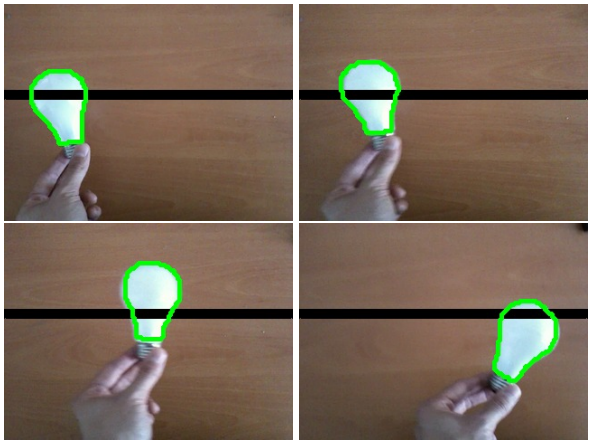

Figure 5. Tracking output on *lamp* sequence.

rigid deformations, we considered the rigid deformations of the shape prior. This limitation can be improved in future by using a non-rigid deformation model such as thin-plate splines.

References

[1] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174–188, Feb 2002.

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1124–1137, Sept. 2004.

[3] A. Dasdan. Experimental analysis of the fastest optimum cycle ratio and mean algorithms. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 9(4):385–418, 2004.

[4] H. Ishikawa and I. Jermyn. Region extraction from multiple images. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 509–516 vol.1, 2001.

[5] I. Jermyn and H. Ishikawa. Globally optimal regions and boundaries as minimum ratio weight cycles. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(10):1075–1088, Oct 2001.

[6] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239, Mar 1998.

[7] M. Sargin. Tracking output videos are available at `http://www.ece.ucsb.edu/~msargin/icpr10`.

[8] T. Schoenemann and D. Cremers. Introducing curvature into globally optimal image segmentation: Minimum ratio cycles on product graphs. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–6, Oct. 2007.

[9] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008.

[10] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Computing Surveys (CSUR)*, 38(4), 2006.