A Probabilistic Shape Filter for Online Contour Tracking

Ibrahima J. Ndiour, Omar Arif, Jochen Teizer and Patricio A. Vela
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250

***Abstract*— Online contour-based tracking is considered through the estimation perspective. We propose a recursive dynamic filtering solution to the tracking problem. The state of the target is described by a pose state which represents the ensemble movement and a shape state which represents the local deformations. The shape state of the filter is described implicitly by a probability field with prediction and correction mechanisms expressed accordingly. The filtering procedure decouples the pose and shape estimation. Experiments conducted with objective measures of quality demonstrate improved tracking.**

Index Terms – Contour tracking, Online visual tracking, Recursive filtering.

I. Introduction

This paper considers the problem of faithful contour-based target tracking under imaging noise and approximate target/background models. Many contour-based techniques view the problem as a detection problem over the individual frames of a given image sequence [2]. However, methods that take advantage of the underlying natural coherence and consistency of the target should provide improved tracking.

Temporal consistency has been used to update the initialization of the detection algorithm at subsequent times [1]. Alternatively, it has been used to process a sequence volumetrically. Such batch processing techniques [10], [12] can successfully recover a solution (through the entire video) that guarantees global coherence and fitness to the measurements by processing the entire sequence at once or several frames before and after the current one. We are interested in an online, recursive method that does not require access to future frames.

The tracking problem can be viewed as an estimation problem given temporally correlated measurements. Recursive filters are derived from a Markov assumption on the temporal state history. Many existing recursive methods usually include reduction of the shape space to a finite-dimensional approximation through PCA, ICA or kernel PCA [4], [5]. Finite-dimensional filtering strategies are then applicable, however, their design requires a careful choice of the training set, a reduction analysis, and possibly a learning phase to estimate the state evolution model in the reduced space [9]. Unfortunately, these techniques are unable to cope with elastic targets whose geometry and shape can drastically change through time; this is known as the out-of-sample problem. The main difficulty resides in the infinite dimensionality of the shape manifold. An alternative consists in avoiding finite-dimensional approximations.

The principal contributions of this paper are: the definition of a recursive estimator on a dynamic target state and the quantitative validation of its performance. The target state model is decomposed into group and shape, with independent filtering stategies on each substate. A second-order model incorporating dynamics is presented that handles non-rigid shape deformations. A probabilistic model is used to describe the shape space and a novel correction mechanism is presented.

The paper is organized as follows. Section II describes the recursive shape filter, which covers the prediction, measurement and correction steps. Section III details experiments conducted and verified using objective performance metrics. Section IV concludes the paper.

II. Recursive Filter Description

A. State Description and General Overview

The state model for the filter must describe the target and the movement of the target in time. The state model chosen follows [13], which defines such movement as being composed of a rigid motion followed by a local deformation. The state decomposition into pose and shape is called *deformotion* [13]. To define a recursive dynamic filter on the dynamic system, the pose velocity and shape velocity must also exist in the state model.

The group variable g is described by $SE(2)$ or its subgroup $E(2)$, describing the group motion as a composition of translation and rotation or as a translation, respectively. Its associated velocity is denoted by ξ. In this work, the shape is implicitly defined by a probability field P representing the probability that each pixel belongs to the target. The 50% probability contour determines the bounding contour of the target. The shape velocity is a vector field Θ defined on the domain of the implicit shape descriptor P. Thus, the internal state of the recursive filter consists of (g, ξ, P, Θ). The recursive filtering structure proposed is depicted in Figure 1. It is a predictor-corrector procedure described below.

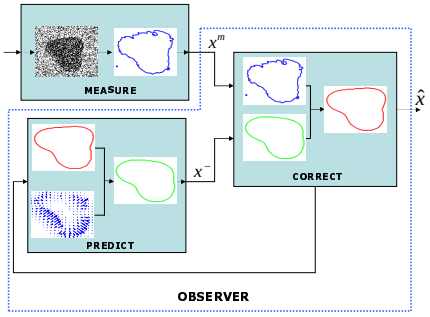


Fig. 1. Visual tracking recursive filter structure.

B. Prediction

The prediction is based on prior knowledge regarding target movement. Here, we provide a general constant velocity motion model for the full filter state

$$\dot{g} = \xi, \qquad \dot{\xi} = 0,$$
$$\dot{P} = \nabla P \cdot \Theta, \qquad \dot{\Theta} = 0.$$

C. Measurement

Measurement of the system involves, at most, the determination of the four sub-states (g, ξ, P, Θ). In practice, ξ is not directly measurable. Note that the measurement block is not a part of the filter, but an external operation providing input to the filter. Any segmentation algorithm can be used to generate the measurements. The segmentation leads to the measured states as follows.

1) Group and Probability Field Measurement: If the segmentation is not a probability field, then additional processing may be required in order to convert the initial output of the segmentation algorithm into a probability field. For example, if an active contour segmentation technique producing signed distance functions was used to generate the measurements, then the processing would transform the signed distance function into a probability field, with the zero level set matching the 50% contour.

The segmentation is registered against the predicted state. The registration procedure thus identifies the coordinates of the two constitutive components (g_m, P_m).

2) Velocity Field Measurement: There are several feasible methods for generating the shape velocity measurement. The method used here is the optical flow field [7] between the current target image data and the previous target image data (after the registration step described above). There is abundant literature about computation of the optical flow field.

D. Correction

Given prediction and measurement, the correction step generates an updated estimate of the target state. In the following, the subscript m, and superscripts $-$ and $+$ denote measurement, prediction and correction, respectively. As the pose and shape corrections are decoupled, we describe them separately.

1) Group: Since the pose and its velocity, (g, ξ) are finite dimensional, correction can be performed through standard filtering methods. For example, given linear dynamics, Kalman filtering can be used.

2) Shape: Due to the infinite dimensional nature of the shape state, a unique correction method does not exist. In the following, we define a novel correction scheme on the probabilistic shape space.

Geometric Blending: Correction of the shape consists of the weighted geometric mean of the predicted and measured probability fields, and their complements,

$$P^+ = (P^-)^{1-K} \cdot (P_m)^K \quad \text{and}$$
$$Q^+ = (1-P^-)^{1-K} \cdot (1-P_m)^K,$$

after which P^+ and Q^+ are normalized, but only P^+ is kept. The parameter K lies in the range $[0, 1]$ and is defined by the user according to measurement noise. Low K indicates high measurement noise, with correction biased to the prediction. High K indicates the opposite.

3) Shape Velocities: The shape velocity is a vector field, thus it resides in a linear space. Corrections are induced on the velocity field through errors in the measured probability fields and/or the measured image velocities:

$$\Theta^+ = \Theta^- + K_{vx} \cdot X_{err}(P_m, P^-) + K_{vv} \cdot (\Theta_m - \Theta^-)$$

One way to generate the error vector field $X_{err}(P_m, P^-)$ is to compute the optical flow between the measured and predicted shapes. The parameters K_{vx} and K_{vv} vary in $[0, 1]$ and are chosen according to the measurement noise.

III. Experiments and Results

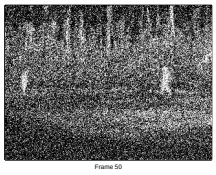
We conducted experiments to demonstrate the ability of the proposed filter to enhance contour-based trackers and provide consistent tracking under perturbations. Due to space constraints, results of three representative video sequences are presented: a synthetically noise-corrupted infrared sequence from the OTCBVS dataset, a color sequence featuring a working man whose body shape varies, and likewise a swimming fish. Ground truth for the sequences consists of manual track point determination and manual segmentations. Using the ground truth, quantitative performance evaluations are obtained through objective measures of quality. For the group variable, we use the L_2 and L_∞ errors of the pose signal. For the shape variable, frame-wise we computed the number of misclassified pixels, the Hausdorff and Sobolev distances [3], [14], and the number of tracked frames per sequence. Strikeouts indicate loss of track (metrics are computed up to track loss).

Several contour-based trackers were implemented to provide comparison. They include Bayesian segmentation [6], active contours (AC) [11], *deformotion* filtering [8], and the shape-based method described in [4] modified to incorporate a Kalman filter on the shape space. Bayesian segmentation is the base measurement strategy for the proposed and *deformotion* filters. Where applicable, the pose state was Kalman filtered.

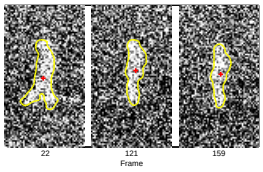
The depicted frames for each sequence (Figs 2-4) are: (left frame) the lowest error obtained, (middle frame) a frame with average error, and (right frame) the largest error obtained. Quantitative results are summarized in Table I, where the average and maximum for each shape metric/sequence are displayed. Visual inspection of the tracked videos corroborates the numbers.

The filter improves upon the unfiltered method (Bayesian segmentation), and equals or outperforms the comparison algorithms (lower metric values are better). Table I-d also shows that the proposed filter can accomodate alternative segmentation strategies.

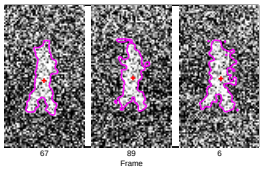
The proposed filter has a low computational cost: our sub-optimal Matlab implementation can process about three frames per second. Significantly higher frame rates (up to real-time) can be achieved by considering implementations on dedicated architectures such as GPU's.



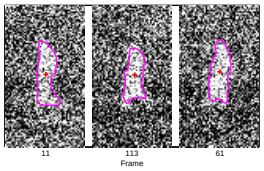
(a) Sequence 1



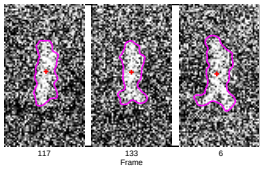
(b) Ground Truth 1



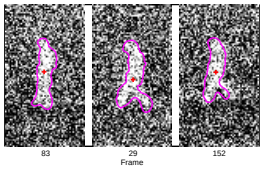
(c) Bayesian



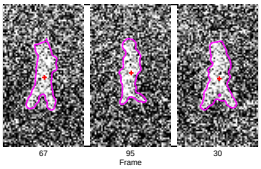
(d) Active Contour



(e) Deformotion



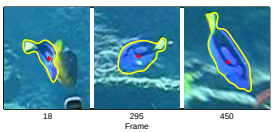
(f) Shape-based



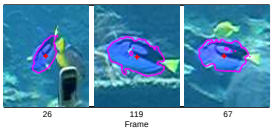
(g) Filtered Bayesian

Fig. 2. Snapshots of Sequence 1.

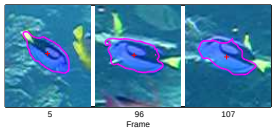
(a) Sequence 2



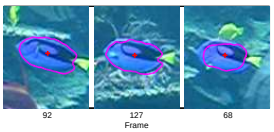
(b) Ground Truth



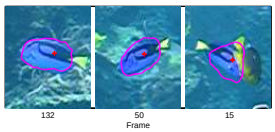
(c) Bayesian Tracker



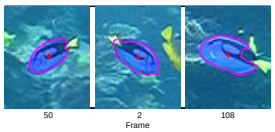
(d) Active Contour



(e) Deformotion



(f) Shape-based



(g) Filtered Bayesian

Fig. 3. Snapshots of Sequence 2.

(a) Sequence 3

(b) Ground Truth 3

(c) Bayesian Tracker

(d) Active Contour

(e) Deformotion-based Tracker

(f) Shape-based

(g) Filtered Bayesian

Fig. 4. Snapshots of Sequence 3.

(a) Sequence 1

Metric \ Algorithm	Bayesian	AC	Deformotion	Shape	Filtered Bayesian
Trackpt error (L_2/L_∞)	1.8 / 4.4	1.4 / 3.9	1.2 / 3.4	3.0 / 10.5	1.2 / 4.4
NMP (avg/max)	129 / 242	91 / 160	116 / 211	105 / 199	90 / 155
Hausdorff (avg/max)	6.2 / 13.5	4.5 / 9.5	4.0 / 6.7	3.9 / 7.7	3.5 / 6.6
Sobolev (avg/max)	3.2 / 10.5	2.4 / 6.9	1.5 / 3.6	1.5 / 5.4	1.2 / 3.3
# Frames tracked	180	180	180	180	180

(b) Sequence 2

Metric \ Algorithm	Bayesian	AC	Deformotion	Shape	Filtered Bayesian
Trackpt error (L_2/L_∞)	8.6 / 13.2	2.8 / 7.0	2.6 / 12.3	5.6 / 15.8	2.7 / 5.8
NMP (avg/max)	251 / 969	244 / 549	248 / 769	575 / 833	279 / 478
Hausdorff (avg/max)	10.9 / 18.4	11.1 / 19.2	12.3 / 19.7	12.0 / 22.5	14.6 / 20.7
Sobolev (avg/max)	8.2 / 52.9	12.9 / 95.8	11.9 / 46.7	13.2 / 43.9	12.9 / 26.9
# Frames tracked	477	478	477	475	478

(c) Sequence 3

Metric \ Algorithm	Bayesian	AC	Deformotion	Shape	Filtered Bayesian
Trackpt error (L_2/L_∞)	16.6 / 24.4	~~11.5 / 52.3~~	7.9 / 16.0	5.4 / 12.3	8.0 / 15.5
NMP (avg/max)	253 / 1420	~~288 / 1328~~	202 / 755	299 / 536	171 / 508
Hausdorff (avg/max)	10.2 / 35.0	~~30.0 / ∞~~	7.8 / 26.2	10.9 / 25.8	7.7 / 27.4
Sobolev (avg/max)	8.2 / 70.6	~~100.0 / ∞~~	5.8 / 35.3	11.7 / 38.1	6.5 / 81.8
# Frames tracked	200	~~150~~	200	200	200

(d) Sequence 3 using Graph Cut and Active Contour segmentations with proposed filter.

Metric \ Algorithm	Filtered AC	Graph Cut	Filtered Graph Cut
Trackpt error (L_2/L_∞)	6.5 / 22.1	7.9 / 31.1	6.9 / 27.4
NMP (avg/max)	192/ 663	288 / 1014	219 /457
Hausdorff (avg/max)	8.3 / 25.4	12.8 / 32.0	11.7/ 25.9
Sobolev (avg/max)	6.2 / 35.8	8.3 / 70.8	10.2/ 80.1
# Frames tracked	200	200	200

TABLE I

TABLES OF PERFORMANCE STATISTICS FOR THE TRACK SEQUENCES.

IV. CONCLUSION

We have presented the design of a recursive dynamic filter for the purpose of tracking consistently through imaging perturbations. Objective measures of quality demonstrated that the proposed filter achieves temporal consistency and is equal to or more effective than other tracking algorithms in an online, recursive estimation setting. It does not require any training and has low computational cost.

REFERENCES

[1] M. Bertalmio, G. Sapiro, and G. Randall. Morphing active contours: a geometric approach to topology-independent image segmentation and tracking. In *ICIP*, pages 318–322, 1998.

[2] V. Caselles. Geometric models for active contours. In *ICIP*, pages 9–12, 1995.

[3] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Found. of Comp. Math.*, 5(2/2):1–58, 2005.

[4] D. Cremers. Dynamical statistical shape priors for level set-based tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1262–1273, 2006.

[5] S. Dambreville, Y. Rathi, and Tannenbaum A. Tracking deformable objects with unscented Kalman filtering and geometric active contours. In *ACC*, pages 2856–2861, 2006.

[6] S. Haker, G. Sapiro, A. Tannenbaum, and D. Washburn. Missile tracking using knowledge-based adaptive thresholding: Tracking of high speed projectiles. In *ICIP*, pages 786–789, 2001.

[7] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[8] J.D. Jackson, A.J. Yezzi, and S. Soatto. Tracking deformable moving objects under severe occlusions. In *CDC*, pages 2990–2995, 2004.

[9] J. Malcolm, Y. Rathi, and A. Tannenbaum. Graph cut segmentation with nonlinear shape priors. In *ICIP*, volume 4, pages 365–368, 2007.

[10] N. Papadakis and E. Memin. Variational optimal control technique for the tracking of deformable objects. In *ICCV*, pages 1–7, 2007.

[11] M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. In *Proceedings IEEE Workshop on Motion and Video Computing*, pages 56–61, 2002.

[12] J. Xiao and M. Shah. Motion layer extraction. in the presence of occlusion using graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1644–1659, 2005.

[13] A.J. Yezzi and S. Soatto. Deformotion: Deforming motion, shape average and the joint registration and approximation of structures in images. *International Journal of Computer Vision*, 53(2/2):153–167, 2003.

[14] L. Younes. Optimal matching between shapes via elastic deformations. *Image and Vision Computing Journal*, 17:381–389, 1999.