A Probabilistic Observer for Visual Tracking

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

***Abstract*— This paper describes an observer for estimation of the rigid pose and shape states associated to an object being tracked in an image sequence. The defined observer utilizes standard estimation strategies for the finite-dimensional rigid pose sub-states, and a novel strategy for the shape sub-states. In particular, the shape sub-state observer utilizes an implicit probability field, where the 50% probability iso-contour defines the object shape. A general purpose second-order model and a corresponding correction scheme are defined for the observer state. The observer is applied to recorded imagery and its performance is examined using objective error metrics.**

I. Introduction

This paper considers the problem of accurate contour-based, or segmentation-based, target tracking when faced with uncertainty in the form of imaging noise and approximate segmentation models. Imaging noise can arise from the actual sensing procedure [9] or from the image handling technique (if lossy compression is required to transmit data over a network). Approximate segmentation models arise from the use of image formation models that are simple relative to the true 3D scene being imaged.

Segmentation-based tracking often views tracking as a series of statically determined measurements for each frame. The introduction of temporal knowledge as derived from the image sequence leads to improvements in the tracking procedure [2]. Alternatively, the tracking problem can be tackled over the entire spatio-temporal volume of image data [17], [22]. Such an approach is warranted when the objective is to post-process a pre-existing video sequence. It is not well-suited, however, to the problem of online, recursive estimation.

The tracking problem can be viewed as an estimation problem given temporally correlated measurements. A Markovian assumption relating the current parameters to the prior parameters simplifies the problem to one of recursive estimation. Recursive methods often incorporate (low) finite-dimensional parametrizations of the target shape space, [6], [7], [21]. 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. Unfortunately, many of these techniques are unable to cope with elastic targets whose geometry and shape drastically change through time or fail to be represented in the training set; this is known as the out-of-sample problem [1].

The work here considers the unconstrained segmentation and tracking problem. Instead of relying on shape information to constrain the segmentation, temporal consistency will be imposed on the segmentations and track points. Temporal consistency is obtained by describing an observer for the generated measurements and using the observer states as the estimated state, rather than the measurements. Related work includes [3], [15], [16], [20], which examined temporal consistency in an infinite-dimensional non-parametrized setting. Alternatively, the overall object motion can be decomposed into a principal fiber, consisting of a group component (the rigid pose) and a shape component, to be filtered over as in [11].

Contributions: The contributions of this paper include (1) the definition of the estimation problem as an observer on the group and shape, (2) the use of an implicit probability field description for the shape state, (3) the incorporation of a dynamical model for the shape space, and (4) a correction method adapted to the probabilistic state model which decouples from the group correction. Furthermore, quantitative evaluation of tracking results for both pose and segmentation are provided, which is unique as segmentation-based papers do not historically quantitatively evaluate the segmentation results against ground truth. Lastly, the observer strategy is agnostic of the segmentation strategy, so long as the measured shape state can be converted into a probability field.

II. State Description

As described in [23], the state description of a tracked object in a video sequence can be decomposed into a group motion and a shape deformation (the decomposition is not unique). In a follow-up, filtering of the signal was achieved through the minimization of an energy in the joint group-shape space [11], which coupled the group and shape correction strategy. The group motion was the special Euclidean group $SE(2)$ or its subgroup $E(2)$, the Euclidean group. The shape description was given by non-parametric active contours [4]. Non-parametric active contours embed the segmented closed curve into a higher-dimensional space, e.g., the level-set of a signed distance function.

In practice, any function capable of implicitly describing a shape through the selection of an iso-contour suffices. This work proposes to utilize a probability field, $P : D \rightarrow [0, 1]$ where $D \subset \mathbb{R}^2$ compact. The implicitly defined contour $\mathcal{C}$ is given by the set $\mathcal{C} = \{r \mid P(r) = c\}$ where $c \in (0, 1)$ and $r \in D$. Pixels with probabilities higher than or equal to c are presumed to belong to the target and those with lower probabilities are presumed to belong to the background object (we use the value $c = 0.5$).

Consequently, the shape space description is the space of probability fields defined on a compact domain. For second-order systems, we will also require the velocity field associated to the probability field, denoted by $X : D \rightarrow \mathbb{R}^2$. Thus, at its most complex, the complete state is described by the group space and its velocity, denoted by (g, ξ), and the probability field plus its associated velocity field, (P, X).

III. Observer Components

An observer requires the definition of three major components plus an input, (1) a dynamic model for the state, (2) a measurement model for the state, and (3) a correction strategy for the internal model given state measurements. The state measurements should occur external to the observer. They will be generated by a segmentation algorithm, called the *sensor measurement*, applied to the current image.

A. Dynamic Update Model

The dynamic model is based on prior knowledge that is available regarding the motion of the target. Describing the dynamic evolution of an unconstrained shape can be somewhat difficult, therefore the motion models proposed will be generic, see Table I. For more precise motion models one can derive a motion model from first principles, given knowledge of the target.

Static Prior: This model simply assumes that there is no change in the state from one time-step to the next.

Constant Group Velocity: This model only considers the temporal evolution of the group space with a constant shape model. It is appropriate for rigid objects, almost-rigid objects, or slowly varying objects (relative to the measurement rate).

Constant Velocity: Here, the group has constant velocity and the shape is deforming with a constant rate of deformation.

B. Sensor Measurement

Measurement of the target can be achieved through any segmentation algorithm applied to the current image, so long as the segmentation is converted to an implicit probability field description (further discussed in Section IV). Candidate algorithms include Bayesian segmentation [8], active contours [2], graph cuts [13], etc.

In a classical observer the measurements would be completely independent of the observer states, however image analysis of video sequences has the unique nature of not explicitly providing the necessary signal. Instead it must be extracted from the image using an image processing or computer vision algorithm. The measurement procedure may not completely determine the necessary target state measurements (due to the non-uniqueness of the group + shape decomposition). Consequently, a registration step is required to describe the predicted and measured shape with respect to the same coordinate frame.

TABLE I
State Motion Models

Static prior	$\begin{cases} \dot{g} = 0 \\ \dot{P} = 0 \end{cases}$
Constant group velocity	$\begin{cases} \dot{g} = \xi, & \dot{\xi} = 0 \\ \dot{P} = 0 \end{cases}$
Constant velocity (1)	$\begin{cases} \dot{g} = \xi, & \dot{\xi} = 0 \\ \dot{P} + \nabla P \cdot X = 0, & \dot{X} = 0 \end{cases}$
Constant velocity (2)	$\begin{cases} \dot{g} = \xi, & \dot{\xi} = 0 \\ \dot{P} + \nabla P \cdot X = 0, & \dot{X} + \nabla X \cdot X = 0 \end{cases}$

Once localization and segmentation are performed on the current image, a registration procedure is applied to match the resulting measured probability field with the predicted probability field, yielding a measurement g_m for the group motion and the measurement P_m for the shape. If desired, the velocity field X_m can be measured by computing the optical flow [10] between two subsequent aligned images. In practice, the group velocity ξ is not directly measurable.

C. Model Measurement

The model measurement is obtained by extracting the components of the internal state model that are equivalent to those obtained from the sensor measurement.

D. Correction

The role of the correction model is to generate an updated estimation of the observer internal state given the predicted measurement and the actual measurement. For the finite-dimensional group component, generating a correction is relatively straightforward. However, for the shape space, there is no unique method for doing so. Instead the infinite-dimensional manifold nature of the unconstrained shape space implies the existence of multiple correction strategies [14], [16].

a) Group Space and Velocities: In this paper, the group space filter is typically chosen to be on either $TE(2)$ or $TSE(2)$, the tangent spaces to $E(2)$ and $SE(2)$, respectively. For the former, a discrete Kalman filter correction is used. For the latter, a discrete extended Kalman filter correction is used.

b) Shape Space: Correction on the probability field is achieved through geometric averaging. Given the predicted

probability field, $\widehat{P}^-$, and the current measured probability field, $\widehat{P}_m$, the current corrected probability field, $\widehat{P}^+$, is

$$\widehat{P}^+(r) = \left(\widehat{P}^-(r)\right)^{1-K_{xx}} \left(\widehat{P}_m(r)\right)^{K_{xx}},$$

where K_{xx} varies in the range $[0, 1]$ and is chosen based on uncertainty estimates of the measurement and the prediction. Low K_{xx} is biased towards the predicted probability, high K_{xx} is biased towards the measured probabilities. This method works when the prediction and measured probability fields dot not differ radically.

When the two fields have sufficient disparity, then the geometric averaging technique no longer works due to non-local shape effects. Instead, an error vector field X_{err} needs to be computed between the predicted and measured densities. Flowing along the error vector field should take P^- to P_m in unit time. The error vector field between the two densities can be computed using a variety of methods, such as optical flow, displacement flow or optimal mass transport [16], [18]. The correction is then given by

$$\widehat{P}^+ = \Phi_{K_{xx}}^{X_{err}}(\widehat{P}^-),$$

where $\Phi_{K_{xx}}^{X_{err}}$ denotes the flow along X_{err} for time K_{xx}. In practice, we did not find this to be necessary.

c) Shape Tangent - Velocities: The velocity field is much simpler to correct on since it is a vector space. There are two ways to induce a correction on the velocity field, one is through an error in the measured probabilities and the other is through an error in the shape velocities:

$$X^+ = X^- + K_{vx} X_{err}(P_m, P^-) + K_{vv}(X_m - X^-)$$

where $X_{err}(P_m, P^-)$ is as defined above. The parameters K_{vx} and K_{vv} vary in the range $[0, 1]$ and are chosen according to uncertainty estimates of the measurement and prediction.

IV. Experiments and Results

This section describes the experimental setup used to implement and validate the observer.

1) Sensor Measurement: As alluded to in Section III-B, the measurements of the shape can be performed using any segmentation algorithm. If the segmentation algorithm does not automatically generate a probability field, then conversion to implicit probability field form is required.

As a case in point, consider an active contour with an implicit representation using the signed-distance function, $\Psi : D \to \mathbb{R}$, then the measurement must be mapped into a probability field description. Such a mapping can be done using the regularized Heaviside function

$$P(\cdot) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{\Psi(\cdot)}{\sigma}\right)\right),$$

or by applying the cumulative density function of the normal distribution (with zero mean) to the negative signed-distance function,

$$P(\cdot) = \mathrm{cdf}(-\Psi(\cdot); \sigma) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\Psi(\cdot)}{\sigma\sqrt{2}}\right)\right),$$

where σ is the standard deviation. In both cases, σ is a regularization parameter.

2) Setup: Tracking experiments were performed on several image sequences using four different algorithms, plus one instantiation of the observer. Two of the algorithms were purely segmentation-based tracking algorithms utilizing, Bayesian segmentation [8] and active contours [19]. The third algorithm run was the deformotion filter technique developed in [11], utilizing the Bayesian segmentations as measurements. The fourth algorithm was a shape-based active contour tracking algorithm [6], however only for the infrared sequence was it realistic to caputre the dynamic motion model. We obtained 7 shapes for the infrared sequence, 67 sample shapes from the construction worker imagery and 60 shapes from the aquarium imagery we have. For infrared data set the five dominant eigenmodes were kept, and ten dominant eigenmodes were kept for the others. Lastly, the probabilistic contour observer with Bayesian segmentation as the measurement strategy was implemented. Note that the deformotion filter and the observer share the same measurement method. Furthermore, all strategies share the same statistical description of the target and background distributions (some use the negative log likelihood for segmentation purposes). For all methods, we applied the same filter to the group space; a second order Kalman filter.

In order to compare the resulting track signals, all of the sequences were hand-segmented to generate a ground-truth signal. Quantitative metrics were used to objectively compare the performances of the different techniques. On the group space, we used the L_2 and L_∞ errors. On the shape, we used the number of misclassified pixels, the Hausdorff distance, and the Sobolev distance. Information about the shape metrics can be found in [5], [12].

3) Results: Figures 1-3 feature, for each sequence and each technique, three frames from the tracking results. The frames from left to right correspond to the best frame tracked, the average tracking behavior, and the worst frame tracked. Table II shows for each metric and each technique, the average error and the maximal error obtained throughout the given sequence (strikeouts indicate loss of track).

The Bayesian segmentation algorithm used as a tracker tends to provide the noisiest contour estimates, due to a moderate smoothing term. Sources of error include image noise, image clutter, and poor modeling of the target and background distributions. Increasing the smoothing term would lead to undersegmentations of the target. For the active contour, the smoothing term was set to provide the best overall segmentation on a frame-by-frame basis. The Bayesian observer is able to utilize temporal consistency to arrive at a smoother result without requiring significant spatial smoothing on a single-frame basis.

Consider now the shape-based and deformotion filter techniques. When the shape-based methods tracks, it is

possible to see incomplete or incorrect segmentations arising from the inability of the tracker to conform to a shape sufficiently outside of its training set. The mismatch is most obvious for Sequence 3 (Figure 3). Even though a couple of the bending segmentations were used, they did not factor into the main eigenmodes. It may be possible to improve the segmentation by allowing for more eigenmodes or using a more complicated shape method [21], however this would increase the computational cost of the algorithm and possibly also the size of the training set. The deformotion-based technique uses an averaging procedure on the shape, due to a static model on the shape dynamics, and is unable to efficiently capture fast shape deformations (see Figure 2 and Table II-b).

The quantitative comparison of the methods is given in Table II. The performance metrics should be examined as an ensemble since they rate different aspects of curve tracking performance. The Bayesian observer performs as well as the other techniques when there is no perturbation (see average error). When perturbations occur and the measurements become unreliable, the observer can attenuate the measurement perturbations (low maximum errors in Table II). Overall, the Bayesian observer demonstrates good temporal consistency in the target contour, given the quality of the measurements. The technique is also able to handle highly variable shape deformations.

Lastly, the observer strategy was applied to the active contour algorithm and also to a graph-cut tracking algorithm [13] for Sequence 3. Results are provided in Table II-d. Note that the original test of the active contour was not able to track, but with the addition of the observer, the target can be tracked completely. Performance is comparable to, or better than, the Bayesian with observer. The observer was also able to improve upon the graph-cut tracking algorithm.

These examples indicate that improvements can be achieved by using an observer in conjunction with standard segmentation algorithms for target tracking. Moreover, the technique is robust to parameter selection. Figure 4 displays the average number of misclassified pixels when the comparison tracking algorithms and the observer are used to track multiple noise-corrupted sequences with a fixed set of parameters.

V. Conclusion

This paper presented the design of a visual tracking observer utilizing an implicit probablistic representation, in the guise of a probability field, for the internal state of the observer. A geometric averaging method was detailed for filtering the shape state, together with a filtering strategy on the velocity field of the shape state. The observer is agnostic of the segmentation strategy. Further, the filtering of the rigid motion and the shape motion are decoupled, providing flexibility in the observer design. Application of the algorithm to infrared, construction site, and aquarium video sequences was performed, including a quantitative performance analysis. When perturbations occur leaving segmentation measurements corrupted, the observer is capable of attenuating them.

References

[1] P. Arias, G. Randall, and G. Sapiro. Connecting the out-of-sample and pre-image problems in kernel methods. In *CVPR*, pages 1–8, 2007.

[2] A. Blake and M. Isard. *Active Contours*. Springer Verlag, 1998.

[3] R. Brockett and A. Blake. Estimating the shape of a moving contour. In *CDC*, pages 3247–3251, 1994.

[4] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 13:5–22, 1997.

[5] G. Charpiat, O. Faugeras, R. Keriven, and P. Maurel. Distance-based shape statistics. In *ICASSP*, pages 925–928, 2006.

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

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

[8] 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.

[9] G.E. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994.

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

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

[12] E. Klassen, A. Srivastava, W. Mio, and S.H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383, 2004.

[13] J. Malcolm, Y. Rathi, and A. Tannenbaum. Tracking through clutter using graph cuts. In *BMVC*, page 116, 2007.

[14] P.W. Michor and D. Mumford. An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23(1):74–113, 2007.

[15] I.J. Ndiour and P.A. Vela. Towards a local Kalman filter for visual tracking. In *CDC*, pages 2420–2426, 2009.

[16] M. Niethammer, P.A. Vela, and A. Tannenbaum. Geometric observers for dynamically evolving curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1093–1108, 2008.

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

[18] G. Pryor, T. Ur-Rehman, S. Lankton, P.A. Vela, and A. Tannenbaum. Fast optimal mass transport for dynamic active contour tracking on the GPU. In *CDC*, pages 2681–2688, 2007.

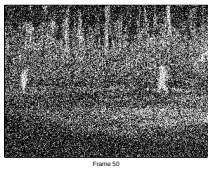
[19] M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. In *Proceedings IEEE Workshop on Motion and Video Computing*, pages 56–61, 2002.

[20] G. Sundaramoorthi, A. Mennucci, S. Soatto, and A.J. Yezzi. Tracking deforming objects by filtering and prediction in the space of curves. In *CDC*, pages 2395–2401, 2009.

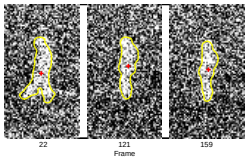
[21] N. Vaswani, A.J. Yezzi, Y. Rathi, and A. Tannenbaum. Time-varying finite dimensional basis for tracking contour deformations. In *CDC*, pages 1665 –1672, 2006.

[22] 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.

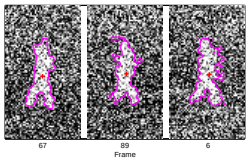
[23] A.J. Yezzi and S. Soatto. Deformotion: Deforming motion, shape average and the joint registration and approximation of structures in images. *International Journal of Computer Vision*, 53(2/2):153–167, 2003.



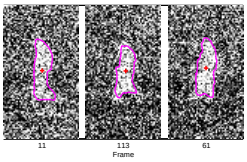
(a) Sequence 1



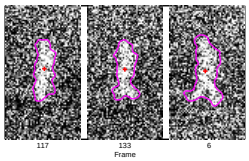
(b) Ground Truth



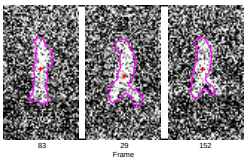
(c) Bayesian



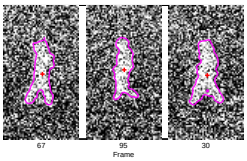
(d) Active Contour



(e) Deformotion Filter



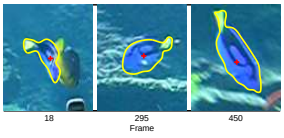
(f) Shape-based



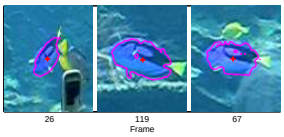
(g) Observer

Fig. 1. Snapshots of Sequence 1.

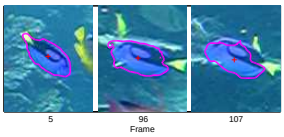
(a) Sequence 2



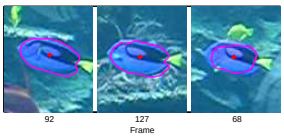
(b) Ground Truth



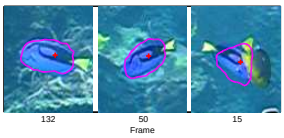
(c) Bayesian



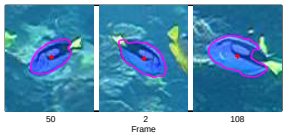
(d) Active Contour



(e) Deformotion Filter



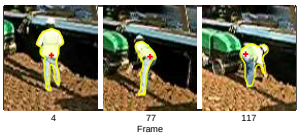
(f) Shape-based



(g) Observer

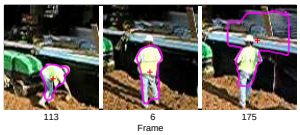
Fig. 2. Snapshots of Sequence 2.

(a) Sequence 3



(b) Ground Truth

(c) Bayesian



(d) Active Contour

(e) Deformotion Filter

(f) Shape-based

(g) Observer

Fig. 3. Snapshots of Sequence 3.

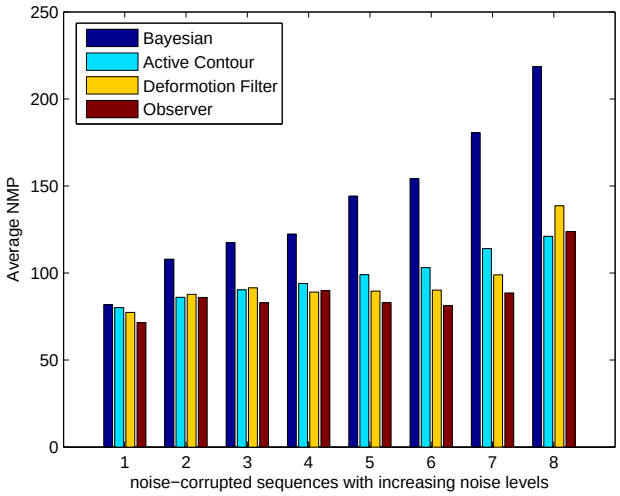


Fig. 4. Robustness to parameter selection. Using a fixed set of parameters for each algorithm, tracking is performed on noise-corrupted sequences with increasing levels of noise. The average number of misclassified pixels is displayed for each technique and each sequence.

TABLE II

COMPARATIVE PERFORMANCES USING OBJECTIVE MEASURES OF QUALITY.

(a) Sequence 1

Metric / Algorithm	Bayesian	AC	Deformotion	Shape	Observer
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	Observer
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	Observer
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 with observer using measurements from Graph Cut and Active Contour segmentations

Metric / Algorithm	Observer with AC measurements	Graph Cut	Observer with Graph Cut measurements
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
Mean Laplace (avg/max)	8.3 / 25.4	12.8 / 32.0	11.7 / 25.9
Max Laplace (avg/max)	6.2 / 35.8	8.3 / 70.8	10.2 / 80.1
# Frames tracked	200	200	200