Optimal Estimation Applied to Visual Contour Tracking

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

***Abstract*— This paper derives an optimal estimator for the purpose of online visual contour tracking. Starting from Bayesian segmentation as the measurement strategy, we use a bottom-up approach to design the estimator. In particular, it is shown that additive imaging noise leads to multiplicative segmentation uncertainty from which a geometric averaging update model is established. Given known noise statistics, the optimal correction gain and associated filtering equations are derived. The optimal estimator is applied to noise-corrupted imagery and its performance compared against a fixed-gain filtering strategy and other visual tracking techniques.**

I. Introduction

This work considers the problem of accurate contour-based tracking in the presence of perturbations caused by imaging noise. Given a sequence of corrupted images $\{I_1, I_2, ... : \Omega \longmapsto \mathbb{R}\}$, performing a series of individual segmentations generates noisy, incorrect contour measurements that may lead, over time, to loss of track. Typical solutions to the problem of noise uncertainty include filtering the image data prior to performing visual tracking, enforcing shape constraints [2] or considering more complex segmentation algorithms [5], [17].

Another class of solutions is given by estimation techniques. Often, these approaches formalize the problem as a nonlinear estimation problem, with the intensity of each pixel being an observation. Most of the related techniques use a top-down approach, featured most prominently in [10], where the general block structure of an observer is proposed and each component is subsequently specified. There are many difficulties associated to these techniques. First, the contour to be estimated resides in an infinite-dimensional manifold space [8] requiring infinite-dimensional filtering design [7], [9], [11]. This leads in general to complex filtering design and possibly large computational cost [14]. The approximation of the shape space through PCA methods simplifies the problem [3], but derived systems are then prone to out-of-sample whenever a shape outside the training set arises during tracking. Secondly, while the selection of a gain parameter is often crucial for the performances, most estimators currently proposed require manual gain selection [7], [10] or quantitative uncertainty levels such as the measurement error covariance matrix [3]. Consequently, these are often chosen by the external user based on a subjective perception of visual perturbations in the sequence of images.

This paper proposes the derivation of an optimal estimator for online visual contour tracking. Instead of a top-down approach, we utilize a bottom-up approach starting from the definition of the measurement strategy, then consider the effects of noise. Bayesian segmentation [6] is chosen as the measurement process. In this setting, the contour encircling the target is given implicitly as the 50% iso-contour of a scalar field describing at each pixel the probability that said pixel belongs to the foreground. We examine how the hypothesis of additive imaging noise affects the classification probabilities, infer the proper update law to be applied under such hypothesis, and derive the resulting optimal filtering scheme. Benefits of this approach include the simplification of a filtering problem on the infinite-dimensional space of closed curves into a series of point-wise filtering tasks. Also, this framework allows the computation of the optimal gain, given knowledge of the uncertainty level on the image data, e.g. the noise variance. Principal contributions of the work include: (1) the formulation of the visual tracking task as a bottom-up filtering design problem, (2) the derivation of an optimal filter, and (3) the quantitative validation of the filter's performance.

The paper is organized as follows. Section II describes the contour measurement algorithm. Sections III and IV discuss the filtering design and provide the algorithmic description for optimal estimation. The experiments are reported in section V. Section VI concludes the paper.

II. The measurement strategy

The Bayesian image segmentation algorithm relies on statistical analysis of the image sequence with classification done through a *maximum a posteriori* approach; it is sometimes referred to as knowledge-based segmentation. The *maximum a posteriori* (MAP) segmentation algorithm with Bayesian update, as implemented for image processing, is an adaptive thresholding algorithm that has found much success in processing and quantizing noise corrupted imagery [6].

The Bayesian segmentation algorithm interprets an image to be the composition of several layers, each of

which is described by a class $c \in C$ from a collection of classes. Each class has associated to it a distribution describing the expected data values v of the class, $\Pr(v|c)$. Such distributions, also called likelihoods, are commonly assumed to be Gaussian. Lastly, for each class c, there is an *a priori* probability of a pixel r being assigned to that particular class $\Pr(c(r) = c)$. The Bayesian classifier selects the most likely class for a given pixel based on the probability $\Pr(c(r) = c \,|\, v(r) = v)$ that a given pixel value $v(r)$ is associated to the class. Classification probabilities are obtained using Bayes' rule

$$\Pr(c_i = c|v_i = v) = \frac{\Pr(v_i = v|c_i = c)\Pr(c_i = c)}{\sum_\gamma \Pr(v_i = v|c_i = \gamma)\Pr(c_i = \gamma)}.$$

The 50% probability contour associated with the target class generates the segmentation. For more details on Bayesian segmentation, we refer to [6].

III. Optimal Observer Design

In this section, we provide the derivation of an optimal estimator for visual tracking. First, we show that corruption by additive imaging noise results in multiplicative uncertainty for the contour measurements when Bayesian segmentation is used. The use of a geometric averaging update model is then justified in this context of filtering with multiplicative noise. Subsequently, the noise statistics are estimated and the filtering equations and optimal gain are derived.

A. From additive imaging noise to multiplicative segmentation uncertainty

Consider an image I defined over a compact domain of the plane and taking values in $\mathbb{R}$. Further, assume that measurement of the pixel intensities has been corrupted by additive Gaussian noise ν with zero mean and variance σ_ν^2. Classification is performed through Bayesian segmentation [6] with two classes: foreground and background. The two classes are modelled with a Gaussian distribution for the pixel intensities. Assuming uniform priors and a normal distribution $\mathcal{N}(\mu_F, \sigma_F^2)$ for the foreground pixels, then the measured likelihood for the corrupted pixel $I(r)$ to be classified as foreground is given by:

$$\zeta_F(r) = \sqrt{\delta} \cdot e^{-\frac{1}{2}\left(\frac{I(r)+\nu(r)-\mu_F}{\sigma_F}\right)^2},$$

where δ is a positive normalizing factor. The expression for the measured likelihood can be expanded further:

$$\zeta_F(r) = \sqrt{\delta}\cdot e^{-\frac{1}{2}\left(\frac{I(r)-\mu_F}{\sigma_F}\right)^2} \cdot e^{-\frac{1}{2}\left(\frac{\nu(r)}{\sigma_F}\right)^2} \cdot e^{-\left(\frac{\nu(r)\,(I(r)-\mu_F)}{\sigma_F^2}\right)}, \tag{1}$$

which can be rewritten as

$$\zeta_F(r) = P_F(r) \cdot \eta(r; \mu_F, \sigma_F),$$

where $P_F(r)$ consisting of the first two terms from (1) is the true classification likelihood, and $\eta(r; \mu_F, \sigma_F)$ consisting of the remaining two terms is the class measurement noise. A similar derivation holds for the background classification densities. Thus, corruption by additive noise on the image data results in multiplicative uncertainty for the foreground/background likelihoods.

B. The geometric averaging update model

We are now interested by the estimation problem of a process given measurements with multiplicative noise. Consider a state $\rho \in (0, 1)$ to be estimated, given a collection of measurements ζ. The measurements are corrupted by multiplicative noise, i.e. $\zeta = \rho \cdot \eta$. A recursive filter following a predictor-corrector structure is proposed. The prediction step of the filtering scheme can be chosen to be static (propagation of the previous state estimate) or dynamic given prior knowledge of the state evolution. The update step of the filtering is described by:

$$\hat{\rho}^+ = \left(\hat{\rho}^-\right)^{1-K} \cdot (\zeta)^K.$$

The previous update equation can be justified by considering the logarithm of the expression:

$$\log(\hat{\rho}^+) = (1 - K) \cdot \log(\hat{\rho}^-) + K \cdot \log(\zeta).$$

Rearranging the terms of the previous equation yields:

$$\log(\hat{\rho}^+) = \log(\hat{\rho}^-) + K \cdot \left[\log(\zeta) - \log(\rho^-)\right],$$

which is the standard linear approach to filtering. The combination of the geometric averaging update model with a prediction model and the Bayesian segmentation as a measurement strategy results in a recursive filter estimating the likelihood that a pixel belong to a given class. The filtering applies to both the foreground and background.

C. Noise statistics estimation

In the log-space associated to the likelihoods, the foreground measurement noise has mean $-\frac{1}{2}\left(\frac{\sigma_\nu}{\sigma_F}\right)^2$. The parameter σ_F is known, being defined by the Gaussian distribution used to model the foreground intensity distribution during the segmentation measurement. The variance of the imaging noise σ_ν can be estimated prior to visual tracking, see [12] and references therein. Given knowledge of the quantitites σ_ν and σ_F, centering of the measurement noise is done by adding the constant factor $\frac{1}{2}\left(\frac{\sigma_\nu}{\sigma_F}\right)^2$ to the measurements in the log-space, which corresponds to multiplying the measured segmentation probabilities by $e^{\frac{1}{2}\left(\frac{\sigma_\nu}{\sigma_F}\right)^2}$. The foreground measurement noise is now expressed as:

$$\eta_F = e^{\frac{1}{2}\left(\frac{\sigma_\nu}{\sigma_F}\right)^2} \cdot e^{-\frac{1}{2}\left(\frac{\nu(r)}{\sigma_F}\right)^2} \cdot e^{-\left(\frac{\nu(r)\,(I(r)-\mu_F)}{\sigma_F^2}\right)}.$$

Using the independence of the imaging noise ν from the image I, the correlation between the foreground likelihood and measurement noise vanishes:

$$S = E\left(\log\left(P_F\right) \cdot \log\left(\eta_F\right)\right) = 0.$$

Similarly, the second order moment of the foreground measurement noise is given by:

$$\begin{aligned} R &= E\left(\left[\log(\eta_F)\right]^2\right) \\ &= \frac{1}{4}\left(\frac{\sigma_\nu}{\sigma_F}\right)^4 + \frac{3}{4}\left(\frac{\sigma_\nu}{\sigma_F}\right)^4 \\ &+ E\left(\left(\frac{\nu(r)}{\sigma_F}\right)^2\right)\cdot E\left(\left(\frac{I(r)-\mu_F}{\sigma_F}\right)^2\right) \\ &- \frac{1}{2}\left(\frac{\sigma_\nu}{\sigma_F}\right)^2 E\left(\left(\frac{\nu(r)}{\sigma_F}\right)^2\right). \end{aligned}$$

Since $\frac{I(r)-\mu_F}{\sigma_F}$ and $\nu(r)$ follow normal distributions $\mathcal{N}(0,1)$ and $\mathcal{N}(0,\sigma_\nu^2)$ respectively, we obtain:

$$R = \frac{1}{2}\left(\frac{\sigma_\nu}{\sigma_F}\right)^4 + \left(\frac{\sigma_\nu}{\sigma_F}\right)^2.$$

The moment R is a function of the ratio $\frac{\sigma_\nu}{\sigma_F}$. A similar analysis is valid for the background measurement noise.

D. Gain computation

Assume a static prediction model with multiplicative process noise τ, i.e. ($\hat{\rho}_k^- = \hat{\rho}_{k-1}^+ \cdot \tau_k$) with $\rho \in \{P_F, P_B\}$. The objective is to find the optimal value of the gain K that minimizes the mean squared logarithmic error $E\left(\left[\log(\rho) - \log(\hat{\rho}^+)\right]^2\right)$. This is a measure of the accuracy of the estimate $\hat{\rho}^+$, it is denoted $\hat{P}^+$:

$$\begin{aligned} \hat{P}^+ &= E\left(\left[\log(\rho) - \log(\hat{\rho}^+)\right]^2\right) \\ &= E\left(\log(\rho) - \log(\hat{\rho}^-) - K\cdot\left[\log(\zeta) - \log(\hat{\rho}^-)\right]^2\right). \end{aligned}$$

Given the setup, in the log-space associated to the likelihoods, the problem considered is one of point-wise linear filtering for a system facing additive noise. Thus, the optimal selection of the gain K is given by the Kalman gain [16]: $K = \hat{P}^-(\hat{P}^- + R)^{-1}$.

E. Filtering equations

In addition to computing the gain, the error variance P needs to be estimated at the prediction step and updated at the correction step. These updates are computed using the state estimate update:

$$\begin{aligned} \hat{P}^+ &= E\left(\left[\log(\rho) - \log(\hat{\rho}^+)\right]^2\right) \\ &= E\left(\left[\log(\rho) - (1-K)\log(\hat{\rho}^-) - K\log(\zeta)\right]^2\right) \\ &= E\left(\left[(1-K)\left[\log(\rho) - \log(\hat{\rho}^-)\right] - K\log(\eta)\right]^2\right) \\ &= (1-K)^2\ \hat{P}^- + K^2\, R. \end{aligned}$$

Assume the process noise τ to be independent from both the process and the observation noise. The predicted covariance $\hat{P}^-$ is then given by:

$$\begin{aligned} P_k^- &= E\left(\left[\log(\rho_k) - \log(\hat{\rho}_k^-)\right]^2\right) \\ &= E\left(\left[\log(\rho_{k-1}) - \log(\hat{\rho}_{k-1}^+) - \log(\tau_k)\right]^2\right) \\ &= E\left(\left[\log(\rho_{k-1}) - \log(\hat{\rho}_{k-1}^+)\right]^2\right) + E\left(\left[\log(\tau_k)\right]^2\right) \\ &= P_{k-1}^+ + Q, \end{aligned}$$

where $Q = E\left(\left[\log(\tau)\right]^2\right)$. These prediction and update calculations complete the derivation of the filtering equations for the system (see Table I).

TABLE I
FILTERING EQUATIONS FOR VISUAL TRACKING

Prediction	$\begin{cases} \hat{\rho}_k^- = \hat{\rho}_{k-1}^+ \\ \hat{P}_k^- = \hat{P}_{k-1}^+ + Q \end{cases}$
Update	$\begin{cases} K_k = \hat{P}_k^-(\hat{P}_k^- + R)^{-1} \\ \hat{\rho}_k^+ = \left(\hat{\rho}_k^-\right)^{1-K_k}\cdot(\zeta_k)^{K_k} \\ \hat{P}_k^+ = (1-K_k)^2\ \hat{P}_k^- + K_k^2\, R \end{cases}$

IV. Algorithm and Implementation

Based on the description of the design, the optimal estimation algorithm can be summarized as follows:

- Estimate the additive imaging noise prior to the visual tracking process.
- For every pixel, run two filters to estimate the foreground and background likelihoods ($\hat{P}_F(r)$ and $\hat{P}_B(r)$):
  1) at the prediction step, run the corresponding equations in Table I to obtain the predictions.
  2) obtain a measurement by taking the classification likelihood given by Bayesian segmentation on the current image.
  3) at the update step, run the corresponding equations in Table I to obtain the updates.
- The estimated classification probability field is obtained by normalizing the likelihood estimates previously obtained: $\frac{\hat{P}_F}{\hat{P}_F+\hat{P}_B}$. The 50% contour of this probability field defines the bounding contour of the target.

The implementation of the algorithm follows the algorithmic steps just described. Given the typically small size of the target relatively to the image dimensions, windowing can be used in order to speed up the technique. In that case, a localization procedure [1], [4], [18] should be applied prior to performing segmentation measurements. Such a localization procedure guarantees that the prediction and measurement are described in a consistent coordinate frame, i.e. prediction and measurement are aligned.

V. Experiments and Results

This section describes the experiments used to test the validity of the estimator design. Manual segmentation is performed on the sequences of images to provide ground truth. To assess the performance, the number of misclassified pixels (NMP) is used as a quantitative metric.

1) Optimality: This experiment is designed to verify the optimality of the gain. From an original high-SNR infra-red sequence of images, we generated multiple noise-corrupted sequences. The noise variance σ_ν of the additive imaging noise $\mathcal{N}(0, \sigma_\nu^2)$ is controlled to vary between 25 and 100. Figure 1(a-c) depicts a sample image from the original sequence and the corrupted sample image at noise levels $\sigma_\nu = 25$ and $\sigma_\nu = 100$ respectively. All sequences are then tracked using the geometric filtering method with constant gains. Subsequently, the optimal filtering method is applied to the sequences. For these experiments and those following, we used the value $Q = 0.3$ for the prediction noise covariance. The results are reported in Figure 1; they show that the best performance is indeed obtained when the optimal gain is used. In presence of severe noise, fixed low-gain filtering strategies have closer performance to the optimal estimator and vice-versa.

2) Comparative performance: Tracking experiments were also conducted to compare the performance of the estimator with other standard tracking techniques. We used the Bayesian segmentation [6], an active contour tracking technique [15] and the filtering method described in [7]. In the following, these methods are labelled respectively as Bayesian, AC and Deformotion filter. The gain parameters of the Deformotion filter were chosen to the best of our understanding. Similarly, the smoothing term of the active contour was chosen to provide the best segmentation possible. All tracking techniques were applied to one noise-corrupted IR sequence ($\sigma_\nu = 50$) and the results obtained were compared to the optimal estimator using the NMP. For this test sequence, the foreground and background were modelled using the respective distributions $\mathcal{N}(\mu_F, \sigma_F^2)$ and $\mathcal{N}(\mu_B, \sigma_B^2)$ with $(\mu_F, \sigma_F) = (202, 68)$ and $(\mu_B, \sigma_B) = (103, 85)$. The parameter σ_ν is equal to 50. The experiment was repeated with a real-life noisy aquarium sequence. A similar modelling was used but with the set of parameters $(\mu_F, \sigma_F) = (30, 14)$ and $(\mu_B, \sigma_B) = (68, 11)$. The parameter σ_ν is estimated to 25. The results are depicted in Figure 2. They clearly indicate that the optimal filtering strategy is a competitive tracking technique.

VI. Conclusion

This paper presented the derivation of an optimal estimator for online visual contour tracking. In contrast to the prevailing methods using a top-down approach, we employ a bottom-up approach starting from the measurement strategy. In this framework, filtering on the infinite-dimensional space of closed curves is reduced to a series of point-wise filtering problems. The optimal gain derivation is formally tied to quantitative uncertainty levels of the image data and, therefore, does not require manual gain tuning. The resulting optimal estimator is able to handle severe noise perturbations, and compares favorably with other estimation-based tracking techniques.

Future work will seek to extend this analysis to color sequences and enforce spatial consistency through the use of distributed filtering methods [13].

References

[1] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–577, 2003.

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

[3] S. Dambreville, Y. Rathi, and A. Tannenbaum. Tracking deformable objects with unscented Kalman filtering and geometric active contours. In *Proceedings of the IEEE American Control Conference*, pages 2856–2861, 2006.

[4] A. Elgammal, R. Duraiswami, and L. S. Davis. Probabilistic tracking in joint feature-spatial spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 781–788, 2003.

[5] D. Freedman and T. Zhang. Active contours for tracking distributions. *IEEE Transactions on Image Processing*, 13(4):518–526, 2004.

[6] S. Haker, G. Sapiro, A. Tannenbaum, and D. Washburn. Missile tracking using knowledge-based adaptive thresholding: Tracking of high speed projectiles. In *Proceedings of the IEEE International Conference on Image Processing*, pages 786–789, 2001.

[7] J.D. Jackson, A.J. Yezzi, and S. Soatto. Tracking deformable moving objects under severe occlusions. In *Proceedings of the IEEE Conference on Decision and Control*, pages 2990–2995, 2004.

[8] 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.

[9] I.J. Ndiour and P.A. Vela. Towards a local Kalman filter for visual tracking. In *Proceedings of the IEEE Conference on Decision and Control*, pages 2420–2426, 2009.

[10] 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.

[11] N. Papadakis and E. Mémin. Variational optimal control technique for the tracking of deformable objects. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1–7, 2007.

[12] K. Rank, M. Lendl, and R. Unbehauen. Estimation of image noise variance. In *Proceedings of the IEE Vision, Image and Signal Processing*, pages 80–84, 1999.

[13] R.S. Rao and H.F. Durrant-Whyte. Fully decentralised algorithm for multisensor Kalman filtering. *Proceedings of the IEE Control Theory and Applications*, 138(5):413–420, 1991.

[14] Y. Rathi, N. Vaswani, and A. Tannenbaum. A generic framework for tracking using particle filter with dynamic shape prior. *IEEE Transactions on Image Processing*, 16(5):1370–1382, 2007.

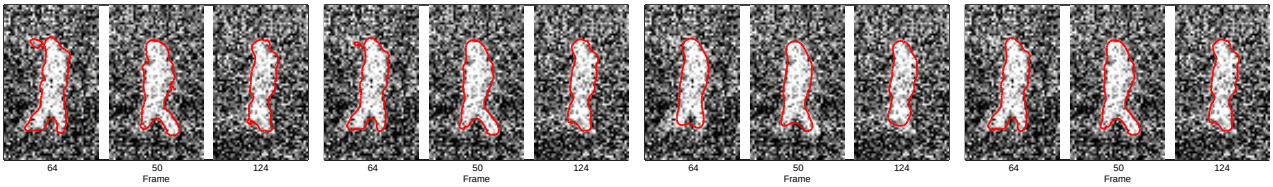
[15] M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. In *Proceedings of the IEEE Workshop on Motion and Video Computing*, pages 56–61, 2002.

[16] D. Simon. *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*. Wiley and Sons, 2006.

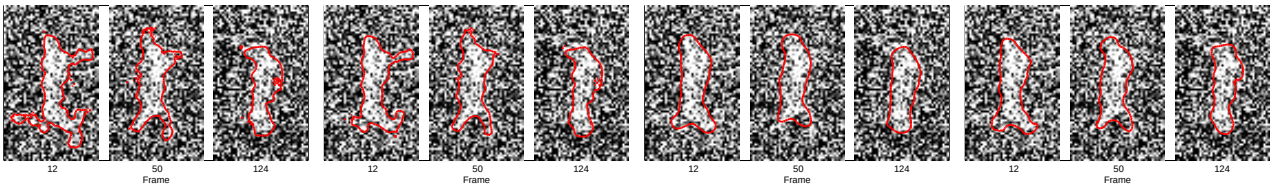
[17] G. Sundaramoorthi, J.D. Jackson, A.J. Yezzi, and A.C. Mennucci. Tracking with Sobolev active contours. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 674–680, 2006.

[18] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Comput. Surv.*, 38(4):13, 2006.

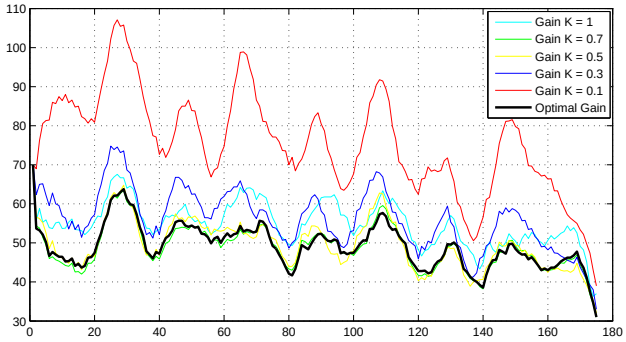
(a) Original IR sequence (b) Noise-corrupted IR ($\sigma_\nu = 25$) (c) Noise-corrupted IR ($\sigma_\nu = 100$)



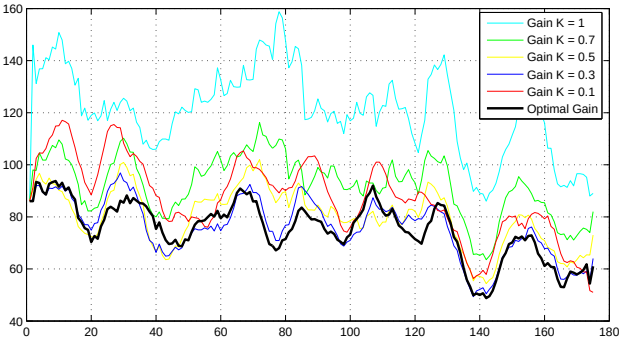
(d) $K = 1,\ \ \sigma_\nu = 25$ (e) $K = 0.7,\ \ \sigma_\nu = 25$ (f) $K = 0.1,\ \ \sigma_\nu = 25$ (g) $K = optimal,\ \ \sigma_\nu = 25$



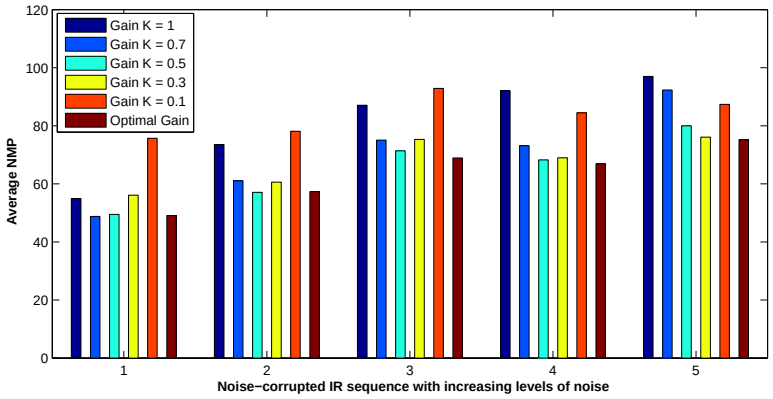
(h) $K = 1,\ \ \sigma_\nu = 100$ (i) $K = 0.7,\ \ \sigma_\nu = 100$ (j) $K = 0.1,\ \ \sigma_\nu = 100$ (k) $K = optimal,\ \ \sigma_\nu = 100$



(l) NMP error vs. time for a given noise level ($\sigma_\nu = 25$)

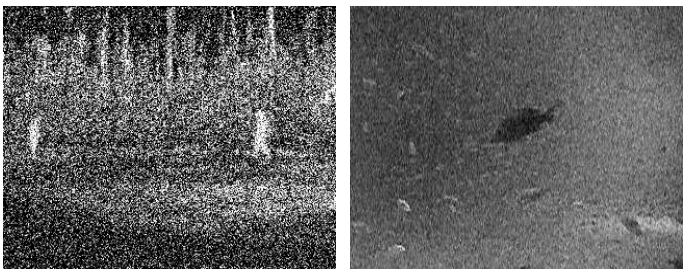


(m) NMP error vs. time for a given noise level ($\sigma_\nu = 100$)

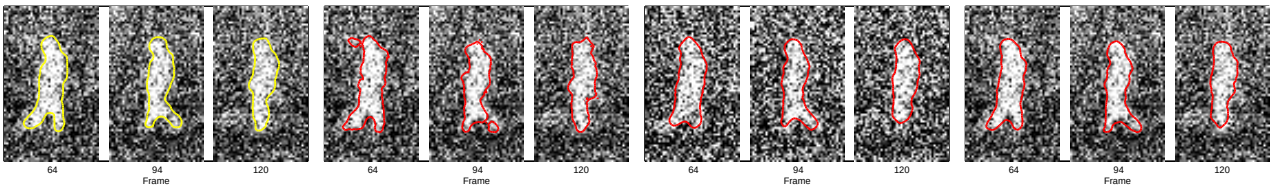


(n) Average NMP error vs. noise level

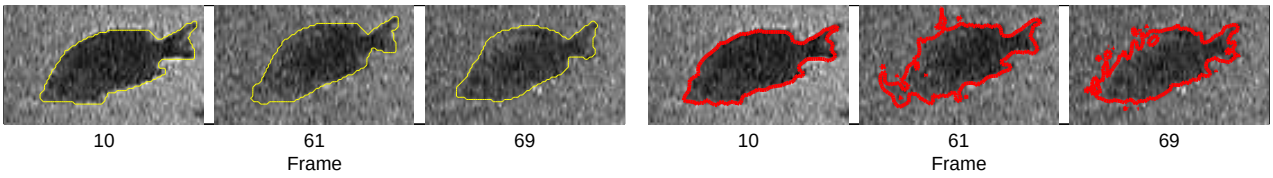
Fig. 1. Quantitative assessment of optimality. On top, figures depict corruption of the original sample with two levels of noise. The next two rows show sample estimates obtained using the filtering scheme with predetermined values of the gain and the optimal gain. Row # 4 compares the number of misclassified pixels for predetermined choices of the gain with the optimal filtering strategy. The bottom figure displays the mean number of misclassified pixels (NMP) for different levels of noise and different configurations of the gain.



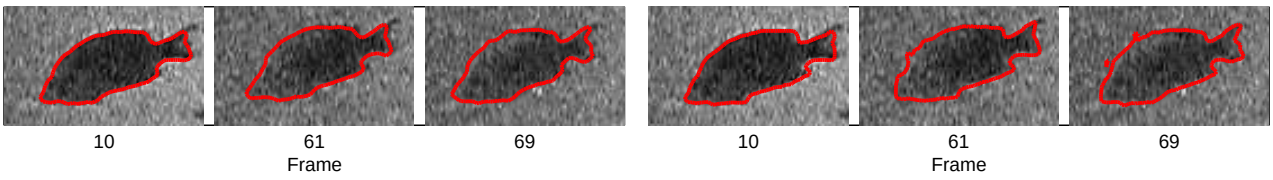
(a) IR sequence (b) Fish sequence



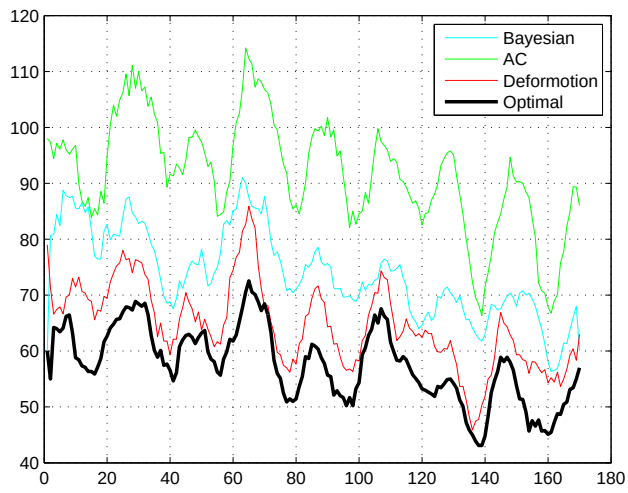
(c) Ground truth (d) Bayesian segmentation (e) Deformotion filter (f) Optimal estimator



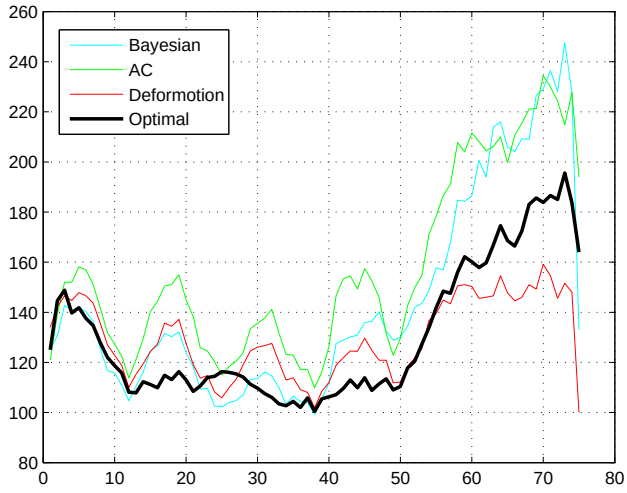
(g) Ground truth (h) Bayesian segmentation



(i) Deformotion filter (j) Optimal estimator



(k) NMP error vs. time for test algorithms on IR sequence



(l) NMP error vs. time for test algorithms on fish sequence

Fig. 2. Quantitative assessment of performance. On top, figures depict samples from the two test sequences. The next three rows show sample estimates obtained using the optimal filtering scheme and other tracking techniques (active contour estimates are similar to the Bayesian segmentation estimates). The bottom figures compare the performances of each technique using a quantitative metric (the number of misclassified pixels).