Catheter Tracking via Online Learning for Dynamic Motion Compensation in Transcatheter Aortic Valve Implantation

Peng Wang[1], Yefeng Zheng[1], Matthias John[2], Dorin Comaniciu[1]

1. Corporate Research and Technology, Siemens Corporation, Princeton NJ, U.S.A.
2. Siemens AG, Healthcare Sector, Siemensstr. 1, Forchheim, Germany

Abstract. Dynamic overlay of 3D models onto 2D X-ray images has important applications in image guided interventions. In this paper, we present a novel catheter tracking for motion compensation in the Transcatheter Aortic Valve Implantation(TAVI). To address such challenges as catheter shape and appearance changes, occlusions, and distractions from cluttered backgrounds, we present an adaptive linear discriminant learning method to build a measurement model online to distinguish catheters from background. An analytic solution is developed to effectively and efficiently update the discriminant model and to minimize the classification errors between the tracking object and backgrounds. The online learned discriminant model is further combined with an offline learned detector and robust template matching in a Bayesian tracking framework. Quantitative evaluations demonstrate the advantages of this method over current state-of-the-art tracking methods in tracking catheters for clinical applications.

1 Introduction

Catheter tracking has found important applications in image guided interventions [2, 5]. In this work, we present the work of tracking pigtail for dynamic motion compensation in the transcatheter aorta valve implantation (TAVI). A pigtail catheter is a type of medical device that is inserted into the aorta or heart chamber to inject dyes in TAVI. Tracking the motion of a pigtail catheter tip, which is the loose circle at a catheter's distal end, can provide real-time motion information for dynamic model overlay in TAVI [5]. Fig. 1.(a) shows the dynamic overlay of an aorta model (i.e., the solid contour), based on tracked pigtail catheter positions. Compared with the static overlay (i.e., the dotted contour), which does not move with breathing and cardiac motions, the dynamic overlay can update clinicians the position of aorta valves in real-time, therefore could help clinicians achieve high precision in TAVI.

Also shown in Fig. 1. (b)(c)(d), the dynamic and complex clinical environment in TAVI presents great challenges to the catheter tracking. When X-ray images are acquired at arbitrary angles under continuous breathing and cardiac motions, a catheter may show a shape of a circle, an ellipse, or even a straight
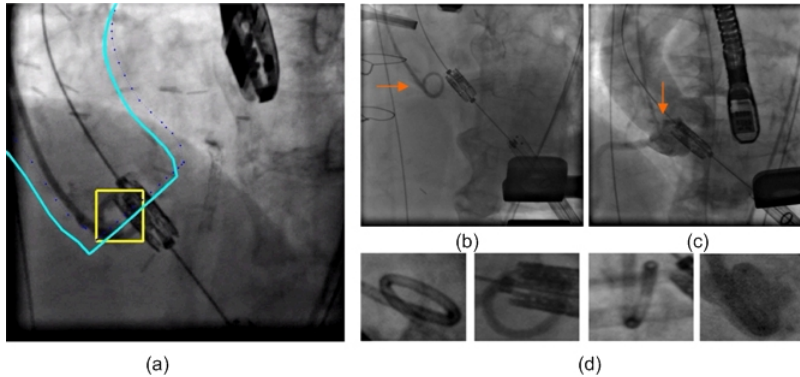
Fig. 1. (a): The dotted contour is a static aorta model projected on 2D, and the solid contour is a dynamically overlayed model based on the tracked catheter (yellow rectangle) ; (b)(c): pigtail catheters in X-ray images; (d): close-up images.

line in a 2D image. Moreover, there also exist in X-ray images many devices, such as other catheters, stents and probes, which could occlude a large part of the pigtail catheter. In past years, there have been work on medical device detection and tracking [2, 5]. In [2], an ellipsoid model is used to fit and tracking a lasso catheter. In [5], a SSD (sum of squared differences) based catheter tracking is used. However, only 10 sequences have been tested in the experiments.

To address aforementioned challenges, we present in this paper an online discriminant learning method to adapt the tracking method to dynamic X-ray images. The contribution of this method is that a closed-form analytical solution is developed to efficiently update a discriminant function directly from online images without updating sample co-variance matrices and subspaces as in previous methods [6, 8]. The solution is effective and efficient in that it directly minimizes the classification error between tracking objects and backgrounds. Compared with existing online learning work [3, 6], our online discriminant learning has a probabilistic formalization that allows us to develop an efficient strategy to adaptively update the discriminant model from online data, instead of selecting from feature candidates [3] or updating scatter matrices and solving eigenvectors as in [6, 8]. To handle dynamic changes of pigtail catheters, we further introduce a Bayesian based fusion of multiple measurement models, including the online discriminant model, an offline learned object detection model, and a robust template matching model.

We validate the presented framework on a set of clinical data containing 198 sequences with totally more than 14,000 frames that are captured in several hospitals during cardiac surgeries. Quantitative evaluations and comparison with the state of the art methods [4, 8, 1] demonstrate that the presented adaptive learning and measurement fusion significantly improve the accuracy and robustness of tracking.

2 Online Discriminant Learning for Catheter Tracking

In this section, we introduce an online discriminant learning method to separate catheters from backgrounds and to handle dynamic environments that could change from one sequence to another. We denote the catheter class as the "positive" class Ω_+ and the background as the "negative" class Ω_-. An observed image is denoted as $\mathbf{Z}$. A data vector extracted from an image patch is denoted as $\mathbf{x}$, and its associated class label is $\ell_{\mathbf{x}}$. $\ell_{\mathbf{x}} = 1$ indicates that $\mathbf{x}$ belongs to the catheter class, while $\ell_{\mathbf{x}} = -1$ indicates that it is from the background.

2.1 Probabilistic Linear Discriminant Analysis

There are many statistical learning methods such as SVM, AdaBoost, k-NN, and Neural Networks. For the consideration of simplicity and computational efficiency, the linear discriminant analysis is used. In the linear discriminant analysis, an original data vector is projected to a subspace of lower dimensionality where objects can be separated from the background. The projection is represented as $\mathbf{y} = \Phi^T \mathbf{x}$, where Φ is the linear transformation vector (or a matrix). Fisher discriminant analysis (FDA) is a commonly used linear discriminant analysis method, which maximizes the ratio of between-class and within-class variance. FDA can be solved as a generalized eigenvector problem:

$$(\Sigma_w^{-1} \Sigma_b)\Phi_k = \lambda_k \Phi_k. \tag{1}$$

where Σ_b and Σ_w are between- and within-class scatter matrices, respectively. The transformation vector Φ_k is the eigenvector corresponding to the k-th eigenvalue. For a two-class problem, only one effective eigenvector can be extracted from FDA. The FDA transformation vector is denoted as $\mathbf{A}_f$. Usually PCA is applied before FDA to reduce the data dimensionality. The overall linear discriminant analysis is the combination of PCA and FDA, as Eqn. (2):

$$y_{\mathbf{x}} = \mathbf{A}_f^T \mathbf{A}_p^T \mathbf{x} = \mathbf{A}^T \mathbf{x}, \tag{2}$$

where $\mathbf{A}_p$ is the PCA eigenspace projection matrix, and $\mathbf{A} = \mathbf{A}_p \mathbf{A}_f$ is the overall linear discriminant vector. After learning the linear discriminant vector $\mathbf{A}$, a sigmoid function is used to model the posterior probability for the two-class classification:

$$P_{\mathbf{A}}(\mathbf{x}) = P(\ell_{\mathbf{x}} = 1|\mathbf{x}) = \frac{1}{1 + e^{(ay_{\mathbf{x}}+b)}} = \frac{1}{1 + e^{(a\mathbf{A}^T\mathbf{x}+b)}} \tag{3}$$

The parameters a and b can be learned from training samples by regression.

2.2 Online Discriminant Learning

The probabilistic linear discriminant can be learned offline from collected training samples. However, the background, and sometimes the object itself, could
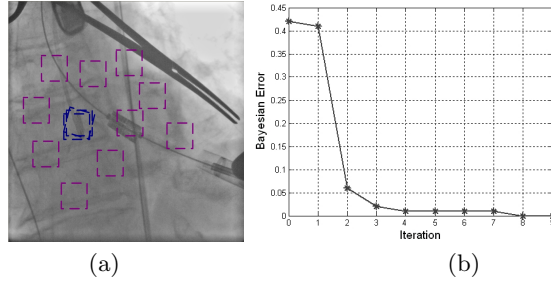
(a) (b)

Fig. 2. Online discriminant learning. (a): positive samples (blue rectangles) and negative samples (purple rectangles) are extracted online; (b): the convergence of online learning at a frame.

change from one tracking scenario to another, and may not be seen in training sets, therefore not available for offline learning. In dynamic environments, the discriminant learning needs to adapt to different tracking scenarios. For this purpose, we present an online method to efficiently and effectively update the discriminant vector $\mathbf{A}$.

Most statistical learning algorithm aims at the minimization of the Bayesian error, i.e., $E_f = \int (1 - p(\ell_\mathbf{x}|\mathbf{x}))p(\mathbf{x})\mathbf{dx}$. In the adaptive discriminant learning, we also try to minimize the Bayesian error, which is equivalent to minimizing the object function $C(\mathbf{A})$ in Eqn. (4):

$$C(\mathbf{A}) = -[\sum_{\mathbf{x}_i \in \Omega_+} P(\mathbf{x}_i)P_\mathbf{A}(\mathbf{x}_i) - \sum_{\mathbf{x}_j \in \Omega_-} P(\mathbf{x}_j)P_\mathbf{A}(\mathbf{x}_j)] = \sum_\mathbf{x} \frac{-\ell_\mathbf{x} P(\mathbf{x})}{1 + e^{(a\mathbf{A}^t\mathbf{x}+b)}}. \tag{4}$$

With new online data available during tracking, to re-train the discriminant analysis needs to access a large number of training samples, and involves re-computation of co-variance metrics and eigenbases. In this method, we online update the discriminant analysis using a gradient descent method. To minimize the Bayesian error, the discriminant analysis is iteratively updated as Eqn. (5):

$$\mathbf{A}^{(k)} = \mathbf{A}^{(k-1)} - \delta_k \nabla_\mathbf{A} C(\mathbf{A}^{(k-1)}), \tag{5}$$

where δ_k is an updating step at the k-th iteration. The updating continues until it converges. The gradient of the object function $\nabla_\mathbf{A} C(\mathbf{A})$ is given by

$$\nabla_\mathbf{A} C(\mathbf{A}) = \sum_{\mathbf{x}_i} \nabla_\mathbf{A} \left(\frac{-\ell_i P(\mathbf{x}_i)}{1 + e^{(a\mathbf{A}_i^t\mathbf{x}_i+b)}} \right) = \sum_{\mathbf{x}_i} \ell_i P(\mathbf{x}_i) \frac{ae^{(a\mathbf{A}^t\mathbf{x}_i+b)}}{(1 + e^{(a\mathbf{A}^t\mathbf{x}_i+b)})^2} \mathbf{x}_i. \tag{6}$$

After $\mathbf{A}$ is updated, a and b are also updated by regression at every iteration.

The online discriminant needs a good starting point to converge to an optimal solution, although a global optimum is not guaranteed. A starting point can be the discriminant function learned offline from collected samples. Even if the

initial discriminant vector may not fit a current environment, the online learning can quickly converge to a good solution. An example is shown in Fig. 2, where the greedy searching converges within a few iterations to achieve a Bayesian error less than 10% for the initial error of 40%. To improve the learning robustness, we keep a history of tracked objects from previous frames in a pool to prevent the learning from being distracted by tracking errors at individual frames. The Bayesian error estimated as Eqn. (4) is used as a criterion to decide if the learning is converged to an optimal solution. For computational efficiency, the online updating of discriminant vector can be performed once every several frames.

3 Bayesian Catheter Tracking Framework

The tracking is to infer unknown object states $\mathbf{m}_t$, i.e., the catheter position in this paper, from an observed image sequence $\mathbf{Z}_{1:t}$. We formalize the parameter inference in a sequential Bayesian inference framework. Assuming a commonly used Markov property for tracking, the posterior probability $P(\mathbf{m}_t|\mathbf{Z}_{1:t})$ is given in Eqn. (7).

$$\begin{aligned} P(\mathbf{m}_t|\mathbf{Z}_{1:t-1}) &= \int P(\mathbf{m}_t|\mathbf{m}_{t-1})P(\mathbf{m}_{t-1}|\mathbf{Z}_{1:t-1})d\mathbf{m}_{t-1} \\ P(\mathbf{m}_t|\mathbf{Z}_{1:t}) &\propto P(\mathbf{m}_t|\mathbf{Z}_{1:t-1})P(\mathbf{Z}_t|\mathbf{m}_t) \end{aligned} \tag{7}$$

The tracking result is the motion parameter corresponding to the maximal posterior probability, i.e, $\hat{\mathbf{m}}_t = \arg\max_{\mathbf{m}_t} P(\mathbf{m}_t|\mathbf{Z}_{1:t})$.

In Eqn. (7), $P(\mathbf{m}_t|\mathbf{m}_{t-1})$ is a dynamic model. We model the dynamic probability as a Gaussian model. The likelihood measurement model $P(\mathbf{Z}_t|\mathbf{m}_t)$ is a fusion of multiple measurements. Assuming that there are K measurement models for an object, i.e., $P^{(k)}(\mathbf{Z}_t|\mathbf{m}_t) = P(\mathbf{Z}_t|\mathbf{m}_t, k), k = 1, ..., K$, a measurement fusion based on the probability marginalization is given in Eqn. (8):

$$P(\mathbf{Z}_t|\mathbf{m}_t) = \sum_k P(\mathbf{Z}_t|\mathbf{m}_t, k)P(k|\mathbf{m}_t) \tag{8}$$

where $P(k|\mathbf{m}_t)$ is the weight for the k-th measurement model.

The online learned discriminant model acts as one measurement model, denoted as $P(\mathbf{Z}_t|\mathbf{m}_t, k = 1) = P_{\mathbf{A}}(\mathbf{x}_{\mathbf{m}_t})$ as in Eqn. (3), where $\mathbf{x}_{\mathbf{m}_t}$ is the data vector extracted from observed images based on the object state $\mathbf{m}_t$. The second measurement model used is an offline learned object detector. A probabilistic boosting tree (PBT) [9] is trained with Haar features to build the catheter detector offline. The probabilistic interpolation of an AdaBoost is used to provide the second measurement model $P(\mathbf{Z}_t|\mathbf{m}_t, k = 2)$. A robust template matching method is used as the third measurement mode in our method. The template matching takes the form in Eqn. (9):

$$P(\mathbf{Z}_t|\mathbf{m}_t, k = 3) \propto G(D(\mathbf{x}_{\mathbf{m}_t}); \sigma_a), \tag{9}$$

where G is a one-dimensional Gaussian kernel with a zero mean and a bandwidth of σ_a. $D(\mathbf{x_{m_t}})$ is a robust distance function [7] to compute the differences between current observations $\mathbf{x_{m_t}}$ and a catheter appearance template, which is updated from the tracking result at a previous frame.

4   Experiments

A set of clinical data has been acquired from TAVI procedures in several hospitals to evaluate the catheter tracking. There are totally 198 sequences acquired at the frame rate between 15fps and 30fps, and more than 14,000 frames in the data set. Each image pixel represents a physical size between 0.154 mm and 0.308 mm. The data set well represents the real surgical scenarios, includes images of poor quality, occlusions, dye injections, and motion blurs. Some exemplar frames in the data set and corresponding tracking results are displayed in Fig. 3. To establish ground truth for evaluation, we manually annotate the pigtail catheters in all the frames. The annotation at the first frame of each sequence is used to initialize the tracking, and the annotations at the remaining frames are used for evaluations. The whole system runs at more than 10 frames per second at an Intel Quad Core 2.5GHz CPU.

In our quantitative evaluation, we measure the **tracking precision** as the Euclidean distance between tracked catheters and the ground truth. The **normalized error** is the distance error divided by corresponding pigtail catheter size (the maximum of the width and height of a catheter). The **tracking success rate** is calculated to measure the percentage of successful tracking. The **frame success rate** is defined as the percentage of the frames whose normalized distance error is less than a threshold (e.g., set as 0.35 in our experiments). The **sequence success rate** is defined as the percentage of sequences where more than a certain percentage of frames (e.g., set as 70% in our experiments) have been successfully tracked. For an unbiased evaluation, we perform a **4-fold cross-validation** to evaluate the tracking accuracy. During evaluation, the algorithm parameters are kept the same for all the sequences, so the method is not over-fitted to individual cases.

Some parameters of our method, such as covariance matrices in Gaussian distributions, can be learned. A few other parameters, such as the weights in Eqn.(8), need to be set empirically. Through our experiments, we found that the accuracy of this method is consistent within a range of parameter settings. Due to limited space, only the accuracy of a single set of parameters (i.e., the weights are set as 0.35, 0.25 and 0.4 for the template based, online learned and offline learned measurement models, respectively) is reported in Table 1.

The quantitative evaluation results in Table 1 show that the online models perform better than offline learned detectors, i.e., 84.1% vs. 82.9% of frame success rate and 79.4% vs 73.9% of sequence success rate. When combining all the measurement models together, this method achieves a 91.0% frame success rate and a 91.9% sequence success rate, respectively. The tracking precision, e.g. the mean error of 1.10mm, further confirms the performance improvement with
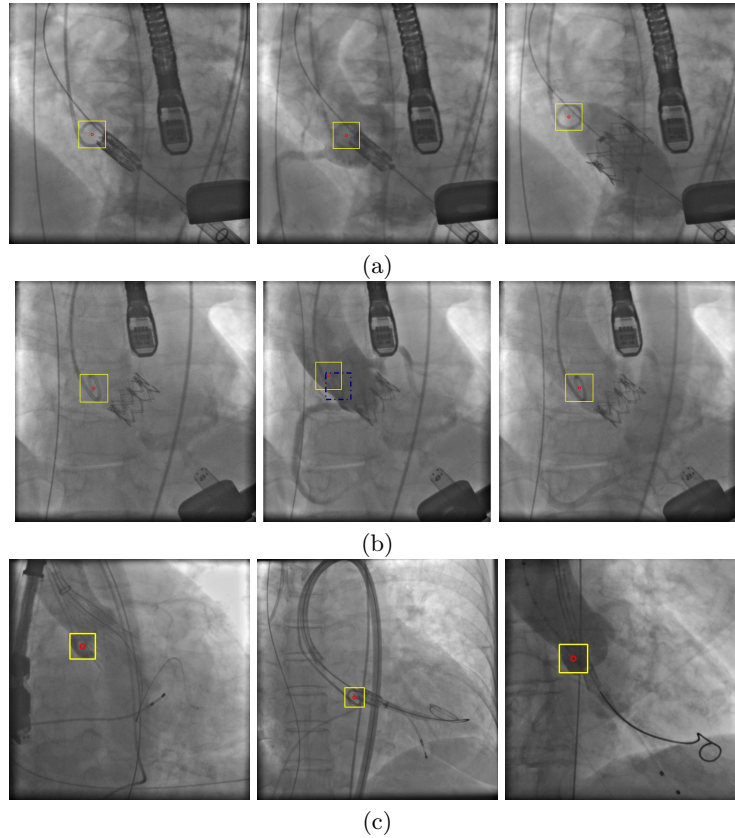
Fig. 3. Exemplar tracking results: (a) handling occlusion and dye injection; (b) recovering from dye injection. The blue rectangle in the middle image shows the ground truth;(c) more tracking results. Each image corresponds to one sequence.

this method. To understand the generalization capability of the method, we also train an object detector with all the annotations and apply it to the tracking. The evaluation, named as "full training", shows that the tracking success rates only drop by 3% with the cross-validation. The tracking failures of our methods are mainly caused by contrast injection and image blurs. In practice, some of tracking failures won't affect the workflow if the tracking resumes correctly after contrast injection, as shown in the example of Fig. 3 (b).

This method is compared with three representative state-of-the-art visual tracking methods: the online boosting method [4], the incremental learning methods [8], and the multiple instant learning method [1], whose implementations are available from authors' homepages. For each algorithm for comparison, we carefully tune the parameters to achieve optimal tracking accuracy. The methods of [4] and [8] can achieve $\sim 65\%$ frame success rate. The multiple instant learning

Table 1. Quantitative evaluation on 198 sequences.

This method	Tracking success rate		Tracking precision (mean errors in mm/normalized errors)	
	Frame	Sequence	median	mean
Using only offline detector(cross-validation)	82.9%	73.9%	1.37/0.08	5.12/0.27
Using only online models	84.1%	79.4%	1.38/0.08	3.23/0.18
Bayesian fusion (cross-validation)	91.0%	91.9%	1.10/0.06	3.19/0.17
Bayesian fusion (full training)	93.3%	94.9%	0.98/0.05	2.43/0.13
Existing methods				
Grabner et. al. [4]	64.4%	52.7%	3.15/0.21	4.55/0.35
Ross et. al. [8]	65.1%	64.3%	2.91/0.17	5.56/0.40

method [1] can only succeed in less than 10% sequences with a mean error of more than 10mm. The poor performance of existing methods shows the challenges of pigtail catheter tracking in X-ray images, and demonstrates the advantages of our method in handling such a challenging environment.

5   Conclusion

In summary, this paper presented a novel and robust method to track pigtail catheters for dynamic motion compensation in TAVI. The validation on a set of 198 sequences demonstrated the performance for this tracking task.

References

1. Babenko, B., Yang, M.H., Belongie, S.: Visual tracking with online multiple instance learning. In: CVPR (2009)
2. Brost, A., Liao, R., Hornegger, J., Strobel, N.: 3-d respiratory motion compensation during EP procedures by image-based 3-d lasso catheter model generation and tracking. In: MICCAI (2009)
3. Collins, R., Liu, Y., Leordeanu, M.: Online selection of discriminative tracking features. PAMI, IEEE Trans. on 27(10), 1631 –1643 (oct 2005)
4. Grabner, M., Grabner, H., Bischof, H.: Learning features for tracking. CVPR (2007)
5. Karar, M.E., John, M., Holzhey, D., Falk, V., Mohr, F.W., Burgert, O.: Model-updated image-guided minimally invasive off-pump transcatheter aortic valve implantation. In: MICCAI (2011)
6. Lin, R.S., Yang, M.H., Levinson, S.: Object tracking using incremental Fisher discriminant analysis. In: ICPR. vol. 2, pp. 757 – 760 Vol.2 (2004)
7. Matthews, I., Ishikawa, T., Baker, S.: The template update problem. IEEE Trans. on Pattern Analysis and Machine Intelligence 26(6), 810–815 (2004)
8. Ross, D., Lim, J., Lin, R.S., Yang., M.H.: Incremental learning for robust visual tracking. International Journal of Computer Vision Special Issue: Learning for Vision (2007)
9. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: ICCV. pp. 1589–1596 (2005)