Learning-based Hypothesis Fusion for Robust Catheter Tracking in 2D X-ray Fluoroscopy

Wen Wu [*]   Terrence Chen [*]   Adrian Barbu [†]   Peng Wang [*]   Norbert Strobel [‡]
Shaohua Kevin Zhou [*]   Dorin Comaniciu [*]

[*]Image Analytics and Informatics, Siemens Corporate Research, Princeton, NJ 08540, USA
[†]Department of Statistics, Florida State University, Tallahassee, FL 32306, USA
[‡]Siemens AG, Forchheim, Germany

Abstract

Catheter tracking has become more and more important in recent interventional applications. It provides real time navigation for the physicians and can be used to control a motion compensated fluoro overlay reference image for other means of guidance, e.g. involving a 3D anatomical model. Tracking the coronary sinus (CS) catheter is effective to compensate respiratory and cardiac motion for 3D overlay navigation to assist positioning the ablation catheter in Atrial Fibrillation (Afib) treatments. During interventions, the CS catheter performs rapid motion and non-rigid deformation due to the beating heart and respiration. In this paper, we model the CS catheter as a set of electrodes. Novelly designed hypotheses generated by a number of learning-based detectors are fused. Robust hypothesis matching through a Bayesian framework is then used to select the best hypothesis for each frame. As a result, our tracking method achieves very high robustness against challenging scenarios such as low SNR, occlusion, foreshortening, non-rigid deformation, as well as the catheter moving in and out of ROI. Quantitative evaluation has been conducted on a database of 13221 frames from 1073 sequences. Our approach obtains 0.50mm median error and 0.76mm mean error. 97.8% of evaluated data have errors less than 2.00mm. The speed of our tracking algorithm reaches 5 frames-per-second on most data sets. Our approach is not limited to the catheters inside the CS but can be extended to track other types of catheters, such as ablation catheters or circumferential mapping catheters.

1. Introduction

Atrial Fibrillation (Afib) is a rapid, highly irregular heartbeat caused by abnormalities in the electrical signals generated by the atria of the heart. It is the most common
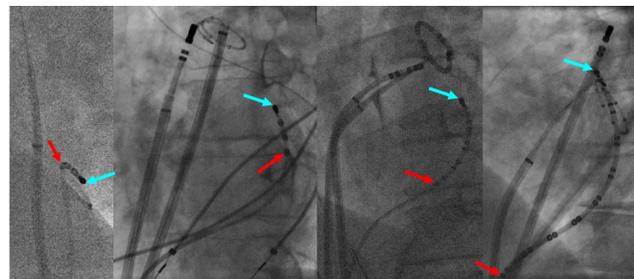

Figure 1. Examples of CS catheters in 2D X-ray fluoroscopy. Catheters demonstrate various appearance and shapes in different contexts. Cyan and red arrows point at the catheter tip and the most proximal electrode (PCS) and in between are other electrodes.

cardiac arrhythmia and involves the two upper chambers (atria) of the heart. Surgical and catheter-based Afib therapies have become common procedures in many major hospitals throughout the world today [4]. One popular treatment is catheter ablation, which modifies the electrical pathways of the heart. To carry out the operation, catheters are inserted and guided to the heart. The entire operation is monitored with real-time fluoroscopic images. The integration of static tomographic volume renderings into three-dimensional catheter tracking systems has introduced an increased need for mapping accuracy during Afib procedures. Current technologies may concentrate on gating catheter positions to a fixed point in time within the cardiac cycle without explicitly taking into account respiration. Clearly, a static positional reference provides only intermediate accuracy in association with ECG gating. For left atrium procedures, a method is known from the literature where a motion-compensated overlay controlled by the most proximal electrode of the CS catheter (PCS) reference was found to be superior to the static reference [8]. Figure 1 shows some examples of CS catheters in 2D X-ray fluoroscopy.

Our goal in this paper is to develop a robust and fast al-

gorithm to track CS catheters to provide accurate real-time information for motion compensation. The task has various challenging characteristics: 1) rapid motion due to cardiac and breathing motion; 2) non-uniform appearance and electrode number; 3) motion variations including catheter foreshortening due to 2D projection, occlusion and non-rigid deformation; 4) diverse background factors including low signal-to-noise ratio (SNR), nearby catheter-like structures and cluttered scenes. State-of-the-art tracking methods [10, 1, 7, 2] may succeed in overcoming some of these challenges, however, our proposed approach is capable to handle these challenges consistently to achieve high performance in tracking CS catheters in continuous 2D X-ray fluoroscopy. Our approach is different from recent work on CS catheter tracking [9] in four aspects: 1) our approach leverages learning-based detectors instead of filter-based blob detectors; 2) our method does not make any assumption of the electrode number and catheter shape; 3) we propose a novel tracking hypothesis generation and evaluation framework; 4) our approach has been extensively evaluated on 1073 sequences while the method in [9] was evaluated on a smaller dataset. Our paper is also different from [14] in terms of problems, proposed methods and contribution. In [14], a wire structure is tracked as a spline curve, which tries to deform current wire to fit the next frame. Our method detects catheter electrodes and tips to efficiently generate tracking hypotheses. A comparison is possible but it would not be fair since the targeted problems are different.

Learning-based methods have demonstrated their strong capabilities to effectively explore object content and context in numerous applications such as segmentation, detection and tracking [13, 12, 3, 14]. In our work, discriminative models are learned based on appearance and contextual features of CS catheter tip and electrodes. Our approach automatically builds the catheter model by analyzing the catheter shape and electrode number from user initialization. In each frame, we first perform catheter tip and electrode detection, and then apply proposed novel schemes to generate tracking hypotheses that are further evaluated by a Bayesian framework. A block diagram of our proposed approach is shown in Figure 2.

2. Learning-based Hypothesis Fusion

In this paper, we represent the CS catheter as an ordered set of electrodes starting from the tip. Only the most proximal electrode (PCS) is important for motion compensation [8]. However, by tracking the whole catheter the proposed approach is capable of fusing more information and obtaining more reliable tracking than by just tracking the PCS.

The non-rigid nature of the CS catheter means that its motion has to be represented in a high dimensional space. Tracking the CS catheter means finding in each frame the location of each electrode. An exhaustive search of the de-
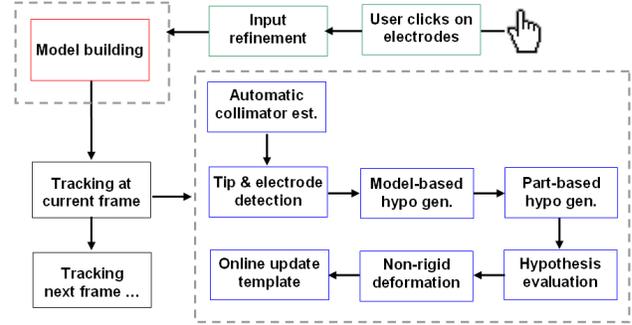

Figure 2. A block diagram of our proposed approach.

formation parameters is not only computationally expensive but also prone to producing false positive matches on other catheters or catheter-like structures present in the image. Even bounding the search range relative to the CS position from the previous frame still leads to a large search space for tracking CS catheters. Furthermore, if the CS position from the previous frame is not accurate, drifting can happen.

To tackle the problem, we propose a novel approach that uses a low dimensional representation that approximates the CS catheter shape with a small error. Then a number of shape hypotheses are generated for finding the CS in the current frame, each shape implicitly containing the template deformation. The hypotheses are then evaluated by information fusion in a Bayesian framework.

The clinical requirements are that the CS catheter is manually initialized by the user in the first frame by marking the electrodes, in the order from the tip to the PCS. The input positions are then refined by local search using trained electrode and tip detectors. In the proposed approach, during the *model building* stage the tracking strategy is selected based on catheter shape and the number of electrodes, with a simpler strategy for shorter catheters with fewer electrodes. The catheter template is then initialized based on the catheter shape and appearance. *Tracking at each frame* consists of a number of steps: automatic collimator estimation by a border detector, learning-based tip and electrode detection, hypothesis generation including model-based and part-based schemes, hypothesis evaluation in a Bayesian formula, non-rigid model deformation and online template update. These steps are described in more detail in the following sections.

Notations are as follow. Z is to denote image observation, D for image intensity data, C for a catheter electrode set. Subscript t denotes t-th frame. Assume that there are K electrodes, $\{e^i, i = 1, ..., K\}$ on the catheter and e^1 and e^K represent the tip and the PCS. We model an electrode as an oriented point as $e^i = [p^i, \theta^i]$ where $p^i = [x^i, y^i]$ is the 2D electrode center and θ^i is electrode orientation which is defined as the catheter curve tangent pointing to the tip. Other catheter body points can be interpolated from the electrode points as $C(b) = \{b = (\gamma^x(\omega), \gamma^y(\omega)), 1 \le \omega \le K\}$ and

```
Data: Initial catheter electrode positions:
        C_0 = {e_0^1, ..., e_0^K}
Result: Tracking strategy for the target catheter
if K ≤ B^1 then
    use a one segment approximation;
end
if B^1 < K ≤ B^2 then
    Find the point with maximal curvature in C_0 and
    approximate C_0 with two segments C_0^1 and C_0^2
    joined at the maximum curvature point;
end
if K > B^2 then
    Find two points with maximal curvature in C_0 and
    approximate C_0 with 3 segments C_0^1, C_0^2 and C_0^3.
end
```

Algorithm 1: Catheter-specific tracking strategy.

$\gamma^x(\omega), \gamma^y(\omega)$ are cubic spline functions and $\omega \in [i-1, i]$ indicates that b is interpolated between two control points (electrodes) e^{i-1} and e^i.

2.1. Automatic Selection of Catheter Shape Representation and Tracking Strategy

As consecutive electrodes are not too distant from each other, the number of electrodes is a good indicator of the model complexity required to approximate the CS catheter. Thus, depending on the number of electrodes, the catheter model is approximated using one, two or three segments, each being a polynomial curve as illustrated in Figure 3. The model representation also drives the tracking strategy, involving tracking one, two or three segments. Algorithm 1 sketches the catheter-specific tracking strategy, in which $B^1 = 8, B^2 = 14$. In our database, a CS catheter can contain up to 20 electrodes.

The catheter segments are approximated as polynomials of degree at most three relative to a system of coordinates centered in the middle of a line segment that connects two given points on the curve, as illustrated in Figure 3.
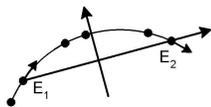

Figure 3. The catheter segment is approximated as a degree two or three polynomial passing through two detected electrodes e^i, e^{i+1}, with given tangents at one or both of the two electrodes.

From the initial shape representation in the first frame, the tracking template is obtained and a system of coordinates that is relative to the shape representation. In this system, a point P_1 near the curve has coordinates $(\frac{l(e^1 P_2)}{l(C_0)}, \|P_1 - P_2\|)$, where P_2 is the closest point on the curve to P_1 and $l(e^1 P_2)$ is the length of the curve from the tip to P_2.

2.2. Detection of Collimator, Catheter Tip and Electrodes

Detection of the collimator is useful for bounding the estimation of catheter motion and location. In our approach, the collimator position on each side is detected using a trained border detector based on Haar features. In real-time fluoroscopy, the collimator can be obtained directly from the imaging device.

Accurate detection of catheter electrodes not only provides robust estimation of the catheter position but also helps prune the search space for catheter tracking. Moreover, it is useful for predicting when the catheter moves out of or back into the view.

The CS tip and electrodes are detected as oriented points (x, y, θ), parameterized by their position (x, y) and orientation θ. For fast detection, we use Marginal Space Learning [16] to first detect just the tip and electrode positions (x, y) and then at promising positions search for all orientations θ. Tip and electrode positions (x, y) are detected using trained binary classifiers. The classifiers use about 100,000 Haar features in a centered window of size 69×69. Each classifier is a Probabilistic Boosting Tree (PBT) [12] and can output a probability $P(e = (x, y)|D)$. The catheter tip is different from the other electrodes in term of context and appearance and it can be detected more reliably. An example of catheter electrode position detection is illustrated in Figure 4. The detected electrode and candidate positions are then augmented with a set of discrete orientations and fed to a trained oriented point detector, and the same applies for the detected tip positions. The oriented point detectors use a richer feature pool including steerable feature responses and image intensity differences relative to the query position and orientation.

The set of detected electrodes and tips at each frame is fed to a non-maximal suppression (NMS) stage that cleans up clustered detections. In each frame, at most I electrodes and F tips are kept as the detection results, denoted as $H_t^E = \{h_t^1, ..., h_t^I\}$ and H_t^T respectively. Any detection at distance at least 250 pixel from the initial CS catheter location are removed. This relies on the observation that during the ablation procedure, the CS catheter has only a limited range of motion due to breathing and the heartbeat.

2.3. Hypothesis Generation

Tracking hypotheses are generated as candidate shapes in the current frame. Given consolidated tip and electrode detection points, we propose two novel schemes to generate catheter tracking hypotheses. For long catheters, these hypotheses are generated for each catheter segment and constrained to be coherent.

One set of hypotheses is generated by parametrically manipulating the catheter model-based on detected tip and electrode point candidates and the assumption that at least one
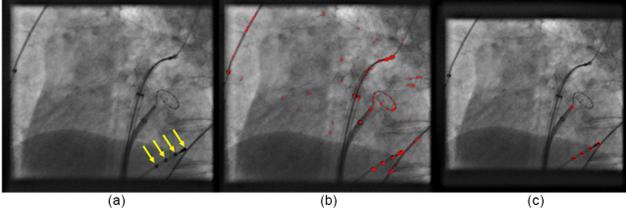
Figure 4. Automatic CS catheter electrode detection. (a) Input image; four yellow arrows point to electrodes. (b) Automatically detected electrode positions (red points). (c) 5 NMS electrode points (red circles) used for model-based hypothesis generation.

electrode detection from H_t^E is correct. The scheme works as follows:

- Input: catheter model $C_0 = e_0^1, ..., e_0^K$;

- Generate seed hypotheses by translating e^j to each detected electrode position h_t^i and obtain a translation vector by which we translate C_0 to get a seed hypothesis Q_j^i. In total we obtain $(K \cdot I)$ seed hypotheses.

- For each Q_j^i, we consider new location of h_t^i as the transformation center and apply a set of affine transformation to generate tracking hypotheses as:

$$\bar{L}_o^s = A \cdot L_o^s, \ A = \begin{bmatrix} \mathcal{C} & d \\ 0^T & 1 \end{bmatrix} \quad (1)$$

where L_o^s represents catheter model coordinates and o indicates the order ($o = 0$ represents the order from the tip to the PCS and $o = 1$ for reverse) and s is the segment index.

This strategy is efficient in generating effective tracking hypotheses. However, it may miss some hypotheses due to catheter motion and shape deformation. Therefore, we add another set of hypotheses that is generated directly from detected oriented tip and electrode points as follows:

- A set of rigid transformation hypotheses that assume that one of electrodes is detected with correct orientation. Thus the hypotheses are obtained by rotating and translating L_o^s to match one of its electrodes and its orientation to the detected oriented electrode.

- Another set of non-rigid transformation hypotheses that assumes that the tip and one of the electrodes are correctly detected and either the tip or the electrode has reliable orientation. In this case all pairs of tip and electrode detections are considered if they are at distance within a range relative to L_o^s. For each such pair, two polynomial curves of degree two and one of degree three are constructed as illustrated in Figure 3. The condition that the curve passes through the two given points imposes two constraints on the polynomial, while each tangent imposes another constraint. Thus, if only one

tangent orientation is known, a degree two polynomial is completely determined, while if both tangents are known, a degree three polynomial can be computed. Curves that differ too much from C_0 are removed from the set of hypotheses.

In sum we obtain a pool of tracking hypotheses, and fusion of two hypothesis generation schemes leads to a near-complete and effective hypothesis pool. Our experiments show that $I = 15, F = 10$ are sufficient for tracking all kinds of CS catheters as seen in our database.

2.4. Learning-based Hypothesis Evaluation

An effective tracking hypothesis evaluation method is necessary to determine the exact position and shape of the CS catheter. Using our notations the object function of the classic mean shift tracking algorithm (MS) [5] at t-th frame is defined as:

$$\begin{aligned} \hat{C}_t &= \arg\min_{C_t} d(C_t, C_0) \\ &= \arg\min_{C_t} \sqrt{1 - \rho[C_t, C_0]} \end{aligned} \quad (2)$$

where $\rho[C_t, C_0]$ is the Bhattacharyya coefficient. MS shows that the most probable location of the target in the current frame is obtained by minimizing the above distance, which is equivalent to maximizing $\rho[C_t, C_0]$.

In some cases, however, the tracking problem cannot be directly formulated as a maximizing-the-Bhattacharyya-coefficient problem. Main reasons include feature representation and problem formulation. Here we introduce a Bayesian framework to evaluate catheter tracking hypotheses. Recent tracking advancements [6, 14, 15] have shown the power of information fusion. The overall goal for evaluating a tracking hypothesis is to maximize the posterior probability:

$$\hat{C}_t = \arg\max_{C_t} P(C_t|Z_{0...t}) \quad (3)$$

where $Z_{0...t}$ is image observation from 0 to t-th frame. By assuming a Markovian representation of the catheter motion the above formula can be expanded as:

$$\begin{aligned} \hat{C}_t &= \arg\max_{C_t} P(C_t|Z_{0...t}) \\ &= \arg\max_{C_t} P(Z_t|C_t)P(C_t|C_{t-1})P(C_{t-1}|Z_{0...t-1}) \end{aligned}$$
$$(4)$$

The above formula essentially combines two parts: the likelihood term, $P(Z_t|C_t)$, which is computed as combination of detection probability and template matching score and the prediction term, $P(C_t|C_{t-1})$, which captures the
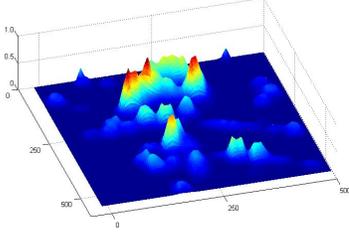
Figure 5. An example of electrode probability map.

motion smoothness. To maximize tracking robustness, the likelihood term $P(Z_t|C_t)$ is estimated by combining tip and electrode detection and catheter body template matching as follows:

$$P(Z_t|C_t) = (1 - \lambda) \cdot P(E_t^*|C_t) + \lambda \cdot P(T_o^s|C_t) \quad (5)$$

where E_t^* is estimated probability measure about electrodes and tips at t-th frame that assists estimation of C_t and λ is defined as:

$$\lambda = \frac{1}{1 + e^{-f(T_o^s, D(C_t))}}, f(T_o^s, D(C_t)) = \frac{cov(T_o^s, D(C_t))}{\sigma(T_o^s) \cdot \sigma(D(C_t))}, \quad (6)$$

where $cov(T_o^s, D(C_t))$ is the intensity cross-correlation between the catheter model template and the image band expanded by C_t. $\sigma(T_o^s)$ and $\sigma(D(C_t))$ are the intensity variance. The detection term $P(E_t^*|C_t)$ is defined in terms of a part model as:

$$P(E_t^*|C_t) = \nu_1 P(E_t^*|e_t^1) + \nu_K P(E_t^*|e_t^K)$$
$$+ \frac{1 - \nu_1 - \nu_K}{K - 2} \sum_{i=2}^{K-1} P(E_t^*|e_t^i), \quad (7)$$

where $P(E_t^*|e_t^1)$ defines the detection probability at the tip, $P(E_t^*|e_t^K)$ defines the probability at the PCS and $P(E_t^*|e_t^i)$ represents the probability at each other electrode. $\nu_1 = 0.3$ and $\nu_K = 0.2$ in our experiments. The similar part-based model has shown effectiveness in [14]. Figure 5 shows an example of electrode probability map.

The prediction term $P(C_t|C_{t-1})$ in Equation (4) is modeled as a zero-mean Gaussian distribution $N(0, \sigma_C)$ with σ_C learned from the training data.

2.5. Non-Rigid Tracking & Online Template Update

The shape of the CS catheter may deform non-rigidly due to the impact of cardiac motion, respiratory motion, and/or projection angulation. In order to handle non-rigid deformation, the algorithm may divide the catheter model into multiple segments based on the number of electrodes and the shape (Algorithm 1). Let $\{e_0^1, e_0^2, ..., e_0^K\}$ represent the

electrodes initialized at frame 0 by the user, the algorithm divides the electrodes into 3 segments if $K > 14$ and 2 segments if $K > 8$. In cases of 2 segments, let $\xi(e^i)$ represent the curvature of the catheter at electrode i, the algorithm finds $e^j = \arg\max_i \xi(e^i)$ as the joint point and cut the electrode set into two segments (sets), $\{e_0^1, e_0^2, .., e_0^j\}$ and $\{e_0^j, e_0^{j+1}, ..., e_0^K\}$. The algorithm then performs tracking on the first segment by the aforementioned tracking approach. After the first segment has been tracked, the location of e^j is served as the transformation center to generate the hypotheses for the second segment. Therefore, the dimension of search space for the second segment is much lower and the search is faster. Using a joint electrode e^j in both segments also guarantees one integrated catheter model as output. To deal with the case when detection misses all the electrodes in the first segment, we perform another tracking from the opposite direction by tracking the second segment first followed by tracking the first segment. The results of these two directions are then evaluated by the overall score, which combines each segment's score as:

$$P(Z_t|C_t) = \sum_{s=1}^{2} \epsilon_s \cdot P(Z_t|C^s) \quad (8)$$

where $P(Z_t|C^s)$ is computed by Equation (5) and ϵ_s is computed as the ratio of the segment length to the sum of all segment lengths. s is the segment index. For 3 segment cases, after the first joint e^{j1} is found, the same curvature analysis is applied to the longer segment to find another joint e^{j2}. Then tracking is performed the same way (bi-directional) as the 2 segment cases.

Since the model-based hypotheses are generated in a discrete space, small errors may be present even for the best candidate. In order to refine the results, after the best hypothesis is found, we adopt the Powell's method [11] to search for the maximum in the parameter space.

Foreground and background structures in fluoroscopy are constantly changing and moving. In order to cope with it dynamically, the catheter model is updated online by:

$$T_{o,t}^s = (1 - \varphi_w) \cdot T_{o,t-1}^s + \varphi_w \cdot D(C_t), \text{ if } P(Z_t|C_t) > \varphi_t, \quad (9)$$

where $T_{o,t}^s$ represents the model template in frame t. $D(C_t)$ is the model obtained at frame t based on the output C_t. $\varphi_w = 0.1$ and $\varphi_t = 0.4$ are set in our algorithm. The impact of online updating the model is evaluated in Table 2.

3. Experiments

3.1. Data and Annotation

1073 fluoroscopic sequences, containing 13221 frames collected from Electrophysiology (EP) Afib procedures are
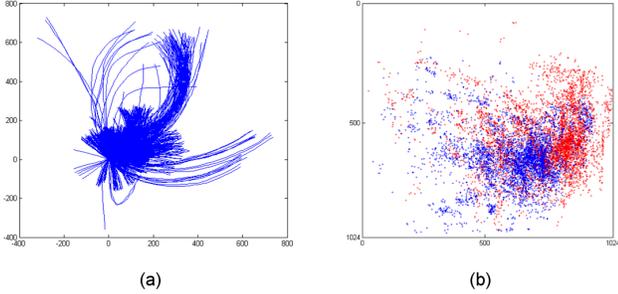
Figure 6. Illustration of our CS catheter dataset: (a) Catheter shapes after aligning to the PCS; (b) Distribution of the tips (red) and the PCSs (blue).
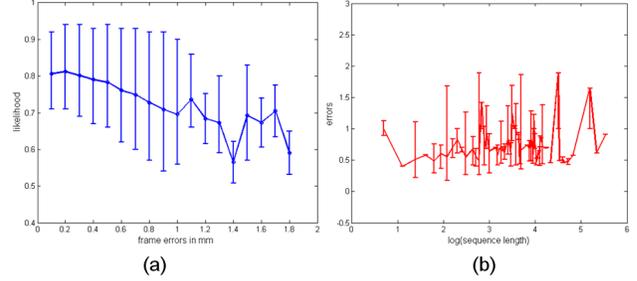

Figure 7. Statistics of errors (mm) on the evaluation set for the **ARO** method: (a) The likelihood (Eq.(5)) versus frame errors and each bar shows max/min likelihood values; (b) Sequence errors (mean and error bar) versus log(sequence length).

used as our database for evaluation. The original image resolutions are either 1024×1024 or 1440×1440 with pixel spacing 0.154, 0.1725 or 0.183 mm/pixel. The electrode and tip detectors are trained from 5103 frames annotated manually. To illustrate the variability of our tracking target, we illustrate the CS catheter shapes and spatial distribution of the catheter tips and PCSs in Figure 6.

3.2. Catheter Tip and Electrode Detection

In the first experiment, we evaluate the trained catheter tip and the electrode detectors. Our tip and electrode detectors are evaluated on 1507 frames. For catheter electrode and tip detection, the top 50 electrode candidates and the top 10 tip candidates are extracted at each frame. Detection rate is measured by (number of ground truth electrodes (tips) that are detected) / (total number of ground truth electrodes (tips)). A candidate which is away from the ground truth location by 3mm is regarded as a false detection. The results are summarized in Table 1. Since the model-based hypotheses are generated in the way that only if none of the electrodes on the catheter is detected, the algorithm could possibly miss the ground truth hypothesis, the probability of missing the ground truth in the proposed framework is significantly low.

	Detection rate	False detection #/frame
Electrode	0.94	23.64
Tip	0.97	1.14

Table 1. Detection rate and false positive number (per frame) for electrode and tip detection.

3.3. Tracking

For all 1073 sequences in our database, we have annotated all the electrodes (from the catheter tip to the PCS) in the first frame and 3337 randomly selected frames. During evaluation, the annotation in the first frame is regarded as user initialization to the algorithm. The algorithm then tracks the catheter as a set of electrodes in the remaining frames. Tracking errors are then evaluated only on the

frames with ground truth annotation. Let the electrodes annotated at i-th frame be $\{a_i^1, a_i^2, ..., a_i^K\}$, the tracking error for i-th frame is defined as:

$$\frac{1}{K} \sum_{k=1}^{K} ||a_i^k - e_i^k||_{L^2}, \qquad (10)$$

where $\{e_i^1, e_i^2, ..., e_i^K\}$ are the tracked electrodes by the algorithm.

The tracking error is summarized in Table 2. We report frame errors in millimeter (mm). All frame errors are sorted in ascending order and Table 2 reports the errors at mean, median, percentile 85 (p85), 90 (p90), 95 (p95) and 98 (p98). Although tracking catheters in real fluoroscopic sequences is a non-trivial task, our algorithm turns out to be very robust against different challenging scenarios and has an error less than 2mm in 97.8% of the total evaluated frames (c.f. the last row in Table 2).

While the major novelty and the tracking power of the proposed tracking algorithm comes from the robust and efficient hypothesis generation and fusion, we illustrate and compare the impact of other important components in Table 2 as well. **DON** is the method by setting $\lambda = 0$ in Eq. (5), which essentially only considers the detection term; **ADD** is the method using Eq. (5) with no input refinement or online template update; **ADR** is **ADD** with input refinement; and **ARO** is **ADR** with online template update. **ARO** is the final complete version of our algorithm. During comparison, the number of detected electrode candidates per frame is set as 15 and all other settings are exactly the same. We have tried other options of fusing detection probability and template matching score, such as multiplication of the two terms in Eq. (5). The effectiveness of Eq. (5) is validated through our batch evaluation over 1000+ sequences.

Due to our robust detection and tracking framework, even the performance of **DON** is already very good in most of the cases. However, improvement due to fusion of learning and image content, automatic user input refinement, and online template updates can still be observed from the rows of

ADD, **ADR**, and **ARO**, respectively.

To investigate our hypothesis evaluation scheme, we depict the relationship between the tracking errors and the likelihood obtained from Eq. (5). Figure 7 (a) shows that the likelihood measure is a good indicator of the tracking error, which demonstrates Eq. (5) as a robust measure for hypothesis evaluation.

To evaluate whether the drifting problem exists in our tracking algorithm, we depict the error versus the length of the sequence on the dataset. Figure 7 (b) shows that the errors stay in the same range regardless the length of a sequence. This result demonstrates that our algorithm has little drifting problem during tracking.

In our last experiment, we try to find the optimal number of electrode candidates during tracking. If the number of candidates is too small, it is possible that none of the ground truth electrode is hit and the ground truth hypothesis is missed in the hypothesis space we generated. On the other hand, if the number of candidates is too large, it increase unnecessary search in the hypothesis space and decreases the tracking speed, it may also introduce more false detections and result in tracking errors. Table 3 compares performance of extracting 3, 7, 11, 15, 20 electrode candidates per frame. While median errors remain mostly the same which implies that increasing detected electrode number does not have much impact on small error data, mean errors are consistently improved until 15 electrode candidates are used. Therefore, the proposed algorithm uses 15 electrode candidates per frame in its final version in order to balance between performance and speed. In Figure 8, we show several single frame results of catheter tracking on challenging scenarios including the target catheter overlapping with other catheters or structures in a cluttered background (A, C, H, I), non-rigid deformation (B, C, E, G, J), foreshortening due to 3D to 2D projection (B, F, G), the target catheter moves out of the image ROI (I), and low SNR (D, H, J), etc. On average, the proposed tracking algorithm reaches 5 frames per second on a desktop machine with Intel Xeon CPU (2.27GHz). If interested, an demo video with many results can be found at https://sites.google.com/site/cvpr111013 /CatheterTracking_CVPR2011.wmv.

It is worth mentioning that the proposed approach is generic and is not limited to track the CS catheters. Although the number and size may be different, electrodes are seen on most catheters in an EP. The same algorithm can easily be generalized to track other catheters, such as circumferential mapping catheters and ablation catheters. Figure 9 shows results of three such examples. Furthermore, our tracking and hypothesis generation scheme can be used to track other types of targets as well by replacing the electrodes with the landmarks on the target.

	mean	median	p85	p90	p95	p98
ARO3	1.95	0.48	0.77	0.90	1.94	10.30
ARO7	0.98	0.49	0.74	0.87	1.43	3.53
ARO11	0.96	0.48	0.74	0.85	1.38	3.72
ARO15	0.76	0.50	0.73	0.82	1.04	2.14
ARO20	0.96	0.48	0.74	0.87	1.38	3.97

Table 3. Evaluation of the impact of electrode candidate number to the tracking performance. ARO3, ARO7, ARO11, ARO15, ARO20 uses top 3, 7, 11, 15, 20 electrode candidates per frame respectively.

4. Conclusion and Future Work

Tracking catheters in the fluoroscopic data is a challenging task due to cardiac and respiratory motion. In addition, the data often contain complex background, constant motion, variations and often suffer from low signal to noise ratio (SNR) due to preferable low radiation in clinics. Our paper focuses on a novel and robust technology to track catheters, which area highly deformable wire structures. We have proposed a robust learning-based hypothesis generation and information fusion framework to automatically detect and track catheters in fluoroscopy. Its unique hypothesis generation and fusion scheme differentiates our work from existing approaches and makes our tracking algorithm efficient and robust. Promising experimental results on a large dataset (1073 sequence) have shown that 97.8% of evaluated data have errors smaller than 2mm. Furthermore, our proposed approach is generic and can be generalized to track other kinds of catheters or to detect and track part-based objects in other types of data.

A 3D position estimation is possible during a bi-plane acquisition. However, the paper focuses on the tracking technology for 2D scenes. Note that bi-plane acquisition or other approaches are not always available in hospitals so 2D image-based tracking of catheters is still necessary in many clinical settings. Our future work includes automation of user initialization. For example, the user only needs to click the catheter tip and all other electrodes are located automatically.

5. Acknowledgement

	mean	median	p85	p90	p95	p98
DON	1.16	0.66	0.98	1.12	1.67	4.26
ADD	0.91	0.45	0.72	0.86	1.56	4.45
ADR	0.78	0.48	0.72	0.81	1.10	2.40
ARO	0.76	0.50	0.73	0.82	1.04	2.14

Table 2. CS catheter tracking performance. The last row shows the best performance including all essential components.
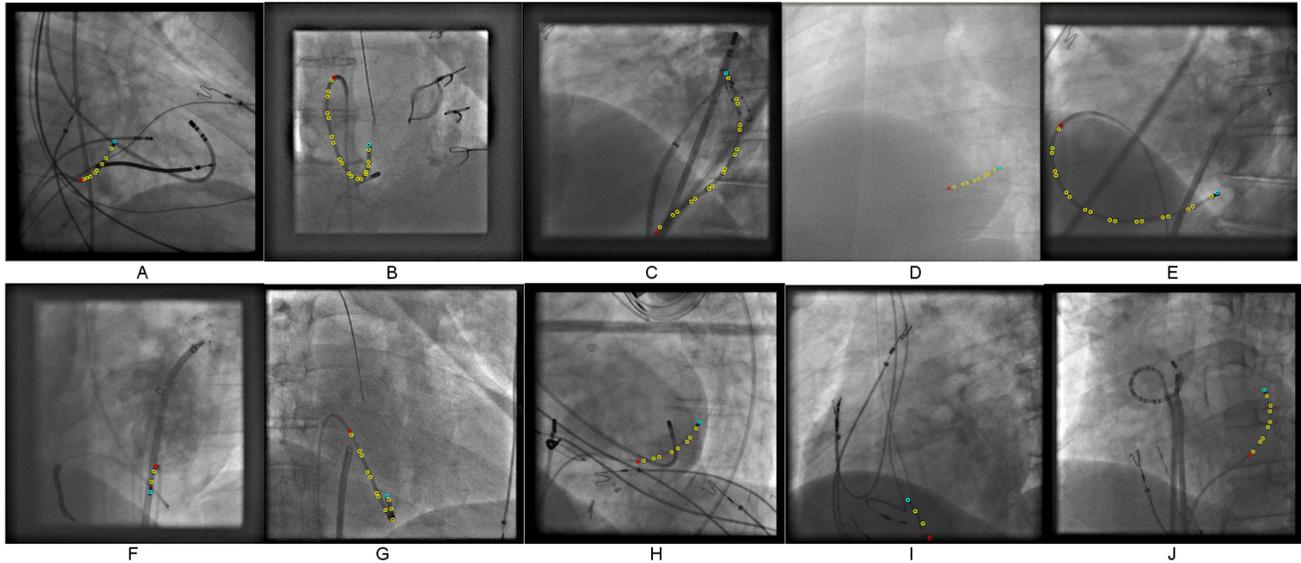
Figure 8. Results of tracking catheters in 10 different sequences. Cyan, yellow, and red circles indicate the catheter tip, intermediate electrodes, and PCSs, respectively.
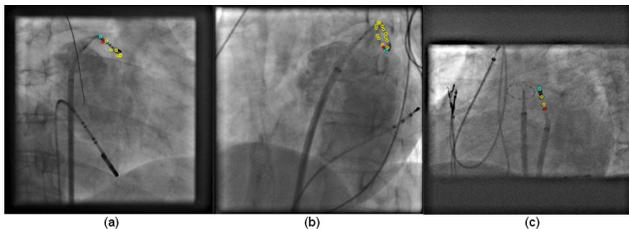

Figure 9. Results of our approach successfully tracking other catheters: circumferential mapping catheters in (a) and (b) and an ablation catheter in (c).

References

[1] S. Avidan. Ensemble tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007. 1098

[2] B. Babenko, M.-H. Yang, and S. Belongie. Visual tracking with online multiple instance learning. In *CVPR*, 2009. 1098

[3] A. Barbu, V. Athitsos, B. Georgescu, S. Boehm, P. Durlak, and D. Comaniciu. Hierarchical learning of curves application to guidewire localization in fluoroscopy. In *CVPR*, 2007. 1098

[4] H. Calkins and et al. HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for personnel, policy, procedures and follow-up. a report of the heart rhythm society (HRS) task force on catheter and surgical ablation of atrial fibrillation. *Heart Rhythm*, 4, 2007. 1097

[5] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *CVPR*, 2000. 1100

[6] D. Comaniciu, X. Zhou, and S. Krishnan. Robust realtime tracking of myocardial border: An information fusion approach. *IEEE Transactions on Medical Imaging*, 2008. 1100

[7] H. Grabner and H. Bischof. On-line boosting and vision. In *CVPR*, 2006. 1098

[8] H. U. Klemm and et al. Catheter motion during atrial ablation due to the beating heart and respiration: Impact on accuracy and spatial referencing in three-dimensional mapping. *Heart Rhythm*, 2007. 1097, 1098

[9] Y. Ma, A. P. King, N. Gogin, C. A. Rinaldi, J. Gill, R. Razavi, and K. S. Rhode. Real-time respiratory motion correction for cardiac electrophysiology procedures using image-based coronary sinus catheter tracking. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2010. 1098

[10] J. Pilet, V. Lepetit, and P. Fua. Real-time non-rigid surface detection. In *CVPR*, 2005. 1098

[11] M. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal*, 1964. 1101

[12] Z. Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *ICCV*, 2005. 1098, 1099

[13] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 2004. 1098

[14] P. Wang, T. Chen, Y. Zhu, W. Zhang, S. K. Zhou, and D. Comaniciu. Robust guidewire tracking in fluoroscopy. In *CVPR*, 2009. 1098, 1100, 1101

[15] Y. Wang, B. Georgescu, D. Comaniciu, and H. Houle. Learning-based 3D myocardial motion flow estimation using high frame rate volumetric ultrasound data. In *IEEE International Symposium on Biomedical Imaging*, 2010. 1100

[16] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE Transactions on Medical Imaging*, 2008. 1099