

Curvilinear Structure Tracking by Low Rank Tensor Approximation with Model Propagation

Erkang Cheng¹, Yu Pang¹, Ying Zhu², Jingyi Yu³, Haibin Ling¹

¹Department of Computer and Information Sciences, Temple University, Philadelphia, PA USA

²Siemens Corporate Technology, Princeton, NJ USA

³Department of Computer and Information Science, University of Delaware, Newark, DE USA

Abstract

Robust tracking of deformable object like catheter or vascular structures in X-ray images is an important technique used in image guided medical interventions for effective motion compensation and dynamic multi-modality image fusion. Tracking of such anatomical structures and devices is very challenging due to large degrees of appearance changes, low visibility of X-ray images and the deformable nature of the underlying motion field as a result of complex 3D anatomical movements projected into 2D images. To address these issues, we propose a new deformable tracking method using the tensor-based algorithm with model propagation. Specifically, the deformable tracking is formulated as a multi-dimensional assignment problem which is solved by rank-1 ℓ_1 tensor approximation. The model prior is propagated in the course of deformable tracking. Both the higher order information and the model prior provide powerful discriminative cues for reducing ambiguity arising from the complex background, and consequently improve the tracking robustness. To validate the proposed approach, we applied it to catheter and vascular structures tracking and tested on X-ray fluoroscopic sequences obtained from 17 clinical cases. The results show, both quantitatively and qualitatively, that our approach achieves a mean tracking error of 1.4 pixels for vascular structure and 1.3 pixels for catheter tracking.

1. Introduction

X-ray fluoroscopy is one of the primary modalities used in image guided interventional procedures. Robust deformable tracking of intravascular devices (e.g. guidewire, catheters as seen in Fig. 1) or vascular structures in dynamic X-ray images is essential in image guided interventions. These structures serve as surrogate signals for dynamic overlay of 3D patient-specific anatomical information over 2D X-ray images. Fig. 1 demonstrates the challenges of

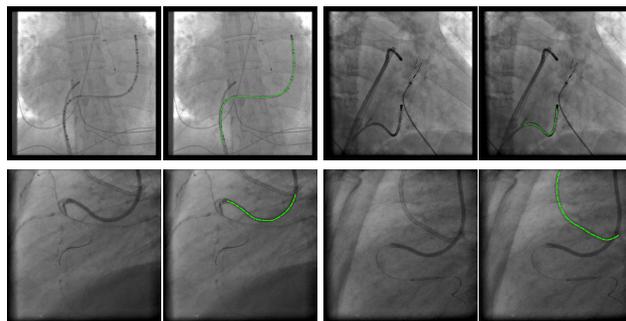


Figure 1. Examples of X-ray fluoroscopic images from different sequences. Images in the first row contain catheters. Vascular structures are shown in the second row. The odd columns are original images, where annotations of catheters or vascular structures are illustrated in the even columns (in green).

the tracking problem. For intravascular devices and vascular structures, large variations in their appearances such as shape and intensity are observed. Low visibility and poor image quality due to a low dose of radiations in interventional imaging make the deformable tracking more difficult. Furthermore, these structures undergo complex deformable motion as a result of 3D deformable anatomical movements projected onto the 2D image plane. In this paper, we address the problem of deformable tracking in dynamic X-ray images.

Object tracking techniques [21] based on appearance features such as intensity and color histograms are not suitable for deformable tracking due to the large variation and non-linear deformation in target shapes. Active contour and level set methods [13] heavily rely on the design of energy function and the choice of object specific parameters. Palti-Wasserman et al. [15] used a filter based method to identify a guidewire in an X-ray image and used Hough transform to fit a polynomial curve to track a guidewire. In [1], guidewire is modeled as a line-like structure and the deformation is achieved by spline optimization. Wang et

al. [19] proposed a learning-based method which aimed at detecting a guidewire in individual frames. The shape of the guidewire is captured by snake-like energy function involving probabilities computed by guidewire detector. These approaches rely on appearance detection for tracking and do not explicitly address the estimation of guidewire deformation.

Image registration and alignment are another set of techniques suitable for deformable tracking. Bhagalia et al. [4] presented an importance sampling approach which used sampled image points with larger gradients to perform registration. To cope with static structures in X-ray images, Zhu et al. [24] introduced a method to separate static structures from moving structures through temporal statistics across frames. The registration is achieved by variational approach [10]. A motion layer separation solution for the similar task was proposed in [23].

In this paper, the objective is to track the curvilinear structure in the dynamic X-ray sequence. The curvilinear structure is represented by a model which is manually labeled in the first frame. Deformable tracking of the curvilinear structure is then estimated during remaining frames. To address the problem, we propose a new approach using tensor-based multi-dimensional assignment (TMDA) framework. First of all, curvilinear structure candidates are detected by a discriminative classifier. Then, we construct a high-order tensor from all trajectory candidates over a time span. Associations of detected candidates between adjacent frames are computed by rank-1 ℓ_1 tensor approximation. Finally, curvilinear structure tracking is achieved by model propagation during the frames. Through model propagation, higher order temporal information helps to reduce ambiguity during object localization across time. To validate the proposed approach, we apply it to curvilinear structure tracking in clinical images obtained from cardiac and liver interventions. A challenging dataset containing 17 clinical X-ray sequences, together with manually annotated ground truth, is used for testing. The proposed method demonstrates very promising results and achieves a mean tracking error of 1.4 pixels for vascular structure and 1.3 pixels for catheter tracking.

In summary, there are two main contributions of our work. Firstly, we propose to use tensor-based multi-dimensional assignment for curvilinear structure tracking. Deformable tracking is estimated by using higher-order information from multiple frames. In contrast, most traditional methods only consider two frames in the tracking. Secondly, model propagation is applied in tensor approximation to preserve the model's spatial constrain during the tracking. Unlike standard tensor approximation, our method does not require the computation of curvilinear structure candidates matching between two consecutive frames. The curvilinear structure can be directly computed

from model propagation. Another benefit of model propagation is to reduce unnecessary associations of detected curvilinear structures. Therefore, only the model is propagated in the tracking.

The remaining of the paper is organized as follows. Related work is reviewed in Sec. 2. In Sec. 3, we formulate the deformable tracking problem into the tensor-based multi-dimensional assignment framework. Experimental results are presented in Sec. 5. Finally, Sec. 6 concludes the paper.

2. Related Work

With observations of a batch of frames, object tracking is formulated as data association of these observations across frames. The classic method is Multiple Hypothesis Tracking (MHT) [16] which finds all possible association combinations and selects the most likely association set as the optimal solution. In general, MHT optimization is an NP-hard problem and the computation is prohibitive when the numbers of objects and frames are large.

Multiple target association across K frames can be formulated as the multiple dimensional assignment (MDA) problem. Although there exists exact solutions with polynomial time for two-frame association. For instance, Hungarian algorithm is one popular method for this special case of MDA, where $K = 2$. It is usually infeasible to search the global optimal solution of MDA when $K > 2$, which is NP-hard when no assumption is provided. Recent studies focus on approximation solution of MDA problem. In [17], a rank constrained continuous formulation of multi-frame multi-target tracking problem is approximated by semi-definite program method. The Lagrange relaxation strategy is applied in [9] to present an efficient and generalized assignment algorithm.

Some works treat MDA as a graph theory problem such as bipartite matching, which incorporates a limited-temporal-locality of the sequence. One of them is to formulate MDA as a network flow problem by decomposing the cost of the trajectory as the product of pairwise terms. Then the network flow problem is solved by linear programming [12], shortest path algorithm [3] etc. Although global optimal solution by polynomial time complexity can be achieved by these methods, higher-order temporal information is missing in the network flow formulation. In order to achieve long-term data association, generalized minimum clique graph is proposed in [22] to incorporate long-term temporal span. Maximum weight independent set of the graph built by pairs of detections is provided to solve similar MDA problem in [5]. Other alternative approaches include Markov Chain Monte Carlo Data Association approaches [2, 14], greedy search [20] and hierarchical target association [11].

Our work is closely related to the framework in [18] by using rank-1 tensor approximation of MDA. Their iterative

approximate solution is inspired by [7], which iteratively solves two-frame assignments in turn while keeping all other assignments fixed. Sharing similar procedure of [18], the curvilinear structure tracking is formulated as a MDA problem. Tensor-based approximation is adopted to estimate the motion of curvilinear structure. Compared to [18], a model is involved in the tensor-based formulation of MDA. In addition, our model prior is preserved during tracking by performing model propagation.

3. Curvilinear Structure Tracking

We first introduce notations used in this paper. We will denote scalars by lower-case italic letters (a, b, \dots); vectors by lower-case boldface letters ($\mathbf{a}, \mathbf{b}, \dots$), matrices by boldface capitals ($\mathbf{A}, \mathbf{B}, \dots$), and tensors by calligraphic letters ($\mathcal{A}, \mathcal{B}, \dots$). The notation is consistently used for lower-order parts of a given structure. For example, the i -th entry of a vector \mathbf{a} is denoted by \mathbf{a}_i , the (i, j) -th entry of a matrix \mathbf{A} by a_{ij} and the (i_0, i_1, \dots, i_K) -th entry of a K -th order tensor \mathcal{A} by $a_{i_0 i_1 \dots i_K}$.

3.1. Problem Formulation

A deformable object is often modeled as a set of elements whose spatial relations vary across time. To track a deformable object, roughly speaking, is to localize or detect its elements across time span. We plan to formulate deformable tracking as a spatial constrained tensor-based multi-dimensional assignment.

The model is manually labeled in the first frame and tracking is performed throughout the following frames. $O^{(0)} = \{\mathbf{o}_{i_0}^{(0)}\}_{i_0=1}^{I_0}$ is denoted as the model to be tracked with I_0 elements. The topology of the object is encoded by an edge set $\mathbb{E} = \{(i_0, i'_0) : \mathbf{o}_{i_0}^{(0)} \text{ and } \mathbf{o}_{i'_0}^{(0)} \text{ are connected}\}$. Then, let $O^{(k)} = \{\mathbf{o}_{i_k}^{(k)}\}_{i_k=1}^{I_k}$ be the set of curvilinear candidate elements for the k -th frame, $k = 1, 2, \dots, K$, and $I_0 \leq I_k$. We assume all sets have equal number of items for notation convenience. This assumption will not affect the algorithms and analysis, since we can always pad the sets with virtual items that are used for handling missing or noisy items. Therefore, we can rewrite it as $O^{(k)} = \{\mathbf{o}_{i_k}^{(k)}\}_{i_k=1}^I$ where each frame has I items.

MDA aims to find a $(K+1)$ -partite assignment from these sets that maximizes the total matched affinity. The problem can be formally defined as the following constrained optimization:

$$\arg \max_{\mathcal{X}^*} f^*(\mathcal{X}^*) = \sum_{i_0} \sum_{i_1} \dots \sum_{i_K} a_{i_0 i_1 \dots i_K}^* x_{i_0 i_1 \dots i_K}^* \quad (1)$$

$$\text{s.t.} \begin{cases} \sum_{i_0, i_1, \dots, i_K \setminus \{i_k\}} x_{i_0 i_1 \dots i_K}^* = 1, & \forall k \\ x_{i_0 i_1 \dots i_K}^* \in \{0, 1\}, & \forall k \end{cases} \quad (2)$$

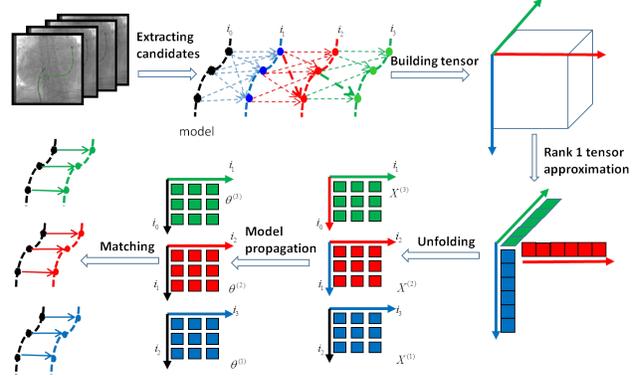


Figure 2. Overview of proposed method.

where $a_{i_0 i_1 \dots i_K}^*$ denotes the affinity for a joint assignment of item $O_{i_0, i_1, \dots, i_K} = \{\mathbf{o}_{i_0}^{(0)}, \mathbf{o}_{i_1}^{(1)}, \dots, \mathbf{o}_{i_K}^{(K)}\}$, $x_{i_0 i_1 \dots i_K}^*$ indicates whether O_{i_0, i_1, \dots, i_K} is selected ($= 1$) or not ($= 0$); and $\mathcal{X}^* = (x_{i_0 i_1 \dots i_K}^*)$. Finally, the tracking result of curvilinear structure $O^{(0)}$ can be retrieved from the joint assignments in k -th frame. The overview of the proposed method is provided in Fig 2.

3.2. Low Rank Tensor Approximation

Following the approach presented in [18], the MDA problem is converted to a new tensor-based MDA formulation. Taking benefit of the non-negativity of $a_{i_0 i_1 \dots i_K}^*$ and $x_{i_0 i_1 \dots i_K}^*$, f^* can be represented in a tensor form as:

$$f^*(\mathcal{X}^*) = \|\mathcal{X}^* \bullet \mathcal{A}^*\|_F, \quad (3)$$

where $\mathcal{A}^* = a_{i_0 i_1 \dots i_K}^*$ is the affinity tensor, ' \bullet ' the Hadamard product and $\|\cdot\|_F$ the Frobenius norm.

Intuitively, to maximize f^* , a large $a_{i_0 i_1 \dots i_K}^*$ encourages a large $x_{i_0 i_1 \dots i_K}^*$, which in turn suggests a sequence of binary matching $(\mathbf{o}_{i_0}^{(0)} \mathbf{o}_{i_1}^{(1)}), (\mathbf{o}_{i_1}^{(1)} \mathbf{o}_{i_2}^{(2)}), \dots, (\mathbf{o}_{i_{K-1}}^{(K-1)} \mathbf{o}_{i_K}^{(K)})$. This suggests us to seek solutions through between-set assignments, hereafter referred as local assignments. By contrast, we refer the joint assignment as global assignment in the rest of the paper.

We first decompose a global assignment variable $x_{i_0 i_1 \dots i_K}^*$ as a sequence of local assignment ones. Let the local assignment between two consecutive sets $O^{(k-1)}$ and $O^{(k)}$ be represented by an assignment matrix $\mathbf{X}^k = (x_{i_{k-1} i_k}^*)^{(k)} \in \mathbb{R}^{I \times I}$, which can be reshaped as a vector denoted as $\mathbf{x}^{(k)} = (x_{j_k}^k) \in \mathbb{R}^J$, where $J \doteq I^2$. Hereafter we use the same scalar symbol x to denote entries in both \mathbf{X} and \mathbf{x} , with double subscripts and a single subscript respectively. In addition, given the matrix entry index (i_1, i_2) , we denote its index after vectorization by $\downarrow(i_1, i_2)$, while given a vector entry index j , we denote its corresponding matrix index as (\bar{j}, \bar{j}) . To summarize,

we have $(\mathbf{X})_{i_1 i_2} = x_{i_1 i_2} = x_{\downarrow(i_1, i_2)} = (\mathbf{x})_{\downarrow(i_1, i_2)}$ and $(\mathbf{x})_j = x_j = x_{\overline{j j}} = (\mathbf{X})_{\overline{j j}}$. With the above notation, we have the following decomposition

$$\begin{aligned} x_{i_0 i_1, \dots, i_K}^* &= x_{i_0 i_1}^{(1)} x_{i_1 i_2}^{(1)} \dots x_{i_{K-1} i_K}^{(K)} \\ &= x_{\downarrow(i_0, i_1)}^{(1)} x_{\downarrow(i_1, i_2)}^{(2)} \dots x_{\downarrow(i_{K-1}, i_K)}^{(K)}. \end{aligned} \quad (4)$$

The K -th order affinity tensor $\mathcal{A} \in \mathbb{R}^{J \times J \times \dots \times J}$ as

$$a_{i_0 i_1, \dots, i_K}^* = \begin{cases} a_{j_1 j_2 \dots j_K}^*, & \text{if } j_k = \overline{j_{k+1}}, \forall k = 1, \dots, K-1 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Combining Eqn. 4 and 5, the MDA problem is reshaped as a new tensor-based MDA (TMDA) in the following form:

$$\arg \max_{\mathbb{X}} f(\mathbb{X}) = \mathcal{A} \times_1 \mathbf{x}^{(1)} \times_2 \mathbf{x}^{(2)} \dots \times_K \mathbf{x}^{(K)}, \quad (6)$$

$$s.t. \begin{cases} \sum_{i_{k-1}} x_{i_{k-1} i_k}^{(k)} = 1, & \forall k = 0, 1, \dots, K \\ \sum_{i_k} x_{i_{k-1} i_k}^{(k)} = 1, & \forall k = 0, 1, \dots, K \\ x_{i_{k-1} i_k}^{(k)} \in \{0, 1\} & \forall k = 0, 1, \dots, K \end{cases} \quad (7)$$

where $\mathbb{X} \doteq \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(K)}\}$ is the set of local assignment vectors and \times_k is the k -mode tensor product.

With the TMDA formulation, the problem can be viewed as to seek vectors \mathbb{X} to maximize the ‘‘joint projection’’ of \mathcal{A} on \mathbb{X} . This way TMDA closely correlates with the rank-1 tensor approximation [8], which aims to approximate a tensor by the tensor product of unit vectors up to a scale factor. By relaxing the integer constraint on the assignment variables, once a real valued solution of \mathbf{X} is achieved, it can be binarized using the Hungarian algorithm. The key issue here is to accommodate the row/column ℓ_1 constraints of Eqn. 7, which are different from the commonly used ℓ_2 norm constraint in tensor factorization. We use the approach developed in [18], which is a tensor power iteration solution with row/column normalization.

The key steps of power iteration are listed below,

$$\mathbf{x}^{(k)} \leftarrow (\mathcal{A} \times_1 \mathbf{x}^{(k)} \bullet \mathbf{x}^{(1)} \times_2 \mathbf{x}^{(2)} \dots \times_{k-1} \mathbf{x}^{(k-1)} \times_k \mathbf{x}^{(k+1)} \dots \times_K \mathbf{x}^{(K)}), \quad (8)$$

$$\mathbf{x}^{(k)} \leftarrow \text{rnorm}(\mathbf{x}^{(k)}), \quad (9)$$

where $\text{rnorm}(\cdot)$ indicates a row/column normalization.

3.3. Model Propagation

To solve the proposed TMDA problem for curvilinear structure tracking, local assignment matrix $\mathbf{X}^{(k)}$ between curvilinear structure sets $\mathbf{o}^{(k-1)}$ and $\mathbf{o}^{(k)}$ is consecutively computed by the power iteration algorithm. Bipartite

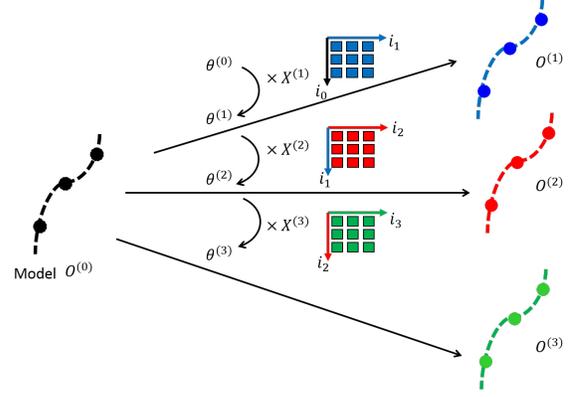


Figure 3. Model propagation example.

matching of $\mathbf{o}^{(k-1)}$ and $\mathbf{o}^{(k)}$ is obtained by Hungarian algorithm. In order to get the tracking result of curvilinear structure of $\mathbf{o}^{(0)}$ in other following frames. The straightforward method is to link these binary assignments together. Curvilinear structure $\mathbf{o}^{(0)}$ is then passed into the sequence by these binary assignments during K frames.

A key issue is to measure, before reaching final results, whether a candidate $\mathbf{o}_{i_k}^{(k)}$ matches a model element part $\mathbf{o}_{i_0}^{(0)}$. We address the issue by dynamically maintain a ‘‘soft assignment’’. In particular, we use $\theta_{i_0 i_k}^{(k)}$ to indicate the likelihood that $\mathbf{o}_{i_k}^{(k)}$ corresponds to $\mathbf{o}_{i_0}^{(0)}$, it can be estimated by

$$\Theta^{(k)} = \Theta^{(k-1)} \mathbf{X}^{(k)}, k = 1, 2, \dots, K, \quad (10)$$

where $\Theta^{(k)} = (\theta_{i_0 i_k}^{(k)}) \in \mathbb{R}^{I_0 I_k}$, and $\Theta^{(0)}$ is fixed as the identity matrix. An illustration example of model propagation is given in Fig. 3.

Together with local assignment matrix $\mathbf{X}^{(k)}$, the model likelihood $\Theta^{(k)}$ is then propagated through the sequence. Therefore, the model likelihood $\Theta^{(k)}$ is incorporated in the power iteration rank-1 tensor approximation for our curvilinear tracking problem. More specifically, the model propagation is updated in each step of the power iteration. The final tracking result of the matching between $\mathbf{o}^{(0)}$ and $\mathbf{o}^{(k)}$ is computed from $\Theta^{(k)}$. Therefore, the curvilinear structure model $\mathcal{CS}^{(k)}$ which is propagated to k -th frame can also be directly computed by $\Theta^{(k)}$. We summarize the proposed method in Algorithm 1.

4. High-Order Tensor Construction

Given a batch of $K+1$ frames from one X-ray sequence, the tensor is constructed from curvilinear structure candidates $\mathbf{O}^{(k)}$, where $k \in \{0, 1, \dots, K\}$. The model, $\mathbf{O}^{(0)}$ is manually annotated in the first image of the batch. Candidates on other frames are detected by a discriminative classifier.

Algorithm 1 Power iteration with model propagation

- 1: Input: Global affinity $\mathcal{A} : a_{j_1 \dots j_k \dots j_K}, k = 1 \dots K$, curvilinear structure candidates $O^{(k)}, k = 0 \dots K$.
 - 2: Output: curvilinear structure $CS^{(k)}, k = 1 \dots K$.
 - 3: Initialize $\mathbf{X}^{(k)}, k = 1 \dots K, CS^{(0)} = O^{(0)}$ and $\Theta^{(0)} = \mathbf{I}$.
 - 4: **repeat**
 - 5: **for** $k = 1, \dots, K$ **do**
 - 6: **for** $j_k = 1, \dots, J$ **do**
 - 7: update $x_{j_k}^{(k)} \doteq x_{i_{k-1}i_k}^{(k)}$ by
$$x_{j_k}^{(k)} \propto x_{j_k}^{(k)} \sum_{j_f: f \neq k}^K a_{j_1 \dots j_k \dots j_K} x_{j_1}^{(1)} \dots x_{j_f}^{(f)} \dots x_{j_K}^{(K)}$$
 - 8: **end for**
 - 9: row/column normalize $\mathbf{X}^{(k)}$
 - 10: update model likelihood: $\Theta^{(k)} = \Theta^{(k-1)} \mathbf{X}^{(k)}$
 - 11: **end for**
 - 12: **until** convergence
 - 13: **for** $k = 1, \dots, K$ **do**
 - 14: discretize $\Theta^{(k)}$ to get curvilinear structure $CS^{(k)}$.
 - 15: **end for**
-

4.1. Curvilinear Structure Candidates Detection

A random forest classifier of curvilinear structure is learned using only the first image of each sequence. Then the classifier is applied to find candidates in other frames.

Local features are proposed to capture the visual appearance of curvilinear structure. A few local features for each sampling point are extracted from the image. Local visual features such as intensity and gradient are fed into the learning process. For example, given a point centered at \mathbf{x} , the feature components are I, g_x, g_y and $\|g\|$. For each feature component f , the following values are computed as additional features: $f, \sqrt{\|f\|}, f^2, f^3$ and $\log(\|f\|)$. All the features are normalized by the variance of intensity values in the local patch.

For each image in the sequence except the first frame, a curvilinear structure probability map is computed by the learned classifier. A threshold is set to eliminated most of the false alarms in the image. These results are further processed by a thinning algorithm. Final curvilinear structure candidates are down-sampled from the thinning results. Down-sampling approach is applied to keep a small number of detection candidates, which is aimed to avoid building huge number of trajectories in tensor construction.

4.2. Trajectory Hypotheses

We follow the conventional method to build trajectory hypotheses based on the association of two adjacent frames [18]. The association between two consecutive frames is created by $O^{(k-1)}$ and $O^{(k)}$. Usually, it is unnecessary to

build the bipartite graph by linking all the candidates together. A link between $\mathbf{o}_{i_{k-1}}^{(k-1)}$ and $\mathbf{o}_{i_k}^{(k)}$ is only established when these two candidates are spatially close. A threshold is taken to eliminate candidates far away from each other, which in turn reduces the number of trajectory hypotheses in multi-frame batch.

As proposed in [6], a practical tracking algorithm must handle missing detections and variable numbers of targets exiting each frame. To address these issues, a virtual dummy node is added to each frame. This dummy node is linked to all the real curvilinear structure candidates in the previous frame and the next frame. A link is also built for dummy observations of two consecutive frames to process short-time missing or exiting candidates. These virtual nodes enlarge the number of potential trajectories.

4.3. Trajectory Affinity

The affinity of one trajectory hypothesis $O_{i_0, i_1, \dots, i_K} = \{\mathbf{o}_{i_0}^{(0)}, \mathbf{o}_{i_1}^{(1)}, \dots, \mathbf{o}_{i_K}^{(K)}\}$ is defined by

$$a_{i_0 i_1 \dots i_K}^* = app_{i_0 i_1 \dots i_K} \times kin_{i_0 i_1 \dots i_K}, \quad (11)$$

where $app_{i_0 i_1 \dots i_K}$ describes the appearance consistence and $kin_{i_0 i_1 \dots i_K}$ is to model the kinetic affinity. Essentially, $app_{i_0 i_1 \dots i_K}$ can be written as

$$app_{i_0 i_1 \dots i_K} = app_{i_0 i_1} \times app_{i_1 i_2} \times \dots \times app_{i_{K-1} i_K}, \quad (12)$$

where $app_{i_{k-1} i_k}$ is the appearance similarity between curvilinear structure candidates $\mathbf{o}_{i_{k-1}}^{(k-1)}$ and $\mathbf{o}_{i_k}^{(k)}$.

Similar as in [7, 18], the motion affinity of a trajectory O_{i_0, i_1, \dots, i_K} is defined as a joint assignment of item $O_{i_0, i_1, \dots, i_K} = \{\mathbf{o}_{i_0}^{(0)}, \mathbf{o}_{i_1}^{(1)}, \dots, \mathbf{o}_{i_K}^{(K)}\}$, $kin_{i_0 i_1 \dots i_K}$ is written as

$$kin_{i_0 i_1 \dots i_K} = \exp(-\alpha E_{cont} - \beta E_{curv}), \quad (13)$$

where

$$\begin{cases} E_{cont} &= \frac{1}{K} \sum_{k=1}^{K+1} \|\mathbf{o}_{i_k}^{(k)} - \mathbf{o}_{i_{k-1}}^{(k-1)}\| \\ E_{curv} &= \frac{1}{K-1} \sum_{k=1}^{K+1} \|\mathbf{o}_{i_{k+1}}^{(k+1)} - 2\mathbf{o}_{i_k}^{(k)} + \mathbf{o}_{i_{k-1}}^{(k-1)}\| \end{cases} \quad (14)$$

Continuous term E_{cont} penalizes large movement of curvilinear structure in position using the average distance between successive pairs of points $\mathbf{o}_{i_k}^{(k)}$ and $\mathbf{o}_{i_{k-1}}^{(k-1)}$. E_{curv} is the curvature measurement computed by the sum of curvature segments over the trajectory. The motion consistence of the trajectory is a variant of internal energy term of the popular ‘‘snake’’ active contour. Parameters α, β are weights of E_{cont} and E_{curv} .

5. Experiments

To investigate the performance of the proposed approach for curvilinear structure deformable motion tracking, we use X-ray images from clinical cases of liver and cardiac intervention. The data is acquired with pixel size of 512×512 and resolution of 0.4313mm, 0.3450mm or 0.3660mm. These cases are chosen due to the visibility of the vascular structure and catheter throughout the entire image, which provide groundtruth for evaluation (as shown in even columns of Fig. 1).

These X-ray sequences are categorized into two groups. One group contains 6 sequences for vascular structure tracking. 11 sequences in the second group are used for catheter tracking. The results of proposed method are compared to registration-based approach and bipartite graph matching.

5.1. Vascular Structure Tracking

Our first experiment is conducted on vascular structure tracking in X-ray image sequences. There are 6 X-ray image sequences which contain vascular structures in all the frames.

Vascular Structure Candidates Detection. The first frames from these sequences are used to generate training samples. Random forest is applied to train a vascular structure detector. An independent training sample set is collected for each tree in the forest. From annotation, 1000 vascular structure points and 1000 negative samples are generated for each image. There are totally $2000 \times 6 \times 2 = 24,000$ training samples to train a decision tree in the forest. Due to the large width of vascular structure, the sampling pattern for local features descriptor is 30×30 pixels, which provides a feature vector with dimension 18,000. A random forest of 50 trees with depth 10 is constructed in the experiment. In the node optimization, 100 active features are randomly sampled from the whole feature space. Stump is chosen in the node optimization.

The vascular structure probability map of X-ray images could be obtained by the learned classifier. Pixels with probability value larger than 0.5 are selected for post-processing. After thinning and down-sampling, around 200 vascular structure candidates left in each frame. Number of points on the model is around 50 for each sequence.

Tensor Parameters. In our experiment, we investigate the case when $K = 3$ which provides 4 frames in the batch. Curvilinear structure model is annotated in the first frame. In order to capture the possible large movement of curvilinear structure, the distance threshold to establish the links between two consecutive frames is 25 pixels. Shape context feature is used to represent the appearance of vascular structure. Appearance similarity $app_{i_{k-1}i_k}$ is calculated by shape context feature of $\mathbf{o}_{i_{k-1}}^{k-1}$ and $\mathbf{o}_{i_k}^k$. The parameter of

$kin_{i_0i_1,\dots,i_K}$ is set to $\alpha = 0.01, \beta = 0.01$ in our experiments.

Tracking Results. Using the proposed method, curvilinear structures model $CS^{(0)}$ is propagated to its consecutive frames. Curvilinear structure $CS^{(k)}$ in the k -th frame is computed by $\Theta^{(k)}$. A B-Spline is then fitted to represent the vascular structure in X-ray images. For these selected sequences, we manually annotate the vascular structure on the X-ray images. Tracking error is defined as the shortest distance between tracked pixels and groundtruth annotations. For each performance metric, we compute its mean and standard deviation.

The proposed method is compared with registration-based approach [24]. In their method, the first image is fixed as the reference image. Registration is conducted between other floating frames and the reference image. Two-level pyramid is applied in the registration-based method to recover large motion of curvilinear structure. The number of iterations for the first level of the pyramid is set to 16 and the second to 24.

Solutions calculated by bipartite graph matching is also compared with our method. We use the same affinity formulation for fair comparison. Hungarian algorithm is applied to compute the bipartite matching. Curvilinear structure $O^{(0)}$ is propagated to $CS^{(k)}$ by bipartite graph matching. The final curvilinear structure is represented by a B-Spline fitted by $CS^{(k)}$.

An example of vascular structure tracking results are illustrated in the first row of Fig. 4. The qualitative experimental results of vascular structure tracking are listed in Table 1. It can be seen that our approach performs better than the registration-based method. The tracking results are better than most of the results computed by bipartite graph matching. The reason is that our proposed method incorporates higher-order information from multiple frames. However, the bipartite graph matching is only computed from 2 frames.

5.2. Catheter Tracking

Our proposed approach is also evaluated on the second group for catheter tracking. We have 11 sequences for catheter tracking.

Catheter Candidates Detection. Similar as vascular structure detection, training samples are generated from 11 images which are the first images in each sequence. 1000 catheter points and 1000 negative samples are generated from each image. $1000 \times 11 \times 2 = 22,000$ training samples are used to train a decision tree in the forest. The feature space has a dimension of 2000 where sampling pattern for local features descriptor is 10×10 pixels. The same parameter of random forest as in vascular structure experiment is set for training a catheter detector.

Table 1. Tracking error of vascular structure.

seq id	registration [24]	bipartite matching	proposed
VAS1	2.7656 ± 3.2464	1.5397 ± 1.5850	1.3346 ± 1.0761
VAS2	2.0180 ± 3.1024	1.4885 ± 1.5743	1.4926 ± 1.7403
VAS3	3.2530 ± 7.6413	1.6519 ± 2.3965	1.4089 ± 1.5470
VAS4	2.1581 ± 2.5227	1.6069 ± 2.2493	1.9900 ± 3.0176
VAS5	3.0399 ± 5.4569	2.7115 ± 4.3623	1.3642 ± 1.4405
VAS6	2.8646 ± 5.5972	1.4048 ± 1.9433	1.3159 ± 1.6788
Over all	2.6881 ± 5.0251	1.7498 ± 2.6012	1.4898 ± 1.8598

The probability map threshold is chosen to 0.5. Consequently, around 200 catheter candidates are kept for each frame. Model in the first frame is annotated to have around 50 points on it.

Tensor Parameters. K is set to 3 in our proposed method which gives a batch with 4 frames. The parameter of $kin_{i_0 i_1, \dots, i_K}$ is set to $\alpha = 0.001, \beta = 0.01$ in our experiments. Other parameters are the same as vascular structure tracking.

Tracking Results. Fig. 4 shows several examples of catheter structure tracking results. We use the same evaluation metric for catheter tracking. Registration-based method [24] and bipartite graph matching are also compared with our method. The comparison results of catheter tracking are listed in Table 2. We can observe that our higher-order tensor-based method outperforms other two approaches.

6. Conclusion and Discussion

In this paper, we model the curvilinear structure tracking as a tensor-based multiple data association problem with model propagation. Rank-1 tensor approximation is used to solve the optimization problem. More specifically, ℓ_1 unit norm tensor power iteration is applied in the tensor decomposition. Our model propagation automatically propagates curvilinear structure to consecutive frames in the sequence. Moreover, curvilinear structure can be obtained directly from model propagation without solving the two adjacent frame matching in standard tensor approximation. Experiments on two groups of curvilinear structure demonstrate the effectiveness of our proposed approach.

Acknowledgement. This work is supported in part by the NSF Grant IIS-1218156 and the NSF CAREER Award IIS-1350521.

References

- [1] S. Baert, M. Viergever, and W. Niessen. Guide-wire tracking during endovascular interventions. *IEEE T. on Medical Imaging*, 22(8):965–972, 2003. 1
- [2] B. Benfold and I. Reid. Stable multi-target tracking in real-time surveillance video. In *CVPR*, 3457–3464. 2011. 2
- [3] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. *Pattern Analysis and Machine Intelligence, IEEE T. on*, 33(9):1806–1819, 2011. 2
- [4] R. Bhagalia, J. Fessler, and B. Kim. Accelerated nonrigid intensity-based image registration using importance sampling. *IEEE T. on Medical Imaging*, 28(8):1208–1216, 2009. 2
- [5] W. Brendel, M. Amer, and S. Todorovic. Multiobject tracking as maximum weight independent set. In *CVPR*, 2011. 2
- [6] A. A. Butt and R. T. Collins. Multiple target tracking using frame triplets. In *ACCV*, 163–176, 2013. 5
- [7] R. T. Collins. Multitarget data association with higher-order motion models. In *CVPR*, 1744–1751, 2012. 3, 5
- [8] L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank-(r_1, r_2, \dots, r_n) approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342, 2000. 4
- [9] S. Deb, M. Yeddanapudi, K. Pattipati, and Y. Bar-Shalom. A generalized sd assignment algorithm for multisensor-multitarget state estimation. *Aerospace and Electronic Systems, IEEE T. on*, 33(2):523–538, 1997. 2
- [10] G. Hermosillo, C. ChefdHotel, and O.D. Faugeras. Variational Methods for Multimodal Image Matching. *IJCV*, 50(3):329–343, 2002. 2
- [11] C. Huang, B. Wu, and R. Nevatia. Robust object tracking by hierarchical association of detection responses. In *ECCV*, 788–801, 2008. 2
- [12] H. Jiang, S. Fels, and J. J. Little. A linear programming approach for multiple object tracking. In *CVPR*, 1–8, 2007. 2
- [13] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int'l J. of Computer Vision*, 1(4):321–331, 1988. 1
- [14] S. Oh, S. Russell, and S. Sastry. Markov chain monte carlo data association for multi-target tracking. *Automatic Control, IEEE T. on*, 54(3):481–497, 2009. 2
- [15] D. Palti-Wasserman, A. Brukstein, and R. Beyar. Identifying and tracking a guide wire in the coronary arteries during angioplasty from x-ray images. *Biomedical Engineering, IEEE T. on*, 44(2):152–164, 1997. 1
- [16] D. Reid. An algorithm for tracking multiple targets. *Automatic Control, IEEE T. on*, 24(6):843–854, 1979. 2
- [17] K. Shafique, M. W. Lee, and N. Haering. A rank constrained continuous formulation of multi-frame multi-target tracking problem. In *CVPR*, 2008. 2

Table 2. Tracking error of catheter.

seq id	registration [24]	bipartite matching	proposed
CAT1	2.8600 ± 3.8342	1.4729 ± 1.5674	1.2890 ± 1.0618
CAT2	1.9765 ± 2.6563	2.3793 ± 5.3322	1.1092 ± 1.5754
CAT3	2.1994 ± 1.5601	1.5535 ± 1.9803	1.3851 ± 1.7042
CAT4	1.0693 ± 0.7569	2.1171 ± 3.3528	1.1484 ± 1.3302
CAT5	2.5391 ± 3.6477	2.0157 ± 4.8478	1.0425 ± 0.8827
CAT6	1.9331 ± 2.1498	2.0575 ± 3.9163	1.1443 ± 0.9488
CAT7	1.3869 ± 2.1759	1.8595 ± 3.7920	1.0013 ± 0.7770
CAT8	2.7393 ± 4.3188	2.3034 ± 5.5272	1.3081 ± 2.2144
CAT9	1.7425 ± 1.2486	2.7979 ± 4.7828	2.0046 ± 2.7392
CAT10	3.1674 ± 5.2587	2.8582 ± 4.3332	2.4766 ± 3.5924
CAT11	3.9647 ± 5.8903	2.6815 ± 4.3589	1.1681 ± 0.9748
Over all	2.3995 ± 3.6241	2.1718 ± 4.1384	1.3782 ± 1.8978

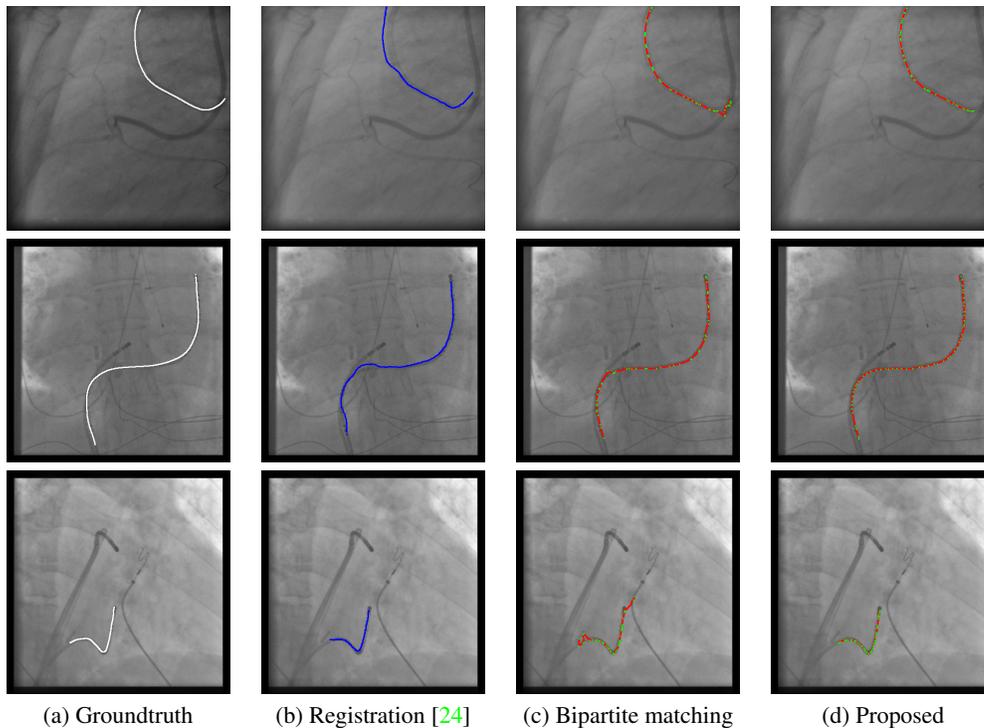


Figure 4. Curvilinear structure tracking results. An example of a vascular structure is displayed in the first row. Other rows give tracking results of catheters. (a) Annotations of curvilinear structure. (b), (c) and (d) provide curvilinear structure tracking results of registration-based method [24], bipartite matching and our proposed method. For bipartite matching and our method, green dot represents propagated model points. Red line is the fitting result of B-Spline of model points.

- [18] X. Shi, H. Ling, J. Xing, and W. Hu. Multi-target tracking by rank-1 tensor approximation. In *CVPR*, 2013. **2, 3, 4, 5**
- [19] P. Wang, T. Chen, Y. Zhu, W. Zhang, S. Zhou, and D. Comaniciu. Robust guidewire tracking in fluoroscopy. In *CVPR*, 2009. **2**
- [20] Z. Wu, N. I. Hristov, T. L. Hedrick, T. H. Kunz, and M. Betke. Tracking a large number of objects from multiple views. In *ICCV*, 1546–1553, 2009. **2**
- [21] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *Acm Computing Surveys*, 38(4):13, 2006. **1**
- [22] A. R. Zamir, A. Dehghan, and M. Shah. Gmcp-tracker: global multi-object tracking using generalized minimum clique graphs. In *ECCV*, 343–356, 2012. **2**
- [23] W. Zhang and H. Ling and S. Zhou and D. Comaniciu. Coronary Tree Extraction Using Motion Layer Separation. In *MICCAI*, 116–123, 2009. **2**
- [24] Y. Zhu, Y. Tsin, H. Sundar, and F. Sauer. Image-based respiratory motion compensation for fluoroscopic coronary roadmapping. In *MICCAI*, 287–294, 2010. **2, 6, 7, 8**