

Structure-Aware Rank-1 Tensor Approximation for Curvilinear Structure Tracking Using Learned Hierarchical Features

Peng Chu¹, Yu Pang¹, Erkang Cheng¹, Ying Zhu², Yefeng Zheng³,
and Haibin Ling¹(✉)

¹ Computer and Information Sciences Department,
Temple University, Philadelphia, USA
hbling@temple.edu

² Electrical and Computer Engineering Department,
Temple University, Philadelphia, USA

³ Medical Imaging Technologies, Siemens Healthcare, Princeton, USA

Abstract. Tracking of *curvilinear structures* (CS), such as vessels and catheters, in X-ray images has become increasingly important in recent interventional applications. However, CS is often barely visible in low-dose X-ray due to overlay of multiple 3D objects in a 2D projection, making robust and accurate tracking of CS very difficult. To address this challenge, we propose a new tracking method that encodes the structure prior of CS in the *rank-1 tensor approximation* tracking framework, and it also uses the learned hierarchical features via a *convolutional neural network* (CNN). The three components, i.e., curvilinear prior modeling, high-order information encoding and automatic feature learning, together enable our algorithm to reduce the ambiguity rising from the complex background, and consequently improve the tracking robustness. Our proposed approach is tested on two sets of X-ray fluoroscopic sequences including vascular structures and catheters, respectively. In the tests our approach achieves a mean tracking error of 1.1 pixels for vascular structure and 0.8 pixels for catheter tracking, significantly outperforming state-of-the-art solutions on both datasets.

1 Introduction

Reliable tracking of vascular structures or intravascular devices in dynamic X-ray images is essential for guidance during interventional procedures and postprocedural analysis [1–3, 8, 13, 14]. However, bad tissue contrast due to low radiation dose and lack of depth information always bring challenges on detecting and tracking those *curvilinear structures* (CS). Traditional registration and alignment-based trackers depend on local image intensity or gradient. Without high-level context information, they cannot efficiently discriminate low-contrasted target structure from complex background. On the other hand, the confounding irrelevant structures bring challenges to detection-based tracking. Recently, a new solution is proposed that exploits the progress in multi-target

tracking [2]. After initially detecting candidate points on a CS, the idea is to model CS tracking as a *multi-dimensional assignment* (MDA) problem, then a tensor approximation is applied to search for a solution. The idea encodes high-order temporal information and hence gains robustness against local ambiguity. However, it suffers from the lack of mechanism to encode the structure prior in CS, and the features used in [2] via random forests lack discrimination power.

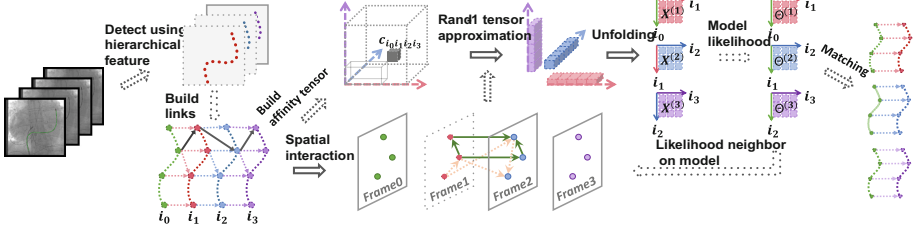


Fig. 1. Overview of the proposed method.

In this paper, we present a new method (refer to Fig. 1 for the flowchart) to detect and track CS in dynamic X-ray sequences. First, a *convolutional neural network* (CNN) is used to detect candidate landmarks on CS. CNN automatically learns the hierarchical representations of input images [6, 7] and has been recently used in medical image analysis (e.g. [9, 10]). With the detected CS candidates, CS tracking is converted to a multiple target tracking problem and then a *multi-dimensional assignment* (MDA) one. In MDA, candidates are associated along motion trajectories cross time, while the association is constructed according to the trajectory affinity. It has been shown in [11] that MDA can be efficiently solved via *rank-1 tensor approximation* (RITA), in which the goal is to seek vectors to maximize the “joint projection” of an affinity tensor. Sharing the similar procedure, our solution adopts RITA to estimate the CS motion. Specifically, a high-order tensor is first constructed from all trajectory candidates over a time span. Then, the model prior of CS is integrated into RITA encoding the spatial interaction between adjacent candidates in the model. Finally, CS tracking results are inferred from model likelihood.

The main contribution of our work lies in two-fold. (1) We propose a structure-aware tensor approximation framework for CS tracking by considering the spatial interaction between CS components. The combination of such spatial interaction and higher order temporal information effectively reduces association ambiguity and hence improves the tracking robustness. (2) We design a discriminative CNN detector for CS candidate detection. Compared with traditional hand-crafted features, the learned CNN features show very high detection quality in identifying CS from low-visibility dynamic X-ray images. As a result, it greatly reduces the number of hypothesis trajectories and improves the tracking efficiency.

For evaluation, our method is tested on two sets of X-ray fluoroscopic sequences including vascular structures and catheters, respectively. Our approach achieves a mean tracking error of 1.1 pixels on the vascular dataset and 0.8 pixels on the catheter dataset. Both results are clearly better than other state-of-the-art solutions in comparison.

2 Candidate Detection with Hierarchical Features

Detecting CS in the low-visibility dynamic X-ray images is challenging. Without color and depth information, CS shares great similarity with other anatomical structures or imaging noise. Attacking these problems, a four-layer CNN (Fig. 2) is designed to automatically learn hierarchical features for CS candidate detection. We employ 32 filters of size 5×5 in the first convolution stage, and 64 filters of the same size in the second stage. Max-pooling layers with a receptive window of 2×2 pixels are employed to down-sample the feature maps. Finally, two fully-connected layers are used as the classifier. Dropout is employed to reduce overfitting. The CNN framework used in our experiments is based on MatConvNet [12].

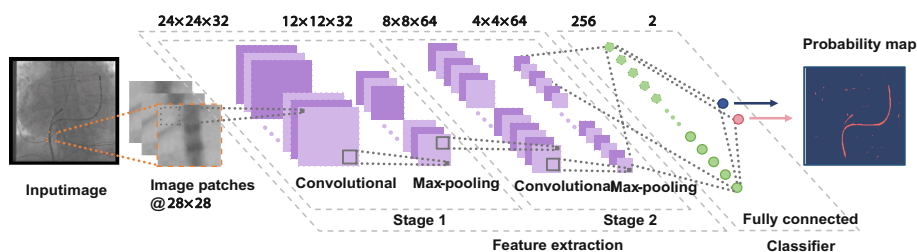


Fig. 2. The CNN architecture for CS candidate detection.

For each image in the sequence except the first one which has groundtruth annotated manually, a CS probability map is computed by the learned classifier. A threshold is set to eliminate most of the false alarms in the image. Result images are further processed by filtering and thinning. Typically, binarized probability map is filtered by a distance mask in which locations too far from the model are excluded. Instead of using a groundtruth bounding box, we take the tracking results from previous image batches. Based on the previously tracked model, we calculate the speed and acceleration of the target to predict its position in next image batch. Finally, after removing isolated pixels, CS candidates are generated from the thinning results. Examples of detection results are shown in Fig. 3. For comparison, probability maps obtained by a random forests classifier with hand-crafted features [2] are also listed. Our probability maps contain less false alarm, which guarantees more accurate candidate locations after post-processing.

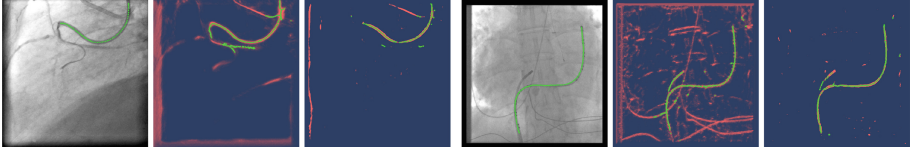


Fig. 3. Probability maps and detected candidates of a vessel (left) and catheter (right). For each example, from left to right are groundtruth, random forests result, and CNN result, respectively. Red indicates region with high possibility, while green dots show resulting candidates.

3 Tracking with Model Prior

To encode the structure prior in a CS model, we use an energy maximization scheme that combines temporal energy of individual candidate and spatial interaction energy of multiple candidates into a united optimization framework. Here, we consider the pairwise interactions of two candidates on neighboring frames. The assignment matrix between two consecutive sets $O^{(k-1)}$ and $O^{(k)}$ (i.e. detected candidate CS landmarks) can be written as $\mathbf{X}^{(k)} = (x_{i_{k-1}i_k})^{(k)}$, where $k = 1, 2, \dots, K$, and $\mathbf{o}_{i_k}^{(k)} \in O^{(k)}$ is the i_k -th landmark candidate of CS. For notation convenience, we use a single subscript j_k to represent the entry index (i_{k-1}, i_k) , such as $x_{j_k}^{(k)} \doteq x_{i_{k-1}i_k}^{(k)}$, i.e., $\text{vec}(\mathbf{X}^{(k)}) = (x_{j_k}^{(k)})$ for vectorized $\mathbf{X}^{(k)}$. Then our objective function can be written as

$$f(\mathcal{X}) = \sum c_{j_1j_2\dots j_K} x_{j_1}^{(1)} x_{j_2}^{(2)} \dots x_{j_K}^{(K)} + \sum_{k=1}^K \sum_{l_k, j_k} w_{l_kj_k}^{(k)} e_{l_kj_k}^{(k)} x_{l_k}^{(k)} x_{j_k}^{(k)}, \quad (1)$$

where $c_{j_1j_2\dots j_K}$ is the affinity measuring trajectory confidence; $w_{l_kj_k}^{(k)}$ the likelihood that candidates $x_{j_k}^{(k)}$ and $x_{l_k}^{(k)}$ are neighboring on the model; and $e_{l_kj_k}^{(k)}$ the spatial interaction of two candidates on two consecutive frames. The affinity has two parts as

$$c_{i_0i_1, \dots, i_K} = \text{app}_{i_0i_1, \dots, i_K} \times \text{kin}_{i_0i_1, \dots, i_K}, \quad (2)$$

where $\text{app}_{i_0i_1, \dots, i_K}$ describes the appearance consistency of the trajectory, and $\text{kin}_{i_0i_1, \dots, i_K}$ the kinetic affinity modeling the higher order temporal affinity as detailed in [2].

Model Prior. CS candidates share two kinds of spatial constrains. First, trajectories of two neighboring elements should have similar direction. Second, relative order of two neighboring elements should not change so that re-composition of CS is prohibited. Thus inspired, we formulate the spatial interaction of two candidates as

$$e_{l_kj_k} \doteq e_{m_{k-1}m_k i_{k-1}i_k} = E_{\text{para}} + E_{\text{order}}, \quad (3)$$

where

$$E_{para} = \frac{(\mathbf{o}_{i_{k-1}}^{(k-1)} - \mathbf{o}_{i_k}^{(k)}) \cdot (\mathbf{o}_{m_{k-1}}^{(k-1)} - \mathbf{o}_{m_k}^{(k)})}{\|(\mathbf{o}_{i_{k-1}}^{(k-1)} - \mathbf{o}_{i_k}^{(k)}) \cdot (\mathbf{o}_{m_{k-1}}^{(k-1)} - \mathbf{o}_{m_k}^{(k)})\|}, \quad E_{order} = \frac{(\mathbf{o}_{i_{k-1}}^{(k-1)} - \mathbf{o}_{m_{k-1}}^{(k-1)}) \cdot (\mathbf{o}_{i_k}^{(k)} - \mathbf{o}_{m_k}^{(k)})}{\|(\mathbf{o}_{i_{k-1}}^{(k-1)} - \mathbf{o}_{m_{k-1}}^{(k-1)}) \cdot (\mathbf{o}_{i_k}^{(k)} - \mathbf{o}_{m_k}^{(k)})\|},$$

such that E_{para} models the angle between two neighbor trajectories, which also penalizes large distance change between them; and E_{order} models the relative order of two adjacent candidates by the inner product of vectors between two neighbor candidates.

Maximizing Eq. 1 closely correlates with the *rank-1 tensor approximation* (R1TA) [4], 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}^k is achieved, it can be binarized using the Hungarian algorithm [5]. The key issue here is to accommodate the row/column ℓ_1 normalization in a general assignment problem, which is different from the commonly used ℓ_2 norm constraint in tensor factorization. We develop an approach similar to [11], which is a tensor power iteration solution with ℓ_1 row/column normalization.

Model Likelihood. Coefficient $w_{l_k j_k}^{(k)} \doteq w_{m_{k-1} m_k i_{k-1} i_k}^{(k)}$ measures the likelihood that two candidates $\mathbf{o}_{i_{k-1}}^{(k-1)}$ and $\mathbf{o}_{m_{k-1}}^{(k-1)}$ are neighboring on model. In order to get the association of each candidate pair in each frame, or in other words, to measure the likelihood a candidate $\mathbf{o}_{i_k}^{(k)}$ matching a model element part $\mathbf{o}_{i_0}^{(0)}$, we 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

$$\Theta^{(k)} = \Theta^{(k-1)} \mathbf{X}^{(k)}, k = 1, 2, \dots, K, \quad (4)$$

where $\Theta^{(k)} = (\theta_{i_0 i_k}^{(k)}) \in \mathbb{R}^{I_0 \times I_k}$ and $\Theta^{(0)}$ is fixed as the identity matrix.

The model likelihood is updated in each step of the power iteration. After the update of the first term in Eq. 1, a pre-likelihood $\Theta'^{(k)}$ is estimated for computing $w_{l_k j_k}^{(k)}$. Since $\Theta^{(k)}$ associates candidates directly with the model, final tracking result of the matching between $\mathbf{o}^{(0)}$ and $\mathbf{o}^{(k)}$ can be derived from $\Theta^{(k)}$.

With $\Theta'^{(k)}$, the approximated distance on model of $\mathbf{o}_{i_{k-1}}^{(k-1)}$ and $\mathbf{o}_{m_{k-1}}^{(k-1)}$ can be calculated as following

$$d_{i_k m_k}^{(k)} = \frac{\sum_{i_0} \|(\mathbf{o}_{i_0}^{(0)} - \mathbf{o}_{i_0+1}^{(0)})\| \theta_{i_0 i_k}^{(k)} \theta_{i_0+1 m_k}^{(k)}}{\sum_{i_0} \theta_{i_0 i_k}^{(k)} \theta_{i_0+1 m_k}^{(k)}}. \quad (5)$$

Thereby, $w_{l_k j_k}^{(k)}$ then can be simply calculated as

$$w_{l_k j_k}^{(k)} \doteq w_{m_{k-1} m_k i_{k-1} i_k}^{(k)} = \frac{2d_{i_{k-1} m_{k-1}}^{(k-1)} \bar{d}}{(d_{i_{k-1} m_{k-1}}^{(k-1)})^2 + (\bar{d})^2}, \quad (6)$$

where \bar{d} is the average distance between two neighboring elements on model $O^{(0)}$. The proposed tracking method is summarized in Algorithm 1.

Algorithm 1. Power iteration with model prior

-
- 1: **Input:** Global affinity $\mathcal{C} = (c_{j_1 j_2 \dots j_K})$, spatial interaction $e_{l_k j_k}$, $k = 1 \dots K$, and CS candidates $O^{(k)}$, $k = 0 \dots K$.
 - 2: **Output:** CS Matching.
 - 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 $x_{j_k}^{(k)} \propto x_{j_k}^{(k)} \sum_{j_f: f \neq k} c_{j_1 \dots j_k \dots j_K} x_{j_1}^{(1)} \dots x_{j_f}^{(f)} \dots x_{j_K}^{(K)}$
 - 8: **end for**
 - 9: row/column normalize $\mathbf{X}^{(k)}$
 - 10: update model pre-likelihood: $\Theta'^{(k)} = \Theta^{(k-1)} \mathbf{X}^{(k)}$
 - 11: **for** $j_k = 1, \dots, J$ **do**
 - 12: update $x_{j_k} \doteq x_{i_{k-1} i_k}^{(k)}$ by $x_{j_k}^{(k)} \propto x_{j_k}^{(k)} \sum_{l_k} w_{l_k j_k}^{(k)} e_{l_k j_k} x_{l_k}^{(k)}$
 - 13: **end for**
 - 14: update model likelihood: $\Theta^{(k)} = \Theta'^{(k)} \mathbf{X}^{(k)}$
 - 15: **end for**
 - 16: **until** convergence
 - 17: discretize $\Theta^{(k)}$ to CS Matching.
-

4 Experiments

We evaluate the proposed CS tracking algorithm using two groups of X-ray clinical data collected from liver and cardiac interventions. The first group consists of six sequences of liver vessel images and the second 11 sequences of catheter images, each with around 20 frames. The data is acquired with 512×512 pixels and physical resolution of 0.345 or 0.366 mm. Groundtruth of each image is manually annotated (Fig. 4(a)).

Vascular Structure Tracking. We first evaluate the proposed algorithm on the vascular sequences. First frame from each sequence is used to generate training samples for CNN. To be specific, 800 vascular structure patches and 1500 negative patches are generated from each image. From the six images, a total of $2300 \times 6 = 13,800$ samples are extracted and split as 75% training and 25% validation. All patches have the same size of 28×28 pixels. Distance threshold of predictive bounding box is set to 60 pixels for enough error tolerance. Finally, there are around 200 vascular structure candidates left in each frame. The number of points on the model is around 50 for each sequence.

In our work, $K = 3$ is used to allow each four consecutive frames to be associated. During tracking, tensor kernel costs around 10s and 100MB (peak value) RAM to process one frame with 200 candidates in our setting running on a single Intel Xeon@2.3GHz core. The tracking error is defined as the shortest distance between tracked pixels and groundtruth annotation. For each performance metric, we compute its mean and standard deviation. For comparison, the registration-based (RG) approach [14], bipartite graph matching [2] (BM) and pure tensor based method [2] (TB) are applied to the same sequences. For BM

and TB, same tracking algorithms but with the CNN detector are also tested and reported. The first block of Fig. 4 illustrates the tracking results of vascular structures. B-spline is used to connect all tracked candidates to represent the tracked vascular structure. The zoom-in view of a selected region (rectangle in blue) in each tracking result is presented below, where portions with large errors are colored red. Quantitative evaluation for each sequence is listed in Table 1.

Catheter Tracking. Similar procedures and parameters are applied to the 11 sequences of catheter images. The second block of Fig. 4 shows example of catheter tracking results. The numerical comparisons are listed in Table 1.

The results show that our method clearly outperforms other three approaches. Candidates in our approach are detected by a highly accurate CNN detector, ensuring most extracted candidates to be on CS, while registration-based method depends on the first frame as reference to identify targets. Our approach is also better than the results of bipartite graph matching where $K = 1$. The reason is that our proposed method incorporates higher-order temporal information from multiple frames; by contrast, bipartite matching is only computed from two frames. Compared with the pure tensor based algorithm, the proposed method incorporates the model prior which provides more powerful

Table 1. Curvilinear structure tracking errors (in pixels)

Dataset	Seq ID	RG [14]	BM [2]	TB [2]	BM+CNN	TB+CNN	Proposed
Vascular structures	VAS1	2.77 ± 3.25	1.54 ± 1.59	1.33 ± 1.08	1.44 ± 2.37	1.15 ± 0.91	1.14 ± 0.84
	VAS2	2.02 ± 3.10	1.49 ± 1.14	1.49 ± 1.74	1.11 ± 0.83	1.30 ± 2.48	1.09 ± 0.83
	VAS3	3.25 ± 7.64	1.65 ± 2.40	1.41 ± 1.54	1.19 ± 0.91	1.17 ± 0.92	1.17 ± 0.91
	VAS4	2.16 ± 2.52	1.61 ± 2.25	1.99 ± 3.02	1.12 ± 1.00	1.95 ± 5.00	1.17 ± 1.53
	VAS5	3.04 ± 5.46	2.71 ± 4.36	1.36 ± 1.44	1.95 ± 3.94	1.14 ± 1.55	1.09 ± 1.42
	VAS6	2.86 ± 5.60	1.40 ± 1.94	1.32 ± 1.68	1.39 ± 2.53	1.09 ± 1.70	1.11 ± 1.90
	75 %ile, 100 %ile	-	2.00, 31.2	2.00, 26.8	1.40, 32.6	1.40, 56.9	1.40, 23.2
	Overall	2.69 ± 5.03	1.75 ± 2.60	1.49 ± 1.86	1.37 ± 2.26	1.30 ± 2.64	1.13 ± 1.30
Catheters	CAT1	2.86 ± 3.83	1.47 ± 1.57	1.29 ± 1.06	1.13 ± 1.19	1.08 ± 0.85	1.00 ± 0.77
	CAT2	1.98 ± 2.66	2.38 ± 5.33	1.11 ± 1.58	1.77 ± 4.11	0.77 ± 1.06	0.56 ± 0.89
	CAT3	2.20 ± 1.56	1.55 ± 1.98	1.39 ± 1.70	0.99 ± 1.52	0.72 ± 0.66	0.74 ± 0.65
	CAT4	1.07 ± 0.76	2.12 ± 3.35	1.15 ± 1.33	0.94 ± 1.37	0.92 ± 1.34	0.76 ± 0.77
	CAT5	2.54 ± 3.65	2.02 ± 4.85	1.04 ± 0.88	1.65 ± 5.36	0.84 ± 1.01	0.83 ± 0.97
	CAT6	1.93 ± 2.15	2.06 ± 3.92	1.14 ± 0.95	1.19 ± 2.03	0.96 ± 0.92	0.93 ± 0.89
	CAT7	1.39 ± 2.18	1.86 ± 3.79	1.00 ± 0.78	0.76 ± 0.72	0.76 ± 0.72	0.73 ± 0.63
	CAT8	2.74 ± 4.32	2.30 ± 5.53	1.31 ± 2.21	1.22 ± 2.21	1.74 ± 3.81	0.96 ± 1.37
	CAT9	1.74 ± 1.25	2.80 ± 4.78	2.00 ± 2.74	1.54 ± 3.44	1.18 ± 2.02	0.99 ± 1.33
	CAT10	3.17 ± 5.26	2.86 ± 4.33	2.48 ± 3.59	0.86 ± 1.26	0.81 ± 1.12	0.86 ± 1.29
	CAT11	3.96 ± 5.89	2.68 ± 4.36	1.17 ± 0.97	3.50 ± 11.3	1.35 ± 3.72	0.80 ± 0.74
	75 %ile, 100 %ile	-	2.00, 47.7	1.40, 24.0	1.00, 70.5	1.00, 48.4	1.00, 19.2
	Overall	2.40 ± 3.62	2.17 ± 4.14	1.38 ± 1.90	1.39 ± 4.16	1.01 ± 1.93	0.83 ± 0.98

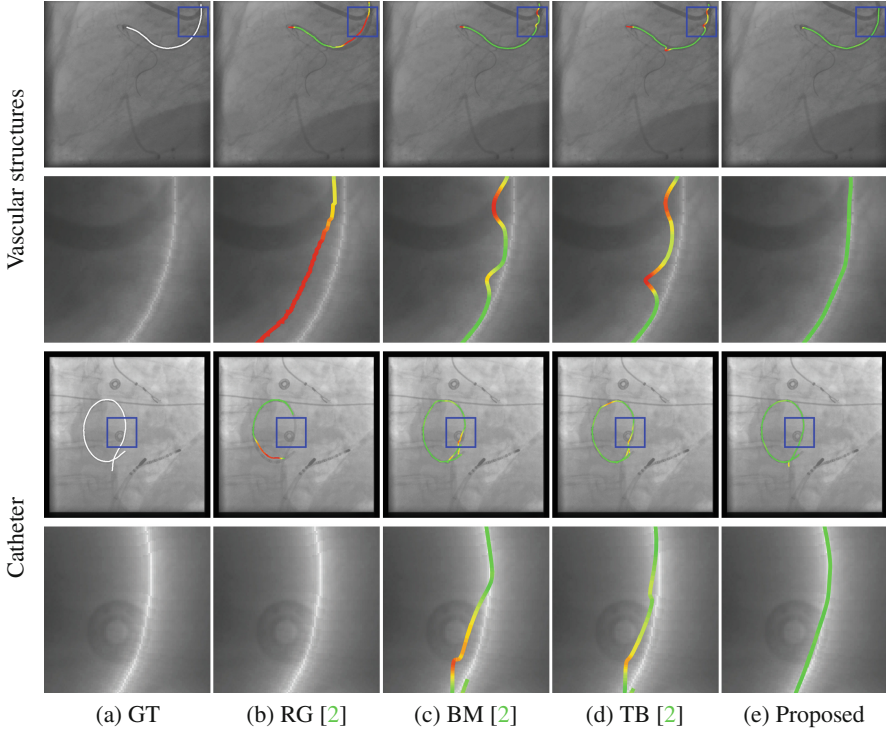


Fig. 4. Curvilinear structure tracking results. (a) groundtruth, (b) registration, (c) bipartite matching, (d) tensor based, and (e) proposed method. Red indicates regions with large errors, while green indicates small errors.

clues for tracking the whole CS. Confirmed by the zoom-in views, with model prior, our proposed method is less affected by neighboring confounding structures.

5 Conclusion

We presented a new method to combine hierarchical features learned in CNN and encode model prior to estimate the motion of CS in X-ray image sequences. Experiments on two groups of CS demonstrate the effectiveness of our proposed approach. Achieving a tracking error of around one pixel (or smaller than 0.5 mm), it clearly outperforms the other state-of-the-art algorithms. For future work, we plan to adopt pyramid detection strategy in order to accelerate the pixel-wised probability map calculation in our current approach.

Acknowledgement. We thank the anonymous reviewers for valuable suggestions. This work was supported in part by NSF grants IIS-1407156 and IIS-1350521.

References

1. Baert, S.A., Viergever, M.A., Niessen, W.J.: Guide-wire tracking during endovascular interventions. *IEEE Trans. Med. Imaging* **22**(8), 965–972 (2003)
2. Cheng, E., Pang, Y., Zhu, Y., Yu, J., Ling, H.: Curvilinear structure tracking by low rank tensor approximation with model propagation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3057–3064 (2014)
3. Cheng, J.Z., Chen, C.M., Cole, E.B., Pisano, E.D., Shen, D.: Automated delineation of calcified vessels in mammography by tracking with uncertainty and graphical linking techniques. *IEEE Trans. Med. Imaging* **31**(11), 2143–2155 (2012)
4. De Lathauwer, L., De Moor, B., Vandewalle, J.: On the best rank-1 and rank- (r_1, r_2, \dots, r_n) approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **21**(4), 1324–1342 (2000)
5. Frank, A.: On Kuhn’s Hungarian method — a tribute from Hungary. *Nav. Res. Logistics* **52**(1), 2–5 (2005)
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
7. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
8. Palti-Wasserman, D., Brukstein, A.M., Beyar, R.P.: Identifying and tracking a guide wire in the coronary arteries during angioplasty from X-ray images. *IEEE Trans. Biomed. Eng.* **44**(2), 152–164 (1997)
9. Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013, Part II. LNCS*, vol. 8150, pp. 246–253. Springer, Heidelberg (2013)
10. Roth, H.R., Wang, Y., Yao, J., Lu, L., Burns, J.E., Summers, R.M.: Deep convolutional networks for automated detection of posterior-element fractures on spine CT. In: *SPIE Medical Imaging*, p. 97850 (2016)
11. Shi, X., Ling, H., Xing, J., Hu, W.: Multi-target tracking by rank-1 tensor approximation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2387–2394 (2013)
12. Vedaldi, A., Lenc, K.: MatConvNet – convolutional neural networks for MATLAB. In: *Proceedings of the ACM International Conference on Multimedia* (2015)
13. Wang, P., Chen, T., Zhu, Y., Zhang, W., Zhou, S.K., Comaniciu, D.: Robust guidewire tracking in fluoroscopy. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 691–698 (2009)
14. Zhu, Y., Tsin, Y., Sundar, H., Sauer, F.: Image-based respiratory motion compensation for fluoroscopic coronary roadmapping. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part III. LNCS*, vol. 6363, pp. 287–294. Springer, Heidelberg (2010)