# LEFT ENDOCARDIUM TRACKING VIA COLLABORATIVE TRACKERS AND SHAPE PRIOR

*Yan Zhou*[⋆] *Shaoting Zhang*[†] *Nikolaos Tsekos*[▽] *Ioannis Pavlidis*[▽] *Dimitris Metaxas*[†]

[⋆] Elekta Inc., Maryland Heights, MO, USA
[†] CBIM, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
[▽] Computer Science Department, University of Houston, Houston, TX, USA

## ABSTRACT

Real-time MRI emerges as a feasible modality for guiding interventions and surgeries on the beating heart. One challenge is the path planning task for intra-cardiac procedures, which requires tracking the inner contour of the left endocardium on the beating heart. We propose a bottom-up approach to solve this tracking problem. On the bottom level, a coarse contour is generated based on a collaborative network of trackers at every time step. On the top level, PCA based shape prior is used to refine the coarse contours, and eventually obtain a smoothed contour of the left endocardium over time. The approach is validated on 14 MRI sequences of a beating heart, with comparisons to the state-of-the-art approaches. Experimental results show that our approach outperforms other approaches, providing a robust, accurate and fast solution to track the left endocardium of the beating heart.

***Index Terms***— Collaborative tracking, shape prior

## 1. INTRODUCTION

Extracting the boundary of a beating heart from cardiac MRI sequences plays an important role in cardiac disease diagnosis and treatment. MRI-guided robotic intervention is potentially important in cardiac procedures such as aortic valve repair [3]. The major difficulty is the path planning of the robotic needle [5], where it requires an accurate contour of the left endocardium on a real-time MRI sequence. The desired algorithm should be automatic, robust, accurate and fast.

Therefore, we consider the problem of tracking the inner boundary of the left endocardium from cardiac MR images. If the motion of the region of interest is constrained (e.g. rigid or approximately rigid), the contour motion can be efficiently represented by a small number of parameters, e.g. the affine group. But this is not the case for a beating heart, where the heart undergoes non-rigid deformations through a heartbeat. The contour deformation then forms an infinite dimensional space. Direct application of motion tracking algorithms (e.g., Kalman filter, particle filter) for large dimensional problems is impractical, due to the reduction in effective particle size as dimension increases. An alternative way is to model the boundary shape to constrain the freedom of how the contour deforms.

Two major types of shape models are statistical shape models and deformable models. Statistical shape models are learned a priori from examples to capture variations in the shape and appearance of the target [1]. These models efficiently constrain the deformation to a specific manner thus better in image interpretation. However, it usually only takes into account local shape or appearance information, thus less robust confronting large appearance and motion changes of the beating heart. Deformable models are curves or surfaces that deform under the influence of internal smoothness and external image forces to delineate the object boundary [2, 4, 10]. They are able to extract smoothed curves but the computational cost can be expensive for temporal data. Some recent work effectively integrates both types [8] and achieve high contour segmentation accuracy, but those methods neglect the temporal connections of the image sequence, and require an additional learning algorithm to generate an initialized contour at each time step.

We propose a bottom-up approach to solve the contour tracking problem. On the bottom level, a collaborative trackers network provides a deformed mesh to generate an initialized contour at each time step [9] (Fig.1). Each tracker in the network is a particle filter tracker featuring the appearance of the tracking region as the template. The coordination of the trackers in the network is modelled by a Bayesian network, which models spatio-temporal dependency relationships of multiple trackers. On the top level, a trained Active Shape Model (ASM)[1] serves as shape prior to refine the initialized contour (Fig.1), and iteratively obtains a smoothed contour. Our approach has the following advantages: 1) The tracking performance of the collaborative trackers network not only depends on the local appearance of individual trackers, but also depends on the performance of its neighbouring trackers, namely its context. Once the ASM shape prior is applied on the top level, the approach integrates shape, appearance and context information in a unified framework, providing a robust solution to the contour tracking problem. 2) The net-

This material is based upon work when Yan Zhou was a PhD student in Computational Physiology Lab, University of Houston.

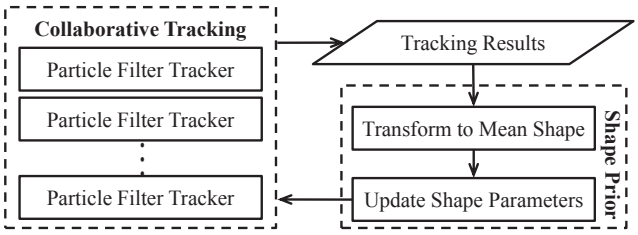



**Fig. 1**. Algorithmic framework, which includes the tracking network and shape prior modules.

work of trackers featuring particle filter and Bayesian network mechanisms automatically enforces temporal consistency between subsequent images. It automatically generates an initialized contour input for the ASM shape refinement at each time step. 3) The approach is computationally light weighted. It takes 0.17 second on average to process one image on a 2.4GHz desktop.

## 2. METHODOLOGY

The contour of the left endocardium at time $t$ is represented by a set of points $C_t = \{(x_1^t, y_1^t), ..., (x_N^t, y_N^t)\}$ on the 2D image plane. The goal of the approach is to accurately estimate the contour $\bar{C}_t$ at each time step. For this purpose, we first use a collaborative tracking network to track the endocardium region (Fig. 1 left). Then an ASM shape prior model is applied to refine the tracking results (Fig. 1 right). The two parts work iteratively to achieve accurate contour tracking performance.

### 2.1. Collaborative Tracking

Given an initial contour region $C_0$, the collaborative tracking approach is used to compute a coarse contour $E_t$, as the base for the shape refinement algorithm. The initial contour $C_0$ is manually annotated on the first frame. There are some collaborative tracking approaches in the literature [6][9], we adopted Zhou et al.'s tracking method [9].

**Tracking Network Computation:** The collaborative trackers network is composed of multiple particle filter trackers $\Gamma_t = \{T_i^t, ..., T_m^t\}$ (Fig. 2 (a)). For each individual tracker $T_i^t$, we measure its tracking performance by computing its surviving probability. The adjacent trackers $\{T_j^t, ..., T_m^t\}$ provide evidence to compute its surviving probability via a Bayesian network:

$$p(W_i^t, \hat{\Theta}_i^t | G_i^{t-1}, Z_i^t) \propto p(G_i^{t-1}) \prod_k N(\hat{\theta}_i^t | \hat{\theta}_k^t, \sigma^2) \prod_k p(\hat{\theta}_k^t | z_k^t) \quad (1)$$

$\hat{\Theta}_i^t = \{\hat{\theta}_i^t, \hat{\theta}_j^t, ..., \hat{\theta}_m^t\}$ and $Z_i^t = \{z_i^t, z_j^t, ..., z_m^t\}$ are the estimated states and observations of $T_i^t$ and its adjacent trackers. $W_i^t$ represents the event that tracker $T_i^t$ survives. $G_i^{t-1}$ is the Bayesian network at time $t-1$, whose probability $p(G_i^{t-1})$ is known at time $t$. $N(\hat{\theta}_i^t | \hat{\theta}_k^t, \sigma^2)$ is the probability density of $\hat{\theta}_i^t$ on the Normal distribution centered at $\hat{\theta}_k^t$ with variance $\sigma^2$. Those terms enforce both spatial and temporal consistency of the trackers network, ensuring robust performance. Detailed explanations can be found in [9].

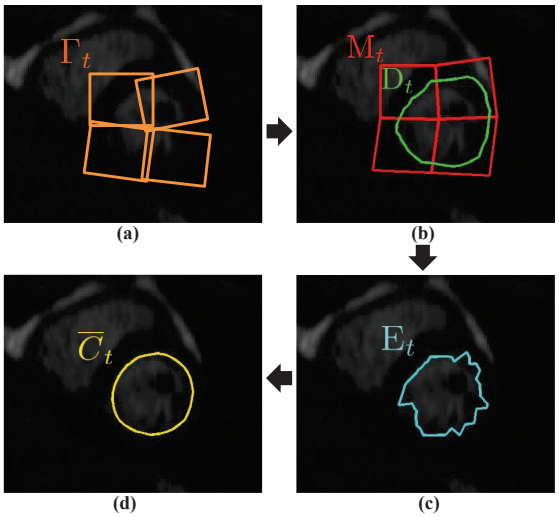



**Fig. 2**. (a) Trackers' network that composes of 4 particle filter trackers $\Gamma_t$; (b) Deformation mesh $M_t$ from the trackers' network (in red), and the deformed contour $D_t$ (in green); (c) Computed edge $E_t$ along the normal direction of the deformed contour; (d) Extracted shape $\overline{C}_t$ after ASM shape refinement.

**Deformation Mesh and Interpolated Contour ($C_0$ to $D_t$):** The deformation mesh (Fig. 2 (b) in red) $M_t$ is composed of a set of nodes $M_t = (c_1^t, c_2^t, ..., c_n^t)$, which are distributed over the targeted heart region. In the initialization step, the nodes coincide with the defining points of the individual trackers. Subsequently,each node's location is decided by the linked trackers as the following:

$$c_i^t = \frac{1}{\sum_{k=1}^{L_i^t} \beta_k^t} \sum_{k=1}^{L_i^t} s_k^t \beta_k^t, \quad (2)$$

where $\{s_k^t\}_{k=1}^{L_i^t}$ indicates all the corner coordinates of the linked trackers at a certain location and $\beta_k^t$ is the impact factor of each linked tracker; we use the surviving probability computed in Equation (1) as the impact factor here. $L_i^t$ is the total number of linked trackers in locale $i$ at time $t$. Given the contour of the previous time step $D_{t-1}$, we are able to linearly interpolate a deformed contour $D_t$ according to the coordinates of the deformed mesh (Fig. 2 (b) in green).

**Edge Detection ($D_t$ to $E_t$):** We use an edge detection method to align $D_t$ to the edge of the tracking region $E_t$. Specifically, $E_t$ is computed by finding the largest first derivative along the normal direction of contour $D_t$ (Fig. 2 (d)). However, the extracted edges always fall into local minima. Therefore it requires a shape refinement procedure to smooth the edges and constrain the deformation.

### 2.2. Refinement using Shape Prior ($E_t$ to $\bar{C}_t$)

The shape refinement procedure is similar to the shape updating step in ASM. Firstly, all the training contour shapes are

aligned using similarity transformation. Then their statistical pattern is captured by principal component analysis (PCA). Given the edge output from the trackers network, it is first transformed to the mean shape, and then mapped onto the PCA space to update the pose and shape parameters. This step enforces shape constraints to prevent over and under-segmentation. To sum, we perform the following steps to extract smoothed contours $\bar{C}_t$ for an MRI sequence:

1. Train the shape prior model. Use PCA to capture statistics (mean and variance) of these contour shapes.
2. Initialize a collaborative tracking network, which consists of multiple simple particle filter trackers $T_t$ and an initialized contour $C_0$.
3. Compute the deformed mesh $M_t$ of the trackers network and interpolated contour $D_t$.
4. Generate an intermediate edge result $E_t$ at each time step.
5. Refine the intermediate result to obtain $\bar{C}_t$ by shape prior constraint.
6. Repeat steps 3-5 until convergence.

## 3. EXPERIMENTS

The performance of the contour tracking approach is critical to MRI-guided robotic interventions. At each time step, the accurate contour of the left endocardium serves as the "boundary" to limit the freedom of the robotic needle so that it will not accidentally touch the cardiac wall. The approach should also be fast enough to minimize the computational cost during a real-time heart surgery. Therefore, accuracy and speed are the most important evaluation factors in our experiments.

We performed experiments on 14 MRI sequences with 12-750 slices (frames) for each sequence and 1728 slices (frames) in total. Data was acquired from 7 patients and by two types of MRI acquisition protocols: (a) CINE MRI sequences where patients hold the breath during MRI scanning; This type of MRI sequences contains relatively few slices (12-25 frames) since the patient is not able to hold the breath for long time. (b) Real-time MRI sequences where patients have natural breathing during MRI scanning. This type of MRI sequences contains more slices (50-750 frames) and the cardiac motion pattern is more complicated because of the additional respiratory motion. From all the images, we randomly selected 150 images for training, and all the images for testing. The proposed method mainly includes two parts: collaborative tracking network and the shape prior. The performance of the tracking network was extensively evaluated in previous work [9]. Thus we mainly evaluate the performance of different shape refinement methods here. We compare the proposed approach with the following approaches: 1) Deformation mesh approach (DM): The contour is computed by interpolating the initialized contour according to the current state of the deformation mesh. 2) Edge detection approach (ED): This approach simply computes the edge along the normal direction of the interpolated contour. All the shape refinement approaches are based on the same intermediate tracking results provided by the collaborative tracking network.

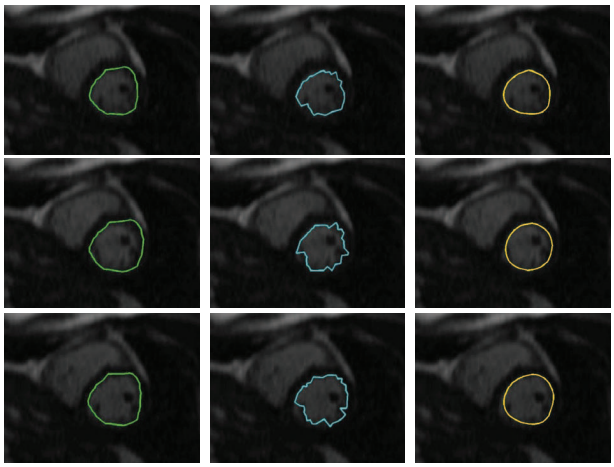



**Fig. 3**. Comparative results on a CINE testing sequence with 15 frames. (a) Tracking results by the deformation mesh approach. (b) Tracking by edge detection. (c) Tracking results using the proposed method. The rows corresponds to frame index 3, 9 and 15.

Some challenging cases are shown in Fig. 3 and Fig. 4. Fig. 3 shows the results from a short CINE sequence. The DM approach performs well visually because the sequence includes just 15 frames with little shape variation over time. The only problem is that the shape of the contour is static no matter how the left endocardium deforms due to the linear interpolation strategy. The ED approach performs worse visually because the edge detection method is easy to drop into local minima. Our proposed approach provides a smoothed shape that matches the deformed left endocardium. Fig. 4 shows the results from a long real-time sequence. The DM approach performs worse than in the CINE sequence because of the large cardiac motion. The deformation mesh is only able to compensate the global motion of each tracking region, but not the accurate local motion of the left endocardium. Therefore, the interpolated shape is not accurate. The ED approach still produces noisy results visually. Our proposed approach not only captures the motion variations over time, but also provide a smoothed accurate contour.

To quantitatively compare different methods, we report the boxplots of sensitivity (P), specificity (Q) and Dice Similarity Coefficient (DSC) in Fig. 5. The computation was performed on all the testing data, and the ground-truth contours were annotated manually by an expert. We prefer the approach with large mean value and small variance since it shows accuracy and stableness. In terms of sensitivity, our method is much better than the ED method and slightly better than the DM method. For the specificity, the result from ED is also good. The reason is that this method always under-

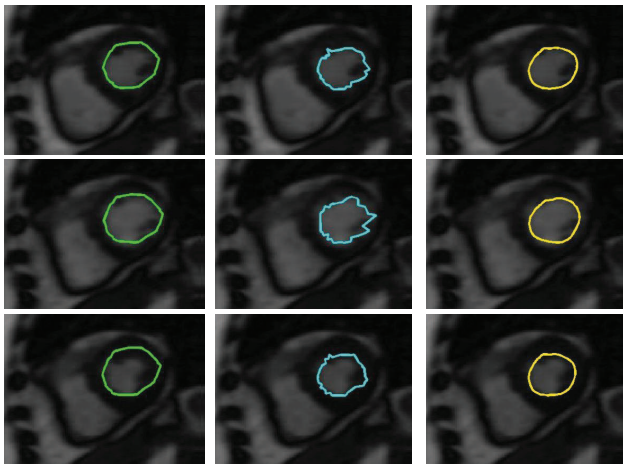



**Fig. 4**. Comparative results on a real-time testing sequence with 750 frames. (a) Tracking results by the deformation mesh approach. (b) Tracking by edge detection. (c) Tracking results using the proposed method. The rows corresponds to frame index 239, 251, 409.

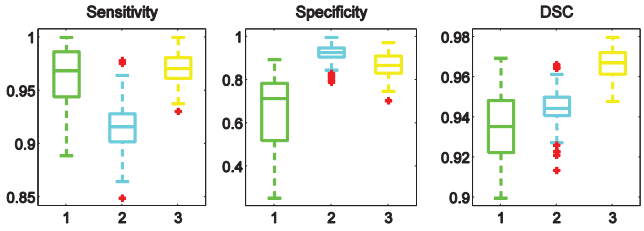



**Fig. 5**. Boxplots of sensitivity, specificity and DSC of all testing data. In each subplot, y-axis is the performance of sensitivity, specificity or DSC. x-axis aligns 3 methods from left to right in the same order as in Fig. 3 and Fig. 4.

segment the ROI. Thus most of the background are correctly predicted. However, as a trade-off, its sensitivity is not as good as the others. Considering both P and Q, our method outperforms DM method and ED method. DSC combines both sensitivity and specificity information. Thus it is a reasonable measurement to compare accuracy. The results of DSC further validate our proposed method.

We evaluate the computing time of the proposed approach by averaging the processing time of all testing cases. It takes 0.17 second to process one image on a 2.4GHz desktop. The reason is that particle filter trackers in the tracking network are computationally light weighted. And the ASM shape refinement process converges quickly with the good quality input shapes from the tracking network. The high speed of the approach makes it promising for real-time applications.

## 4. DISCUSSIONS

Integrating the tracking network and the shape prior is more robust to noise and can handle complex textures than purely using the tracking algorithm. Our model converges quickly and robustly towards the boundary because of the good-quality tracking results and the constraint of the shape prior. Another benefit of our model is that the size of the training data can be small (150 training images in our case), because the tracking algorithm already enforces both spatial and temporal consistency, thus robust to image noise. In the future, we will try to use more sophisticated shape prior methods such as sparse shape composition[7] to refine shape.

## 5. ACKNOWLEDGEMENTS

This material is supported by the National Science Foundation under grants No. #IIS-0812526 and #CNS-0932272.

## 6. REFERENCES

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: Their training and application. *CVIU*, 61:38–59, 1995.

[2] T. Kohlberger, D. Cremers, M. Rousson, R. Ramaraj, and G. Funka-Lea. 4D shape priors for a level set segmentation of the left myocardium in SPECT sequences. *MICCAI*, pages 92–100, 2006.

[3] M. Li, D. Mazilu, and K. Horvath. Robotic system for transapical aortic valve replacement with MRI guidance. *MICCAI*, pages 476–484, 2008.

[4] J. I. McEachen and J. Duncan. Shape-based tracking of left ventricular wall motion. *TMI*, 16(3):270 –283, 1997.

[5] E. Yeniaras, N. V. Navkar, A. E. Sonmez, D. J. Shah, Z. Deng, and N. V. Tsekos. MR-based real time path planning for cardiac operations with transapical access. *MICCAI*, pages 25–32, 2011.

[6] T. Yu and Y. Wu. Decentralized multiple target tracking using netted collaborative autonomous trackers. In *CVPR*, pages 939–946, 2005.

[7] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical Image Analysis*, 16(1):265–277, 2012.

[8] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *TMI*, 27(11):1668 –1681, 2008.

[9] Y. Zhou, E. Yeniaras, P. Tsiamyrtzis, N. Tsekos, and I. Pavlidis. Collaborative tracking for MRI-guided robotic intervention on the beating heart. *MICCAI*, pages 351–358, 2010.

[10] Y. Zhu, X. Papademetris, A. Sinusas, and J. Duncan. Segmentation of the left ventricle from cardiac MR images using a subject-specific dynamical model. *TMI*, 29(3):669 –687, 2010.