

Prediction Based Collaborative Trackers (PCT): A Robust and Accurate Approach Toward 3D Medical Object Tracking

Lin Yang, *Member, IEEE*, Bogdan Georgescu, *Member, IEEE*, Yefeng Zheng, *Senior Member, IEEE*, Yang Wang, *Member, IEEE*, Peter Meer, *Senior Member, IEEE*, and Dorin Comaniciu, *Senior Member, IEEE*

Abstract—Robust and fast 3D tracking of deformable objects, such as heart, is a challenging task because of the relatively low image contrast and speed requirement. Many existing 2D algorithms might not be directly applied on the 3D tracking problem. The 3D tracking performance is limited due to dramatically increased data size, landmarks ambiguity, signal drop-out or complex non-rigid deformation. In this paper we present a robust, fast and accurate 3D tracking algorithm: Prediction Based Collaborative Trackers (PCT). A novel one-step forward prediction is introduced to generate the motion prior using motion manifold learning. Collaborative trackers are introduced to achieve both temporal consistency and failure recovery. Compared with tracking by detection and 3D optical flow, PCT provides the best results. The new tracking algorithm is completely automatic and computationally efficient. It requires less than 1.5 seconds to process a 3D volume which contains millions of voxels. In order to demonstrate the generality of PCT, the tracker is fully tested on three large clinical datasets for three 3D heart tracking problems with two different imaging modalities: endocardium tracking of the left ventricle (67 sequences, 1134 3D volumetric echocardiography data), dense tracking in the myocardial regions between the epicardium and endocardium of the left ventricle (503 sequences, roughly 9000 3D volumetric echocardiography data), and whole heart four chambers tracking (20 sequences, 200 cardiac 3D volumetric CT data). Our datasets are much larger than most studies reported in the literature and we achieve very accurate tracking results compared with human experts' annotations and recent literature.

Index Terms—Tracking, Ultrasound, CT, Heart, Left Ventricle, Motion Learning.

I. INTRODUCTION

The 3D echocardiography and cardiac computed tomography (CT) are emerging diagnostic tools among modern imaging modalities for visualizing cardiac structure and diagnosing cardiovascular diseases. The echocardiography [1] is real-time, noninvasive imaging modality which is less expensive than CT and magnetic resonance imaging (MR). Ultrasound normally produces noisy images with poor object boundaries. Compared with other imaging modalities (such as ultrasound and MR), cardiac CT can provide detailed anatomic information about the heart chambers, large vessels, and coronary arteries [2].

L. Yang is with the Department of Radiology in the University of Medicine and Dentistry of New Jersey, B. Georgescu, Y. Zheng, Y. Wang and D. Comaniciu are with Siemens Corporate Research, P. Meer is with the Department of Electrical and Computer Engineering in Rutgers University.

This research was completed when the first author was in the Integrated Data Systems Department of Siemens Corporate Research.

Further author and paper information: please send correspondence to Lin Yang at yangli@umdnj.edu or linyang711@gmail.com.

Recently, the problem of automatic detection, segmentation and tracking of heart chambers in 3D radiological imaging, such as ultrasound and CT, have received considerable attentions. A multiple-template based tracking approach is proposed in [3] for tracking cardiac structure in MRI. A new deformable model technique based on a snake-like approach and Fourier shape descriptors parameterization is used in [4] for MRI, and a statistical point distribution model based deformable model is proposed in [5] for both MRI and CT. A 3D slice cut algorithm is proposed in [6] which transfers the 3D segmentation problem into several 2D segmentation problems in orthogonal planes. Marginal space learning [7] and collaborative trackers [8] are proposed for fast and robust segmentation and tracking of heart chambers, in 3D CT and ultrasound, respectively. The accurate tracking of the human's heart motion has clear clinical significance for radiologists to evaluate dangerous cardiac disease, such as acute myocardial infarction. Compared with traditional 2D tracking applications, one of the major considerations for developing a practically useful 3D tracking algorithm is speed, as the computational demand is much higher for 3D volumetric data compared with 2D.

The widely used 2D tracking algorithms [9] won't provide good results if directly applied on 3D tracking applications. In order to achieve robust tracking in 3D radiology images, learning based detector and boundary classifiers are proposed to robustly track the boundary of the heart in each frame. This tracking by detection strategy can avoid accumulating errors and is proven to be quite effective in recent literature [10], [11], [12]. However, it still has several problems:

- The boundary classifiers are sensitive to initial positions [13] and good initializations have to be provided because we can not exhaustively search all the possible configurations in the whole 3D volume due to speed consideration.
- Tracking by detection applies a universal description of the objects without considering any temporal relationship, which leads to temporal inconsistency between adjacent frames.

In this paper, we propose a fast and novel automatic 3D tracking algorithm called Prediction Based Collaborative Trackers (PCT), which addresses these difficulties. PCT applies a motion prior learned on a low dimensional manifold to constrain the collaborative trackers including a detection tracker and a template tracker. Robust information fusion is

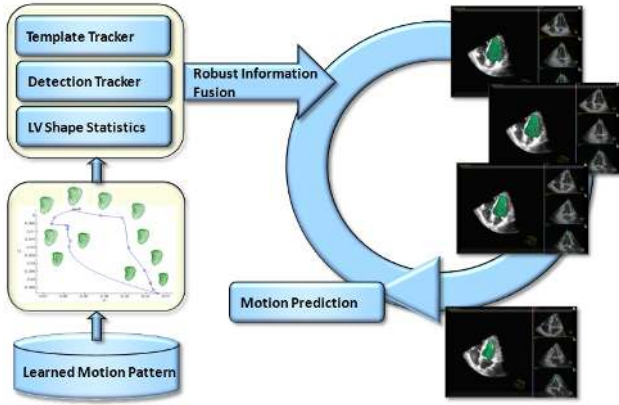


Fig. 1. The flow chart of the predication based collaborative trackers (PCT) using one step forward prediction, marginal space learning, motion manifold learning and detection/template trackers.

applied to generate the joint posterior probability. Motion prediction is performed using registration guided one step forward prediction. The whole procedure is performed periodically. A flowchart of the PCT is shown in Figure 1.

The contributions of this paper are:

- A novel one-step forward prediction using motion manifold learning, which respects the geodesic distances of both the shape and motion of the heart. The motion priors can provide good initial positions for the boundary classifiers.
- A collaborative 3D template tracker is applied to erase the temporal inconsistency introduced by 3D detection tracker.
- A rectangle filter is used to reject outliers and achieve robust data fusion. Smooth boundary tracking is obtained by projecting the tracking points in each frame to the constrained PCA shape boundary.
- The algorithm can process a 3D volume, e.g., $160 \times 144 \times 208$ voxels, in less than 1.5 seconds and we obtain subvoxel tracking accuracy.
- The algorithm is a general framework which can be applied to diverse tracking applications in different imaging modalities, as we shown in the experimental part.

To our knowledge, this is the first report to illustrate the usage of motion manifold learning to provide the one step forward motion prior for a set of collaborative trackers in heart chamber's tracking. This is also a novel and critical step which increases not only the accuracy but also the tracking speed. The paper is organized as follows. Section 2 introduces the Bayesian tracking framework. The learning methods and tracking algorithm are described in Section 3 and Section 4, respectively. Section 5 provides the experimental results for endocardium tracking, dense myocardial region tracking between endocardium and epicardium, and the whole four chambers tracking. Section 6 concludes the paper.

II. BAYESIAN TRACKING FRAMEWORK

Define \mathbf{x}_t as the true position of each 3D boundary point of the heart at time t and let \mathbf{z}_t represent the measurement. The tracking problem can be formulated as an estimation of a

posteriori probability $p(\mathbf{x}_t|\mathbf{z}_{1:t})$, where $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \dots, \mathbf{z}_t\}$ represents the past t measurements. Sequential Bayesian tracking based on Markovian assumption is performed recursively in a *prediction*

$$p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})d\mathbf{x}_{t-1} \quad (1)$$

and *updating* step

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) \propto p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1}). \quad (2)$$

Bayesian tracking assumes that the following densities are known. The $p(\mathbf{x}_0)$ denotes the distribution of the 3D LV surface points in the first frame. In our algorithm $p(\mathbf{x}_0)$ is automatically calculated using the trained detector and boundary classifiers. The $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ represents the motion prior (state model) and is predicted for the next frame. The $p(\mathbf{z}_t|\mathbf{x}_t)$ represents the measurement density.

Both the Kalman filter and particle filter [14], [15] assume a Markov model, which only considers the previous state \mathbf{x}_{t-1} to estimate the density of current state \mathbf{x}_t . In real tracking problems, the motion prior (state model) $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1})$ may not follow a Markovian assumption and it could be of any form. In our algorithm, we model the $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1})$ to be dependent on both \mathbf{x}_{t-1} and $\mathbf{z}_{1:t-1}$. One-step forward prediction using motion manifold learning is applied to estimate this critical motion prior.

For the measurement densities $p(\mathbf{z}_t|\mathbf{x}_t)$, we select two collaborative trackers: the detection tracker and the template tracker, which can mutually benefit each other. The detection tracker can discriminate the 3D target from background in low image quality and noisy environment. The template tracker respects the local image information and preserves the temporal consistence between adjacent frames. The trackers are modeled as r_k , $k = 1$ for the detection tracker and $k = 2$ for the template tracker, then

$$p(\mathbf{z}_t|\mathbf{x}_t) = p(\mathbf{z}_t|\mathbf{x}_t, r_1)p(r_1) + p(\mathbf{z}_t|\mathbf{x}_t, r_2)p(r_2). \quad (3)$$

Substituting (3) in (2) and replacing $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ with $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1})$ in (1), the final posterior probability $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ is obtained from the robust data fusion and the $\mathbf{x}_t = \arg \max_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{z}_{1:t})$.

III. MOTION LEARNING AND CLASSIFIERS TRAINING

The proposed prediction based collaborative trackers (PCT) is not restricted to any specific heart chamber. It can be used for the endocardium of the left ventricle, the myocardial region between the endocardium and epicardium, the right ventricle (RV), the left atrium (LA) and right atrium (RA), or the arbitrary combination of four heart chambers. In the experimental part, we will report three applications: tracking of the endocardium of the left ventricle, dense tracking of the myocardial region of the left ventricle, and the four chambers tracking in 3D cardiac CT, using the proposed collaborative tracking and motion learning algorithm.

For illustration purpose, tracking the endocardium of the left ventricle is utilized as an example to describe the proposed method. The tracking of the myocardial region and the other chambers follow the same training and testing procedures.

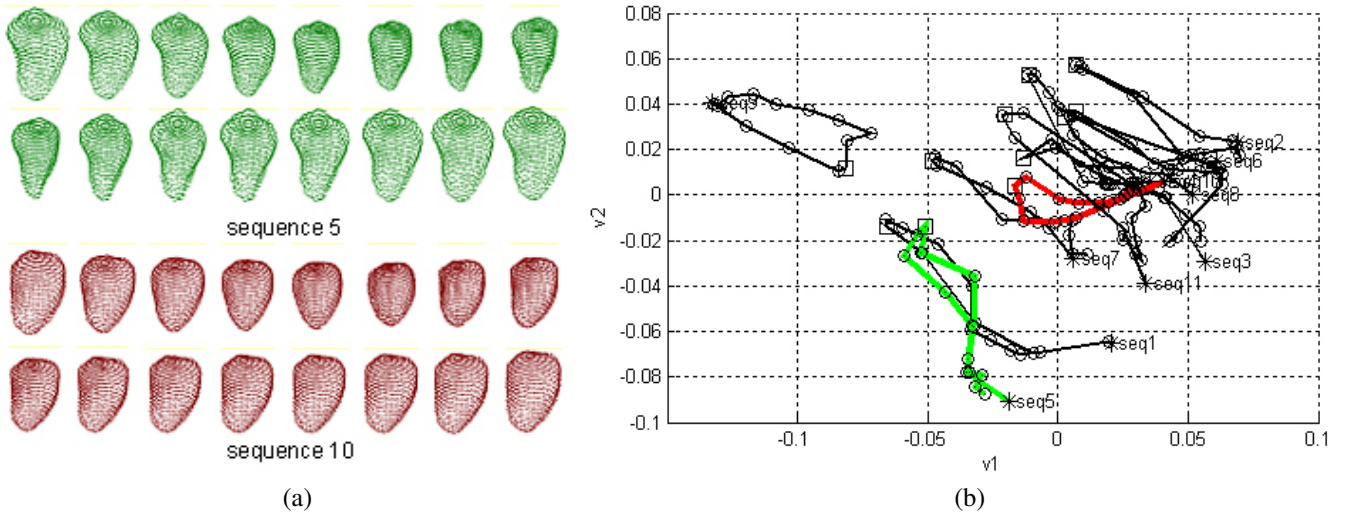


Fig. 2. Manifold embedding for heart motion patterns. (a) Two left ventricle surface mesh sequences. (b) 11 sequences embedded in a 2D subspace. Note: The end diastolic (ED) phase has larger volumes and represented as stars in (b), while the end systolic (ES) phase has smaller volumes and represented as squares in (b).

Furthermore, in the algorithm description part, although we just use the endocardium of the left ventricle as an example, we refer to left ventricle tracking throughout the paper for simplicity.

Each left ventricle training sequence contains a heart motion cycle which starts from the end-diastolic (ED) phase, passes through the end-systolic (ES) phase and comes back to ED. The learning contains three steps. First the motion modes are learned using manifold learning and hierarchical K-means. Next, an ED detector is trained to locate the position of the object in the first frame. Finally, two boundary classifiers (one for ED and one for ES) are trained using the annotated sequences to delineate the boundary.

A. Learning the Motion Modes

In this section we explain the learning steps using the left ventricle (LV) as an example. After motion alignment, multiple motion modes are learned using motion manifold learning and hierarchical K-means.

1) *Motion Alignment Using 4D Generalized Procrustes Analysis:* The training sequences contain 11-25 time frames from one cardiac cycle, and each sequence is resampled to 16 frames using interpolation along the timeline. In this way we generate 4D motion vectors containing the same dimensionality d , where $d = N_f \times 3 \times 16$, $N_f = 289$ is the number of boundary points and three represents x , y and z dimensions.

Generalized Procrustes analysis (GPA) is used to align all resampled motion vectors to remove the translation, rotation and scaling [16, ch. 5]. Please be aware that the shape variations and motion patterns inside the motion vectors are still kept, GPA just removed the difference of the training motion vectors in the global coordinate systems. After the 4D GPA, these aligned motion vectors are decomposed into separated 3D shapes. All the following learning steps are performed on the aligned 3D shape vectors.

2) *Motion Manifold Learning:* Because the actual number of constraints that control the LV motion are much less than its original dimensionality, the aligned 3D shape vectors lie on a low-dimensional manifold, where geodesic distance has to be used to measure the similarities. Given the whole set of 3D training shape vectors, $M = \{\mathbf{m}_0, \dots, \mathbf{m}_i, \dots, \mathbf{m}_n\}$ where $\mathbf{m}_i \in R^d$, there exists a mapping F which represents \mathbf{m}_i in the low-dimensions as

$$\mathbf{m}_i = F(\mathbf{v}_i) + \mathbf{u}_i \quad i = 1, 2, \dots, n, \quad (4)$$

where $\mathbf{u}_i \in R^d$ is the sampling noise and $\mathbf{v}_i \in R^q$ denotes the original i -th shape \mathbf{m}_i in the low-dimensional manifold. The nonlinear mapping F is the transformation from the low-dimensional manifold to the original space.

Unsupervised manifold learning is capable of discovering the nonlinear degrees of freedom that underlie complex natural observations. We apply ISOMAP [17] to embed the nonlinear manifold into a low-dimensional subspace. We start by finding the seven closest neighbors (seven was found to be the best) of each point \mathbf{m}_i in the original space R^d and connect the neighbors to form a weighted graph G . The weights are calculated based on the Euclidean distance between each connected pair of vectors. We then calculate the shortest distance $d_G(i, j)$ between any pair of points \mathbf{m}_i and \mathbf{m}_j in the graph G . The final step is to apply the standard multiple dimensional scaling (MDS) to the matrix of graph distance $\{d_G(i, j)\}$. In this way, the ISOMAP applies a linear MDS on the local patch but preserves the geometric distance globally using the shortest path in the weighted graph G .

Figure 2a shows two annotated LV motion sequences. Figure 2b shows several LV motion representations in a low-dimensional manifold. An interesting but expected observation is illustrated in Figure 2b. The LV motion is almost periodic because one cycle of heart beat starts from ED and returns to ED. In total we applied manifold learning on 36 annotated LV motion sequences from the total 67 sequences. In order to make the figure readable, we only show 11 sequences in

Figure 2b.

3) *Hierarchical K-means Clustering*: Given all the motion cycles shown on the embedded subspace, a hierarchical K-means clustering method is proposed to learn the motion modes. First, we apply K-means on all the training ED shapes. After this step, we align all the trained motion sequences by moving their motion vectors to their own cluster center based on the ED clustering results. In this way we cancel the translation among training motions with similar ED shapes, and different cluster center represents the difference among the ED shapes of the left ventricle. In the second round of K-means, the learning procedure focuses on separating those left ventricles with similar ED shapes, but different motion patterns. Each aligned motion sequence is transformed to a $q \times 16$ dimensional vector, where q represents the reduced dimensionality and 16 represents the number of frames. K-means is applied again within each group, and the K-means cluster center in this round represents a learned motion mode. Each motion mode is a weighted sum of all sequences that are clustered into the same group. The weights are proportional to their Euclidean distance from the cluster center. The geodesic distance in the original space is modeled by Euclidean distance in this embedded low-dimensional manifold. In the left ventricle tracking, we tested different dimensions, such as two, three and four as the dimensionality in the low dimensional manifold, and no big difference is discovered regards to the performance. Therefore $q = 2$ is chosen as the reduced subspace dimensionality for the left ventricle tracking in 3D echocardiography. For the twist motion tracking in the myocardial region which contains large degree of freedoms, we performed the motion manifold learning on a manifold with higher dimension $q = 20$. The dimensionality of the embedded manifold depends on the application and can be determined heuristically. We want to clarify that low dimensional nonlinear motion manifold is just used in the training stage to cluster the motion patterns. After the clustering is done, we will record the clustering index of each motion, and the averaged motions with original dimensionality (e. g. $289 \times 3 \times 16$ for left ventricle) are used as the learned motion patterns. The one-step forward motion prior is generated by nonlinear registration between the learned motion patterns and the testing data (shown in details in Section IV-B).

B. Learning the ED Detector

In this step we train a 3D detector to locate the pose of the left ventricle (LV) in the motion sequence. We first calculate a mean shape by averaging all the left ventricles in the ED frames of the annotated training sequences. A principal component analysis (PCA) shape space is calculated for all the ED shapes at the same time and the first 100 components are kept. In order to automatically initialize the tracker, we need to find the similarity transformation from the mean shape to the LV in the ED frame for each sequence as the testing volume can have arbitrary pose. Discriminative learning based approaches have proven to be efficient and robust for 2D object detection [18]. The object is found by scanning the classifier over an exhaustive range of possible

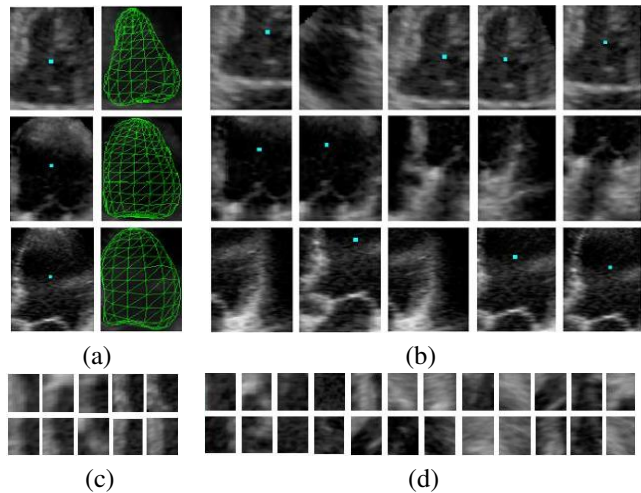


Fig. 3. The positive (a), (c) and negative samples (b), (d) used for training. (a) and (b) Training samples for the detector. (c) and (d) Training samples for the boundary classifiers.

locations, orientations and scales in an image. However, it is challenging to extend them to 3D objection detection problems since the number of hypotheses increases exponentially with respect to the dimensionality of the parameter space. As the posterior distribution is often clustered in a small region, it is not necessary to search the parameter space exhaustively.

Recently, we proposed marginal space learning (MSL) [19], [7] to efficiently detect a 3D anatomic structure in various medical imaging modalities. Here, we apply MSL to learn an ED detector to locate the LV in the first frame efficiently. The idea for MSL is to incrementally learn classifiers on projected marginal spaces. We split the estimation into position detection, position-orientation detection and full similarity transformation detection. The MSL reduces the number of testing hypotheses by six orders of magnitude in our applications, which makes the directly training of the detector on 3D volumetric data a feasible procedure.

C. Learning the Boundary Classifiers

After we obtain the pose of the LV in the ED frame, we need to segment its boundary to automatically start the trackers. The active shape model (ASM) [13] is used to deform an initial estimate of a nonrigid shape. The non-learning based boundary classifier in the original ASM does not work in our application due to the complex background and weak edges in 3D ultrasound. Learning based methods can exploit image evidences to achieve robust boundary detection. Boundaries with different orientation are usually detected on pre-rotated images [20]. Since 3D volume rotation is very time consuming, we use steerable features [7] for boundary detection and avoid rotating the 3D volume. For each boundary point (289 in total), we sample several points from the volume around it under a special pattern, which embed the orientation information into the distribution of the sampling points. A few local features at each sampling point, such as voxel intensity and gradients, etc., are calculated. The advantages of steerable features are that they combine the advantages of both local and global features.

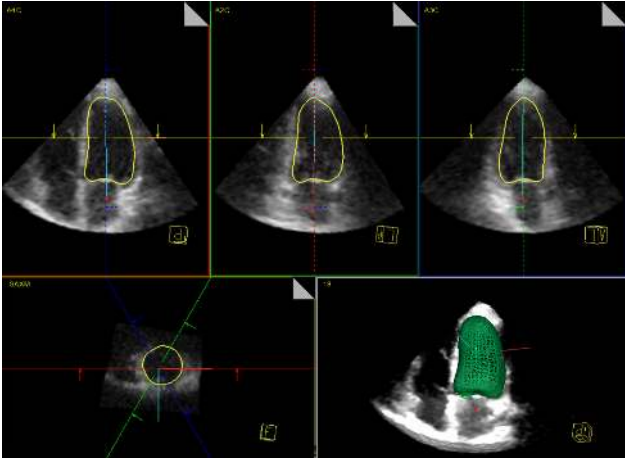


Fig. 4. The four canonical views and a 3D representation of the segmentation result (automatic tracking initialization) of LV.

Two boundary classifiers, one for LV motion close to the ED phase and the other for LV motion close to ES, are trained using the probabilistic boosting tree (PBT) [21]. The PBT ensembles many strong classifiers into a tree structure. The widely used cascade boosting [18] can be treated as a special case in PBT. The learned boundary classifiers are used to automatically segment the boundary of the LV, to initialize the trackers, and also used as the detection tracker for the following frame. In Figure 3 we show some positive and negative training samples used for training both the detector and the boundary classifiers.

IV. TRACKING PROCEDURE

The tracking procedure on a testing sequence contains four steps. The $p(\mathbf{x}_0)$ is initialized using the learned ED detector and the ED boundary classifier. At time $t-1$, registration based reverse mapping and one-step forward prediction are used to estimate the next state $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1})$. We then apply two collaborative trackers and robust data fusion to estimate the measurement density $p(\mathbf{z}_t|\mathbf{x}_t)$. In order to obtain smooth tracking of the LV, each boundary point is mapped to a shape constrained 3D boundary. The final results are obtained by maximizing the posterior probability in Equation (2). These *prediction* (1) and *updating* (2) steps are performed recursively for each frame in a sequence.

A. Initialization of Tracking

In order to initialize the boundary tracking of the LV in an automatic manner, we need to automatically detect and segment the LV in the ED frame of a testing sequence. Given the ED frame, all positions are scanned by the trained position detector. The top 100 candidates (x_i, y_i, z_i) , $i = 1, 2, \dots, 100$ are kept. Each candidate is then expanded using 1000 hypothesis on orientations $(\psi_{ij}, \phi_{ij}, \theta_{ij})$, $j = 1, 2, \dots, 1000$. The trained position-orientation detector is applied for each candidate and the best 50 are kept. These 50 candidates are expanded using 1000 scale hypothesis $(sx_{kl}, sy_{kl}, sz_{kl})$, $k = 1, 2, \dots, 50$, $l = 1, 2, \dots, 1000$. Evaluated by the position-orientation-scale detector, the best 100 candidates are averaged to produce the final similarity transformation estimation.

After the similarity transformation is calculated, the LV mean shape is registered and superimposed on the ED frame of the testing sequence as the initial position. For each boundary point we search ± 12 mm range on the normal directions of the boundary. (The value ± 12 mm was set considering both speed and accuracy.) The learned boundary classifier is used to move each boundary point to its optimal position where the estimated boundary probability is maximized. Figure 4 shows the tracking initialization result.

B. One Step Forward Prediction

In this step we calculate the motion prior (state model) $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1})$ using the learned motion modes. At time $t-1$, we first transform the current 3D shape $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ to the corresponding frame of *each* motion mode in the 4D GPA coordinate system. Thin plate spline (TPS) transformation [22] is applied to perform this mapping. The TPS is a nonrigid transformation between two 3D point sets. The transformation T contains an affine mapping plus a warping coefficient matrix. We used 72 uniformly sampled points from 289 boundary points to estimate the TPS transformation T by minimizing

$$E_{TPS}(T) = \sum_{i=1}^{72} \|\mathbf{w}_i - T(\mathbf{b}_i)\|^2 + \lambda f(T), \quad (5)$$

where \mathbf{w}_i denotes the 3D mesh point on the learned motion modes and \mathbf{b}_i denotes the point on the testing object's boundary. The $f(T)$ is a function containing a kernel which represents the internal structure relationship of the point set. The regularization parameter λ is chosen as 1.5. We refer the readers to [22] for more details.

The prediction is applied on the motion mode which minimizes the previous 1 to $t-1$ *accumulated* TPS registration errors

$$I = \arg \min_i \sum_{j=0}^{t-1} \min_T E_{TPS}(\mathbf{x}_j, \mathbf{q}_{i,j}, T), \quad (6)$$

where $i = 1, \dots, Q$ represents the number of motion modes. The $\mathbf{q}_{i,j}$ is the i -th motion mode in the j -th frame and the $E_{TPS}(\mathbf{x}_j, \mathbf{q}_{i,j}, T)$ is the registration error.

After we found the correct motion mode $\mathbf{q}_{I,t}$ using (6), the final prediction result in the real world coordinate system of the $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_{1:t-1})$ is obtained using the reverse mapping T^{-1} . A motion mode generated by reverse mapping using TPS is shown in Figure 5.

There could be motion mode changes during the prediction when the prediction starts from one motion mode and jumps to another mode during tracking. This corresponds to the LV motion which starts from an ED shape in one learned motion mode, but has a motion trajectory close to another mode. This is the reason we calculated the *accumulated* TPS registration error, and just perform one-step forward prediction instead of calculating the whole cycle motion prior one time. The one-step forward prediction step (illustrated in Figure 6) provides accurate motion prior in PCT.

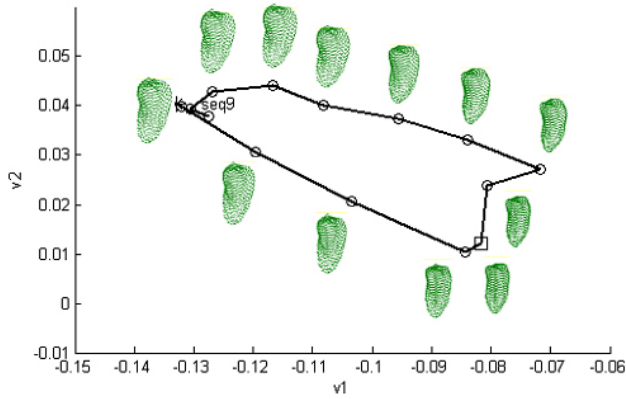


Fig. 5. The LV boundaries in 3D world coordinates of a motion mode. The results are calculated using TPS reverse mapping and superimposed in the two-dimensional reduced motion manifold.

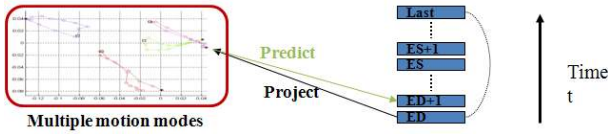


Fig. 6. The procedure for one step forward prediction in PCT.

C. Collaborative Trackers

Given the motion prior $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{z}_{1:t-1})$ learned using one-step forward prediction on the motion manifold, for each boundary point the learned boundary classifiers are used to search in its ± 12 mm range on the normal direction. The optimal position is found by maximizing the boundary probability. The ED boundary classifier is used when the frame index is close to ED and the ES boundary classifier is used when it is close to ES. The final position using *detection* tracker is obtained by maximizing $p(\mathbf{z}_t | \mathbf{x}_t, r_1)$ in (3).

In order to compensate the disadvantages of detection tracking mentioned in the introduction, a 3D *template* tracker is also applied. Given \mathbf{x}_t a 3D boundary point and its neighborhood $N(\mathbf{x}_t)$, let $G(N(\mathbf{x}_t), \mu)$ denotes the transformation of the template. (The neighborhood was chosen to be a cube of $13 \times 13 \times 13$ voxels based on experiments.) The goal is to search the best transformation parameters μ which minimize the error between $N(\mathbf{x}_t)$ and $G(N(\mathbf{x}_t), \mu)$.

$$\mu = \arg \min_{\mu} \sum_{\mathbf{x}_t \in N(\mathbf{x}_t)} [G(N(\mathbf{x}_t), \mu) - N(\mathbf{x}_t)]^2. \quad (7)$$

Because there is only a small change of parameter μ between adjacent frame, the minimization of (7) can be solved by linearizing the expression

$$G(N(\mathbf{x}_t), \mu) = G(N(\mathbf{x}_t)) + \frac{\partial G(N(\mathbf{x}_t), \mu)}{\partial \mu} d\mu. \quad (8)$$

At the end the result $p(\mathbf{z}_t | \mathbf{x}_t, r_2)$ is obtained.

Although the template matching algorithm is not robust and only works under the assumption of small inter-frame motions, it preserves temporal consistence and its disadvantages can be compensated by the one-step forward prediction and the detection tracker. Template updating is a key issue

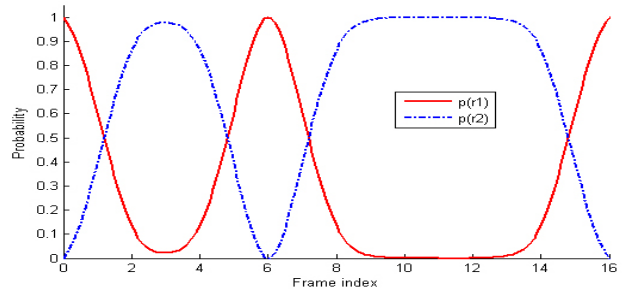


Fig. 7. The prior $p(r_1)$ for detection tracker (red solid line) and $p(r_2)$ for template tracker (blue dotted line). The ED phase has frame index zero and ES phase is around frame six.

in template tracking. If we update the template in each frame only based on the previous template tracking result, the error will be inevitably accumulated and finally results in the template drifting [23], [24]. Generally it is difficult for template tracking to recover from drifting without special processing. In our method we update the template using the previous collaborative tracking result. Because the learned motion prior is enforced and detection is used, this updating scheme can help the template tracker to recover from the template drifting. As shown in [25], [26], [27], learned motion prior is quite effective to help tracking to recover from failures. Collaborative kernel trackers are also successfully used in [28].

D. Data Fusion

Fusion of the collaborative tracking is obtained by defining prior distribution $p(r_1)$, $p(r_2)$ in (3). Based on domain expert's knowledge, both priors were designed as the exponential functions of t , which is illustrated in Figure 7. We show only one heart beat cycle which contains 16 frames. In order to reject outliers and achieve robustness, we apply a rectangle filter of $[-12 \ 12]^3$ mm on the final data fusion results to erase the motion replacements which are larger than this size between adjacent frames. The corresponding position of the eliminated boundary point is recalculated based on the bicubic interpolation of its neighbors. After this step we obtained the $p(\mathbf{z}_t | \mathbf{x}_t)$.

E. Postprocessing and Projection

Due to speed considerations, the detection tracker is designed to search on the normal direction of the boundary, and the template tracker is searching along the direction of the gradients. Both of them can not provide smooth tracking results. Let B_f denotes the 3D boundary point-set after the data fusion step. We project B_f onto the PCA shape space calculated from the training stage, and obtain a smooth boundary point set B_s . Please be aware that here the PCA shape model indicates the heart chamber models built for the ASM based boundary classifier in Section III-C. The first 100 components are saved to keep enough details and provide smooth boundaries as well. A surface of the smooth boundary B_s is constructed using the 3D triangulation. Each boundary point P on B_f is projected onto the smooth surface by finding the triangle $S \in T = \{\text{all}$

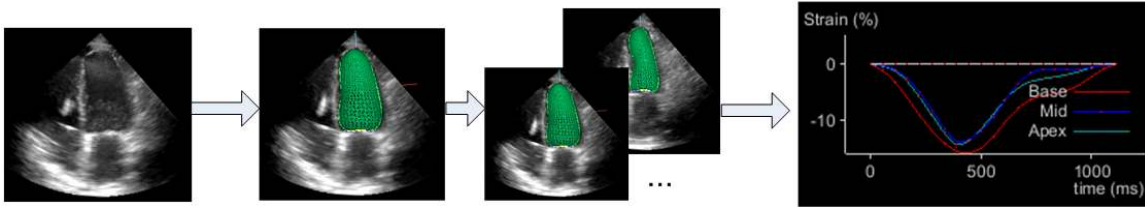


Fig. 8. The working flow chart of the automatic endocardium tracking of the left ventricle and the measurement of the ejection fraction and volume-time curve.

triangles of B_s } which minimize the square distance

$$dist(s, t) = [S(s, t) - P]^2, \quad (9)$$

where $S(s, t) = \mathbf{b} + s\mathbf{e}_0 + t\mathbf{e}_1$ with $(s, t) \in D = \{(s, t) : s \in [0, 1], t \in [0, 1], s + t \leq 1\}$. Here, \mathbf{b} is one vertex in the triangle and \mathbf{e}_0 and \mathbf{e}_1 represent two edges. In this way, we maintain both the tracking accuracy and the mesh smoothness.

V. EXPERIMENTAL RESULTS

The proposed predication based collaborative trackers (PCT) is a general framework which can be applied to track individual heart chamber, the combination of heart chambers, the epicardium and endocardium or the myocardial region, etc. In this section, we demonstrate the experimental results applying PCT on endocardium tracking, the myocardial region dense tracking, and four chambers whole heart tracking.

A. Endocardium Tracking of The Left Ventricle

In total 67 3D ultrasound endocardium motion sequences are collected. Each subject contributes one sequence, therefore we have 67 different subjects. Each 4D (x, y, z, t) motion sequence contains from 11 to 25 3D frames from one cardiac cycle. In total we have 1143 ultrasound 3D volumetric data. Our dataset is much larger than those reported in the literature, e.g., 29 cases with 482 3D frames in [5], 21 cases with about 400 3D frames in [29] and 22 cases with 328 3D frames in [30], etc. The working procedure is shown in Figure 8.

The imaging protocols are heterogeneous with different capture ranges and resolutions along each dimension. The dimensionality of 27 sequences is $160 \times 144 \times 208$ and the other 40 sequences is $160 \times 144 \times 128$. The x, y and z resolution ranges are $[1.24, 1.42]$, $[1.34, 1.42]$ and $[0.85, 0.90]$ mm. In our experiments, we randomly select 36 sequences for training and the rest are used for testing.

The accuracy is measured by the *point-to-mesh* (PTM) error. All 3D points on each frame of the testing sequence are projected onto the corresponding annotated boundary of the test set. In order to make the measurement symmetric we also calculate the reverse mapping distance. The average projection distance from the point to boundary is recorded as the PTM error, e_{ptm} . For a perfect tracking, the e_{ptm} should be equal to zero for each 3D frame. In Table I, we compared the quantitative e_{ptm} using the proposed predication based collaborative trackers (PCT) with tracking by 3D optical flow and tracking by detection.

The 80% column in Table I represents the 80-th percentile of the sorted e_{ptm} , and is commonly used by doctors to evaluate

the usability of the system. For example, if the doctor can tolerate an error of 1.5 mm, they normally expect 80% of the errors to be smaller than this number. The mean e_{ptm} we obtained is 1.28 mm with a 80% error below 1.47 mm. In this experiment we show that collaborative trackers is obviously superior to each individual detection or template tracker.

In Table II we list the quantitative comparison results with other eight algorithms. PCT provides one of the best results in the recent heart segmentation/tracking literature. Although the algorithm in [31] reports better mean PTM errors, there are several differences between these two studies: 1) PCT is completely automatic including the initialization step while [31] initializes the tracker in a semi-automatic manner; 2) PCT is roughly three times faster than [31]; 3) PCT contains 67 patients which is two times larger; 4) The best resolution of the data that PCT had is $1.24 \times 1.34 \times 0.85$ mm while the data in [31] had resolution $1 \times 1 \times 0.7$ mm; 5) The accuracy reported in PCT is based on a random half training, half testing manner while [31] reports the leave-one-out result.

The systolic-diastolic function can be visualized using the volume-time curve which represents continuous endocardium volume change over the time t . It is an important diagnosis term to evaluate the health condition of the heart. In Figure 9, we illustrate one heart cycle of two systolic-diastolic functions. The curves for all three tracking algorithms and the ground-truth annotation are shown. Our algorithm (PCT) provides the most similar functions to the ground truth curves. Given the systolic-diastolic functions, we had calculated the ejection-fraction (EF) rate compared with the ground-truth annotations and the results are listed in Table III. The ejection fraction is defined as $EF = (v_{ED} - v_{ES})/v_{ED}$, where v_{ED} and v_{ES} represent the volumes of the left ventricle at the ED and ES phase, respectively. A normal LV ejection fraction is 55% to 70% percent. The decreased ejection fraction may indicate problems such as weakness of heart muscle, problem with heart's valves or long-standing, uncontrolled high blood pressure, etc.

Two types of errors are frequent in tracking endocardium in 3D ultrasound. The first type is the leakage error, $e_{leakage}$, which often happens on the mitral valve region. This is introduced by the similar appearance of the mitral valve compared to the endocardium boundary. A good left ventricle boundary tracking algorithm should not follow the motion of the leaflets of the mitral valve. Tracking by detection failed on frame 8 (row 3, columns 3 and 4 in Figure 10b) because it always searches for what it learned in the training stage, but ignores all the local image information and temporal consistency.

The second type is the shrinkage error, e_{shrink} , which often

TABLE I
THE *point-to-mesh* (PTM) ERRORS MEASURED IN MILLIMETERS USING THREE TRACKING ALGORITHMS.

	Mean	Variance	Median	Min	Max	80%
3D optical flow	2.68	1.28	2.39	0.94	10.38	3.23
Tracking by detection	1.61	1.24	1.31	0.59	9.89	1.89
PCT	1.28	1.11	1.03	0.38	9.80	1.47

TABLE II
THE *point-to-mesh* (PTM) ERRORS OF PCT COMPARED WITH EIGHT DIFFERENT SEGMENTATION AND TRACKING ALGORITHMS. EMPTY FIELD MEANS THE INFORMATION IS NOT AVAILABLE. BY DEFAULT THE ERROR IS MEASURED WITH MILLIMETERS. IN [32] THE AUTHOR ONLY REPORTS VOXEL ERROR WITHOUT RESOLUTION INFORMATION AVAILABLE, AND IT WAS TEST ON 10 SCANS FROM THE SAME SUBJECT, DENOTED AS 1(10) IN THE TABLE

	Zhu [30]	Wolf [33]	Myronenko[32]	Lin [34]	Hansegård1 [35]	Hansegård2 [36]	Orderud [29]	Leung [31]	PCT
Mean	1.45	3.44	1.03 voxel	1.64	3.4	2.2	2.7	1.19	1.28
Variance	0.3	1.18	0.63 voxel	0.50	2.3	0.56		0.47	1.11
Segmentation	×	×		×	×	×	×	×	×
Tracking			×		×	×	×	×	×
#Subjects	22	20	1(10)	24	36	21	21	35	67

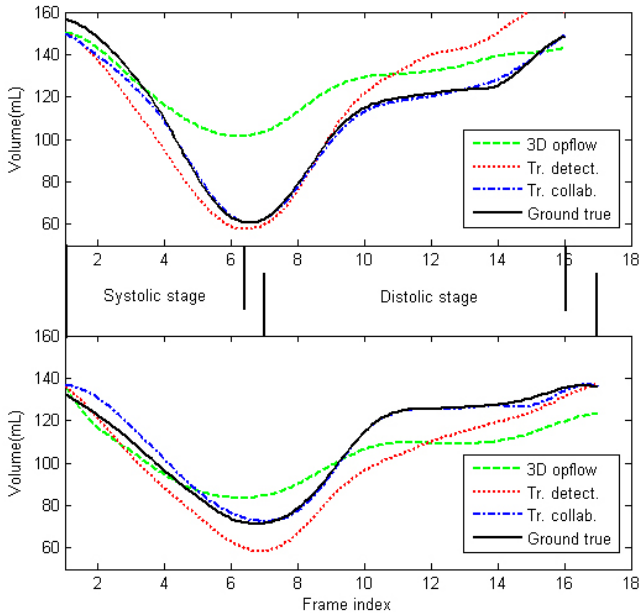


Figure 9. Two volume-time curves demonstrate the whole cardiac cycle, which includes the systole stage and the diastole stage.

happens on the apex region. The 3D optical flow failed on frame 6 (row 2, columns 5 and 6 in Figure 10c) because of the low image quality around the apex region. This proves that motion prior is necessary to obtain enough shrinkage for endocardium tracking in the 3D echocardiography because of the low quality of ultrasound imaging.

Using our algorithm, shown in Figure 10a, neither of the errors are observed. Therefore we achieve more accurate tracking results both visually and quantitatively.

The most important practical consideration for 3D tracking is computational complexity. One of the major reasons to propose marginal space learning, steerable features, registration based reverse mapping and one-step forward prediction in PCT is the speed consideration. The currently C++ implementation of PCT requires 1 – 1.5 seconds to process one 3D volume containing $160 \times 148 \times 208 = 4,925,440$ voxels. It takes

about 20 seconds for the whole motion sequence. PCT is at least two times faster than the slice-cut algorithm presented in [6] which is even not directly working on 3D volumetric data, three times faster than [31], and about hundreds of times faster than [32] which reported using a MATLAB implementation.

TABLE III
THE *ejection-fraction* (EF) RATE ESTIMATION ERRORS (%) AND LEFT VENTRICLE VOLUME ESTIMATION ERROR (ML) USING PCT COMPARED WITH THE GROUND-TRUTH.

	Mean	Std	Median	Min	Max	80%
EF(%)	1.01	5.23	0.82	0.22	9.89	1.41
Volume(ml)	1.32	6.27	1.15	0.25	9.54	1.53

B. Dense Myocardial Region Tracking of The Left Ventricle

In this section, we demonstrate the performance of the proposed tracking algorithm for dense tracking in myocardial region. The extension of the endocardium tracking to myocardial region tracking is straightforward. In the starting frame (typically the end-systole or end-diastole cardiac phase), we initialize the tracking process automatically by detecting the endocardial and epicardial boundary of the left ventricle (LV), the same tracking procedure described in Section IV is followed to track the epicardium, endocardium and the myocardial region of the left ventricle. The final dense tracking results are achieved by tessellating the whole myocardium into a 3D dense mesh. The procedure is shown in Figure 11.

In the myocardial region dense tracking, high frame-rate 3D ultrasound motion sequences were acquired with the average volume size of $200 \times 200 \times 140$. The average temporal resolution is 44 frames per second.

1) *In-Vitro Study*: To evaluate the accuracy of PCT for dense myocardial region tracking, we performed an *in-vitro* experiment on animals. The ground-truth motion was generated by a rotation device and a water pump controlling the stroke volume. Two crystals were implanted in the apical and middle regions of the left ventricle, respectively, to measure the myocardial movement. The average distance between two crystals is 30mm. Four volumetric ultrasound sequences were

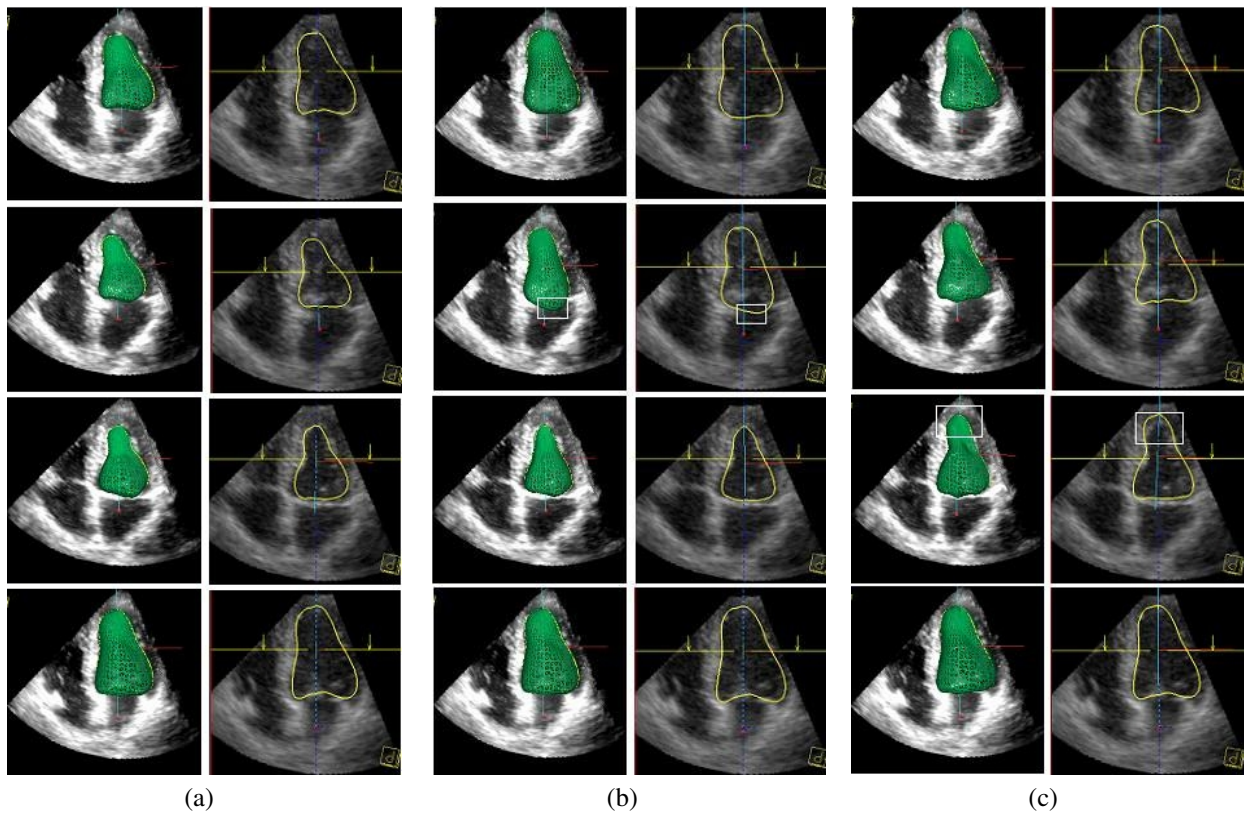


Figure 10. Comparative tracking results on a testing sequence with 12 frames. (a) Tracking results using the proposed method. (b) Tracking by detection. (c) Tracking using 3D optical flow. The rows correspond to frame indices 1, 6, 8 and 10, respectively.



Figure 11. Example initializations of the endocardial and epicardial boundary of the left ventricle (LV). From left to right, the endo- and epi-cardial mesh shown in multi-planar reformatted planes (MPRs): apical four chamber plane, apical three chamber plane, apical two chamber plane, and short axis middle plane. The last subfigure shows the resulting tessellated 3D mesh.

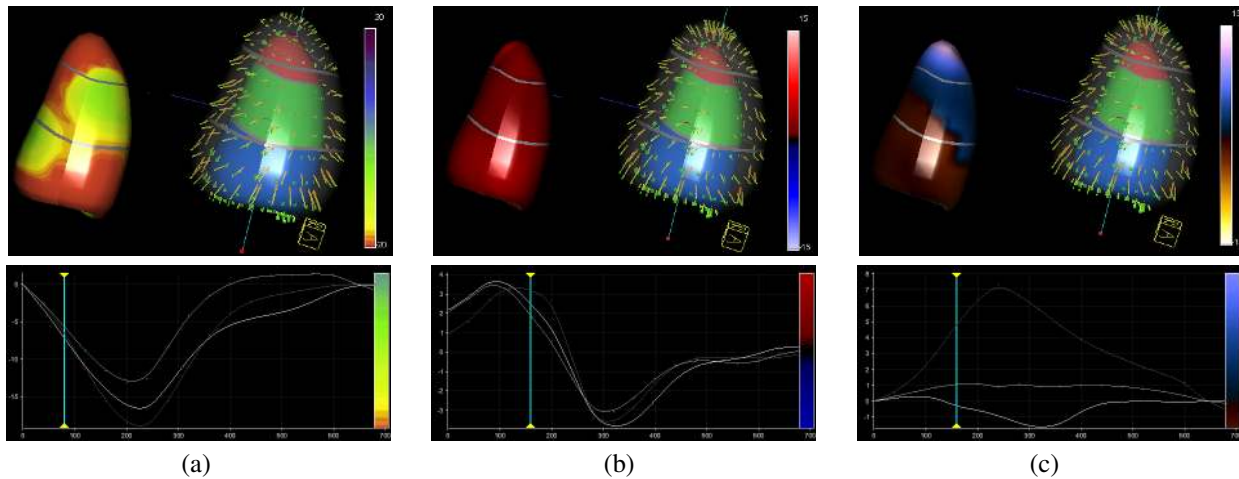


Figure 12. Many clinically relevant measurements can be calculated after myocardium tracking, including: (a) strain (a) strain, (b) velocity, and (c) displacement values in the longitudinal, radial, and circumferential direction respectively.

TABLE IV
IN-VITRO EXPERIMENTS ON BOTH (A) ROTATION AND (B) DISPLACEMENT DATA.

Rotation(degrees)	10	15	20	25	Displacement(mm)	0.82	1.29	2.02
Estimation	9.3	13.5	18.1	21.8	Estimation	0.9	1.54	2.31
Accuracy	93%	90%	91%	87%	Accuracy	90%	81%	91%

(a) Rotation Data

(b) Displacement Data

acquired with 10, 15, 20, and 25 rotation degrees, respectively, and three sequences with different stroke volumes. As reported in Table IV, our tracking results are consistent with the ground-truth measurements on both rotation and displacement data. The displacements in Table IVb were computed based on a 30mm reference length.

Furthermore, to evaluate the PCT tracking results for myocardial strain estimation, we compare them against the crystal measurements for the same subjects in the *in-vitro* study. The ground-truth longitudinal Lagrangian strain [37] can be computed based on the displacement reported in Table IVb, where the two crystals were implanted in the apical and middle regions of the left ventricle, respectively. Table V reports the comparison between our estimated strain values and the ones from crystal measurements. The two crystals were implanted in the apical and middle regions of the left ventricle, such that the longitudinal Lagrangian strain can be computed based on the displacement as the ground-truth measurement in the top row. The estimation results in the middle row are computed based on the proposed tracking method. The low difference values in Table V show clearly that the results from our method are consistent with the clinical measurements.

2) *In-Vivo Study*: To evaluate the robustness of the algorithm, we tested it on a large data set including 503 volumetric ultrasound sequences from human subjects. The data set was randomly split into a training set and a testing set, where the training set was used to train the detectors, generate the shape model and produce the motion priors in Section III, while the testing set reflected the performance of the algorithm for unseen data. The results on both the training and testing sets are reported in Table VI. In the first experiment, the data set was evenly split into a training set with 239 sequences and a testing set with the remaining 264 sequences, while in the second experiment the training set (434) and the testing set (69) were not balanced. The error measurements were computed as the *point-to-mesh* (PTM) error, defined in Section V-A, between our estimated mesh and the ground-truth annotations provided by experts. The low error values on both the training and testing data demonstrate the high accuracy and robust performance of the proposed tracking algorithm on both seen and unseen data. An example of strain, velocity and displacement estimation results is shown in Figure 12. In the top row, the left picture in each pair shows the estimated values mapped to the endocardial boundary of the left ventricle, while the right one shows the direction and magnitude of the dense velocity field. The apical, middle, and basal regions are marked in red, green, and blue, respectively. The bottom row shows the plot on each region, where the horizontal axis is time, and the vertical axis is the estimated mechanical parameter value. The vertical blue bar indicates the time stamp of the frame

displayed in the top row. Please note that in Figure 12c the recovered rotation motion in the apical and basal regions are in opposite directions, which shows that our method can recover the twist motion of the left ventricle.

TABLE V
COMPARISON OF THE LONGITUDINAL STRAIN ESTIMATION BETWEEN OUR METHOD AND THE CRYSTAL MEASUREMENTS IN THE *in-vitro* STUDY.

Longitudinal Strain	2.63%	4.11%	6.68%
Estimation	3.43%	5.19%	8.25%
Difference	0.8%	1.08%	1.57%

C. Four Chambers Tracking of The Heart

In order to show the generality of PCT in handling different imaging modality, we also tested it on four chambers tracking in 3D cardiac CT. The same procedures described in Section IV are applied on each heart chamber separately, and the whole heart mesh is a tessellation of each individual heart chamber’s mesh.

The ED detector and boundary classifier were trained on 323 static cardiac CT volumes from 137 patients with various cardiovascular diseases. The cardiac motion model was trained on additional 20 sequences (each with 10 frames). Our dataset is much larger than those reported in the literature, e.g., 13 in [38], 18 in [5], 27 in [39], and 30 in [40].

The data was collected from 27 institutes over the world (mostly from Germany, the USA, and China) using Siemens Somatom Sensation or Definition scanners. The imaging protocols are heterogeneous with different capture ranges and resolutions. A volume may contain 80 to 350 slices, while the size of each slice is the same with 512×512 pixels. The resolution inside a slice is isotropic and varies from 0.28 mm to 0.74 mm for different volumes. The slice thickness (distance between neighboring slices) is larger than the in-slice resolution and varies from 0.4 mm to 2.0 mm for different volumes.

The four chambers mesh models are shown in Figure 13. The left ventricle, including both epicardium (magenta) and endocardium (green), is shown in Figure 13a. The left atrium is represented by an open mesh separated by a mitral valve, shown in Figure 13b. The combined left heart mesh is shown in Figure 13c. The right ventricle has a more complicated shape and its model is represented by an open mesh, shown in Figure 13d. The mesh model for the right atrium and the combined mesh are shown in Figure 13e and Figure 13f, respectively. The detailed mesh modeling can be found in [41].

The motion model for each chamber in Figure 13 is learned using the same motion manifold learning method described in Section III-A. Object detector and boundary classifier are

TABLE VI
PERFORMANCE ANALYSIS OF MYOCARDIUM TRACKING ON A LARGE DATA SET INCLUDING 503 3D+T ULTRASOUND MOTION SEQUENCES.

measure(mm)	training (239)	testing (264)	training (434)	testing (69)
mean/std	2.21/1.57	2.68/2.63	2.26/1.42	2.64/2.23

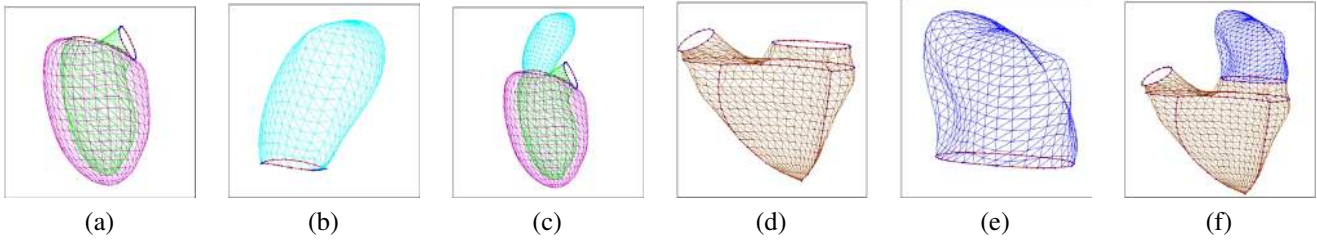


Figure 13. The heart mesh models for the four chambers (a) The mesh model for the left ventricle (LV), with green for the LV endocardium and magenta for the LV epicardium. (b) The mesh model for the left atrium (LA). (c) The combined LV and LA mesh. (d) The mesh model for the right ventricle (RV). (e) The mesh model for the right atrium (RA). (f) The combined mesh model for both RV and RA.

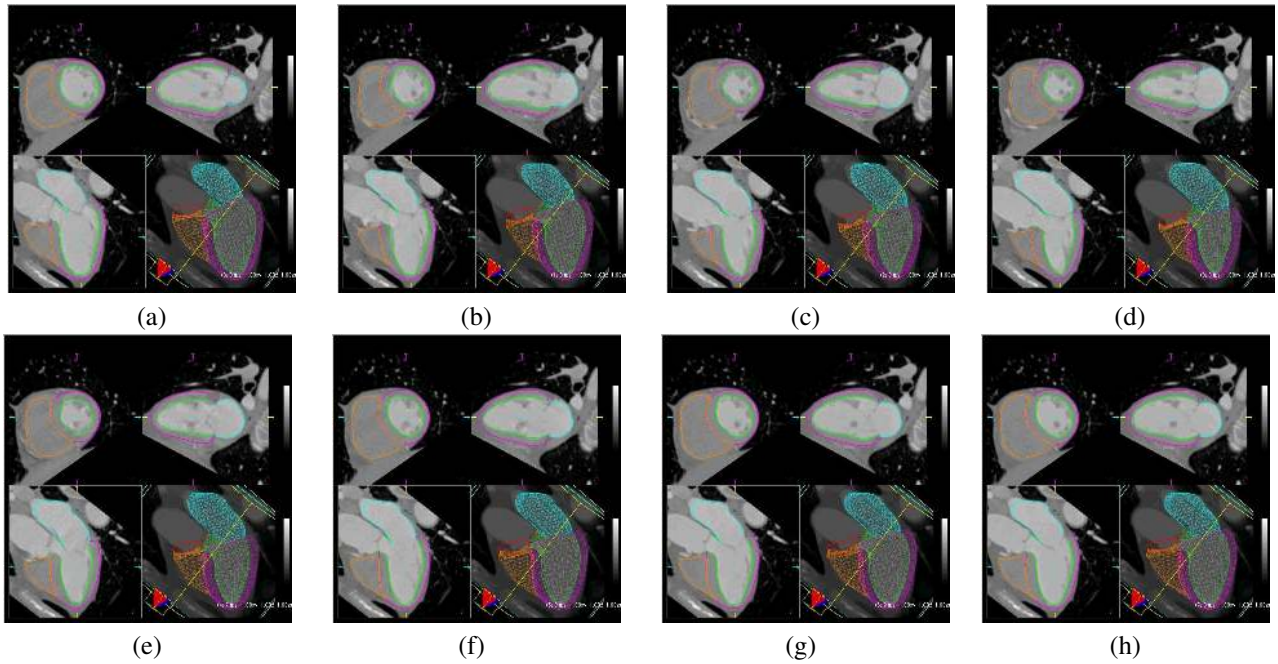


Figure 14. The four chambers tracking results on a testing 3D CT sequence with 8 frames. (a)-(h) represent the 1-8 frame index of the whole sequence.

trained for each heart chamber, respectively. In the tracking stage, the same one step forward prediction procedure described in Section IV is utilized to provide motion prior during the tracking procedure. The robust data fusion is applied to calculate the final motion displacement. The 3D cardiac CT four chambers (LV-epicardium, LV-endocardium, LA, RV, and RA) tracking results for each frame in a CT sequence containing eight frames are shown in Figure 14.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we presented a robust, fast and accurate predication based collaborative track (PCT) and tested in both 3D ultrasound and CT. PCT can process a 3D volume in less than 1.5 seconds and provides subvoxel accuracy. We demonstrated that PCT increases the tracking accuracy and especially speed dramatically. The final accurate results

are achieved by introducing the one-step forward prediction motion prior. The robustness to complex background and weak edges is obtained from the learned discriminative detector and boundary classifiers. The temporal consistency is preserved by the template tracker in the collaborative trackers.

Instead of building specific models manually, all the shape priors and motion patterns in PCT are based on learning. The algorithm is therefore general enough to be extended to other 3D medical tracking problems. The generality of PCT is already proven by a diverse and extensive experiments using three challenging heart tracking applications. Currently, we are working on extending PCT to other tracking applications, such as lung tumor tracking in continuously acquired fluoroscopic video sequences.

REFERENCES

[1] B. Anderson, *Echocardiography: The Normal Examination and Echocardiographic Measurements*, 2nd ed. MGA Graphics, Australia,

- 2007.
- [2] P. Schoenhausen, S. S. Halliburton, A. E. Stillman, and R. D. White, "CT of the heart: Principles, advances, clinical uses," *Cleveland Clinic Journal of Medicine*, vol. 72, no. 2, pp. 127–138, 2005.
 - [3] M. Dewan, C. H. Lorenz, and G. D. Hager, "Deformable motion tracking of cardiac structures (DEMOTRACS) for improved MR imaging," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 2007.
 - [4] P. F. U. Gotardo, K. L. Boyer, J. Saltz, and S. V. Raman, "A new deformable model for boundary tracking in cardiac MRI and its application to the detection of intra-ventricular dyssynchrony," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 736–743, 2006.
 - [5] M. P. Jolly, "Automatic segmentation of the left ventricles in cardiac MR and CT images," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 151–163, 2006.
 - [6] W. Hong, B. Georgescu, X. S. Zhou, S. Krishnan, Y. Ma, and D. Comaniciu, "Database-guided simultaneous multi-slice 3D segmentation for volumetric data," *Proc. European Conference on Computer Vision*, vol. 4, pp. 397–409, 2006.
 - [7] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuring, and D. Comaniciu, "Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features," *Proc. IEEE International Conference on Computer Vision*, 2007.
 - [8] L. Yang, B. Georgescu, Y. Zheng, P. Meer, and D. Comaniciu, "3D ultrasound tracking of the left ventricle using one-step forward prediction and data fusion of collaborative trackers," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 2008.
 - [9] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computer Survey*, vol. 38, no. 4, pp. 13.1–13.45, 2006.
 - [10] S. Avidan, "Ensemble tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 261–271, 2007.
 - [11] Y. Li, H. Ai, T. Yamashita, S. Lao, and M. Kawade, "Tracking in low frame rate video: A cascade particle filter with discriminative observers of different lifespans," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 2007.
 - [12] T. Zhao and R. Nevatia, "3D tracking of human locomotion: A tracking as recognition approach," *Proc. IEEE International Conference on Pattern Recognition*, pp. 1054–1060, 2002.
 - [13] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models: Their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
 - [14] M. Isard and A. Blake, "CONDENSATION — Conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 28, no. 1, pp. 5–28, 1998.
 - [15] Z. Khan, T. Balch, and F. Dellaert, "An MCMC-based particle filter for tracking multiple interacting targets," *Proc. European Conference on Computer Vision*, vol. 4, pp. 279–290, 2004.
 - [16] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. John Wiley and Sons, 1998.
 - [17] J. Tenebaum, V. de Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
 - [18] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 511–518, 2001.
 - [19] A. Barbu, V. Athitsos, B. Georgescu, S. Boehm, P. Durlak, and D. Comaniciu, "Hierarchical learning of curves application to guidewire localization in fluoroscopy," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 2007.
 - [20] P. Dollar, Z. Tu, and S. Belongie, "Supervised learning of edges and object boundaries," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 1964–1971, 2006.
 - [21] Z. Tu, "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering," *Proc. IEEE International Conference on Computer Vision*, vol. 2, pp. 1589–1596, 2005.
 - [22] J. Lim and M. Yang, "A direct method for modeling non-rigid motion with thin plate spline," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1196–1202, 2005.
 - [23] J. Chen and Q. Ji, "Online spatial-temporal data fusion for robust adaptive tracking," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 2007.
 - [24] I. Matthews, T. Ishikawa, and S. Baker, "The template update problem," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 810–815, 2004.
 - [25] D. Ormoneit, M. J. Black, T. Hastie, and H. Kjellström, "Representing cyclic human motion using functional analysis," *Image and Vision Computing*, vol. 23, no. 14, pp. 1264–1276, 2005.
 - [26] R. Urtasun, D. J. Fleet, and P. Fua, "Temporal motion models for monocular and multiview 3D human body tracking," *Computer Vision and Image Understanding*, vol. 104, pp. 157–177, 2006.
 - [27] S. Zhou, J. Shao, B. Georgescu, and D. Comaniciu, "Pairwise active appearance model and its application to echocardiography tracking," *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. 4190, pp. 736–743, 2006.
 - [28] Z. Fan, Y. Wu, and M. Yang, "Multiple collaborative kernel tracking," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 502–509, 2005.
 - [29] F. Orderud, J. Hansgård, and S. I. Rabben, "Real-time tracking of the left ventricle in 3D echocardiography using a state estimation approach," *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. 4791, pp. 858–865, 2007.
 - [30] Y. Zhu, X. Papademetris, A. Sinusas, and J. S. Duncan, "Segmentation of myocardial volumes from real-time 3D echocardiography using an incompressibility constraint," *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. 4791, pp. 44–51, 2007.
 - [31] K. Y. E. Leung, M. G. Danilouchkine, M. van Stralen, N. de Jong, A. F. W. van der Steen, and J. G. Bosch, "Probabilistic framework for tracking in artifact-prone 3D echocardiograms," *Medical Image Analysis*, vol. 14, no. 5, pp. 750–758, 2010.
 - [32] A. Myronenko, X. B. Song, and D. J. Sahn, "LV motion tracking from 3D echocardiography using textural and structural information," *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. 4792, pp. 428–435, 2007.
 - [33] I. Wolf, M. Hastenteufel, R. de Simone, M. Vetter, G. Glombitza, S. Mottl Link, C. Vahl, and H. Meinzer, "ROPES: A semiautomated segmentation method for accelerated analysis of three-dimensional echocardiographic data," *IEEE Trans. Medical Imaging*, vol. 21, no. 9, pp. 1091–1104, September 2002.
 - [34] N. Lin, W. Yu, and J. S. Duncan, "Combinative multi-scale level set framework for echocardiographic image segmentation," *Medical Image Analysis*, vol. 7, no. 4, pp. 529–537, 2002.
 - [35] J. Hansegård, F. Orderud, and S. I. Rabben, "Real-time active shape models for segmentation of 3D cardiac ultrasound," *Proceedings of the 12th International Conference on Computer Analysis of Images and Patterns*, pp. 157–164, 2007.
 - [36] J. Hansegård, S. Urheim, K. Lunde, and S. I. Rabben, "Constrained active appearance models for segmentation of triplane echocardiograms," *IEEE Trans. Medical Imaging*, vol. 26, no. 10, pp. 1391–1400, 2007.
 - [37] J. D'hooge, A. Heimdal, F. Jamal, T. Kukulski, B. Bijnens, F. Rademakers, L. Hatle, P. Suetens, and G. R. Sutherland, "Regional strain and strain rate measurements by cardiac ultrasound: Principles, implementation and limitations," *European Journal of Echocardiography*, vol. 1, no. 3, pp. 154–170, 2000.
 - [38] O. Ecabert, J. Peters, and J. Weese, "Modeling shape variability for full heart segmentation in cardiac computed-tomography images," *SPIE Medical Imaging*, pp. 1199–1210, 2006.
 - [39] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Modeling shape variability for full heart segmentation in cardiac computed-tomography images," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 193–199, 1997.
 - [40] D. Fritz, D. Rinck, R. Unterhinninghofen, R. Dillmann, and M. Scheuring, "Automatic segmentation of the left ventricle and computation of diagnostic parameters using regiongrowing and a statistical model," *SPIE Medical Imaging*, pp. 1844–1854, 2005.
 - [41] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuring, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features," *IEEE Trans. Medical Imaging*, vol. 27, no. 11, pp. 1668–1881, 2008.