Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal Space Learning and Steerable Features

Yefeng Zheng[1], Adrian Barbu[1], Bogdan Georgescu[1], Michael Scheuering[2], and Dorin Comaniciu[1]
[1]Integrated Data Systems Department, Siemens Corporate Research, USA
[2]Siemens Medical Solutions, Germany
{yefeng.zheng, adrian.barbu, bogdan.georgescu, michael.scheuering, dorin.comaniciu}@siemens.com

Abstract

Multi-chamber heart segmentation is a prerequisite for global quantification of the cardiac function. The complexity of cardiac anatomy, poor contrast, noise or motion artifacts makes this segmentation problem a challenging task. In this paper, we present an efficient, robust, and fully automatic segmentation method for 3D cardiac computed tomography (CT) volumes. Our approach is based on recent advances in learning discriminative object models and we exploit a large database of annotated CT volumes. We formulate the segmentation as a two step learning problem: anatomical structure localization and boundary delineation. A novel algorithm, Marginal Space Learning (MSL), is introduced to solve the 9-dimensional similarity search problem for localizing the heart chambers. MSL reduces the number of testing hypotheses by about six orders of magnitude. We also propose to use steerable image features, which incorporate the orientation and scale information into the distribution of sampling points, thus avoiding the time-consuming volume data rotation operations. After determining the similarity transformation of the heart chambers, we estimate the 3D shape through learning-based boundary delineation. Extensive experiments on multi-chamber heart segmentation demonstrate the efficiency and robustness of the proposed approach, comparing favorably to the state-of-the-art. This is the first study reporting stable results on a large cardiac CT dataset with 323 volumes. In addition, we achieve a speed of less than eight seconds for automatic segmentation of all four chambers.

1. Introduction

Cardiac computed tomography (CT) is an important imaging modality for diagnosing cardiovascular disease and it can provide detailed anatomic information about the cardiac chambers, large vessels or coronary arteries. Segmentation of cardiac chambers is a prerequisite for quantitative functional analysis and various approaches have been proposed in the literature [6, 7]. Except for a few works [5, 24], most of the previous research focuses on the left ventricle (LV) segmentation. However, complete segmentation of all four heart chambers, as shown in Fig. 1, can help to diagnose diseases in other chambers, *e.g.*, left atrium (LA) fibrillation, right ventricle (RV) overload or to perform dyssynchrony analysis.

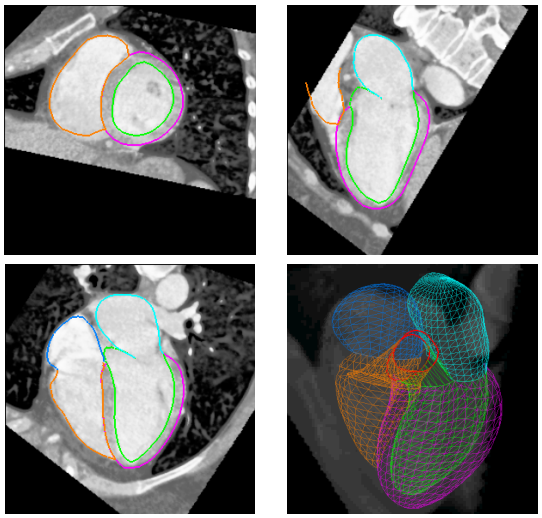


Figure 1. Complete segmentation of all four chambers in a CT volume with green for the left ventricle (LV) endocardial surface, magenta for LV epicardial surface, cyan for the left atrium (LA), brown for the right ventricle (RV), and blue for the right atrium (RA).

There are two tasks for a non-rigid object segmentation problem: object localization and boundary delineation. Most of the previous approaches focus on boundary delineation based on active shape models (ASM) [22], active appearance models (AAM) [1, 13], and deformable models [2, 4, 5, 8, 12, 17]. There are a few limitations inherent in these techniques: 1) Most of them are semi-

automatic and manual labeling of a rough position and pose of the heart chambers is needed. 2) They are likely to get stuck in local strong image evidence. Other techniques are straightforward extensions of 2D image segmentation to 3D [10, 18, 25]. The segmentation is performed on each 2D slice and the results are combined to get the final 3D segmentation. However, such techniques cannot fully exploit the benefit of 3D imaging in a natural way. Lorenzo-Valdés *et al*. [11] proposed a registration based approach, but its performance is not clear for large datasets.

Object localization is required for an automatic segmentation system and discriminative learning approaches have proved to be efficient and robust for solving 2D problems. In these methods, shape detection or localization is formulated as a classification problem: whether an image block contains the target shape or not [16, 23]. To build a robust system, a classifier only has to tolerate limited variation in object pose. The object is found by scanning the classifier over an exhaustive range of possible locations, orientations, scales or other parameters in an image. This searching strategy is different from other parameter estimation approaches, such as deformable models, where an initial estimate is adjusted (*e.g*., using the gradient descent technique) to optimize a predefined objective function.

Exhaustive searching makes the system robust under local minima, however there are two challenges to extend the learning based approaches to 3D. First, the number of hypotheses increases exponentially with respect to the dimensionality of the parameter space. For example, there are nine degrees of freedom for the anisotropic similarity transformation[1], namely three translation parameters, three rotation angles, and three scales. Suppose we search n discrete values for each dimension, the number of tested hypotheses is n^9 (for a very coarse estimation with a small n=5, n^9=1,953,125).The computational demands are beyond the capabilities of current desktop computers. Due to this limitation, previous approaches often constrain the search to a lower dimensional space. For example, only the position and isotropic scaling (4D) is searched in the generalized Hough transformation based approach [19]. Hong *et al*. [9] extended the learning based approach to a 5D parameter space for semi-automatic segmentation. The second challenge is that we need efficient features to search the orientation and scale spaces. Haar wavelet features can be efficiently computed for translation and scale transformations [15, 23]. However when searching for rotation parameters one either has to rotate the feature templates or rotate the volume which is very time consuming. The efficiency of image feature computation becomes more important when combined with a very large number of test hypotheses.

[1]The ordinary similarity transformation allows only isotropic scaling. In this paper, we search for anisotropic scales to cope better with the non-rigid deformation of the shape.

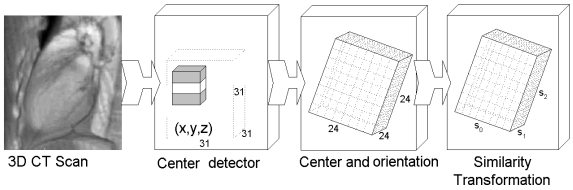


Figure 2. 3D object localization using marginal space learning.

1.1. Overview of Our Approach

In this paper, we propose two simple but elegant techniques, marginal space learning (MSL) and steerable features, to solve the above challenges. The idea for MSL is not to learn a classifier directly in the full similarity parameters space but to incrementally learn classifiers on projected sample distributions. As the dimensionality increases, the valid (positive) space region becomes more restricted by previous marginal space classifiers. In our case, we split the estimation into three problems: translation estimation, translation-orientation estimation, and full similarity estimation (Fig. 2). After each step, we maintain multiple candidates to increase the robustness.

Besides reducing the searching space significantly, there is another advantage using MSL: we can use different features or learning methods in each step. For example, in the translation estimation step, since we treat rotation as an intra-class variation, we can use the efficient 3D Haar features [21]. In the translation-orientation and similarity transformation estimation steps, we introduce the steerable features, another major contribution of this paper. Steerable features constitute a very flexible framework where the idea is to sample a few points from the volume under a special pattern. We extract a few local features for each sampling point, such as voxel intensity and gradient. To evaluate the steerable features under a specified orientation, we only need to steer the sampling pattern and no volume rotation is involved.

After similarity transformation estimation, we get an initial estimate of the non-rigid shape. We use learning based 3D boundary detection to guide the shape deformation in the ASM framework. Again, steerable features are used to train local detectors and find the boundary under any orientation, therefore avoiding time consuming volume rotation.

In summary, we make the following contributions:

1. We propose MSL to search the shape space efficiently.
2. We introduce steerable features, which can be evaluated efficiently under any orientation and scale without rotating the volume. These features are also exploited in a learning-based 3D boundary detection scheme.
3. Combining the above techniques, we have implemented a fully automatic, fast, and robust system for multi-chamber heart segmentation in CT volumes.

In the remaining of the paper, we first present our two major contributions, marginal space learning in Section 2 and steerable features in Section 3. Their application to 3D object localization is discussed in Section 4. The learning based 3D boundary detection and its application for non-rigid deformation estimation are discussed in Section 5. We demonstrate the robustness of the proposed method on heart chamber segmentation in Section 6. This paper ends with a discussion of the future work in Section 7.

2. Marginal Space Learning

In many cases, the posterior distribution is clustered in a small region in the high dimensional parameter space. It is not necessary to search the whole space uniformly and exhaustively We propose a novel efficient parameter searching method, marginal space learning, to search such clustered space. In MSL, the dimensionality of the search space is gradually increased. Let Ω be the space where the solution to the given problem exists and let P_Ω be the true probability that needs to be learned. The learning and computation are performed in a sequence of marginal spaces

$$\Omega_1 \subset \Omega_2 \subset ... \subset \Omega_n = \Omega \tag{1}$$

such that Ω_1 is a low dimensional space (*e.g.*, 3-dimensional translation instead of 9-dimensional similarity transformation), and for each k, $\dim(\Omega_k) - \dim(\Omega_{k-1})$ is small. A search in the marginal space Ω_1 using the learned probability model finds a subspace $\Pi_1 \subset \Omega_1$ containing the most probable values and discards the rest of the space. The restricted marginal space Π_1 is then extended to $\Pi_1^e = \Pi_1 \times X_1 \subset \Omega_2$. Another stage of learning and testing is performed on Π_1^e obtaining a restricted marginal space $\Pi_2 \subset \Omega_2$ and the procedure is repeated until the full space Ω is reached. At each step, the restricted space Π_k is one or two orders of magnitude smaller than $\Pi_{k-1} \times X_k$. This results in a very efficient algorithm with minimal loss in performance.

Fig. 3 illustrates a simple example for 2D space searching. A classifier trained on $p(y)$ can quickly eliminate a large portion of the search space. We can then train a classifier in a much smaller region (region 2 in Fig. 3) for joint distribution $p(x, y)$. **Note that MSL is significantly different from a classifier cascade [23]** . In a cascade the search and learning are performed in the *same* space while for MSL the learning and search space is gradually increased.

MSL is similar to particle filters [14] in the way we handle multiple hypotheses. Both approaches keep a limited number of samples to represent underlying probability distributions. Samples are propagated sequentially to the following stages and pruned by the model.

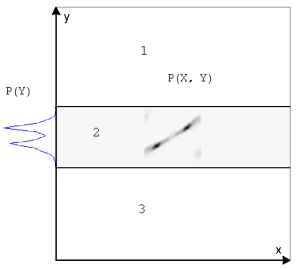
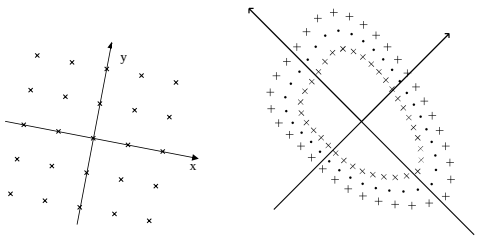


Figure 3. Marginal space learning. A classifier trained on $p(y)$ can quickly eliminate a large portion (regions 1 and 3) of the search space. Another classifier is then trained on restricted space for $p(x, y)$.



(a) (b)

Figure 4. Sampling patterns in the steerable features (visualized in 2D for clearance). (a) A regular sampling pattern. (b) Sampling pattern with points around the shape boundary.

3. Steerable Features

In the section, we present another contribution of this paper, steerable features, which enjoys the advantages of both global and local features. Global features, such as 3D Haar wavelet features, are effective to capture the global information (*e.g.*, orientation and scale) of an object. As shown in [21], pre-alignment of the image or volume is important for a learning based approach. However, it is very time consuming to rotate a 3D volume, so 3D Haar wavelet features are not efficient for orientation estimation. Local features are fast to evaluate but they lose the global information of the whole object. In this paper, we propose a new framework, steerable features, which can capture the orientation and scale of the object and at the same time be very efficient.

Basically, we sample a few points from the volume under a special pattern. We then extract a few local features for each sampling point, such as voxel intensity and gradient. Fig. 4a shows a regular sampling pattern. Suppose we want to test if hypothesis $(X, Y, Z, \psi, \phi, \theta, S_x, S_y, S_z)$ is a good estimation of the similarity transformation of the object in the volume. We put a local coordinate system centered on the candidate position (X, Y, Z) and align the axes with the hypothesized orientation (ψ, ϕ, θ). We uniformly sample a few points along each coordinate axis inside a rectangle (represented as 'x' in Fig. 4a). The sampling step along an axis is proportional to the scale (S_x, S_y, and S_z) of the shape in that direction to incorporate the scale infor-

mation. The steerable features are a general framework and different sampling patterns can be defined depending on the application to incorporate the orientation and scale information. For many shapes, since the boundary provides critical information about the orientation and scale, we can strategically put sampling points around the boundary, as shown in Fig. 4b.

For each sampling point, we extract a set of local features based on the intensity and gradient. For example, given a sampling point (x, y, z), if its intensity is I and the gradient is $g = (g_x, g_y, g_z)$, the following features are used: I, $\sqrt{I}$, I^2, I^3, $\log I$, g_x, g_y, g_z, $\|g\|$, $\sqrt{\|g\|}$, $\|g\|^2$, $\|g\|^3$, $\log \|g\|$, ..., etc. In total, we have 24 local features for each sampling point. Suppose there are P sampling points (often in the order of a few hundreds to a thousand), we get a feature pool containing $24 \times P$ features. This features are used to train simple classifiers and we use probabilistic boosting-tree (PBT) [20] to combine them to get a strong classifier for the given parameters.

Instead of aligning the volume to the hypothesized orientation to extract Haar wavelet features [21], we steer the sampling pattern. This is where the name "steerable features" comes from[2]. In the steerable feature framework, each feature is local, therefore efficient. The sampling pattern is global to capture the orientation and scale information. In this way, it combines the advantages of both global and local features.

4. 3D Object Localization

In this section, we present our 3D object localization scheme using MSL and steerable features. To increase the speed, we use a pyramid-based coarse-to-fine strategy and the similarity transformation estimation is performed on a low-resolution (3 mm) volume.

4.1. Training of Object Position Estimator

As shown in Fig. 2, first, we estimate the position of the object inside the volume. We treat the orientation and scale as the intra-class variations, therefore learning is constrained in a marginal space with three dimensions. Haar wavelet features are very fast to compute and have been shown to be effective for many applications [15, 23]. Therefore, we use 3D Haar wavelet features for learning in this step. Readers are referred to [15, 21, 23] for details about 3D Haar wavelet features.

Given a set of candidates, we split them into two groups, positive and negative, based on their distance to the ground truth. The error in object position and scale estimation is not comparable with that of orientation estimation directly. Therefore, we define a normalized distance measure using the searching step size.

$$E = \max_{i=1,\ldots,N} |V_i^e - V_i^t| / \text{SearchStep}_i, \qquad (2)$$

where is V_i^e is the estimated value for dimension i and V_i^t is the ground truth. A sample is regarded as a positive one if $E \leq 1.0$ and all the others are negative samples. The searching step for position estimation is one voxel, so a positive sample (X, Y, Z) should satisfy

$$\max\{|X - X_t|, |Y - Y_t|, |Z - Z_t|\} \leq 1 \text{ voxel}, \qquad (3)$$

where (X_t, Y_t, Z_t) is the ground truth of the object center.

Given a set of positive and negative training samples, we extract 3D Haar wavelet features and train a classifier using the probabilistic boosting-tree (PBT) [20]. Given a trained classifier, we use it to scan a training volume and preserve a small number of candidates (100 in our experiments), such that the solution is among top hypotheses.

4.2. Training of Position-Orientation and Similarity Transformation Estimators

Suppose for a given volume, we have 100 candidates, (X_i, Y_i, Z_i), $i = 1 \ldots 100$, for the object position. We then estimate both the position and orientation. The hypothesized parameter space is six dimensional so we need to augment the dimension of candidates. For each candidate of the position, we scan the orientation space uniformly to generate hypotheses for orientation estimation. It is well-known that the orientation in 3D can be represented as three Euler angles, ψ, ϕ, and θ. We scan the orientation space using a step size of 0.2 radians (11 degrees). For each candidate (X_i, Y_i, Z_i), we augment it with N (about 1000) hypotheses about orientation, $(X_i, Y_i, Z_i, \psi_j, \phi_j, \theta_j)$, $j = 1 \ldots N$. Some hypotheses are close to the ground truth (positive) and others are far away (negative). The learning goal is to distinguish the positive and negative samples using image features (here, steerable features). A hypothesis $(X, Y, Z, \psi, \phi, \theta)$ is regarded as a positive sample if it satisfies both Eq. 3 and

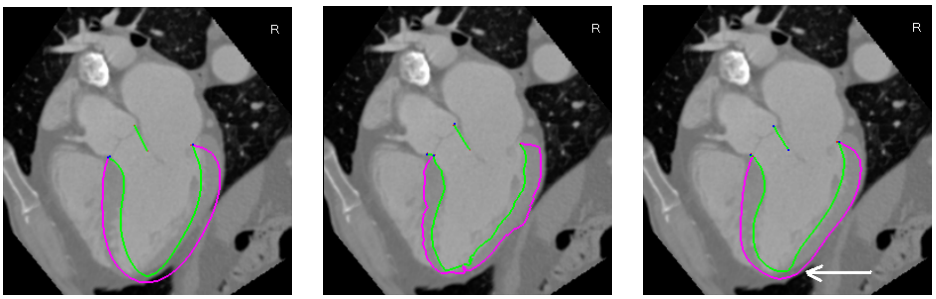
$$\max\{|\psi - \psi_t|, |\phi - \phi_t|, |\theta - \theta_t|\} \leq 0.2, \qquad (4)$$

where $(\psi_t, \phi_t, \theta_t)$ represent the orientation ground truth. All the other hypotheses are regarded as negative samples.

Since aligning 3D Haar wavelet features to a specified orientation is not efficient, we use the proposed steerable features in the following steps. We train a classifier using PBT and the steerable features. The trained classifier is used to prune the hypotheses to preserve only a few candidates (50 in our experiments).

The similarity (adding the scale) estimation step is analogous except learning is performed in the full nine dimensional similarity transformation space. The dimension of each candidate is augmented by scanning the scale subspace uniformly and exhaustively.

[2] It has no relationship with the well known steerable filters



(a) (b) (c)

Figure 5. Example of non-rigid deformation estimation for LV with green for endocardial surface and magenta for epicardial surface. (a) Detected mean shape. (b) After boundary adjustment. (c) Final delineation by projecting the adjusted shape onto a shape subspace (50 dimensions).

4.3. Testing Procedure

This section provides a summary about the testing procedure on an unseen volume. The input volume is first normalized to 3 mm isotropic resolution, and all voxels are scanned using the trained position estimator. Top 100 candidates, (X_i, Y_i, Z_i), $i = 1 \dots 100$, are kept. Each candidate is augmented with N (about 1000) hypotheses about orientation, $(X_i, Y_i, Z_i, \psi_j, \phi_j, \theta_j)$, $j = 1 \dots N$. Next, the trained translation-orientation classifier is used to prune these $100 \times N$ hypotheses and the top 50 candidates are retained, $(\hat{X}_i, \hat{Y}_i, \hat{Z}_i, \hat{\psi}_i, \hat{\phi}_i, \hat{\theta}_i)$, $i = 1 \dots 50$. Similarly, we augment each candidate with M (also about 1000) hypotheses about scaling and use the trained classifier to rank these $50 \times M$ hypotheses. The goal is to obtain a single estimate of the similarity transformation. We tried several methods to aggregate multiple candidates and found a simple averaging of the top K ($K = 100$) gives the best estimate.

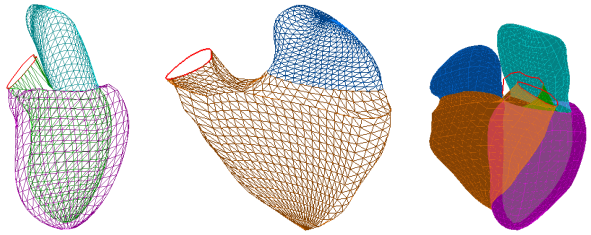
In terms of computational complexity, for translation estimation, all voxels are scanned (about 260,000 for a small $64 \times 64 \times 64$ volume at the 3 mm resolution) for possible object position. There are about 1000 hypotheses for orientation and scale each. If the parameter space is searched uniformly and exhaustively, there are about 2.6×10^{11} hypotheses to be tested! However, using MSL, we only test about $260,000 + 100 \times 1000 + 50 \times 1000 = 4.1 \times 10^5$ hypotheses and reduce the testing by almost six orders of magnitude.

5. Non-Rigid Deformation Estimation

After the first stage, we get the position, orientation, and scale of the object. We align the mean shape with the estimated transformation to get a rough estimate of the object shape. Fig. 5a shows the aligned left ventricle (LV) for heart chamber segmentation in a cardiac CT volume.

We train a set of local boundary detectors using the proposed steerable features with the regular sampling pattern (as shown in Fig. 4a). The boundary detectors are then used to move each landmark point to the optimal position where the estimated boundary probability is maximized. Since more accurate delineation of the shape boundary is desired, this stage is performed on the original high resolution volume. Fig. 5b shows the adjusted shape of LV, which follows the boundary well but is not smooth and unnatural shape may be generated. Shape constraint is enforced by projecting the adjusted shape onto a shape subspace to get the final result [3], as shown in Fig. 5c. The arrow in the figure indicates the region with better boundary delineation.

Our non-rigid deformation estimation approach is within the ASM framework. The major difference is that we use a learning based 3D boundary detector, which is more robust under complex background. Readers are referred to [3] for more details about ASM.



(a) (b) (c)

Figure 6. Triangulated heart surface model. (a) LV and LA. (b) RV and RA. (c) Combined four-chamber model.

6. Experiments

In this section, we demonstrate the performance of the proposed method for multi-chamber localization and delineation in cardiac CT volumes. As shown in Fig. 6, triangulated surface meshes are used to represent the anatomical structures. We delineate both the endo- and epi-cardial surfaces for LV, but only the endocardial surface for other chambers. During manual labeling, we establish correspondence between mesh points crossing volumes, therefore, we can build the statistical shape model for ASM [3]. Details

about correspondence establishment are out the scope of this paper. In the following experiments, each chamber is processed independently. The detected meshes from different chambers may cross each other, natural constraints are imposed as post-processing to solve the conflict.

6.1. Data Set

We collected and annotated 323 cardiac CT volumes from 137 patients with various cardiovascular diseases. The number of patients is significantly larger than those reported in the literature, for example, 10 in [19], 13 in [5], and 18 in [10]. The imaging protocols are heterogeneous with different capture ranges and resolutions. A volume contains 80 to 350 slices and the size of each slice is 512×512 pixels. The resolution inside a slice is isotropic and varies from 0.28 mm to 0.74 mm, while slice thickness varies from 0.4 mm to 2.0 mm for different volumes. Four-fold cross validation is performed to evaluate our algorithm. Special care is taken to prevent volumes from the same patient appear in both the training and test sets. In the following, all the evaluation is done based on four-fold cross validation.

6.2. Experiments on Heart Chambers Localization

In this section, we evaluate the proposed approach for the similarity transformation estimation, using the error measure defined in Eq. 2. Comparing to other error measures (*e.g.*, the weighted Euclidean distance), an advantage of our error measure is that we can easily distinguish optimal and non-optimal estimates. The optimal estimate under any specified searching grid is up-bounded by 0.5, while the error of a non-optimal one is larger than 0.5.

To efficiently explore the high-dimensional searching space using MSL, we keep a small number of candidates after each step. One concern about MSL is that since the space is not fully explored, it may miss the optimal solution at an early stage. In the following, we demonstrate that accuracy only deteriorates slightly in MSL. Fig. 7 shows the error of the best candidate after each step with respect to the number of candidates preserved. The curves are calculated on all 323 volumes based on cross validation. The red line shows the error of the optimal solution under the searching grid. As shown in Fig. 7a for translation estimation (where the curves almost overlap each other), if we keep only one candidate, the average error may be as large as 3.5 voxels. However, by keeping more candidates, the minimum errors decrease quickly. We have a high probability to keep the optimal solution when 100 candidates are preserved. Therefore, after this step, we can reduce the candidates dramatically. For translation-orientation estimation, as shown in Fig. 7b, the errors of the best candidates also decrease quickly with more candidates preserved. Based on the trade-off between accuracy and speed, we preserve 50 candidates. Similarly, after full similarity transformation estimation, the best candidates we get have an error ranging from 1.0 to 1.4 searching steps as shown in Fig. 7c.

Finally, we use simple averaging to aggregate the multiple candidates into the final single estimate. As shown in Fig. 8a, the errors decrease quickly with more candidates for averaging until 100 and after that they saturate. Using 100 candidates for averaging, we achieve an error of about 1.5 to 2.0 searching steps for different chambers. Fig. 8b shows the cumulative errors on all volumes. Without any major failure, our approach is more robust than [5], where the success rate of heart localization is about 90%.

The conclusion of these experiments is that only a small number of candidates are necessary to be preserved after each step, without deteriorating accuracy much.

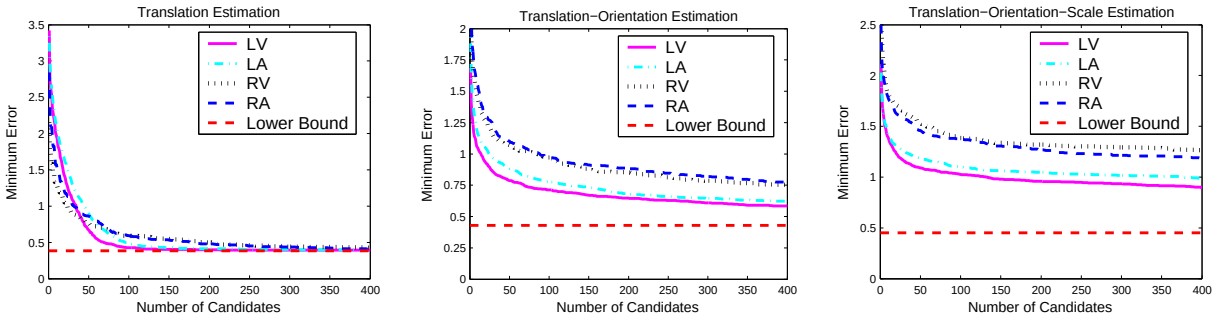
6.3. Experiments on Boundary Delineation

After we get the position, orientation, and scale of the object, we align the mean shape with the estimated transformation. We train five boundary detectors (one for each surface) and use them to guide the shape deformation to fit the boundary (as presented in Section 5).

The accuracy of boundary delineation is measured with the point-to-mesh distance, E_{p2m}. For each point on a mesh, we search for the closest point on the other mesh to calculate the minimum distance. We calculate the point-to-mesh distance from the detected mesh to the ground-truth and vice verse to make the measurement symmetric. Table 1 shows the mean and variance of E_{p2m}. The mean E_{p2m} error of the initialization ranges from 2.78 mm to 3.23 mm. By deforming the mean shape to fit the boundary, we can reduce the error by a half. The mean E_{p2m} error ranges form 1.29 mm to 1.57 mm for different chambers. LV and LA have smaller errors than RV and RA since the contrast of the blood pool in the left side of a heart is consistently higher than the right side due to the using of contrast agents (as shown in Fig. 10).

We also compare our approach to the baseline ASM using non-learning based boundary detection scheme [3]. The same detected mean shape is used to initialize the deformation, and the iteration number in the baseline ASM is tuned to give the best performance. As shown in Table 1, the baseline ASM only slightly reduces the error for weak boundaries (such as LV epicardial, RV, and RA surfaces). It performs much better for strong boundaries, such as LV endocardial and LA surfaces, but it is significantly worse than the proposed method. Fig. 9 shows the cumulative errors of E_{p2m} for the baseline ASM and the proposed approach. Due to the space limit, we only show the results for LV, both endo- and epi-cardial surfaces.

Fig. 10 shows several examples for heart chamber segmentation using the proposed approach. The second row shows a volume with low contrast, our segmentation result is quite good. Our approach is robust even under severe

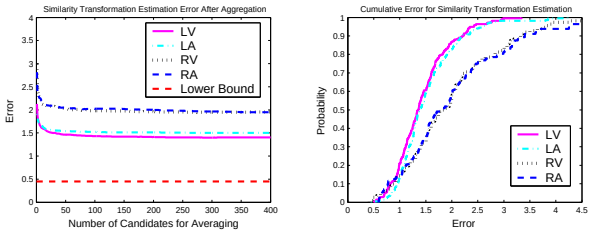


(a) (b) (c)

Figure 7. The error of the best candidate with respect to the number of candidates preserved after each step. (a) Translation, (b) translation-orientation, and (c) full similarity transformation estimation, respectively. The red line shows the lower bound of the error.

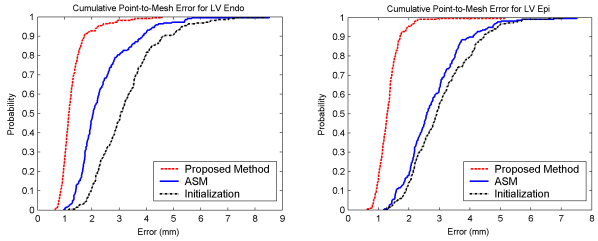
Table 1. Mean and variance (in parentheses) of the point-to-mesh error (in millimeters) for the segmentation of heart chambers on 323 volumes based on cross validation.

	Initialization	Baseline ASM [3]	Our Approach
LV Endo	3.23 (1.17)	2.37 (1.03)	**1.29 (0.53)**
LV Epi	3.05 (1.04)	2.78 (0.98)	**1.33 (0.42)**
LA	2.78 (0.98)	1.89 (1.43)	**1.32 (0.42)**
RV	2.93 (0.75)	2.69 (1.10)	**1.55 (0.38)**
RA	3.09 (0.86)	2.81 (1.15)	**1.57 (0.48)**



(a) (b)

Figure 8. Similarity transformation estimation error by aggregating multiple candidates. (a) Error vs. the number of candidates for averaging. (b) Cumulative errors on 323 test cases using 100 candidates for averaging.



(a) (b)

Figure 9. Cumulative errors of point-to-mesh distance, E_{p2m}, for (a) LV endocardial surface and (b) LV epicardial surface.

streak artifacts as shown in the third example. Please refer to the supplementary materials for more examples.

Our approach is fast with an average speed of 7.9 seconds for automatic segmentation of all four chambers (on a computer with a 3.2 GHz CPU and 3 GB memory). The computation time is roughly equally split on the MSL based similarity transformation estimation and the non-rigid deformation estimation. Our approach is sensibly faster comparing to other reported results, *e.g.*, 3 seconds for LV using a semi-automatic approach in [9], 15 seconds for non-rigid deformation in [24], 50 seconds for heart localization in [19], and 2-3 minutes for a 3D AAM based approach in [13].

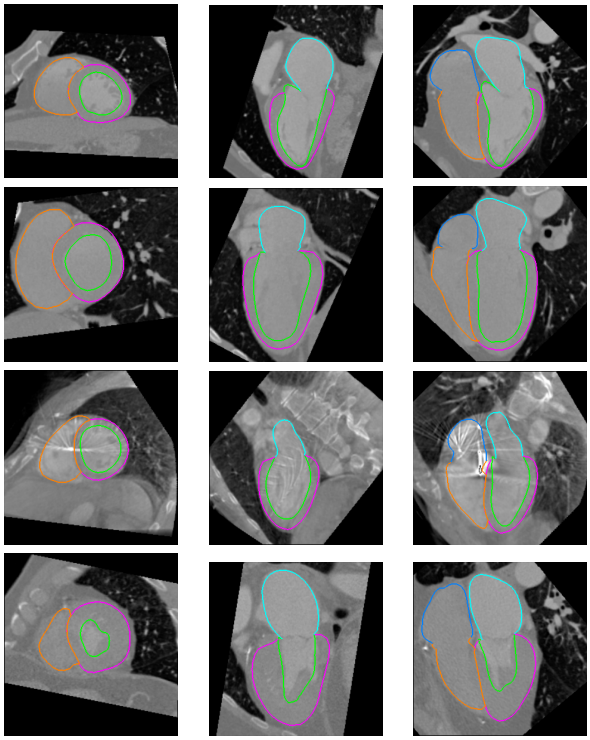


Figure 10. Examples of heart chamber segmentation in 3D CT volumes with green for LV endocardial surface, magenta for LV epicardial surface, cyan for LA, brown for RV, and blue for RA. Each row represents three orthogonal views of a volume.

7. Conclusions and Future Work

In this paper, we proposed an efficient and robust approach for automatic heart chamber segmentation in 3D CT volumes. The efficiency of our approach comes from the two new techniques named marginal space learning and steerable features. Robustness is achieved by using recent advances in learning discriminative object models and exploiting large volumetric images databases. All major steps in our approach are learning-based therefore minimizing the number of underlying model assumptions. According to our knowledge, this is the first study reporting stable results on a large cardiac CT data set. Our approach is general and we have extensively tested it on many challenging 3D detection and segmentation tasks in medical imaging (e.g., ileocecal valves, polyps, and livers in abdominal CT, brain tissues and heart chambers in ultrasound images, and heart chambers in MRI). In our current system, each heart chamber is detected independently. This is by no means optimal. In the future, we will exploit the geometric constraints among different chambers to improve the system on both speed and accuracy.

References

[1] A. Andreopoulos and J. K. Tsotsos. A novel algorithm for fitting 3-D active appearance models: Application to cardiac MRI segmentation. In *Proc. Scandinavian Conf. Image Analysis*, pages 729–739, 2005.

[2] Z. Bao, L. Zhukov, I. Guskov, J. Wood, and D. Breen. Dynamic deformable models for 3D MRI heart segmentation. In *SPIE Medical Imaging*, pages 398–405, 2002.

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

[4] C. Corsi, G. Saracino, A. Sarti, and C. Lamberti. Left ventricular volume estimation for real-time three-dimensional echocardiography. *IEEE Trans. Medical Imaging*, 21(9):1202–1208, 2002.

[5] O. Ecabert, J. Peters, and J. Weese. Modeling shape variability for full heart segmentation in cardiac computed-tomography images. In *SPIE Medical Imaging*, pages 1199–1210, 2006.

[6] A. F. Frangi, W. J. Niessen, and M. A. Viergever. Three-dimensional modeling for functional analysis of cardiac images: A review. *IEEE Trans. Medical Imaging*, 20(1):2–25, 2001.

[7] A. F. Frangi, D. Rueckert, and J. S. Duncan. Three-dimensional cardiovascular image analysis. *IEEE Trans. Medical Imaging*, 21(9):1005–1010, 2002.

[8] O. Gerard, A. C. Billon, J.-M. Rouet, M. Jacob, M. Fradkin, and C. Allouche. Efficient model-based quantification of left ventricular function in 3-D echocardiography. *IEEE Trans. Medical Imaging*, 21(9):1059–1068, 2002.

[9] W. Hong, B. Georgescu, X. S. Zhou, S. Krishnan, Y. Ma, and D. Comaniciu. Database-guided simultaneous multi-slice 3D segmentation for volumetric data. In *ECCV*, pages 397–409, 2006.

[10] M.-P. Jolly. Automatic segmentation of the left ventricle in cardiac MR and CT images. *IJCV*, 70(2):151–163, 2006.

[11] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, R. Mohiaddin, and D. Rueckert. Atlas-based segmentation and tracking of 3D cardiac MR images using non-rigid registration. In *MICCAI*, pages 642–650, 2002.

[12] T. McInerney and D. Terzopoulos. A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4D image analysis. *Computerized Medical Imaging and Graphics*, 19(1):69–83, 1995.

[13] S. C. Mitchell, J. G. Bosch, B. P. F. Lelieveldt, R. J. van Geest, J. H. C. Reiber, and M. Sonka. 3-D active appearance models: Segmentation of cardiac MR and ultrasound images. *IEEE Trans. Medical Imaging*, 21(9):1167–1178, 2002.

[14] P. D. Moral, A. Doucet, and A. Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[15] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *CVPR*, pages 193–199, 1997.

[16] R. Osadchy, M. Miller, and Y. LeCun. Synergistic face detection and pose estimation with energy-based model. In *NIPS*, pages 1017–1024, 2005.

[17] K. Park, A. Montillo, D. Metaxas, and L. Axel. Volumetric heart modeling and analysis. *Communications of the ACM*, 48(2):43–48, 2005.

[18] G. I. Sanchez-Ortiz, G. J. T. Wright, N. Clarke, J. Declerck, A. P. Banning, and J. A. Noble. Automated 3-D echocardiography analysis compared with manual delineations and SPECT MUGA. *IEEE Trans. Medical Imaging*, 21(9):1069–1076, 2002.

[19] H. Schramm, O. Ecabert, J. Peters, V. Philomin, and J. Weese. Towards fully automatic object detection and segmentation. In *SPIE Medical Imaging*, pages 11–20, 2006.

[20] Z. Tu. Probabilistic boosting-tree: Learning discriminative methods for classification, recognition, and clustering. In *ICCV*, pages 1589–1596, 2005.

[21] Z. Tu, X. S. Zhou, A. Barbu, L. Bogoni, and D. Comaniciu. Probabilistic 3D polyp detection in CT images: The role of sample alignment. In *CVPR*, pages 1544–1551, 2006.

[22] H. C. van Assen, M. G. Danilouchkine, A. F. Frangi, S. Ordas, J. J. M. Westernberg, J. H. C. Reiber, and B. P. F. Lelieveldt. SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. *Medical Image Analysis*, 10(2):286–303, 2006.

[23] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR*, pages 511–518, 2001.

[24] J. von Berg and C. Lorenz. Multi-surface cardiac modelling, segmentation, and tracking. In *Proc. Functional Imaging and Modeling of the Heart*, pages 1–11, 2005.

[25] I. Wolf, M. Hastenteufel, R. D. Simone, M. Vetter, G. Glombitza, S. Mottl-Link, C. F. Vahl, and H.-P. Meinzer. ROPES: A semiautomated segmentation method for accelerated analysis of three-dimensional echocardiographic data. *IEEE Trans. Medical Imaging*, 21(9):1091–1104, 2002.