Constrained Marginal Space Learning for Efficient 3D Anatomical Structure Detection in Medical Images

Yefeng Zheng[1], Bogdan Georgescu[1], Haibin Ling[2]*, S. Kevin Zhou[1],
Michael Scheuering[3], and Dorin Comaniciu[1]

[1]Integrated Data Systems Department, Siemens Corporate Research, Princeton, NJ, USA
[2]Computer and Information Sciences Department, Temple University, PA, USA
[3]Computed Tomography, Siemens Healthcare Sector, Forchheim, Germany

Abstract

Recently, we proposed marginal space learning (MSL) as a generic approach for automatic detection of 3D anatomical structures in many medical imaging modalities. To accurately localize a 3D object, we need to estimate nine parameters (three for position, three for orientation, and three for anisotropic scaling). Instead of uniformly searching the original nine-dimensional parameter space, only low-dimensional marginal spaces are uniformly searched in MSL, which significantly improves the speed. In many real applications, a strong correlation may exist among parameters in the same marginal spaces. For example, a large object may have large scaling values along all directions. In this paper, we propose constrained MSL to exploit this correlation for further speed-up. As another major contribution, we propose to use quaternions for 3D orientation representation and distance measurement to overcome the inherent drawbacks of Euler angles in the original MSL. The proposed method has been tested on three 3D anatomical structure detection problems in medical images, including liver detection in computed tomography (CT) volumes, and left ventricle detection in both CT and ultrasound volumes. Experiments on the largest datasets ever reported show that constrained MSL can improve the detection speed up to 14 times, while achieving comparable or better detection accuracy. It takes less than half a second to detect a 3D anatomical structure in a volume.

1. Introduction

Efficiently localizing an anatomical structure (e.g., heart, liver, and kidney) in medical images is often a prerequisite for the subsequent procedures, e.g., segmentation, measuring, and classification. Albeit important, automatic object detection is largely ignored in previous work. Most existing 3D segmentation methods focus on boundary delineation using active shape models (ASM) [2], active appearance models (AAM) [1], and deformable models by assuming that a rough pose estimate of the object is available. Sometimes, heuristic methods may be used for automatic object localization by exploiting the domain specific knowledge [3]. Recently, we proposed a generic 3D object detection framework, marginal space learning (MSL) [14, 15], based on learning discriminative classifiers. The full parameter space for 3D object localization has nine dimensions: three for position, three for orientation, and three for anisotropic scaling. To efficiently localize the object, we perform parameter estimation in a series of marginal spaces with increasing dimensionality. To be specific, the task is split into three steps: object position estimation, position-orientation estimation, and similarity transformation estimation (as shown in Fig. 1). After each step, a few promising candidates (e.g., 100) are retained for the next estimation step. Instead of uniformly searching the original 9D parameter space, only low-dimensional marginal spaces are uniformly and exhaustively searched in MSL. It has been shown that MSL can reduce the number of testing hypotheses by about six orders of magnitude [15], compared to a naive implementation of full space searching. Later, MSL has been successfully applied to many 3D anatomical structure detection problems in medical imaging, e.g., ileocecal valves [7], polyps [8], and livers [6] in abdominal CT, and heart chambers [13] in ultrasound.

Previously, we used three Euler angles to represent the 3D orientation space [15], which have several limitations. (1) There are multiple sets of values that can yield the same orientation, leading to a fundamental ambiguity. During training, we assign a hypothesis to the positive or negative

*Haibin Ling contributed to this work when he was with the Integrated Data Systems Department of Siemens Corporate Research.

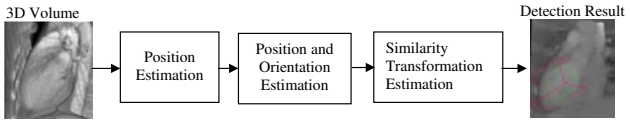


Figure 1. 3D object localization using marginal space learning [15].

set based on its distance to the ground truth. The Euclidean distance in the Euler angle space was used in [15] as the distance measurement between two orientations. However, it is well known that the Euclidean distance is not a good orientation distance measurement [4]. (2) Uniform sampling in the Euler angle space is not uniform in the orientation space [5]. (3) There are many widely used conventions for Euler angles. To reduce the search range, the best convention need to be selected depending on application scenario.

In this paper, we propose to use *quaternions* [4] to overcome all the above limitations. Quaternions provide an elegant conceptual framework, which can solve many problems involving rotation. For example, it is much easier to use quaternions to calculate the correct distance between two orientations.

In terms of computational efficiency, MSL significantly outperforms a brute-force full space search. However, it still has much room for improvement since the marginal spaces are exhaustively searched, though in a lower dimension (three in this case). The variations of the object orientation and its physical size are normally bounded. The distribution range of a parameter can be estimated from the training set. During searching, each parameter is uniformly sampled within that range to generate testing hypotheses. Each of the three subspaces (the translation, orientation, and scale spaces) are uniformly sampled without considering the correlation among parameters in the same marginal space. However, in many real applications, the parameters are unlikely to be independent. For example, a large object (the heart of an adult) is likely to have larger values than a small object (the heart of a baby) in all three directions. Independent sampling of each parameter will result in much more testing hypotheses than necessary. Because the detection speed is roughly proportional to the number of hypotheses, reducing the testing hypothesis set can speed up the system.

In this paper, we propose to further constrain the search by exploiting the correlation among object pose parameters. Due to the heterogeneity in the capture range of the scanning protocols, the position of an organ may vary significantly in a volume. We propose a generic way to reduce the search range for object position. To study an organ, normally we need to capture the whole organ in the volume. Therefore, the center of the organ cannot be arbitrarily close to the volume border. Using this observation, we can safely skip those hypotheses around volume margin for object position estimation. To constrain the search of the orientation or scale space, we can estimate the joint distribution of parameters using the training set. We then sample only the region with large probabilities. However, it is not trivial to estimate the joint probability distributions reliably since, usually, only a limited number of training samples (a couple of hundreds or even less) are available. To solve this problem, we propose an *example-based strategy* to constrain the search to a region with large probabilities. We first uniformly sample the space to get a large set, S_u. For each training sample, we add its neighboring hypotheses in S_u to the test set S_t. Repeating the above process for all training samples and removing redundant hypotheses from S_t, we can get a much smaller test set than S_u. Using the constrained marginal space learning, overall, we can improve the detection speed by an order of magnitude further. Besides speed-up, constraining the search to a small valid region can reduce the likelihood of detection outliers, therefore improve the detection accuracy.

In summary, we make two major improvements to the original MSL [15].

1. After analyzing the drawbacks of Euler angles for 3D orientation representation, we propose to use quaternions to overcome the limitations of Euler angles.
2. We propose efficient ways to further constrain the search spaces in MSL and improve the detection speed by an order of magnitude. It takes less than half a second to detect a 3D anatomical structure in a volume.

This paper is organized as follows. In the next section, we give a brief overview of marginal space learning. In Section 3, we compare Euler angles and quaternions for 3D orientation representation and discuss the limitations of Euler angles as used in [15]. Section 4 presents our approach to constrain the search space in MSL. Extensive comparison experiments on large medical datasets in Section 5 demonstrate the efficiency of the proposed method. This paper concludes with Section 6.

2. Overview of Marginal Space Learning

In this section, we briefly review the marginal space learning (MSL) based 3D object detection method. Interested readers are referred to [15] for more details. To localize a 3D object, we need to estimate nine parameters (three for position, three for orientation, and three for anisotropic scaling). As a straightforward application of a learning-based approach (e.g., [12]), we can train a discriminative classifier that assigns a high score to a hypothesis closing to the true object pose and a low score to those far away. During testing, we exhaustively search the 9D parameter space (i.e., all possible combinations of position, orientation, and scales) under a specified searching step to gener-

ate a huge number of hypotheses and test each one using the trained classifier. The hypothesis with the highest classification score can be taken as the final detection result. Due to the exponential increase of hypotheses w.r.t. the dimension of the search space, the computation demand of this naive implementation is well beyond the current personal computers.

In MSL, we split the task into three steps: object position estimation, position-orientation estimation, and similarity transformation estimation (as shown in Fig. 1). For each step, we train a classifier using the probabilistic boosting tree (PBT) [10] to assign a high score to a correct hypothesis. The 3D Haar wavelet features are used to train the position detector, while efficient steerable features are used for the other two detectors to avoid time-consuming volume rotation. During testing, all voxels are scanned using the trained position classifier and top K (e.g., $K = 100$) candidates, (X_i, Y_i, Z_i), $i = 1, \ldots, K$, are kept. Next, each candidate is augmented with M hypotheses about orientation by exhaustively searching the orientation space, $(X_i, Y_i, Z_i, \psi_j, \phi_j, \theta_j)$, $j = 1, \ldots, M$. The trained position-orientation classifier is used to prune these $K \times M$ hypotheses and the top N (e.g., $N = 50$) candidates are retained, $(\hat{X}_i, \hat{Y}_i, \hat{Z}_i, \hat{\psi}_i, \hat{\phi}_i, \hat{\theta}_i)$, $i = 1, \ldots, N$. Similarly, we augment each position-orientation candidate with P hypotheses about scaling by exhaustively searching the scale space. The position-orientation-scale classifier is used to pick the final best estimate. As shown in [15], MSL can reduce the number of testing hypotheses by six orders of magnitude compared the full space searching approach.

3. Orientation Representation for 3D

In this section, we first analyze the drawbacks of Euler angles in 3D orientation representation in the original MSL [15] and then propose to use quaternions to solve all the limitations of Euler angles.

3.1. Drawbacks of Euler Angles

It is well known that 3D orientation has three degrees of freedom and can be represented as three Euler angles. An advantage of Euler angles is that they have an intuitive physical meaning. For example an orientation with Euler angles of ψ, ϕ, and θ in the ZXZ convention is achieved by rotating the original coordinate system around the z axis with an amount ψ, followed by a rotation around the x axis with an amount ϕ, and lastly a rotation around the z axis again with an amount θ. The rotation operation is not commutable. That means we cannot change the order of rotations.

To train a classifier for distinguishing correct orientation estimates from wrong ones, we need to provide both positive and negative training samples. The Euclidean distance[1] in the Euler angle space can be used to measure the distance of a hypothesis $O^h = (\psi^h, \phi^h, \theta^h)$ to the ground truth $O^t = (\psi^t, \phi^t, \theta^t)$

$$D_e(O^h, O^t) = \sqrt{\|\psi^h - \psi^t\|^2 + \|\phi^h - \phi^t\|^2 + \|\theta^h - \theta^t\|^2}. \quad (1)$$

If the distance is less than a threshold, it is taken as a positive sample, otherwise negative. Though convenient, the Euclidean distance is not a good distance measurement of orientations. There are multiple sets of Euler angles which yield the same orientation, leading to a fundamental ambiguity. For example, Euler angles (α, 0, β) and (γ, 0, θ) in the ZXZ convention represent the same orientation when $\alpha + \beta = \gamma + \theta$. That means two close orientations may have a large distance in the Euler angle space. Therefore, the collected positive and negative sets may be confusing, which makes the learning difficult. To estimate the orientation of an object, we need to uniformly sample the orientation space to generate a set of hypotheses. Each hypothesis is then tested with the trained classifier to pick the best one. However, uniform sampling of the Euler angle space under the Euclidean distance is not truly uniform in the orientation space [5].

Another drawback of using Euler angles is that there are 12 possible conventions and a handful are widely used [5]. In our previous work[15], the range of Euler angles are computed from the training set. Each Euler angle is then uniformly sampled within that range to generate hypotheses. Different conventions may give quite different search ranges. For example, if the rotation is only around the y axis, the XYZ convention (where two Euler angles are zero) is more compact than the ZXZ conventions since the latter needs three rotations to generate a pure rotation around the y axis. Fig. 2 shows the statistics of the Euler angles with the ZXZ and XYZ conventions for the LV orientation in ultrasound volumes (see Section 5.3). In this case, the representation with the XYZ convention is more compact. However,

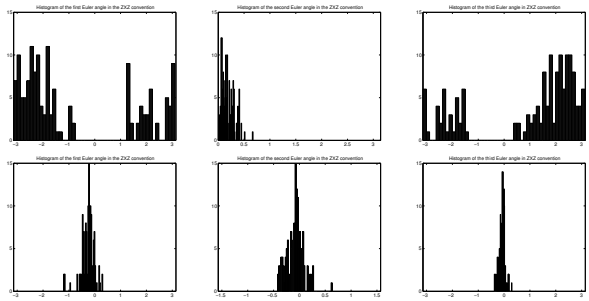


Figure 2. Histograms of Euler angles of the left ventricle orientation in ultrasound volumes (see Section 5.3). First row shows three Euler angles of the ZXZ convention and the second row shows the XYZ convention.

[1] In [15], a special city-block distance is actually used to split the orientation hypotheses into the positive and negative sets. The city-block distance is less accurate than the Euclidean distance and can be taken as an approximation of the latter.

for the application of heart chamber detection in cardiac CT volumes (Section 5.2), we find that the ZXZ convention is more efficient than the XYZ convention. In practice, for a new application, we need to try different conventions to select the most compact one.

In summary, previous use of Euler angles for orientation representation in [15] has the following drawbacks.

1. Due to the inherent ambiguity in Euler angle representation, the same orientation can be represented with multiple value sets.
2. The Euclidean distance in the Euler angle space is not a good distance measurement of orientations.
3. Naive uniform sampling of the Euler angle space is not uniform in the orientation space due to the use of a wrong distance measurement.
4. Among many conventions, we need to manually select the Euler angle convention that represents the search range most compactly.

3.2. Quaternions for Orientation Representation

In this paper, we propose to use quaternions to overcome all the drawbacks of the Euler angles as used in [15]. Introduced in mid-1800s by Hamilton, quaternions provide an elegant conceptual framework, which can solve many problems involving rotation [4]. A quaternion is represented by four numbers

$$\mathbf{q} = [w, x, y, z], \tag{2}$$

or as a scalar and a vector

$$\mathbf{q} = [s, \mathbf{v}]. \tag{3}$$

In the scalar-vector representation, multiplication of two quaternions becomes

$$\mathbf{q_1 q_2} = [s_1 s_2 - \mathbf{v_1}.\mathbf{v_2}, \quad s_1\mathbf{v_2} + s_2\mathbf{v_1} + \mathbf{v_1} \times \mathbf{v_2}], \tag{4}$$

where $\mathbf{v_1}.\mathbf{v_2}$ is the vector inner-product and $\mathbf{v_1} \times \mathbf{v_2}$ is the vector cross-product. The multiplication of two quaternions is also a quaternion.

To represent an orientation, we use unit quaternions,

$$|\mathbf{q}| = w^2 + x^2 + y^2 + z^2 = 1. \tag{5}$$

Therefore, a unit quaternion also has three degrees of freedom, the same as the Euler angles.

A unit quaternion can also be represented in the scalar-vector form as

$$\mathbf{q} = [\cos(\theta/2), \mathbf{v}\sin(\theta/2)], \tag{6}$$

where $\mathbf{v}$ is a three-dimensional unit vector. Given a quaternion $\mathbf{p}$, if we left-multiple it with $\mathbf{q} = [\cos(\theta/2), \mathbf{v}\sin(\theta/2)]$, we get a new quaternion $\mathbf{qp}$. The physical meaning of this operation is that $\mathbf{qp}$ represents the orientation after we rotate $\mathbf{p}$ around axis $\mathbf{v}$ with the amount of rotation θ [4]. The conjugate of a quaternion is defined as

$$\bar{\mathbf{q}} = [w, -x, -y, -z] = [\cos(-\theta/2), \mathbf{v}\sin(-\theta/2)]. \tag{7}$$

Here, $\bar{\mathbf{q}}$ represents a rotation around axis $\mathbf{v}$ with an amount $-\theta$.

Given two orientations, we can rotate one along an axis to align it with the other [9]. The amount of rotation provides a more natural definition of the distance between two orientations. Using quaternions, we can calculate the amount of rotation between two orientations easily. The rotation, $\mathbf{q} = \mathbf{q_1}\bar{\mathbf{q}_2}$, moves $\mathbf{q_2}$ to $\mathbf{q_1}$. Therefore, the amount of rotation between quaternions $\mathbf{q_1}$ and $\mathbf{q_2}$ using the scalar-vector representation in Eq. (3) is

$$D_q(q_1, q_2) = \arccos(|s_1 s_2 - \mathbf{v_1}.\mathbf{v_2}|). \tag{8}$$

4. Constrained Search Space for MSL

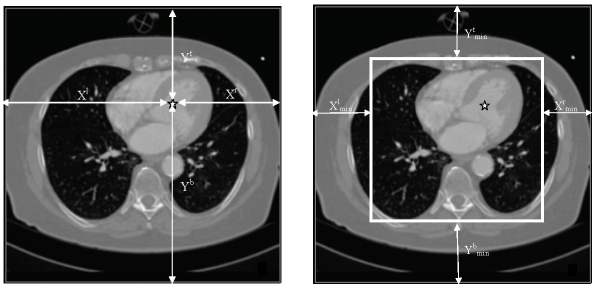
In this section, we present two methods to effectively constrain the search space in MSL.

4.1. Constrained Space for Object Position

Due to the heterogeneity in scanning protocol, the position of an object-of-interest may vary significantly in a volume. As shown in Fig. 5, the first volume focuses on the liver, while the second volume captures almost the full torso. A learning based object detection system [15] normally tests all voxels as hypotheses of the object center. Therefore, for a big volume, the number of hypotheses is quite large. It is preferable to constrain the search to a smaller region. The challenge is that the scheme should be generic and works for different application scenarios.

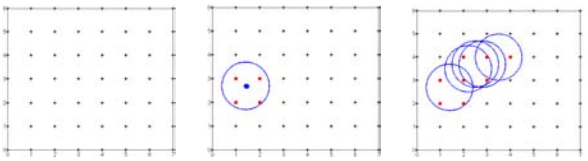
In this paper, we propose a generic way to constrain the search space. Our basic assumption is that, to study an organ, normally we need to capture the whole organ in the volume. Therefore, the center of the organ cannot be arbitrarily close to the volume border. As shown in Fig. 3a, for each training volume, we can measure the distance of the object center (e.g., the left ventricle in this case) to the volume border in six directions (e.g., X^l for the distance to the left volume border, X^r for right, Y^t for top, Y^b for bottom, Z^f for front, and Z^b for back). All the distances should be measured in physical units (e.g., millimeters) to handle different resolution settings in different volumes. The minimum value (e.g., X^l_{min} for the left margin) for each direction can be easily calculated from a training set. These minimum margins define a region (as shown in the white box in Fig. 3b) and we only need to test voxels inside the region as possible position hypotheses.

Using the proposed method, we reduce the number of testing hypotheses by 75% to 98% in our empirical studies. On average, we achieve a reduction of 91% for liver



(a) (b)

Figure 3. Constraining the search for object center in a volume, illustrated for the left ventricle (LV) detection in a CT volume. (a) Distances of the object center to the volume borders. (b) Constrained search space (the region enclosed by the white box).



(a) (b) (c)

Figure 4. Example-based selection of testing hypotheses. (a) Uniformly sampled hypotheses, shown as black '+'s. (b) After processing the first training sample. The blue dot shows the ground truth and the circle shows the neighborhood range. All hypotheses inside the circle (represented as red dots) are added to the testing hypothesis set. (c) The testing hypothesis set after processing five training samples.

detection (Section 5.1), 89% for LV detection in CT (Section 5.2), and 84% for LV detection in ultrasound (Section 5.3). That means we can speed up the position estimation step about 10 times. Our strategy is generic. If there is any application-specific prior knowledge available, it can be combined with our strategy to further constrain the position search space.

4.2. Constrained Spaces for Orientation

In this section, we present our example-based strategy to effectively constrain the orientation search spaces. We use LV detection in CT volumes (Section 5.2) to illustrate the efficiency of the proposed method. Similar analysis can also be performed for the other two applications presented in Section 5.

For many problems, the orientation of an object is well constrained in a small region. It is not necessary to test the whole orientation space. For example, on the cardiac CT dataset the ranges of Euler angles for the LV using the ZXZ convention are [-0.9, 39.1], [-60.0, 88.7], and [-68.8, -21.9] degrees, respectively. In [15], Euler angles are sampled independently within the region to generate testing hypotheses. However, since three Euler angles should be combined to define an orientation, they are not independent. Sampling each Euler angle independently will generate far more hypotheses than necessary.

To constrain the search space, we can estimate the joint distribution of parameters using the training set. We then sample only the region with large probabilities. However, it is not trivial to estimate the joint probability distributions reliably since, usually, only a limited number of training samples (a couple of hundreds or even less) are available. In this paper, we propose to use an example-based strategy to generate testing hypotheses (as shown in Fig. 4). The procedure is as follows.

1. Uniformly sample the parameter space with a certain resolution r to generate S_u (as shown in Fig. 4a).
2. Set the selected hypothesis set S_t to empty.
3. For each training sample, we add its neighboring samples in S_u (which have a distance no more than d) into S_t (as shown by red dots in Fig. 4b). Here, $d \geq r/2$, otherwise, there may be no sample satisfying the condition. In our experiments, we set $d = r$.
4. Remove redundant elements in S_t to get the final testing hypothesis set.

To generate the constrained testing hypothesis set for orientation, we first need to uniformly sample the whole orientation space to generate the set S_u. The problem is formulated as, given N sample orientations, we want to distribute them as uniformly as possible. We can define "a covering radius, α, as the maximum rotation needed to align an arbitrary orientation with one of the sample orientations" [4]. For uniform sampling, we want to find an optimal configuration of N sample orientations that gives the smallest α. The optimization procedure is a little bit involved, refer to [4] for more details. Near-optimal configurations for some N's are available from a companion website of [4] (http://charles.karney.info/orientation/). We start from a uniform set of 7416 samples distributed in the whole orientation space with $\alpha = 9.72$ degrees as S_u. On a dataset of 457 cardiac CT volumes, S_t of the LV orientation has only 66 unique orientations, which is much smaller than S_u (7416) and also smaller than the number of the training volumes (457).

In [15], each Euler angle was uniformly sampled at a step size β to generate hypotheses. The maximum distance for an arbitrary orientation to the closest hypothesis is $\frac{\sqrt{3}}{2}\beta$ using the Euclidean distance measurement. Since different distance measurements are used in our case (the quaternion distance) and the previous work [15] (the Euclidean distance), we cannot compare them directly. The Euclidean distance measurement tends to over-estimate the true distance. However, in the worst case, it is as large as the quaternion distance. To achieve a nominally equivalent sampling resolution, the search step size β should be $\beta = \frac{2}{\sqrt{3}}\alpha$, that is 11.2 degrees. Suppose the range of a

parameter is within $[V_{min}, V_{max}]$. We sample N points, $P_{min}, P_{min}+r, \ldots, P_{max} = P_{min}+(N-1)r$, under resolution r. To fully cover the whole range, we must have $P_{min} \leq V_{min}$ and $P_{max} \geq V_{max}$. Therefore, the number of samples should be

$$N = \left\lceil \frac{V_{max} - V_{min}}{r} \right\rceil + 1, \tag{9}$$

where $\lceil x \rceil$ returns the smallest integer that is no less than x. On the cardiac CT dataset, sampling Euler angles under the resolution of 11.2 degrees, we need $5 \times 15 \times 6 = 450$ samples to cover the orientation space. Using the quaternion representation and exploiting the correlation among orientation parameters, we reduce the number of hypotheses by 85% to 66 samples.

4.3. Constrained Spaces for Scale

The same technique can also be applied for the scale space. The LV has a roughly rotation symmetric shape. Therefore, two scales (we denote them as S_x and S_y) perpendicular to the LV long axis (represented as the z axis) are highly correlated since they are roughly the same. The range of the scales calculated from the 457 training volumes are [53.0, 91.1] mm for S_x, [49.9, 94.0] mm for S_y, and [72.3, 128.4] mm for S_z. If we uniformly sample each scale independently using a resolution of 6 mm [15], we need $8 \times 9 \times 11 = 792$ samples. Using our example-based strategy, we only need 240 samples to cover the whole training set.

5. Experiments

In this section, we present three experiments to demonstrate the improved efficiency of constrained MSL, compared to the original MSL [15].

5.1. Liver Detection in CT Volumes

Our database contains 226 3D CT volumes. The dataset is very challenging because the volumes come from largely diverse sources. Due to the difference in scanning protocol, the volumes have various dimensionality: the inter-slice resolution varies from 1.0 mm to 8.0 mm; the number of slices varies from 40 to 524; and the actual volume height varies from 122 mm to 766 mm.

After object localization, we align the mean shape (a surface mesh) with the estimated transformation. The accuracy of initial shape estimate is measured with the symmetric 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. After initialization, we can deform the mesh to fit the image boundary to further reduce the error. In this paper, we focus on object localization. Therefore, in the following we only measure the error of the initialized shapes for comparison.

Table 1. Comparison of unconstrained and constrained MSL on the number of testing hypotheses and computation time.

(a) Liver detection in CT volumes.

	Unconstrained MSL [15]		Constrained MSL	
	#Hypotheses	Speed	#Hypotheses	Speed
Position	∼403,000	2088.7 ms	∼38,000	167.1 ms
Orientation	2686	2090.0 ms	42	59.5 ms
Scale	1664	1082.8 ms	303	243.7 ms
Overall		6590.8 ms		470.3 ms

(b) Left ventricle detection in CT volumes.

	Unconstrained MSL [15]		Constrained MSL	
	#Hypotheses	Speed	#Hypotheses	Speed
Position	∼158,000	784.3 ms	∼18,000	75.9 ms
Orientation	450	351.5 ms	66	52.6 ms
Scale	792	193.3 ms	240	60.4 ms
Overall		1329.3 ms		188.9 ms

(c) Left ventricle detection in ultrasound volumes.

	Unconstrained MSL [15]		Constrained MSL	
	#Hypotheses	Speed	#Hypotheses	Speed
Position	∼233,000	1487.3 ms	∼37,000	163.6 ms
Orientation	882	696.3 ms	99	12.7 ms
Scale	2340	769.7 ms	296	219.3 ms
Overall		2953.3 ms		395.6 ms

Table 2. Comparison of unconstrained and constrained MSL on detection accuracy. Average point-to-mesh error E_{p2m} (in millimeters) of the initialized shape is used for evaluation.

(a) Liver detection in CT volumes.

	Mean	Standard Deviation	Median	Mean of Worst 10%
Unconstrained MSL [15]	7.44	2.26	6.99	12.32
Constrained MSL	7.12	2.15	6.73	11.88

(b) Left ventricle detection in CT volumes.

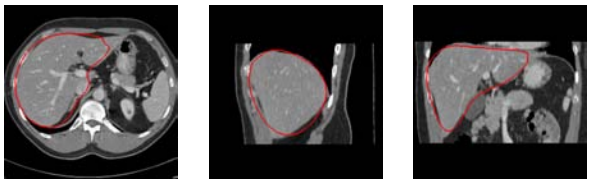
	Mean	Standard Deviation	Median	Mean of Worst 10%
Unconstrained MSL [15]	2.66	1.00	2.45	4.73
Constrained MSL	2.62	0.84	2.45	4.43

(c) Left ventricle detection in ultrasound volumes.

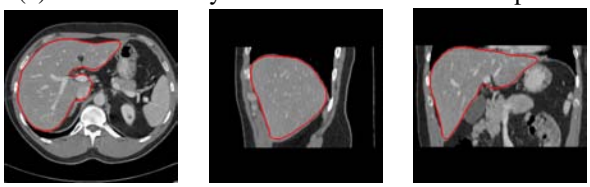
	Mean	Standard Deviation	Median	Mean of Worst 10%
Unconstrained MSL [15]	3.28	2.50	2.76	7.89
Constrained MSL	3.25	2.09	2.74	7.46

The detection speed of MSL is roughly proportional to the number of testing hypotheses. The analysis presented in Section 4 shows that constrained MSL significantly reduces the number of testing hypotheses. Table 1a shows the break-down computation time for three steps in MSL (see Fig. 1). Overall, constrained MSL uses only 470.3 ms to process one volume, while unconstrained MSL uses 6590.8 ms. Using constrained MSL, we achieve a speed-up by a factor of 14.

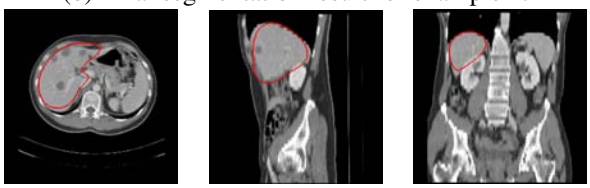
Constrained MSL also slightly improves detection accuracy. As shown in Table 2a, constrained MSL reduces the mean error E_{p2m} from 7.44 mm to 7.12 mm, and the me-



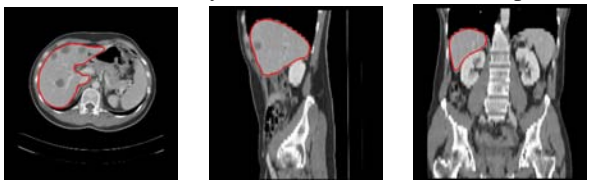
(a) Initialization by constrained MSL for example 1.



(b) Final segmentation result for example 1.

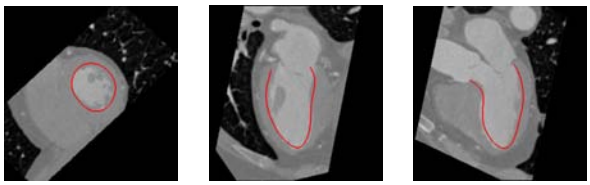


(c) Initialization by constrained MSL for example 2.

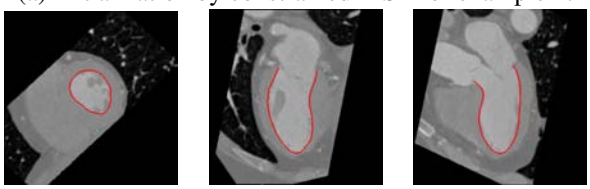


(d) Final segmentation result for example 2.

Figure 5. Typical liver segmentation results. From left to right: transversal, sagittal, and coronal views.



(a) Initialization by constrained MSL for example 1.



(b) Final segmentation result for example 1.

(c) Initialization by constrained MSL for example 2.

(d) Final segmentation result for example 2.

Figure 6. Typical left ventricle segmentation results in cardiac CT volumes. Three standard cardiac views are shown in each row.

dian error E_{p2m} from 6.99 mm to 6.73 mm, in a three-fold cross-validation. The accuracy improvement arises from two parts. First, as we constrain the search to a smaller but more meaningful region, the likelihood of detection outliers is reduced. Second, quaternions are used for orientation distance measurement, which reduces the confusion caused by the wrong measurement used in [15].

Fig. 5 shows typical liver segmentation results on two volumes. Accurate boundary delineation is achieved starting from the good initial estimate of the shape generated by constrained MSL. After applying our learning-based non-rigid deformation estimation method [6], the final E_{p2m} error is 1.45 mm, which is comparable or better than the state-of-the-art [11].

5.2. Left Ventricle Detection in CT Volumes

We collected 457 expert-annotated cardiac CT volumes from 186 patients with various cardiovascular diseases. 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 the distance between neighboring slices varies from 0.4 mm to 2.0 mm.

Similar to the previous experiments, constrained MSL also significantly reduces the number of testing hypotheses and speeds up the detection consequently. Table 1b shows the break-down computation time for three steps in MSL. Overall, constrained MSL uses only 188.9 ms to process one volume, while unconstrained MSL uses 1329.3 ms. Using constrained MSL, we achieve a speed-up by a factor of seven.

Table 2b shows the quantitative evaluation based on a four-fold cross validation for both unconstrained [15] and constrained MSL. The average errors by constrained MSL is slightly better than unconstrained MSL (2.62 mm vs. 2.66 mm). We also study the worst 10% cases (46 cases). The mean error for the worst 10% cases is 4.43 mm for constrained MSL, about 6.3% less than 4.73 mm for unconstrained MSL. It is clear that constraining the search space, we can reduce the likelihood of detection outliers.

Fig. 6 shows typical segmentation results for the LV on two volumes. Constrained MSL can provide a quite good initial estimate of the shape. After non-rigid deformation estimation [15], we can achieve accurate boundary delin-

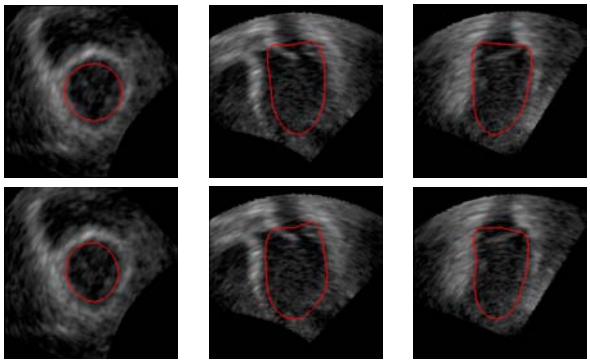


Figure 7. Typical left ventricle segmentation result in an ultrasound volume. Three standard cardiac views are shown in each row. The first row shows the initialization after object pose detection and the second row shows the final segmentation results.

eation results. Mean E_{p2m} error of 0.84 mm has been achieved, which compares favorably with the state-of-the-art [15, 3].

5.3. Left Ventricle Detection in Ultrasound Volumes

We collected 505 expert-annotated 3D ultrasound volumes. A typical volume size is about $160 \times 144 \times 208$ voxels and the resolution ranges are [1.24, 1.42], [1.34, 1.42], and [0.85, 0.90] mm along different directions. A four-fold cross validation is performed to evaluate our algorithm. Similar to the previous experiments, using constrained MSL we can speed up the detection by a factor of seven (Table 1c), while achieving a comparable detection accuracy (Table 2c). Fig. 7 shows the segmentation result for one example volume.

6. Conclusion

In this paper, we proposed two major improvements for the original marginal space learning (MSL) [15]. First, we proposed to use quaternions for 3D orientation representation to overcome the limitations of Euler angles. Second, a novel constrained MSL technique was introduced to reduce the search space. Based on the statistics of the distance from the object center to the volume border, we proposed a generic method to effectively constrain the object position space. Instead of quantizing each orientation and scale parameter independently, an example-based strategy is used to constrain the search to a small region with a high distribution probability. Extensive comparison experiments on three 3D anatomical structure detection tasks demonstrated the efficiency of the proposed method. It significantly accelerates the detection speed by an order of magnitude, resulting in a system that can process one volume in less than half a second. At the same time, constrained MSL can also improve the detection accuracy by reducing the likelihood for detection outliers.

References

[1] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(6):681–685, 2001.

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

[3] O. Ecabert, J. Peters, and H. Schramm et al. Automatic model-based segmentation of the heart in CT images. *IEEE Trans. Medical Imaging*, 27(9):1189–1201, 2008.

[4] C. F. F. Karney. Quaternions in molecular modeling. *Journal of Molecular Graphics and Modeling*, 25(5):595–604, 2007.

[5] J. J. Kuffner. Effective sampling and distance metrics for 3D rigid body path planning. In *Proc. IEEE Int'l Conf. Robotics and Automation*, pages 3993–3998, 2004.

[6] H. Ling, S. K. Zhou, Y. Zheng, B. Georgescu, M. Suehling, and D. Comaniciu. Hierarchical, learning-based automatic liver segmentation. In *CVPR*, 2008.

[7] L. Lu, A. Barbu, M. Wolf, J. Liang, L. Bogoni, M. Salganicoff, and D. Comaniciu. Simultaneous detection and registration for ileo-cecal valve detection in 3D CT colonography. In *ECCV*, 2008.

[8] L. Lu, A. Barbu, M. Wolf, J. Liang, M. Salganicoff, and D. Comaniciu. Accurate polyp segmentation for 3D CT colongraphy using multi-staged probabilistic binary learning and compositional model. In *CVPR*, 2008.

[9] K. Shoemake. Animating rotation with quaternion curves. In *Proc. SIGGRAPH*, pages 245–254, 1985.

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

[11] B. van Ginneken, T. Heimann, and M. Styner. 3D segmentation in the clinic: A grand challenge. In *MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge*, 2007.

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

[13] L. Yang, B. Georgescu, Y. Zheng, P. Meer, and D. Comaniciu. 3D ultrasound tracking of the left ventricles using one-step forward prediction and data fusion of collaborative trackers. In *CVPR*, 2008.

[14] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features. In *ICCV*, 2007.

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