Marginal Space Learning for Efficient Detection of 2D/3D Anatomical Structures in Medical Images

Yefeng Zheng, Bogdan Georgescu, and Dorin Comaniciu

Integrated Data Systems Department, Siemens Corporate Research, USA
{yefeng.zheng, bogdan.georgescu, dorin.comaniciu}@siemens.com

Abstract. Recently, marginal space learning (MSL) was proposed as a generic approach for automatic detection of 3D anatomical structures in many medical imaging modalities [1]. To accurately localize a 3D object, we need to estimate nine pose parameters (three for position, three for orientation, and three for anisotropic scaling). Instead of exhaustively searching the original nine-dimensional pose parameter space, only low-dimensional marginal spaces are searched in MSL to improve the detection speed. In this paper, we apply MSL to 2D object detection and perform a thorough comparison between MSL and the alternative full space learning (FSL) approach. Experiments on left ventricle detection in 2D MRI images show MSL outperforms FSL in both speed and accuracy. In addition, we propose two novel techniques, constrained MSL and nonrigid MSL, to further improve the efficiency and accuracy. In many real applications, a strong correlation may exist among pose parameters in the same marginal spaces. For example, a large object may have large scaling values along all directions. Constrained MSL exploits this correlation for further speed-up. The original MSL only estimates the rigid transformation of an object in the image, therefore cannot accurately localize a nonrigid object under a large deformation. The proposed nonrigid MSL directly estimates the nonrigid deformation parameters to improve the localization accuracy. The comparison experiments on liver detection in 226 abdominal CT volumes demonstrate the effectiveness of the proposed methods. Our system takes less than a second to accurately detect the liver in a volume.

1 Introduction

Efficiently detecting 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) [3], and deformable models [4] 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 [5]. As a more generic approach, the discriminative learning based method has been proved to be efficient and robust for many 2D object detection problems [6]. In this method, object detection is formulated as a classification problem: whether an image block contains the target object or not. To build a robust system, a classifier only tolerates limited variation in object pose. The object

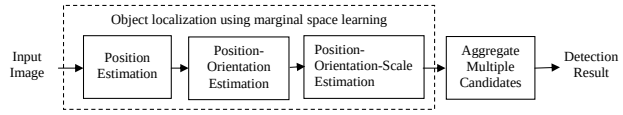


Fig. 1. Object localization using marginal space learning [1].

is found by scanning the classifier exhaustively over all possible combinations of position, orientation, and scale. Since both learning and searching are performed in the full pose parameter space, we call this method full space learning (FSL). This search strategy is different from other parameter estimation approaches, such as deformable models [4], where an initial estimate is adjusted (e.g., using the gradient descent technique) to optimize a predefined objective function. Exhaustive search makes the learning based system [6] robust under local minima. However, the number of testing hypotheses increases exponentially with respect to the dimension of the parameter space. We cannot directly apply FSL to 3D object detection since the pose parameter space for a 3D object has nine dimensions: three for position, three for orientation, and three for anisotropic scaling. Recently, we proposed an efficient learning-based technique, marginal space learning (MSL), for 3D object detection [1]. MSL performs 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 position-orientation-scale estimation (as shown in Fig. 1). Instead of exhaustively searching the original nine-dimensional parameter space, only low-dimensional marginal spaces are searched in MSL. Mathematical analysis shows that MSL can reduce the number of testing hypotheses by about six orders of magnitude [1], compared to a naive implementation of FSL.

MSL was originally proposed for efficient 3D object detection. Due to the huge number of hypotheses, FSL does not work for a 3D object detection problem. Therefore, there is no direct comparison experiment between MSL and FSL. The comparison on computation time of MSL and FSL in [1] was based on mathematical analysis. In this paper, we apply MSL to 2D object detection to estimate five object pose parameters (two for translation, one for orientation, and two for anisotropic scaling). In this low-dimensional space, it is possible to apply FSL with a coarse-to-fine strategy to achieve a reasonable detection speed. FSL is currently the state-of-the-art for 2D object detection [6]. There is a wide interest for a direct comparison between FSL and MSL. As a contribution of this paper, we perform a thorough comparison experiment on left ventricle (LV) detection in 2D magnetic resonance images (MRI). The experiment shows that MSL significantly outperforms FSL on both speed and accuracy.

In terms of computational efficiency, MSL outperforms a brute-force full space search by a significant margin on 3D object detection. However, it still has much room for improvement since the pose subspaces are exhaustively searched, though in a lower dimension. The variations of the object orientation and its physical size are normally bounded. The distribution range of a pose parameter can be estimated from the training set. During searching, each parameter is independently sampled within its own range to generate testing hypotheses [1]. Each of the three subspaces (the translation, orien-

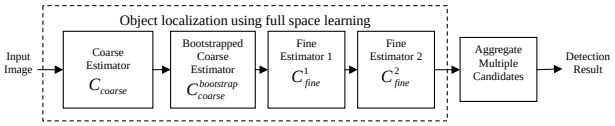


Fig. 2. Object localization using full space learning with a coarse-to-fine strategy.

tation, and scale spaces) is exhaustively sampled without considering the correlation among parameters in the same space. However, in many real applications, the pose parameters are unlikely to be independent. For example, a large object (e.g., the liver of an adult) is likely to have larger scales than a small object (e.g., the liver of a baby) in all three directions. Independent sampling of each parameter results in much more testing hypotheses than necessary. Because the detection speed is roughly proportional to the number of testing hypotheses, reducing the hypotheses can speed up the system. In this paper, we propose to further constrain the search space using an example-based strategy to exploit the correlation among object pose parameters. Using constrained marginal space learning, we can further improve the detection speed by an order of magnitude. Besides speed-up, constraining the search to a small valid region can reduce the likelihood of detection outliers, therefore improve the detection accuracy.

The original MSL was proposed to estimate the rigid transformation (translation, rotation, and scaling) of an object [1]. To better localize a nonrigid object, we may need to further estimate its nonrigid deformation. In this paper, we apply the marginal space learning principle to directly estimate the nonrigid deformation parameters. Within the steerable feature framework [1], we propose a new sampling pattern to efficiently incorporate nonrigid deformation parameters into the image feature set. These image features are used to train a classifier to distinguish a correct estimate of the deformation parameters from the wrong estimates.

In summary, we make three major contributions in this paper.

1. We perform a thorough comparison experiment between MSL and FSL on 2D object detection. The experiment shows that MSL outperforms FSL on both speed and accuracy.
2. We propose constrained MSL to further reduce the search spaces in MSL, therefore improve the detection speed by an order of magnitude.
3. We propose nonrigid MSL to directly estimate nonrigid deformation parameters of an object to improve the localization accuracy.

2 Full Space Learning vs. Marginal Space Learning

2.1 Full Space Learning

Object detection is equivalent to estimating the object pose parameters in an image. As a straightforward application of a learning-based approach [6], we can train a discriminative classifier that assigns a high score to a hypothesis that is close to the true object pose and a low score to those far away. During testing, we need to exhaustively search the

Table 1. Parameters for full space learning. The "# Hyph" columns show the number of hypotheses for each parameter. The "Step" columns show the search step size for each parameter. The "# Total Hyph" column lists the total number of hypotheses tested by each classifier. The "# Preserve" column lists the number of candidates preserved after each step.

	X		Y		θ		S_x		a		# Total Hyph	# Preserve
	# Hyph	Step	# Hyph	Step	# Hyph	Step	# Hyph	Step	# Hyph	Step		
C_{coarse}	36	8	23	8	18	20^o	15	16	6	0.2	1,341,360	10,000
$C_{coarse}^{bootstrap}$	1	8	1	8	1	20^o	1	16	1	0.2	$10,000 \times 1$	200
C_{fine}^{1}	3	4	3	4	3	10^o	3	8	3	0.1	200×243	100
C_{fine}^{2}	3	2	3	2	3	5^o	3	4	3	0.05	100×243	100

full parameter space (i.e., all possible combinations of position, orientation, and scale) to generate testing hypotheses. Suppose there are P pose parameters and parameter i is discretized to H_i values, the total number of testing hypotheses is $H_1 \times H_2 \ldots \times H_P$. Each hypothesis is then tested by the trained classifier to get a score and the hypothesis with the highest score is the best estimate of the pose parameters. (Please refer to Fig. 9 of [1] for an illustration of the basic idea). Since both learning and searching are performed in the full parameter space, we call it full space learning (FSL).

Due to the exponential increase of hypotheses w.r.t. the dimension of the parameter space, the computation demand of this naive implementation of FSL for 3D object detection is well beyond the current personal computers. For 2D object detection, we only need to estimate five object pose parameters, (X, Y, θ, S_x, S_y), with (X, Y) for the object position, θ for orientation, and (S_x, S_y) for anisotropic scaling. (Alternatively, we can use the aspect ratio $a = S_y/S_x$ to replace S_y as the last parameter.) A coarse-to-fine strategy can be exploited to accelerate the detection speed. The system diagram for left ventricle (LV) detection in 2D MRI images is shown in Fig. 2. In this particular implementation, in total, we train four classifiers.

At the coarse level, we use large search steps to reduce the total number of testing hypotheses. For example, the search step for position is set to eight pixels and the orientation search step is set to 20 degrees to generate 18 hypotheses for the whole orientation range. Even with these coarse search steps, the total number of hypotheses can easily exceed one million. As shown by the row labeled "C_{coarse}" in Table 1, in total, we search $36 \times 23 \times 18 \times 15 \times 6 = 1,341,360$ hypotheses at the coarse level. A classifier is trained to distinguish the hypotheses close to the ground truth from those far away. Interested readers are referred to [1, 6] for more details about the training of a classifier. Each hypothesis is then tested using the trained C_{coarse} classifier and the top 10,000 candidates are preserved. A bootstrapped classifier $C_{coarse}^{bootstrap}$ can be further exploited to reduce the number of candidates to 200.

As shown in Fig. 2, we use two iterations of fine level search to improve the estimation accuracy. In each iteration, the search step for each parameter is reduced by half. Around a candidate, we search three hypotheses for each parameter. In total, we search $3^5 = 243$ hypotheses around each candidate. Therefore, for the first fine classifier C_{fine}^{1}, in total we need to test $200 \times 243 = 48,600$ hypotheses. We preserve the top 100 candidates after the first fine-search step. After that, we reduce the search step by half again to start the second round refinement. The number of hypotheses and search step sizes for each classifier are listed in Table 1. In total, we test $1,341,360 + 10,000 + 46,800 + 23,400 = 1,424,260$ hypotheses.

2.2 Marginal Space Learning

Instead of exhaustively searching the full parameter space directly, MSL splits the task into three steps: object position estimation, position-orientation estimation, and position-orientation-scale estimation (as shown in Fig. 1). For each step, we train a classifier to assign a high score to a correct hypothesis. After each step, only a limited number of hypotheses are obtained for the following processing. Please refer to [1] for more details about MSL.

Following is an analysis on the number of testing hypotheses for MSL on LV detection in 2D MRI images. First all pixels are tested using the trained position classifier and the top 1000 candidates, (X_i, Y_i), $i = 1, \ldots, 1000$, are kept. Next, the whole orientation space is discretized under the resolution of five degrees, resulting in 72 orientation hypotheses. Each position candidate is augmented with all orientation hypotheses, (X_i, Y_i, θ_j), $j = 1, \ldots, 72$. The trained position-orientation classifier is used to prune these $1000 \times 72 = 72,000$ hypotheses and the top 100 candidates are retained, $(\hat{X}_i, \hat{Y}_i, \hat{\theta}_i)$, $i = 1, \ldots, 100$. Similarly, we augment each position-orientation candidate with a set of hypotheses about scaling. For LV detection, we have 182 scale combinations, resulting in a total of $100 \times 182 = 18,200$ hypotheses. The position-orientation-scale classifier is then used to pick the best hypothesis. For a typical image of 300×200 pixels, in total, we test $300 \times 200 + 1000 \times 72 + 100 \times 182 = 150,200$ hypotheses. For comparison, a total of 1,424,260 hypotheses need to be tested in FSL. Since the speed of the system is roughly proportional to the number of hypotheses, MSL is about an order of magnitude faster than FSL.

3 Constrained Marginal Space Learning

In this section, we present our approach to effectively constraining the search of MSL in all three subspaces (i.e., translation, orientation, and scale spaces) for further speed-up in 3D object detection.

Due to the heterogeneity in scanning protocol, the position of an object may vary significantly in a volume. As shown in Fig. 6, the first volume focuses on the liver, while the second volume captures almost the full torso. A learning based object detection system [1, 6] 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 liver 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). 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 (the white box in Fig. 3b) and we only need to test voxels inside the region for the object center. Using the proposed

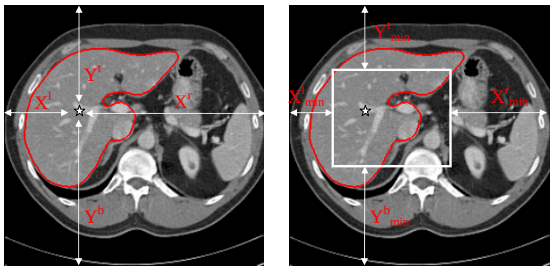


Fig. 3. Constraining the search for object center in a volume, illustrated for liver detection in a CT volume. (a) Distances of the object center to the volume borders. (b) Constrained search space (the region enclosed by the white box) based on the minimum distances to the volume borders.

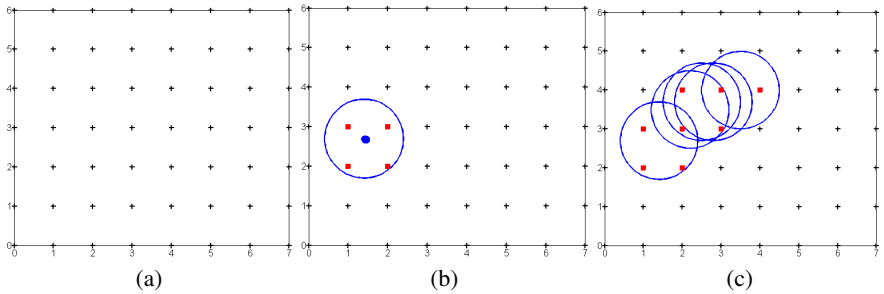


Fig. 4. Example-based selection of testing hypotheses. (a) Uniformly sampled hypotheses, shown as black '+'s. (b) After processing the first training sample. The blue dot shows the ground truth and the circle shows the neighborhood range. All hypotheses inside the circle (represented as red squares)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

method, on average, we get rid of 91% of voxels for liver detection (see Section 5.2), resulting in a speed-up in position estimation about 10 times.

In many applications, the orientation of an object is also constrained. Three Euler angles were used to represent the 3D orientation in [1]. The distribution range for an Euler angle can be calculated on a training set. Each Euler angle is then sampled independently within its own range to generate testing hypotheses. However, since the Euler angles are unlikely to be independent in a real application, sampling each Euler angle independently generates far more hypotheses than necessary. To constrain the search space, we should estimate the joint distribution of orientation parameters using the training set. We then generate hypotheses only in the region with a large probability. However, it is not trivial to estimate the joint probability distribution 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. The procedure is as follows (also illustrated in Fig. 4),

1. Uniformly sample the parameter space with a certain resolution r to generate S_u.

2. Set the selected hypothesis set S_t to empty.
3. For each training sample, its neighboring samples in S_u (with a distance no more than d to it) are added into S_t. Here, neighborhood size d should be large enough, 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 S_t.

Using a discretization resolution of 9.72 degrees, we get S_u of 7416 samples uniformly distributed in the whole orientation space [7]. On a dataset of 226 abdominal computed tomography (CT) volumes, S_t of the liver orientation has only 42 unique orientations, which is much smaller than S_u (7416) and also smaller than the number of the training volumes (226). For comparison, if we sample each Euler angle independently under the same resolution, we get 2,686 hypotheses for orientation.

A big object normally has large scaling values along all three directions. Therefore, the same technique can also be applied to constrain the scale space by exploiting the strong correlation among three scaling parameters. On the same liver dataset, if we uniformly sample each scale independently using a resolution of 6 mm, we get 1,664 hypotheses. Using our example-based strategy, we only need 303 hypotheses to cover the whole training set.

4 Nonrigid Marginal Space Learning

The original MSL [1] only estimates the rigid transformation for object localization. In many cases, we want to delineate the nonrigid boundary of an organ. For this purpose, the mean shape was aligned with the estimated object pose as an initial rough estimate of the shape in [1]. The active shape model (ASM) was then exploited to deform the initial shape to achieve the final boundary delineation. Since ASM only converges to a local optimum, the initial shape needs to be close to the true object boundary. Otherwise, the deformation is likely to get stuck in a wrong configuration. This problem is manifest in liver segmentation since the liver is the largest organ in human body and it is cluttered with several other organs (e.g., heart, kidney, stomach, and diaphragm). As a soft organ, it deforms significantly under the pressure from the neighboring organs. As noted in [8], for a highly deformable shape, the pose estimation can be improved by further initialization. In this paper, we propose *nonrigid MSL* to directly estimate the nonrigid deformation of an object for better shape initialization.

There are many ways to represent a nonrigid deformation. We use the statistical shape model [2] since it can capture the major deformation modes with a few parameters. To build a statistical shape model, we need N shapes and each is represented by M points with correspondence in anatomy. Stacking the 3D coordinates of these M points, we get a $3M$ dimensional vector $X_i, i = 1, 2, \ldots, N$, to represent a shape. To remove the relative translation, orientation, and scaling, we first jointly align all shapes using generalized Procrustes analysis [2] to get the aligned shapes $x_i, i = 1, 2, \ldots, N$. The mean shape $\bar{x}$ is calculated as the simple average of the aligned shapes, $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. The shape space spanned by these N aligned shapes can be represented as a linear space with $K = \min\{3M - 1, N - 1\}$ eigen vectors, $V_1, \ldots, V_K$, based on principal component analysis (PCA) [2]. A new shape y in the aligned shape space can be represented

as

$$y = \bar{x} + \sum_{i=1}^{K} c_i V_i + e, \tag{1}$$

where c_i the so-called PCA coefficient, and e is a $3M$ dimensional vector for the residual error. Using the statistical shape model, a nonrigid shape can be represented parametrically as $(T, R, S, c_1, \ldots, c_K, \bar{x}, e)$, where T, R, S represents the translation, rotation, and scaling to transfer a nonrigid shape in the aligned shape space back to the world coordinate system.

With this representation, we can convert the segmentation (or boundary delineation) problem to a parameter estimation problem. Among all these parameters, $\bar{x}$ is fixed and e is sufficiently small if K is large enough (e.g., with enough training shapes). The original MSL only estimates the rigid part (T, R, S) of the transformation. Here, we extend MSL to directly estimate the parameters for nonrigid deformation $(c_1, \ldots, c_K)$. Given a hypothesis $(T, R, S, c_1, \ldots, c_K)$, we train a classifier based on a set of image features F to distinguish a correct hypothesis from a wrong one. The image features should be a function of the hypothesis, $F = F(T, R, S, c_1, \ldots, c_K)$, to incorporate sufficient information for classification. Steerable features were proposed in [1] to efficiently embed the object pose information into the feature set. The basic idea of steerable features is to steer (translate, rotate, and scale) a sampling pattern w.r.t. the testing hypothesis. On each sampling point, 24 local image features (e.g., intensity and gradient) are extracted. A regular sampling pattern was used in [1] to embed the object pose parameters (T, R, S). Here, we need to embed the nonrigid shape parameters $c_i, i = 1, \ldots, K$, into the sampling pattern too. For this purpose, we propose a new sampling pattern based on the synthesized nonrigid shape. Each hypothesis $(T, R, S, c_1, \ldots, c_K)$ corresponds to a nonrigid shape using the statistical shape model (see Eq. (1)). We use this synthesized shape as the sampling pattern and extract the local image features on its M points, resulting in a feature pool with $24 \times M$ features. If the hypothesis is close to the ground truth, the sampling points should be close to the true object boundary. The image features (e.g., gradient of the image intensity) extracted on these sampling points can help us to distinguish it from a wrong hypothesis where the sampling points are far from the object boundary and likely to lie in a smooth region. Similar to [1], we use the boosting technique to learn the classifier.

Due to the exponential increase of testing hypotheses, we cannot train a monolithic classifier to estimate all nonrigid deformation parameters simultaneously. Using the marginal space learning principle, we split the nonrigid deformation parameters into groups and estimate them sequentially. To be specific, we train a classifier in the marginal space of (T, R, S, c_1, c_2, c_3), where (c_1, c_2, c_3) correspond to the top three deformation modes. Given a small set of candidates after position-orientation-scale estimation, we augment them with all possible combinations of (c_1, c_2, c_3) and use the trained nonrigid MSL classifier to prune these hypotheses to a manageable number. In theory, we can apply the MSL principle to estimate more and more nonrigid deformation parameters sequentially. In practice, we find that with the increase of the dimensionality of the marginal spaces, the classifier is more likely to over-fit the data due to the limited number of training samples. No significant improvement has been achieved by estimating more than three nonrigid deformation parameters.

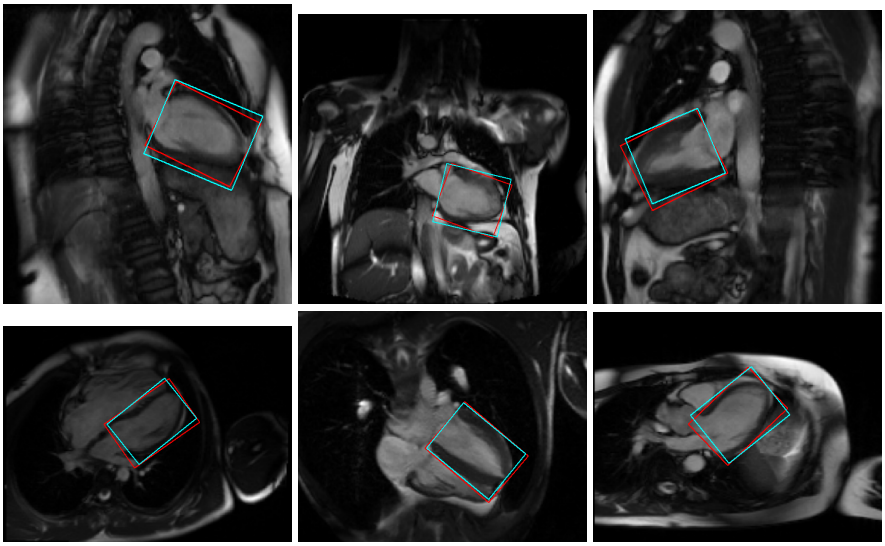


Fig. 5. Left ventricle detection results on 2D MRI images. Red boxes show the ground truth, while cyan boxes show the detection results.

Table 2. Comparison of marginal space learning (MSL) and full space learning (FSL) for LV bounding box detection in 2D MRI images on both the training (400 images) and test (395 images) sets. The errors are measured in millimeters.

	Training Set				Test Set			
	Center-Center Distance		Vertex-Vertex Distance		Center-Center Distance		Vertex-Vertex Distance	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Full Space Learning	9.73	1.79	17.31	5.07	43.88	21.01	63.26	46.49
Marginal Space Learning	**1.31**	**1.15**	**3.09**	**2.82**	**13.49**	**5.77**	**21.39**	**10.19**

5 Experiments

5.1 Left Ventricle Detection in 2D MRI Images

In this experiment, we quantitatively evaluate the performance of marginal space learning (MSL) and full space learning (FSL) for left ventricle (LV) detection in 2D MRI images. We collect 795 images of the LV long-axis view. Among them, 400 images are randomly selected for training and the remaining 395 images are reserved for testing. Two error measurements are used for quantitative evaluation, the center-center distance and the vertex-vertex distance (which is defined as the mean Euclidean distance between the corresponding vertices of the box). Table 2 shows detection errors of the LV bounding box obtained by MSL and FSL. It is quite clear that MSL achieves much better results than FSL. The mean center-center and vertex-vertex errors on the test set are 13.49 mm and 21.39 mm for MSL, respectively, which are about one third of the corresponding errors of FSL. MSL was originally proposed to accelerate 3D object detection [1]. In this comparison experiment, we find it also significantly improves the accuracy for 2D object detection.

Table 3. Comparison of unconstrained and constrained MSL on the number of testing hypotheses and computation time for liver detection in CT volumes.

	Unconstrained MSL [1]		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

Table 4. Comparison of constrained MSL and nonrigid MSL against the baseline version [1] on liver detection in 226 CT volumes. "Constrained + Nonrigid MSL" is the version combining both constrained MSL and nonrigid MSL. Average point-to-mesh error E_{p2m} (in millimeters) of the initialized shape is used for evaluation.

	Mean	Standard Deviation	Median
Unconstrained MSL [1]	7.44	2.26	6.99
Constrained MSL	7.12	2.15	6.73
Constrained + Nonrigid MSL	6.65	1.96	6.25

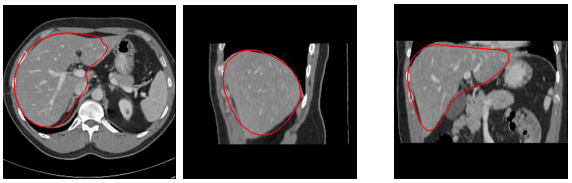
MSL is faster than FSL since much fewer hypotheses need to be tested. On a computer with a 3.2 GHz processor and 3 GB memory, the detection speed of MSL is about 1.49 seconds/image, while FSL takes about 13.12 seconds to process one image.

5.2 Liver Detection in 3D CT Volumes

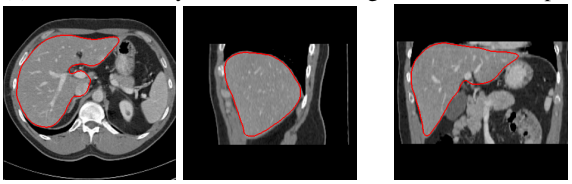
In this experiment, we compare constrained MSL and nonrigid MSL against the baseline version [1] on liver detection in 226 3D CT volumes. Our dataset is very challenging (including both contrasted and non-contrasted scans) because the volumes come from diverse sources. After object localization, we align the mean shape (a surface mesh) with the estimated transformation. Similar to [1], the accuracy of the initial shape estimate is measured with the symmetric point-to-mesh distance E_{p2m}. We can then 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 purpose.

The detection speed of MSL is roughly proportional to the number of testing hypotheses. The analysis presented in Section 3 shows that constrained MSL significantly reduces the number of testing hypotheses. Table 3 shows the break-down computation time for all 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 improves detection accuracy marginally. Since we constrain the search to a smaller but more meaningful region, the likelihood of detection outliers is reduced. As shown in Table 4, constrained MSL reduces the mean E_{p2m} error from 7.44 mm to 7.12 mm, and the median error from 6.99 mm to 6.73 mm, in a three-fold cross-validation.

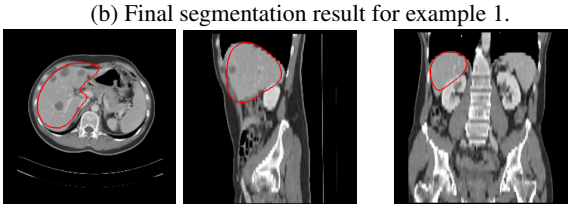
To further reduce the initialization error, we estimate three more nonrigid deformation parameters (top three PCA coefficients) to improve the object localization accuracy. As shown in the last row of Table 4, combining nonrigid MSL with constrained MSL, we can further reduce the average E_{p2m} to 6.65 mm (about 11% improvement compared to the original error of 7.44 mm).



(a) Initialization by constrained+nonrigid MSL for example 1.



(b) Final segmentation result for example 1.



(c) Initialization by constrained+nonrigid MSL for example 2.

(d) Final segmentation result for example 2.

Fig. 6. Typical liver segmentation results on two CT volumes. From left to right: transversal, sagittal, and coronal views.

Fig. 6 shows typical liver segmentation results on two volumes. Accurate boundary delineation is achieved starting from the good initial estimate of the shape achieved by the proposed method. After applying the learning-based non-rigid deformation estimation method [1], we achieve a final E_{p2m} error of 1.45 mm on 226 CT volumes (based on a three-fold cross-validation), which compares favorably with the state-of-the-art [9]. Our overall system runs as fast as ten seconds per volume (1 s for object localization using constrained+nonrigid MSL and 9 s for boundary delineation), while the state-of-the-art solutions take at least one minute [10], often up to 15 minutes [8, 11], to process a volume.

6 Conclusion

In summary, we made three major contributions in this paper. First, we performed a direct comparison experiment between MSL and FSL on 2D object detection, which

shows that MSL outperformed FSL on both speed and accuracy. 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 sampling 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. Last, nonrigid MSL was proposed to directly estimate the nonrigid deformation parameters to improve the localization accuracy for a nonrigid object. Comparison experiments on liver detection demonstrated the effectiveness of both the constrained and nonrigid versions of MSL.

Acknowledgments

The authors would like to thank Dr. Adrian Barbu for discussion on nonrigid MSL, Dr. Xiaoguang Lu for the comparison experiment on LV detection, and Dr. Haibin Ling for the help on the liver detection experiment.

References

1. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: 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) (2008) 1668–1681
2. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. Computer Vision and Image Understanding **61**(1) (1995) 38–59
3. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Pattern Anal. Machine Intell. **23**(6) (2001) 681–685
4. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Int. J. Computer Vision **1**(4) (1988) 321–331
5. Ecabert, O., Peters, J., H. Schramm et al.: Automatic model-based segmentation of the heart in CT images. IEEE Trans. Medical Imaging **27**(9) (2008) 1189–1201
6. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. (2001) 511–518
7. Karney, C.F.F.: Quaternions in molecular modeling. Journal of Molecular Graphics and Modeling **25**(5) (2007) 595–604
8. Heimann, T., Münzing, S., Meinzer, H.P., Wolf, I.: A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation. In: Proc. IPMI. (2007)
9. van Ginneken, B., Heimann, T., Styner, M.: 3D segmentation in the clinic: A grand challenge. In: MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge. (2007)
10. Ruskó, L., Bekes, G., Németh, G., Fidrichf, M.: Fully automatic liver segmentation for contrast-enhanced CT images. In: MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge. (2007)
11. Kainmueller, D., Lange, T., Lamecker, H.: Shape constrained automatic segmentation of the liver based on a heuristic intensity model. In: MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge. (2007)