

Deformable Segmentation via Sparse Shape Representation

Shaoting Zhang², Yiqiang Zhan¹, Maneesh Dewan¹, Junzhou Huang²,
Dimitris N. Metaxas², and Xiang Sean Zhou¹

¹Siemens Medical Solutions, Malvern, PA, USA

²Department of Computer Science, Rutgers University, Piscataway, NJ, USA

Abstract. Appearance and shape are two key elements exploited in medical image segmentation. However, in some medical image analysis tasks, appearance cues are weak/misleading due to disease/artifacts and often lead to erroneous segmentation. In this paper, a novel deformable model is proposed for robust segmentation in the presence of weak/misleading appearance cues. Owing to the less trustable appearance information, this method focuses on the effective shape modeling with two contributions. First, a shape composition method is designed to incorporate shape prior on-the-fly. Based on two sparsity observations, this method is robust to false appearance information and adaptive to statistically insignificant shape modes. Second, shape priors are modeled and used in a hierarchical fashion. More specifically, by using affinity propagation method, our deformable surface is divided into multiple partitions, on which local shape models are built independently. This scheme facilitates a more compact shape prior modeling and hence a more robust and efficient segmentation. Our deformable model is applied on two very diverse segmentation problems, liver segmentation in PET-CT images and rodent brain segmentation in MR images. Compared to state-of-art methods, our method achieves better performance in both studies.

1 Introduction

In various applications of medical image segmentation, deformable model has achieved tremendous success, which should be contributed to its joint employment of shape and appearance characteristics. While appearance features provide low level clues of organ boundaries, shape imposes high level knowledge to infer and refine deformable model. However, in some medical image analysis, appearance cues are relatively weaker or even misleading (Fig. 1). In those cases, the best “guess” of the organ boundaries can only come from shape priors, which should be effectively modeled from training shapes. However, effective shape modeling is confronting these challenges, 1) shape variation is complex and cannot always be modeled by a parametric probability distribution; 2) a shape instance derived from image appearance cues (input shape) may have gross errors; and 3) local details of the input shape are difficult to preserve if they are not statistically significant in the training data. Traditional deformable model, e.g., Active Shape Model its extensions [1,6], can not tackle them uniformly.

In this paper, we propose a deformable model aiming to achieve robust segmentation in the presence of weak/misleading appearance cues. In particular, two novel methods are designed for robust and effective shape prior modeling. *First*, instead of assuming any parametric model of shape statistics, we propose to incorporate shape priors on-the-fly through sparse representation. More specifically, we have two sparsity observations: 1) Given a large shape repository of an organ, a shape instance of the same organ can be approximated by the composition of a sparse set of instances in the shape repository; and 2) gross errors from local appearance cues might exist but these errors are sparse in spatial space. Incorporating these two sparsity priors, our deformable model becomes robust to gross errors and can preserve shape details even they are not statistically significant in the training repository. This shape composition method benefits both the model initialization and refinement. *Second*, instead of modeling global shape priors, we propose to decompose the deformable surface to multiple parts and build shape models on them independently. The partition is accomplished by affinity propagation method [4] based on image and geometry features. Since the shape statistics of local structures often has more compact distribution than global structures, this strategy facilitates better shape modeling and increases algorithm runtime efficiency.

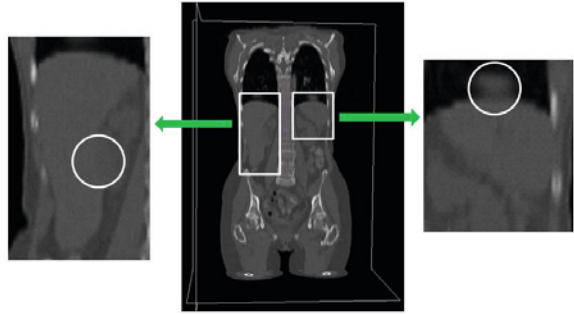


Fig. 1. Middle: whole body low-dose CT data. Left: zoom in of the liver and lung. In the marked region the boundary between the liver and the kidney is hardly observed. The appearance cue is weak because of the low contrast around the boundary. Right: zoom in of the spleen and lung. In the marked region, there is artifact induced by breath. It is part of the lung. Since the image information is misleading here, segmentation methods solely relying on appearance cue may accidentally include this region as spleen.

2 Methodology

Segmentation Framework: To achieve generality, our segmentation framework is designed in the spirit of “data-driven”. Fig. 2 shows the workflow of our segmentation system, which consists of offline learning and runtime segmentation stages. In offline learning, 3D volume images along with the manually labeled ground truths are employed to learn the appearance and shape characteristics of the organ under study. More specifically, methods proposed in [10,9] are used to learn landmark detectors and a set of spatially adaptive boundary detectors. Meanwhile, organ surfaces are stored in a shape repository, which will

be exploited to derive shape priors during runtime. Runtime segmentation starts from the initialization of the model surface (represented by a triangular surface) based on automatically detected landmarks and shape priors. The surface then deforms under the guidance of both image appearance cues and shape priors. More specifically, two steps are performed iteratively until convergence. First, the surface model deforms to local places where the learning-based boundary detectors generate higher responses. Next, the locally deformed surface is refined by the shape priors derived from the shape repository.

As discussed before, although learning-based landmark/boundary detectors can tackle reasonable appearance variations [10,9], they might generate wrong responses in the presence of severe imaging artifacts/diseases, and hence mislead the deformable model. In this scenario, shape prior is the only information source to initialize/correct the deformable surface. (Note that shape priors are employed in both landmark-based model initialization and shape refinement in Fig. 2.) Therefore, the effective modeling of shape priors becomes extremely critical to achieve a robust segmentation. Due to page limits, we will focus on the modeling of shape priors in the remainder of this paper.

Shape Prior Modeling via Sparse Composition: Instead of assuming any parametric probabilistic distributions of the shape statistics, our shape prior model is based on two observations: 1) After being aligned to a common canonical space, any shape can be approximated by a sparse linear combination of other shape instances in the same shape category. Approximation residuals might come from inter-subject variations. 2) If the shape to be approximated is derived by appearance cues, residual errors might include gross errors from landmark/boundary detections. But such errors are sparse as well. Accordingly, we aim to incorporate shape priors *on-the-fly* through shape composition, i.e., a shape derived by appearance cues is refined by the approximation of a set of annotated shape instances following the two sparsity observations. It is worth mentioning that sparsity has been adopted in segmentation algorithms in different manners, such as the sparse information models [3], which reconstruct a 3D surface from sparse subcomponents.

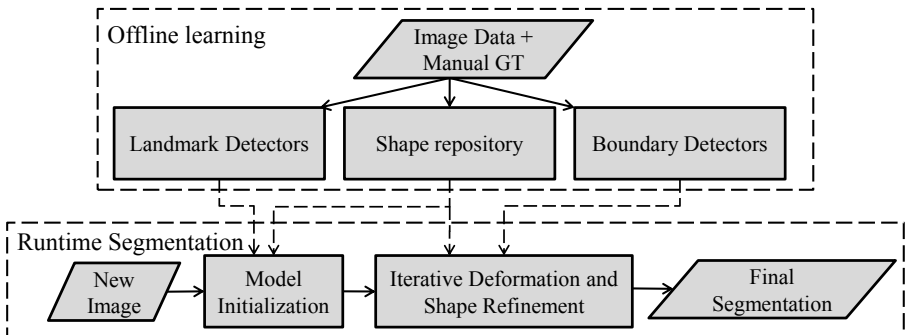


Fig. 2. The workflow of our segmentation framework

In this work, a shape is represented by a triangle mesh which consists of a set of vertices. Denote the input shape as \mathbf{v} , where $\mathbf{v} \in \mathbf{R}^{3N}$ is a vector concatenated by 3D coordinates of its N vertices. (In the remainder of this paper, any shape instance is defined as a vector in the same way and has the same dimensionality.) Assume $D = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K] \in \mathbf{R}^{3N \times K}$ is a large shape repository that includes K accurately annotated shape instances \mathbf{d}_i . Note that $\{\mathbf{d}_i, i = 1, 2, 3, \dots, K\}$ are pre-aligned using generalized Procrustes analysis [5,1]. The approximation of \mathbf{v} by D is then formulated as an optimization problem:

$$\arg \min_{\mathbf{x}, \mathbf{e}, \beta} \|T(\mathbf{v}, \beta) - D\mathbf{x} - \mathbf{e}\|_2^2, \quad s.t. \|\mathbf{x}\|_0 < k_1, \|\mathbf{e}\|_0 < k_2, \quad (1)$$

where $T(\mathbf{v}, \beta)$ is a global transformation operator with parameter β , which aligns the input shape \mathbf{v} to the common canonical space of D . The key idea of our shape prior modeling lies in the two constraints of (1). In the first constraint, $\mathbf{x} \in \mathbf{R}^K$ denotes the coefficient/weights of linear combination. The L_0 -norm of \mathbf{x} ensures that the number of nonzero elements in \mathbf{x} is less than k_1 . In other words, only a sparse set of shape instances can be used to approximate the input shape, which prevents the overfitting to errors from missing/misleading appearance cues. In the second constraint, $\mathbf{e} \in \mathbf{R}^{3N}$ is a vector that models the large residual errors. The sparsity constraint is imposed on \mathbf{e} to incorporate the observation that gross errors might exist but are occasional.

(1) is a NP hard problem owing to the non-convex L_0 norm. Thanks to the recent proof of the sparse representation theorem [2], L_1 norm relaxation can be employed to make the problem convex while still preserving the sparsity property. However, to solve (1), we still need to simultaneously optimize multiple variables and deal with the nonlinearity if $T(\mathbf{v}, \beta)$ is modeled as a rigid or similarity transformation. Our solution is to use an Expectation-Maximization (EM) style algorithm (or alternating minimization) to solve (1). It is divided into two sub-problems: 1) estimate the transformation parameter β and 2) efficiently minimize the simplified linear inverse problem with the aligned shape. In the ‘‘E’’ step, β is estimated using Procrustes analysis, which aligns the shape \mathbf{v} to the canonical space as $\mathbf{v}' = T(\mathbf{v}, \beta)$. In the ‘‘M’’ step, the following simplified problem is minimized:

$$\arg \min_{\mathbf{x}, \mathbf{e}} \|\mathbf{v}' - D\mathbf{x} - \mathbf{e}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{e}\|_1, \quad (2)$$

Since (2) now becomes a typical linear inverse problem, it can be solved using existing solvers. The ‘‘E’’ and ‘‘M’’ steps are iteratively performed until \mathbf{x} , \mathbf{e} and β converge. $D\mathbf{x}$ is then computed as a refined version of the input shape, which imposes the shape priors on-the-fly.

Multi-resolution Shape Refinement and Local Shape Priors: It has been widely accepted that multiresolution/hierarchical scheme should be employed to improve the efficiency and robustness of deformable segmentation [7]. In a multiresolution scheme, only a small set of sparsely distributed vertices are used as driving vertices to estimate a rough segmentation of the initial stages. As the

iterations increase, more and more vertices join the driving set to gradually reach accurate segmentation. Our sparse shape composition method naturally supports this scheme by estimating a sparse linear combination from an incomplete input. Assume $\mathbf{v}_{sub} = \mathbf{S}\mathbf{v}$ is a subset of all vertices in shape \mathbf{v} , where \mathbf{S} is a binary diagonal matrix which indicates if the i th vertex is in the subset ($\mathbf{S}_{ii} = 1$). (1) can then be naturally extended as:

$$\arg \min_{\mathbf{x}, \mathbf{e}, \beta} \|T(\mathbf{v}_{sub}, \beta) - \mathbf{S}D\mathbf{x} - \mathbf{S}\mathbf{e}\|_2^2, \quad s.t. \|\mathbf{x}\|_0 < k_1, \|\mathbf{e}\|_0 < k_2, \quad (3)$$

(3) can be solved using the same EM optimization. The only difference is that the optimized \mathbf{x} will be finally applied on the full space of D , such that the entire input shape is refined.

One extreme situation of (3) is that \mathbf{S} becomes very sparse and only includes a few vertices (usually with the most distinctive appearance/geometry characteristics). In this situation, (3) indeed becomes the formula of landmark-based surface initialization, which is the first step of our runtime segmentation system. Again, by incorporating shape priors with the assumption of “sparse gross errors”, our initialization method becomes robust to erroneous landmark detections due to severe diseases/imaging artifacts.

The merit of (3) is actually beyond the support of multiresolution deformation scheme. In practice, many 3D deformable models include many thousands of points to give an accurate description of organ shapes. The optimization of (2) thus has high computational complexity. In addition, since local shape statistics often lie in a more compact space than global ones, shape priors built on sub-surface are expected to improve the performance. To achieve this goal, we propose a “mesh partitioning” method, which can also be seamlessly incorporated in our sparse shape composition formula. Affinity propagation clustering [4] is employed to divide the model shape into multiple partitions. Since one-to-one correspondences are already constructed among all shapes, affinity propagation only needs to perform once for the model shape. The similarity used in the affinity propagation is defined as the combination of the image similarity and geodesic distances between vertices [9].

In our implementation, each divided partition is further “dilated” for several levels to produce overlaps with neighboring partitions. Finally, partitions are converted to a set of indication matrices $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_p$ used in (3). The optimization problem defined on the entire surface is thus decomposed to a set of sub-problems. Each partition is refined independently but the refined partitions are averaged in these overlapping regions to guarantee the smoothness of the entire surface.

The computational complexity of an existing solver (e.g., interior point method) is $\mathcal{O}(N^3)$, where N is the number of vertices of the whole surface. After dividing the whole surface into p partitions with about $\frac{N}{p}$ vertices in each partition. The computational complexity is decreased to only $\frac{1}{p^2}$ of the original one, which highly improves the efficiency.

3 Experiments

Liver Segmentation from

Low-dose CT: We evaluate the segmentation performance of our system, using 3D low-dose CT data from PET-CT. In wholebody PET-CT scan, CT images usually have low dose and large slice thickness, which result in low contrast and fuzzy boundaries between organs. Hence, organ segmentation in PET-CT becomes more challenging than traditional CT. In our experiment, the 3D ground truth of low-dose CT is manually segmented by multiple clinical experts. 40 out of 67 CT scans are used to train the landmark detector and also

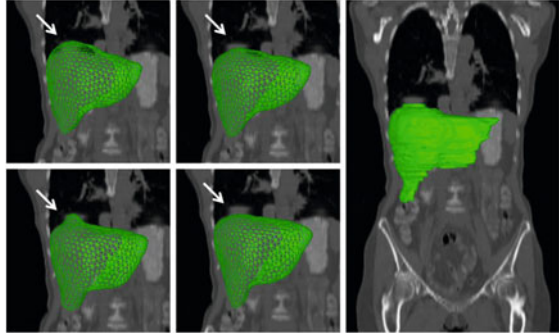


Fig. 3. Initialization results (1st row) and deformation results (2nd row) from the corresponding initialization. Results from PA (1st column) and our method (2nd column) and ground truth (right-most figure). PA incorrectly includes part of the lung because of the artifacts inducing by breath (see arrows).

used to construct the shape repository D . The other 27 are left for testing. To obtain the one-to-one correspondence for vertices among all shapes, we choose one shape as a reference and register it to all the others using adaptive-focus deformable model (AFDM) [8]. The shape has around 1,000 vertices, and 20 are selected as landmarks for model initialization. The proposed method is compared to three popular algorithms: 1) PA: Procrustes Analysis [5] is used to find a similarity transformation to fit a mean shape to detected landmarks. There is no shape refinement during deformation. 2) SMS: It is the Shape Model Search module in ASM [1], which employs the PCA method to learn shape statistics and refine the input shape. 3) SI-NN: k -nearest neighbors method, which uses nearest neighbors to find the closest prototypes in the expert's structure annotations. For a fair comparison, same landmark/boundary detectors and deformation strategy are used in all methods. They only differ in model initialization and model refinement, which involve shape priors.

Fig. 3 compares the landmark detection based initialization. Since the image contrast of low-dose CT is very low and there are breathing artifacts in the lung region, the landmark detector may easily fail to locate correct positions. Our method is less sensitive to such errors. Its initialization result is already very close to the object boundary. We also compare the deformation results starting from different initializations. A better initialized model also benefits the deformation performance.

To quantitatively evaluate the 3D segmentation accuracy, we report the mean value and standard deviation of the distances between shape surfaces in Tab. 1.

Our proposed framework achieves the best performance. The standard deviations in Tab. 1 show that our method also achieves the best stability among all methods. To evaluate the benefit of mesh partitioning, the surface mesh is divided into 30 regions. The shape refinement step takes several minutes when applied to the whole surface directly. Using mesh partitioning, it significantly improves the efficiency and only takes 2-3 seconds. The whole system takes around 20 seconds (a Python implementation on a PC with 2.4GHz Intel Quad CPU) to segment liver in a 512x512x300 CT volume. Note that the shape refinement module not only improves the robustness of the deformable model, but also decreases the iteration times of deformation since it helps avoid local minima of image information.

Table 1. Quantitative comparisons of the mean values and standard deviations of the distances (voxel) between surfaces. 1st column compares initialization results. Note that SMS and PA uses the same initialization.

Method	Fig. 3(Init)	All data
PA	2.26 ± 1.72	3.96 ± 3.21
SMS	2.26 ± 1.72	2.16 ± 1.68
SI-NN	4.88 ± 3.61	3.82 ± 3.12
Ours	1.31 ± 0.95	1.13 ± 0.83

Rodent Brain Structure Segmentation from MRI: In this study, we use the proposed method to segment rodent brain structures in MR images. In our experiments, 58 data are delineated by clinical experts. 40 are used as training data, and the rest 18 are used as testing. We focus on the 3D segmentation of the cerebellum (Fig. 4). This task is challenging in two aspects. First, there are complex textures and high gradient values inside of the cerebellum region, which adversely affect the deformation module. Second, rodent cerebellum contains two protruding parts, which are easily to be falsely “smoothed out” by traditional shape prior modeling. The visual comparison of the segmentation results is shown in Fig. 4. With regular shape constraint, the protruding parts are shrunk, and

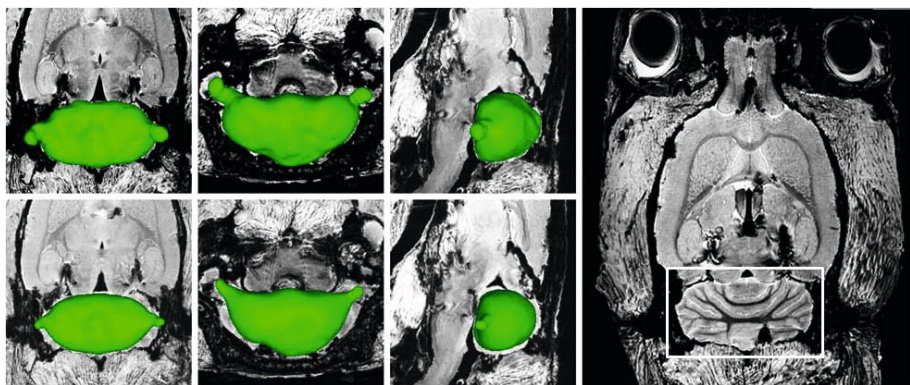


Fig. 4. Segmentation results of the rodent cerebellum. The 1st row is from the proposed method. The 2nd row is from the same framework but without shape prior constraint. The rightmost figure is one slice of the MR image with cerebellum highlighted.

the whole shape is attracted by the high internal gradient. Most regions are under-segmented. The error is 5.86 ± 3.68 , in terms of voxel distance. These problems are well tackled by the proposed method. The protruding parts are well preserved, and the global shape are properly constrained. It achieves better segmentation results, with error 1.31 ± 0.91 .

4 Conclusions

In this paper, we proposed a deformable model aiming to robustly segment organs in the presence of weak/misleading appearance cues. A sparse shape composition method is proposed to model and incorporate shape priors on-the-fly. It is able to tackle three challenges in a unified framework and naturally supports multi-resolution deformation scheme. Furthermore, we use the affinity propagation method to partition the surface shape local shape priors. Besides a more efficient shape prior modeling, this strategy also dramatically increase run-time efficiency. The majority of the work was carried out when Shaoting Zhang was a research intern at Siemens Medical Solutions, USA. <http://www.research.rutgers.edu/~shaoting/research/siemens2010/project.htm>

References

1. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape model - their training and application. *Computer Vision and Image Understanding* 61, 38–59 (1995)
2. Donoho, D.: For most large undetermined systems of equations, the minimal ℓ_1 -norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics* 59(7), 907–934 (2007)
3. Florin, C., Paragios, N., Funka-Lea, G., Williams, J.: Liver segmentation using sparse 3D prior models with optimal data support. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007*. LNCS, vol. 4584, pp. 38–49. Springer, Heidelberg (2007)
4. Frey, B., Dueck, D.: Clustering by passing messages between data points. *Science* 315(5814), 972 (2007)
5. Goodall, C.: Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society* 53, 285–339 (1991)
6. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis* 13(4), 543–563 (2009)
7. Langs, G., Paragios, N., Essafi, S.: Hierarchical 3D diffusion wavelet shape priors. In: *IEEE 12th International Conference on Computer Vision*, pp. 1717–1724 (2010)
8. Shen, D., Davatzikos, C.: An adaptive-focus deformable model using statistical and geometric information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8), 906–913 (2000)
9. Zhan, Y., Dewan, M., Zhou, X.: Cross modality deformable segmentation using hierarchical clustering and learning. In: Yang, G.Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 1033–1041. Springer, Heidelberg (2009)
10. Zhan, Y., Zhou, X., Peng, Z., Krishnan, A.: Active scheduling of organ detection and segmentation in whole-body medical images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I*. LNCS, vol. 5241, pp. 313–321. Springer, Heidelberg (2008)