

Cross Modality Deformable Segmentation Using Hierarchical Clustering and Learning

Yiqiang Zhan, Maneesh Dewan, and Xiang Sean Zhou

CAD R&D, Siemens Healthcare, Malvern, PA USA

Abstract. Segmentation of anatomical objects is always a fundamental task for various clinical applications. Although many automatic segmentation methods have been designed to segment specific anatomical objects in a given imaging modality, a more generic solution that is directly applicable to different imaging modalities and different *deformable surfaces* is desired, if attainable. In this paper, we propose such a framework, which learns from examples the *spatially adaptive appearance* and *shape* of a 3D surface (either open or closed). The application to a new object/surface in a new modality requires *only* the annotation of training examples. Key contributions of our method include: (1) an automatic clustering and learning algorithm to capture the spatial distribution of appearance similarities/variations on the 3D surface. More specifically, the model vertices are hierarchically clustered into a set of anatomical primitives (sub-surfaces) using both geometric and appearance features. The appearance characteristics of each learned anatomical primitive are then captured through a cascaded boosting learning method. (2) To effectively incorporate non-Gaussian shape priors, we cluster the training shapes in order to build multiple statistical shape models. (3) To our best knowledge, this is the first time the same segmentation algorithm has been directly employed in two very diverse applications: a. Liver segmentation (closed surface) in PET-CT, in which CT has very low-resolution and low-contrast; b. Distal femur (condyle) surface (open surface) segmentation in MRI.

1 Introduction

In recent decades, automatic/semi-automatic algorithms for the delineation of anatomical structures has become more and more important in assisting and automating specific radiological tasks. Hence, it is not surprising that there have been multitude of algorithms developed in recent years, each tailored towards a particular anatomical structure and imaging modality, typically with a few key parameters to tweak for the algorithm to work. In spite of the availability of all these algorithms [1][2][3][4], generally, it is not easy to make these algorithms to work on another structure/organ and/or imaging modality.

To that end, in this paper, we propose a learning-based hierarchical deformable model to segment various organs (or structures) from different medical imaging modalities. Compared to the existing methods in the “deformable

model” family, our method has three hallmarks. *First*, the boundary appearance is hierarchically modeled and learned in a spatially adaptive way. More specifically, the vertices of the deformable model are hierarchically clustered based on both the geometric and appearance similarity. Boundary characteristics of each cluster is then captured by training a boundary detector using a cascade boosting method, finally providing a hierarchically learned model for deformation. *Second*, shape priors are modeled by multiple statistical shape models built upon clustered shape instances. Since each cluster of shape instances represents one of the distribution modes in the shape space, multiple statistical shape models are able to provide more “specific” refinement to the deformable model. *Third*, and finally we demonstrate the efficacy of our algorithm (without changing any parameters), on two different anatomical structures (liver and distal end of femur) in two contrasting imaging modalities (PET-CT and MR).

2 Related Work

Deformable model is a vigorously studied model-based approach in the area of medical image segmentation. The widely recognized potency of deformable models comes from their ability to segment anatomic structures by exploiting constraints derived from the image data (bottom-up) together with prior knowledge about these structures (top-down). The deformation process is usually formulated as an optimization problem whose objective function consists of external (image) term and internal (shape) term. While internal energy is designed to preserve the geometric characteristics of the organ under study, the external energy is defined to move the model toward organ boundaries. Traditional external energy term usually comes from edge information [5], e.g., image gradient. In recent years, more effort has been invested on the integration of other image features, e.g., local regional information [1][2] and texture models [3]. By combining different image features as the external energy, deformable model achieves tremendous success in various clinical practices. However, as these external energy terms are usually designed for specific imaging modality and organ, they lack scalability to different medical imaging modalities. Machine learning technologies have opened the door for a more generic external energy design. By using learning-based methods, boundary characteristics can be learned from training data [6][7]. In other words, the “design” of external energy becomes data driven and extensible to different imaging modalities. A potential problem is that the boundary characteristics of organs can seldom be

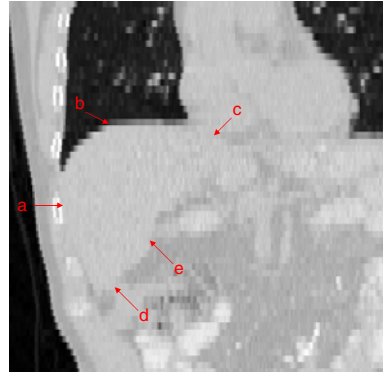


Fig. 1. An example of liver CT images. Arrows *a-e* point to boundaries between liver and rib, lung, heart, abdomen and colon that show heterogeneous appearance.

learned by a single classifier due to heterogenous characteristics along organ boundaries (c.f. Fig 1). To address this problem, a “divide-and-conquer” strategy is desired. More specifically, the deformable model should be decomposed into a set of sub-surfaces with relatively similar boundary characteristics.

Moreover, by hierarchically decomposing the deformable model into a set of deformation units, the speed and the robustness of segmentation can be highly improved [8]. However, the hierarchical structure was designed heuristically in [8]. Hence, a more principled way to generate the hierarchical structure, as presented in this work, would be highly desirable.

3 Method

3.1 Overview

In our study, we aim to develop a deformable model that is extensible to different imaging modalities. To achieve this purpose, we propose a learning-based hierarchical model, which is purely data-driven. Given a set of manually segmented training data, the hierarchical structure of the deformable model is constructed through an iterative clustering and feature selection method. As shown in Fig. 3, every node of the hierarchical structure represents a sub-surface of the deformable model. For each primitive sub-surface, i.e., leaf nodes in the hierarchical tree, a boundary detector is learned using a cascade boosting method. The ensemble of these learned boundary detectors actually capture the appearance characteristics of a specific organ in a specific imaging modality. Their responses guide the deformable model to the desired organ boundary. In addition to the hierarchical model and the boundary detectors, a set of statistical shape models are built upon clustered shape instances in the learning stage. These shape priors will be used to constrain the deformable model at run-time. The diagram of our method is shown in Fig 2.

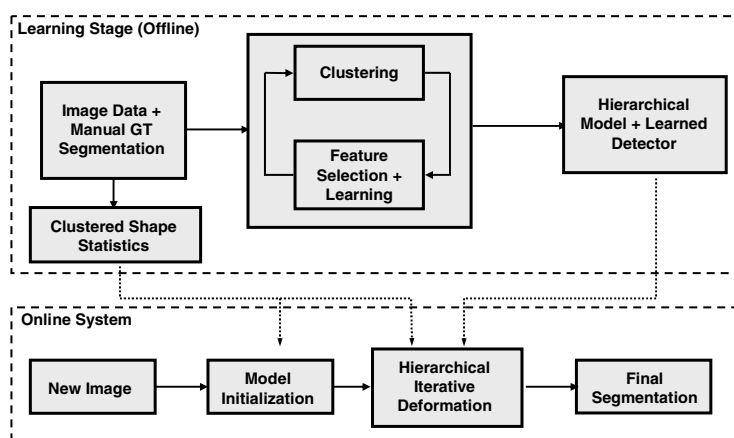


Fig. 2. Flowchart of the learning-based hierarchical model showing both the offline learning and the online testing system

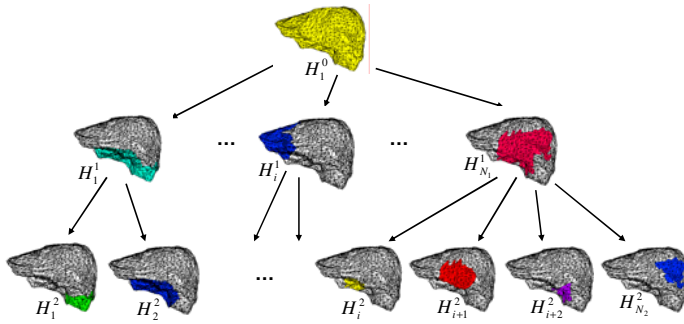


Fig. 3. Hierarchical structure of the deformable model. Color patches depict the sub-surfaces (H_i^j) at j th hierarchical level comprised of vertices in the i th cluster.

Model Description. Our deformable model is represented by a triangle mesh: $S \equiv (V, T)$, where $V = \{v_i | i = 1, \dots, N\}$ denotes the vertices of the surface and $T = \{t_j | j = 1, \dots, M\}$ denotes the triangles defined by vertices. Mathematically, the segmentation problem is formulated as the minimization of an energy function defined as:

$$E(S) = E_{ext}(S) + E_{int}(S) = \sum_{i=1}^N E_{ext}(v_i) + E_{int}(S) \tag{1}$$

where E_{ext} and E_{int} are the image (external) energy term and shape (internal) energy term. A hierarchical deformation strategy is employed to solve this high dimensional optimization problem (Refer [6] for details). Due to page limits, we will focus on the energy terms in the remainder of this paper.

3.2 Hierarchical Model with Learning-Based Boundary Detectors

Our external energy is defined by the responses of a set (or ensemble) of boundary detectors built upon the hierarchical deformable model. The following steps are used to generate the hierarchial model and the boundary detectors.

Affinity Propagation Clustering. ‘‘Affinity propagation’’ method [9] is employed to cluster vertices. Affinity propagation method models each data point as a node in a network. During the clustering process, real-valued messages are recursively exchanged between data points until a high quality set of exemplars and corresponding clusters emerge. Compared to other clustering methods, affinity propagation considers each data point as a potential cluster center and gradually generate clusters. Therefore, the solution is not sensitive to bad initialization and hard decision.

In our study, it is important to design an appropriate similarity between vertices since it determines the clustering results of affinity propagation. We put two constraints on the clustered vertices. First, to facilitate the characterization of heterogenous boundary, vertices in the same cluster should have relatively

similar image features. Second, the hierarchical deformable model requires the vertices within a cluster to be proximal to each other on the surface. In this way, the cluster center can be treated as a “driving vertex” and drive its neighborhood in the deformation process. To achieve these two purposes, the similarity between vertices is defined as follows.

$$s(v_i, v_j) = 1 - (1/K) \sum_{k=1}^K [\alpha G(v_i^k, v_j^k) + (1 - \alpha)C(\mathcal{F}(v_i^k), \mathcal{F}(v_j^k))] \quad (2)$$

Here, K is the number of training subjects, v_i^k denotes the i th vertex of the k th subject. $G(v_i^k, v_j^k)$ denotes the geodesic distance between v_i^k and v_j^k . $C(\mathcal{F}(v_i^k), \mathcal{F}(v_j^k))$ denotes the Euclidean distance between image feature vectors calculated at v_i^k and v_j^k .

Iterative Feature Selection/Clustering. To construct the hierarchical structure of the deformable model, vertices are recursively clustered. Assume H_i^j is the i th cluster at the j th hierarchical level, vertices belonging to H_i^j are further clustered to a set of sub-clusters $\{H_k^{j+1}, k = 1, \dots, N_i\}$:

$$H_i^j = \bigcup_{k=i_1}^{i_{N_i}} H_k^{j+1} \text{ and } \bigcap_{k=i_1}^{i_{N_i}} H_k^{j+1} = \emptyset \quad (3)$$

The remaining problem is the selection of appropriate $\mathcal{F}(\cdot)$ in Eq. 2. This is actually an “egg-and-chicken” problem. On one hand, to achieve the desired clusters, we need to know the distinctive feature sets for boundary description. On the other hand, distinctive features for local boundary can be obtained only after we have the vertices cluster. To address this problem, we propose an iterative clustering and feature selection method.

For the first level of cluster, we use intensity profile along normal of the vertices as $\mathcal{F}(\cdot)$. After that, assume $H_i^j = \{v_l\}$, we use *Adaboost* method to select the features that are most powerful to distinguish $\{v_l\}$ from the points along their normal directions, both inside and outside of the surface. The selected feature set are used as $\mathcal{F}(\cdot)$ in Eq. 2 to further cluster $\{v_l\}$ to a set of sub-clusters $\{H_k^{j+1}, k = i_1, \dots, i_{N_i}\}$. Feature selection and clustering are iteratively executed until boundary characteristics within a cluster becomes learnable.

Learn Boundary Detectors. For each primitive cluster, i.e., the leaf node of the hierarchical tree, a boundary detector is learned to characterize local boundary. In principle, we follow the idea of [10] to use an extensively redundant feature pool and a cascade *Adaboost* method to learn a boundary detector. Given an image I , $\mathfrak{F}(\mathbf{x}; I)$ denotes the redundant feature vector of \mathbf{x} . (In practice, we use 2D, 3D or 4D Haar-like features depends on the dimensionality of different image modalities.) In run-time system, each learned classifier generates a boundary probability map $P(\mathbf{x}|I)$. Hence, the external energy term in Eq. 1 is defined as:

$$E_{ext}(v_i) = 1 - P(v_i|I) = 1 - C_{\tilde{h}_{v_i}}(\mathfrak{F}(v_i; I)) \quad (4)$$

where \tilde{h}_{v_i} is the cluster index of v_i and $C_{\tilde{h}}$ defines the corresponding classifier.

3.3 Multiple Statistical Atlas Built on Clustered Shape Instances

In the well-known active shape model [11], shape prior is modeled by a statistical model built upon the whole population of shape instances. However, the assumption that the shape instances follow a mono-Gaussian distribution in the shape space might not always hold for some organs. An effective solution is to build sub-population models [12]. Inspired by this idea, we build multiple shape models upon clustered shape instances. Given a set of training shapes, they are firstly clustered according to pair-wise shape similarity, which is defined by the Euclidean distance between shape vectors. A set of statistical models are then built on each clusters, respectively. In the run time, the deformable model is constrained by the most similar statistical shape model. More specifically, the deformed model is mapped to the eigen-spaces of each shape cluster, respectively. The shape statistical model that gives the most compact description is selected to refine the deformed model. Hence, the internal energy of Eq. 1 is formulated as:

$$E_{int}(S) = 1 - \max_i e^{(-S-S_i)^T \Xi_i (-S-S_i)} \quad (5)$$

where S_i and Ξ_i denote the average shape and the covariance matrix of the i th shape cluster, respectively.

4 Results

Liver segmentation in wholebody PET-CT. Wholebody PET-CT provides fused morphological and functional information, which benefits cancer diagnosis

and therapy evaluation. As the standardized uptake value of liver is usually higher than surrounding tissues, it is desired to segment liver from PET-CT for an organ-specific PET-CT interpretation. In this study, the learning-based hierarchical model was trained by 20 wholebody PET-CT scans with manually delineated liver surfaces. (To jointly exploit CT-PET information, 4D Haar-like filters are used for feature extractor.[13]) As shown in Fig. 4, the generated model has two hierarchical levels with 8 and 25 vertices clusters, respectively. The automatic segmentation results on 30 testing dataset

(PET: $5 \times 5 \times 5mm$; CT: $1.3 \times 1.3 \times 5mm$) are compared with manually delineated organ surfaces (see Fig. 5). Accuracy measurements include median distance between surfaces, average distance between surfaces, volume difference and volume overlap difference. In Table 1, we compare our proposed

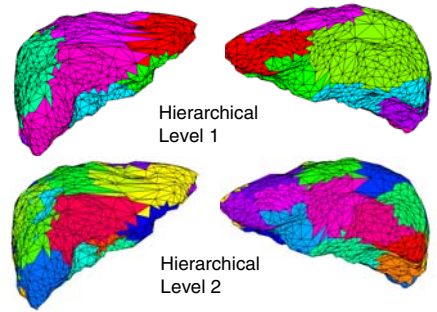


Fig. 4. 3D rendering of the hierarchical structure of a liver model. Color patches denote the vertices belonging to the same cluster. Left: Anterior View. Right: Posterior view.

Table 1. Quantitative comparison of learning based methods on PET-CT liver segmentation

	Med Surf. Dist.(voxel)	Avg Surf. Dist (voxel)	Vol. Diff. (%)	Overlap Diff. (%)
<i>Method1</i>	0.84	1.01	3.13	7.61
<i>Method2</i>	1.27	1.61	5.16	12.1

method (*Method1*) with *Method2*, which is a learning-based deformable model with *heuristically* designed hierarchical structure. More specifically, in *Method2* the hierarchical structure is determined by clustering neighboring vertices, based only on geodesic information¹. Therefore, the spatially clustered vertices in *Method2* might have larger appearance variation, which is difficult to learn. Hence, *Method2* shows inferior performance.

Distal femur condyle surface segmentation in MR T1 image. The morphological shape and geometry of the condyle surface is not only important for

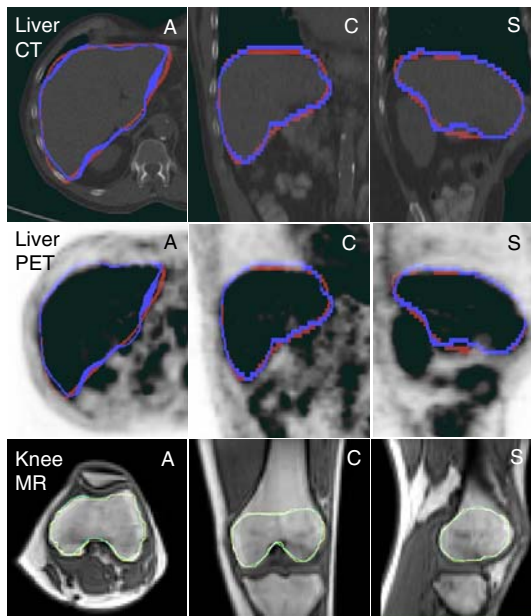


Fig. 5. Comparison of manual and automatic segmentation. Blue/yellow and red/blue contours denote the automatic and manual segmentation, respectively.

¹ Note that *Method2* only differs from *Method1* in the way their hierarchical structures are built.

² As the distal femur surface is an open surface, we cannot provide the volume error in this experiment.

understanding the kinematic function of the knee, but also has clinical significance in areas of total knee arthroplasty and anterior cruciate ligament reconstruction [14]. The automatic segmentation of condyle surface is useful to knee disease diagnosis and therapy planning. We apply the same method on the segmentation of distal femur MR T1 images. As shown in Fig. 5, the automatic segmentation is very close to manual segmentation. Tested on 21 knee MR images (T1, $1.953 \times 1.953 \times 2mm$), the average distance between manually the automatically delineated surfaces is **1.87mm** (in sub-voxel precision).²

Importantly, for both liver PET-CT and femur distal surface segmentation, we use exactly same parameters. The accurate results shows that our method is transparent to different imaging modalities and different organs.

5 Conclusion

In this paper, we proposed a learning-based hierarchical deformable model for automatically segmenting anatomic structures across different imaging modalities. The boundary appearance of the anatomical structure is hierarchically modeled through an iterative clustering and learning framework. In addition, the non-gaussian shape variations of these anatomical structures are modeled through multiple statistical shape models, built upon clustered shape instances. The algorithm does not require tweaking of any parameters and the only input required by the algorithm is the image data along with the corresponding annotations. Experimental results on two different structures (liver and distal femur) demonstrated the potential of this method in dealing with segmentation problems in contrasting image modalities.

References

1. Vese, L., Chan, T.F.: A multiphase level set framework for image segmentation using the mumford and shah model. *International Journal of Computer Vision* 50(3), 271–293 (2002)
2. Ronfard, R.: Region-based strategies for active contour models. *International Journal of Computer Vision* 13(2), 229–251 (1994)
3. Huang, X., Metaxas, D., Chen, T.: Metamorphs: Deformable shape and texture models. In: *CVPR 2004*, pp. 496–503 (2004)
4. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. *Annual Review of Biomedical Engineering* 2, 315–338 (2000)
5. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing* 7(3), 359–369 (1998)
6. Zhan, Y., Shen, D.: Deformable segmentation of 3-d ultrasound prostate images using statistical texture matching method. *IEEE TMI* 25, 256–272 (2006)
7. Zheng, Y., Barbu, A., et al.: Four-chamber heart modeling and automatic segmentation for 3-d cardiac ct volumes using marginal space learning and steerable features. *IEEE Transactions on Medical Imaging* 27(11), 1668–1681 (2008)
8. Shen, D., Davatzikos, C.: An adaptive-focus deformable model using statistical and geometric information. *IEEE PAMI* 22(8), 906–913 (2000)
9. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* 315, 972–976 (2007)
10. Viola, P., Jones, M.J.: Robust real-time face detection. *International Journal of Computer Vision* 57, 137–154 (2004)
11. Cootes, T., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. *CVIU* 61(1), 38–59 (1995)

12. Shi, Y., Shen, D.: Hierarchical shape statistical model for segmentation of lung fields in chest radiographs. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 417–424. Springer, Heidelberg (2008)
13. Zhan, Y., Peng, Z., Zhou, X.: Towards organ-specific pet-ct interpretation: generic organ segmentation using joint pet-ct information. In: MICCAI 2008 Workshop on Analysis of Functional Medical Images (2008)
14. Eckhoff, D.G., Bach, J.M., et al.: Three-dimensional morphology and kinematics of the distal part of the femur viewed in virtual reality. *Journal of Bone and Joint Surgery* 85, 97–104 (2003)