EFFICIENT SPARSE SHAPE COMPOSITION WITH ITS APPLICATIONS IN BIOMEDICAL IMAGE ANALYSIS: AN OVERVIEW

Shaoting Zhang[1], *Yiqiang Zhan*[2], *Yan Zhou*[3], *Dimitris N. Metaxas*[1]

[1]Department of Computer Science, Rutgers University, Piscataway, NJ, USA
[2]CAD R&D, Siemens Healthcare, Malvern, PA USA
[3]CMS Software, Elekta Inc, MO, USA

ABSTRACT

Shape information plays an important role in biomedical image analysis because of the strong shape characteristics of biological structures. It is often used as a prior to constrain or refine the intermediate shape information derived from low-level image features. In this paper, we give an overview of the sparse shape composition based prior modeling method and its various applications of biomedical image analysis. Instead of learning a generative shape model, it incorporates shape priors on-the-fly through the sparse shape composition. Particularly, a shape instance derived from low level image features is refined by a sparse linear combination of a sparse set of shapes in the repository. We also design three strategies to improve run-time efficiency. 1) When the shape repository contains a large number of instances, K-SVD can be used to learn a more compact but still informative shape dictionary. 2) If the derived shape instance has a large number of vertices, which often appears in 3D problems, affinity propagation method can be used to partition the surface into small sub-regions. 3) When there are multiple structures, hierarchical scheme is employed to model them simultaneously. These strategies decrease the scale of the sparse optimization problem and thus speed up the algorithm. Our method is applied on different biomedical image analysis problems, including localization, tracking and segmentation of anatomical structures. In all of these applications, this method achieves promising results.

Index Terms— Shape prior, sparse representation, dictionary learning, segmentation, tracking

1. INTRODUCTION

In many clinical practices organ shape plays an important role, which is usually derived from medical images using low level appearance cues. However, due to diseases and imaging artifacts, low level appearance cues are often weak or misleading. In this situation, shape priors become critical to infer and refine the shape derived from image appearances. Effective modeling of shape priors is challenging because: 1) shape variations are complex and cannot always be modeled by parametric probability distributions; 2) a shape instance derived from image appearance cues may have significant errors; and 3) local details of an input shape may be important for clinical purposes but difficult to preserve if they are not statistically significant in the training data.

The most popular shape prior based method, perhaps, is Active Shape Model (ASM) [3], whose shape prior module is from using principal component analysis (PCA). Many variations have been proposed to improve this statistical shape model [6] by solving three these challenges, *i.e.*, handling gross errors [7], modeling complex shape variations, and preserving local shape details [9]. However, these methods usually only focus on alleviating one specific problem.

To address these challenges in a unified framework, sparse shape representation has been proposed [15][1]. A sparse set of shapes is selected from the shape repository and composed together to infer and refine an input shape. This way, the prior information is implicitly incorporated on-the-fly. This model leverages two sparsity observations of the input shape instance: 1) the input shape can be approximately represented by a sparse linear combination of shapes in the shape repository; 2) parts of the input shape may contain large errors but such errors are sparse. This model is formulated as a sparse learning problem. Using $L1$ norm relaxation, it can be solved by an efficient expectation-maximization (EM) framework. Furthermore, this model can be extended to effectively handle multi-resolution, local shape priors [14], large scale training dataset [1], and hierarchical priors [13].

This shape prior model has been applied to solve several diverse medical applications, including 2D lung localization in chest X-ray images, 3D liver segmentation in low-dose Computed Tomography (CT) scans, 3D segmentation of multiple rodent brain structures in Magnetic Resonance (MR) microscope, and real time tracking of left ventricles in Magnetic Resonance Imaging (MRI). Compared to state-of-the-art methods, our model exhibits better performance in all these studies.

[1]The concept of sparse representation has been investigated in computer vision community, such as, but not limited to, face recognition [11], super resolution [12], and image bias field estimation [16].

2. SPARSE SHAPE REPRESENTATION

Formulation: We use explicit parametric shape representation to model a shape instance (i.e., a curve (2D) or a triangular mesh (3D) consisting of a set of vertices). To describe the ith shape in the training data, the coordinates of all its vertices are concatenated into a vector $d_i \in \mathbf{R}^n$, where n is the product of the number of vertices in each shape by the dimension. Thus the training repository can be represented as a matrix $D = [d_1, d_2, ..., d_k] \in \mathbf{R}^{n \times k}$, where k is the number of shapes. In our framework, all $d_i, i = 1, 2, 3, ..., k$ are pre-aligned using generalized Procrustes analysis [5]. $y \in \mathbf{R}^n$ is the vector of a newly-input shape which needs to be constrained or refined. Our sparse shape representation assumes that after proper alignment, any input shape y can be approximately represented as a sparse linear combination of existing data $d_i, i = 1, 2, 3, ..., k$ and sparse gross errors $e \in \mathbf{R}^n$. We denote $x = [x_1, x_2, ..., x_k]^T \in \mathbf{R}^k$ as the coefficients or weights. Thus the value of x for this sparse linear combination is found by minimizing the following loss function:

$$\begin{aligned} &\arg\min_{x,e,\beta} \{\|T(y,\beta) - Dx - e\|_2^2\}, \qquad (1) \\ &s.t.\ \|x\|_0 \le k_1, \|e\|_0 \le k_2 \end{aligned}$$

where $T(y, \beta)$ is a global transformation operator with parameter β, which aligns the input shape y to the mean shape of existing data D, $\|\cdot\|_0$ is the $L0$ norm counting the nonzero entries of a vector, k_1 is the pre-defined sparsity number and k_2 is the sparsity number of e. e captures sparse but large errors which are caused by occlusion or point missing. When there is no such error, the $L2$ norm loss function can deal with it well and e will be all zeros. Thus e is a good supplement which specifically handles non-Gaussian and sparse errors. After obtaining x, the input shape y is constrained or refined as Dx and transformed back by the inverse of the transformation matrix using parameter β.

Optimization: Due to the nonconvexity of $L0$ norm, the constraints in Eq. 1 are not directly tractable. Fortunately, recent developments in sparse representation provide a theorem to efficiently solve this kind of problems through $L1$ norm relaxation [10]. Thus Eq. 1 is reformulated as:

$$\arg\min_{x,e,\beta} \{\|T(y,\beta) - Dx - e\|_2^2 + \lambda_1\|x\|_1 + \lambda_2\|e\|_1\}, \quad (2)$$

where λ_1 and λ_2 control how sparse x and e are, respectively. After relaxation, $\lambda_1\|x\|_1 + \lambda_2\|e\|_1$ is non-smooth but continuous and convex.

To solve Eq. 2, we need to simultaneously recover the alignment parameter β and error e. It is a typical Chicken-and-Egg problem. Furthermore, to efficiently optimize Eq. 2, we need to deal with the nonlinearity of $T(y, \beta)$ if the transformation is rigid or a similarity. We use EM types of algorithms (or alternating minimization). Eq. 2 is divided into two sub-problems: 1) estimating β and computing $T(y, \beta)$, 2) efficiently minimizing this simplified linear inverse problem. In the "E" step, β is estimated using Procrustes analysis, which aligns the shape y to the mean shape. Then vector $y' = T(y, \beta)$ is obtained. In the "M" step, the following simplified problem is minimized:

$$\arg\min_{x,e} \{\|y' - Dx - e\|_2^2 + \lambda_1\|x\|_1 + \lambda_2\|e\|_1\}, \quad (3)$$

which is now a linear inverse problem. It is then efficiently solved using existing methods such as the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) [2]. Two procedures are iteratively employed to obtain x, e and β. Then Dx is computed as the approximated shape and is transformed to its original coordinate system.

Dictionary Learning: In many applications of medical image analysis, the number of training data is usually less or around hundreds. In such cases, we can just put all data into the matrix D, since Eq. 2 can be still efficiently solved. However, if there are thousands or millions training data, it may not be feasible to select models fast. Thus, dictionary learning technique is employed to compute a compact codebook (i.e., dictionary), which can well represent the original database.

We use K-SVD [1] to achieve this. It includes two steps, sparse coding and codebook update. Sparse coding is a greedy method which can approximate an input data by finding a sparse set of elements from the codebook. Codebook update is to generate a better dictionary given sparse coding results. Two steps are alternately employed until converge. The resulting codebook can have much less number of elements (i.e., less columns) while is still generative enough to represent input data. Once the dictionary is learned from this procedure, we use it directly in the sparse shape model. Compared to the whole database, the compact dictionary reduces the redundancy of the shape repository. K-SVD algorithm minimizes the reconstruction errors, and this compact D can highly improve the computational efficiency.

Local Shape Prior: Another challenge is that many 3D deformable models include many thousands of points to give an accurate description of organ shapes. The optimization of Eq. 3 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 use 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. The reason we also need to consider image information is that organ boundaries may have heterogeneous appearance. Thus such partitioning also benefits the perfor-

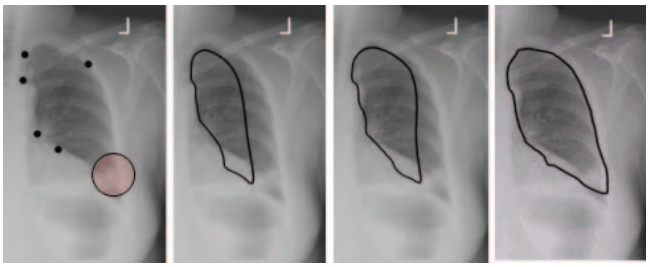


Fig. 1. From left to right: detected landmarks; shape fitting using procrustes analysis; ASM; sparse shape representation.

mance of boundary detectors if we train different detectors for each local shape patch.

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. This setting highly improves the computational efficiency of solving Eq. 3.

Hierarchical Shape Prior: In some applications, we need to effectively model multiple structures simultaneously. Thus hierarchical shape prior is employed. First, shape prior module is applied on each structure separately as the traditional way. Thus, shape variation of individual structure can be well discovered even with limited number of samples. Second, their relative locations with respect to the mass centroid are also modeled using the same prior module. This global statistics is used to put high-level constraints, such as placing structures and preventing intersection among structures.

3. MEDICAL APPLICATIONS

2D Lung Localization: Radiography (X-ray) is the most frequently used medical imaging modality due to its fast imaging speed and low cost. In this application, we aim to roughly locate the left or right lung, which can be further used as an input of computer aided diagnostic programs. We first detect several landmarks using learning based methods, and then infer a shape to fit these landmarks using sparse shape representation.

The major challenges of lung segmentation in radiography come from large variations of lung shapes, lung disease and pseudo-boundary close to diaphragm. Fig. 1 shows a case of complex shape variations. There is one point missing (marked by a circle), and this lung has a very special shape, which is not captured by the mean shape or its variations. The sparse shape representation is still able to roughly locate it.

2D Left Ventricle Tracking: Extraction of the boundary contour of a beating heart from cardiac MRI image sequences plays an important role in cardiac disease diagnosis and treatments. The algorithm should be robust, accurate and fast. We use a bottom-up approach to solve the contour tracking problem. On the bottom level, a collaborative trackers network provide a deformed mesh which is able to predict an initialized contour at each time step [17]. Each tracker in the network is a particle filter tracker featuring the appearance of the tracking region as the template. The coordination of the trackers in the network is modeled by Bayesian network, which models spatio-temporal dependency relationships of multiple trackers. On the top level, the shape prior module is employed to refine the initialized contour provided by the collaborative trackers network, and eventually obtains a smoothed contour of the left endocardium over time. Fig. 2 shows the tracking framework.

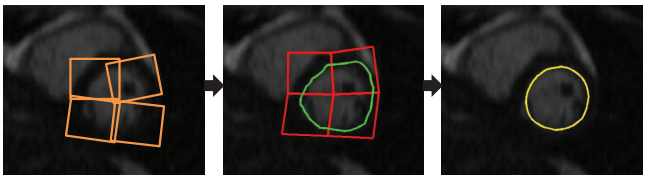


Fig. 2. Tracking framework using collaborative tracking and shape refinement.

3D Liver Segmentation: Whole body PET-CT is an emerging medical imaging modality that combines a Positron Emission Tomography (PET) and an x-ray Computed Tomography (CT) scan. The co-registered anatomical (CT) and functional (PET) information benefits various clinical practices, especially for oncology studies. Due to the high variations of F-fluorodeoxyglucose (FDG) uptakes across different organs, the preferred way to interpret PET-CT images is in an organ-specific fashion, which requires organ segmentation. In this application, we try to segment the liver from the low-dose whole body CT, using deformable models and shape refinement. AFDM method [8] is employed to obtain one-to-one correspondence among training shapes. The initialization step is based on landmark detection and similar to the 2D lung localization step. A surface model is then fitted into the image, and is hierarchically deformed to the image gradient information. During the deformation procedure, our sparsity-based shape prior modeling is used to refine each shape partition individually, and then these partitions are merged together. Compared to applying shape prior method on the whole shape, this shape partitioning approach can significantly improve the computational efficiency. This shape prior step is actually a regularization step to avoid the deformable model getting stuck in local minima of the image information. This method is fully automatic without any manual interaction. Fig. 3 shows one comparison of shape refinement methods.

3D Rodent Brain Structure Segmentation: Magnetic resonance imaging (MRI) at a spatial resolution of at least 100um in one dimension is frequently referred to as MR microscopy (MRM) and is currently available with the use of high magnetic field images. This technical achievement has

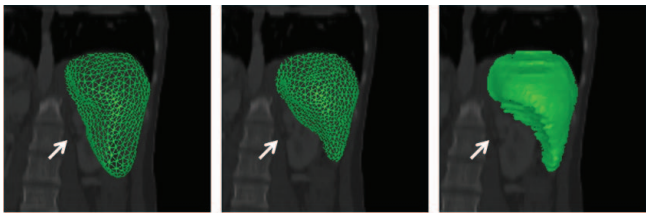


Fig. 3. Comparison of shape refinement for liver segmentation. From left to right: ASM, sparse shape representation, and ground truth. ASM result includes a large part of kidney. The proposed method is relatively accurate.

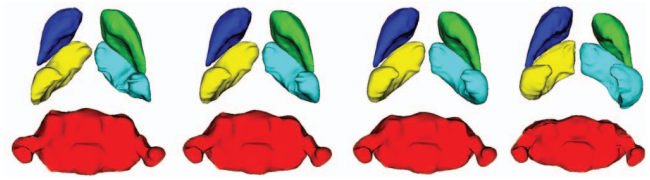


Fig. 4. Hierarchical shape prior model of multiple rodent brain structures.

permitted the detailed anatomical study of the rodent brain, which is much smaller than the human brain, and requires a small voxel size, in order to be imaged. Rodents are often used as models of human disease not only because they frequently exhibit key features of abnormal neurological conditions but also because they are a convenient starting point for novel studies. In this application, we aim to segment multiple rodent brain structures simultaneously. The challenge is to effectively model shape prior given a limited number of training data. We use hierarchical shape priors to achieve this. Fig. 4 shows the shape modeling results for multiple structures.

4. CONCLUSION AND FUTURE WORK

In this article we give an overview of sparse shape composition based shape prior modeling. It is able to model complex shape variations, handle gross errors, and preserve shape details in a unified framework. We also extend this method by using dictionary learning, mesh partitioning, and hierarchical shape priors. This method is applied in several diverse medical applications, in both 2D and 3D. In the future, we would like to investigate online dictionary learning, and apply this method to more applications including shape registration.

5. REFERENCES

[1] M. Aharon, M. Elad, , and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. In *TSP*, pages 4311–4322, 2006.

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[3] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape model - their training and application. *CVIU*, 61:38–59, 1995.

[4] B. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972, 2007.

[5] C. Goodall. Procrustes methods in the statistical analysis of shape. *JRSS*, 53:285–339, 1991.

[6] T. Heimann and H.-P. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *MedIA*, 13(4):543 – 563, 2009.

[7] M. Rogers and J. Graham. Robust active shape model search. In *ECCV*, pages 517–530, 2002.

[8] D. Shen and C. Davatzikos. An adaptive-focus deformable model using statistical and geometric information. *TPAMI*, 22(8):906 –913, 2000.

[9] K. Sjostrand and et.al. Sparse decomposition and modeling of anatomical shape variation. *TMI*, 26(12):1625 –1635, 2007.

[10] J. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *TIP*, 14:1570–1582, 2004.

[11] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *TPAMI*, 31(2):210 –227, 2009.

[12] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *TIP*, 19, 2010.

[13] S. Zhang, J. Huang, M. Uzunbas, T. Shen, F. Delis, X. Huang, N. Volkow, P. Thanos, and D. N. Metaxas. 3D segmentation of rodent brain structures using hierarchical shape priors and deformable models. In *MICCAI*, pages 611–618, 2011.

[14] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. Metaxas, and X. Zhou. Deformable segmentation via sparse shape representation. In *MICCAI*, pages 451–458, 2011.

[15] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou. Towards robust and effective shape modeling: Sparse shape composition. *MedIA*, 16(1):265 – 277, 2012.

[16] Y. Zheng and J. Gee. Estimation of image bias field with sparsity constraints. In *CVPR*, pages 255–262, 2010.

[17] Y. Zhou, E. Yeniaras, P. Tsiamyrtzis, N. Tsekos, and I. Pavlidis. Collaborative tracking for MRI-guided robotic intervention on the beating heart. *MICCAI*, pages 351–358, 2010.