

Sparse Shape Composition: A New Framework for Shape Prior Modeling

Shaoting Zhang^{1,2}, Yiqiang Zhan¹, Maneesh Dewan¹, Junzhou Huang²,
Dimitris N. Metaxas², and Xiang Sean Zhou¹

¹CAD R&D, Siemens Healthcare, Malvern, PA, USA

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

{shaoting, jzhuang, dnm}@cs.rutgers.edu

{yiqiang.zhan, Maneesh.Dewan, xiang.zhou}@siemens.com

Abstract

Image appearance cues are often used to derive object shapes, which is usually one of the key steps of image understanding tasks. However, when image appearance cues are weak or misleading, shape priors become critical to infer and refine the shape derived by these appearance cues. Effective modeling of shape priors is challenging because: 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. In this paper we propose a novel Sparse Shape Composition model (SSC) to deal with these three challenges in a unified framework. In our method, training shapes are adaptively composed to infer/refine an input shape. The a-priori information is thus implicitly incorporated on-the-fly. Our model leverages two sparsity observations of the input shape instance: 1) the input shape can be approximately represented by a sparse linear combination of training shapes; 2) parts of the input shape may contain gross errors but such errors are usually sparse. Using L_1 norm relaxation, our model is formulated as a convex optimization problem, which is solved by an efficient alternating minimization framework. Our method is extensively validated on two real world medical applications, 2D lung localization in X-ray images and 3D liver segmentation in low-dose CT scans. Compared to state-of-the-art methods, our model exhibits better performance in both studies.

1. Introduction

Shape is a distinctive characteristic of many objects and plays a critical role in various computer vision tasks used for image interpretation. Here, image appearance cues pro-

vide low-level evidence to derive object shapes. However, the derived shape instances may be *incomplete* in the presence of weak (missing) appearance cues, or *incorrect* when misleading appearance cues are present. Shape models are thereby designed to *infer* and *refine* the object shape in an optimal sense using high-level priors. The success of these models is highly dependent on the way shape priors are modeled and on the optimization method used. One of the seminal work in this area, “Snake” [14], models the shape prior as a general regularity term in the optimization, which assumes that the shape should deform like a membrane or a thin plate. Subsequently, more object-specific shape priors became prevalent, where the shape priors are learned from a set of training samples, such as Active Shape Model (ASM) [4]. Many adaptations of these algorithms have been proposed over the years (see Sec. 1.1 for details).

Shape prior based models confront three major challenges. *First*, shape variations are usually complex and therefore difficult to model using a parametric probability distribution. *Second*, image appearance information can be highly misleading and non-Gaussian errors frequently appear in the input shape. Shape models have to be robust to handle these errors. *Third*, shape models should be adaptive in order to preserve local detail information in the input shape, provided such details are present in the training data, even if they are not statistically significant. While several methods have been proposed to address one or two of those challenges, it remains that, to the best of our knowledge, none of them can deal with all three simultaneously.

In this paper, we propose a novel Sparse Shape Composition model (SSC), to address the above-mentioned challenges in a unified framework. Instead of explicitly learning shape priors from a set of training samples offline, we propose to adaptively approximate the input shape on-the-fly, by a sparse linear combination of training shapes. Hence, the shape prior constraint is implicitly applied. Our method is inspired by recently proposed sparsity theories in the compressive sensing community, *i.e.*, the problem of

computing sparse linear representations with respect to an overcomplete dictionary of base elements [2]. It has been successfully applied in many computer vision applications, such as, but not limited to, face recognition [26], background subtraction [13] and image annotation [27]. Yet, to our knowledge, such techniques have not been used in the context of shape priors.

There are two “sparsity” observations behind our method. *First*, given a large enough training dataset (dictionary), an instance can be approximately represented by a sparse linear combination of instances in the dictionary. Similarly, in our application each given shape is approximated by a sparse linear combination of training shapes. Without any assumption of a parametric distribution model (*e.g.*, a unimodal distribution assumption in ASM), it becomes general to objects whose shape statistics can be very complex. Moreover, such a setting is able to recover detail information even if the detail of the input shape is only present in a small number of training data and is not statistically significant. *Second*, the given shape information may contain gross errors, but such errors are often very sparse, *e.g.*, there is an object occluded in the image or a point missing in the input shape. Combining these two, we formulate the shape prior task as an optimization problem, and efficiently solve it by an alternating minimization framework. Furthermore, we explicitly model the nonlinear shape transformation in the optimization framework without assuming the misalignment is small. In our experiments, the proposed method shows improved accuracy and robustness compared to some widely used approaches. Note that although we validate our method on two applications of medical image analysis, the proposed SSC is actually general and can be directly extended to other applications.

The main contributions of our work are threefold: 1) SSC is proposed to model shapes and implicitly incorporate the shape prior constraint effectively. It is based on sparse representations and our unified framework is able to handle non-Gaussian errors, multimodal distributions of shapes and detail information recovery. 2) The problem is efficiently solved by an alternating minimization framework. 3) It is successfully applied to two medical applications, 2D lung localization from X-ray image and 3D liver segmentation from low-dose CT. The extensive experiments demonstrate the superior performance of our method.

1.1. Relevant Work

As discussed earlier, many shape prior based approaches have been proposed in different contexts to incorporate the shape prior constraint. In the context of medical image analysis, ASM [4] and its variations [12] are probably the most widely used approaches. Many methods have been proposed to improve the shape prior module of ASM. They mainly focus on three aspects:

Modeling complex shape variations. A significant effort has been put on handling multimodal distribution of shapes, which cannot be represented by their mean shape and variations. A classical solution is to use a mixture of Gaussians to represent shape variation [5]. Manifold learning techniques can also be used to learn a non-linear shape prior to alleviate this problem [7]. However, it is possible that shape variation is too complex to model with a parametric probability distribution. Some notable work uses patient-specific shape statistics [23] or subject-specific dynamical model [28] to constrain the deformable contours. Thus each data has a particular shape even if it cannot be approximated by the mean shape and its variations. It is also worth mentioning that there are some successful work in the face alignment field that attacks similar problems, such as the multi-level generative model [11] and subspace constrained mean-shift [21]. Shape inference is also a potential solution. It constructs a surface from a set of 3D points. In [9], to infer the shape, a nearest-neighbor approach is used by finding the closest instance in a database, and the database is based on the expert’s structure annotations.

Handling non-Gaussian errors. A large number of proposed modifications on the original ASM algorithm tries to improve the stability against outliers. Some methods have tried to identify and correct outliers. For example, Lekadir *et al.* [15] employ a shape metric based on the ratio of landmark distances to detect outliers. Other methods try to decrease outliers’ influence using the weighting of residuals. Rogers and Graham [20] evaluate the use of M-estimators, image match and random sample consensus (RANSAC) [8] for this purpose. In a concluding evaluation, RANSAC was the most effective of these three methods. Nahed *et al.* [17] proposed to use a robust point matching algorithm which rejects outliers and finds the best-fitting model.

Preserving Local detail information. Another difficulty is to preserve local details of the input shape when such details are also present in the training data but not statistically significant. PCA performs eigen-analysis and extracts eigenvectors with the largest eigenvalues. The discarded eigenvectors are statistically insignificant, but they may contain important local details. Some relevant work can alleviate this problem. Sparse PCA [24] obtains sparser modes and produces near-orthogonal components. Thus each mode only affects locally clustered landmarks and captures more detail information. Some other methods divide the shape model into several independently modeled parts, such as the hierarchical approach [6]. Since the smaller parts exhibit less variation, they can be captured with fewer training samples than the variations for the full shape.

However, most discussed methods focus on solving one or two limitations. It is not trivial to handle all of them simultaneously. In our work, we address these challenges in a unified framework as outlined in the next sections.

2. Sparse Shape Composition Model

2.1. Problem Formulation

Notations and basic framework: In this study, we want to model the shape of an object using existing training data. Explicit parametric shape representation is employed here, *i.e.*, a curve (2D) or a triangular mesh (3D) consisting of a set of vertices. To describe the i th shape in the training data, the coordinates of all its vertices are concatenated into a vector $d_i \in \mathbb{R}^n$, where n is the product of the number of vertices in each shape by the dimension. Thus the whole training data can be represented as a matrix $D = [d_1, d_2, \dots, d_k] \in \mathbb{R}^{n \times k}$, where k is the number of shapes. In our framework, all $d_i, i = 1, 2, 3, \dots, k$ are pre-aligned using generalized Procrustes analysis [10]. $y \in \mathbb{R}^n$ is the vector of a newly-input shape which needs to be constrained or refined. Our basic framework assumes that after proper alignment, any input shape y can be approximately represented as a weighted linear combination of existing data $d_i, i = 1, 2, 3, \dots, k$, and the parts which cannot be approximated are noises. We denote $x = [x_1, x_2, \dots, x_k]^T \in \mathbb{R}^k$ as the coefficients or weights. Thus the value of x for the linear combination is found by minimizing the following loss function:

$$\arg \min_{x, \beta} \|T(y, \beta) - Dx\|_2^2, \quad (1)$$

where $T(y, \beta)$ is a global transformation operator with parameter β . It aligns the input shape y to the mean shape of existing data D . x and β are computed by solving (1). Thus the input shape y is constrained or refined as Dx and transformed back by the inverse of the transformation matrix using parameter β .

Sparse linear combination: The limitations of (1) are twofold. First the data matrix D may be overcomplete ($k > n$) when the number of models is larger than the length of d_i . Thus the system may not have a unique solution. More constraints of the coefficient x are needed. Second, the input shape, including the noises, may be perfectly represented if any linear combination can be used. A more appropriate assumption is that the input shape can be approximately represented by a *sparse* linear combination of existing data. This way, the problem is reformulated as:

$$\begin{aligned} \arg \min_{x, \beta} \|T(y, \beta) - Dx\|_2^2, \\ \text{s.t. } \|x\|_0 \leq k_1 \end{aligned} \quad (2)$$

where $\|\cdot\|_0$ is the L^0 norm counting the nonzero entries of a vector, k_1 is the pre-defined sparsity number. Such formulation ensures that the number of nonzero elements in x is smaller than k_1 . [16] proposed a cost function which is in spirit similar to (2).

Non-Gaussian errors: The formulation (2) works well for many scenarios. However, there is still one limitation in (2). Since the loss function is based on L_2 norm, it assumes that the error model follows a Gaussian distribution. Thus it is sensitive to large noises or gross errors of the input shape, caused by image occlusion or points missing. Such problem happens frequently in many applications. In these cases, some errors can be very large, but they are relatively sparse compared to the whole data. To alleviate this problem, we explicitly model the error as a sparse vector $e \in \mathbb{R}^n$ by reformulating the problem as:

$$\begin{aligned} \arg \min_{x, e, \beta} \|T(y, \beta) - Dx - e\|_2^2, \\ \text{s.t. } \|x\|_0 \leq k_1, \|e\|_0 \leq k_2 \end{aligned} \quad (3)$$

where k_2 is the sparsity number of e . When solving (3), e captures sparse but large errors which are caused by occlusion or point missing. When there is no such error, the L_2 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. Note that unlike the formulation in the robust face recognition [26], we do not assume that the misalignment is small and thus explicitly model the transformation with parameter β in (3).

Convex relaxation: The constraints in (3) are not directly tractable because of the nonconvexity of L^0 norm. Greedy algorithms can be applied to this NP-hard L^0 norm minimization problem, but there is no guarantee to capture the global minima. In the general case, no known procedure can correctly find the sparsest solution more efficiently than exhausting all subsets of the entries for x and e . Furthermore, in practice the sparsity numbers k_1 and k_2 may change for different data in the same application. For example, some data have errors while others do not. Fortunately, recent developments in sparse representation provide a theorem to efficiently solve this kind of problems through L_1 norm relaxation [25]. Thus (2) is reformulated as:

$$\arg \min_{x, \beta} \|T(y, \beta) - Dx\|_2^2 + \lambda_1 \|x\|_1, \quad (4)$$

and is named as SSC(4). Similarly, (3) is reformulated as:

$$\arg \min_{x, e, \beta} \|T(y, \beta) - Dx - e\|_2^2 + \lambda_1 \|x\|_1 + \lambda_2 \|e\|_1, \quad (5)$$

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. (5) is our objective function of our proposed SSC.

Connections to other methods: It is interesting to look into (5) by adjusting λ_1 and λ_2 into some extreme values.

- If λ_2 is extremely large, e will be all zeros. Thus SSC is similar to methods which do not model non-Gaussian errors.

- If both λ_1 and λ_2 are large enough, e will be all zeros and x may have only one nonzero element. Thus SSC becomes the nearest neighbor method.
- If λ_2 is extremely large and λ_1 is small, a dense linear combination of shapes is used, which is able to perfectly approximate the transformed input shape. Thus SSC degenerates to the Procrustes analysis.

The insight of (5) indeed reveals the connections of our SSC with some other popular methods. Those methods can be regarded as special cases of SSC. In other words, SSC provides a unified framework to deal with different challenges of shape prior modeling simultaneously. SSC can also provide flexibility to meet the requirements of different applications by adjusting the sparsity of x and e .

2.2. Optimization Framework

To solve (5), we need to simultaneously recover the alignment parameter β and error e . It is a typical Chicken-and-Egg problem. Furthermore, to efficiently optimize (5), we need to deal with the nonlinearity of $T(y, \beta)$ if the transformation is rigid or a similarity. A notable approach is to use iterative linearization and optimize all variables simultaneously, which was proposed and successfully applied in image alignment [18]. However, this algorithm assumes that the initial misalignment is not too large, which may not be held in our problem. Furthermore, it focuses on rigid transformation in 2D images, while we deal with nonrigid transformation in arbitrary dimensions for shapes. The efficiency of the optimization framework is also important.

Our solution is to use alternating minimization. (5) is divided into two sub-problems: 1) estimating β and computing $T(y, \beta)$, 2) efficiently minimizing this simplified linear inverse problem. In the first 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 second step, the following simplified problem is minimized:

$$\arg \min_{x, e} \|y' - Dx - e\|_2^2 + \lambda_1 \|x\|_1 + \lambda_2 \|e\|_1, \quad (6)$$

which is now a linear inverse problem. It can be optimized using existing solvers. 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.

3. SSC in Medical Image Analysis

Medical image analysis has become an important application area for computer vision algorithms. Different from general computer vision problems, medical images have their unique challenges and opportunities. Due to imaging artifacts and diseases, appearance cues in medical images might be unreliable or misleading. On the other hand,

however, strong shape priors of human anatomy provide opportunities to shape prior-based methods. To evaluate the capability of SSC in shape prior modeling, we apply it to two medical image analysis tasks: organ localization using shape inference and segmentation using shape refinement.

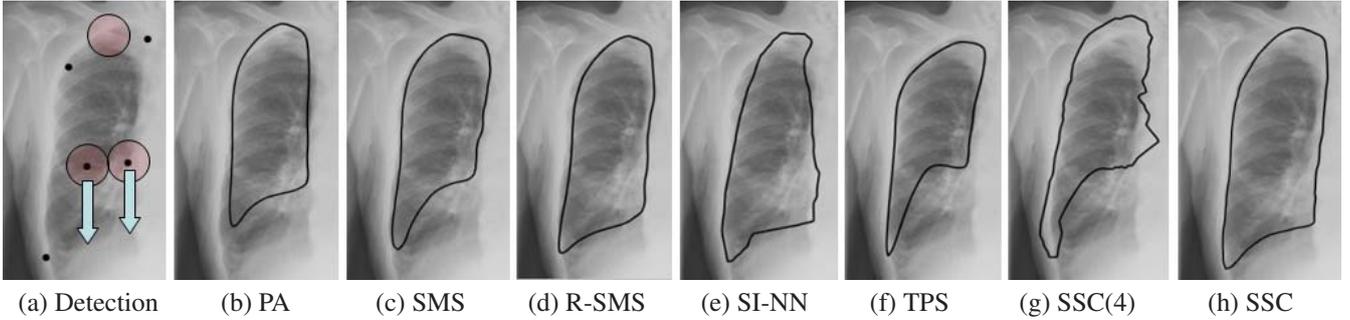
Organ localization based on landmark detection and shape inference: The positions and orientations of the same organ vary significantly in medical image data. Quickly and accurately locating the organ is crucial to image segmentation. One approach is to find a similarity transformation matrix, and then use this matrix to align a mean shape to the organ. Generally it achieves good performance. However, similarity transform only has nine degrees of freedom. Thus it may not be able to represent some specific data or shapes by transforming a mean shape (in Sec. 4.1).

To solve this problem, we propose a landmark detection and shape inference based localization method. A learning-based method is employed for landmark detection. Detected landmarks can be very sparse compared to the whole shape. Furthermore, there may be gross errors or point missing from the detection results. We use SSC to infer a shape to fit these detected landmarks. Compared to solely using similarity transformation to transform the mean shape, our approach has more degrees of freedom and is able to better fit to the actual shape in the image.

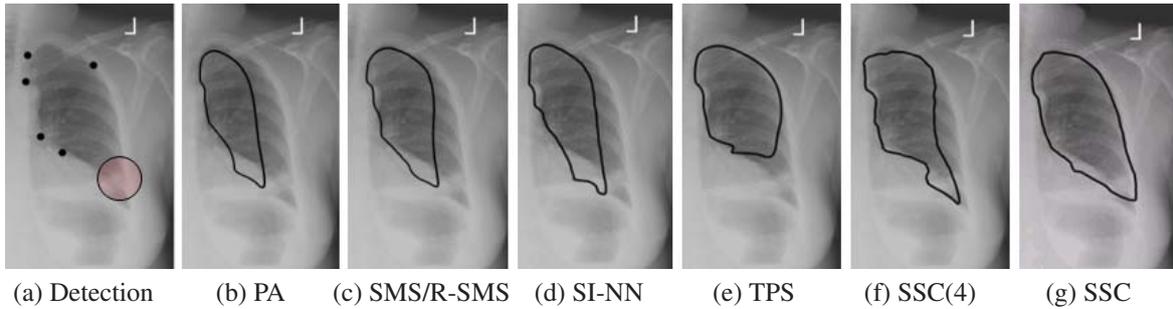
In this framework, we assume that ground truths for shapes are available from training data, and the one-to-one correspondence is already obtained. We then automatically or manually choose some specific points (*e.g.*, corner points or high curvature points) as landmarks on the shape of each data. Such training landmarks and shapes are fed into data matrices denoted as D_L and D_S , respectively. Then a landmark detector is trained using image data and annotated landmarks. Given a testing image, its landmarks y_L are estimated using the detector. Then x and β is computed by optimizing (5) with D_L and y_L . Finally $D_S x$ is computed as the refined shape and transformed back to its coordinate system using inverse of the transformation matrix with parameter β . Such localization can also be used as the initialization of many segmentation algorithms.

In Sec. 4.1, this framework is employed to locate 2D lung from X-ray image. In Sec. 4.2, the result of this framework is used as 3D initialization for segmentation algorithms.

Organ segmentation based on deformable model and shape refinement: Deformable model-based organ segmentation is very popular. However, deformable models may get trapped in local minima when the image contains noise. Thus high level shape information may be a valuable complement. Many methods have been proposed to incorporate shape prior information and achieved good performance. We propose an SSC based shape refinement method as a regularization step during deformation. An initialized shape is deformed following the image gradient informa-



(a) Detection (b) PA (c) SMS (d) R-SMS (e) SI-NN (f) TPS (g) SSC(4) (h) SSC
 Figure 1. Comparisons of the right lung localization. (a) Detected landmarks are marked as black dots. There are two detection errors and one point missing (marked as circles, and the arrows point to the proper positions). (b) Similarity transformation from Procrustes Analysis. (c) Active Shape Model. (d) Robust Active Shape Model with RANSAC. (e) Shape inference method using nearest neighbors. (f) Thin-plate-spline. (g) Sparse representation without modeling e , by solving (4). (h) The proposed method by solving (4).



(a) Detection (b) PA (c) SMS/R-SMS (d) SI-NN (e) TPS (f) SSC(4) (g) SSC
 Figure 2. Comparisons of the left lung localization. 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. Compared methods are the same as Fig. 1.

tion. During the deformation procedure, the shape refinement is employed as a high level constrain to avoid getting stuck in local minima of the image information. Denote the training shape matrix as D_S , and the intermediate deformation result as y_S . Then x is computed by solving (5) with D_S and y_S . $D_S x$ is used as the refined shape and transformed back. In this refinement procedure, e may not have large values since the model is already roughly aligned after initialization. However, modeling e is still necessary to capture small errors not following Gaussian distribution. In Sec. 4.2, this framework is used to segment 3D liver.

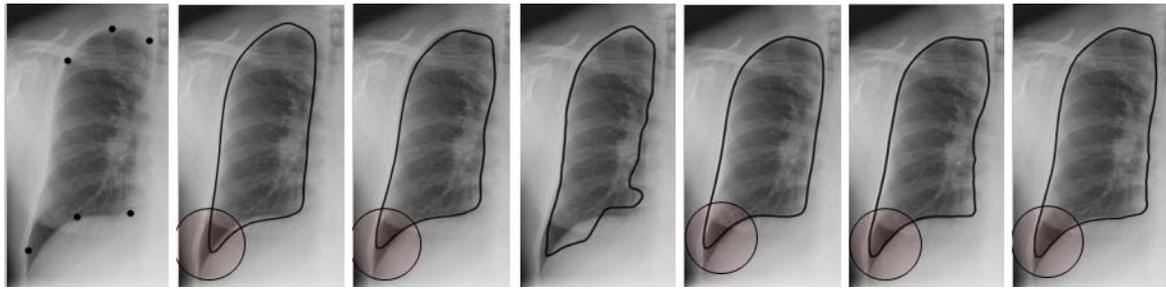
4. Experiments

4.1. 2D Lung Localization from X-ray Image

Background and experimental setting: Radiography (X-ray) is the most frequently used medical imaging modality due to its fast imaging speed and low cost. About one third of radiograph exams are chest radiographs. It is used to reveal various pathologies including abnormal cardiac sizes, pneumonia shadow and mass lesions. In chest X-ray, the position, size and shape of lungs often provide important clinical information. Therefore, in this experiment we try to locate the left or right lung using landmark detection and SSC based shape inference. Out of 367 X-ray images (all images are from different patients), 200 are used

as training data, and the rest 167 are used for testing purpose. The ground truths are binary masks of segmentation results. A 2D contour is extracted from each mask. To obtain the landmarks for training purpose, we manually select six specific points (e.g., corner points) on the contour, and then evenly and automatically interpolate a fixed amount of points between two neighboring landmarks along the contour. Thus a rough one-to-one correspondence is obtained for both landmarks and shapes. Since the detected landmarks may not be accurate or complete, SSC is necessary to infer a shape from them. When applying this model, we constantly use the same parameter values for all X-ray images since we observed that the model is not sensitive to different samples in the same application. The experiments are performed on a PC with 2.4GHz Intel Quad CPU, 8GB memory, with Python 2.5 and C++ implementations. The whole framework is fully automatic.

In this study, we compare the proposed sparsity-based shape prior modeling with other state-of-art methods. For a fair comparison, we intentionally embed different shape models to the same organ localization framework. (It is not fair to compare completely different end-to-end systems, e.g., our system vs. ASM system, since the performance difference, if any, cannot be solely attributed to shape prior modeling.) Thus we use the same input from the landmark detection module, and validate different approaches



(a) Detection (b) PA (c) SMS/R-SMS (d) SI-NN (e) TPS (f) SSC(4) (g) SSC

Figure 3. Comparisons of the right lung localization. All six detections are roughly accurate. Thus there is no gross error. The regions marked by circles show the difference of captured details. Compared methods are the same as Fig. 1.

of shape inference. The following are compared methods:

- PA: Procrustes Analysis is used to find a similarity transformation to fit a mean shape to detected landmarks.
- SMS: It is the Shape Model Search module in ASM, which employs the PCA method to learn shape statistics and refine the input shape.¹
- R-SMS: The shape model search step in the robust ASM [20] method uses the RANSAC framework to remove the influence of outliers.
- SI-NN: It stands for shape inference using nearest neighbors. It is similar to [9], which uses nearest neighbors to find the closest prototypes in the expert's structure annotations. The distance metric we used is based on the L_2 distance between corresponding points.
- TPS: Thin-plate-spline [1] is used to deform the mean shape to fit detected landmarks. TPS is a non-rigid and local deformation technology and has been used in robust point matching application (TPS-RPM) [3].
- SSC(4): It is the sparse representation method without modeling e . The result is computed by solving (4).

Some representative and challenging cases are shown in Fig. 1, 2 and 3. In Fig. 1, there are some mis-detections which are considered as gross errors. The Procrustes analysis, SMS method, SI-NN algorithm and TPS cannot handle such cases. R-SMS is not sensitive to outliers and performs better. SSC(4) also fails to handle such non-Gaussian errors since e is not modeled. The proposed method SSC can successfully capture such mis-detected points in e and generate a reasonable shape. In Fig. 2, the underlying shape of the lung is special and different from most other lung shapes (see the mean shape in Fig. 2 (b)). Furthermore, there is a missing point. Neither a transformed mean shape or its

¹Since the detected landmarks are very sparse compared to the whole contour, directly fitting ASM to the small number of landmarks result in poor performance. To achieve reasonable performance, a contour is approximated by interpolating points in-between landmarks. When there is a point missing, the mean position of that point is used instead.

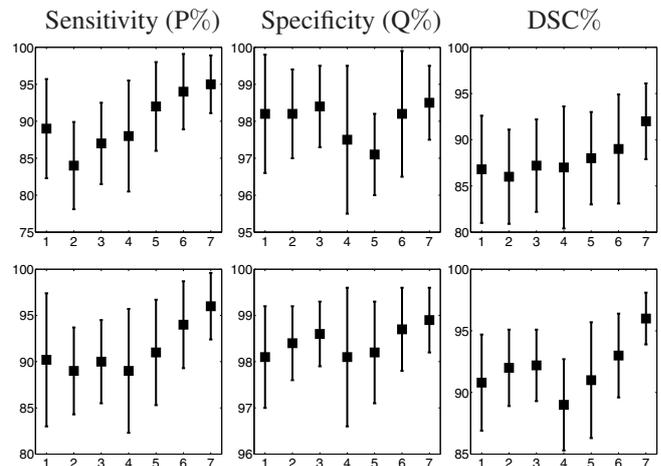


Figure 4. Mean values (μ) and standard deviations (σ) of P, Q and DSC from the left lung (1st row) and right lung (2nd row) of all testing data. In each figure, y-axis is the performance of P, Q or DSC. x-axis means 7 methods (from left to right) in the same order as Fig. 1. Squares denote μ , and the lengths of segments denote σ .

variations can represent such shape. TPS is very flexible and able to generate special shapes. However, it fails to handle the missing point. SSC roughly captures the correct shape and generates a better result than the others. In Fig. 3, all six detections are correct. However, the shape's details are not obtained using the mean shape or its variations. Both SSC(4) and SSC discover more detail information than other methods. Thus a sparse linear combination is sufficient to recover such details even the gross error e is not modeled.

To quantitatively compare different methods, we report the mean values and standard deviations of sensitivity and specificity between binary masks in Fig. 4. Note that the specificity is always good in all methods. The reason is that the size of either left or right lung is relatively small compared to the whole chest X-ray image. Hence, most "true negative" can be correctly found. Thus we also report Dice Similarity Coefficient (DSC) [19], which is a good

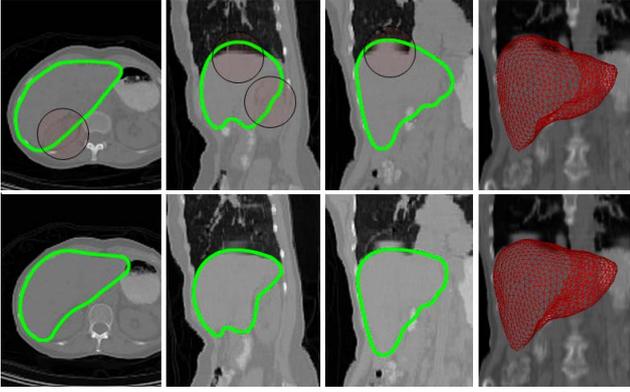


Figure 5. 3D initialization results for the segmentation framework. The first row is from the similarity transformation. The second row is from the proposed method (SSC). Results in the first row incorrectly include part of the kidney and the artifact caused by breath (marked by circles). The fourth column is the corresponding 3D view of the third column.

complement to the other two measurements. DSC is defined as: $2 \times TP / (2 \times TP + FP + FN)$, where TP , FP and FN stand for true positive, false positive and false negative, respectively. Generally Procrustes analysis, TPS and SMS achieve good performances, especially when landmarks are correctly detected. However, they are sensitive to non-Gaussian errors. R-SMS can handle this problem because of the RANSAC method, but sometimes it fails to deal with the multimodal distribution of shapes. SI-NN is a good nonparametric method. However, it may not be able to represent shapes which do not appear in the training data. The sparse linear combination by SSC(4) can approximate such shape and it generally performs better than the others. Without modeling error e , this method still fails to recover a correct shape. Our proposed method performs the best in terms of sensitivity and DSC, without sacrificing the specificity. The standard deviations in Fig. 4 show that SSC also achieves the best stability among all compared methods.

4.2. 3D Liver Segmentation from Low-dose CT

Background and experimental setting: 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. However, to decrease radiations to patients, CT images in PET-CT scans usually have low dose and large slice thickness, which result in low contrast and fuzzy boundaries between organs. Hence, organ segmentation in whole body PET-CT becomes more chal-

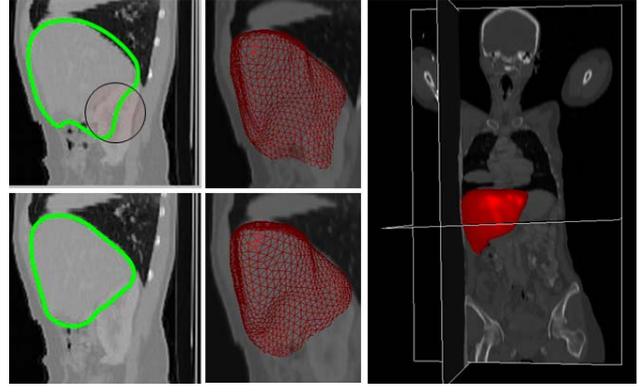


Figure 6. The refinement results after deformation. The first row is from SMS, and the second row is from SSC. Results of SMS incorrectly includes part of the kidney (marked by circles). The second column is the corresponding 3D view of the first column. The third column displays our result with the image data.

lenging than traditional CT. In this experiment we try to segment the liver from the low-dose whole body CT, using deformable models and SSC based shape refinement. The 3D ground truth of low-dose CT is manually segmented by multiple clinical experts. 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) [22]. The shape has around 1,000 vertices, and 20 are selected as landmarks. The initialization step is based on landmark detection and similar to Sec. 4.1. A surface model is then fitted into the image, and is deformed following the image gradient information. During the deformation procedure, SSC based shape refinement is employed as a regularization step to avoid getting stuck in local minima of the image information. We also test other methods to infer the shape for initialization and to refine the intermediate deformation result after several iterations as a regularization step. Note that all deformation modules are the same. The whole segmentation framework is fully automatic without any manual interaction.

Fig. 5 and 6 show some visual results in 3D. Fig. 5 compares the landmark detection based initialization. Since the image contrast of low-dose CT is very low, 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, which can benefit the deformation procedure. Fig. 6 compares the refinement results after deformation of a different case, starting from the same initialization. Note that the refined shape may not be exactly on the image boundary since this part is just a regularization step without considering any image information. The refined shape of ASM method follows the mean shape and the its variations, but it incorrectly includes a large part of the kidney. The shape from the proposed

method is more specific to this image and is more accurate.

5. Conclusions

In this paper, we proposed Sparse Shape Composition to implicitly model shape priors. It alleviates three problems in a unified framework, *i.e.*, modeling complex shape variations, handling non-Gaussian errors and preserve local detail information of the input shape. It is successfully applied to two important problems of medical image analysis. The experimental results demonstrate the high performance of our method.

In the future we would like to improve the algorithm in several directions. First, when there is a huge number of training samples, it may not be feasible to use all of them in the data matrix. It is desirable to learn a dictionary from the data. Second, the shape can be divided into parts, and the sparse representation is discovered for each part. Thus local features can be better preserved. Last but not the least, we plan to apply this shape model to more applications such as shape matching and registration.

Acknowledgement

The majority of the work was carried out when Shaoting Zhang was a summer intern at Siemens Healthcare, Malvern, PA, USA. We thank Dr. Bing Jian, Ting Chen (University of Florida), Yan Zhou (University of Houston) and Hongsheng Li (Lehigh University) for helpful discussions.

References

- [1] F. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *TPAMI*, 11(6), 1989.
- [2] E. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, 2006.
- [3] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *CVIU*, 89(2-3):114–141, 2003.
- [4] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape model - their training and application. *CVIU*, 61:38–59, 1995.
- [5] T. F. Cootes and C. J. Taylor. A mixture model for representing shape variation. In *IVC*, pages 110–119, 1997.
- [6] C. Davatzikos, X. Tao, and D. Shen. Hierarchical active shape models, using the wavelet transform. *TMI*, 22(3):414–423, 2003.
- [7] P. Etyngier, F. Segonne, and R. Keriven. Shape priors using manifold learning techniques. In *ICCV*, pages 1–8, 2007.
- [8] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.
- [9] B. Georgescu, X. Zhou, D. Comaniciu, and A. Gupta. Database-guided segmentation of anatomical structures with complex appearance. In *CVPR*, volume 2, pages 429–436, 2005.
- [10] C. Goodall. Procrustes methods in the statistical analysis of shape. *J. Roy. Statistical Society*, 53:285–339, 1991.
- [11] L. Gu and T. Kanade. A generative shape regularization model for robust face alignment. In *ECCV*, pages 413–426, 2008.
- [12] T. Heimann and H.-P. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Med. Image Anal.*, 13(4):543–563, 2009.
- [13] J. Huang, X. Huang, and D. Metaxas. Learning with dynamic group sparsity. In *ICCV*, pages 64–71, 2009.
- [14] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1:321–331, 1987.
- [15] K. Lekadir, R. Merrifield, and G. zhong Yang. G.-z.: Outlier detection and handling for robust 3D active shape models search. *TMI*, 26:212–222, 2007.
- [16] Y. Li and J. Feng. Sparse representation shape model. In *ICIP*, pages 2733–2736, 2010.
- [17] J. A. Nahed, M. pierre Jolly, and G. zhong Yang. Robust active shape models: A robust, generic and simple automatic segmentation tool. In *MICCAI*, 2006.
- [18] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment via sparse and low-rank decomposition for linearly correlated images. In *CVPR*. 2010.
- [19] A. Popovic, M. de la Fuente, M. Engelhardt, and K. Radermacher. Statistical validation metric for accuracy assessment in medical image segmentation. *Int J Comput Assist Radiol Surg*, 2:169–181, 2007.
- [20] M. Rogers and J. Graham. Robust active shape model search. In *ECCV*, pages 517–530, 2002.
- [21] J. Saragih, S. Lucey, and J. Cohn. Face alignment through subspace constrained mean-shifts. In *ICCV*, pages 1034–1041, 2009.
- [22] D. Shen and C. Davatzikos. An adaptive-focus deformable model using statistical and geometric information. *TPAMI*, 22(8):906–913, 2000.
- [23] Y. Shi, F. Qi, Z. Xue, L. Chen, K. Ito, H. Matsuo, and D. Shen. Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. *TMI*, 27(4):481–494, 2008.
- [24] K. Sjostrand and et.al. Sparse decomposition and modeling of anatomical shape variation. *TMI*, 26(12):1625–1635, 2007.
- [25] J. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *TIP*, 14:1570–1582, 2004.
- [26] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *TPAMI*, 31(2):210–227, 2009.
- [27] S. Zhang, J. Huang, Y. Huang, Y. Yu, H. Li, and D. N. Metaxas. Automatic image annotation using group sparsity. In *CVPR'10*, pages 3312–3319, 2010.
- [28] Y. Zhu, X. Papademetris, A. J. Sinusas, and J. S. Duncan. A dynamical shape prior for LV segmentation from RT3D echocardiography. In *MICCAI*, pages 206–213, 2009.