Medical Image Analysis 16 (2012) 1385–1396

Contents lists available at SciVerse ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Deformable segmentation via sparse representation and dictionary learning

Shaoting Zhang [a], Yiqiang Zhan [b,*], Dimitris N. Metaxas [a]

[a] Department of Computer Science, Rutgers University, Piscataway, NJ, USA
[b] Syngo R&D, Siemens Healthcare, Malvern, PA, USA

ARTICLE INFO

Article history:
Received 21 January 2012
Received in revised form 4 July 2012
Accepted 27 July 2012
Available online 23 August 2012

Keywords:
Shape prior
Segmentation
Sparse representation
Dictionary learning
Mesh partitioning

ABSTRACT

"Shape" and "appearance", the two pillars of a deformable model, complement each other in object segmentation. In many medical imaging applications, while the low-level appearance information is weak or mis-leading, shape priors play a more important role to guide a correct segmentation, thanks to the strong shape characteristics of biological structures. Recently a novel shape prior modeling method has been proposed based on sparse learning theory. Instead of learning a generative shape model, shape priors are incorporated on-the-fly through the sparse shape composition (SSC). SSC is robust to non-Gaussian errors and still preserves individual shape characteristics even when such characteristics is not statistically significant.

Although it seems straightforward to incorporate SSC into a deformable segmentation framework as shape priors, the large-scale sparse optimization of SSC has low runtime efficiency, which cannot satisfy clinical requirements. In this paper, we design two strategies to decrease the computational complexity of SSC, making a robust, accurate and efficient deformable segmentation system. (1) When the shape repository contains a large number of instances, which is often the case in 2D problems, K-SVD is 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, an affinity propagation method is used to partition the surface into small sub-regions, on which the sparse shape composition is performed locally. Both strategies dramatically decrease the scale of the sparse optimization problem and hence speed up the algorithm. Our method is applied on a diverse set of biomedical image analysis problems. Compared to the original SSC, these two newly-proposed modules not only significant reduce the computational complexity, but also improve the overall accuracy.

1. Introduction

In various applications of medical image segmentation, deformable models have achieved tremendous success, thanks to the combined use of shape and appearance characteristics. While appearance features provide low level clues of organ boundaries, shapes impose high level knowledge to infer and refine deformable models. However, in some medical image analysis problems, appearance cues are relatively weaker or even misleading. In those cases, the best "guess" of the organ boundaries can only come from shape priors, which should be effectively modeled from training shapes. Effective shape modeling faces the following challenges, (1) shape variations are 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) the 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 (Cootes et al., 1995), as well as its extensions (Heimann and Meinzer, 2009; Nahed et al., 2006), cannot tackle them in a uniform way. (More detailed reviews are presented in Section 2.1.)

Recently, a new non-parametric method has been proposed to tackle these three challenges in a unified framework (Zhang et al., 2011a, 2012). Instead of using any parametric model of shape statistics, this method incorporates shape priors on-the-fly through *sparse shape composition*. More specifically, there are two general 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 repository; and (2) gross errors from local appearance cues might exist but these errors are sparse in spatial space. This setting shows three major advantages: (1) General: There is no assumption of a parametric distribution model (e.g., a unimodal distribution assumption in ASM), which facilitates the modeling of complex shape statistics. (2) Robust: Since it explicitly models gross errors, erroneous information from appearance cues can be effectively detected and removed. (3) Comprehensive: It exploits information from all training shapes. Thus it is able to recover detail information even if the detail is not statistically significant in training data.

* Corresponding author. Tel.: +1 610 574 1968.
E-mail address: yiqiang@gmail.com (Y. Zhan).

1361-8415/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.media.2012.07.007

Combining of SSC and deformable models provides more robust and accurate segmentation, especially when low level appearance cues are missing or misleading.

However, sparse shape composition has inferior run-time efficiency, in particular under two situations. *First*, there are a large number of training data available and all of them are straightforwardly included in the training repository. *Second*, in practice, many 3D deformable models have many thousands of points in order to give an accurate description of organ shapes. In both scenarios, Sparse Shape Composition needs to solve a large scale sparse optimization problem, which has high computational complexity. Compared to traditional shape modeling methods, a deformable segmentation system with sparse shape composition might exhibit low speed, which cannot satisfy many clinical applications.

In this paper, we propose a deformable model aiming to achieve robust segmentation in the presence of weak/misleading appearance cues. In order to integrate sparse shape composition as a shape prior modeling component, we designed two strategies to improve its run-time efficiency. *First*, a sparse dictionary learning technique is employed to compute a compact codebook (i.e., dictionary), which can represent well the original database. Instead of using all data for model selection, we perform a sparse shape composition on this learned dictionary. Although the dictionary contains less information than the initial complete data, it is still able to capture shape details, and improves the computational efficiency. *Second*, instead of modeling global shape priors, we propose to decompose the deformable surface into multiple parts and build shape models on them independently. The partition is accomplished by affinity propagation method (Frey and Dueck, 2007) using both geometry and appearance features. Besides the run-time efficiency, this strategy also leads to a better shape modeling, since the shape statistics of local sub-surfaces often have more compact distributions than global surfaces. Using these two strategies, we achieve a robust, accurate, and most importantly, *efficient*, deformable segmentation system, which is clinically acceptable.

The rest of the paper is organized as follows. Section 2 reviews relevant work in both shape prior models and sparsity methods. Section 3 details our algorithm, including the segmentation framework, basic sparse shape composition and its two key optimizations, (1) dictionary learning, and (2) local shape prior modeling using mesh partitioning. Section 4 shows the experimental results of three applications, including 2D lung localization using dictionary learning, 3D liver segmentation using mesh partitioning, and 3D rat cerebellum segmentation. Section 5 concludes this paper.

2. Related work

Since we use sparsity and dictionary learning methods to model shape priors, we review relevant work in both shape prior and sparsity communities.

2.1. Shape prior models

Following the seminal work on active shape models, many other methods have been proposed to alleviate problems in three categories (1) complex shape variations, (2) gross errors of input, and (3) loss of shape details.

(1) To model complex shape variations, one needs to improve the traditional single Gaussion distribution assumption. A mixture of Gaussians (Cootes and Taylor, 1997) is a natural extension. It is more flexible to represent 2D shape variation than single Gaussian. However, some complex shape variations may not be well represented by any parametric probability distribution. Thus, instead of generically modeling a whole group of shapes, one can model the specific shape. For example, Pohl et al. (2004) have proposed to couple the PCA based shape modeling with a maximum a posteriori estimation problem which is solved through an alternating optimization framework. This allows the system to accommodate shapes that differ from those modeled by the PCA. Thus it is not restricted to the training data, but can model patient-specific abnormalities. Some methods were also proposed to decompose the shape space into multiple sub-spaces, such as patient-specific shape statistics (Shi et al., 2008; Yan and Kruecker, 2010) and subject-specific dynamical model (Zhu et al., 2009, Zhu et al., 2010). Another approach is to non-linearly model shapes using manifold learning techniques, such as in Etyngier et al. (2007) and Zhang et al. (2011b).

(2) To handle outliers of input data, one type of methods is to detect and correct the outliers. For example, Duta and Sonka (1998) propose to use the variance information from the PDM to detect outliers, and use their neighboring information to correct detected outliers. Lekadir et al. (2007) consider the relationship among landmarks (i.e., the ratio of landmark distance) to detect outliers. It is worth mentioning that missing landmarks can also be treated as a special case of outliers. It is solved by using the partial ASM techniques (Yan et al., 2010). The other type of methods uses the weighting of residuals to decrease outliers' influence. For example, Rogers and Graham (2002) have proposed a robust ASM which uses random sample consensus (RANSAC) (Fischler and Bolles, 1981) to remove the influence of outliers. Chui and Rangarajan (2003) have proposed to use robust point matching algorithm to find the best-fitting model and thus reject outliers.

(3) To preserve local shape details, one needs to well model the local shape information even such information is not statistically significant. Since the traditional PCA method has no constraint of the changing scale of each mode, it usually depicts the global shape variance. Thus, sparse PCA (Sjostrand et al., 2007) has been proposed. Different from traditional PCA, sparse PCA ensures that each mode only affects locally clustered landmarks (i.e., neighboring elements in the 2D PCA modes), while the rests are flat zeros. In other words, it captures the local shape variance in each mode. However, it only handles 2D cases since it is ambiguous to map 3D neighbors to 2D vectors. Hierarchical modeling is another approach to incorporate local information. For example, Davatzikos et al. (2003) represents 2D shapes as their wavelet transform, followed by a PCA on the wavelet coefficients. By doing this, local shapes can be well represented using wavelets. In general, it is challenging to preserve local shape details in 3D space, since the 3D neighborhood information is not straightforward to encode explicitly.

2.2. Sparsity methods and dictionary learning

Sparsity methods have been widely investigated recently. It has been shown that a sparse signal can be recovered from a small number of its linear measurements with high probability (Candes et al., 2006; Donoho, 2006). To solve these problems of sparsity priors, one can either use greedy methods such as basis pursuit (BP) (Chen et al., 2001), matching pursuit (Mallat and Zhang, 1993), orthogonal matching pursuit (OMP) (Chen et al., 1896) and stagewise OMP (stOMP) (Donoho et al., 2006), or use $L1$ norm relaxation and convex optimization (Candes et al., 2006; Kim et al., 2007; Figueiredo et al., 2007).

The sparsity prior has been widely used in computer vision and multimedia communities, such as, but not limited to, robust face recognition (Wright et al., 2009), image restoration (Mairal et al., 2009), image bias estimation (Zheng and Gee, 2010), MR reconstruction (Lustig et al., 2007; Huang et al., 2011), automatic image annotation (Zhang et al., 2010), and expression recognition (Zhong et al., 2012). Specifically, Wright et al. (2009) have shown both theoretically and experimentally that sparse representation is critical for high-performance classification of high-dimensional data such as face images. They also demonstrated that the choice of features is less important than the number of features used, and that occlusion and corruption can be handled uniformly and robustly with this framework. We have investigated sparse representation for shapes (Zhang et al., 2011a, 2012) instead of images or videos. In this context, the main challenges are twofold: (1) shape data may have arbitrary dimensions, e.g., 2D contour or 3D mesh; (2) different from the assumption in (Wright et al., 2009), the input shape data may have large misalignments. Thus transformation parameters should also be modeled explicitly.

Furthermore, when training datasets have thousands or millions of samples, it may not be feasible to use all of them due to the computational considerations. It is desirable to learn a compact dictionary to represent the original dataset by minimizing the reconstruction errors. Finding a compact dictionary has been extensively studied in signal processing community. We briefly introduce some relevant techniques. Dictionary learning typically consists of sparse coding and codebook update. Greedy algorithms such as matching pursuit (MP) (Mallat and Zhang, 1993) and orthogonal matching pursuit (OMP) (Chen et al., 1896) can be employed for finding sparse coefficients (coding). Extensive studies of these algorithms show that if the sought solution is sparse enough, these greedy techniques can obtain the optimal solution (Tropp, 2004). To update a codebook, optimal direction (MOD) (Engan et al., 1999) and K-SVD (Aharon et al., 2006) are two effective approaches. Although both of them achieve similar results, we use K-SVD to learn our dictionary because of its better convergence rate.

It is also worth mentioning that sparsity has been adopted in segmentation algorithms in different manners, such as the sparse information models (Florin et al., 2007), which reconstruct a 3D surface from 2D sparse subcomponents.

3. Methodology

3.1. Segmentation framework

To remain general, our segmentation framework is designed as "data-driven". Fig. 1 shows the workflow of our segmentation system, which consists of offline learning and runtime segmentation stages. In offline learning, images along with the manually labeled ground truths are employed to learn the appearance and shape characteristics of the organ under study.

Appearance characteristics are obtained through learning landmark detectors and a set of spatially adaptive boundary detectors. More specifically, landmark/boundary detection is formulated as a classification problem. A large number of features are extracted using over-complete Haar wavelets. These features are selected and optimally aggregated using Adaboost classification algorithm. Shape characteristics are extracted from a set of manually delineated organ contours or surfaces. A shape repository/dictionary is constructed using these organ contours/surfaces. It will be exploited to derive shape priors during runtime.

Runtime segmentation starts from the initialization of the model (represented by a 2D contour or a 3D triangular surface) based on automatically detected landmarks and shape priors. The contour/surface then deforms under the guidance of both image appearance cues and shape priors. Deformation is usually formulated as the optimization of an energy function consisting of external/image and internal/shape terms, and in practical implementations it is usually performed in two steps iteratively. 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. These two steps will be repeated until convergence.

Although learning-based landmark/boundary detectors can tackle reasonable appearance variations (Zhan et al., 2008, 2009), 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. 1.) Therefore, the effective modeling of shape priors becomes extremely critical to achieve a robust segmentation. This will be the major topic in the remainder of this paper. (For more details of appearance modeling, refer to Zhan et al. (2008, 2009, 2011).)

3.2. Sparse shape composition

In this subsection, we briefly introduce the sparse shape composition (SSC) model proposed in Zhang et al. (2012). Instead of assuming any parametric probabilistic distributions of the shape statistics, SSC is designed according to 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, shape priors are incorporated *on-the-fly* through shape composition, i.e., a shape derived

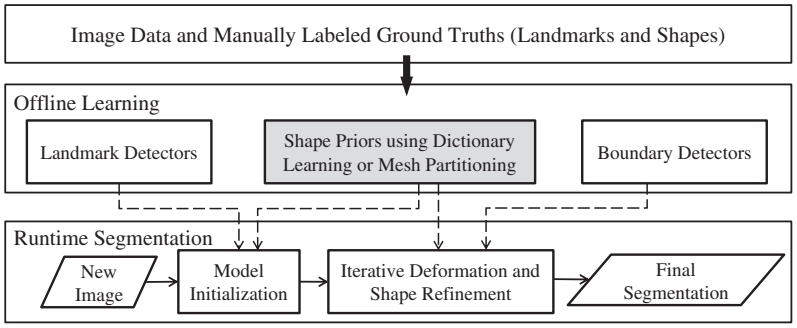

Fig. 1. The workflow of our segmentation framework consisting the offline learning and online testing modules. The grey block indicates our major novelty and contribution in this system.

by appearance cues is refined by the approximation of a set of annotated shape instances following the two sparsity observations.

In this study, a shape is represented by a contour (2D) or a triangle mesh (3D) which consists of a set of vertices and edges. Denote the input shape as $\mathbf{v}$, where $\mathbf{v} \in \mathbf{R}^{\mathfrak{D}N}$ is a vector concatenated by coordinates of its N vertices, and $\mathfrak{D} = \{2, 3\}$ denotes the dimensionality of the segmentation problem. (In the remainder of this paper, any shape instance is defined as a vector in the same way.) Assume $D = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K] \in \mathbf{R}^{\mathfrak{D}N \times K}$ is a large shape repository that includes K accurately annotated shape instances $\mathbf{d}_i$. Note that $\{\mathbf{d}_i, i = 1, 2, 3, \ldots, K\}$ are pre-aligned using generalized Procrustes analysis (Goodall, 1991; Cootes et al., 1995). The approximation of $\mathbf{v}$ by D is then formulated as an optimization problem:

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

where $T(\mathbf{v}, \beta)$ is a global transformation operator with parameter β. It aligns the input shape $\mathbf{v}$ to the common canonical space of D. The key idea of SSC lies in the two conditions of the objective function. In the first condition, $\mathbf{x} \in \mathbf{R}^K$ denotes the coefficient/weights of linear combination. The $L0$ norm of $\mathbf{x}$ ensures that the number of non-zero elements in $\mathbf{x}$ is less than a sparsity number (i.e., k_1 or k_2). 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 condition, $\mathbf{e} \in \mathbf{R}^{\mathfrak{D}N}$ 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.

Eq. (1) is known as a NP hard problem owing to the non-convex $L0$ norm. $L1$ norm relaxation (Donoho, 2004) can be employed to make the problem convex while still preserving the sparsity property. The optimization problem is then transformed to Eq. (2)

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

Eq. (2) is minimized using a alternating minimization scheme. In first step, the transformation parameter β is estimated using Procrustes analysis. In second step, with a fixed β, Eq. (2) becomes a typical linear inverse problem, which can be solved by a typical convex solver, such as the interior-point convex optimization solver (Nesterov and Nemirovsky, 1994; Grant and Boyd, 2008). These two steps are iteratively performed until $\mathbf{x}, \mathbf{e}$ and β converge.

3.3. Computational complexity analysis

By optimizing Eq. (2), shape priors are in fact incorporated *on-the-fly* through shape composition. Compared to traditional statistical shape models, e.g., active shape model, SSC is able to remove gross errors from local appearance cues and preserve shape details even if they are not statistically significant. However, on the other hand, the optimization of Eq. (2) increases the computational cost, which limits the run-time efficiency of the sparse shape composition. From the definition of the objective function, the computational cost is determined by the size of the shape repository matrix $D \in \mathbf{R}^{\mathfrak{D}N \times K}$. More specifically, the computational complexity of the interior-point convex optimization solver is $\mathcal{O}(N^2K)$ per iteration (Nesterov and Nemirovsky, 1994), which means the computational cost will increase quickly with the increase of (1) K, number of the shape instances in the shape repository; and (2) N, number of vertices of a shape. Note that $\mathcal{O}(N^2K)$ is the computational complexity for *one* iteration. Empirically, with larger K and N, it usually takes more iterations to converge, which further decreases the algorithm speed.

Accordingly, the run-time efficiency of SSC will be particularly low in two scenarios. *First*, in some applications, e.g., 2D lung localization in X-ray images, a large amount of training shapes (a large K) are available due to the relatively lower cost of manual/semi-automatic annotations. *Second*, a large number of vertices (a large N), are necessary to describe shape details. This is usually the case in 3D segmentation problem. Therefore, a straightforward integration of SSC with a deformable model might have a very low segmentation speed, which is not clinically accepted. To alleviate this problem, we design two strategies as detailed next.

3.4. Dictionary learning for compact shape representations

Algorithm 1. The framework of the OMP algorithm.

Input: Dictionary $D \in \mathbf{R}^{n \times k}$, input data $y_i \in \mathbf{R}^n$.
Output: coefficients $x_i \in \mathbf{R}^k$.
$\Gamma = \emptyset$.
repeat
Select the atom which most reduces the objective

$$\arg\min_j \left\{ \min_{x'} |y_i - D_{\Gamma \cup \{j\}} x'|_2^2 \right\} \tag{3}$$

Update the active set: $\Gamma \leftarrow \Gamma \cup \{j\}$.
Update the residual using orthogonal projection

$$r \leftarrow \left(I - D_\Gamma \left(D_\Gamma^T D_\Gamma \right)^{-1} D_\Gamma^T \right) y_i \tag{4}$$

where r is the residual and I is the identity matrix.
Update the coefficients

$$x_\Gamma = \left(D_\Gamma^T D_\Gamma \right)^{-1} D_\Gamma^T y_i \tag{5}$$

until Stop criteria

Algorithm 2. The framework of the K-SVD algorithm.

Input: dictionary $D \in \mathbf{R}^{n \times k}$, input data $y_i \in \mathbf{R}^n$ and coefficients $x_i \in \mathbf{R}^k$.
Output: D and X.
repeat
Sparse coding:
use OMP to compute coefficient x_i for each signal y_i, to minimize

$$\min_{x_i} \left\{ \|y_i - Dx_i\|_2^2 \right\} \; subject \; to \; \|x_i\|_0 \leqslant L \tag{6}$$

Codebook update:
for $i = 1, 2, \ldots, k$, update each column d_i in D and also x_T^i (ith row)
Find the group using d_i ($x_T^i \neq 0$), denoted as ω_i
Compute error matrix E_i as in Eq. (10)
Restrict E_i by choosing columns corresponding to ω_i.
The resized error is denoted as E_i^R
Apply SVD and obtain

$$E_i^R = U \Sigma V^T \tag{7}$$

Update d_i as the first column of U. Update non-zero elements in x_T^i as the first column of V multiplied by $\Sigma(1,1)$

until Stop criterions

Assume K is the number of training samples, the computational complexity of the sparse optimization for each iteration is comparable to the cost of solving a least-squares problem (*i.e.*, $\mathcal{O}(N^2K)$), when using interior-point methods (Nesterov and Nemirovsky, 1994). One factor that affects the run-time efficiency is K, the number of shape instances in matrix D. In many applications of medical image analysis, the number of training data is usually limited ($\sim$100) due to the high cost of manual annotations. In such cases, we can simply include all samples in the matrix D, since Eq. (2) can be still efficiently solved. However, in some other applications, e.g., lung localization in chest X-ray images, one can potentially get thousands of training data using some semi-automatic tools. In this scenario, including all shape instances in D will dramatically decrease the algorithm speed.

In fact, owing to the similar shape characteristics across the population, these thousands of shape instances usually contain significant redundant information. Instead of including all of them, D should be a shape "dictionary" with a significantly more compact size. However, compact size is not the only requirement of the shape dictionary. Since sparse shape composition is our ultimate goal (see Eq. (2)), this shape dictionary should also have the sparse representation capability, i.e., any shape in the population can be approximated by a sparse set of elements in the dictionary. To learn such a shape dictionary, we employ a recently proposed algorithm K-SVD (Aharon et al., 2006).

Mathematically, K-SVD aims to optimize the following objective function with respect to dictionary D and coefficient X:

$$\arg\min_{D,X}\left\{\|Y - DX\|_2^2\right\} \tag{8}$$

$$s.t.\ \forall i, \|x_i\|_0 \leqslant L \tag{9}$$

where matrix $Y \in \mathbf{R}^{n\times K}$ represents the complete dataset (all training shapes in our case), $D \in \mathbf{R}^{n\times k}(k \ll K)$ is the unknown overcomplete dictionary, matrix X contains the sparse coefficients. Denote y_i as the ith column of Y, x_i as the ith column of X, then y_i and x_i are the ith shape vector and coefficient vector, respectively, with dimensionality, $y_i \in \mathbf{R}^n$ and $x_i \in \mathbf{R}^k$. This equation contains two important properties of the learned dictionary D. *First*, $k \ll K$ indicates the dictionary has a significantly more compact size. *Second*, $\forall i$, $\|x_i\|_0 \leqslant L$ guarantees the sparse representation capability of the dictionary.

In K-SVD algorithm, Eq. (9) is optimized by two alternative 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 used to generate a better dictionary given sparse coding results. These two steps are alternately performed until convergence.

Sparse coding stage: K-SVD algorithm starts from a random D and X and the sparse coding stage uses pursuit algorithms to find the sparse coefficient x_i for each signal y_i. OMP (Chen et al., 1896) is employed in this stage. OMP is an iterative greedy algorithm that selects at each step the dictionary element that best correlates with the residual part of the signal. Then it produces a new approximation by projecting the signal onto those elements already selected (Tropp, 2004). The pseudo-codes of OMP is listed in Algorithm 1.

Codebook update stage: In the codebook update stage K-SVD aims to update D and X iteratively. In each iteration D and X are fixed except only one column d_i and the coefficients corresponding to d_i (ith row in X), denoted as x_T^i. Eq. (8) can be rewritten as

$$\left\|Y - \sum_{j=1}^{k} d_j x_T^j\right\|_F^2 = \left\|\left(Y - \sum_{j\neq i} d_j x_T^j\right) - d_i x_T^i\right\|_F^2 \tag{10}$$

$$= \left\|E_i - d_i x_T^i\right\|_F^2 \tag{11}$$

We need to minimize the difference between E_i and $d_i x_T^i$ with fixed E_i, by finding alternative d_i and x_T^i. Since SVD finds the closest rank-1 matrix that approximates E_i, it can be used to minimize Eq. (10). Assume $E_i = U\Sigma V^T$, d_i is updated as the first column of U, which is the eigenvector corresponding to the largest eigenvalue. x_T^i is updated as the first column of V multiplied by $\Sigma(1,1)$. The updated x_T^i may not always guarantee sparsity. A simple but effective solution is to discard the zero entries corresponding to the old x_T^i. The detail algorithms of K-SVD are listed in the Algorithm 2.

The learned dictionary D will be used in Eq. (2) at run-time. It is worth noting that an element in D might not be the same as any shape instances in the training set. In other words, the learned shape dictionary consists of virtual shape instances which might not exist in the real world. However, these virtual shapes do have sparse composition capabilities with a significantly more compact size, which can highly improve the run-time efficiency of our sparse shape composition.

Fig. 2 shows an illustration of dictionary learning for 2D lung shapes. A compact dictionary is generated from input samples, such that each input sample can be approximately represented as a sparse linear combination of dictionary elements. This compact dictionary is used as data matrix D for our model.

3.5. Mesh partition for local sparse shape composition

Recall the computational complexity of the solver $\mathcal{O}(N^2K)$. In practice, many 3D deformable models include many thousands of vertices (*i.e.*, large N) to give an accurate description of organ shapes. The optimization of Eq. (2) thus has high computational complexity. If the whole surface can be divided into p partitions with about $\frac{N}{p}$ vertices in each partition, the computational complexity is decreased to only $\mathcal{O}\left(\left(\frac{N}{p}\right)^2 K\right)$ for each partition, which is $\frac{1}{p^2}$ of the original one.

Our sparse shape composition method inherently supports the partition of surface 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 ith vertex is in the subset ($\mathbf{S}_{ii} = 1$). Eq. (1) can then be naturally extended as:

$$\arg\min_{\mathbf{x},\mathbf{e},\beta} \|T(\mathbf{v_{sub}},\beta) - \mathbf{S}D\mathbf{x} - \mathbf{S}\mathbf{e}\|_2^2, \quad s.t. \|\mathbf{x}\|_0 < k_1, \quad \|\mathbf{e}\|_0 < k_2 \tag{12}$$

Eq. (12) can be solved using the same $L1$ norm relaxation and alternating 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.

The remaining problem is to divide the surface efficiently. In this study, affinity propagation clustering (Frey and Dueck, 2007) is employed to divide the model shape into multiple partitions. Affinity propagation is a general clustering method, which only requires the definition of similarity/affinity between data pairs (vertices pairs in our study). During the clustering process, two types of messages derived from the affinity definition are exchanged between data points (vertices in our study). (1) "Responsibility" is sent from data points to exemplar candidates. It measures how well-suited a point is to serve as the exemplar for another. (2) "Availability" is sent from exemplar candidates to data points, which reflects how appropriate it would be for a point to choose another one as its exemplar. Messages are iteratively exchanged until a high-quality set of exemplars and corresponding clusters gradually emerges. It is worth noting that affinity propagation does not require a pre-specific cluster number. Instead, the number of cluster is determined by the definition of affinity between data points.

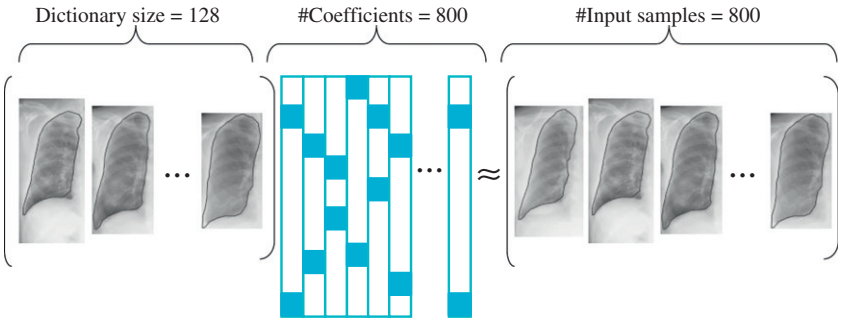

Fig. 2. An illustration of dictionary learning for 2D lung shapes. The input data contains 800 training samples, and a size 128 compact dictionary is generated. Each input shape can be approximately represented as a sparse linear combination of dictionary elements. Note that dictionary elements are "virtual shapes", which are not the same as any input shape but can generalize them well.

The similarity used in the affinity propagation is defined as the combination of the image similarity and geodesic distances between vertices (Zhan et al., 2009):

$$s(v_i, v_j) = 1 - \frac{1}{K}\sum_{k=1}^{K}\left[\alpha G\left(v_i^k, v_j^k\right) + (1-\alpha)C\left(F(v_i^k), F\left(v_j^k\right)\right)\right] \quad (13)$$

where K is the number of training subjects, v_i^k denotes the ith vertex of the kth subject. $G\left(v_i^k, v_j^k\right)$ denotes the geodesic distance between v_i^k and v_j^k. $F(v_i^k)$ and $F\left(v_j^k\right)$ denote the image feature vectors on vertices v_i^k and v_j^k. $C\left(F(v_i^k), F\left(v_j^k\right)\right)$ denotes the Euclidean distance between image feature vector calculated at v_i^k and v_j^k.

In Eq. (13), the first term ensures vertices of the same cluster/sub-surface are spatially continuous. The second term is designed to group vertices with similar image appearance into the same cluster. These two terms are balanced by parameter α, which is set as 0.2 in this study. Note that the affinity propagation only needs to perform once for the model shape, since one-to-one correspondences have already been constructed among all shapes.

It is worth noting that there are two additional benefits of partitioning the entire surface in our deformable segmentation framework. *First*, since organ boundaries may have heterogeneous appearance, as shown in Fig. 3, the partitioning of surface also facilitates the training of local boundary detectors, which only deal with training samples with less feature variations. *Second*, 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 of shape modeling as well.

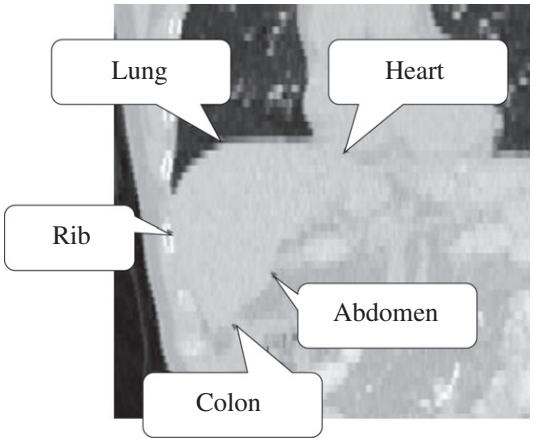

Fig. 3. An example of liver boundaries from CT images. It includes boundaries between liver and rib, lung, heart, abdomen and colon that show heterogeneous appearance in local shapes.

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 $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_p$ used in Eq. (12). The optimization problem defined on the entire surface is thus decomposed into 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 merit of Eq. (12) is actually beyond the support of surface partition. One extreme situation of Eq. (12) is that $\mathbf{S}$ becomes very sparse and only includes a few vertices (usually with the most distinctive appearance/geometry characteristics). In this situation, Eq. (12) is equivalent to the shape inference method, 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 or imaging artifacts.

In addition, Eq. (12) also supports multi-resolution/hierarchical scheme, which has been widely employed to improve the efficiency and robustness of deformable segmentation (Langs et al., 2010). In our implementation, 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.

4. Experiments

In this section, we evaluate the proposed method in three applications. (1) To locate the right lung, we use landmark detector to find several anatomical points, and then infer a shape to fit these points. Since there are around 1000 data with lung shape delineated, we use dictionary learning to generate a compact dictionary. Note that this localization corresponds to the model initialization in our segmentation framework (Fig. 1). (2) To segment the liver in whole body CT, we use a similar scheme as the lung application, *i.e.*, initialization using landmark detection and shape inference, and then segment it using deformable models. Since the mesh has thousands of vertices, we use a mesh partitioning algorithm to obtain multiple local shapes, which significantly improves the computational efficiency. (3) We also segment rat cerebellum using shape priors and deformable models. The prior module is able to preserve shape details.

4.1. 2D lung localization using dictionary learning

In this experiment we aim to roughly locate the right lung from chest X-ray, using landmark detections and sparse dictionary learning. Radiography (X-ray) is widely used for screening because of the fast imaging speed, low cost and low radiations. Various

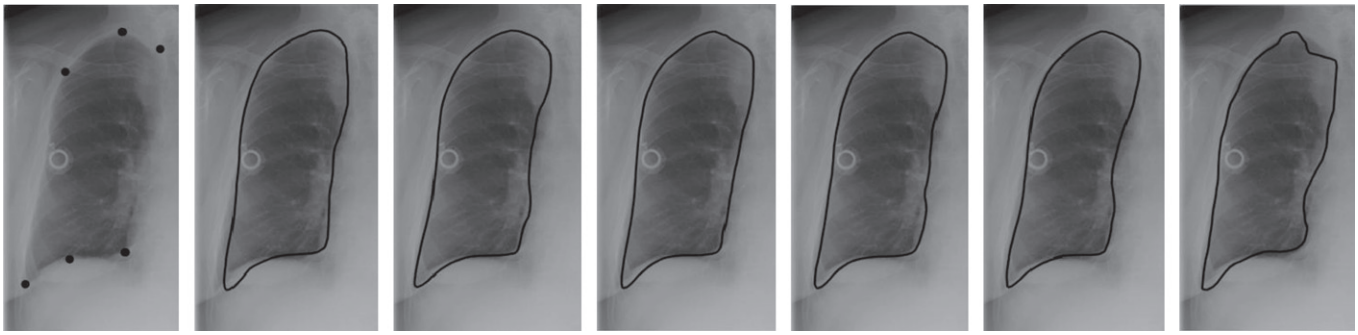

Fig. 4. Comparison of sparse shape composition using different dictionaries. (a) is the detected landmarks. (b) is the shape inference result of using all training data. (c)–(g) are results of using different sizes of dictionaries learned from K-SVD.

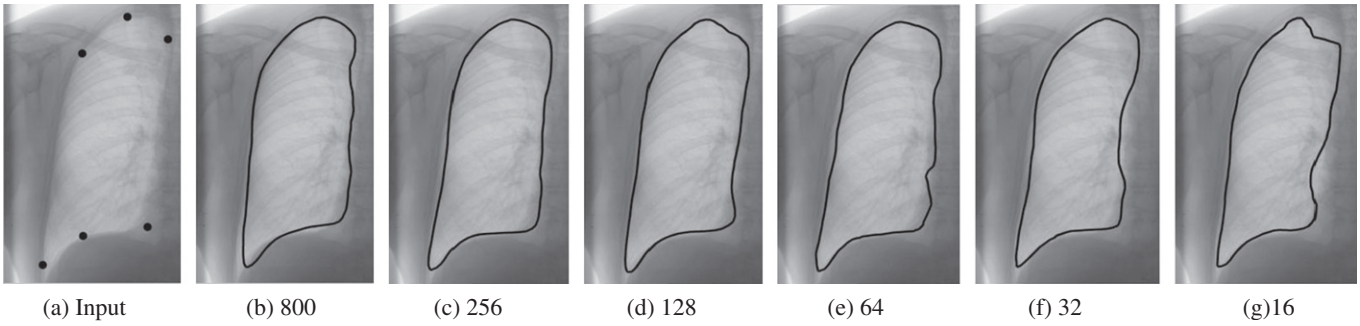

Fig. 5. Comparison of sparse shape composition using different dictionaries. The same setting is used as in Fig. 4.

studies have been conducted for the automatic detection of pathologies, including abnormal cardiac sizes, pneumonia shadow and mass lesions by analyzing X-ray images. As a common first step of these algorithms, lungs need to be automatically localized in a robust and efficient way.

Our experiment dataset includes 1037 scans generated by different X-ray machines, e.g.,Siemens FD-X, GE Revolution XRd and Canon Axiom-Muttix, etc., with resolution from 0.143 mm to 0.195 mm. The ground truths are binary masks of manual segmentation results, labeled by a clinical expert. From the binary masks, 2D contours can be easily extracted. We then manually select six specific landmarks (e.g., lung apex, lung lateral tips and hepatic dome) on the contour as the anatomical landmarks for training purpose. After that, a fixed amount of points between two neighboring landmarks are evenly and automatically interpolated along the contour. Thus, a one-to-one correspondence is obtained for all landmarks and shapes. When a new image is presented, we first detect the anatomical landmarks (which may contain errors), and then use shape priors to infer a shape from these detected landmarks.

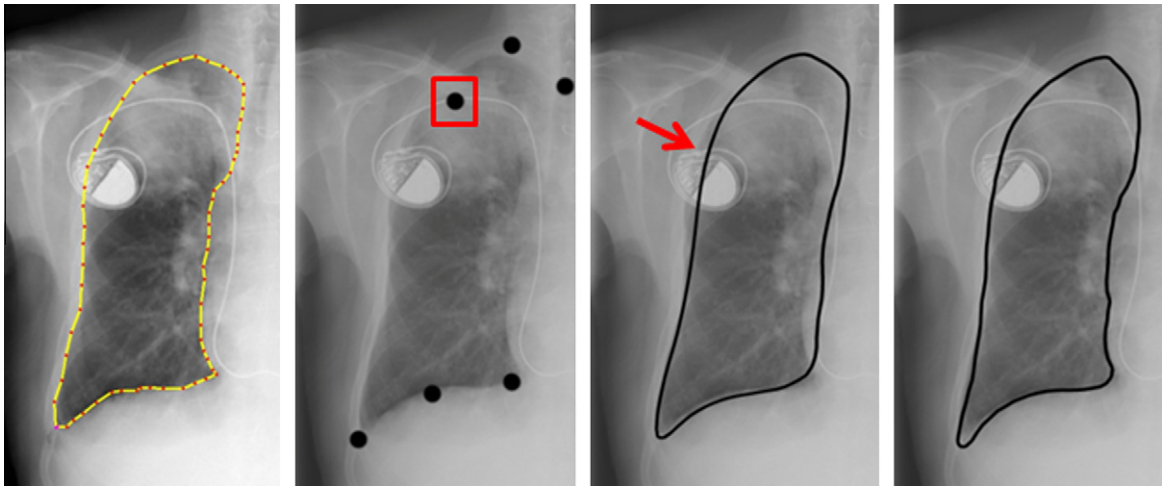

Fig. 6. Comparisons of the localization results. From left to right: manual label, detection results, ASM's shape prior, the proposed shape prior method with a learned dictionary. Due to the erroneous detection (marked by the red box), PCA result moves to the right and is not on the boundary (see the red arrow). Zoom in for better view. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

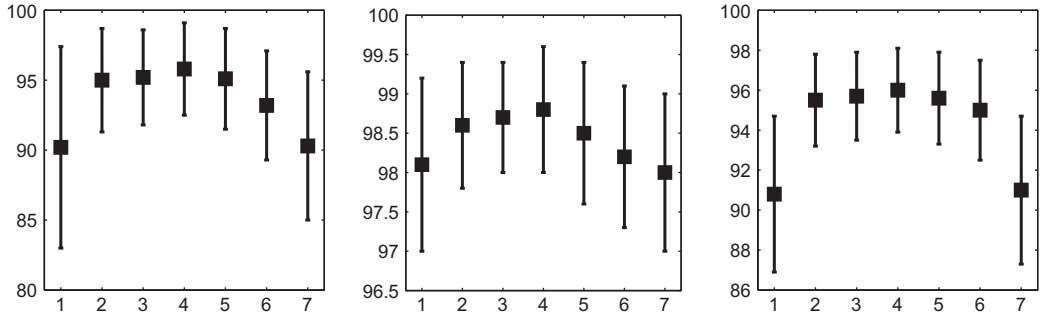

Fig. 7. Quantitative comparison of the mean values (black square) and the standard deviations (the lengths of the lines) in terms of sensitivity, specificity, and dice similarity coefficient (DSC). From left to right: baseline method using Procrustes analysis (mean shape only), sparse shape composition using all data, or using dictionaries with size 256, 128, 64, 32, 16, respectively.

Among all data, 800 are used as training, and the remaining 237 are used as testing. Five dictionaries are learned from 800 training data, whose sizes are 256, 128, 64, 32, 16, respectively. We compare the performance of sparse shape composition using either these dictionaries or using all data. Figs. 4 and 5 show two visual comparisons of shape inference results. It is worth noting that using compact dictionary sometimes achieves better results than using all data. The reason is that K-SVD not only generalizes the training data, but also removes noise when computing the compact dictionary. When the dictionary size is smaller than a threshold (16 in our experiment), however, its performance drops quickly. Thus a certain number of elements is still necessary for a meaningful dictionary. Fig. 6 compares the proposed method with the traditional PCA-based shape prior method, which has been widely used in ASM and its variations. Due to the erroneous detection (marked by the red box in Fig. 6), the PCA result shifts to the right and is not on the boundary (see the red arrow). Our method is more robust to detection outliers and thus generates a more accurate result.

Fig. 7 shows quantitative comparisons of the proposed method. The mean values and the standard deviations of sensitivity, specificity, and dice similarity coefficient (DSC) are reported. We choose Procrustes analysis (Goodall, 1991) as the baseline method, which fits the mean shape to detected landmarks using the similarity transformation found by minimizing least square errors. We also compare the performance of sparse shape composition using all training data (Zhang et al., 2012), or using compact dictionaries with sizes 256, 128, 64, 32 and 16. The results are consistent with the visual comparison in Figs. 4 and 5, *i.e.*, compact dictionaries can achieve comparable or even better performance as using all data, when the dictionary size is larger than a threshold. When only few instances are used as the dictionary, it actually degenerates to the simplest case, *i.e.*, only using mean shape as the Procrustes analysis. Thus the performance of dictionary with size 16 achieves similar performance as Procrustes method.

Another significant benefit of using compact dictionary is the computational efficiency. The experiments of computational time are performed on a PC with 2.4 GHz Intel Quad CPU, 8 GB memory, with Python 2.5 and C++ implementations. CVX (Grant and Boyd, 2008) is employed to solve convex functions. Table 1 shows the computational time of the compared methods. The traditional SSC cannot handle a large number of training data. It takes more than half a minute to infer a shape when using all 800 training data. This is not fast enough for real world applications since CAD program usually needs very efficient localization process. Using compact dictionary significantly decreases the computational time of shape inference to less than one second without adversely affect the accuracy. Thus, this proposed dictionary learning based shape prior can be potentially used in realtime systems.

Table 1
Comparison of computational efficiency of using sparse shape composition with all data or with compact dictionaries. We also list the running of time of Procrustes analysis (PA) for reference.

All data	Dict (256)	Dict (128)	Dict (64)	Dict (32)	Dict (16)	PA
31.7 s	6.8 s	4.3 s	2.1 s	0.5 s	0.1 s	0.07 s

4.2. 3D liver segmentation using mesh partitioning

To evaluate the mesh partitioning based shape priors, we use 3D low-dose CT data from PET-CT. CT images in PET-CT scans usually have low contrast and fuzzy boundaries between organs since PET-CT uses low dose and has large slice thickness in order to decrease radiations to patients. Thus, it is more challenging to accurately segment organs in whole body PET-CT compared to using the traditional CT.

Our experiment dataset includes 67 CT datasets from different Siemens scanner (SOMATOM Emotion Duo and Sensation 16). The inslice resolution ranges from 0.97 mm to 1.36 mm and the slice thickness is inbetween 3.37 and 5 mm. The 3D ground truth of all data are provided by clinical experts using itk-SNAP.[1] Fourty out of 67 CT scans are used to train the landmark detectors and construct the shape repository D. The other 27 are left for testing.

In the training stage, to obtain the one-to-one correspondence for vertices among all shapes, we use adaptive-focus deformable model (AFDM) (Shen and Davatzikos, 2000) to register a reference shape to all the others. Thus all registered shapes share the same topology and connectivity. Each shape has 1096 vertices and 2088 triangles. Twenty vertices which are critical anatomical landmarks or have discriminative appearance features are selected as landmarks for model initialization, since they are relatively easy to detect using our landmark detectors.

Since sparse shape composition is not efficient to deal with a large number of vertices, we evenly select a subset of vertices for Sparse Shape Composition for acceleration purpose, in a multi-resolution scheme. However, this approach ignores much information since only a part of the shape is considered. Instead of using a subset, we use mesh partitioning to generate local shape partitions, and then apply Sparse Shape Composition on these local shapes. This approach not only improves the computational efficiency, but also models well the local shape statistics. Fig. 8 shows an example of a partition result on the liver data.

[1] http://www.itksnap.org/pmwiki/pmwiki.php.

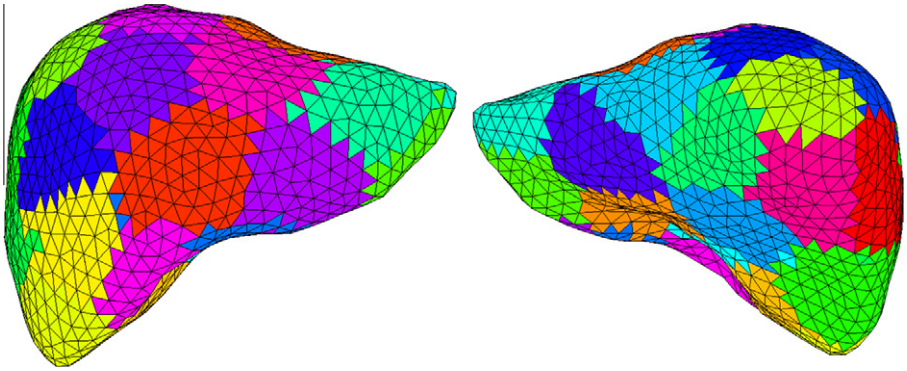

Fig. 8. Mesh partitioning using affinity propagation. This local shape setting can learn detail information better, and is more efficient for the optimization solver.

Four methods (including the proposed one) are compared:

1. The baseline is a learning based deformable model with Procrustes analysis as the shape prior, which finds a similarity transformation to fit a mean shape to detected landmarks.
2. The hierarchical clustering and learning based deformable model with PCA shape priors (Zhan et al., 2009). In the same way as the widely used ASM, PCA is applied to learn shape statistics and then refine the intermediate shape result.
3. The traditional SSC based shape prior, which is applied on the whole shapes (Zhang et al., 2012). The disadvantage is that the shape contains thousands of vertices and thus the optimization is slow.
4. The proposed method, i.e. the mesh partitioning based SSC. It not only models local shape priors, but also improves the computational efficiency. The surface mesh is divided into 30 regions as shown in Fig. 8.

Note that the same landmark/boundary detectors and deformation strategy are used in all methods for fair comparisons. These compared methods only differ when shape priors are involved, i.e., model initialization and model refinement. The reason to compare with (2) and (3) is that they achieve state-of-the-art quality in this application. We aim to improve the accuracy by modeling local shape priors and also accelerate the shape prior procedure.

Fig. 9 compares the landmark detection based initializations. These cases are challenging since the landmark detector may easily fail to locate correct positions due to the low image contrast and the breathing artifacts, at the boundary between the lung and the liver. SSC based prior is less sensitive to such errors. Thus its initialization result is already very close to the object boundary, which also benefits the deformation performance. Fig. 10 compares the refinement methods which apply shape priors on intermediate results. Since this shape refinement is just a regularization step without considering any image information, the refined shape may not be exactly on

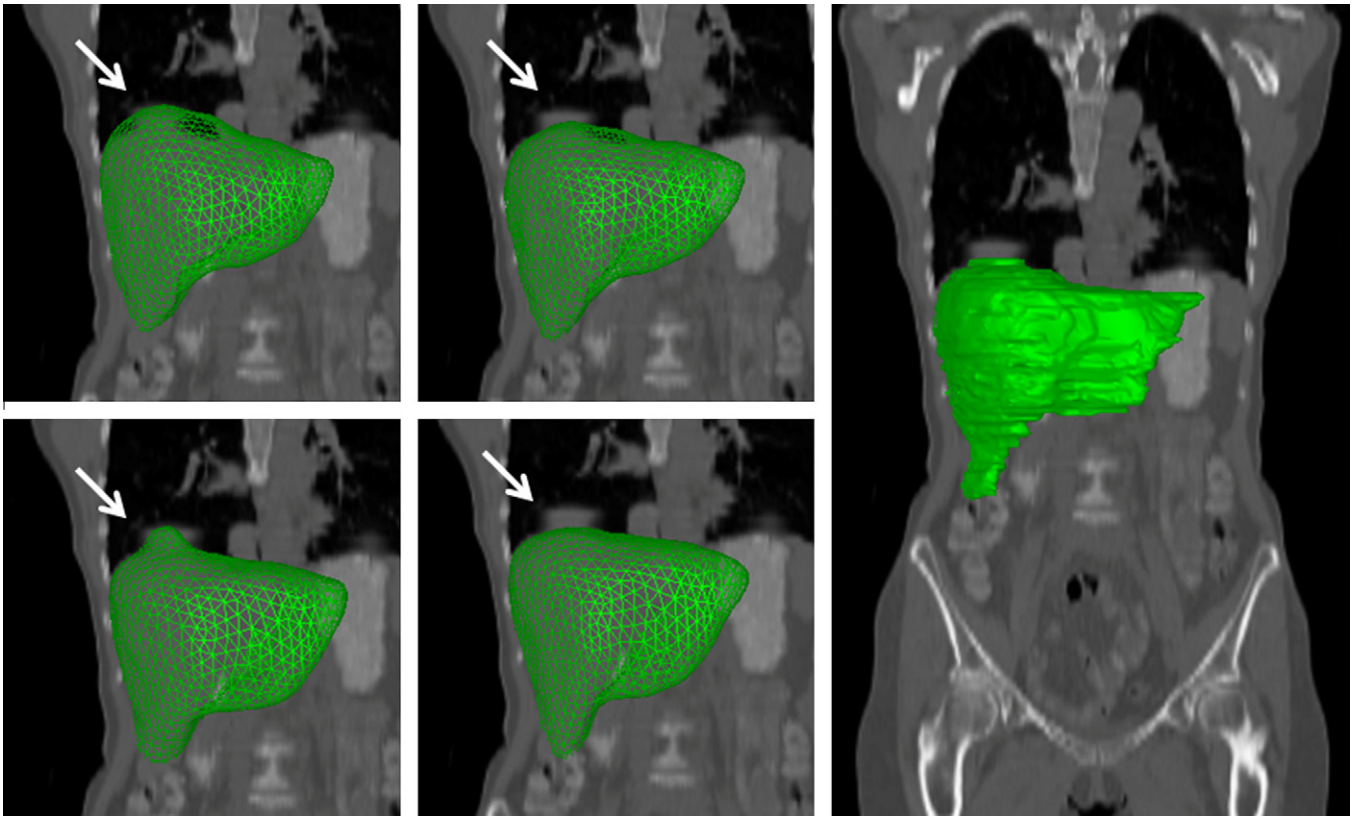

Fig. 9. Initialization results (1st row) and deformation results (2nd row) from the corresponding initialization. Compared methods are Procrustes analysis (1st column) and our proposed method (2nd column). The rightmost figure is the ground truth. Procrustes analysis incorrectly includes part of the lung because of the artifacts inducing by breath (see the marked arrow).

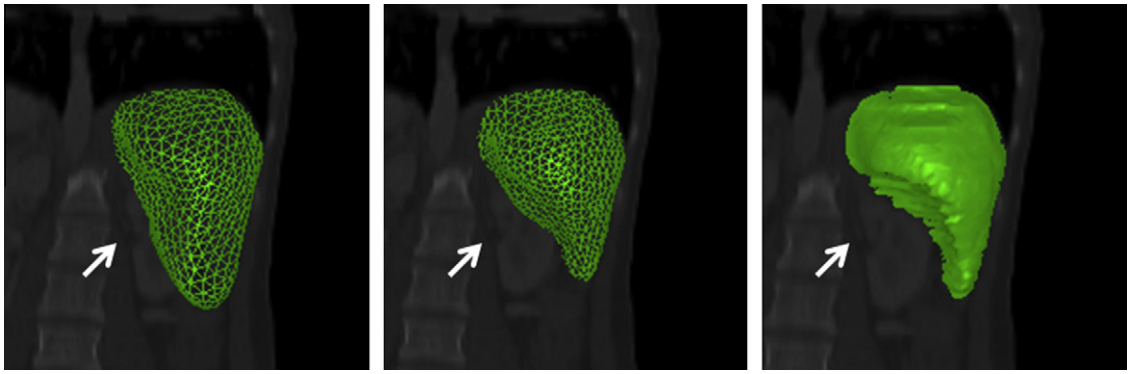

Fig. 10. The refinement results. From left to right: PCA results, the proposed method, the ground truth. PCA method incorrectly includes a large part of the kidney.

Table 2
Quantitative comparisons of the mean values and standard deviations of the distances (voxel) between surfaces. The 1st column compares initialization results. The baseline method is the deformable model with procrustes analysis as shape priors. The hierarchical model (Zhan et al., 2009) is based on clustering, learning, and PCA shape priors. The other two methods use the same deformation scheme but with SSC as the shape prior, which is applied on the whole shape (Zhang et al., 2012) or on shape partitions (the proposed method). Bold values mean the best performance in that column.

Method	Fig. 9(Init)	Fig. 10	All data
Baseline method	2.26 ± 1.72	3.67 ± 3.34	3.96 ± 3.21
Hierarchical model	2.26 ± 1.72	1.81 ± 2.10	2.16 ± 1.68
SSC on the whole shape	**1.31 ± 0.95**	1.37 ± 0.96	1.25 ± 0.92
SSC with mesh partition	**1.31 ± 0.95**	**1.10 ± 0.87**	**1.13 ± 0.83**

the image boundary. It still requires several iterations of deformation to reach the correct boundary. The method in Zhan et al. (2009) uses PCA based shape prior, which incorrectly includes a large part of the kidney. Our segmentation method with SSC prior is more specific to this image and is more accurate.

To quantitatively evaluate the 3D segmentation accuracy, we report the mean value and standard deviation of the distances between shape surfaces in Table 2. The reason of not using the same measurements as in the 2D application (i.e., sensitivity, specificity, and DSC) is that the values of these measurements are always very high in 3D cases, since the 3D segmentation results usually have very large overlap with the ground truth in volume space. Thus, they are not adapted for the comparison of different methods and show the improvements using these measurements. Instead, the distances between shape surfaces are chosen to show the quantitative results, which have been widely used in 3D segmentation tasks. Our proposed framework achieves the best performance. The standard deviations in Table 2 show that our method also achieves the best stability. Sparse prior with mesh partitioning is slightly better than the same prior with whole shape since local shape priors are better modeled. Note that the initialization step is the same for both settings, *i.e.*, using multi-resolution scheme.

To further evaluate the benefit of mesh partitioning, we also compare the running time. The shape refinement step takes several minutes when applied to the whole surface directly. Again, it is not fast enough for real world applications. The mesh partitioning based shape prior significantly improves the efficiency. It only takes 2–3 s to refine each shape. The whole segmentation system takes around 10–20 s (a Python implementation on a PC with 2.4 GHz Intel Quad CPU, without any code optimizations) to segment the liver in a $512 \times 512 \times 300$ CT volume.

4.3. Rat cerebellum segmentation

In this study, we use sparse shape composition and deformable models to segment 3D rat cerebellum in MR microscopy images. MRI at a spatial resolution of at least 100um in one dimension is frequently referred to as magnetic resonance microscopy (MRM) and is currently available with the use of high magnetic field images. This technical achievement has permitted the detailed anatomical study of the rat brain, which is much smaller than the human brain (2.8 cm^3 rat brain volume vs. 1500 cm^3 human brain volume), and requires a small voxel size (0.001 mm^3 for rat vs. 1 mm^3 for human brain), in order to be imaged. Rats 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. The analysis of rat brain image faces similar challenges to human imaging, with individual variation in size, morphology, and topology of the brain structures complicating the neuroanatomical studies. Such analysis is frequently performed by segmenting the regions of interest (ROI) in rat brain images. Thus, automatic 3D segmentation is important for clinical analysis.

In our experiments, 58 datasets are used. To obtain the data, the adult male Sprague–Dawley rats were transcardially perfused with 4% paraformaldehyde. Their heads were stored in paraformaldehyde, and then scanned for magnetic resonance microscopy (MRM). To avoid tissue and shape distortions during brain extraction, the brains remained in the heads during scanning. The heads were scanned on a 21.1T Bruker Biospin Avance scanner (Bruker Biospin Corporation, Massachusetts, USA), and the protocol consisted of a 3D T2-weighted scan with echo-time (TE) 7.5 ms, repetition time (TR) 150 ms, 27.7 kHz bandwidth, field of view (FOV) of $3.4 \times 3.2 \times 3.0$ mm, and voxel size 0.08 mm, isotropic. 3D annotation is manually performed on all 58 data by clinical experts. Fourty are used as training data, and the remaining 18 are used as testing.

We focus on the 3D segmentation of the cerebellum (Fig. 11). The challenge of this segmentation task is twofold. (1) The image information is sometimes incomplete or misleading. For example, the cerebellum contains interleaving textures. These textures are complex and there are high gradient values inside of the cerebellum region, which adversely affect the deformation module. (2) It is important to discover and/or preserve some complex local shape details, such as the paraflocculi (two protruding features of the cerebellum), which are tend to be falsely "smoothed out" by traditional shape prior modeling. Thus we need effective shape priors and deformable models to segment it. We show that this shape prior can robustly segment the rat cerebellum, and also preserve shape details during deformable segmentation.

The visual comparison of the segmentation results is shown in Fig. 11. We compared with a recently proposed deformable model based method, 3D Metamorphs (Huang and Metaxas, 2008, 2011), which uses a regular shape smoothness constraint. With this constraint, the protruding parts are shrunk, and the whole shape is attracted by the high internal gradient. Most regions are under-segmented. To quantitatively evaluate the accuracy of this

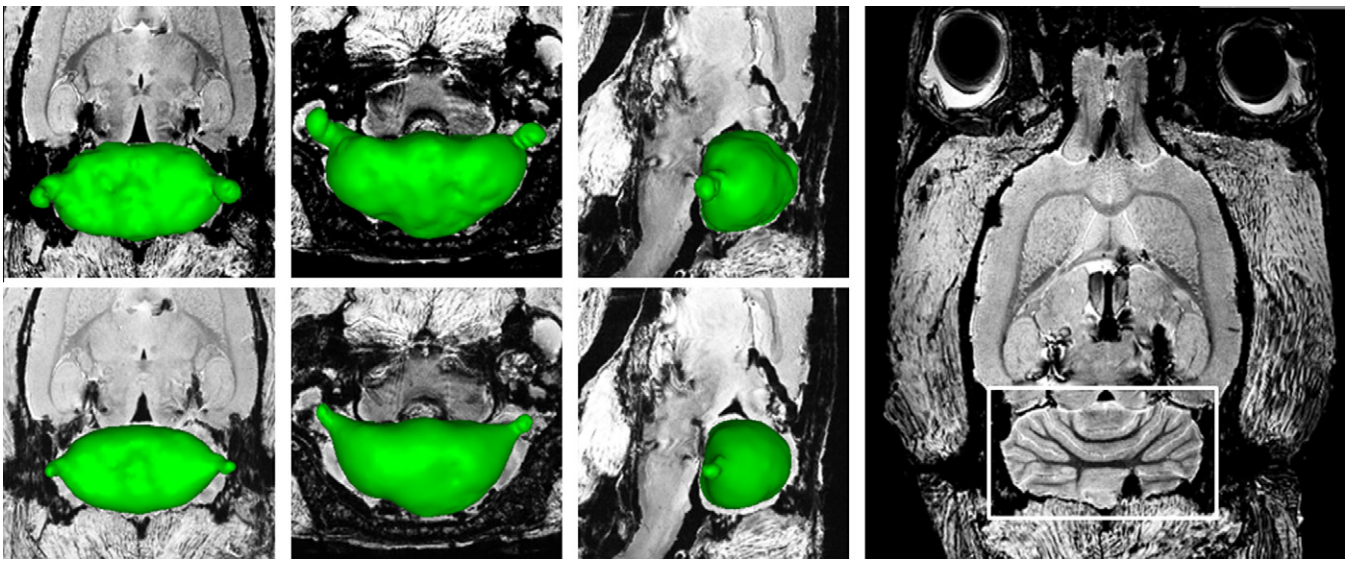

Fig. 11. Segmentation results of the rat cerebellum. The first row is from the proposed method. The second row is from the same framework but without shape prior constraint. The rightmost figure is one slice of the MR image with cerebellum highlighted.

3D segmentation task, we use the same measurement as in Section 4.2, i.e., the mean value and standard deviation of the distances between shape surfaces. The error of using regular shape smoothness constraint is 5.86 ± 3.68 voxels. 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 voxels.

5. Conclusions

In this paper, we proposed a robust and efficient deformable segmentation framework based on our previous work Sparse Shape Composition (SSC) (Zhang et al., 2012). Contrary to traditional shape models, SSC incorporates shape priors on-the-fly through sparse shape composition. It is able to handle gross errors, to model complex shape variations and to preserve shape details in a unified framework. Owing to the high computational cost of standard SSC, a straightforward integration might has low runtime efficiency, which impedes the segmentation system being clinically accepted. Therefore, we designed two strategies to improve the computational efficiency of the sparse shape composition. First, given a large number of training shape instances, K-SVD is used to learn a much more compact but still informative shape dictionary. Second, if shape instances have a large number of vertices, affinity propagation method is used to partition the surface into small sub-regions, on which the sparse shape composition is performed locally. Both strategies dramatically decrease the scale of the sparse optimization problem. An efficient deformable segmentation system that is robust to weak/misleading appearance cures is thereby realized.

We validate these two newly-proposed modules with three very diverse medical applications, 2D lung localization in X-ray, 3D liver segmentation in PET-CT, and 3D Rat Cerebellum Segmentation in MRM. The experimental results are promising and show the following facts:

1. Sparse shape composition is able to effectively model shape priors. Our segmentation system based this shape prior achieves better performance than several newly proposed segmentation systems (Shen et al., 2011; Zhan et al., 2009) which use the smoothness shape prior, Procrustes analysis, and PCA based shape prior.
2. When the number of training data is very large (e.g., thousands), it is infeasible to simply stack all shapes into the data matrix since sparse shape composition could not handle them efficiently. In this case, dictionary learning technique is employed to learn a compact dictionary, whose size is much smaller than the whole dataset. This compact dictionary highly improves the computational efficiency without sacrificing segmentation accuracy.
3. When the shape data contains thousands of vertices, solving the convex problem becomes time consuming. Multi-resolution scheme can decrease the computational time. However, it also ignores some information since it only uses a subset of the whole shape. We use mesh partitioning method to model local shape priors instead of the whole shape. In this way, we can decrease the running time and effectively model the local priors as well.
4. This proposed shape prior method is applied to handle problems related to diverse medical imaging modalities, e.g., X-ray, CT and MR. It shows good generality to different imaging modalities and shape dimensions. We expect to apply this method to more applications in the future.

The proposed method can be further extended in two directions. First, in practice, training shape instances often do not come in one batch but gradually. Therefore, online dictionary learning is worth investigating as we would rather *update* than *re-learn* the dictionary whenever a new training shape becomes available. Second, since the scope of shape modeling is beyond localization/segmentation, we plan to apply this proposed method to other medical applications, and integrate it into other methods, such as tracking (Zhou et al., 2010) and registration.

References

Aharon, M., Elad, M.,, Bruckstein, A., 2006. K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. In: IEEE Transaction Signal Processing, pp. 4311–4322.

Candes, E., Romberg, J., Tao, T., 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory 52, 489–509.

Chen, S., Billings, S.A., Luo, W., 1989. Orthogonal least squares methods and their application to non-linear system identification. In: International Journal of Control, pp. 1873–1896.

Chen, S.S., Donoho, D.L., Saunders, M.A., 2001. Atomic decomposition by basis pursuit. SIAM Journal on Scientific Computing 43, 129–159.
Chui, H., Rangarajan, A., 2003. A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding 89, 114–141.
Cootes, T., Taylor, C., Cooper, D., Graham, J., 1995. Active shape model – their training and application. Computer Vision and Image Understanding 61, 38–59.
Cootes, T.F., Taylor, C.J., 1997. A mixture model for representing shape variation. In: Image and Vision Computing, pp. 110–119.
Davatzikos, C., Tao, X., Shen, D., 2003. Hierarchical active shape models, using the wavelet transform. IEEE Transactions on Medical Imaging 22, 414–423.
Donoho, D.L., 2004. For most large underdetermined systems of equations, the minimal l1-norm near-solution approximates the sparsest near-solution. Technical Report, Communications on Pure and Applied Mathematics.
Donoho, D.L., 2006. Compressed sensing. IEEE Transactions on Information Theory 52, 1289–1306.
Donoho, D.L., Tsaig, Y., Drori, I., luc Starck, J., 2006. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Technical Report.
Duta, N., Sonka, M., 1998. Segmentation and interpretation of MR brain images: an improved active shape model. IEEE Transactions on Medical Imaging 17, 1049–1062.
Engan, K., Aase, S., Hakon Husoy, J., 1999. Method of optimal directions for frame design. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE, pp. 2443–2446.
Etyngier, P., Segonne, F., Keriven, R., 2007. Shape priors using manifold learning techniques. In: International Conference on Computer Vision, pp. 1–8.
Figueiredo, M., Nowak, R., Wright, S., 2007. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. IEEE Journal of Selected Topics in Signal Processing 1, 586–597.
Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24, 381–395.
Florin, C., Paragios, N., Funka-Lea, G., Williams, J., 2007. Liver segmentation using sparse 3D prior models with optimal data support. In: Information Processing in Medical Imaging, pp. 38–49.
Frey, B., Dueck, D., 2007. Clustering by passing messages between data points. Science 315, 972.
Goodall, C., 1991. Procrustes methods in the statistical analysis of shape. Journal of the Royal Statistical Society 53, 285–339.
Grant, M., Boyd, S., 2008. Cvx: Matlab software for disciplined convex programming.
Heimann, T., Meinzer, H.P., 2009. Statistical shape models for 3D medical image segmentation: a review. Medical Image Analysis 13, 543–563.
Huang, J., Zhang, S., Metaxas, D., 2011. Efficient MR image reconstruction for compressed MR imaging. Medical Image Analysis 15, 670–679.
Huang, X., Metaxas, D., 2008. Metamorphs: deformable shape and appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 1444–1459.
Kim, S., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D., 2007. An interior-point method for large-scale l1-regularized least squares. IEEE Journal of Selected Topics in Signal Processing 1, 606–617.
Langs, G., Paragios, N., Essafi, S., 2010. Hierarchical 3D diffusion wavelet shape priors. In: International Conference on Computer Vision, pp. 1717–1724.
Lekadir, K., Merrifield, R., Yang, G., 2007. G.-z.: outlier detection and handling for robust 3D active shape models search. IEEE Transactions on Medical Imaging 26, 212–222.
Lustig, M., Donoho, D., Pauly, J., 2007. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine 58, 1182–1195.
Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A., 2009. Non-local sparse models for image restoration. In: International Conference on Computer Vision, pp. 2272–2279.
Mallat, S., Zhang, Z., 1993. Matching pursuits with time-frequency dictionaries. In: IEEE Transaction Signal Processing, pp. 3397–3415.
Nahed, J.A., pierre Jolly, M., Yang, G., 2006. Robust active shape models: A robust, generic and simple automatic segmentation tool. In: International Conference on Medical Image Computing and Computer Assisted Intervention.
Nesterov, Y., Nemirovsky, A., 1994. Interior point polynomial methods in convex programming. Studies in Applied Mathematics 13, 1993.
Pohl, K., Warfield, S., Kikinis, R., Grimson, W., Wells, W., 2004. Coupling statistical segmentation and pca shape modeling. Medical Image Computing and Computer-Assisted Intervention, 151–159.
Rogers, M., Graham, J., 2002. Robust active shape model search. In: European Conference on Computer Vision, pp. 517–530.
Shen, D., Davatzikos, C., 2000. An adaptive-focus deformable model using statistical and geometric information. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 906–913.
Shen, T., Li, H., Huang, X., 2011. Active volume models for medical image segmentation. IEEE Transactions on Medical Imaging 30, 774–791.
Shi, Y., Qi, F., Xue, Z., Chen, L., Ito, K., Matsuo, H., Shen, D., 2008. Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. IEEE Transactions on Medical Imaging 27, 481–494.
Sjostrand, K. et al., 2007. Sparse decomposition and modeling of anatomical shape variation. IEEE Transactions on Medical Imaging 26, 1625–1635.
Tropp, J.A., 2004. Greed is good: Algorithmic results for sparse approximation. In: IEEE Transaction Information Theory, pp. 2231–2242.
Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y., 2009. Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 210–227.
Yan, P., Kruecker, J., 2010. Incremental shape statistics learning for prostate tracking in trus. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 42–49.
Yan, P., Xu, S., Turkbey, B., Kruecker, J., 2010. Discrete deformable model guided by partial active shape model for trus image segmentation. IEEE Transactions on Biomedical Engineering 57, 1158–1166.
Zhan, Y., Dewan, M., Harder, M., Krishnan, A., Zhou, X.S., 2011. Robust automatic knee mr slice positioning through redundant and hierarchical anatomy detection. IEEE Transactions on Medical Imaging 30, 2087–2100.
Zhan, Y., Dewan, M., Zhou, X.S., 2009. Cross modality deformable segmentation using hierarchical clustering and learning. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 1033–1041.
Zhan, Y., Zhou, X.S.Z., Peng, Z., Krishnan, A., 2008. Active scheduling of organ detection and segmentation in whole-body medical images. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 313–321.
Zhang, S., Huang, J., Huang, Y., Yu, Y., Li, H., Metaxas, D., 2010. Automatic image annotation using group sparsity. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 3312–3319.
Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D., Zhou, X., 2011a. Deformable segmentation via sparse shape representation. In: Fichtinger, G., Martel, A., Peters, T. (Eds.), Medical Image Computing and Computer-Assisted Intervention C MICCAI 2011, Lecture Notes in Computer Science, vol. 6892. Springer, Berlin/Heidelberg, pp. 451–458.
Zhang, W., Yan, P., Li, X., 2011b. Estimating patient-specific shape prior for medical image segmentation. In: International Symposium on Biomedical Imaging. IEEE, pp. 1451–1454.
Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D., Zhou, X., 2012. Towards robust and effective shape modeling: sparse shape composition. Medical Image Analysis 16, 265–277.
Zheng, Y., Gee, J., 2010. Estimation of image bias field with sparsity constraints. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 255–262.
Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J., Metaxas, D., 2012. Learning active facial patches for expression analysis. In: IEEE Conference on Computer Vision and Pattern Recognition.
Zhou, Y., Yeniaras, E., Tsiamyrtzis, P., Tsekos, N., Pavlidis, I., 2010. Collaborative tracking for MRI-guided robotic intervention on the beating heart. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010, 351–358.
Zhu, Y., Papademetris, X., Sinusas, A., Duncan, J., 2009. A dynamical shape prior for LV segmentation from RT3D echocardiography. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 206–213.
Zhu, Y., Papademetris, X., Sinusas, A., Duncan, J., 2010. Segmentation of the left ventricle from cardiac MR images using a subject-specific dynamical model. IEEE Transactions on Medical Imaging 29, 669–687.