

13

From point to surface: Hierarchical parsing of human anatomy in medical images using machine learning technologies

Y. Zhan¹, M. Dewan², S. Zhang³, Z. Peng¹, B. Jian⁴, X.S. Zhou¹

Siemens Healthcare, Malvern, PA, United States¹ Flipkart, Palo Alto, CA, United States²

University of North Carolina at Charlotte, Charlotte, NC, United States³ Google, Mountain View, CA, United States⁴

CHAPTER OUTLINE

13.1 Introduction	384
13.2 Literature Review	385
13.3 Anatomy Landmark Detection	386
13.3.1 Learning-Based Landmark Detection	386
13.3.2 Application: Automatic Localize and Label of Vertebrae.....	389
13.4 Detection of Anatomical Boxes	392
13.4.1 Robust Anatomical Box Detection Using an Ensemble of Local Spatial Configurations.....	392
13.4.2 Application: Autoalignment for MR Knee Scan Planning	393
13.5 Coarse Organ Segmentation	395
13.5.1 Sparse Shape Composition for Organ Localization.....	395
13.5.2 Lung Region Localization in Chest X-Ray	398
13.6 Precise Organ Segmentation	401
13.6.1 Deformable Segmentation Using Hierarchical Clustering and Learning.....	401
13.6.2 Liver Segmentation in Whole-Body PET-CT Scans	406
13.7 Conclusion	408
References	408

13.1 INTRODUCTION

In various medical imaging applications, from body-part identification to lung nodule detection, one critical task is to localize and interpret specific anatomical structures in medical images. For an experienced professional, this task is often accomplished by leveraging the “anatomy signatures” in medical images—the unique shape or appearance characteristics of the anatomies of interest. In fact, during clinical training and practice, a medical professional learns many of these “anatomy signatures” from seeing a large number of medical images. While computer-aided algorithms have become more and more popular in the medical imaging domain, it is not surprising that the research community has become interested in developing algorithms to automatically localize and interpret different anatomies in medical images. The success of these algorithms is essentially dependent on if/how many anatomy signatures can be well extracted by the algorithms. While earlier studies often crafted specific image filters to extract anatomy signatures, a more recent research trend shows the prevalence of learning-based approaches for two reasons: (1) Machine learning technologies have matured to solve real-world problems. (2) More and more medical image datasets are available to facilitate the “exploration” of complex medical imaging data.

In this chapter, we will present a learning-based framework that is able to “learn” anatomy signatures and localize anatomical structures in medical images. As shown in Fig. 13.1, this framework has a hierarchical structure and parses human anatomies in a coarse-to-fine fashion. Specifically, we start from the smallest anatomical

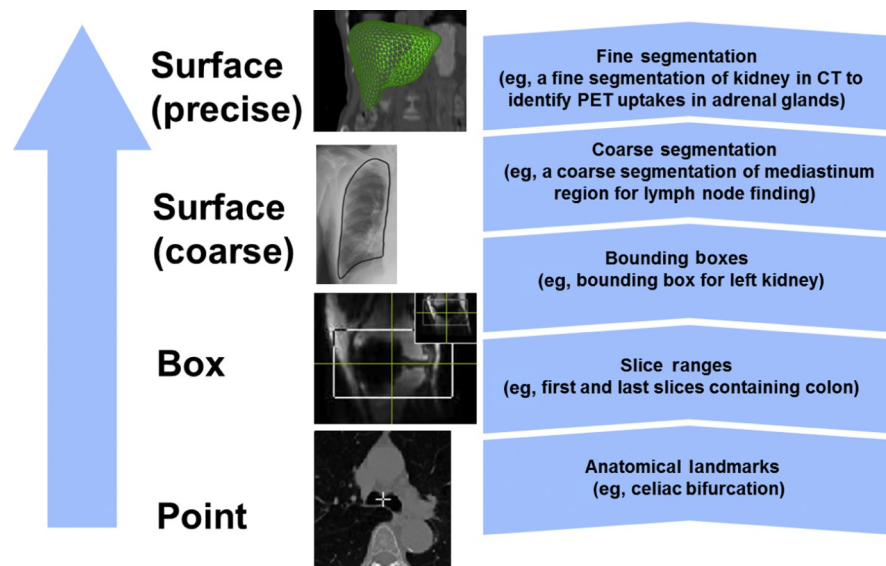


FIG. 13.1

Infrastructure of our anatomy interpretation system.

entity—anatomical landmarks. Our system is then extended to the detection of anatomy boxes (region of interests), coarse organ segmentation, and finally reaches the precise delineation of organ boundaries. On one hand, algorithms developed on each layer may benefit different image analysis applications, for example, bounding box detection of knee meniscus helps to automate magnetic resonance (MR) imaging planning. On the other hand, algorithms developed on lower layers also provides intermediate results for the higher layers. For example, landmark detection provide basic image appearance cues for coarse organ segmentation. It is worth noting that since algorithms in this framework aim to extract anatomy signatures through learning, they are not restricted to specific organ characteristics. Thus the entire framework is scalable to different imaging modalities and organ systems.

The rest of this chapter is organized as follows. We start in [Section 13.2](#) by reviewing some related works. In [Sections 13.3–13.6](#), we present algorithms from landmark detection to precise organ segmentation. In each of these sections, we start by introducing the methods followed by some relevant medical imaging use cases. Conclusions are drawn in [Section 13.7](#).

13.2 LITERATURE REVIEW

Since our framework includes anatomy parsing at different levels, from landmarks to organ surfaces, the literature review will cover all these different aspects.

Since anatomical landmarks are used in various clinical use cases, automatic landmark detection has gained significant interests in the medical imaging research community. While traditional methods aim to design filters or templates for specific landmarks ([Betke et al., 2003](#)), more recent studies have used machine learning technologies for a solution scalable to different landmarks in different imaging modalities. Representative works include [Liu and Zhou \(2012\)](#) and [Criminisi et al. \(2011\)](#), which employ probabilistic boosting tree and random forests to learn the appearance characteristics of landmarks.

Automatic detection of anatomical boxes has also been extensively investigated. A key problem is how to address the appearance variability resulting from different object orientations. In [Schneiderman and Kanade \(2000\)](#), object detectors for different poses are exclusively performed and the pose corresponding to the detector with the strongest response is regarded as the aligned pose. [Lv et al. \(2000\)](#) employs local orientation analysis to estimate object poses. Recently, marginal space learning (MSL) ([Zheng et al., 2008](#)) has shown promise in estimating rough transformations of anatomical structures in medical images. The basic idea is to decompose the transformation parameters into location, orientation, scale, etc., and train detectors to estimate each of them sequentially. Since this method treats the target anatomy (eg, a heart chamber or the knee menisci) as a whole, the burden of learning is high, especially for the first round of learning (location detectors).

The success of coarse organ segmentation relies on the modeling of shape priors. Following the pioneering use of the active shape model ([Cootes et al., 1995](#)), various

methods have been proposed to improve the effectiveness of shape prior modeling in different circumstances. For example, Cootes and Taylor (1997) and Etyngier et al. (2007) aims to model the nonlinear shape variations through a mixture of Gaussians and manifold learning. Yan et al. (2010) tried to use partial active shape model (ASM) to make the shape model robust to missing boundaries. Sjostrand and et.al. (2007) employed a sparse principal component analysis (PCA) to obtain sparser modes and produce near-orthogonal components. Davatzikos et al. (2003) divides the shape model into several independently modeled parts for hierarchical shape representation. Both of these works aim to capture small shape variations and preserve shape details.

For precise organ segmentation, deformable models have been extensively studied, particularly in the area of medical image segmentation. The widely recognized potency of deformable models comes from their ability to segment anatomic structures by exploiting constraints derived from the image data (bottom-up) together with prior knowledge about these structures (top-down). The deformation process is usually formulated as an optimization problem whose objective function consists of an external (image) term and an internal (shape) term. While internal energy is designed to preserve the geometric characteristics of the organ under study, the external energy is defined to move the model toward organ boundaries. Traditionally, the external energy term usually comes from edge information (Xu and Prince, 1998), for example, image gradient. In recent years, more effort has been invested on the integration of other image features, for example, local regional information (Vese and Chan, 2002; Ronfard, 1994) and texture models (Huang et al., 2004). By combining different image features as the external energy, deformable models have achieved tremendous success in various clinical practices. Machine learning technologies have opened the door to a more generic external energy design. By using learning-based methods, boundary characteristics can be learned from training data (Zhan and Shen, 2006; Zheng et al., 2008). In this way, the “design” of external energy becomes data driven and can be extended to different imaging modalities.

13.3 ANATOMY LANDMARK DETECTION

13.3.1 LEARNING-BASED LANDMARK DETECTION

Anatomical landmarks are biologically meaningful points existing in various organ systems. They are defined to describe the morphological characteristics of anatomical structures and facilitate communication between scientists in the fields of biology and medicine, etc. In the medical imaging community, while most anatomical landmarks become visible in vivo, they play important roles in the interpretation of medical images. In the same way that geographic landmarks guide travelers in exploring the earth, anatomical landmarks provide guidance to navigate the medical images. For example, anatomical content within a medical image can be determined by the locations of anatomical landmarks. In addition, some anatomical landmarks also

provides critical clues to diagnose diseases. For example, the posterior junction point of a lumbar vertebra and the spinal cord can be used to diagnose spondylolisthesis and evaluate the disease stage. Hence algorithms that can automatically detect anatomical landmarks can directly benefit various clinical use cases. Besides the direct impact on clinical use cases, automatic detection of anatomical landmarks also paves the way for other medical image analysis tasks. For example, automatically detected landmarks can be used to initialize deformable models for organ segmentation (Zhan et al., 2009; Zhang et al., 2012). They can also provide an initial transformation for image registration (Zhan et al., 2011).

Due to the complex appearance of different anatomical landmarks, we use a learning-based approach for individual landmark detection. Because of its data-driven nature, learning-based approaches also make our method highly scalable to different imaging modalities, for example, computed tomography (CT), MR, positron emission tomography (PET), ultrasound, etc. We formulate landmark detection as a voxel-wise classification problem. Specifically, voxels close to the landmark are considered as positive samples and voxels away from the landmark are regarded as negative ones. To learn a specific landmark detector, we first annotate the landmark in a set of training images. The positive and negative samples/voxels are determined based on their distances to the annotated landmark. For each training sample (voxel), a set of elementary features is extracted in its neighborhood. Our elementary features are generated by a set of mother functions, $\{H_l(\mathbf{x})\}$, extended from the Haar wavelet basis. As shown in Eq. (13.1) and Fig. 13.2, each mother function consists of one or more 3D rectangle functions with different polarities:

$$H(\mathbf{x}) = \sum_{i=1}^N p_i R(\mathbf{x} - \mathbf{a}_i), \quad (13.1)$$

where polarities $p_i = \{-1, 1\}$,

$$R(\mathbf{x}) = \begin{cases} 1, & \|\mathbf{x}\|_\infty \leq 1, \\ 0, & \|\mathbf{x}\|_\infty > 1 \end{cases}$$

denotes rectangle functions, and \mathbf{a}_i is the translation.

By scaling the mother functions and convoluting them with the original image, a set of spatial-frequency spaces are constructed:

$$F_l(\mathbf{x}, s) = H_l(s\mathbf{x}) * I(\mathbf{x}), \quad (13.2)$$

where s and l denote the scaling factor and index of mother functions respectively.

Finally, for any voxel $\mathbf{x}_0 \in \mathcal{R}^3$, its feature vector $\mathfrak{F}(\mathbf{x}_0)$ is obtained by sampling these spatial-frequency spaces in the neighborhood of \mathbf{x}_0 (Eq. 13.3). It provides cross-scale appearance descriptions of voxel \mathbf{x}_0 .

$$\mathfrak{F}(\mathbf{x}_0) = \bigcup_{l=1 \dots L} \{F_l(\mathbf{x}_i, s_j) | \mathbf{x}_i \in \mathbb{N}(\mathbf{x}_0), s_{\min} < s_j < s_{\max}\}. \quad (13.3)$$

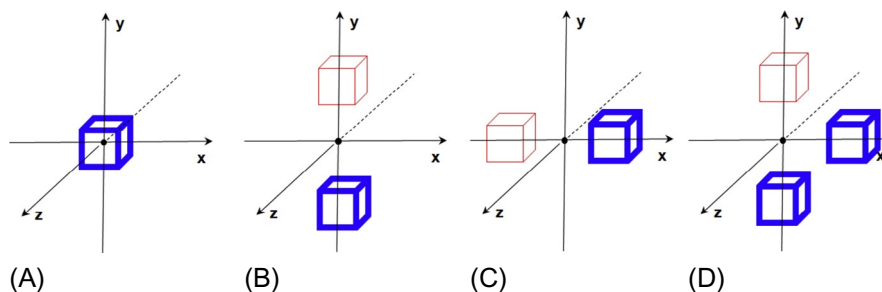


FIG. 13.2

Some examples of Haar-based mother functions.

Compared to the standard Haar wavelet, the mother functions we employed are not orthogonal. However, they provide more comprehensive image features to characterize different anatomy primitives. For example, as shown in Fig. 13.2, mother function (A) potentially works as a smoothing filter, which is able to extract regional features. Mother functions (B) and (C) can generate horizontal or vertical “edginess” responses, which are robust to local noises. More complicated mother functions like (D) are able to detect “L-shaped” patterns, which might be useful to distinguish some anatomy primitives. In addition, our features can be quickly calculated through integral images (Crow, 1984). This paves the way for an efficient landmark detection system.

All elementary features are then fed into a cascade classification framework (Viola and Jones, 2004), as shown in Fig. 13.3. The cascade framework is designed to address the highly unbalanced positive and negative samples. In fact, since only voxels around landmarks under study are positives and all other voxels are negatives, the ratio of positives to negatives is often less than $1:10^5$. In the training stage, all positives and a small proportion of negatives are used at every cascade level. The training algorithm is “biased” to positives, such that each positive has to be correctly classified but the negatives are allowed to be misclassified. These misclassified negatives, that is, false positives, will be further used to train the following cascades. At run-time, while positives are expected to go through all cascades, most negatives can be rejected in the first several cascades and do not need further evaluation.

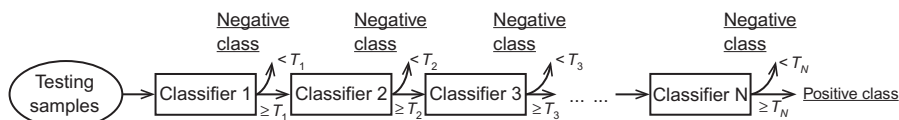


FIG. 13.3

Schematic explanation of cascade Adaboost classifiers.

In this way, the run-time speed can be dramatically increased. In our study, we use *Adaboost* (Freund and Schapire, 1997) as the basic classifier in the cascade framework. The output of the learned classifier $\mathcal{A}(\mathcal{F}(\mathbf{x}))$ indicates the likelihood of the specific landmark appearing at \mathbf{x} .

At run-time, a sliding window approach is employed to apply the learned classifiers on each voxel in the image. The voxel that has the highest response of $\mathcal{A}(\mathcal{F}(\mathbf{x}))$ can be considered as the detected landmark.

13.3.2 APPLICATION: AUTOMATIC LOCALIZE AND LABEL OF VERTEBRAE

The spine is one of the major organs in the human body. It includes the vertebral column and spinal cord. A human vertebral column typically consists of 33 vertebrae. Twenty-four of them are articulating (7 cervical, 12 thoracic, and 5 lumbar vertebrae) and nine of them are fused vertebrae in the sacrum and the coccyx. As the spine is strongly correlated to both neural and skeletal systems, various neurological, orthopedic, and oncological studies involve investigations of spinal anatomies. In addition, due to the strong spatial correlations between specific vertebrae and their surrounding organs, the spine may also be used as a vertical reference framework to describe the locations of other organs in the trunk, for example, the transpyloric plane. In spine image analysis, localization and labeling of vertebrae are often the first steps, and are tedious and time consuming for manual operators. Accordingly, we employ the landmark detection algorithm to automatically localize and label vertebrae. (Besides the algorithms introduced in this section, we also treat different vertebrae as “anchor” and “bundle” and train vertebrae detectors in a hierarchical way. In addition, an articulated model is used to model the spine geometry. Please refer to Zhan et al. (2012) for more details.)

Our method is evaluated on both CT and MR datasets. The CT dataset includes 189 randomly selected CT cases with partial or whole spine coverage. A wide range of imaging parameters are used for this dataset, including different reconstruction kernels and slice thicknesses. Our MR dataset includes 300 T1-weighted 3D MR scout scans. These scout scans have relatively low but isotropic resolution of 1.7 mm and large fields of view. These datasets come from different clinical sites and were generated by different types of Siemens MR Scanners (Avanto 1.5T, Verio 3T, Skyra 3T, etc.). Both MR and CT datasets include partial or whole spines.

The automatic vertebrae labeling results were shown to experienced radiologists and rated as “perfect” (no manual editing required), “acceptable” (minor manual editing required), and “rejected” (major manual editing required).

As shown in Table 13.1, our method achieves “perfect” detection in 95+% of cases. It is worth noting that our testing set includes CT/MR scans with severe diseases or imaging artifacts. However, our method can still detect vertebrae and discs robustly. Figs. 13.4 and 13.5 show the results of our method on challenging CT and

Table 13.1 Evaluation of Spine Detection in CT and MR Scans

	Number of Cases	Perfect	Acceptable	Reject
CT	189	180 (95.2%)	6 (3.2%)	3 (1.6%)
MR	300	293 (97.7%)	4 (1.3%)	3 (1.0%)

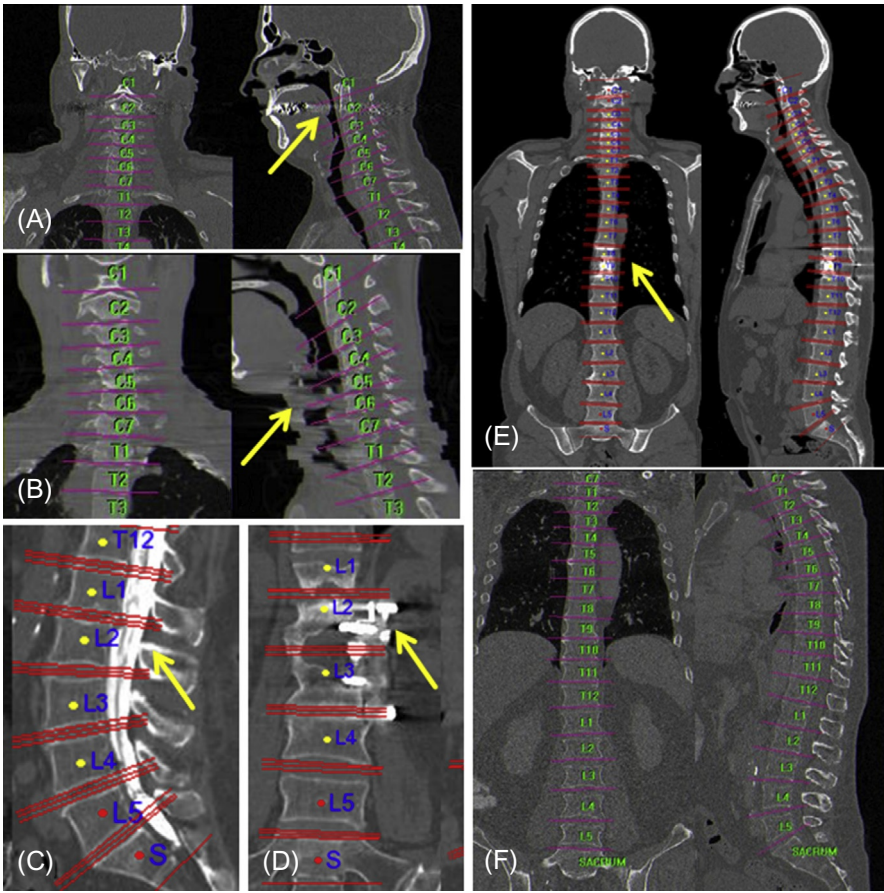
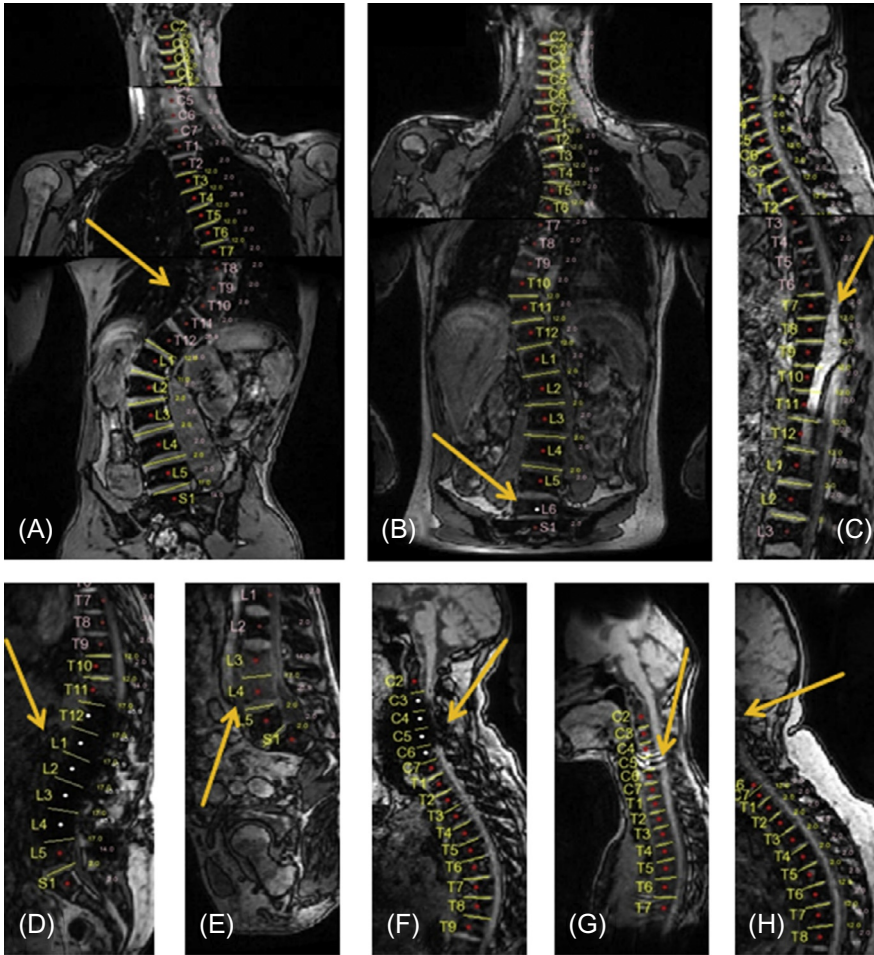


FIG. 13.4

Examples of vertebrae labeling results in challenging CT scans. (A) C-spine scan with metal artifacts. (B) C-spine scan with motion artifacts. (C) L-spine scan with spinal cord disease. (D) L-spine scan with metal implant. (E) Whole-spine scan with metal artifacts. (F) Whole-spine scan with large imaging noises.

**FIG. 13.5**

Examples of vertebrae labeling results in challenging MR scans. (A) Whole-spine scan with strong scoliosis. (B) Whole-spine scan with congenital abnormality (six lumbar vertebrae). (C) Whole-spine scan with folding artifact. (D) L-spine scan with metal implant. (E) L-spine scan with vertebra pathology. (F) C-spine scan with metal artifact. (G) C-spine scan with ring artifact. (H) C-spine scan where anchor vertebra is out of field of view.

MR scans. As shown in these cases, our method is robust to different kinds of imaging artifacts (Figs. 13.4A, B and 13.5C, G), large imaging noises (Figs. 13.4F), metal implants (Figs. 13.4D, E and 13.5D, F), severe scoliosis (Fig. 13.5A), pathologies (Figs. 13.4C and 13.5E) congenital abnormality (Fig. 13.5B) and scans that have anchor vertebrae out of field of view (Fig. 13.5H).

13.4 DETECTION OF ANATOMICAL BOXES

13.4.1 ROBUST ANATOMICAL BOX DETECTION USING AN ENSEMBLE OF LOCAL SPATIAL CONFIGURATIONS

An anatomical box is one level beyond anatomical landmarks. It is usually defined by a rectangle (2D) or a cuboid (3D) that contains anatomy of interest and is oriented according to specific anatomical structures, for example, knee meniscus plane, femoral neck axis, intervertebral discs, etc. Autodetection of anatomical boxes is necessary to improve radiological workflow. For example, MR imaging processes often consist of scout scanning and high-resolution scanning stages. In between these two stages, technicians need to manually position high-resolution slice group boxes in scout scans. Appropriate positioning of the slice group is very critical to the quality of the high-resolution MR images (more details will be presented in [Section 13.4.2](#)). If an algorithm can automatically detect these high-resolution slice group boxes, which are essentially anatomical boxes, the speed and quality of MR imaging process can be significantly improved.

Since an anatomical box is often defined by anatomical landmarks, it is natural to derive it based on the landmarking technology introduced in [Section 13.3](#). However, a straightforward extension of landmarking may not be enough to derive the anatomical box robustly. For example, in theory, a cuboid can be derived by four noncoplanar points. If we invoke the four landmark detectors independently to derive the box, assuming the detection rate of each landmark is as high as 95% (which is quite high given the variability of diseases and imaging artifacts in medical images), the detection rate of the box falls to 81%. To achieve high robustness, we go beyond the “necessary” number of landmarks and employ “more than enough” landmarks to derive anatomical boxes. The key principle here is to detect “redundant” landmarks and prune the erroneous detections using a spatial model of multiple landmarks.

In contrast to the active shape model that learns global shape statistics, we propose to learn distributed spatial configuration models, that is, for each landmark p_i , we aim to learn its spatial relations with other landmarks in a group-wise fashion. Assume p_i is the landmark under study, then $U(\mathbf{p} \setminus p_i)$ is a subset of landmarks which does not contain p_i , that is, $U(\mathbf{p} \setminus p_i) \subset \{\mathbf{p} \setminus p_i\}$. The group-wise spatial configuration between p_i and $U(\mathbf{p} \setminus p_i)$ is modeled as a conditional probability following a multivariate Gaussian distribution:

$$s(p_i | U(\mathbf{p} \setminus p_i)) = \frac{1}{(2\pi)^{(3/2)|\Sigma|^{1/2}}} \exp\left(-\frac{1}{2}(p_i - \mu)^T \Sigma^{-1}(p_i - \mu)\right), \quad (13.4)$$

where μ and Σ are two statistical coefficients that are learned as follows.

We employ a linear model to capture the spatial correlation between p_i and $U(\mathbf{p} \setminus p_i)$:

$$p_i = \mathcal{C} \cdot \mathbb{U}, \quad (13.5)$$

where \mathbb{U} is a vector concatenated by $\{p_j | p_j \in U(\mathbf{p} \setminus p_i)\}$ and \mathfrak{C} denotes the linear correlation matrix. Given a set of training samples, \mathfrak{C} can be learned by solving a least squares problem. Furthermore, μ and Σ are calculated as

$$\begin{aligned}\mu &= E[\mathfrak{C} \cdot \mathbb{U}] \\ \Sigma &= E[(\mathfrak{C} \cdot \mathbb{U} - \mu)(\mathfrak{C} \cdot \mathbb{U} - \mu)^T].\end{aligned}\tag{13.6}$$

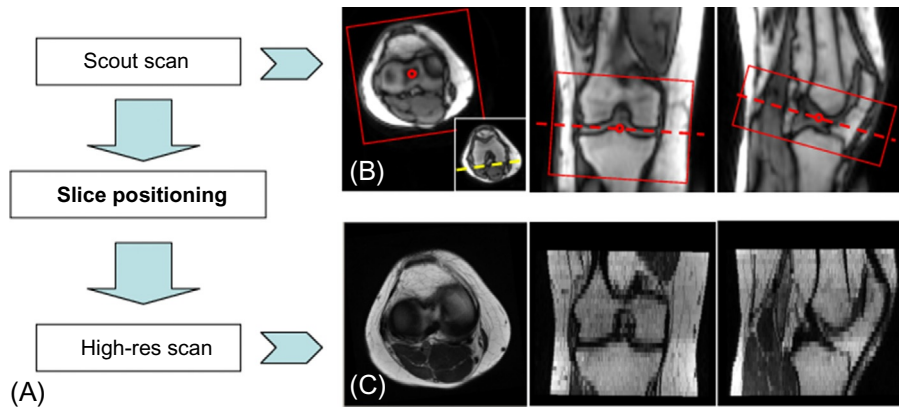
At run-time, we will first detect a “redundant” set of landmarks. Each landmark p_i will then receive a set of “votes” from other landmark groups $U(\mathbf{p} \setminus p_i)$ (Eq. 13.4) based on their relative spatial locations. It is worth noting that a correctly detected p_i may still receive a low conditional probability from a subset $U(\mathbf{p} \setminus p_i)$ that includes erroneous detections. In contrast, the conditional probabilities received by an erroneously detected p_i will *all* be low. (The only exception is that the erroneous detections happen to construct a correct spatial configuration, which very rarely happens in reality.) Therefore the maximum value of all votes received by p_i is used to measure the “eligibility” of the spatial location of the p_i . A landmark p_i whose maximum “vote” value is less than a threshold will be considered as an erroneous detection and excluded from the detected landmark set. This process is iteratively conducted until erroneous detections are gradually “peeled” away.

The advantage of our local spatial correlation modeling is twofold. First, instead of learning global spatial statistics, we learn spatial correlations within small groups of landmarks. At run-time, the decisions from these distributed models are assembled in a “democratic” way. This makes our method robust to missing or gross detection failures. Even when only a minority of landmarks are correctly detected, they will form a consistent clique to robustly remove erroneous detections. Second, by constraining the cardinality of U_j and using a linear model, our spatial model will not overfit erroneous detections. The run-time efficiency is also guaranteed in this way.

13.4.2 APPLICATION: AUTOALIGNMENT FOR MR KNEE SCAN PLANNING

Magnetic resonance imaging (MRI) has been successfully used to diagnose various knee diseases ranging from acute knee injuries to chronic knee dysfunction, for example, ligament tears, patella dislocation, meniscal pathology, and arthropathies (Ostlere, 2007). The inherent imaging physics and speed limitations of MR typically constrain the diagnostic MR images to have isotropic high resolution in 3D. Instead, the diagnostic MR image is a 2.5D modality with high in-slice resolution and low through-slice resolution (Fig. 13.6C). Hence their diagnostic quality is highly dependent upon the positioning accuracy of the slice groups (imaging planes). Good centering and orientation ensures that the anatomy of interest is optimally captured within the high-resolution (high-res) imaging plane.

Traditionally, this achieved by acquiring 2D scout images (called “localizers”): technicians can then plan out high-res slice groups based on relevant anatomies visible in these scouts. Recently, the 3D knee scout scan has been introduced to

**FIG. 13.6**

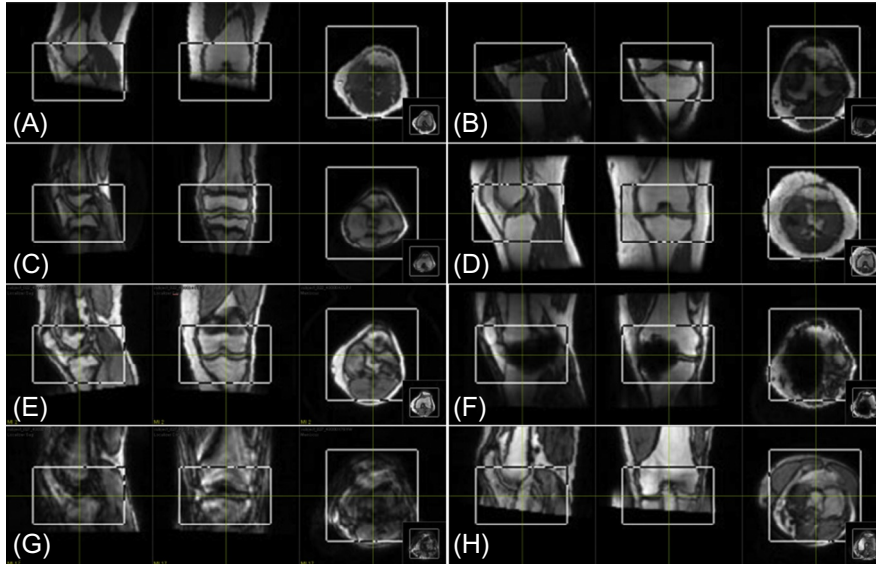
MR slice positioning using scout scans. (A) Workflow of knee MR scans. (B) An MR knee scout scan. Rectangular boxes: coverage of MR slice group. Circles: imaging centers. Yellow dashed line (in the small thumbnail) defines the in-plane rotation. (C) A typical high-resolution diagnostic MR knee transversal scan.

improve the quality of the workflow. A 3D knee scout scan has low but isotropic resolution. Although it might not be of diagnostic quality, it provides complete 3D context, which enables a human operator, or a computer, to plan out all required high-res slice groups without any additional scout scans. Fig. 13.6 shows an example of how to plan a high-res slice group using the 3D scout in a reproducible way: to image the menisci, the transversal slice group should be parallel to the meniscus plane, and the in-plane rotation is determined by aligning with the line connecting the two lower edges of the femoral condyle (the yellow dashed line).

Technically, the slice positioning problem is equivalent to detecting a 3D double oblique anatomical box based on several anatomical landmarks. Therefore we employed the aforementioned anatomical box detection framework for the auto-slice positioning. (Besides the aforementioned algorithms, hierarchical learning is also used in the system. Please refer to Zhan et al. (2011) for more details.)

We evaluate our method on 744 knee scout scans. The alignment results were visually checked and evaluated by two experienced professionals as “accurate (AC)”: accurate positioning that is acceptable by technicians; “reasonable (RE)”: reasonable good positioning but still has the space to improve; and “gross failure (GF)”. Our method can achieve “AC” in 736(98.9%) cases, and has only 1(0.1%) “GF”. As shown in Fig. 13.7, our system is able to achieve robust knee alignment even for challenging cases, for example, severe bone diseases, metal implants, etc.

We also quantitatively evaluate the accuracy of the alignment using the detected landmarks on 50 cases with manual alignment. As shown in Table 13.2, the errors are very limited and our system can satisfy clinical requirements.

**FIG. 13.7**

Robust slice positioning of our algorithm on “stress testing” cases. Each part shows the three orthogonal multiplanar reconstructions (MPRs) of the aligned volume at the center of the knee meniscus slice box: (A) tibia out of field of view, (B) femur out of field of view, (C) a pediatric patient, (D) an old patient, (E) a patient with bone disease, (F) a patient with metal implant, (G) motion artifacts, and (H) folding artifacts.

Table 13.2 Quantitative Errors (Average and Standard Deviation) of Our Alignment Method Compared to Manual Alignment

	Knee Meniscus	Femur Cartilage	Patella Cartilage
Trans. (mm)	0.93 ± 0.27	1.53 ± 0.42	1.53 ± 0.38
Rot. tra>>cor (°)	0.83 ± 0.31	1.99 ± 0.60	2.01 ± 0.77
Rot. sag>>tra (°)	0.54 ± 0.28	0.72 ± 0.25	3.14 ± 1.16
Rot. sag>>cor (°)	0.63 ± 0.27	0.82 ± 0.35	1.35 ± 0.52

Rot. tra>>cor, sag>>tra and sag>>cor denote the rotation angles from transversal to coronal, from sagittal to transversal, and from sagittal to coronal respectively.

13.5 COARSE ORGAN SEGMENTATION

13.5.1 SPARSE SHAPE COMPOSITION FOR ORGAN LOCALIZATION

While the detection of anatomical boxes benefits radiological workflow in many ways, some use cases require more precise interpretation of anatomical characteristics—organ shapes. For example, the localization of lung regions in

chest X-ray images (not necessary to be very precise) not only directly assists clinicians, for example, auto-measurement of cardiac-thoracic ratio, but also helps following intelligent algorithms, for example, reduction of false positives of lung CAD (computer aided detection).

Organ localization in medical images (in the remainder of this chapter, “coarse organ segmentation” and “organ localization” refer to the same thing by default) requires both appearance and shape information. In our framework, the appearance cues come from the learning-based landmark detectors introduced in [Section 13.3](#). These pretrained detectors essentially capture local appearance cues and automatically identify anatomical landmarks at run-time. Based on these detected landmarks, an organ shape model can be transformed to the subject image under study and considered as a coarse organ segmentation. The remaining question is how to effectively model shape priors and make the organ localization robust to missing/misleading local appearance cues (see [Fig. 13.8](#)).

Effective modeling of shape priors faces three challenges: (1) complex shape variation cannot always be modeled by a parametric probability distribution; (2) local image appearance cues (input shape) may have gross errors, for example, detected lung landmarks may be wrong; and (3) local details of the input shape are difficult to preserve if they are not statistically significant in the training data.

We propose a sparse shape composition model (SSC) to address the aforementioned challenges in a unified framework. Given that a shape repository consists of a number of annotated shape instances, instead of explicitly learning shape priors from these shapes offline, we propose to adaptively approximate the input shape on-the-fly, by a sparse linear combination of a subset of shapes in the shape repository. Hence the shape prior constraint is implicitly applied. Our method is inspired by recently proposed sparsity theories in the compressive sensing community, that is, the

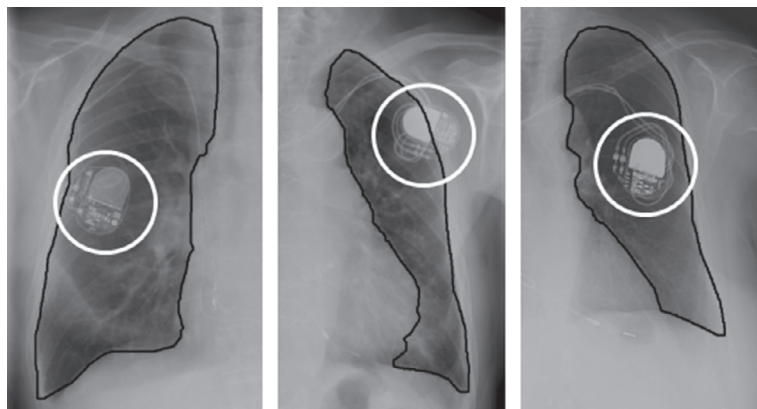


FIG. 13.8

Chest X-ray images with annotated boundaries. The appearance cue is misleading because of the instruments in the marked regions.

problem of computing sparse linear representations with respect to an overcomplete dictionary of base elements (Candes and Tao, 2006; Donoho, 2004). It has been successfully applied in many computer vision applications, such as, but not limited to, face recognition (Wright et al., 2009), image restoration (Mairal et al., 2009), and background subtraction (Huang et al., 2009).

Our method is designed based on two “sparsity” observations. First, given a large enough training dataset (repository), an instance can be approximately represented by a sparse linear combination of instances in the shape repository. Similarly, in our application each given shape is approximated by a sparse linear combination of annotated shapes. Without any assumption of a parametric distribution model (eg, 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 detailed information even if the detail of the input shape is only present in a small proportion of the shape repository and is not statistically significant. Second, the given shape information may contain gross errors resulting from the wrong detections, but such errors are often very sparse, for example, 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 a sparse learning problem, and efficiently solve it by an expectation-maximization (EM) type of framework.

In the SSC model, we employ explicit parametric shape representation to model a shape instance, that is, 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 training repository 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. Here, all d_i , $i = 1, 2, 3, \dots, k$, are prealigned using generalized Procrustes analysis (Goodall, 1991). $y \in \mathbb{R}^n$ 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 combination coefficients or weights.

To incorporate the two “sparsity” observations, we derive x by minimizing the loss function as

$$\begin{aligned} \arg \min_{x, e, \beta} \|T(y, \beta) - Dx - e\|_2^2, \\ \text{s.t. } \|x\|_0 \leq k_1, \|e\|_0 \leq k_2. \end{aligned} \quad (13.7)$$

Here, $T(y, \beta)$ is a global transformation operator with parameter β . It aligns the input shape y to the mean shape of existing data D . $\|x\|_0 \leq k_1$ defines the first sparsity constraint, that is, an instance can be approximately represented by a sparse linear combination of instances in the shape repository. $\|e\|_0 \leq k_2$ corresponds to the

second sparsity constraint; the given shape information may contain gross errors, but such errors are often very sparse.

The constraints in Eq. (13.7) 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 of capturing 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 problem through L^1 norm relaxation (Starck et al., 2004). Thus Eq (13.7) is reformulated as

$$\arg \min_{x,e,\beta} \|T(y, \beta) - Dx - e\|_2^2 + \lambda_1 \|x\|_1 + \lambda_2 \|e\|_1, \quad (13.8)$$

where λ_1 and λ_2 respectively control how sparse x and e are. After relaxation, $\lambda_1 \|x\|_1 + \lambda_2 \|e\|_1$ is nonsmooth but continuous and convex. Eq. (13.8) is our objective function of our proposed SSC. The deviation from Eq. (13.7) to Eq. (13.8) relaxes the absolute sparseness constraints of the objective function (L^0 norm to L^1 norm). From the shape modeling perspective, we might use more shape instances for shape composition by optimizing Eq. (13.8). However, since this deviation converts an NP-hard problem to a continuous and convex optimization problem which can be solved efficiently using an expectation-maximization optimization framework (refer to (Zhang et al., 2012) for more details).

It is interesting to look into Eq. (13.8) 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 Eq. (13.8) 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 .

13.5.2 LUNG REGION LOCALIZATION IN CHEST X-RAY

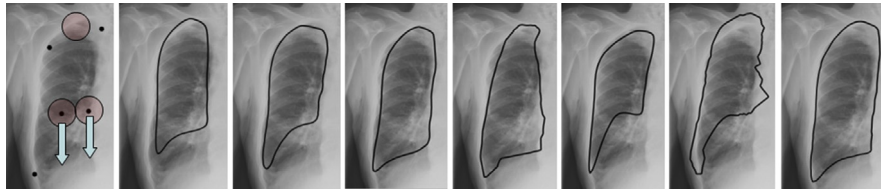
Radiography (X-ray) is the most frequently used medical imaging modality due to its fast imaging speed and low cost. About one-third of radiograph exams are

chest radiographs. It is used to reveal various pathologies including abnormal cardiac sizes, pneumonia shadow, and mass lesions. The automation of pathology detection often requires robust and accurate lung segmentation. The major challenges of lung segmentation in radiography come from large variations of lung shapes, lung disease, and pseudo-boundary close to the diaphragm. 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 shape inference. Of 367 X-ray images (all images are from different patients), 200 are used as training data, and the other 167 are used for testing purposes. In this study, we select training samples to ensure a good coverage of different ages and genders (according to information from DICOM header). The number of training samples is determined empirically. The ground truths are binary masks of manual segmentation results. A 2D contour is extracted from each mask. To obtain the landmarks for training purposes, we manually select six specific points (eg, 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, a shape prior is necessary to infer a shape from them. When applying this model, we constantly use the same parameter values for all X-ray images, that is, $\lambda_1 = 50$ and $\lambda_2 = 0.15$.

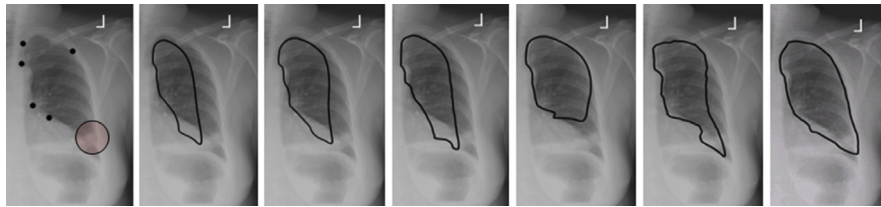
In this study, we compare the proposed sparsity-based shape prior modeling with other state-of-the-art methods listed below:

1. PA: Procrustes analysis is used to find a similarity transformation to fit a mean shape to detected landmarks.
2. SMS: This is the shape model search module in ASM, which employs the PCA method to refine the input shape. Note that we are not using the entire ASM framework including boundary detection and iterative fitting.
3. R-SMS: The shape model search step in the robust ASM (Rogers and Graham, 2002) method uses the RANSAC framework to remove the influence of erroneous detections.
4. SI-NN: This stands for shape inference using k nearest neighbors. It is similar to Georgescu et al. (2005), 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.
5. TPS: Thin-plate-spline (Bookstein, 1989) is used to deform the mean shape to fit detected landmarks.
6. SSC*: This is a variant of the proposed SSC algorithm without modeling e .

Some representative and challenging cases are shown in Figs. 13.9–13.11. In Fig. 13.9, there are some misdetections 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* also fails to handle such non-Gaussian errors since e is not modeled. SSC can successfully capture such misdetections points in e and generate a reasonable shape. In Fig. 13.10,

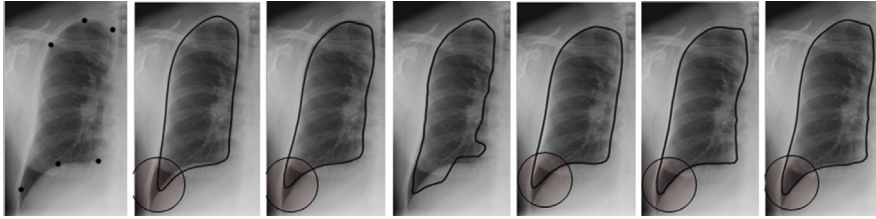
**FIG. 13.9**

Comparisons of 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 correct positions). (B) Similarity transformation from Procrustes analysis. (C) Shape model search module in ASM, using a PCA-based method. (D) Shape model search in robust ASM, using RANSAC to improve the robustness. (E) Shape inference method using nearest neighbors. (F) Thin-plate-spline. (G) Sparse representation without modeling e . (H) The proposed method by solving Eq. (13.8).

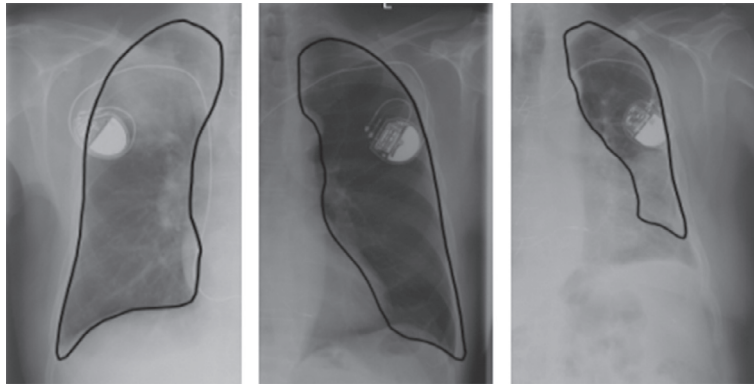
**FIG. 13.10**

Comparisons of 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 in Fig. 13.9.

the underlying shape of the lung is special and different from most other lung shapes (see the mean shape in Fig. 13.10). Furthermore, there is a missing point. Neither a transformed mean shape nor its variations can represent such a 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. 13.11, all six detections are correct. However, the shape's details are not preserved using the mean shape or its variations. Both SSC* and SSC discover more detailed information than the other methods. Thus a sparse linear combination is sufficient to recover such details even when the gross error e is not modeled. Fig. 13.12 shows some results from our proposed method for challenging cases with medical instruments. The shape prior contributes to the stability of the system. It still generates reasonable results with such misleading appearance cues (Table 13.3).

**FIG. 13.11**

Comparisons of right lung localization. All six detections are roughly accurate. Thus there is no gross error. The regions marked by circles show the difference of preserved details. Compared methods are the same as in [Fig. 13.9](#).

**FIG. 13.12**

Some localization results from our proposed method on challenging cases with medical instruments. Note that the localized shape may not be exactly on the boundary, since the shape module does not use image information. However, such results are good enough for the input of a CAD program or initialization of segmentation algorithms.

13.6 PRECISE ORGAN SEGMENTATION

13.6.1 DEFORMABLE SEGMENTATION USING HIERARCHICAL CLUSTERING AND LEARNING

Although rough organ boundaries/shapes from coarse segmentation approaches can benefit multiple use cases, they are still one step from precise delineation of organ boundaries, which are important for quantitative studies. Therefore, at the final level of our anatomical parsing system, we developed a precise organ segmentation algorithm based on a deformable model.

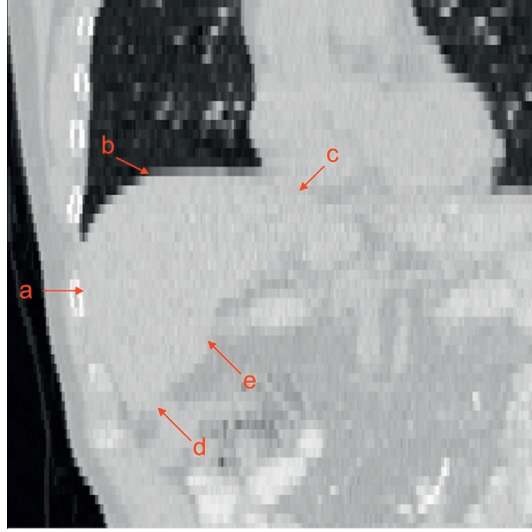
Table 13.3 Quantitative Comparisons of Seven Methods

	Fig. 13.9			Fig. 13.10			Fig. 13.11		
Method	P	Q	DSC	P	Q	DSC	P	Q	DSC
PA	62	99	76	50	99	64	93	99	94
SMS	66	99	78	61	99	72	93	99	95
R-SMS	81	99	88	61	99	72	93	99	95
SI-NN	81	99	87	63	98	73	87	99	90
TPS	59	99	74	75	99	79	97	98	94
SSC*	63	98	71	73	99	79	97	99	96
SSC	87	99	91	92	99	91	98	99	96

The sensitivity (P%), specificity (Q%), and Dice Similarity Coefficient (DSC%) are reported for the cases in Figs. 13.9–13.11.

Deformable modeling is a vigorously studied model-based approach in the area of medical image segmentation. The widely recognized potency of deformable models comes from their ability to segment anatomic structures by exploiting constraints derived from the image data (bottom-up) together with prior knowledge about these structures (top-down). The deformation process is usually formulated as an optimization problem whose objective function consists of an external (image) term and an internal (shape) term. While internal energy is designed to preserve the geometric characteristics of the organ under study, the external energy is defined to describe the appearance characteristics of the organ under study and drive the model toward organ boundaries. Recently, machine learning technologies have opened the door to a more generic external energy design. By using learning-based methods, boundary characteristics can be learned from training data (Zhan and Shen, 2006). In other words, the originally hand-crafted external energy is derived in a data-driven fashion and can be extended to different imaging modalities and organ systems. A potential problem is that the boundary characteristics of organs may show heterogeneous characteristics along organ boundaries. As shown in Fig. 13.13, the appearance characteristics of liver-lung boundaries is quite different from those of liver-heart boundaries. Hence, it is very challenging to learn the organ boundary characteristics using a single classifier. To address this problem, a “divide-and-conquer” strategy is needed. More specifically, the deformable model should be decomposed into a set of subsurfaces with relatively similar boundary characteristics (Fig. 13.14).

To this end, we propose a learning-based hierarchical deformable model. A key hallmark of our model is that its hierarchical structure is constructed through an iterative clustering and feature selection method. As shown in Fig. 13.15, every node of the hierarchical structure represents a subsurface of the deformable model. For each primitive subsurface, that is, leaf nodes in the hierarchical tree, a boundary

**FIG. 13.13**

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

detector is learned using a cascade boosting method. The ensemble of these learned boundary detectors actually captures the appearance characteristics of a specific organ in a specific imaging modality. Their responses guide the deformable model to the desired organ boundary.

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

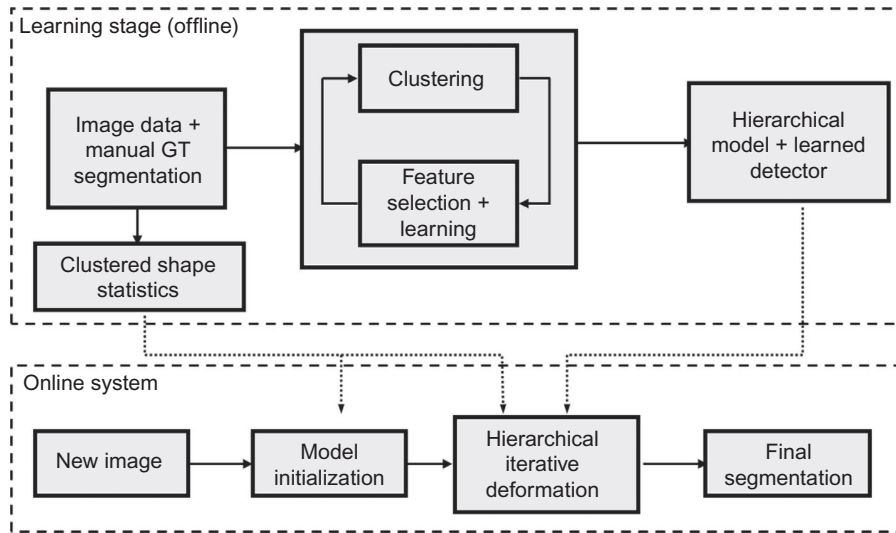
$$E(S) = E_{\text{ext}}(S) + E_{\text{int}}(S) = \sum_{i=1}^N E_{\text{ext}}(v_i) + E_{\text{int}}(S), \quad (13.9)$$

where E_{ext} and E_{int} are the image (external) energy term and shape (internal) energy term. A hierarchical deformation strategy is employed to solve this high-dimensional optimization problem (refer to [Zhan and Shen \(2006\)](#) for details).

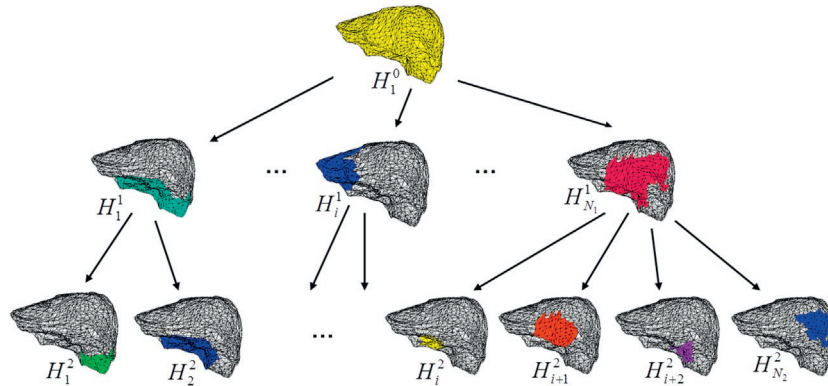
The external energy is defined by the responses of a set (or ensemble) of boundary detectors built upon the hierarchical deformable model. The following steps are used to generate the hierarchical model and the boundary detectors.

13.6.1.1 Affinity propagation clustering

The construction of model hierarchy is equivalent to the clustering of vertices. In this study, we employ “affinity propagation,” a generic clustering method proposed by

**FIG. 13.14**

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

**FIG. 13.15**

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

[Frey and Dueck \(2007\)](#). The affinity propagation method models each data point as a node in a network. During the clustering process, real-valued messages are recursively exchanged between data points until a high quality set of exemplars and corresponding clusters emerges.

In the “affinity propagation” framework, the quality of clustering is determined by the affinity between data points—the similarity between vertices in our study. We design the similarity based on two rationales. First, to facilitate the characterization of the heterogeneous boundary, vertices in the same cluster should have relatively similar image features. Second, the hierarchical deformable model requires the vertices within a cluster to be proximal to each other on the surface. In this way, the cluster center can be treated as a “driving vertex” and drive its neighborhood in the deformation process. Mathematically, the similarity between vertices is defined as follows:

$$s(v_i, v_j) = 1 - (1/K) \sum_{k=1}^K [\alpha G(v_i^k, v_j^k) + (1 - \alpha) C(\mathcal{F}(v_i^k), \mathcal{F}(v_j^k))]. \quad (13.10)$$

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

13.6.1.2 Iterative feature selection/clustering

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

$$H_i^j = \bigcup_{k=i_1}^{i_{N_i}} H_k^{j+1} \text{ and } \bigcap_{k=i_1}^{i_{N_i}} H_k^{j+1} = \emptyset. \quad (13.11)$$

The remaining problem is the selection of appropriate $\mathcal{F}(\cdot)$ in Eq. (13.10). This is actually an “egg-and-chicken” problem. On one hand, to achieve the desired clusters, we need to know the distinctive feature sets for boundary description. On the other hand, distinctive features for a local boundary can be obtained only after we have the vertex cluster. To address this problem, we propose an iterative clustering and feature selection method.

For the first level of cluster, we use the intensity profile along the normal of the vertices as $\mathcal{F}(\cdot)$. After that, assume $H_i^j = \{v_l\}$, and we use the *Adaboost* method to select the features that are most powerful to distinguish $\{v_l\}$ from the points along their normal directions, both inside and outside of the surface. The selected feature set is used as $\mathcal{F}(\cdot)$ in Eq. (13.10) to further cluster $\{v_l\}$ to a set of subclusters $\{H_k^{j+1}, k = i_1, \dots, i_{N_i}\}$. Feature selection and clustering are iteratively executed until boundary characteristics within a cluster become learnable.

13.6.1.3 Learn boundary detectors

For each primitive cluster, that is, the leaf node of the hierarchical tree, a boundary detector is learned to characterize the local boundary. We use the similar

Haar-like features and a cascade *Adaboost* method as our landmarking method (see Section 13.3) to learn a boundary detector. Given an image I , $\mathfrak{F}(\mathbf{x}; I)$ denotes the redundant feature vector of \mathbf{x} . (In practice, we use 2D, 3D, or 4D Haar-like features depending on the dimensionality of different image modalities.) In the run-time system, each learned classifier generates a boundary probability map $P(\mathbf{x}|I)$. Hence, the external energy term in Eq. (13.9) is defined as

$$E_{\text{ext}}(v_i) = 1 - P(v_i|I) = 1 - C_{h_{v_i}}(\mathfrak{F}(v_i; I)), \quad (13.12)$$

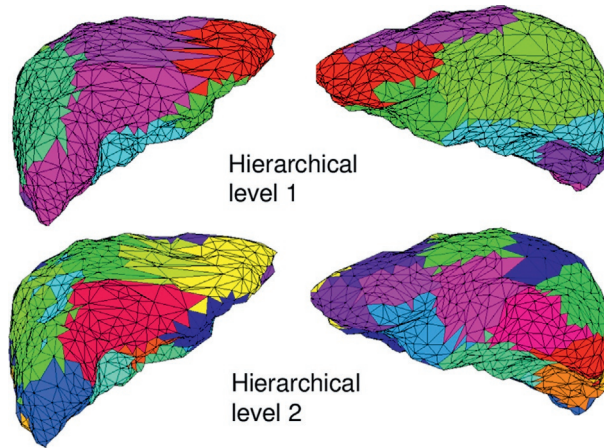
where h_{v_i} is the cluster index of v_i and C_h defines the corresponding classifier.

13.6.2 LIVER SEGMENTATION IN WHOLE-BODY PET-CT SCANS

PET-CT is a medical imaging modality that combines PET and an X-ray CT. As the acquired PET and CT images are inherently co-registered, PET-CT provides fused morphological and functional information, which potentially benefits various medical studies. After the first PET-CT prototype was introduced to clinical practice in 1998, PET-CT has triggered a revolution in image-based diagnosis for cancer patients. Many clinical studies reported that PET-CT has superior diagnostic value than mono-modalities, for example, CT (Shim et al., 2005), MR (Antoch et al., 2003) and PET (Bar-Shalom et al., 2003), and separate dual-modalities, for example, PET+morphological image (Pelosi et al., 2004).

In PET-CT-based cancer diagnosis, ^{18}F -fluoro-deoxyglucose (FDG) is the most widely used tracer. Since glucose utilization is known to be enhanced in many malignant tissues, uptake of FDG is usually considered as an indicator of malignance. In particular, the standardized uptake value (SUV) is employed to provide semiquantitative interpretation of malignance (Thie, 2004). However, glucose utilization is not entirely specific to malignant tissues. Normal tissues, such as the cerebral cortex, left ventricular myocardium and renal system, may also have high FDG uptake. A more troublesome fact is that FDG uptake shows large variations across different organs due to different glucose utilization. This is one of the major pitfalls of PET-CT interpretation. For example, while a spot in a lung with $\text{SUV} = 3.0$ is a highly suspicious lesion, a spot in the liver with the same level of SUV is completely normal.

An effective solution is to segment PET-CT and interpret it in an organ-specific fashion. Based on different physiological characteristics of organs, radiologists can use different thresholds/strategies to detect hot spots in different organs. The organ-specific PET-CT interpretation presumes the organs to be investigated have been segmented. Since the liver is one of the most important organs in oncology studies, we apply the learning-based hierarchical model to segment the liver in whole-body PET-CT scans. It is worth noting that the whole-body CT often has larger slice thickness and low dose (lower contrast), both of which make the segmentation task difficult. We trained the learning-based hierarchical model using 20 whole-body PET-CT scans with manually delineated liver surfaces. (To jointly exploit CT-PET information, 4D Haar-like filters are used for feature extractor (Zhan et al., 2008).) As shown in Fig. 13.16, the generated model has two hierarchical levels with 8 and

**FIG. 13.16**

3D rendering of the hierarchical structure of a liver model. Color patches denote the vertices belonging to the same cluster. Left: anterior view. Right: posterior view.

25 vertex clusters respectively. The automatic segmentation results on 30 testing datasets (PET: $5\text{ mm} \times 5\text{ mm} \times 5\text{ mm}$; CT: $1.3\text{ mm} \times 1.3\text{ mm} \times 5\text{ mm}$) are compared with manually delineated organ surfaces. Accuracy measurements include median distance between surfaces, average distance between surfaces, volume difference, and volume overlap difference. In Table 13.4, we compare our proposed method (Method 1) with Method 2, which is a learning-based deformable model with *heuristically* designed hierarchical structure. More specifically, in Method 2 the hierarchical structure is determined by clustering neighboring vertices, based only on geodesic information.¹ Therefore the spatially clustered vertices in Method 2 might have larger appearance variation, which is difficult to learn. Hence Method 2 shows inferior performance.

Table 13.4 Quantitative Comparison of Learning-Based Methods on PET-CT Liver Segmentation

	Med Surf. Dist. (voxel)	Avg Surf. Dist. (voxel)	Vol. Diff. (%)	Overlap Diff. (%)
Method 1	0.84	1.01	3.13	7.61
Method 2	1.27	1.61	5.16	12.1

Method 1: Our hierarchical deformable model. Method 2: A heuristically designed hierarchical deformable model.

¹Note that Method 2 only differs from Method 1 in the way their hierarchical structures are built.

13.7 CONCLUSION

In this chapter, we introduced a framework to parse human anatomies in medical images. Our framework is designed in a coarse-to-fine fashion to localize and identify human anatomies from landmarks to precise organ boundaries. In each layer, different machine learning technologies, including Adaboost, sparse representation, iterative clustering, etc., are used to learn “anatomical signatures.” In this way, our framework can be extended to diverse clinical applications in different image modalities and can be easily adapted to new applications. Although the different layers independently benefit different clinical use cases, they also help each other from a technical point of view. For example, the landmarking algorithm provides appearance cues for coarse organ segmentation and coarse organ segmentation paves the way for precise segmentation.

We have prototyped different applications based on this framework. Some representative examples are presented in this chapter. These applications are very diverse, covering different imaging modalities (CT, MR, PET, radiography, etc.) and clinical fields (oncology, orthopedics, neurology, etc.). The preliminary results show that our methods are able to achieve robust and accurate results and can potentially benefit clinical workflow in various ways.

REFERENCES

- Antoch, G., Vogt, F.M., Freudenberg, L.S., Nazaradeh, F., Goehde, S.C., Barkhausen, J., Dahmen, G., Bockisch, A., Debatin, J.F., Ruehm, S.G., 2003. Whole-body dual-modality PET/CT and whole-body MRI for tumor staging in oncology. *J. Am. Med. Assoc.* 290, 3199–3206.
- Bar-Shalom, R., Yefremov, N., Guralnik, L., Gaitini, D., A.Frenkel, Kuten, A., Altman, H., Keidar, Z., Israel, O., 2003. Clinical performance of PET/CT in evaluation of cancer: additional value for diagnostic imaging and patient management. *J. Nucl. Med.* 44 (8), 1200–1209.
- Betke, M., Hong, H., Thomas, D., Prince, C., Ko, J.P., 2003. Landmark detection in the chest and registration of lung surfaces with an application to nodule registration. *Med. Image Anal.* 7 (3), 265–281.
- Bookstein, F.L., 1989. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (6), 567–585.
- Candes, E.J., Tao, T., 2006. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory* 52 (12), 5406–5425.
- Cootes, T.F., Taylor, C.J., 1997. A mixture model for representing shape variation. In: *Image and Vision Computing*, pp. 110–119.
- Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape model—their training and application. *Comput. Vision Image Understanding* 61, 38–59.
- Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E., 2011. Regression forests for efficient anatomy detection and localization in CT studies. In: *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*. Springer, New York, pp. 106–117.

- Crow, F.C., 1984. Summed-area tables for texture mapping. In: ACM SIGGRAPH Computer Graphics 18.3, 207–212.
- Davatzikos, C., Tao, X., Shen, D., 2003. Hierarchical active shape models, using the wavelet transform. *IEEE Trans. Med. Imaging* 22 (3), 414–423.
- Donoho, D.L., 2004. For most large underdetermined systems of equations, the minimal ℓ_1 -norm near-solution approximates the sparsest near-solution. *Commun. Pure Appl. Math.* 59, 797–829.
- Etyngier, P., Segonne, F., Keriven, R., 2007. Shape priors using manifold learning techniques. In: *International Conference on Computer Vision*, pp. 1–8.
- Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* 55, 119–139.
- Frey, B.J., Dueck, D., 2007. Clustering by passing messages between data points. *Science* 315, 972–976.
- Georgescu, B., Zhou, X.S., Comaniciu, D., Gupta, A., 2005. Database-guided segmentation of anatomical structures with complex appearance. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 429–436.
- Goodall, C., 1991. Procrustes methods in the statistical analysis of shape. *J. R. Stat. Soc.* 53, 285–339.
- Huang, X., Metaxas, D., Chen, T., 2004. Metamorphs: deformable shape and texture models. In: *Computer Vision and Pattern Recognition 2004*, pp. 496–503.
- Huang, J., Huang, X., Metaxas, D., 2009. Learning with dynamic group sparsity. In: *International Conference on Computer Vision*, pp. 64–71.
- Liu, D., Zhou, S.K., 2012. Anatomical landmark detection using nearest neighbor matching and submodular optimization. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2012*. Springer, New York, pp. 393–401.
- Lv, X., Zhou, J., Zhang, C.S., 2000. A novel algorithm for rotated human face detection. In: *Computer Vision and Pattern Recognition 2000*.
- 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.
- Ostlere, S., 2007. Imaging the knee. *Imaging* 15, 217–241.
- Pelosi, E., Messa, C., Sironi, S., Picchio, M., Landoni, C., Bettinardi, V., Gianolli, L., Maschio, A.D., Gilardi, M.C., Fazio, F., 2004. Value of integrated PET/CT for lesion localisation in cancer patients: a comparative study. *Eur. J. Nucl. Med. Mol. Imaging* 31 (7), 1619–1707.
- Rogers, M., Graham, J., 2002. Robust active shape model search. In: *European Conference on Computer Vision*, pp. 517–530.
- Ronfard, R., 1994. Region-based strategies for active contour models. *Int. J. Comput. Vision* 13 (2), 229–251.
- Schneiderman, H., Kanade, T., 2000. A statistical method for 3d object detection applied to faces and cars. In: *Computer Vision and Pattern Recognition novel algorithm for rotated human face detection 2000*.
- Shim, S.S., Lee, K.S., Kim, B.T., Chung, M.J., Lee, E.J., Han, J., Choi, J.Y., Kwon, O.J., Shim, Y.M., Kim, S., 2005. Nonsmall cell lung cancer: prospective comparison of integrated FDG PET/CT and CT alone for preoperative staging. *Radiology* 236, 1011–1019.
- Sjostrand, K., et.al., 2007. Sparse decomposition and modeling of anatomical shape variation. *IEEE Trans. Med. Imaging* 26 (12), 1625–1635.
- Starck, J.L., Elad, M., Donoho, D.L., 2004. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Trans. Image Process.* 14, 1570–1582.

- Thie, J., 2004. Understanding the standardized uptake value, its methods, and implications for usage. *J. Nucl. Med.* 45 (9), 1431–1434.
- Vese, L.A., Chan, T.F., 2002. A multiphase level set framework for image segmentation using the Mumford and Shah model. *Int. J. Comput. Vision* 50 (3), 271–293.
- Viola, P., Jones, M.J., 2004. Robust real-time face detection. *Int. J. Comput. Vision* 57, 137–154.
- Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., 2009. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2), 210–227.
- Xu, C., Prince, J.L., 1998. Snakes, shapes, and gradient vector flow. *IEEE Trans. Image Process.* 7 (3), 359–369.
- Yan, P., Xu, S., Turkbey, B., Kruecker, J., 2010. Discrete deformable model guided by partial active shape model for trus image segmentation. *IEEE Trans. Biomed. Eng.* 57 (5), 1158–1166.
- Zhan, Y., Shen, D., 2006. Deformable segmentation of 3-d ultrasound prostate images using statistical texture matching method. *IEEE Trans. Med. Imaging* 25, 256–272.
- Zhan, Y., Peng, Z., Zhou, X., 2008. Towards organ-specific PET-CT interpretation: generic organ segmentation using joint PET-CT information. In: *MICCAI2008 Workshop on Analysis of Functional Medical Images*.
- Zhan, Y., Dewan, M., Zhou, X.S., 2009. Cross modality deformable segmentation using hierarchical clustering and learning. In: *Medical Image Computing and Computer-Assisted Intervention*, pp. 1033–1041.
- 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 Trans. Med. Imaging* 30, 2087–2100.
- Zhan, Y., Dewan, M., Harder, M., Zhou, X.S., 2012. Robust MR spine detection using hierarchical learning and local articulated model. In: *Medical Image Computing and Computer-Assisted Intervention*, pp. 141–148.
- Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S., 2012. Towards robust and effective shape modeling: sparse shape composition. *Med. Image Anal.* 16, 265–277.
- Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D., 2008. Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans. Med. Imaging* 27.