

# Segmenting Lung Fields in Serial Chest Radiographs Using Both Population-Based and Patient-Specific Shape Statistics

Yonghong Shi, Feihu Qi, Zhong Xue, Liya Chen, Kyoko Ito, Hidenori Matsuo, and Dinggang Shen\*

**Abstract**—This paper presents a new deformable model using both population-based and patient-specific shape statistics to segment lung fields from serial chest radiographs. There are two novelties in the proposed deformable model. First, a modified scale invariant feature transform (SIFT) local descriptor, which is more distinctive than the general intensity and gradient features, is used to characterize the image features in the vicinity of each pixel. Second, the deformable contour is constrained by both population-based and patient-specific shape statistics, and it yields more robust and accurate segmentation of lung fields for serial chest radiographs. In particular, for segmenting the initial time-point images, the population-based shape statistics is used to constrain the deformable contour; as more subsequent images of the same patient are acquired, the patient-specific shape statistics online collected from the previous segmentation results gradually takes more roles. Thus, this patient-specific shape statistics is updated each time when a new segmentation result is obtained, and it is further used to refine the segmentation results of all the available time-point images. Experimental results show that the proposed method is more robust and accurate than other active shape models in segmenting the lung fields from serial chest radiographs.

**Index Terms**—Deformable model, scale invariant feature transform (SIFT) local descriptor, segmentation, serial chest radiographs, shape statistics.

## I. INTRODUCTION

CHEST radiograph is a very popular diagnostic modality, and it provides sufficient pathological information about cardiac size, pneumonia-shadow, and mass-lesions, with low

Manuscript received June 6, 2007; revised August 10, 2007. Asterisk indicates corresponding author.

Y. Shi is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200030, China and with the Digital Medical Research Center, Fudan University, Shanghai 200032, China (e-mail: shi-yh@cs.sjtu.edu.cn).

F. Qi is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200030, China (e-mail: fhqi@cs.sjtu.edu.cn).

Z. Xue is with the Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: zhong.xue@uphs.upenn.edu).

L. Chen is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200030, China (e-mail: liyachen@sjtu.edu.cn).

K. Ito and H. Matsuo are with Hidaka Hospital, 886 Nakaomachi, Takasaki Gunma-Prefecture 370-0001, Japan (e-mail: itokyon@da3.so-net.ne.jp; hidakadc@cocoa.ocn.ne.jp).

\*D. Shen is with the Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, 300 Market Street, Suite 380, Philadelphia, PA 19104-0366 USA (e-mail: dinggang.shen@uphs.upenn.edu).

Color versions of one or more of the figures and tables in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2007.908130

cost and high reproducibility [1]–[5]. For example, the evaluation of monthly changes of cardiac size by taking serial chest radiographs remains crucial for the treatment of hemodialysis patients. This is because the evaluation of a patient's dry weight (DW) is usually based on the cardiac size measured during each dialysis session [2], [3], and the miscalculation of DW often results in a fatal illness such as patient's death. Since segmentation of the lung fields in chest radiographs provides a means to measure the cardiac size, it is important to accurately segment the lung fields from regularly captured serial chest radiographs of each patient.

Many methods have been proposed in the literature [1], [6]–[10], [15]–[22] to segment lung fields from posterior–anterior chest radiographs, and they can be classified into four categories: 1) rule-based segmentation methods have been used to detect the outline of ribcage or the diaphragm [6], [7]; 2) pixel-based methods were proposed to classify each pixel of an image into either lung field or background based on a multiscale filter bank of Gaussian derivatives and a K-NN classifier [8], [9]; 3) hybrid methods were formulated by combining rule-based methods and pixel-based classifications for lung field segmentation [10]; and 4) deformable model-based methods, such as active shape model (ASM) and active appearance model (AAM) [11]–[14], have been successfully applied in lung field segmentation [8].

Among these four categories of the segmentation algorithms, the active shape model developed by Cootes *et al.* [11]–[14] was a prosperous starting point because of its ability to incorporate *a priori* information extracted from a training set and its flexibility to represent object shapes. In ASM, the statistics of image intensities and gradients along the profiles of contour points (or a local gray-level appearance model for each contour point) are used to drive the contour toward the boundary of the object, and PCA-based shape statistics is used to constrain the contour. Many improvements have been made to improve the performance of the matching strategy. For example, Seghers *et al.* [16] proposed the simultaneous optimization of shape and gray-level appearance models based on noniterative dynamic programming. In this algorithm, an object shape is described by multiple landmark-specific statistical models that capture local dependencies between adjacent landmarks on the shape. Gleason *et al.* [17] had developed a probabilistic-based deformable model which simultaneously optimizes a single objective function generated by the global and local shape model, as well as the gray-level appearance model. Here, the local shape model is introduced to preserve the local shape

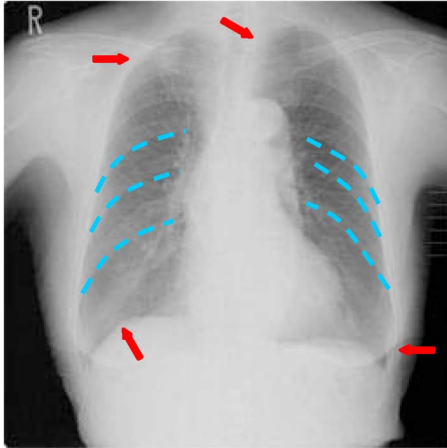


Fig. 1. Example of the lung fields in a chest radiograph. Red arrows indicate that the edge features are inconsistent across different locations. Light blue lines indicate the repeated patterns on rib boundaries. .

information around each user-defined critical landmark points. Other researches suggest that the local gray-level appearance model should be created by using some optimal features, in order to drive contour points to desired object boundaries. For example, van Ginneken *et al.* [18] determined the distinct set of optimal features by means of machine learning. Therefore, it is worth considering the following two key strategies. First, distinctive image features around each contour point should be used in order to accurately and robustly match the deformable contour onto the object in an image. Second, both global and local shape statistics should be used in order to capture the shape variability more accurately. These two strategies are both used to design the segmentation method proposed in this paper.

The first strategy aims to characterize rich image features around each point on the boundaries of lung fields and then use these image features to facilitate correspondence detection during the segmentation of serial lung fields. Notice that generic image features along the boundaries of lung fields are not always consistent; for example, the outline of ribcage is much fuzzier compared to that of diaphragm, as indicated by red arrows in Fig. 1. Also there are some repeated patterns like ribs in the lung fields (e.g. light blue dotted lines in Fig. 1). All of these render generic features like edges not sufficiently distinctive to discriminate rib boundaries from lung field boundaries. Ideally, a distinctive local descriptor should be built for each point on lung field boundary in order to reliably differentiate it from other boundary points, thus helping detect correspondences during the deformable segmentation procedure. In the literature, a number of local descriptors have been designed to characterize structures around lung field boundaries. For example, the profile of normalized first-order derivatives along the normal direction of the lung field boundary has been used in [8] and [11]. However, there is no *a priori* reason why this should be an optimal choice. Thus, optimal local features based on machine learning were extensively studied. For example, a general set of local image structure descriptors, such as the moments of local histograms extracted from filtered versions of the images using a filter bank of Gaussian derivatives [18], were considered. Also, statistical

analysis methods have been applied to learning the local descriptors with the most informative features, i.e., optimal local features in [18]. On the other hand, various features have been designed in computer vision, such as local spatial histograms [23], scale invariant feature transform (SIFT) [24], PCA-SIFT [25], steerable filters [26], differential invariants [27], complex filters [28], and moment invariants [29]. In our method, SIFT is used to capture image features for guiding the lung field segmentation, since it has been validated as one of the best performing feature sets among many local descriptors in a recent comparative study [30].

The second strategy aims to use both population-based and patient-specific shape statistics to adaptively constrain the deformable model. Because of the distortions in the serial images due to the mapping of the differently posed 3-D body to the 2-D images, as well as the anatomical changes of lung fields due to the development of disease or health recovery, a deformable contour is used to match the lung field boundary and to adapt to these shape variations. Different from the general snake method that uses smoothness constraint to regularize the shape of the contour, shape statistics trained from a number of contours can be used to constrain the deformable matching procedure. In this work, we used two kinds of shape statistics, the *population-based* and the *patient-specific* shape statistics. The former is trained from the manually marked lung field contours from a population using PCA method, and the latter is obtained from the segmentation results of a specific patient available during serial lung field segmentation procedure. In particular, for segmenting initial time-point images of a new patient, the population-based shape statistics takes the role to constrain the deformable contour; as more and more images of the same patient are acquired and segmented, it is possible to train a patient-specific shape statistics from the available segmentation results, and gradually use this patient-specific statistics to guide the lung field segmentation procedure of future images of the same patient. While the patient-specific shape statistics takes more roles in the segmentation as more patient data are collected, the population-based shape statistics will take less roles. Each time when the segmentation result of a new image of the patient is obtained, the patient-specific shape statistics is updated and is then used to refine the segmentation results of all the previous time-point images. Effectively, the deformable model gradually adapts to the shape statistics of a patient, and hence yields more accurate segmentation results, since the patient-specific shape statistics reflects the shape variability of the same patient more accurately than the population-based shape statistics. All of these will be detailed in the next section.

The rest of the paper is organized as follows. Section II introduces the proposed deformable model in detail. The performance of this deformable model is validated in Section III, and this paper concludes in Section IV.

## II. METHOD

### A. Description

In the clinical setting, serial chest radiographs are regularly acquired from each patient at consecutive time-points (i.e., each month). Our goal is therefore to accurately segment the lung

fields from the serial images and provide measures such as cardiac size for quantitative estimation of DW. Two major steps are performed at each time-point  $t$ : 1) the newly captured image is first normalized by *the preprocessing step* and 2) then it is segmented by *the deformable segmentation step*. The segmentation results of previous time-point images are refined by the incrementally updated patient-specific shape statistics, if it is available. These two steps are detailed next.

*The preprocessing step:* The intensities of the first time-point image of a patient are first normalized to have the same distribution as the model image of a population [33]. Then, the image acquired at time-point  $t$  from the same patient is rigidly transformed onto the space of the first time-point image by using a registration method described in [31], [32]. Moreover, its intensities are also globally normalized to have a similar distribution as the first time-point image [33]. Thus, we can finally obtain an intensity and pose normalized image at time  $t$ , denoted as  $I_t$ . It is worth noting that, since the intensities of the first time-point image have been normalized to have the same distribution as the model image, the intensities of image  $I_t$  thus also have the same distribution as the model image. In this paper, all the intensity and pose normalized serial images are denoted as  $\{I_i, i = 1, 2, \dots, t\}$ , and can have up to 17 monthly scanned images in our database. For avoiding the size difference between patient's image and model image, the serial images  $\{I_i, i = 1, 2, \dots, t\}$  can be also rescaled to the size of model image.

*The deformable segmentation step:* A new deformable model is utilized to segment the current time-point image  $I_t$ , by using both the statistical information collected from a population and the statistical information collected from  $t - 1$  previously segmented images of the same patient (if it is available). Denoting the deformable contour as an ordered point set,  $\mathbf{f}_t = \{v_i = (x_i, y_i) | i = 1, 2, \dots, M\}$ , where  $M$  is the number of contour points, and given the image  $I_t$ ,  $\mathbf{f}_t$  can be obtained by minimizing the following energy function:

$$E_t = \omega_s E_t^{\text{SIFT}} + (1 - \omega_t) E^{\text{P}} + \omega_t E_t^{\text{I}} \quad (1)$$

where  $\omega_s$  is the weight factor for balancing between the energy term of SIFT features and the shape statistical model, and  $\omega_t$  is the weight factor for balancing between two different shape statistical models.  $E_t^{\text{SIFT}}$  represents the overall difference between the SIFT features of the contour points and those of the corresponding points in image  $I_t$ .  $E^{\text{P}}$  is the constraint energy term derived from the population-based shape statistics [8], [11], [34].  $E_t^{\text{I}}$  is the constraint energy term derived from the patient-specific shape statistics, collected from the segmentation results of  $t - 1$  previous time-point images of the same patient. Therefore, the energy term  $E_t^{\text{I}}$  is not existent for several initial time-point images, simply because the available segmentation results for the specific patient are not sufficient to capture the patient-specific shape statistics. We can actually begin to collect the patient-specific shape statistics once  $N_s$  time-point images have been acquired and segmented,

and  $N_s$  is the minimal number of segmentation results for a patient. Notice that we can gradually increase the weight ( $\omega_t$ ) of the energy term  $E_t^{\text{I}}$  and simultaneously decrease the weight of the energy term  $E^{\text{P}}$ , as more and more new images of the same patient are acquired and segmented. The weight  $\omega_t$  is defined as follows:

$$\omega_t = \begin{cases} 0 & t \leq N_s \\ (t - N_s)/(N_b - N_s) & N_s < t \leq N_b \\ 1 & N_b < t \end{cases} \quad (2)$$

If we have a sufficient number of images, i.e., over  $N_b$  time-point images, segmented from the same patient, we will stop using the population-based shape constraint term  $E^{\text{P}}$ , since the patient-specific shape statistics collected from over  $N_b$  samples is considered sufficient to capture the variability of lung fields in a specific patient.

The deformable segmentation of the  $t$ th time-point image,  $I_t$ , can be summarized as follows.

- 1) If  $t \leq N_s$ , we directly segment  $I_t$  by minimizing (1), using the same optimization strategy of ASM; here,  $E_t^{\text{I}}$  is not used since  $\omega_t$  is zero according to (2).
- 2) If  $t > N_s$ , the following four steps are performed (refer to Fig. 2 for schematic explanations).
  - Step 1) Segment image  $I_t$  by incorporating the patient-specific constraint term  $E_t^{\text{I}}$ , which is derived from the segmentation results of  $t - 1$  previous time-point images,  $I_1, \dots, I_{t-1}$ .
  - Step 2) Update the patient-specific shape statistics by adding a new segmentation result of  $I_t$  to the patient-specific training set.
  - Step 3) Use this updated patient-specific shape statistics to refine the segmentations of all  $t$  time-point images,  $I_1, \dots, I_t$ .
  - Step 4) Update the patient-specific shape statistics again, which will be used for segmentation of next time-point image.

Notice that, in order to collect the patient-specific shape statistics from a limited number of rigidly-aligned training samples, a hierarchical shape representation method [35] is adopted in this study. We will describe this method in Section II-C.

In the following subsections, we first explain the necessity and performance of the SIFT local descriptor for correspondence detection in the chest radiographs (Section II-B), and the strategy of using the online learned patient-specific shape statistics for accurate segmentation of lung fields (Section II-C). Then, we describe each energy term in Section II-D. Finally, we describe our used optimization algorithm in Section II-E.

## B. SIFT Local Descriptor

We first describe the SIFT local descriptor for characterizing local image features around each pixel in an image, and then demonstrate the performance of the SIFT local descriptor in detecting point correspondences in the chest radiographs.

SIFT, as detailed in [24], consists of four major steps: 1) scale-space peak selection; 2) key point localization; 3) orientation assignment; 4) key point descriptor calculation. The first two steps detect the key points in the scale space, and they are not required in this study since we already have the

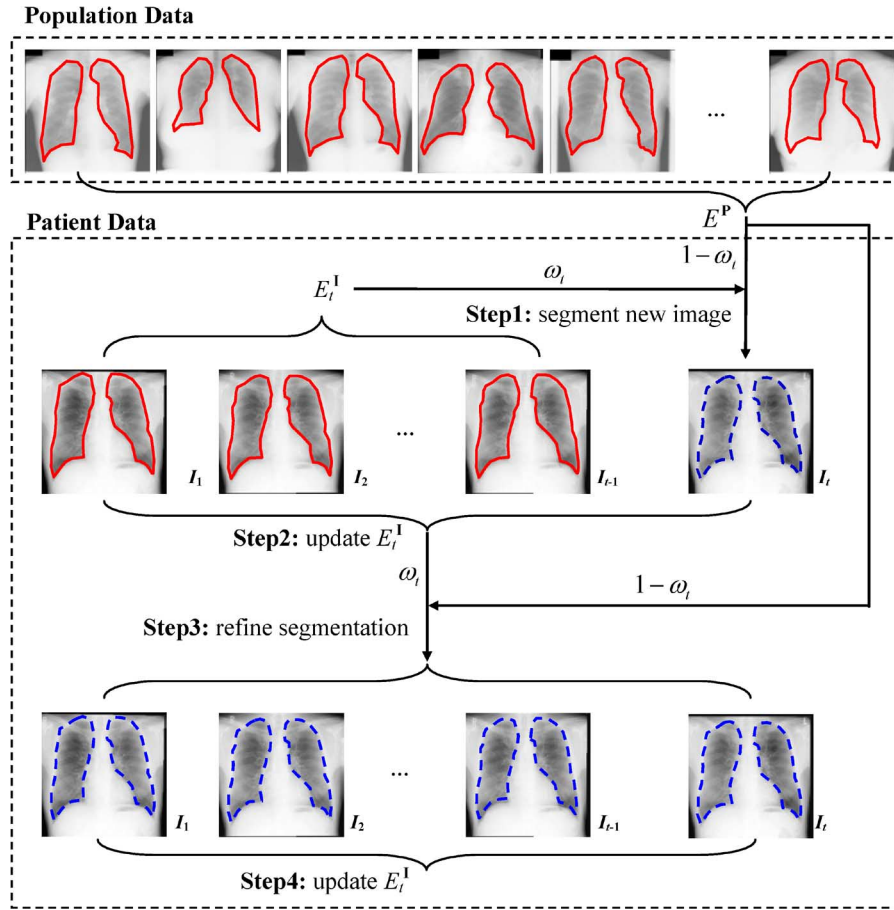


Fig. 2. Four major steps used in our deformable segmentation method for (Step 1) segmenting a new image  $I_t$ , (Step 2) updating shape statistics, (Step 3) refining the results, and (Step 4) updating shape statistics again.

contour points along the boundaries of lung fields in the chest radiographs as the key points. Therefore, we use the last two steps to compute the SIFT features for each contour point. The dominant orientation for each point is computed based on all the gradients inside the local neighborhood around that point, and the SIFT features are computed from the histograms of the gradient magnitudes within the neighborhood, as detailed next.

The dominant orientation for each point of the deformable contour is generated as follows. The input image  $I_t$  is first smoothed using a Gaussian filter, i.e.,  $L_t(x, y) = I_t(x, y) * G(x, y, \sigma)$ , where “\*” is the convolution operator, and  $G(x, y, \sigma) = e^{-(x^2+y^2)/2\sigma^2} / 2\pi\sigma^2$ . The width of the filter  $\sigma$  is determined empirically for ensuring that the computed SIFT feature can detect corresponding points in serial chest radiographs. Since all serial chest radiographs have been globally aligned and further rescaled to the same size

of the model image,  $\sigma$  can be fixed in the whole serial image segmentation procedure. The gradient magnitude  $m_t(x, y)$  and orientation  $\theta_t(x, y)$  can be computed from  $L_t(x, y)$ , as shown in (3) at the bottom of the page.

The dominant direction is defined as the peak in the orientation histogram, which is formed from the gradient orientations of all the pixels within a neighborhood of the point under consideration. In particular, the orientation histogram has 36 bins covering the  $360^\circ$  range of orientations. Each neighboring pixel added to the histogram is weighted by its gradient magnitude and by its distance to the point under consideration at a Gaussian way. Thus, pixels with larger gradient magnitude and smaller distance to the point under consideration will contribute more to the calculation of orientation histogram, and eventually the calculation of dominant direction.

In summary, the SIFT local descriptor for each point in the image is computed as follows.

$$\begin{aligned}
 m_t(x, y) &= \frac{1}{2} \sqrt{(L_t(x+1, y) - L_t(x-1, y))^2 + (L_t(x, y+1) - L_t(x, y-1))^2} \\
 \theta_t(x, y) &= \tan^{-1}((L_t(x, y+1) - L_t(x, y-1)) / (L_t(x+1, y) - L_t(x-1, y)))
 \end{aligned} \tag{3}$$

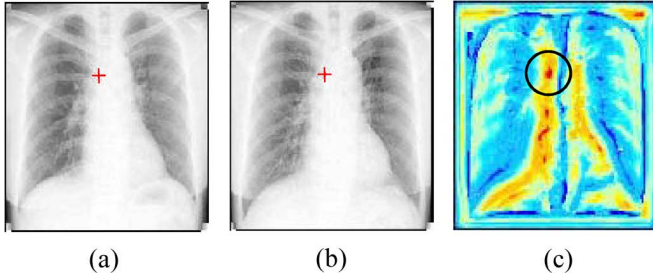


Fig. 3. Demonstration of using SIFT features for correspondence detection in two different time-point images of the same patient. SIFT features of the red crossed point in (a) is compared with the SIFT features of all the points in (b). Resultant similarity map is color-coded and shown in (c), where red represents high similarity. Red crossed point in (b) is the corresponding point detected, which has the highest similarity within the circular neighborhood in (c).

- 1) An image patch of the gradients around each point is first obtained. Note that the orientation of the image patch is determined according to the direction of the dominant gradients, as described above.
- 2) The SIFT local descriptor is then created based on the magnitudes and the orientations in the image gradient patch. First, the image gradient patch is spatially divided into  $4 \times 4$  equal-sized square regions, and then the histogram of the orientations within each region is computed, each with 8 orientation bins. Thus, the SIFT local descriptor of each point owns a 128-element feature vector.
- 3) This 128-element feature vector is finally normalized to a unit length.

The SIFT features have been proved to be one of the best to represent distinctive image features, compared to other features [30]. Fig. 3 shows that SIFT features are able to detect corresponding points in the chest radiographs. In this figure, the SIFT features of the red crossed point in one time-point image, as shown in Fig. 3(a) are compared with the SIFT features of all the points in another time-point image as shown in Fig. 3(b), by using a Euclidean distance measure between two SIFT feature vectors. As indicated by the color-coded map in Fig. 3(c), there are only a very small number of points in Fig. 3(b) similar to the red crossed point in Fig. 3(a). Since the two images have been rigidly aligned, the corresponding point of the red crossed point of Fig. 3(a) can be detected in Fig. 3(b) by searching for the point with the maximal similar SIFT features in a local neighborhood.

### C. Learning Patient-Specific Shape Statistics

Although we have globally transformed all the subsequent images onto the space of the first time-point image and normalized their intensities, there exist distortions in the chest radiographs due to 3-D to 2-D mapping, as well as the anatomical changes of lung fields due to the development of disease or health recovery. Therefore, we cannot directly use the segmentation results of the previous images to predict the segmentation in the current image, e.g., using linear prediction along temporal dimension. In order to reflect the longitudinal correlations and variability of the shapes of the lung fields, we use the patient-specific shape statistics, learned online from the segmentation results of previous time-point images of the same patient, to

constrain the segmentation of current time-point image. Moreover, once a new image is acquired and segmented, its result can be inserted into the training set to update the patient-specific shape statistics. This updated shape statistics will be further used to refine the segmentation results of all the previous time-point images. This is the main idea of our online learning method, i.e., online updating the patient-specific shape statistics and applying it for future image segmentation.

Compared to the shapes of lung fields from a population, the shapes of the serial lung fields of the same patient are very similar, as shown in Fig. 4. Thus, it is relatively easy to capture the patient-specific shape statistics from a small number of shape samples, compared to the case of collecting population-based shape statistics. However, initially the number of training samples is far smaller than the number of points comprising the shape. If PCA is used to capture the shape variability, the degree-of-freedom or the number of principal components is very small (at most equal to the number of training samples minus one). This limits the “allowable shapes” spanned by the relatively few eigenvectors and leads to the inability of patient-specific shape statistics to match the desired contour, i.e., the boundary of lung fields. To effectively estimate the shape variability from a limited number of training samples, in this paper, we used a hierarchical formulation of PCA [35], referred to as *hierarchical PCA*. In the *hierarchical PCA* method, a shape is spatially partitioned into a number of segments and then PCA is performed for each segment, as well as for the middle point sets of all segments, as detailed below. Compared to the *global PCA*, *hierarchical PCA* increases the degree-of-freedom and can capture the shape variability more accurately, even from a relatively small number of training samples. In experiments, we demonstrate the performances of using *hierarchical PCA* to capture the patient-specific shape statistics by comparing with *global PCA*.

Specifically, at time-point  $t$ , we have  $t - 1$  previously segmented images and a current time-point image  $I_t$ . We need to train the *hierarchical PCA* model using either the  $t - 1$  segmentation results from the images  $I_1, \dots, I_{t-1}$  (in Step 1 of Fig. 2), or the  $t$  segmentation results from the images  $I_1, \dots, I_t$  (in Step 3 of Fig. 2). We first normalize each segmentation result (i.e., the resultant contour) to the space of the model image using a rigid transformation, and then capture the shape variations of the normalized contours using *hierarchical PCA*. In particular, for performing *hierarchical PCA*, we partition the contours along the boundaries of lung fields into  $G$  overlapping segments, each of which consists of  $g$  contour points. Therefore, the top level of the contour consists of  $G$  middle points of  $G$  segments, reflecting the spatial relationships between different segments, while the bottom level has  $G$  segments, each with  $g$  contour points to capture local shape changes. If  $g$  is comparable with the number of samples, i.e.,  $N_s$ , then  $N_s$  samples might be adequate to statistically represent and capture the local variability of each segment. Thus, by computing the local shape statistics for each segment and also the global shape statistics for the  $G$  middle points of  $G$  segments, we can capture both local and global shape variation information from the training samples using  $G + 1$  PCA models. In this way, this hierarchical representation of shape statistics can be used to effectively guide the

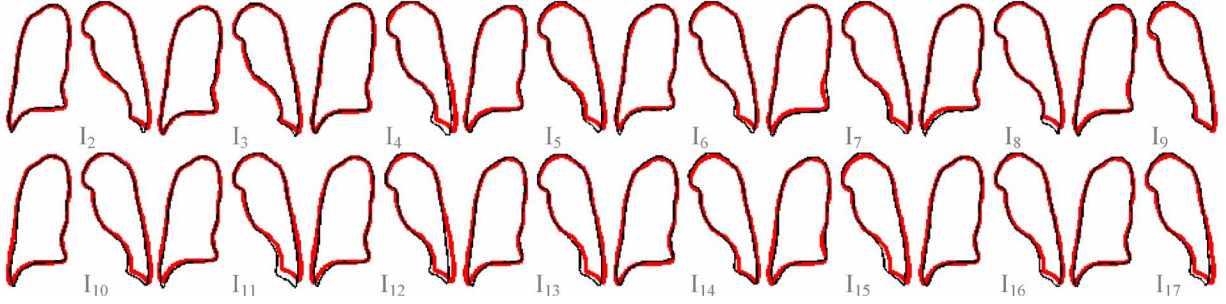


Fig. 4. Demonstration of local deformations of lung field shapes from the serial images of a patient. Black contours are the segmentation results of the serial images, and the red contours are identical, denoting the same shape segmented from the first time-point image.

deformable segmentation of lung fields in the current time-point image.

It is worth noting that this online and incrementally learned patient-specific shape statistics can be used only after a sufficient number of segmentation results, i.e.,  $N_s$ , are obtained. That means the segmentation for initial  $N_s$  time-point images will be fully constrained by the population-based shape statistics, and we start to capture the patient-specific shape statistics and use it to constrain the segmentation after  $N_s$  time points. As more time-point images are collected and processed, the constraint derived from the population-based shape statistics becomes less important, while at the same time the constraint derived from the patient-specific shape statistics increases gradually. Finally, the weight for population-based statistical shape constraint is set to 0, if we have at least  $N_b$  training samples from the same patient.  $N_b$  should be large enough to allow the *hierarchical PCA* to capture the fine variations of the whole shape. Equation (2) mathematically summarizes this strategy. In summary, the proposed strategy enables the deformable model to adapt to the patient-specific shape variability, and thus it is more accurate and robust for lung field segmentation.

#### D. Energy Term

In this subsection, the energy terms in (1) are described in detail. The first energy term  $E_t^{\text{SIFT}}$  is a distance measure between SIFT features, and minimizing it requires that the SIFT features of each contour point from the model image (of population or patient) match the SIFT features of the corresponding point in the image  $I_t$ . As we have described in Section II-B, SIFT features are distinctive and thus facilitate the correspondences detection across different time-point chest radiographs, particularly when patient's specific shape model is used after  $N_s$  time points. As mentioned above, for each contour point  $v_i$ , its SIFT feature vector  $s_i$  consists of 128 elements, representing the spatial distribution of the gradient orientations within a local neighborhood around this contour point. In order to match the lung fields more robustly, we use SIFT features on a profile of each contour point, i.e., all the SIFT features of  $2k + 1$  points selected evenly along both sides of the normal direction of the deformable contour at this point. Then, the energy term  $E_t^{\text{SIFT}}$  can be defined as follows:

$$E_t^{\text{SIFT}} = \frac{1}{M(2k+1)} \sum_{i=1}^M \sum_{j=1}^{2k+1} \sum_{l=1}^{128} \left( \frac{s_{ij}^t(l) - \bar{s}_{ij}(l)}{\sigma_{ij}(l)} \right)^2 \quad (4)$$

where  $\bar{s}_{ij}(l)$  represents the  $l$ th element of the average SIFT feature vector of the  $j$ th point along the profile of contour point  $v_i$  in the model image space, and  $s_{ij}^t(l)$  is the  $l$ -th element of the SIFT feature vector of the corresponding point in image  $I_t$ .  $\sigma_{ij}(l)$  is the standard deviation estimated from the training samples.  $M$  is the total number of contour points in the deformable contour as defined before.

Notice that the statistics of SIFT features, i.e., the average SIFT feature vectors and the standard deviations of each element, i.e.,  $\bar{s}_{ij}(l)$  and  $\sigma_{ij}(l)$ , are used for feature matching according to (4). The use of the statistics of SIFT features tends to be more robust in feature matching and also it is less biased, compared to the use of SIFT features from a single image. Similar to the use of the patient-specific shape statistics, the statistics of SIFT features obtained from the population is used for segmentation of the initial time-point images, i.e.,  $t \leq N_s$ , while the statistics of SIFT features that can be calculated from the available segmentation results of the patient is applied for the segmentation of the later time-point images, i.e.,  $t > N_s$ . The latter particularly makes the correspondence detection by SIFT features relatively easy (as demonstrated in Fig. 3), since SIFT features used in the model and the SIFT features calculated from the testing image are from the same patient. In summary, the selection of the average SIFT feature vectors and the standard deviations is dependent on the number of images we have captured and processed from the same patient. This is consistent with the use of the patient-specific shape statistical model in our deformable segmentation procedure.

The second energy term is the energy function of the population-based shape statistics. We use the same PCA-based method as [11]–[13] to capture the shape statistics from the population, i.e., a set of manually segmented training samples obtained from [8] are used to construct the population-based shape statistics in this study. According to PCA, the energy term is defined as,

$$E^{(P)} = \frac{1}{2} \sum_{i=1}^{C^{(P)}} \frac{(b_i^{(P)})^2}{e_i^{(P)}} \quad (5)$$

where  $\mathbf{b}^{(P)}$  is the vector of the shape parameters of the current contour  $\mathbf{f}_t$ , which is calculated using  $\mathbf{b}^{(P)} = (\Phi^{(P)})^T (\mathbf{f}'_t - \bar{\mathbf{f}}^{(P)})$ . Here,  $\mathbf{f}'_t$  is a rigidly-normalized contour of  $\mathbf{f}_t$  in the population space,  $\Phi^{(P)}$  is the matrix composed of the eigenvectors corresponding to the largest  $C^{(P)}$  eigenvalues of the covariance matrix of all the sample contours of the population,  $\bar{\mathbf{f}}^{(P)}$  denotes

TABLE I  
PARAMETERS USED IN THIS PAPER

Symbol	Value and Description
$N_s$	<b>5</b> (Start to collect patient-specific shape statistics if we have over $N_s$ serial images from a patient)
$N_b$	<b>10</b> (Use only the patient-specific shape statistics if we have over $N_b$ serial images from a patient)
$M$	<b>94</b> (Number of contour points used to match the lung fields)
$G$	<b>9</b> (Number of segments in a deformable contour)
$g$	<b>15</b> (Number of points for each segment in a deformable contour)
$k$	<b>2</b> (Number of points on each side of the profile of a contour point)
$N_{\max}$	<b>20</b> (Maximum number of iterations at each resolution)

the average shape of all the sample contours of the population, and  $e_j^{(P)}$  represents the corresponding eigenvalue.

The third energy term is the energy function of patient-specific shape statistics. Since there are a limited number of training samples for the patient-specific statistical shape model, we use the *hierarchical PCA* in [35] to effectively capture the detailed shape variability from a relatively small number of training samples. Please refer to Section II-C for a brief description of the *hierarchical PCA*. Since the *hierarchical PCA* consists of  $G + 1$  PCA models, respectively, for  $G$  segments and one global model of the middle points of all segments, the patient-specific energy function is, therefore, defined as a summation of the  $G + 1$  respective PCA-based energy terms

$$E_t^{(I)} = \frac{1}{G+1} \sum_{i=1}^{G+1} \sum_{j=1}^{C^{(i)}} \frac{(b_j^{(i)})^2}{e_j^{(i)}} \quad (6)$$

where  $\mathbf{b}^{(i)}$  is the vector of the shape parameters, and  $e_j^{(i)}$  is the  $j$ th eigenvalue in the PCA model of segment  $i$ . The definitions of these shape parameters are the same as those for the population-based shape statistical model detailed above.

### E. Optimization Algorithm

The algorithm is implemented in a multiresolution framework in order to prevent local minima and to improve the robustness of the proposed algorithm. In addition to the original image level, we used three subsampled resolution levels, and the down sampling rate is 2:1. The algorithm starts from the lowest resolution and use the matching results from the lower resolution to initialize the contour of next higher resolution. At each resolution, we first calculate the SIFT feature for every point in the image and then compute the deformable contour by minimizing the energy function (1) at that resolution. The iterative optimization strategy used in ASM [11]–[14] is adopted here to solve the deformable segmentation. It is worth noting that the statistics of SIFT features, i.e., mean and standard deviation, are separately calculated for each resolution, thus exactly capturing the statistics of SIFT features at particular resolutions. This is important for achieving successful segmentation at each resolution.

At each resolution, an iterative optimization strategy is used to compute the deformable contour by repeatedly minimizing  $E_t^{\text{SIFT}}$  and  $(1 - \omega_t)E_t^{\text{P}} + \omega_t E_t^{\text{I}}$ . First, the deformable contour is updated via a local search around its current location, so that

the newly updated contour has better match to the SIFT features of the testing image. In particular, each contour point is updated by searching along the direction perpendicular to the contour. In this way, a smaller  $E_t^{\text{SIFT}}$  is thus obtained. Then, the population-based and patient-specific shape statistics are used to constrain the newly updated deformable contour as follows.

- Given a new (normalized) deformable contour  $\mathbf{f}'_t$  at time point  $t$ , we first constrain it to lie within an acceptable shape range of the population-based and the patient-specific statistical shape spaces, respectively, and hence obtain two new contours,  $\mathbf{f}'_t|_{\text{P}}$  and  $\mathbf{f}'_t|_{\text{I}}$ , in the population-based and the patient-specific shape spaces, respectively.
- Then, we allow the deformable contour  $\mathbf{f}'_t$  to move toward the positions of these two new constrained contours at step sizes of  $(1 - \omega_t)$  and  $\omega_t$ , respectively, and thus obtain the final constrained deformable contour as  $((1 - \omega_t)\mathbf{f}'_t|_{\text{P}} + \omega_t\mathbf{f}'_t|_{\text{I}})$ .

This iterative procedure continues until the deformable contour converges or the maximal number of allowed iterations  $N_{\max}$  is reached. For lower resolution images, this contour is then used for initializing the contour in the next higher resolution, and for the finest resolution, this contour is the final matching result. Typically, the process ends within 20 iterations, and hence we set  $N_{\max}$  to 20. Finally, it should be indicated that this optimization algorithm does not guarantee the optimal solution, although it was used widely in many applications [11]–[13], [35]. But in our application, we found that its performance is satisfied.

## III. EXPERIMENT

Two sets of experiments are performed to respectively evaluate the performance of the proposed segmentation method. We first evaluate our algorithm without patient-specific shape statistics on a publicly available Japanese Society of Radiological Technology (JSRT) database [34], and compare its performance with that by ASM [11]–[14]. Then, we evaluate our full algorithm (with patient-specific shape statistics) on our database of serial chest radiographs.

Before providing the detailed experimental results, we first summarize the parameters used in these experiments, and then introduce the measures used to evaluate the goodness of segmentations in this study. Table I gives the values of important parameters we used in the experiments. These parameters

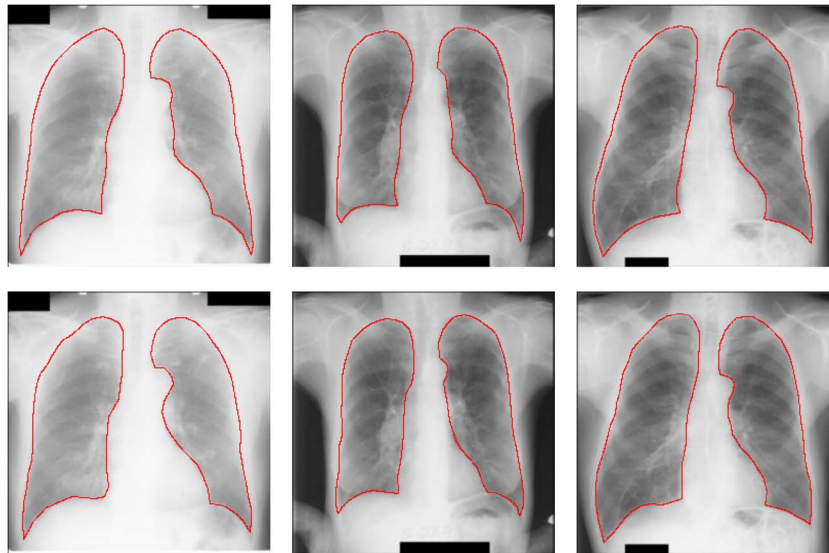


Fig. 5. Three chest radiographs with (top) manual and (bottom) automated lung fields segmentation by *ASM SIFT*.

specify when to collect and use the patient-specific shape statistics, how to decompose a deformable contour into different segments in *hierarchical PCA*, and how to define the profile of each contour point.

Two quantitative measures are used to evaluate the performance of the algorithms, i.e., the overlay percentage and the average contour distance between automated segmentation result and manual segmentation result [36]. The overlay percentage is computed as

$$\Omega = \frac{TP}{TP + FP + FN} \quad (7)$$

where TP stands for true positive (area correctly classified as object), FP for false positive (area incorrectly classified as object), and FN for false negative (area incorrectly classified as background).

For calculating the average contour distance between two contours A (i.e., automated segmentation) and B (i.e., manual segmentation), the distance between each point on contour A and its closest point on contour B is first computed, and then averaged over all points on contour A. For symmetric measurement, the above step is further repeated by interchanging contours A and B, and the overall average distance is finally reported with a unit of pixel.

#### A. Comparison on Public Database

This experiment evaluates our algorithm on the public JSRT/SCR database [8], [34]. The 247 cases in the database were subdivided in two folds. Each fold contains an equal amount of normal cases and abnormal cases. Images in one fold were segmented with the images in the other fold as training set, and vice versa. All of the original radiographs were down-sampled to the  $256 \times 256$  resolution images.

Since these public data are individual images, not serial images, we evaluated our algorithm without patient-specific shape statistics on these data. That is, our algorithm actually minimizes the energy function (1) with  $\omega_t = 0$ . The difference between our algorithm and ASM is that our algorithm uses SIFT

TABLE II  
AVERAGE OVERLAY PERCENTAGE BETWEEN THE MANUAL SEGMENTATION AND AUTOMATED SEGMENTATION OF ALL 247 IMAGES IN A PUBLIC DATABASE, BY *ASM SIFT* AND *ASM Intensity*, RESPECTIVELY (%)

Algorithm	Mean $\pm$ std	Minimum	Median	Maximum
<i>ASM SIFT</i>	92.0 $\pm$ 3.1	78.3	92.8	96.1
<i>ASM Intensity</i>	87.0 $\pm$ 7.4	60.8	89.2	95.4

TABLE III  
AVERAGE CONTOUR DISTANCE BETWEEN THE MANUAL SEGMENTATION AND AUTOMATED SEGMENTATION OF ALL 247 IMAGES IN A PUBLIC DATABASE, BY *ASM SIFT* AND *ASM Intensity*, RESPECTIVELY (UNIT IN PIXEL)

Algorithm	Mean $\pm$ std	Minimum	Median	Maximum
<i>ASM SIFT</i>	1.78 $\pm$ 0.78	0.96	1.64	6.79
<i>ASM Intensity</i>	3.10 $\pm$ 1.95	1.03	2.36	9.69

features for lung field matching, while ASM uses image intensity and gradient features. For convenience, we denote our algorithm as *ASM SIFT*, and the standard ASM as *ASM Intensity*.

Some qualitative segmentation results by *ASM SIFT* are illustrated in Fig. 5. The top row shows the manual segmentation results, and the bottom row shows the *ASM SIFT* results. It can be seen that the results of *ASM SIFT* are similar to the manually segmented contours.

For quantitative results, the average overlay percentage and the average contour distance between the segmentation results and the ground truth of all 247 images are reported in Tables II and III, respectively. It can be seen that the average overlay for the lung fields is 92.0% for *ASM SIFT* compared to 87.0% for *ASM Intensity*, and the average contour distance for *ASM SIFT* is about 1.78 pixels compared to 3.10 pixels for the *ASM Intensity*. Statistical test shows that the difference between *ASM Intensity* and *ASM SIFT* is statistically significant ( $p < 0.05$ ). Notice that *ASM Intensity* is the same algorithm termed as *ASM default* in

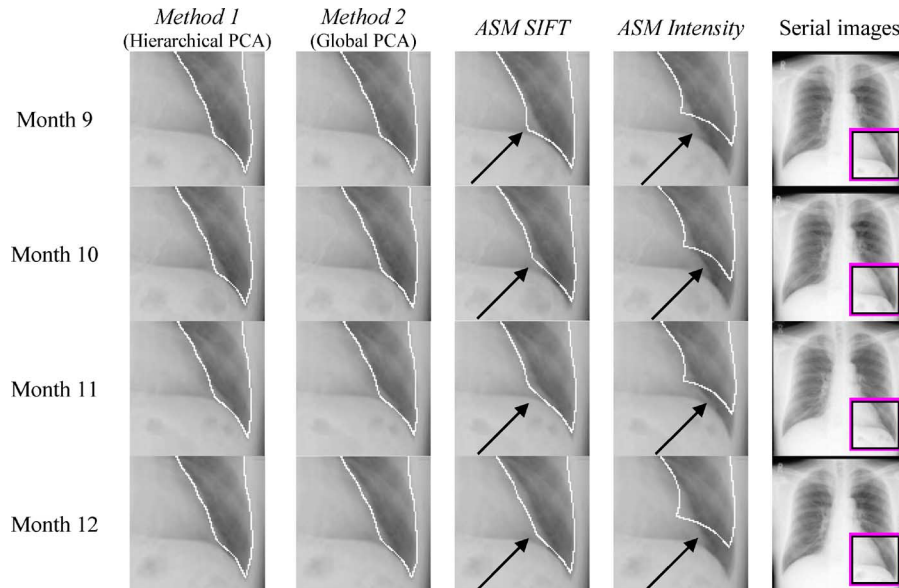


Fig. 6. Some segmentation results produced by four different segmentation methods.

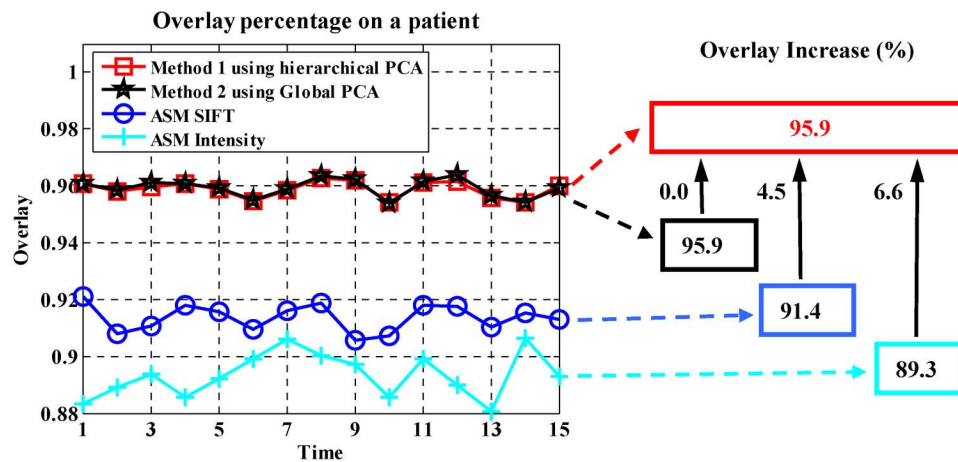


Fig. 7. Quantitative comparison of average overlay percentages on the serial images of a patient, by four different segmentation methods.

[8], and the accuracy we obtained is slightly worse than that in [8]. This could be the use of different contour models in two separate studies. Our model includes only the contours of left and right lung fields, without including the contours of heart and clavicles, since the boundaries of heart and some clavicles are not clearly seen in our serial data. Therefore, in our method, the contours of the left lung and the right lung are apart from each other, which might lead to weak correlation between two lungs and thus slightly affect our segmentation results.

### B. Comparison on Serial Chest Radiograph Database

This experiment evaluates our full algorithm (with patient-specific shape statistics) on the serial frontal chest radiographs of 39 patients. Most of these 39 patients have up to 17 monthly scanned images, each with  $256 \times 256$  pixels. The lung fields of all these serial images have been manually delineated by a human observer, thus these manual segmentation results can be used as a golden standard to evaluate the automatic segmentation results. Both qualitative and quantitative results are provided in this section to show the advantage of using *hierarchical*

*PCA* to capture patient-specific shape statistics and also the effectiveness of using both population-based and patient-specific shape statistics for chest radiograph image segmentation. The qualitative comparison is provided in Section III-B1, and the quantitative comparison is reported in Section III-B2.

In order to evaluate the performance of SIFT features as well as patient-specific shape statistics, in this experiment we compared the following four methods using our serial chest radiograph database. *Method 1* is our full method, which uses both SIFT features and the patient-specific shape statistics trained using *hierarchical PCA*. *Method 2* uses *global PCA* to replace *hierarchical PCA* (in *Method 1*) for patient-specific shape statistics estimation. In addition, the *ASM SIFT*, which is our method without patient-specific shape statistics, and the *ASM Intensity*, which is the standard ASM algorithm, are also compared.

All of these four methods were implemented using standard C programs. The computation time for segmenting one chest radiograph image is around 75 s for our methods using *hierarchical PCA* (*Method 1*) or using *global PCA* (*Method 2*), 75 s

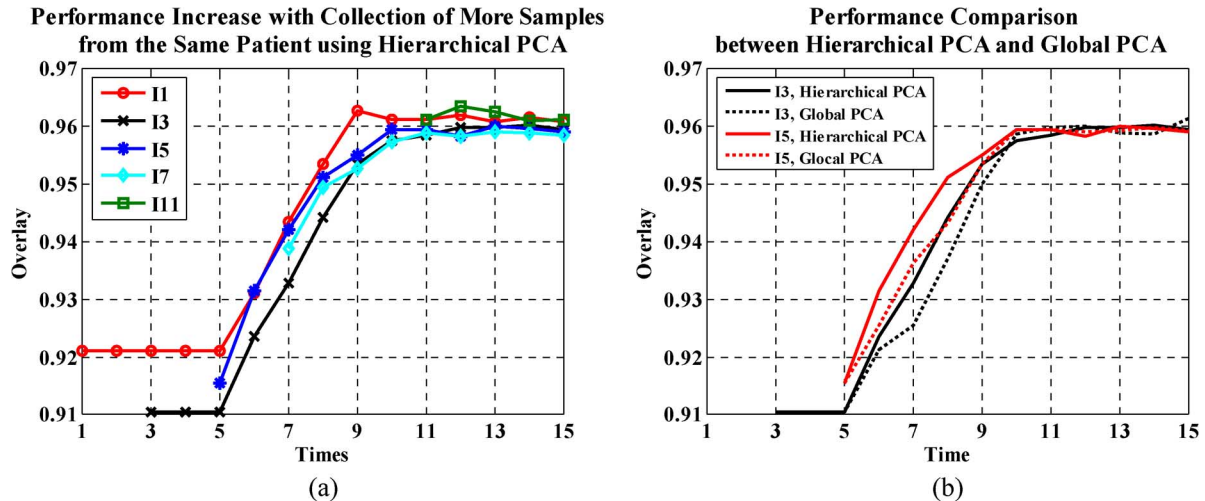


Fig. 8. Process of updating segmentations for five selected time-point images of a patient by *Method 1* using *Hierarchical PCA* (a), evaluated by average overlay percentage. Performance comparison between *Hierarchical PCA* and *Global PCA* is also provided in (b), when segmenting images of time points 3 and 5 at different segmentation stages. .

TABLE IV  
ILLUSTRATION OF THE PROCESS OF UPDATING SEGMENTATIONS FOR SERIAL IMAGES BY *Method 1* IN FIG. 7.  
PERFORMANCE IS EVALUATED BY AVERAGE OVERLAY PERCENTAGE (%)

Algor. / Time	Segmenting serial images of a patient by <i>Method 1</i> using <i>Hierarchical PCA</i> since $t=6$														
	$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	$I_7$	$I_8$	$I_9$	$I_{10}$	$I_{11}$	$I_{12}$	$I_{13}$	$I_{14}$	$I_{15}$
ASM Intensity	88.3	88.9	89.4	88.6	89.2	89.9	90.6	90.0	89.7	88.6	89.9	89.0	88.0	90.7	89.3
ASM SIFT	92.1	90.8	91.0	91.8	91.6	91.0	91.6	91.9	90.6	90.7	91.8	91.7	91.0	91.5	91.3
$t \leq N_s = 5$	92.1	90.8	91.0	91.8	91.6										
$t=6$	93.1	92.6	92.4	93.4	93.1	92.3									
$t=7$	94.3	93.2	93.3	94.4	94.2	93.2	93.9								
$t=8$	95.4	93.9	94.4	95.3	95.1	93.8	94.9	94.9							
$t=9$	96.3	95.6	95.3	96.0	95.5	95.0	95.3	96.0	95.8						
$t=10$	96.1	95.8	95.8	96.3	96.0	95.6	95.7	96.1	96.2	95.4					
$t=11$	96.1	95.8	95.9	96.3	96.0	95.5	95.9	96.3	96.2	95.5	96.1				
$t=12$	96.2	95.8	96.0	96.2	95.8	95.5	95.8	96.3	96.2	95.5	96.4	96.6			
$t=13$	96.1	95.7	96.0	96.2	96.0	95.4	95.9	96.2	96.2	95.5	96.3	96.3	95.5		
$t=14$	96.2	95.8	96.0	96.2	96.0	95.6	95.9	96.2	96.2	95.4	96.1	96.0	95.6	95.5	
$t=15$	96.1	95.8	96.0	96.1	96.0	95.5	95.8	96.3	96.2	95.4	96.1	96.2	95.6	95.4	96.0

for *ASM SIFT*, and around 30 s for *ASM Intensity* on a 3.0 GHz Intel PC with 1GB RAM. It is obvious that the methods using SIFT features took longer times, since 128 SIFT features are used to characterize the vicinity of each point in the deformable shape model, which is much more than the use of only intensity and gradient features in *ASM Intensity*.

Finally, it is worth noting that, if the segmentations from the previous time-points are wrong, the acquired patient-specific shape statistics will affect the segmentation of the later time-point images. Therefore, we allow users to check the results before adding the new segmentations to update the patient-specific shape statistics. This is particularly possible for our clinical setting, since images are acquired monthly. On the other hand, we would like to mention that we did not find any obvious wrong segmentation cases in our serial image database on 39 patients.

1) *Qualitative Comparison*: Fig. 6 shows the segmentation results of serial chest radiographs of a patient, using the four methods. The original serial images are shown in the fifth column. In order to be able to visually compare the segmentation results, we enlarged the images in the small rectangles, and the white contours represent the final matching results of different methods. From these results, it can be observed that our segmentation results (*Method 1* and *Method 2*) are generally very close to the lung field boundaries, while the two ASM-based segmentation methods are not always accurate, particularly in some regions as indicated by black arrows. At the same time, we can find that *Method 1* and *Method 2* are longitudinally more stable than *ASM SIFT* and *ASM Intensity*. These results indicate that the effectiveness of using patient-specific shape statistics in improving the segmentation accuracy. Moreover, by comparing the respective segmentation

TABLE V

AVERAGE OVERLAY PERCENTAGE BETWEEN THE MANUAL SEGMENTATION AND AUTOMATED SEGMENTATION BY *Method 1* USING SIFT FEATURE AND *Hierarchical PCA*, *Method 2* USING SIFT FEATURE AND *Global PCA*, *Method 3* USING INTENSITY FEATURE AND *Hierarchical PCA*, *Method 4* USING INTENSITY FEATURE AND *Global PCA*, *ASM SIFT*, AND *ASM Intensity*, RESPECTIVELY, ON THE SERIAL CHEST RADIOGRAPHS DATABASE OF 39 PATIENTS (%)

Algorithm	Mean $\pm$ std	Minimum	Median	Maximum
<i>Method 1</i> ( using SIFT feature and <i>Hierarchical PCA</i> )	94.9 $\pm$ 1.3	88.4	95.0	97.0
<i>Method 2</i> (using SIFT feature and <i>Global PCA</i> )	94.8 $\pm$ 1.4	88.2	95.1	96.9
<i>Method 3</i> (using intensity feature and <i>Hierarchical PCA</i> )	94.3 $\pm$ 1.8	84.9	94.7	97.1
<i>Method 4</i> (using intensity feature and <i>Global PCA</i> )	93.5 $\pm$ 1.8	84.4	93.8	96.9
<i>ASM SIFT</i>	91.0 $\pm$ 1.7	85.0	91.4	94.8
<i>ASM Intensity</i>	87.3 $\pm$ 4.9	76.9	89.2	93.7

TABLE VI

AVERAGE CONTOUR DISTANCE BETWEEN THE MANUAL SEGMENTATION AND AUTOMATED SEGMENTATION BY *Method 1* USING SIFT FEATURE AND *Hierarchical PCA*, *Method 2* USING SIFT FEATURE AND *Global PCA*, *Method 3* USING INTENSITY FEATURE AND *Hierarchical PCA*, *Method 4* USING INTENSITY FEATURE AND *Global PCA*, *ASM SIFT*, AND *ASM Intensity*, RESPECTIVELY, ON THE SERIAL CHEST RADIOGRAPHS DATABASE OF 39 PATIENTS (UNIT IN PIXEL)

Algorithm	Mean $\pm$ std	Minimum	Median	Maximum
<i>Method 1</i> (using SIFT feature and <i>Hierarchical PCA</i> )	1.10 $\pm$ 0.31	0.62	1.06	2.72
<i>Method 2</i> ( using SIFT feature and <i>Global PCA</i> )	1.10 $\pm$ 0.30	0.62	1.07	2.75
<i>Method 3</i> (using intensity feature and <i>Hierarchical PCA</i> )	1.17 $\pm$ 0.35	0.65	1.09	3.56
<i>Method 4</i> (using intensity feature and <i>Global PCA</i> )	1.34 $\pm$ 0.36	0.72	1.27	3.72
<i>ASM SIFT</i>	2.06 $\pm$ 0.46	1.06	1.97	3.76
<i>ASM Intensity</i>	3.07 $\pm$ 1.35	1.32	2.45	6.05

results obtained by *ASM SIFT* and *ASM Intensity*, we can see that the use of SIFT features effectively improves the performance of *ASM SIFT*, since the use of SIFT features facilitates the correspondence detection in the serial image segmentation, which is in consistence with the comparison performed in Section III-A. In the next section, we will compare the segmentation results of these four methods quantitatively.

## 2) Quantitative Comparison:

*On individual patient:* We used quantitative measures to evaluate the segmentation results obtained by the four segmentation methods. Figs. 7 and 8, and Table IV show the overlay percentages between manual segmentation and automated segmentation on serial chest radiographs of a patient. From Fig. 7, it can be observed that, compared to *ASM Intensity* and *ASM SIFT*, *Method 1* and *Method 2* that use patient-specific shape statistics provide higher overlay degrees with the manual segmentations. For example, the average overlay percentages by *Method 1* and *Method 2* are both 95.9%, which are much higher than those by *ASM SIFT* and *ASM Intensity* (91.4% and 89.3%).

Notice that Fig. 7 shows the final updated results for the serial images, after we have processed 15 serial images. For our full method, *Method 1*, since we started using the patient-specific shape statistics (trained using *hierarchical PCA*) after  $N_s =$

5 images were processed, and stopped using the population-based shape statistics after  $N_s = 10$  images were processed, it would be interesting to look into the detailed results during this progressive procedure. The intermediate results are shown in Table IV. Each row shows the final results at a particular stage. It can be seen that when we only have 5 or less images, we obtain the same results as the *ASM SIFT*, since no patient-specific shape statistics is used in this stage. As more and more new images are processed, the overall accuracy for matching the lung field is gradually improved. In addition, by comparing all segmentation results in Table IV, we can observe that 1) the use of SIFT features can produce more accurate segmentation, and 2) patient-specific shape statistics can significantly improve the accuracy of segmentation.

Furthermore, if we look at the data along each column, we can find that most of the improvements happened between time points 6 and 10. This can be clearly seen if we plot some of the columns of Table IV in Fig. 8(a). It can be seen that the segmentation accuracy for each image increases very quickly between time points 6 and 10, while the improvement of accuracy is very little after time point 10. This might be because 10 time-point images would be sufficient to capture the shape variations within the data of that patient.

TABLE VII  
PERFORMANCE COMPARISON BETWEEN *Method 1 (Hierarchical PCA)* AND *Method 2 (Global PCA)* ON SEGMENTATION OF DIFFERENT NUMBER OF SERIAL IMAGES IN 39 PATIENTS, USING AVERAGE OVERLAY PERCENTAGE AND AVERAGE CONTOUR DISTANCE

All seg. results at time point	Average overlay percentage (%)		Average contour distance (Unit in pixel)	
	<i>Hierarchical PCA</i>	<i>Global PCA</i>	<i>Hierarchical PCA</i>	<i>Global PCA</i>
6 <sup>th</sup>	92.5±1.2	91.0±1.4	1.68±0.32	2.06±0.49
7 <sup>th</sup>	93.4±1.2	91.0±1.7	1.46±0.28	2.03±0.47
8 <sup>th</sup>	94.1±1.4	92.0±1.8	1.34±0.34	1.83±0.45
9 <sup>th</sup>	94.8±1.3	93.7±1.4	1.11±0.28	1.37±0.31
10 <sup>th</sup>	94.8±1.4	94.8±1.4	1.11±0.31	1.12±0.31
11 <sup>th</sup>	94.8±1.4	94.8±1.4	1.11±0.31	1.11±0.33
12 <sup>th</sup>	94.8±1.4	94.8±1.4	1.10±0.31	1.10±0.31
13 <sup>th</sup>	94.9±1.4	94.9±1.3	1.09±0.31	1.08±0.31
14 <sup>th</sup>	94.8±1.5	94.8±1.6	1.09±0.32	1.10±0.31
15 <sup>th</sup>	94.9±1.3	94.8±1.4	1.10±0.31	1.10±0.30

In Fig. 8(b), we demonstrate the advantage of using *hierarchical PCA (Method 1)* to effectively capture the patient-specific shape statistics, when the patient-specific training set is relatively small, i.e., between time points 6 and 10. It can be observed that the segmentation results by *Method 1* (using *hierarchical PCA*) is relatively better than *Method 2* (using *global PCA*) for time points between 6 and 10. After time point 10, it seems that *global PCA* starts to effectively capture the patient-specific shape statistics, thus producing the similar results as using *hierarchical PCA*.

Similar to the overlay percentage measure, we also calculated the average contour distances between the manual segmentation and the automated segmentation obtained by the four segmentation methods on the same patient. It turns out that the average contour distances by *Method 1* and *Method 2* are much less than those by *ASM SIFT* and *ASM Intensity*. For example, for the same patient, after processing 15 serial images, the average contour distances of our methods (*Method 1* and *Method 2*) are 0.96 pixels, and those of *ASM SIFT* and *ASM Intensity* are 2.08 and 2.74 pixels, respectively.

*On all 39 patients:* We have segmented all the serial images of 39 patients (up to 17 images per patient) in our database using different methods, and quantitatively compared the average overlay percentages and the average contour distances of all the segmentation results. In addition to the four methods used above, two other methods, called as *Method 3* and *Method 4*, are also compared to evaluate whether the use of patient-specific shape statistics or the use of SIFT features contributes more to the improvement of the accuracy of segmentation. In particular, *Method 3* uses image intensity features, instead of SIFT features, along with the patient-specific shape statistics trained by *hierarchical PCA*. *Method 4* is similar to *Method 3*, but it uses *global PCA*, instead of *hierarchical PCA* for patient-specific shape statistics estimation.

The average overlay percentages and the average contour distances of all the six methods are shown in Tables V and VI, respectively. It can be seen that, when the same SIFT features are used, the average overlay percentages of *Method 1* (94.9%) and

*Method 2* (94.8%) that use patient-specific statistics are much higher than that of *ASM SIFT* (91.0%). Also, when the same image intensity features are used, the average overlay percentages of *Method 3* (94.3%) and *Method 4* (93.5%) are much higher than that of *ASM intensity* (87.3%). Similar conclusion can be drawn for the average contour distances in Table VI. All of these results indicate that, no matter which kind of image features are used, the methods using the patient-specific statistics (*Methods 1~4*) outperform those without using patient-specific statistics (*ASM SIFT* and *ASM Intensity*).

Similarly, by comparing the results obtained by *Methods 1~2* and *Methods 3~4* as well as the results obtained by *ASM SIFT* and *ASM Intensity* in Tables V and VI, it can be concluded that the use of SIFT features produces more accurate segmentation results, compared to the use of image intensity features. Since the relative improvement achieved by *Methods 1~2*, compared to *Methods 3~4*, is relatively small, it seems that the use of patient-specific shape statistics contributes more than the use of SIFT features in serial chest image segmentation.

Finally, we compare the performances of *Method 1* and *Method 2*, to demonstrate the advantage of using *Hierarchical PCA (Method 1)* to effectively capture the patient-specific shape statistics, when the patient-specific training set is relatively small. Table VII lists the average overlay percentages and the average contour distances of the segmentation results for all the serial images of 39 patients, obtained at different time points. For example, the line of the sixth time point gives the results achieved when we have six serial images for each of 39 patients. It can be seen that *Method 1* is more accurate than *Method 2*, especially for smaller number of images (6~9 serial images). This indicates that *hierarchical PCA* captures patient-specific shape statistics more effectively than *global PCA*, when the patient-specific training set is relatively small.

#### IV. CONCLUSION

We have presented a new deformable model to segment the lung fields in serial chest radiographs by using both population-

based and patient-specific shape statistical constraints. The patient-specific shape statistics is online and incrementally learned from the segmentation results of previous time-point images of the same patient. In particular, for initial time-point images, the population-based shape statistics plays the major role for statistically constraining the deformable contours. As more and more segmentation results are obtained, the patient-specific shape statistics starts to constrain the segmentation of lung fields and gradually take the major role for the statistical constraining. In order to effectively capture the patient-specific shape variability from a relatively small number of training samples, a hierarchical shape representation method, *i.e.*, *hierarchical PCA*, is employed. Also, the SIFT local descriptor is used to characterize the image features in the vicinity of each contour point, in order to facilitate the correspondence detection during the deformable segmentation procedure. Compared to other generic features, SIFT features are relatively distinctive and thus make the correspondence detection across different time-point chest radiographs more reliable.

Both qualitative and quantitative results show that the use of SIFT features improves the accuracy for lung field segmentation, *e.g.*, *ASM SIFT* has better performance than *ASM Intensity*. Moreover, patient-specific shape statistics can further improve the accuracy of lung field matching, *e.g.*, the performance of our full method is much better than that of *ASM SIFT*.

In the future works, we plan to further improve our method in the following aspects. First, the training samples obtained from the previous segmentation results of the same patient might be adaptively treated when using them to capture the patient-specific shape statistics. This is because the previous segmentation results from the neighboring time-points should have more contributions for capturing the patient-specific shape statistics for the current time-point, since they are more correlated with the current time-point images. Second, the learning-based method can be used to adaptively separate lung field boundaries into short segments, thus all points on each segment having longitudinally similar deformations. In the current study, the lung field boundaries are evenly separated into a number of short segments. In the future, the lung field boundaries can be separated into short segments according to longitudinal deformation learned for each boundary point. Thus, the relatively small variations on each short segment can be easily captured by a particular PCA.

## REFERENCES

- [1] B. van Ginneken, B. ter Haar Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: A survey," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1228–1241, Dec. 2001.
- [2] J. Q. Jaeger and R. L. Mehta, "Assessment of dry weight in hemodialysis: An overview," *J. Amer. Social Nephrol.*, vol. 10, pp. 392–403, 1999.
- [3] N. D. Vaziri and R. Prakash, "Echocardiographic evaluation of the effect of hemodialysis on cardiac size and function in patients with end-stage renal disease," *Amer. J. Med. Sci.*, vol. 278, pp. 201–206, 1979.
- [4] K. Inoue, K. Yoshii, and H. Ito, "Effect of aging on cardiothoracic ratio in women: A longitudinal study," *Gerontology*, vol. 45, pp. 53–58, 1999.
- [5] T. Suwatanapongched, D. S. Gierada, R. M. Slone, T. K. Pilgram, and P. G. Tuteur, "Variation in diaphragm position and shape in adults with normal pulmonary function," *Chest*, vol. 123, pp. 2019–2027, 2003.
- [6] M. S. Brown, L. S. Wilson, B. D. Doust, R. W. Gill, and C. Sun, "Knowledge-based method for segmentation and analysis of lung boundaries in chest x-rays images," *Comput. Med. Imag. Graph.*, vol. 22, pp. 463–477, 1998.
- [7] J. Duryea and J. M. Boone, "A fully automatic algorithm for the segmentation of lung fields in digital chest radiographic images," *Med. Phys.*, vol. 22, no. 2, pp. 183–191, 1995.
- [8] B. van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database," *Med. Image Anal.*, vol. 10, no. 1, pp. 19–40, 2006.
- [9] M. F. McNitt-Gray, H. K. Huang, and J. W. Sayre, "Feature selection in the pattern classification problem of digital chest radiographs segmentation," *IEEE Trans. Med. Imag.*, vol. 14, no. 3, pp. 537–547, Sep. 1995.
- [10] B. van Ginneken and B. M. ter Haar Romeny, "Automatic segmentation of lung fields in chest radiographs," *Med Phys.*, vol. 27, no. 10, pp. 2445–2455, 2000.
- [11] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.
- [12] T. Cootes, G. J. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. Pattern Ana. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
- [13] T. F. Cootes and C. J. Taylor, Statistical models of appearance for computer vision Wolfson Image Analysis Unit, Univ. Manchester, Manchester, U.K., 2001.
- [14] M. B. Stegmann, B. K. Ersboll, and R. Larsen, "FAME—A flexible appearance modeling environment," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1319–1331, Oct. 2003.
- [15] B. van Ginneken and M. Loog, "Pixel position regression-application to medical image segmentation," in *Proc. 17th Int. Conf. Pattern Recognition*, 2004, vol. 3, pp. 718–721.
- [16] D. Seghers, D. Loechx, F. Maes, and P. Suetens, "Image segmentation using local shape and gray-level appearance models," *Proc. SPIE*, vol. 6144, p. 614401.
- [17] S. S. Gleason, M. Paulus, D. Johnson, H. Sari-Sarraf, and M. A. Abidi, "Statistical-based deformable models with simultaneous optimization of object gray-level and shape characteristics," in *Proc. 4th IEEE Southwest Symp. Image Anal. Interpretation*, 2000, pp. 93–95.
- [18] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. M. T. H. Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 924–933, Aug. 2002.
- [19] M. J. Carreira, D. Cabello, and A. Mosquera, "Automatic segmentation of lung fields on chest radiograph images," *Comput. Biomed. Res.*, vol. 32, pp. 283–303, 1999.
- [20] S. G. Armato, M. Giger, and H. MacMahon, "Automated lung segmentation in digitized posteroanterior chest radiographs," *Academic Radiol.*, vol. 5, pp. 245–255, 1998.
- [21] H. Luo, R. Gaborski, and R. Acharya, "Automatic segmentation of lung regions in chest radiographs: A model guided approach," in *Proc. Int. Conf. Image Process.*, Vancouver, BC, Canada, 2000, vol. 2, pp. 483–486.
- [22] L. Li, Y. Zheng, M. Kallergi, and R. A. Clark, "Improved method for automatic identification of lung regions on chest radiographs," *Acad. Radiol.*, vol. 8, no. 7, pp. 629–638, 2001.
- [23] D. Shen, "Image registration by local histogram matching," *Pattern Recognit.*, vol. 40, pp. 1161–1172, 2007.
- [24] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [25] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *Comput. Vis. Pattern Recognit.*, vol. 2, pp. 506–513, 2004.
- [26] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.
- [27] J. J. Koenderink and A. J. van Doorn, "Representation of local geometry in the visual system," *Biol. Cybern.*, vol. 55, pp. 367–375, 1987.
- [28] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or how do I organize my holiday snaps?," in *Proc. Eur. Conf. Comput. Vision (ECCV)*, Copenhagen, Denmark, 2002, pp. 414–431.

- [29] L. J. van Gool, T. Moons, and D. Ungureanu, "Affine/photometric invariants for planar intensity patterns," in *Proc. 4th Eur. Conf. Comput. Vision*, Cambridge, England, 1996, vol. 1, pp. 642–651.
- [30] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
- [31] M. Jenkinson and S. M. Smith, "A global optimization method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, 2001.
- [32] M. Jenkinson, P. R. Bannister, J. M. Brady, and S. M. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, no. 2, pp. 825–841, Oct. 2002.
- [33] L. G. Nyul, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 143–150, Feb. 2000.
- [34] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *Amer. J. Roentgenol.*, vol. 174, pp. 71–74, 2000.
- [35] C. Davatzikos, X. Tao, and D. Shen, "Hierarchical active shape models, using the wavelet transform," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 414–423, Mar. 2003.
- [36] G. Gerig, M. Jomier, and M. Chakos, "VALMET: A new validation tool for assessing and improving 3D object segmentation," in *MICCAI 2001*. New York: Springer, vol. 2208, Lecture Notes in Computer Science, pp. 516–528.