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

Yonghong Shi[1], Feihu Qi[1], Zhong Xue[2], Kyoko Ito[3],
Hidenori Matsuo[3], and Dinggang Shen[2,*]

[1] Department of Computer Science and Engineering
Shanghai Jiao Tong University, Shanghai, China 200030
`{shi-yh, fhqi}@cs.sjtu.edu.cn`
[2] Section of Biomedical Image Analysis, Department of Radiology
University of Pennsylvania, Philadelphia, PA 19104,
`{zhong.xue, dinggang.shen}@uphs.upenn.edu`
[3] Hidaka Hospital, 886 Nakaomachi, Takasaki Gunma-Prefecture, 370-0001, Japan
`{itokyon@da3.so-net.ne.jp, hidakadc}@cocoa.ocn.ne.jp`

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. *First*, a modified scale-invariant feature transform (SIFT) local descriptor is used to characterize the image features in the vicinity of each pixel, so that the deformable model deforms in a way that seeks for the region with similar SIFT local descriptors. *Second*, the deformable model is constrained by both population-based and patient-specified shape statistics. Initially, population-based shape statistics takes most of the rules when the number of serial images is small; gradually, patient-specific shape statistics takes more rules after a sufficient number of segmentation results on the same patient have been obtained. The proposed deformable model can adapt to the shape variability of different patients, and obtain more robust and accurate segmentation results.

1 Introduction

Evaluation of the variation of cardiac size from month to month by taking serial chest radiographs [1-3] 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 [4], and the misevaluation of DW often results in a fatal illness such as patient's death. Since segmentation of lung fields in the chest radiographs provides a means to measure cardiac size, it is important to accurately segment lung fields from the regularly captured serial chest radiographs of each patient. Methods for segmenting lung fields in chest radiographs [1] can be classified into three groups, i.e., the rule-based methods [3], the pixel-based classification methods [2], and the deformable model-based methods, including ASM [5] and AAM [6]. In this paper, a new deformable model is proposed for robust and accurate serial lung field segmentation.

---

* Corresponding author. E-mail: dinggang.shen@uphs.upenn.edu (Dinggang Shen).

In deformable segmentation, it is important to characterize the relatively rich image features around the boundaries of lung fields, and then use these image features to guide the image segmentation. This is because the generic image features, such as edges along the boundaries of lung fields, are always inconsistent, and not sufficiently distinctive to discriminate, for example, rib edges from lung field edges. Therefore, the complex local descriptors, such as SIFT [7], might be suitable to characterize the image features around each point along the boundaries of lung fields. In the proposed deformable model, SIFT is used to capture image features for guiding the segmentation of lung fields, since it has been validated as the best among many local descriptors [8].

Another important idea in the proposed method is that the deformable model is constrained by both population-based and patient-specific shape statistics. The shape statistics collected from the segmentation results of a population, i.e., the shapes of lung fields of individuals, can be used for constraining the segmentation of lung fields in the initial time-point images of a specific patient. As the number of serial images of the same patient increases with time, it is possible to collect the patient-specific shape statistics from previous segmentation results of the same patient and use it to guide the segmentation of lung fields in the serial chest radiographs of the same patient. In this way, the deformable model gradually adapts to the shape statistics of the specific patient and yields more robust segmentation results. In our method, the patient-specific shape statistics is updated each time after a new image is acquired and segmented; then the segmentation results of all the previous time-point images are further refined by using the updated shape statistics.

2   Method

2.1   Description

In the clinical setting, serial chest radiographs are acquired monthly from each patient. 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. At each time-point t, two major steps are performed for the newly captured image, i.e., the preprocessing step and the deformable segmentation step, as detailed next.

• The preprocessing step. The first time-point image is selected as the template image, and the image captured at time t is rigidly transformed onto the space of the template image by using rigid transformation [9]. Moreover, the image intensities are globally normalized to have a similar distribution to the one of the template image [10]. Thus, we obtain a normalized image at time t. All the normalized serial images are denoted as $\{I_i, i=1, 2,\ldots, t\}$, where I_t is the current time-point image.

• The deformable segmentation step. A new deformable model is utilized to segment the current image I_t, by using the statistical information collected from a population and also from t-1 previously segmented images of the same patient. The energy function of the deformable model is defined as,

$$E_t = E_t^{\text{Smth}} + E_t^{\text{SIFT}} + (1 - \omega_t)E_t^{\text{P}} + \omega_t E_t^{\text{I}} \tag{1}$$

where E_t^{Smth} denotes the smoothness constraint of the deformable contour, and it requires that the first and the second derivatives of the deformable contour be close to

zero. E_t^{SIFT} denotes the overall difference of the SIFT features of the corresponding model points in the image I_t and in the template I_0, which is actually a mean shape generated from a public database [2, 11]. E_t^{P} denotes the constraint derived from the population-based shape statistics, while E_t^{I} denotes the constraint required by the patient-specific shape statistics, trained from t-1 segmentation results on previous time-point images of the same patient. The energy term E_t^{I} is not existent for several initial time-point images, simply because the available segmentation results for the specific patient are not sufficient to train a statistical shape model. We can actually begin to train the patient-specific statistical shape model once N_s time-point images have been collected and segmented. Here, N_s is the minimal number of segmentation results for the same patient, and $N_s = 5$ is used in our study. Afterward, we can gradually increase the weight of energy term E_t^{I} and simultaneously decrease the weight of energy term E_t^{P}, as more and more new images are added. This is controlled by parameter ω_t:

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

If we have more than N_b images from the same patient, we will stop using the population-based shape constraint term E_t^{P}, since the patient-specific statistics collected from over N_b samples is sufficient to capture the variation of lung fields of the patient.

The deformable segmentation of image I_t is summarized as follows:

(1) If $t \leq N_s$, segment I_t by minimizing Eq. (1), where ω_t is 0 and E_t^{I} is not used.

(2) If $t > N_s$, the following two steps are performed (refer to Fig.1 for details):

Step (2.1), segment image I_t by incorporating the patient-specific constraint term E_t^{I}, which is derived from t-1 segmentation results on previous time-point images $I_1,...,I_{t-1}$.

Step (2.2), update the patient-specific shape statistics by adding a new segmentation result of I_t to the training set, and then use this updated patient-specific shape statistics to refine the segmentations on all t images $I_1,...,I_t$. Finally, the patient-specific shape statistics is updated again, which will be used for segmenting the next time-point image.
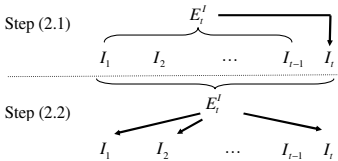

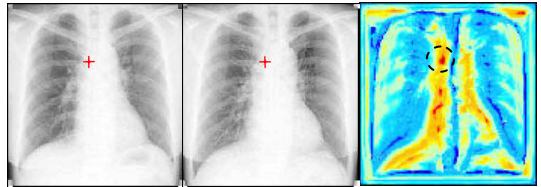**Fig. 1.** Two major steps for segmenting the current image at t

Fig. 2. Demonstration of using SIFT features for correspondence detection in two different time-point images

2.2   Energy Terms

Denoting the deformable contour as M ordered points, i.e., $\{v_i = (x_i, y_i) \mid i = 1, 2, ..., M\}$, the smoothness constraint term E_t^{Smth} can be represented as,

$$E_t^{\text{Smth}} = \sum_{i=1}^{M} \left(\alpha_i E_{t,i}^{\text{Cont}} + \beta_i E_{t,i}^{\text{Curv}} \right) = \sum_{i=1}^{M} \alpha_i (|\overline{d}_t - | v_{t,i} - v_{t,i-1} ||) + \beta_i (| v_{t,i-1} - 2v_{t,i} + v_{t,i+1} |) \quad (3)$$

where $E_{t,i}^{\text{Cont}}$ and $E_{t,i}^{\text{Curv}}$ are the continuity and the curvature constraints on the i-th point of the contour during the segmentation of lung fields at time t. α_i and β_i are the weights, and $\overline{d}_t$ is the average distance between neighboring points in the contour.

The second energy term in Eq. (1) E_t^{SIFT} is a distance measure between two SIFT features, and minimizing it enables us to determine point correspondences between the template I_0 and image I_t. As we will describe in the next subsection, SIFT features are distinctive, thus facilitating the correspondence detection in the chest radiographs. The SIFT features used in our study are designed as: for each point v_i, a 128-element vector s_i is used to represent the histogram of orientations in the local image around the point v_i. Then, the energy term E_t^{SIFT} can be defined as,

$$E_t^{\text{SIFT}} = \sum_{i=1}^{M} \prod_{l=1}^{128} \left| s_i^{\text{Mdl}}(l) - s_i^t(l) \right| \quad (4)$$

where s_i^{Mdl} and s_i^t represent the SIFT feature vectors of point v_i in the template image I_0 and its corresponding point in the image I_t. $s_i^{\text{Mdl}}(l)$ and $s_i^t(l)$ represent the l-th element of s_i^{Mdl} and s_i^t, respectively.

The third and forth energy terms are population-based and patient-specific statistical constraints, respectively. We use [5] to capture shape statistics from the population, i.e., a set of manually segmented training samples [2]. In contrast, we use a hierarchical shape representation in [12] to effectively capture the patient-specific shape statistics, since the number of segmentation results from the same patient is not large.

2.3   SIFT Local Descriptor

SIFT, as detailed in [7], consists of four major steps: (1) scale-space peak selection; (2) key point localization; (3) orientation assignment; (4) key point description. The first two steps detect the key points in the scale space, and they are not required in this study since we only focus on using SIFT features of the contour points along the boundaries of lung fields. Thus, we use the last two steps to compute the SIFT features for each contour point. The SIFT local descriptor for a point is computed as follows:

(1) A patch is centered on the point, rotated according to the dominant gradient orientation around the point, and scaled to an appropriate size.
(2) The SIFT local descriptor is then created by sampling the magnitudes and the orientations of the image gradients in the patch around the point, to build the smoothed orientation histograms for capturing the important information in this patch of image. A 4x4 array of histograms, each with 8 orientation bins, is obtained, capturing the rough spatial structure of this patch of image.

(3) This 128-element vector is then normalized to a unit length and further thresholded to remove the elements with small values.

Fig. 2 shows the performance of SIFT features in detecting 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 the left image are compared with the SIFT features of all the points in another time-point image as shown in the middle image. As indicated by the color-coded similarity map in the right image, a very small number of points in the middle image are similar to the red crossed point in the left image. Since the two images have been rigidly aligned, the corresponding point of the red crossed point in left image is detected in the middle image by searching the maximal similarity within a local neighborhood.

2.4 Learning Patient-Specific Shape Statistics

For serial image segmentation, it is important to use the segmentation results from the previous time-point images to guide the segmentation of the current time-point image. Similarly, the segmentation results of previous images can also be refined using the information collected in the later time-points. Since we have normalized intensities and poses of other time-point images to the first time-point image, all the lung fields are globally overlapped well. However, although the globally aligned serial lung field shapes are very similar, there still exist visible differences, due to distortion, pose change, 3D-2D projection, and normal or abnormal development. Fig. 3 shows an example of the serial lung field shapes. The thin black contours represent the boundaries of lung fields in each image, while the thick red contours are identical and they are actually the same contour segmented from the first time-point image. It can be seen that the segmentation results obtained from the serial images of the same patient are globally similar while locally different, because of the reasons mentioned above. Note that it is relatively easy to capture the shape statistics from the shapes of the same patient, even using a small number of shape samples. This type of patient-specific shape statistics, collected from the previous segmentation results, can guide the segmentation of the current and future time-point images. This is the main idea of our on-line learning method for capturing the patient-specific shape statistics.
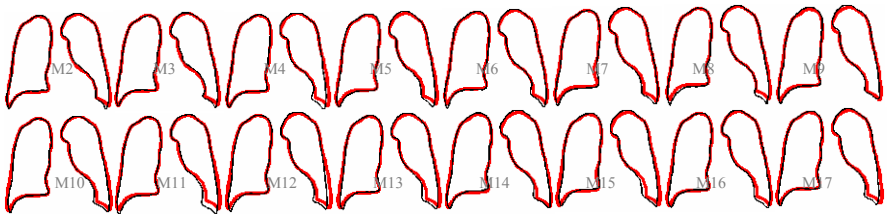

Fig. 3. Demonstration of local deformations of lung field shapes from the serial images of the same patient. The thin black contours are the segmentation results of the serial images, and the thick red contours are identical, denoting the same shape segmented from the first time-point image.

At time-point t, we have t-1 previously segmented images and a current time-point image I_t. We can use the t-1 segmentation results from images $I_1,...,I_{t-1}$ (in Step (2.1) of Fig.1), or the t segmentation results from images $I_1,...,I_t$ (in Step (2.2) of Fig.1), as the training samples. In order to capture the patient-specific shape variability information from the training samples, we adopt the hierarchical representation of shape statistics that can effectively guide the deformable segmentation of lung fields in the current time-point image [12].

It should be noted that this on-line and incrementally learned patient-specific shape statistics can be used only after a sufficient number (N_s) of segmentation results are obtained. As more images are processed, the constraint derived from the population-based shape statistics becomes less and less important, while the constraint derived from the patient-specific shape statistics increases gradually. Finally, the weight for population-based statistical constraint is 0 if we have N_b or more training samples. This mechanism enables the deformable model to adapt to the patient-specific shape variability, thus it is more robust and accurate for lung field segmentation.

3   Experimental Results

The serial frontal chest radiographs of 39 patients, most with up to 17 monthly scans, are used. The lung fields of all serial images have been manually delineated by a human observer, which are used as a gold standard to validate the segmentation results obtained by our method, a standard Snake [13], and ASM [6] in the next.

3.1   Segmentation of Serial Chest Radiographs

Visual comparison. Fig.4 qualitatively compares the performances of our method, the standard Snake, and ASM in segmenting the lung fields from the serial chest radiographs (256x256). Initializations for these three methods are displayed in the top left image. The top right image is a selected time-point image from the serial images. The red, green and black contours denote the segmentation results by the standard Snake, ASM, and our method, respectively. It can be observed that our results are very close to the lung field boundaries, and they are temporally more stable than the ASM results (please refer to the four small serial images shown on the bottom of the figure). These results indicate that the use of patient-specific shape statistics improves the accuracy of image segmentation. They also indicate that a simple feature, such as a local gradient used in the standard Snake, is not able to describe the complex and inconsistent image content along the boundaries of lung fields. On the other hand, SIFT features facilitate the detection of correspondences in the serial images, which ensures the relatively better segmentations by our method.

Quantitative comparison. We also used quantitative measures to compare the segmentation results obtained by the three segmentation algorithms. We use the overlay percentage to compare the automated segmentation results with the manual segmentation results for each image. It turns out that the mean and standard deviation of overlay percentage are 93.8%±0.58% by our method, 92.6%±0.8% by ASM, and 78.7%±1.3% by the standard Snake. We also compute the average contour distance between automated segmentations and manual segmentations [2]. The mean and standard deviation

of average contour distance are 1.76±0.23 pixels by our method, 2.06±0.25 pixels by ASM, and 5.31±0.29 pixels by the standard Snake, respectively. These results show that our method achieved better performance than the standard Snake and ASM.

3.2 1D and 2D Cardiothoracic Ratios

By segmenting lung fields from chest radiographs, we can compute the diagnostic measures, i.e., cardiothoracic ratio (CTR), referred to as 1D-CTR. 1D-CTR is defined as the ratio of the transverse diameter of the heart to the transverse diameter of the thorax, as shown in Fig.5. Recently, the extraction of 2D cardiothoracic ratio (2D-CTR) from the chest radiographs has been extensively studied [14], since it is believed to be more robust than 1D-CTR. In our study, 2D-CTR is defined as the squared root of the ratio between the area of white region and the area of both white and grey regions, as shown on the right of Fig.5.

The importance of 1D-CTR has been validated in many clinical applications. For evaluating the usefulness of 2D-CTR, we need to measure the relationship between 1D-CTR and 2D-CTR from the same serial chest radiographs. As shown by the left plot in Fig.6, 2D-CTR is highly related to 1D-CTR, which indirectly indicates that 2D-CTR can be also used as a diagnostic measure.

Moreover, the Bland and Altman plots [15] of 1D-CTR and 2D-CTR are provided in the middle and right plots in Fig.6, respectively, by using the manual segmentation results as the gold standard. These two plots indicate that the distribution of 2D-CTR is more compact than that of 1D-CTR. Also, the mean difference value by 2D-CTR measurement is smaller than that by 1D-CTR measurement. Moreover, the confidence interval of 2D-CTR is tighter than that of 1D-CTR.
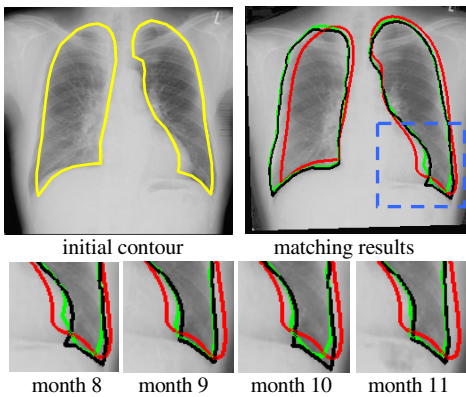
initial contour   matching results

month 8   month 9   month 10   month 11

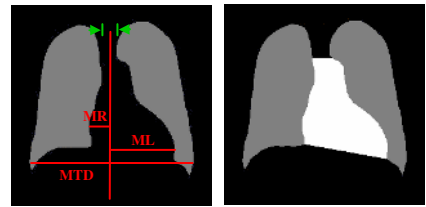**Fig. 4.** Comparison of automated segmentations

Fig. 5. 1D-CTR and 2D-CTR. In left image, MR+ML denotes the cardiac diameter, and MTD denotes the thoracic diameter. In right image, white region represents the central shadow in the chest radiograph, while gray region represents the bilateral lung fields.

4 Conclusion

We have presented a new deformable model to segment the lung fields from serial chest radiographs by using both population-based and patient-specific shape statistical constraints. The patient-specific shape statistics is on-line and incrementally learned

from the segmentation results of the previous time-point images of the same patient. For the initial time-point images, the population-based shape statistics plays the major rule for statistically constraining the deformable contours. As more and more segmentation results are obtained, the patient-specific shape statistics starts to constrain the segmentation, and it gradually takes the major rule for statistical constraining. Also, SIFT features, used to characterize image points, are relatively distinctive, thus making the correspondence detection in the two chest radiographs relatively easy. Experimental results show that our method can segment the lung fields more accurately, compared to the standard Snake and ASM. Finally, we demonstrated that 2D-CTR is highly related to 1D-CTR, but it is more robust to compute the diagnostic measure.
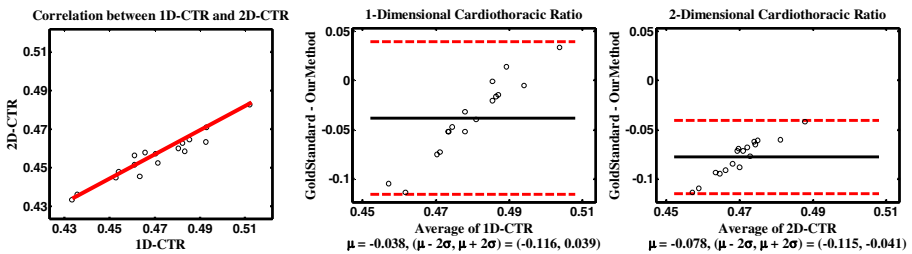

Fig. 6. Relationship between 1D-CTR and 2D-CTR (left), the Bland and Altman plots of 1D-CTR and 2D-CTR (middle and right), respectively

References

1. B. van Ginneken, B.M. ter Haar Romeny, M.A. Viergever. Computer-aided diagnosis in chest radiography: a survey. *IEEE Trans. on Med Imaging*, 20(12): 1228-1241, 2001.
2. 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. *Medical Image Analysis*, 10:19-40, 2006.
3. 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-ray images. *Computerized Medical Imaging and Graphics*, 22:463-477, 1998.
4. J.W. Jaeger and R.L. Mehta. Assessment of dry weight in hemodialysis: an overview. *Journal of American Social Nephrology*, 10:392-403, 1999.
5. T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38-59, 1995.
6. T.F. Cootes and C.J. Taylor. Statistical models of appearance for computer vision. *Technical Report, Wolfson Image Analysis Unit, University of Manchester*, 2001.
7. D.G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, 1999.
8. K. Mikolajczyk and C.Schmid. A performance evaluation of local descriptors. In *IEEE conference on Computer Vision and Pattern Recognition*, 2003.
9. M. Jenkinson and S. Smith. A global optimization method for robust affine registration of brain images. *Medical Image Analysis*, 5:143-156, 2001.
10. L.G. Nyul, J.K. Udupa, and X. Zhang. New variants of a method of MRI scale standardization. *IEEE Trans. on Medical Imaging*, 19(2): 143-150, 2000.

11. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., matsui, M., fujita, H., Kodera, Y., Doi, K., Development of a digital image database for chest radiographs with and without a lung nodule: receiver operation characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, 174: 71-74, 2000.
12. C. Davatzikos, X. Tao, and D. Shen. Hierarchical active shape models using the wavelet transform. *IEEE transaction on medical imaging*, 22(3):414-423, 2003.
13. M. Kass, A.Witkin, and D. Terzopoulos. Snakes: active contour models. *Internal Journal of Computer Vision*, 1(4): 321-331, 1987.
14. R.F.J. Browne, G. O'Reilly, and D. Mclnerney. Extraction of the two-dimensional cardiothoracic ratio from digital PA chest radiographs: correlation with cardiac function and the traditional cardiothoracic ratio. *Journal of Digital Imaging*, 17(2): 120-123, 2004.
15. J.M. Bland and D.G. Altman. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 346 (8982), 1085-1087, 1995.