Lung Segmentation in Chest Radiographs Using Anatomical Atlases With Nonrigid Registration


Abstract—The National Library of Medicine (NLM) is developing a digital chest X-ray (CXR) screening system for deployment in resource constrained communities and developing countries worldwide with a focus on early detection of tuberculosis. A critical component in the computer-aided diagnosis of digital CXRs is the automatic detection of the lung regions. In this paper, we present a nonrigid registration-driven robust lung segmentation method using image retrieval-based patient specific adaptive lung models that detects lung boundaries, surpassing state-of-the-art performance. The method consists of three main stages: 1) a content-based image retrieval approach for identifying training images (with masks) most similar to the patient CXR using a partial Radon transform and Bhattacharyya shape similarity measure, 2) creating the initial patient-specific anatomical model of lung shape using SIFT-flow for deformable registration of training masks to the patient CXR, and 3) extracting refined lung boundaries using a graph cuts optimization approach with a customized energy function. Our average accuracy of 95.4% on the public JSRT database is the highest among published results. A similar degree of accuracy of 94.1% and 91.7% on two new CXR datasets from Montgomery County, MD, USA, and India, respectively, demonstrates the robustness of our lung segmentation approach.

Index Terms—Chest X-ray imaging, computer-aided detection, image registration, image segmentation, tuberculosis (TB).

I. INTRODUCTION

Detecting the lung regions in chest X-ray images is an important component in computer-aided diagnosis (CAD) of lung health. In certain diagnostic conditions the relevant image-based information can be extracted directly from the lung boundaries without further analysis. For example, shape irregularity, size measurements, and total lung volume [1], [2] provide clues for serious diseases such as cardiomegaly [3], [4], pneumothorax, pneumoconiosis, or emphysema [5]–[7]. In the case of CAD-based identification of lung diseases, accurate lung boundary segmentation plays an important role in subsequent stages of automated diagnosis [8]–[10].

The National Library of Medicine, in collaboration with Indiana University School of Medicine, AMPATH (The Academic Model Providing Access to Healthcare), is developing a computer-aided system for screening and detecting the pulmonary pathologies in chest radiographs. This system is being developed as part of a project aimed at screening of tuberculosis (TB) patients in regions of the world with high incidence of disease but inadequate healthcare facilities. The initial screening region will be rural areas of western Kenya, using light weight portable X-ray scanners. The shortage of radiological infrastructure and radiologists in rural areas of Kenya necessitates an automated TB screening approach in such resource constrained regions.

One of the important steps in automatic analysis of chest X-ray images is to detect the lung boundaries accurately. There are a number of anatomical challenges and subtle cues involved in segmenting the lung region within a CXR. For example, for lung segmentation, the strong edges at the rib cage and clavicle region results in local minima in many minimization approaches. Segmenting the lung apex is another difficult problem because of the varying intensities in the upper clavicle bone region. Additional challenges include segmenting the small costophrenic angle, making allowances for anatomical shape variations such as varying heart dimensions or other pathology, and X-ray imaging inhomogeneities. Fig. 1 shows some examples of such variations in lung appearance.

In this paper, we present a lung boundary detection system incorporating nonrigid registration with a CXR database of pre-segmented lung regions to build an anatomical atlas as a guide combined with graph cuts based image region refinement. We presented a preliminary version of this paper in [11] and [12]. The initial work is significantly expanded in this paper to incorporate a deformable anatomical lung model using a novel nonrigid registration approach based on SIFT-flow, a detailed assessment of the approach compared to other state-of-the-art
methods using the validated Japanese Society of Radiological Technology (JSRT) dataset [13], and further experimental validation of the approach using two additional CXR databases.

A. Related Research in Lung Boundary Segmentation

Over the past decade, a number of research groups have worked on chest X-ray analysis, and various methods have been proposed for lung boundary segmentation. Ginneken et al. [14] classified the early segmentation methods for chest X-rays into roughly four categories: 1) rule-based methods, 2) pixel classification-based methods, 3) deformable model-based methods, and 4) hybrid methods. Our method is a hybrid approach to detect, register and robustly segment lung organ boundaries across a large patient population. We survey some of the recent results in X-ray image based lung segmentation; earlier work in computer analysis of chest X-ray images is covered in [14] and more recent work in [15].

Rule-based segmentation methods [16] contain sequences of steps and rules such as thresholding or morphological operations. These methods have mostly heuristic assumptions and compute approximate solutions that can be far from the global optimum. Therefore, they are generally used as an initialization stage of more robust segmentation algorithms [17].

Pixel classification-based methods are more general than rule-based methods. They mainly model the intensities of inside and outside of the lung regions, and classify the image pixels into either object (lung inside) or background [18]–[20], [15], [17].

Deformable models have been extensively studied and used in medical image segmentation because of their shape flexibility. Active shape models (ASM) [21] and active appearance models (AAM) [22] have been successfully applied to lung region segmentation [23], [24], [15], [25]–[27]. Although active shape and appearance modeling approaches have become popular for biomedical applications, they have several limitations and shortcomings including: 1) they can become trapped at local minima in chest X-rays due to high contrast and strong rib cage edges, 2) segmentation performance relies on the approximation accuracy of the initial model, and 3) they have many internal parameters which produces highly variable solutions. For example, the two implementations of ASM in [15] and in [25] produced significantly different results on the same dataset because of the free parameters in the scheme. Extensions of ASMs for lung field segmentation are actively being investigated [27], [28], [29] in order to overcome these limitations. For example, in [30], a shape particle filtering approach is used to prevent getting trapped at local minima. Another recent effort uses SIFT descriptors and a shape representation utilizing both population-based and patient-specific shape statistics for matching serial images of the same patient [25]. A modified ASM approach presented in [31] optimizes shape and intensity characteristics simultaneously during ASM iterations. In addition, a similar effort in [26] incorporates global edge and region forces as part of the objective function to reach the global minimum.

Hybrid methods aim to produce better results by fusing several techniques. In [32], a rule-based and a pixel-based approach are combined. In [15], researchers propose three hybrid approaches, fusing deformation-based (active shape model, active appearance model) and pixel classification methods by choosing the best performing approach using majority voting.

In [8], the lung region is extracted using a combination of an intensity mask, a lung model mask derived from a training set, and a Log-Gabor mask. Recently, a new algorithm has been proposed for emphysema detection, in which the lung boundaries are modeled as a closed fuzzy-curve. The fuzzy-membership is estimated with Kohonen networks [7].

B. System Overview and Our Contribution

We propose a robust automated lung segmentation system for chest X-ray images. Our method shown in Fig. 2 consists of three main stages. First we use a content-based image retrieval approach to identify a small set of lung CXR images that are most similar to the patient X-ray using partial Radon transforms combined with a Bhattacharyya similarity measure. The partial Radon transform based retrieval method is fast and can accommodate small affine distortions in the CXR. The highly ranked lung images retrieved by this method are usually a good fit for the patient lung and is a successful application of medical CBIR methods to anatomical atlas construction. After ranking, we compute an anatomically guided patient-specific lung model by warping the training database of chest X-rays to the patient X-ray using a state-of-the-art deformable registration algorithm [33]. Finally, the lung boundaries are determined using a graph cuts discrete optimization approach [34]–[36], [11] with a customized energy function. The graph cuts energy function includes a novel anatomical atlas shape prior term that ensures close adherence to normal lung anatomy.

In order to estimate an approximate lung model, we employ a deformable registration method that solves the alignment process in an energy minimization framework. It first calculates the corresponding pixels between training and patient X-ray which provides the transformation mapping for each pixel. Then it aligns the training masks using the transformation mapping. We use the average of warped masks as a lung atlas model for the patient. Registration follows content-based ranking and selection from the training database in order to speed up the overall segmentation algorithm, since nonrigid registration is computationally expensive. Finally, our system detects the lung boundaries with a segmentation algorithm. We use the graph cuts algorithm, which models the segmentation process using an objective function in terms of boundary, region, and lung model properties. The graph cuts algorithm computes a global binary segmentation by minimizing the objective function.
Although active shape and appearance models are popular for segmenting anatomical parts, we show that our graph cuts approach does not get trapped at local minima and outperforms these methods.

The proposed system combines two state-of-the-art algorithms in a novel manner for organ segmentation in radiographical images. Both algorithms model the deformable registration and segmentation processes in an energy minimization framework. We test the system on three challenging datasets, one of which has been publicly available and widely used in the literature (JSRT). We use two additional CXR datasets from Montgomery County, MD, USA, and India, which we are making publicly available to the research and clinical communities to improve lung segmentation algorithms for regions of the world where pulmonary diseases such as TB remains an endemic health concern. These are the first new datasets for CXR lung segmentation application since the JSRT became available more than a decade ago, as well as the first to include abnormal lung boundaries and the first publicly available CXR datasets specifically for TB screening.

The paper is structured as follows. Section II, describes how the lung model is calculated using the deformable registration algorithm. In Section III, we first give the basic terminology of the graph cuts method and then describe the segmentation algorithm in detail. Section IV provides a description of the datasets and experimental results. Conclusions are given in Section V.

II. PATIENT-SPECIFIC STATISTICAL LUNG ATLAS MODEL USING NONRIGID REGISTRATION

Segmentation in medical imaging poses a number of challenges including multiplicative noise, motion during imaging, sampling artifacts caused by the acquisition equipment, low contrast, deformation of tissues and anatomical shape variations due to normal anatomy and disease. Therefore, classical segmentation techniques, which make simplifying assumptions of rigid motion or additive noise for example, do not use a priori information, usually produce unsatisfactory results on medical images. In order to provide a priori information for improved segmentation, we incorporate a lung atlas model into the system. Since the X-ray images contain variable lung shapes, a static model is not sufficient to describe the lung regions. Our system therefore estimates a statistical model for each patient X-ray using a training set of segmented images (atlases) to identify the most similar images followed by a nonrigid registration algorithm to warp the most similar training masks to the patient CXR.

A. CBIR Paradigm for Inter-Patient Matching

We first identify a small subset of images (i.e., five) in the training database that are most similar to the patient query image, using a content-based image retrieval (CBIR) inspired approach, and use this subset of training images including corresponding lung masks to develop a patient-specific lung model. Using a small subset of images from the database is sufficient to build an accurate lung model while significantly speeding up the step of nonrigid registration between the training and the patient query images. Ranking precedes registration, otherwise we would need to extract SIFT features and compute SIFT-flow deformable registration models for every image in an extensive training database which is prohibitively expensive and impractical for a fieldable system.

Unlike other patient-specific lung models in the literature that use intra-patient image information [25], we develop an inter-patient matching and image retrieval system that follows the CBIR paradigm to guide segmentation. CBIR systems are designed to be fast for online retrieval applications with an offline preprocessing step to extract signature features for each image in the database [37]–[39] and can incorporate multimodal information to improve precision [40]. CBIR systems usually produce a ranked subset of images most similar to the query which in our case is a new patient CXR image. We assume that the CXR database has been appropriately preprocessed and consists of globally aligned and normalized CXRs.

We use partial Radon transforms, or orthogonal projection profiles, to compare and rank the similarity between two patient’s lung images. The Radon transform projection along an arbitrary line in the x-y plane is defined as

$$\mathcal{R}(\rho, \theta) = \int \int I(x, y) \delta(\rho - x \cos \theta - y \sin \theta) dx \, dy$$

(1)

where $\delta(\cdot)$ is the 2-D impulse function

$$\int_{-\infty}^{\infty} \delta(x, y) dx \, dy = 1$$

(2)

and has the shifting property

$$\int_{-\infty}^{\infty} I(x, y) \delta(x - x_0, y - y_0) dx \, dy = I(x_0, y_0)$$

(3)
The Radon transform computes a projection of the image as a sum of line integrals accumulating pixel intensities along rays defined by $\rho = x \cos \theta + y \sin \theta$ in the $x$-$y$ plane. Instead of a complete set of $360^\circ$ angles, we use a small subset of angles which we refer to as the partial Radon transform and use these few projections for image matching and similarity assessment. The partial Radon transform projection method is fast to compute and only an approximate matching atlas set of lung segmentations from the CXR database is needed to compute a spatial prior that can be refined in the subsequent phase of the algorithm. Our X-ray sets contain only a small number of slightly rotated images. Therefore, in the experiments two orthogonal projections, horizontal and vertical, are sufficient to accommodate small patient rotations around the projection axis and translations in the CXR. Large image rotations and shears can result in a higher degree of global alignment mismatch between the patient CXR and the normalized set of CXRs in the database. Input images with large rotations can be dealt with by using a larger number of projections in the Radon transform at a slightly increased computational cost. Such rotation in X-rays can be also calculated using rib cage or clavicle bone direction [41], [42], [15].

The horizontal and vertical projection profiles are precomputed for all images in the training database to speed up the CBIR search process. We first calculate the intensity projection of the histogram-equalized images in the vertical and the horizontal directions. Then we measure the similarity of each projection profile between the atlas database and the patient chest X-ray using the average Bhattacharyya coefficient

$$BC(I_1, I_2) = \alpha \sum_{x=1}^{n} \sqrt{p_1(x)p_2(x)} + (1 - \alpha) \sum_{y=1}^{m} \sqrt{q_1(y)q_2(y)},$$

where $p_1(x)$ and $p_2(x)$ are the horizontal projections, $q_1(y)$ and $q_2(y)$ are the vertical projections of images $I_1$ and $I_2$, respectively, $x$ and $y$ are the histogram bins of the projection profiles, $n$ and $m$ are the number of bins in the profile histograms, and $\alpha = n/(n + m)$ is the weight for each profile; when $n = m$, $\alpha = 1/2$. Fig. 3 demonstrates the horizontal and vertical profile histograms of two example images. The Bhattacharyya coefficient measures the overlap between two distributions and ranges between zero and one. It is easy to interpret and fast to compute. Other metrics between histograms such as entropy or cross correlation could be used to measure the degree of similarity between the patient X-ray and the atlas. But these are computationally more expensive compared to the partial Radon transform, which just involves integer sums.

We select a set of best fit training atlases from the anatomical database of segmented lung images to learn a patient specific lung model. The registration performance is significantly improved when a personalized lung model is designed by comparing the patient X-ray with presegmented lung images in the CXR database using a fast shape similarity measure based on partial Radon transforms.

### B. SIFT-Flow Deformable Warping of Lung Atlas

Image registration is an important task for many medical applications such as comparing/fusing images from different modalities, tracking temporal changes in medical images collected at different times. A registration scheme calculates a transformation mapping from source image to target image by matching corresponding pixels of images. Correspondences can be calculated either for each pixel or only for salient locations such as edge points or corners. Images can be modeled using local feature descriptors such as Scale Invariant Feature Transform (SIFT) [43], or using a combination of gradient, shape, and curvature descriptors, as in [44], [45]. In this work, we use the SIFT descriptor which is among the best performing local feature descriptors [46].

Nonrigid registration is a challenging problem as each pixel has a greater degree of independent motion and global geometric constraints cannot be enforced. Several nonrigid registration algorithms have been reported in the literature [47]–[56]. Some of them focused on registering different views of the same scene in which a relatively simple transformation will be sufficient for registration. In our case, in order to create a lung model, we register chest X-rays from different patients. In other words, the image pairs are similar, but have different objects. Therefore, the registration process needs an elaborate transformation mapping. Our work is focused on inter-patient similarity with deformable warping for creating a patient specific lung shape atlas. We found that the SIFT-flow algorithm [33], which is designed to register images with different scene/object appearances, worked well for this task.

The SIFT-flow algorithm models local gradient information of the observed image using the Scale Invariant Feature Transform (SIFT) [43]. The SIFT features of the X-rays are calculated as follows. First, the gradient orientations and magnitudes are computed at each pixel. The gradients are weighted by a Gaussian pyramid in a $K \times K$ region (e.g., $K = 16$) in order to increase the influence of the gradient in the center. Then, the regions are subdivided into $k \times k$ (e.g., $k = 4$) quadrants. In each quadrant, a gradient orientation histogram is formed by adding the gradient values to one of eight orientation histogram bins. The concatenation of orientation histograms of the quadrants form the SIFT descriptor vector for the center pixel of the $K \times K$ region. Once we have calculated the SIFT features for the
image pair, the registration algorithm computes pixel-to-pixel correspondences by matching the SIFT descriptors. The correspondence matching is formulated using the following objective function:

$$E(w) = \sum_{p \in P} \min(||S_1(p) - S_2(p + w(p))||, \gamma) + \sum_{p \in P} (u(p) + v(p)) + \sum_{(p,q) \in N} \min(|u(p) - u(q)|, d) + \min(|v(p) - v(q)|, d)$$

where $P$ is the set of pixels in the X-ray; $N$ is the spatial neighborhood set, $S_1$ and $S_2$ are the SIFT images in which each pixel is represented by a SIFT descriptor vector; $w(p) = (u(p), v(p))$ are the flow vectors at $p$; $t$ and $d$ are the truncated thresholds. The minimization algorithm calculates the SIFT-flow $w$ by minimizing the objective function. The first term of the objective function forces the algorithm to match pixels according to their SIFT descriptors, with warping based on the registration flow vector $w(p)$. The second term constrains the flow vectors to be as small as possible. The third term constrains the flow vectors of neighboring pixels to be similar.

Fig. 4 illustrates the registration stage of the proposed system. Fig. 4(b) is the patient X-ray. Fig. 4(a) is the most similar X-ray to the patient X-ray in the database. Colored markers indicate corresponding matches for sample pixels. We see that the lung boundary in one X-ray image approximately matches the lung boundary in the other X-ray. The spatial shifts between corresponding matches define the transformation mapping for pixels. The algorithm applies the transformation mapping by simply shifting each pixel in the training mask according to the calculated shift distance [Fig. 4(c)]. The registered mask is shown in Fig. 4(d). The registration stage is repeated for each of the top-$n$ (e.g., $n=5$) similar X-rays to the patient X-ray. The lung model for the patient X-ray is built-up using the mean of the top-ranked registered masks. The computed patient specific lung model is a probabilistic shape prior in which each pixel value is the probability of the pixel being part of the lung field.

Feature based registration algorithms are not accurate within textureless regions of images due to a lack of features. In our case, we are only interested in registering the lung regions which contain enough texture for a reliable correspondence calculation. There may be some exceptional cases, such as dense breasts (in women's X-rays) or fluid in the lung space (in abnormal X-rays), which decrease the visibility of texture. However, adjusting the contrast with a histogram equalization algorithm in preprocessing stage helps to uncover some texture information in these areas if the opacity is not severe.

Experimental results in Section IV-C validate the adequacy of the nonrigid registration approach using SIFT-flow for anatomically driven alignment of lung images.
Fig. 5. (a) Red contour is the nonsmoothed boundary after the warping stage. (b)–(d) Blue dots are the critical points at different iterations. The green curve is the cubic spline interpolation of all critical points. The numbers of critical points on each lung are 100, 60, and 30, respectively. At each iteration, a point with the lowest relevance value is removed from the contour. Outer turn angles close to 180° and short line segments have a low relevance value. Note that, decreasing the number of critical points does not smooth the costophrenic angle region because of the sharp angle.

Fig. 6. (a) Randomly selected chest X-ray image from the JSRT dataset. (b) Patient-specific lung model after registration. Each pixel intensity in the lung model image represents the probability of the pixel being part of the lung region.

C. Anatomical Lung Atlas: An Example

In this section, we illustrate the image retrieval framework followed by the nonrigid registration stage with an example. Fig. 6(a) shows a random X-ray from the JSRT dataset, and the corresponding estimated lung model is shown in Fig. 6(b). The algorithm chooses the most similar X-rays by comparing intensity projections of images. The top five most similar X-rays to the patient X-ray are shown in Fig. 7(a). The SIFT-flow algorithm computes correspondences between the patient and the training X-rays. The spatial distances between the corresponding matches are given by the transformation mapping between the pixels. In order to show the visual success of the registration stage, we warped the training images with the calculated transformations. The resulting X-rays are shown in Fig. 7(c). Note the similarity between the patient and the warped X-rays. Our system computes the prior lung model by applying the transformation mapping to the training masks [Fig. 7(d)]. The average sum of the registered masks forms the lung model for the patient X-ray. The calculated lung model is shown in Fig. 6(b).

III. GRAPH CUTS BASED LUNG BOUNDARY SEGMENTATION

The system detects the lung boundary of X-ray images using image properties and the lung model calculated in the previous stage. We perform image segmentation using graph cuts [34]–[36] and model the segmentation process with an objective function. The max-flow min-cut algorithm [59] minimizes the objective function to find a global minimum which corresponds to the foreground (fg) and the background (bg) labeling of the pixels. This section provides the details of the segmentation component of our system.

A. Graph Cuts: Terminology

The graph cuts algorithm models computer vision problems using an undirected graph $G = (V, E)$. The set of vertices $V$ represents the pixel properties such as intensity; and the set of edges $E$ connects these vertices. The edge weights represent a spatial proximity measure between the vertices. The graph has two extra vertices (terminals) representing the fg and the bg labels. There are two types of edges: 1) neighborhood edges denoted as $\{p, q\}$ where $p, q \in V$ model the boundary properties of objects; and 2) edges between terminals and pixels denoted as $\{p, S\}$ and $\{p, T\}$, where $S$ and $T$ represent the fg and the bg terminals. The graph structure is formulated in an objective function that consists of a data and a smoothness term. The data term forces the algorithm to produce a solution that is consistent with the data (e.g., image intensities). On the other hand, the smoothness term encourages the algorithm to favor a smooth solution (e.g., assigning similar labels to neighborhood pixels). The edge weights between the terminals and the pixels
are integrated into the data energy term; the neighborhood edges are integrated into the smoothness energy term of the objective function. To minimize the objective function, we compute the min-cut which partitions the graph into two subgraphs such that each pixel is connected to either the $S$ or the $T$ terminal, and thus is either labeled as fg or bg.

### B. Objective Function

The objective function is formulated based on the desired segmentation criteria including: 1) the segmentation labels (fg/bg) should be consistent with the image intensities of the lung; 2) the neighborhood labels should be consistent with each other, and 3) the resulting segmentation should fit the calculated shape model. Let $f = \{f_1, \ldots, f_p, \ldots, f_P\}$ be a binary vector whose components $f_p$ correspond to fg/bg label assignments to pixels $p \in P$, where $P$ is the set of pixels of the image. The algorithm aims to find an optimal configuration of $f$ according to the specified constraints. Based on the segmentation criteria, we define the objective function in terms of boundary, region, and shape model properties of the pixels as follows:

$$E(f) = \alpha_1 E_d(f) + \alpha_2 E_s(f) + \alpha_3 E_m(f)$$

(8)

where $E_d$, $E_s$, and $E_m$ represent the data, smoothness and lung model terms of the objective function. The scalar weights $\alpha_1$, $\alpha_2$, $\alpha_3$ are empirically determined; we use $\alpha_1 = \alpha_3$ and a relatively low weight $\alpha_2$ on the smoothness constraint in order to preserve features such as the sharp costophrenic angles. In order to confine the segmentation labels to be close to the image intensities, we formulate the edge weights between the terminals and the pixels as follows:

$$E_d(f) = \sum_{p \in P} \left( E_{\{p, S\}} + E_{\{p, T\}} \right)$$

(9)

$$E_{\{p, S\}} = I_p - I_T / I_{max}$$

(10)

$$E_{\{p, T\}} = I_p - I_S / I_{max}$$

(11)

where $I_p$ denotes the intensity of pixel $p$, $I_S$ and $I_T$ are the virtual intensities of object and background terminals, and $I_{max}$ is the maximum intensity value of the observed image. We model the terminal intensities using our training masks. $E_{\{p, S\}}$ and $E_{\{p, T\}}$ measure how well the assigned labels fit the image intensities.

We assign a high weight to neighborhood pixels that have similar intensities to ensure that nearby pixels have similar labels. The boundary penalties between pixel $p$ and $q$ are formulated as follows:

$$E_s(f) = \sum_{p,q \in N(p)} E_{\{p,q\}} \delta_{f_p \neq f_q}$$

$$= \sum_{p,q \in N(p)} \exp(-|I_p - I_q|^2) \delta_{f_p \neq f_q}$$

(12)

where $p$ is the current pixel, and $q \in N(p)$ are its neighbors, and

$$\delta_{f_p \neq f_q} = \begin{cases} 1, & \text{if } f_p \neq f_q \\ 0, & \text{if } f_p = f_q. \end{cases}$$

(13)

The exponential term is inversely proportional to edge strength. Flat regions in the image have high values to prevent cuts that break up such smooth regions. Whereas along edges the smoothness term has low values and cuts can separate the edge set into separate regions.

We incorporated the patient specific lung atlas model into the graph edge weights between the terminal nodes and pixel nodes $p$. As explained in Section II-B, the lung model is calculated by registering the top $n$ most similar X-rays to the patient X-ray. It is formed as a 2-D array that has same size as the observed image and contains the probabilities of the pixels being part of the lung field. The lung model energy is incorporated as weights associated with the source (fg) and terminal (bg) nodes as follows:

$$E_m(f) = \sum_{p \in S} \Pr_p + \sum_{p \in T} (1 - \Pr_p)$$

(14)

where $\Pr_p$ indicates the probability of pixel $p$ being part of the lung model.

After formulating the objective function, the next step is to compute the global minimum (min-cut) corresponding to the optimal labeling configuration satisfying the formulated constraints. We use a fast implementation of min-cut/max-flow [59]. The global minimum separates the graph into two subgraphs in which some pixels are connected to the fg terminal and the other pixels are connected to the bg terminal.

### IV. Experimental Results

#### A. Digital Chest X-Ray Datasets

In this work, we evaluated the proposed lung segmentation algorithm using three different CXR datasets. To the best of our knowledge, the only publicly available database for evaluating lung segmentation in chest X-ray imagery is the JSRT dataset which became available more than a decade ago in 2000 [13]. Most authors evaluate their lung segmentation methods on the publicly available JSRT set. However, we think that evaluations based on this dataset are flawed. While segmentations of healthy lungs can be properly evaluated on the JSRT set, evaluating the segmentations of abnormal lungs is a problem. The reason for this lies in the limited number of abnormalities contained in the JSRT set. The only abnormality covered by the set are nodules, which in most cases do not affect the shape of the lung, especially when they are small or are not located in the peripheral lung region. Most of the lung shapes in the JSRT set can therefore be considered normal. Other lung diseases such as tuberculosis and pneumonia, on the other hand, can severely affect the lung shape. In the extreme case of a collapsed lung, or an effusion, the lung shape will differ dramatically from the lung shape of a healthy lung. These abnormal lung shapes can cause problems for segmentation methods using lung models that have been computed on healthy lungs only. It is therefore essential to include abnormal lung shapes in the evaluation set of any automatic lung segmentation algorithm.

At NLM, we used two additional chest X-ray datasets from Montgomery County, Maryland, and India. The datasets were de-identified by the data providers and were exempted from IRB review at their institutions. The datasets were exempted from IRB review (No. 5357) by the NIH Office of Human Research Protections Programs. The new datasets are more challenging...
than the JSRT dataset. For example, the JSRT dataset contains X-rays captured using an analog imaging system by exposing film then later scanned into digital format [15]. These film-based images have a homogeneous intensity appearance. On the other hand, the chest X-rays in the India and Montgomery datasets were acquired using digital scanners. The images have much higher contrast in which rib cage edges are stronger and intensity variations are more heterogeneous compared to JSRT. These sets allow more realistic evaluations of our segmentation method because they contain both normal and abnormal lung shapes.

1) JSRT Set: Set compiled by the Japanese Society of Radiological Technology (JSRT) [13]. The set contains 247 chest X-rays, among which 154 have lung nodules (100 malignant cases, 54 benign cases), and 93 have no nodules. All X-ray images have a size of 2048 × 2048 pixels and a grayscale color depth of 12 bit. The pixel spacing in vertical and horizontal directions is 0.175 mm. The JSRT set is publicly available and has gold standard masks [15] for performance evaluation. In our experiments, we use JSRT masks as training masks for the registration stage.

2) Montgomery Set: Set from the Department of Health and Human Services, Montgomery County, Maryland. The set contains X-rays collected over many years under Montgomery County’s Tuberculosis Control program. The dataset consists of 138 X-rays, 80 of them are normal and 58 X-rays are abnormal with manifestations of tuberculosis. The X-ray images are in 12-bit grayscale and their dimensions are either 4020 × 4892 or 4892 × 4020. The pixel spacing in vertical and horizontal directions is 0.0875 mm.

3) India Set: Contains 397 chest X-rays from a private clinic in India with resolutions of 2446 × 2010, 1772 × 1430, or 2010 × 1572. The gray-scale color depth is 12 bit. The pixel spacing in vertical and horizontal directions is 0.175 mm. It contains more challenging X-ray images due to rotated torsi scans, stronger edges within the rib cage, denser shade effects at apices, wide variations of lung sizes and large lung deformations. In the experiments, we used a subset of the India set, which contains arbitrarily selected 100 normal and 100 abnormal chest X-rays with gold standard masks.

We manually generated gold standard segmentations for the chest X-ray sets under the supervision of a radiologist (JPM). The gold standard lung boundary segmentations for the JSRT dataset are from [15], and we used similar conventions in outlining the boundary and determining what parts of the CXR image to include as part of the lung organ and what regions to exclude. Both posterior and anterior ribs are readily visible in the CXRs, the part of the lung behind the heart is difficult to see and is excluded. We follow anatomical landmarks such as the boundary of the heart, aortic arch/line, pericardium line, or hilar and draw an inferred boundary when the pathology is severe, including pleural effusion, pneumothorax, pneumonia, or dense hilar affecting the morphological appearance of the lungs, and sharp costophrenic angle that follow the diaphragm boundary. We first used an interactive segmentation tool [60] for a quick first segmentation pass through the dataset. The tool detects edges automatically, allowing us to obtain a rough outline of the lung boundaries with just a few mouse clicks. However, these contours are not accurate and jagged, so in a second pass, under the supervision of a radiologist, we corrected any deviations from the proper anatomic lung boundary in these rough outlines. For the second pass, we used the web-based labeling tool FireFly [61], [62], which offers more precise drawing and editing functionality, powerful data management and online labeling capabilities. The net improvement in the foreground lung mask between the two passes, using (1 − Ω) measure [see (15)], was around 5.8% and 4.0% for Montgomery and India datasets respectively.

Benchmarking was done at a reduced image resolution of 256 × 256 to be compatible with published results. The lower resolution provides automatic boundary smoothing, slightly better segmentation results due to improved signal-to-noise ratio with better suppression of distracting features from the rib cage and clavicle bones, and yields faster performance.

B. Evaluation Metrics

Literature proposed several algorithms with different evaluation metrics. In order to compare our segmentation quality with the segmentation performances in the literature, we used three commonly used metrics.

1) The Jaccard Similarity Coefficient (overlap measure): is the agreement between the ground truth (GT) and the estimated segmentation mask (S) over all pixels in the image. We formulate it as follows:

Ω = \frac{|S \cap GT|}{|S \cup GT|} = \frac{|TP|}{FP + |TP| + |FN|} \quad (15)

where TP (true positives) represents correctly classified pixels, FP (false positives) represents pixels that are classified as object but that are in fact background, and FN (false negatives) represents pixels that are classified as background but that are in fact part of the object.

2) Dice’s Coefficient [63]: is the overlap between the ground truth GT and the calculated segmentation mask S

DSC = \frac{2|TP|}{2|TP| + |FP| + |FN|} \quad (16)

3) Average Contour Distance (ACD): is the average distance between the segmentation boundary S and the ground truth boundary GT. Let \( a_i \) and \( b_j \) be the points on the boundary S and GT, respectively. The minimum distance of point \( a_i \) on S to the GT boundary is defined as follows:

\[ d(a_i, GT) = \min_j ||b_j - a_i||. \quad (17) \]

For ACD computation, the minimum distance for each point on the boundary S to the contour GT is computed. Then, the distances are averaged over all points of boundary S. In order to make the similarity measure symmetric, the computation is repeated from contour GT to contour S

\[ \text{ACD}(S, GT) = \frac{1}{2} \left( \sum_i d(a_i, GT) + \sum_j d(b_j, S) \right) \quad (18) \]

where \(| \cdot |\) is the cardinality of the set.
TABLE I
PARAMETER VALUES USED IN DIFFERENT MODULES OF THE LUNG BOUNDARY SEGMENTATION PIPELINE

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape Model Parameters:</td>
<td></td>
</tr>
<tr>
<td># of training masks</td>
<td>5</td>
</tr>
<tr>
<td>SIFT Feature Parameters:</td>
<td></td>
</tr>
<tr>
<td>window around the keypoint (pixel)</td>
<td>4 x 4 array</td>
</tr>
<tr>
<td># of orientation quantization bins</td>
<td>8-bins</td>
</tr>
<tr>
<td>Non-rigid Registration Parameters:</td>
<td></td>
</tr>
<tr>
<td>α, the weight of the truncated L1-norm regularization</td>
<td>2</td>
</tr>
<tr>
<td>γ, the weight of the magnitude of the flow</td>
<td>0.005</td>
</tr>
<tr>
<td>θd truncation thresholds</td>
<td>40</td>
</tr>
<tr>
<td># of iterations</td>
<td>30</td>
</tr>
<tr>
<td>Graph Cuts Parameters:</td>
<td></td>
</tr>
<tr>
<td>α1, data term weight</td>
<td>1</td>
</tr>
<tr>
<td>α2, smoothness term weight</td>
<td>0.05</td>
</tr>
<tr>
<td>α3, lung model term weight</td>
<td>1</td>
</tr>
</tbody>
</table>

TABLE II
OVERLAP, DICE, AND AVERAGE CONTOUR DISTANCE SCORES BETWEEN THE GOLD STANDARD AND THE CALCULATED SEGMENTATIONS FOR EACH X-RAY DATASET

<table>
<thead>
<tr>
<th></th>
<th>JSRT (all)</th>
<th>JSRT right</th>
<th>JSRT left</th>
<th>Montgomery (all)</th>
<th>Montgomery right</th>
<th>Montgomery left</th>
<th>India (all)</th>
<th>India right</th>
<th>India left</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ω</td>
<td>0.954 ± 0.015</td>
<td>0.967 ± 0.008</td>
<td>1.321 ± 0.316</td>
<td>0.941 ± 0.034</td>
<td>0.960 ± 0.018</td>
<td>1.599 ± 0.742</td>
<td>0.917 ± 0.048</td>
<td>0.947 ± 0.028</td>
<td>2.567 ± 1.454</td>
</tr>
<tr>
<td>DSC</td>
<td>0.951 ± 0.020</td>
<td>0.967 ± 0.010</td>
<td>1.216 ± 0.368</td>
<td>0.935 ± 0.042</td>
<td>0.959 ± 0.025</td>
<td>1.453 ± 0.875</td>
<td>0.913 ± 0.054</td>
<td>0.948 ± 0.031</td>
<td>1.760 ± 1.193</td>
</tr>
<tr>
<td>ACD</td>
<td>0.957 ± 0.017</td>
<td>0.967 ± 0.011</td>
<td>1.426 ± 0.453</td>
<td>0.948 ± 0.041</td>
<td>0.961 ± 0.023</td>
<td>1.744 ± 1.024</td>
<td>0.922 ± 0.055</td>
<td>0.946 ± 0.034</td>
<td>1.887 ± 1.144</td>
</tr>
</tbody>
</table>

C. Segmentation Performance of the System

We measure the segmentation performance of the system on the aforementioned datasets. Table I summarizes the main parameter values used in different modules of the lung segmentation pipeline.

Table II lists the average Jaccard, Dice score, and contour distances of segmentation performances of three datasets, including left and right lung segmentation performances. Due to homogenous intensity appearance of JSRT set, the segmentation performance is higher compared to the segmentation performance on Montgomery and India sets. The quantitative results of all segmentations are shown in Fig. 8. We see that the overlap scores of almost all segmentations (93.5% of all cases) are higher than 0.90. Most scores are around the typical human observer score (Ω = 0.94). We get a few cases around Ω = 0.80. Fig. 9 shows the visual quality of computed lung boundaries for the three datasets. The green and red contours represent the gold standard and the calculated segmentations, respectively. Fig. 10 shows some example results where the method failed due to abnormalities in the lung region.

D. Computational Speed of the Lung Segmentation Algorithm

The lung segmentation algorithm is only one component of a full CAD pipeline for tuberculosis screening using CXRs. A fielded system working in rugged conditions with minimal access to technical and healthcare experts needs to be robust in terms of accuracy and near realtime in terms of performance. Image subsampling to a lower resolution prior to segmentation speeds up runtime significantly while having a negligible impact on accuracy. The computationally expensive numerical algorithms for energy optimization are implemented in C++ and other parts in Matlab. We report the execution times of our lung segmentation algorithm on a desktop personal computer with a 2.53-GHz Intel Xeon CPU and 4 GB of memory in Table III.

E. Comparison With Existing Algorithms in the Literature

Many papers have reported good segmentation results on chest X-ray images. Ginneken et al. [32] reported several pixel classifier algorithm scores to compare their rule-based scheme. In their subsequent work [14], quantitative results of some early segmentation algorithms are listed. However, these algorithms are conducted on nonpublic datasets. Therefore, comparison of our segmentation results with the segmentation results of these algorithms is not possible. In 2006, the same research group [15] delineated lung, heart, and clavicle boundaries of the JSRT set [13] under the supervision of a radiologist. Lung boundary contours were drawn by two observers independently, with one segmentation set done by a medical student trained by a radiologist being utilized as the gold standard for evaluation. The second observer’s segmentations (a nonmedical student also trained by a radiologist) are used to measure inter-observer accuracy in order to compare the algorithm results with human performance. Tables IV and V list the results of our system on the JSRT dataset together with the results reported in the literature. The accuracy of our proposed system is Ω = 0.954 ± 0.015. (It is even higher at 0.961 ± 0.020 without the smoothing stage of the registered lung model). The improved pixel classification, hybrid voting algorithms, and our system achieve a human observer performance level, which indicates the success of the computerized methods. Two different observer accuracies are reported in the literature. In [32], 98.4% accuracy is calculated on an internal set; and in [15], 94.6% accuracy is calculated on the JSRT set. Every system that has an accuracy between these two observer performances can be regarded as accurate and robust for lung boundary detection on chest X-rays. According to the table, hybrid methods produce
better results than other methods. MISCP [31], ASM-SIFT [25], and ShRAC [29] are the modified versions of ASM [21], and produce better results than traditional ASM [25], [15]. In Table IV, we report three different ASM results. Because of the internal parameters of the ASM method, researchers obtained different segmentation results in [25] and in [15]. In [15], the best parameter set is tuned for this method (ASM-tuned in Table IV), resulting in better results than the ones achieved by other implementations.

F. Segmentation Performance of Lung Model

In atlas-based segmentation, the registered atlas (or combination of registered atlases) can be treated as a final segmentation as in [64], [65]. Therefore, we also measure the segmentation accuracy of the lung model. In the proposed system, the lung model for each patient is formed by following these two steps: 1) selecting the most similar X-rays in the training database, and 2) warping the selected X-rays with a nonrigid registration algorithm. The average sum of the warped masks is used as lung model which is a probability image in which each intensity (between 0–1) indicates the probability of being part of the lung region (see Fig. 6). In our experiment, we first converted the probability image (lung model) to a binary image by thresholding the probabilities with 0.5. Then, we calculated the segmentation accuracy of the binary image. We measured $0.946 \pm 0.016$ average overlap score $\Omega$ for the JSRT dataset which is comparable with
Fig. 10. Particularly difficult lung segmentation cases. (a) The left diaphragm is elevated and there is a large air-distended colon loop below the lung boundary which is incorrectly combined with the lobe into a single region by the automatic algorithm. (b) Detected lung boundary includes the air cavity below left lung. (c)–(e) The algorithm could not detect the lung boundary correctly due to opacity caused by fluid in the lung space. The radiologist “estimated” the ground truth lung boundary (green contour).

### TABLE III

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Execution Time</th>
<th>( \Omega ) of JSRT dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>256 x 256</td>
<td>20-25s</td>
<td>0.954 ± 0.015</td>
</tr>
<tr>
<td>512 x 512</td>
<td>85-90s</td>
<td>0.953 ± 0.016</td>
</tr>
<tr>
<td>1024 x 1024</td>
<td>355-360s</td>
<td>0.937 ± 0.026</td>
</tr>
</tbody>
</table>

### TABLE IV

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>( \Omega )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed System</td>
<td>0.954 ± 0.015</td>
</tr>
<tr>
<td>Hybrid Voting [15]</td>
<td>0.949 ± 0.020</td>
</tr>
<tr>
<td>PC postprocessed [15]</td>
<td>0.945 ± 0.022</td>
</tr>
<tr>
<td>Human Observer [15]</td>
<td>0.946 ± 0.018</td>
</tr>
<tr>
<td>Fusing-Intensity&amp;ShapePriors [28]</td>
<td>0.940 ± 0.053</td>
</tr>
<tr>
<td>Hybrid ASM-PC [15]</td>
<td>0.934 ± 0.037</td>
</tr>
<tr>
<td>Hybrid AAM-PC [15]</td>
<td>0.933 ± 0.026</td>
</tr>
<tr>
<td>MISCP [31]</td>
<td>0.930 ± 0.045</td>
</tr>
<tr>
<td>ASMOF [23]</td>
<td>0.927 ± 0.032</td>
</tr>
<tr>
<td>Fuzzy-Curve [7]</td>
<td>0.927 ± 0.033</td>
</tr>
<tr>
<td>ASM-SIFT [25]</td>
<td>0.920 ± 0.031</td>
</tr>
<tr>
<td>ShRAC [29]</td>
<td>0.907 ± 0.033</td>
</tr>
<tr>
<td>ASM-tuned [15]</td>
<td>0.903 ± 0.057</td>
</tr>
<tr>
<td>ASM [25]</td>
<td>0.870 ± 0.074</td>
</tr>
<tr>
<td>AAM [15]</td>
<td>0.847 ± 0.095</td>
</tr>
<tr>
<td>Mean shape [15]</td>
<td>0.713 ± 0.075</td>
</tr>
</tbody>
</table>

### TABLE V

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>ACD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed System</td>
<td>1.321 ± 0.316</td>
</tr>
<tr>
<td>Fuzzy-Curve [7]</td>
<td>1.730 ± 0.870</td>
</tr>
<tr>
<td>Fusing-Intensity&amp;ShapePriors [28]</td>
<td>2.460 ± 2.060</td>
</tr>
</tbody>
</table>

The selection stage of the X-rays most similar to the patient X-ray, the performance of the registration algorithm and the combination of registered atlases increase the lung model accuracy. However, there is a possibility that the training atlas set could be anatomically unrepresentative of the patient X-ray and could produce a poor lung model. Therefore, we prefer to use the lung model as another term in the objective function of the graph cuts algorithm. The graph cuts refinement step increased the segmentation performance to 0.954 ± 0.015 as we listed in Table II.

We measure the segmentation performance of the system with respect to the number of training masks. Fig. 11(a) shows the average \( \Omega \) overlap score in the JSRT dataset, with the number of masks varying from one to ten. Using a small subset of images is sufficient (at least three) to build an accurate lung model atlas while significantly speeding up the step of nonrigid registration between the training and patient query images [Fig. 11(b)].

### G. Costophrenic Angle and Apical Region of Lung

Robust segmentation of costophrenic angles (CP) and apices are important for the subsequent processes, especially for tuberculosis which often begins at the apices. Costophrenic angle blunting may indicate the presence of pleural effusion [66], [67]. However, segmenting these regions is more challenging compared to the other parts of the lung. The clavicle bone causes intensive intensity variation at the apex, which can mislead the algorithm to label the pixels at this region incorrectly. Energy minimization algorithms generally have problems in segmenting lung and thin regions, because of the smoothness constraint in the objective function [68]–[70]. Therefore, segmenting the lower lobes of the lung is more challenging due to the small CP-angle. In order to observe the performance of the system for these regions, we evaluate the segmentation algorithm at the top 20% and bottom 20% of the lung, as shown in Fig. 12. Figs. 13 and 14 show visual segmentation results at the clavicle and costophrenic angle regions. Green and red contours represent the gold standard and automatic segmentation, respectively. Table VI summarizes the overlap score in these regions. According to the overlap scores, our algorithm robustly segments the clavicle region even in the India dataset which has strong intensity inhomogeneities at the apex due to the presence of the clavicle bone. The algorithm performance
Fig. 12. Segmenting the apex and costophrenic angle regions is more challenging than segmenting the other parts of the lung. These regions correspond approximately to the top 20% and bottom 20% of the lung.

Fig. 13. Segmentation results for apical regions.

Fig. 14. Segmentation results for costophrenic angle regions.

decreases in the costophrenic angle region compared to all other regions (Table VI).

V. CONCLUSION

We have presented a robust lung boundary detection method that is based on a patient-specific lung atlas using fast partial Radon profile similarity selection and SIFT-flow nonrigid registration with refinement using a graph cuts segmentation algorithm. We evaluated the algorithm using three different datasets containing 585 chest radiographs from patients with normal lungs and various pulmonary diseases. On the publicly available JSRT dataset, experimental results showed an accuracy of 95.4% ($\Omega$ overlap measure), compared to the expert segmentation gold standard, which is the highest machine performance reported in the literature. On the other CXR datasets from Montgomery County and India, with more challenging pathologies including abnormal lung boundaries, the same algorithm shows consistently high detection accuracies of 94.1% and 91.7%, respectively. These are the first results reported for automatic lung boundary segmentation that include abnormal lung shapes. The results indicate the robustness and effectiveness of the proposed approach when applied to CXRs collected in different geographical regions. A point to note here is that fluid-filled lungs are radio-opaque, and any radiologist-marked “ground-truth” lung boundary is only an estimate. Therefore, one can assert that the algorithm is being criticized too severely in such cases.

We are making our CXR datasets from Montgomery County, MD, USA and India publicly available to the research community. This includes the X-ray images as well as associated metadata and ground-truth lung segmentations. These two new CXR sets provide for the first time a publicly available research collection of CXRs with manifestations of tuberculosis. The X-rays cover different regions of the world, including high incidence regions. The ground-truth segmentations are provided by a board certified radiologist, and represents the first publicly available collection to include abnormal lung shapes. These sets will enable new research opportunities, and they will improve the existing benchmark evaluations.

APPENDIX

The X-ray sets as well as their corresponding lung boundaries are publicly available for research purposes upon review of request for data. To submit the request, please visit the following webpage: http://archive.nlm.nih.gov/. Under the “Repositories” tab, a link points to a page with more information on our chest images, including contact information.

ACKNOWLEDGMENT

The authors would like to thank Dr. S. Qasba, Medical Director of Montgomery County’s TB Control program, for providing them with the Montgomery dataset and medical advice. The authors would also like to thank Dr. C. S. Durgi, Consulting Radiologist, Mediscan Diagnostic Care, and Associate Professor of Radiology, Mahadevappa Rampure Medical College, Gulbarga, India for the India dataset.