Original Research

Ventilation-Based Segmentation of the Lungs Using Hyperpolarized ^{3}He MRI

Nicholas J. Tustison, DSc,[1]* Brian B. Avants, PhD,[2] Lucia Flors, MD,[1,3] Talissa A. Altes, MD,[1] Eduard E. de Lange, MD,[1] John P. Mugler III, PhD,[1] and James C. Gee, PhD[2]

Purpose: To develop an automated segmentation method to differentiate the ventilated lung volume on ^{3}He magnetic resonance imaging (MRI).

Materials and Methods: Computational processing (CP) for each subject consisted of the following three essential steps: 1) inhomogeneity bias correction, 2) whole lung segmentation, and 3) subdivision of the lung segmentation into regions of similar ventilation. Evaluation consisted of two comparative analyses: i) comparison of the number of defects scored by two human readers in 43 subjects, and ii) simultaneous truth and performance level estimation (STAPLE) in 18 subjects in which the ventilation defects were manually segmented by four human readers.

Results: There was excellent correlation between the number of ventilation defects tabulated by CP and reader #1 (intraclass correlation coefficient [ICC] = 0.86), CP and reader #2 (ICC = 0.85), and between the two readers (ICC = 0.97). The STAPLE results from the second analysis yielded the following sensitivity/specificity numbers: CP (0.898/0.905), radiologist #1 (0.743/0.897), radiologist #2 (0.501/0.985), radiologist #3 (0.898/0.848), and the first author (0.600/0.984).

Conclusion: We developed and evaluated an automated method for quantifying the ventilated lung volume on ^{3}He MRI. The findings strongly indicate that our proposed algorithmic processing may be a reliable, automatic method for quantitating ventilation defects.

Key Words: hyperpolarized ^{3}He; segmentation; ventilation defects
J. Magn. Reson. Imaging 2011;34:831–841.
© 2011 Wiley-Liss, Inc.

[1]Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, Virginia, USA.

[2]Penn Image Computing and Science Laboratory, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

[3]Department of Medicine, Universidad Autónoma de Barcelona, Barcelona, Spain.

*Address reprint requests to: N.J.T., 480 Ray C Hunt Dr., PO Box 801339, Charlottesville, VA 22932. E-mail: ntustison@virginia.edu

DEVELOPMENTS IN MAGNETIC RESONANCE IMAGING (MRI) research utilizing noble gases, such as ^{3}He and ^{129}Xe, have demonstrated the capability of visualizing alveolar and bronchial air spaces (1). Currently, hyperpolarized ^{3}He MRI is a low-risk (2) investigatory technique that provides high spatial and temporal resolution images of the air spaces of the lungs and has been used to investigate a variety of lung diseases. Automated or semiautomated approaches for classifying areas of varying degrees of ventilation are of potential benefit for facilitating such investigation.

Although various approaches have been previously proposed in the literature, several potential confounds continue to complicate the task. These confounds include the presence of a low-frequency intensity bias due to the inhomogeneity field, ventilation defects on the boundary of the lung complicating whole lung segmentation from the background, and the intensity signature of the vasculature, which appear in the same intensity range as ventilation defects. Further complicating ^{3}He MRI quantitative assessment is that signal intensity is not solely dependent on the density of ^{3}He atoms in each pixel and, therefore, does not directly reflect regional ventilation. This is due to the fact that the coil transmit and receive sensitivity and the regional partial pressure of oxygen within the lung contribute to the measured signal intensity. For this reason, the ^{3}He spin-density images provide information about the homogeneity of ventilation within the lung but do not provide a quantitative measure of absolute regional ventilation.

We present an automated algorithmic pipeline for ventilation-based partitioning of the lungs in hyperpolarized ^{3}He MRI which attempts to address the aforementioned complexities. The workflow of the major components of this pipeline is illustrated in Fig. 1. Offline processing includes building the unbiased template and statistical description of the lung shape from sets of normal data. These two descriptions of the data are then used for individual subject processing. We describe the major components of this pipeline and compare its effectiveness in identifying ventilation defects with human readers. Our software pipeline is made available to the research community as open source software through our Advanced Normalization Tools (ANTs) package (http://www.picsl.upenn.edu/ANTS), a suite
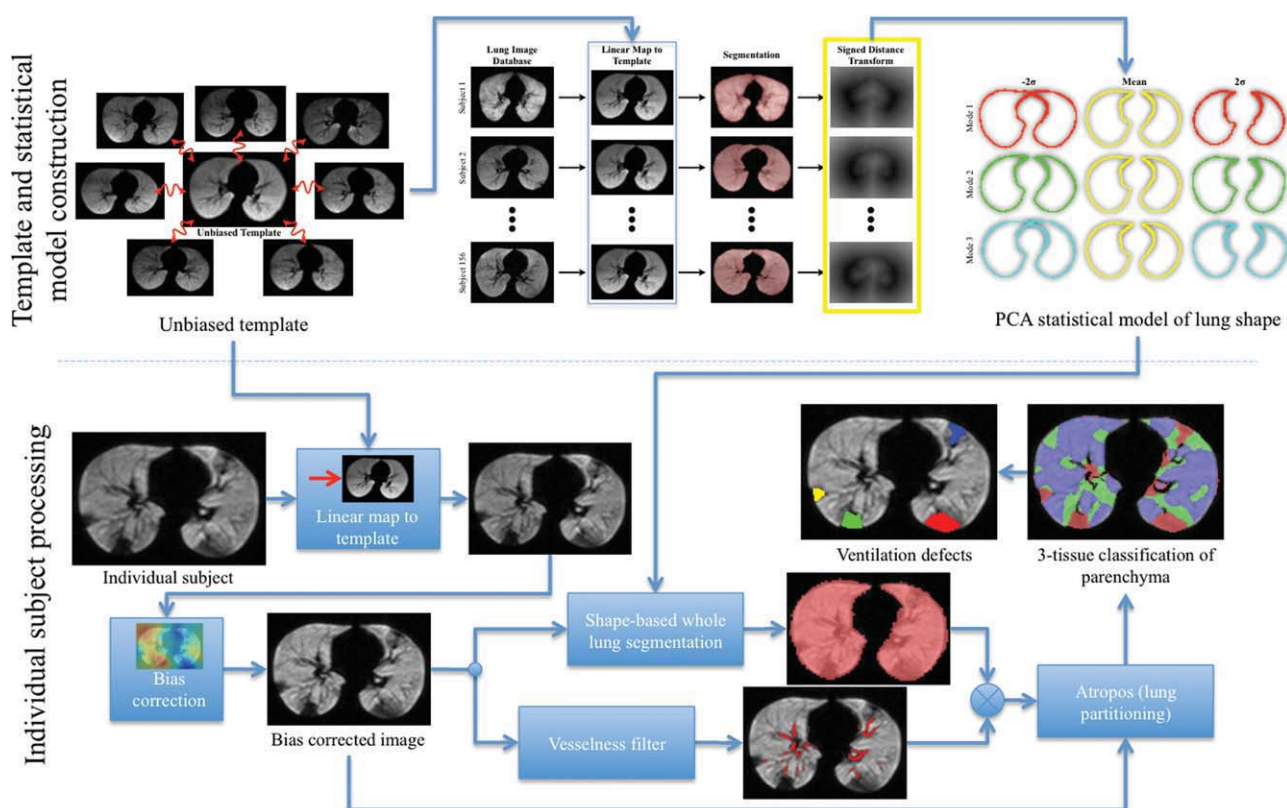
Figure 1. Diagrammatic workflow for ventilation-based segmentation. The top row illustrates the various processes used to construct the template and statistical shape model in preparation for individual subject processing. The final output consists of a set of posterior probabilities at each voxel location indicating membership in a particular ventilation class. From these posterior probability images, hard classifications can be inferred including the location of ventilation defects. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

of software tools for registration and segmentation of medical images based on the Insight Toolkit of the United States National Library of Medicine of the National Institutes of Health. The remaining components not contained in the ANTs package are available directly through the Insight Toolkit (http://www.itk.org).

MATERIALS AND METHODS

Image Acquisition

Imaging with hyperpolarized ^{3}He was performed under an Institutional Review Board (IRB)-approved protocol with written informed consent obtained from each subject. In addition, all imaging was performed under an Food and Drug Administration (FDA)-approved physician's Investigational New Drug application (IND 57866) for hyperpolarized ^{3}He. MRI data were acquired on a 1.5 T whole-body MRI scanner (Siemens Sonata, Siemens Medical Solutions, Malvern, PA) with broadband capabilities and a flexible ^{3}He chest radiofrequency coil (RF; IGC Medical Advances, Milwaukee, WI; or Clinical MR Solutions, Brookfield, WI). During a 10–20-second breath-hold following the inhalation of $\approx$300 mL of hyperpolarized ^{3}He mixed with $\approx$700 mL of nitrogen, a set of 19–28 contiguous axial sections were collected. Parameters of the fast low angle shot sequence for ^{3}He MRI were as follows: repetition time msec / echo time msec, 7/3; flip angle,

10$°$; matrix, 80 $\times$ 128; field of view, 26 $\times$ 42 cm; section thickness, 10 mm; and intersection gap, none. The data were deidentified prior to analysis.

Whole Lung Segmentation

Shape/Intensity Unbiased Template Construction

Several steps in our processing pipeline require the use of a template for normalization purposes. This includes steps for both offline preprocessing and individual subject processing. For example, the removal of global shape differences in creating our statistical shape model of the normal lung requires all images in the database to be registered to a normalized space. In addition, during individual subject processing each image needs to be warped to such a space so that the mediastinum can be identified for further processing.

ANTs contains a suite of normalization tools including software for unbiased template construction using a symmetric diffeomorphic registration algorithm (3). Although other strategies can be used for identifying templates, such as random subject selection, unbiased template construction provides a shape/intensity normalization space that resides at approximately the mean of the population of interest (4). In other words, the total deformation (ie, the shape distance) between the template and the subject population is minimized. The results of our template construction on 3D data can be seen in Fig. 2, where we
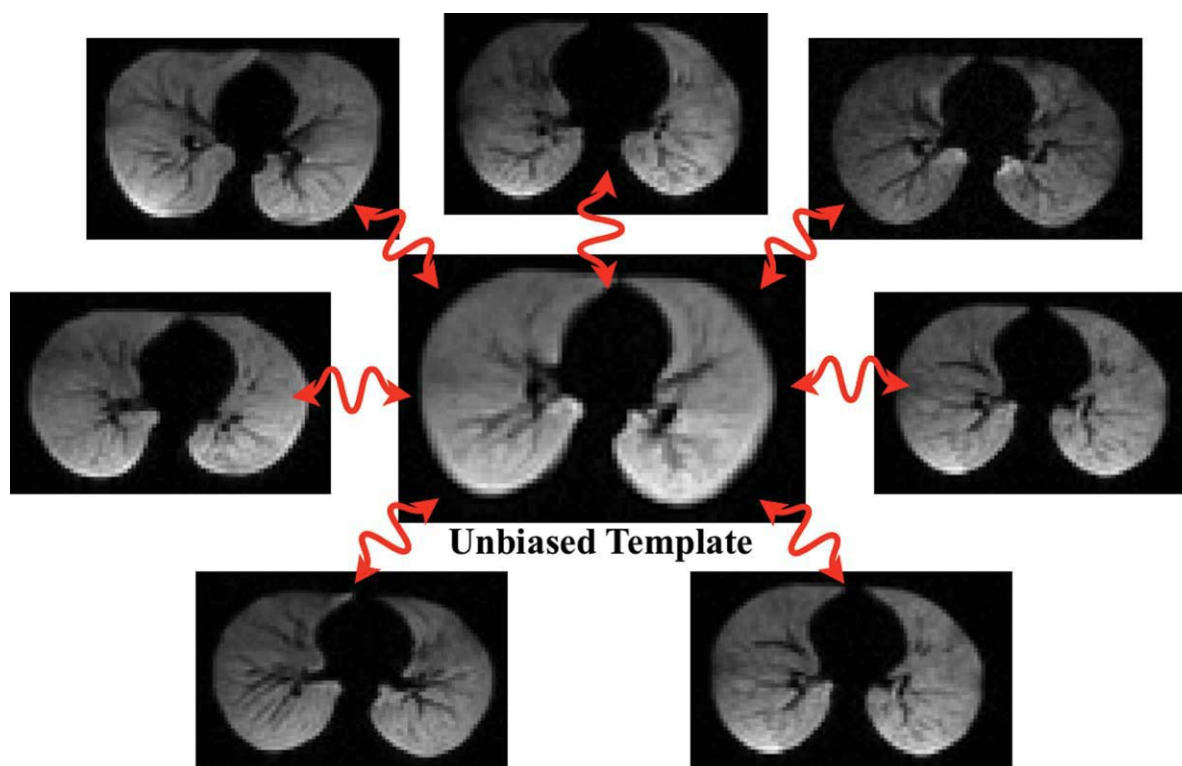
Figure 2. Unbiased template creation from a subset of seven subjects. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

used seven randomly selected subjects to create the unbiased template seen in the center of the figure. Note that this is a purely synthetic image, which can be understood intuitively as a single image that best represents the seven input images. Experimentally, it was seen that seven subjects provided a satisfactory compromise between quality of results and required computational time.

Statistical Shape Description of the Lung Boundary

In normal subjects lung boundary delineation is relatively straightforward, as one can perform segmentation by sampling the image in an area of known background to determine statistical properties of this region (5). This seeds an iterative region-growing strategy that can be used to segment the background from the lung. However, ventilation defects on the periphery obscure the true boundary of the lung preventing application of the statistical region-growing segmentation algorithms to all subjects.

We use control subject lung segmentations and principal components analysis (PCA) to build a statistical model of lung boundaries that are free from ventilation defects. Because the PCA model stores the normal variation in anatomical lung boundaries, we can then later apply this model to robustly segment lungs with defects (6). We delay discussion of the segmentation of an individual subject to the next section but first describe the construction of the PCA statistical model, which is laid out in Fig. 3. The PCA model used in the evaluation study is constructed from a ³He image database of 156 normal subjects. Each image of the database is transformed to the template using an affine

transformation. This minimizes the presence of any global shape differences in the statistical model. In this transformed space, each image is segmented using our region-growing segmentation algorithm. The resulting binary image is used to create its corresponding signed distance transform (SDT) (7) shown in the last column where the value at each voxel in the SDT is the Euclidean distance to the nearest lung boundary voxel.

PCA image decomposition (8) was performed on the 156 SDT images to determine the mean and modes of variation. It was found that 52 modes account for 98% of the variation of the SDT images, which is what was retained to account for the description of lung boundary shape. The first three modes of the model are shown in Fig. 4. Since the PCA statistical model describes the shape of the lung in a quantitative way, it can be incorporated into the well-established level set framework for an application-specific, top-down strategy for segmentation (9). Segmentation of each set of lungs begins with the mean lung shape (shown in the middle column of Fig. 4). Iterations of the level set method, which refine the segmentation, involve solving a differential equation that incorporates both the PCA statistical model and the given image intensity information. Thus, the PCA statistical model is used to constrain the level set to produce a segmentation result that exists within the variation allowed by the shape statistics. The utility of this approach can be seen in the example given in Fig. 5.

Individual Subject Processing

Retrospective Bias Correction

Low-frequency intensity artifacts are present in hyperpolarized noble gas MR primarily due to flip
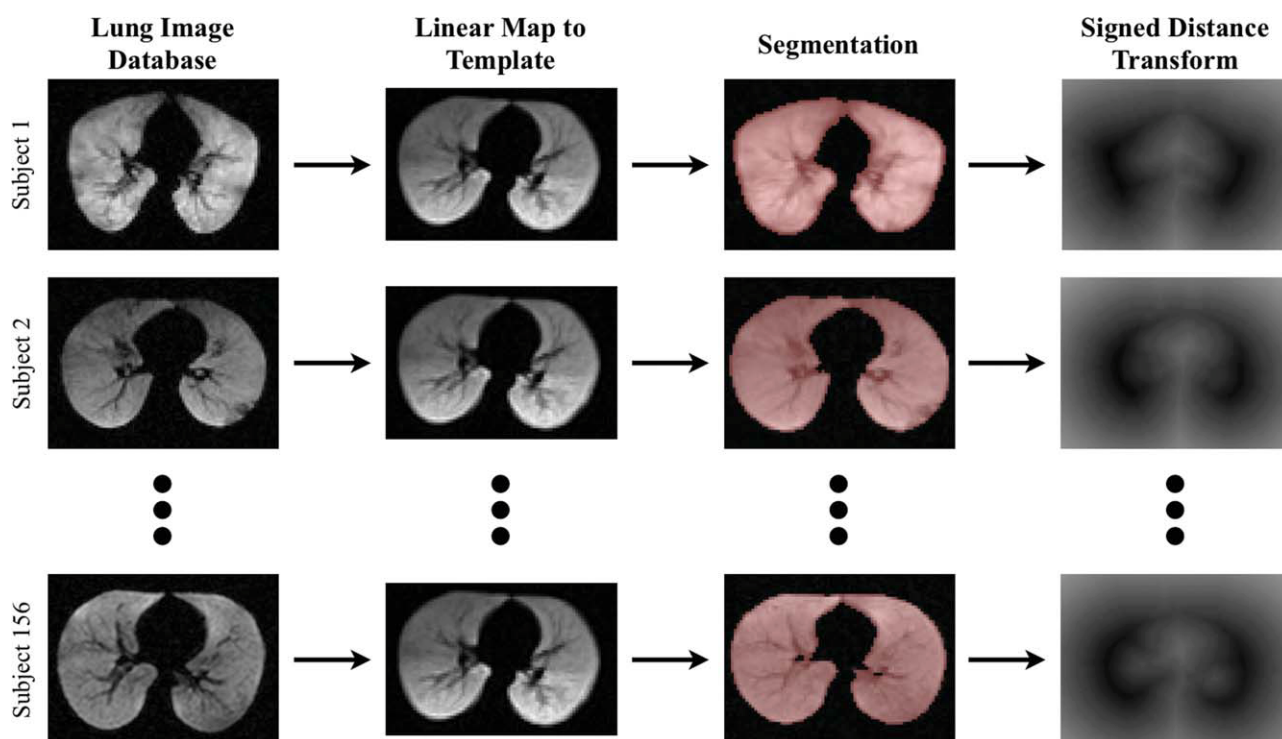
Figure 3. Construction of a statistical shape description of the lung boundary consists of the following steps: 1) Each image within an image database of 156 normal subjects (first column) is warped to the shape/intensity template (second column) using an affine transformation; 2) a connected components algorithm is used to automatically segment the lung boundaries from the background; and 3) each binary segmentation is converted to its signed distance transform (last column). A PCA decomposition of the latter set of images is used to constrain the level set algorithm for whole lung segmentation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

angle variations caused by the inhomogeneity introduced by the RF coil. Other possible causes of nonuniform bias include the anatomical diffusion gradient (10) and posture-related dependencies (11). These artifacts are potential confounders in image segmentation and correction of such artifacts is an important preprocessing step. A novel algorithm was proposed (12) for retrospective inhomogeneity correction in MRI. This algorithm is based on the observation that the effect of the artifactual bias field is to smooth out the intensity histogram. Thus, tissue (or ventilated-based) peaks which correspond to a particular intensity signature (eg, normally ventilated regions) are eroded as a result of the bias field. To estimate the bias field and subsequently correct the resulting artifact, a solution is obtained by iterating between sharpening the intensity histogram and smoothing the current bias field estimation by an approximating/smoothing B-spline scalar field. Sample effects of this algorithm as seen on an axial slice of a ^{3}He image are illustrated in Fig. 6.

"Vesselness" Filtering for Identification of the Pulmonary Vasculature

Since the vasculature appears as dark regions in the ^{3}He images, identification of the vasculature is necessary for restricting ventilation analysis to other parenchymal regions. The application of the vessel-enhancing work of Frangi et al (13) and its generalization (14) is used to segment out the vasculature from the rest of the parenchyma. First, the Hessian matrix is calculated from the image following convolution with a Gaussian kernel at a feature-based scale. The eigenvalues of the Hessian matrix are then calculated from which a measure of "vesselness" is quantified. This image filtering operation is performed over several feature-based scales where the maximal response is maintained. Two examples are shown in Fig. 7. By removing the vasculature regions, a better estimate of the nonventilated regions can be obtained.

Ventilation-Based Segmentation

Once the masked region (parenchyma minus the vasculature) is defined, we can then apply the main component of our contribution, which is the ventilation-based segmentation algorithm known as Atropos.* Atropos encodes a family of Bayesian segmentation techniques that may be configured in an application-specific manner. The theory underlying Atropos dates back 20+ years and is representative of some of the most innovative work in the field. Capabilities include both the conventional Gaussian mixture modeling and nonparametric approaches. A priori strategies including template (15) and/or Markov random field (16) priors can be used. Initialization of the algorithm (including both initial classification and class

---

*Atropos is one of the three Fates from Greek mythology characterized by her dreaded shears used to decide the destiny of each mortal.
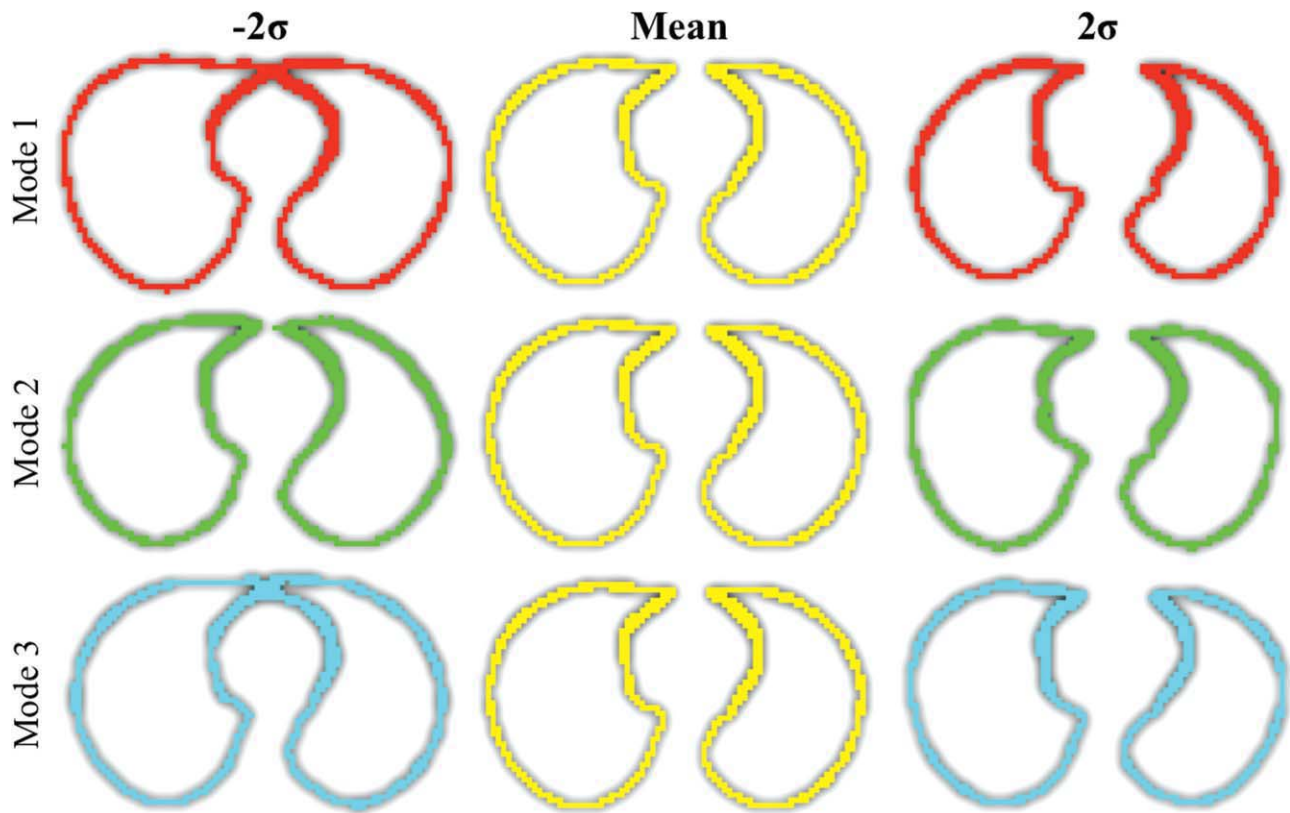
Figure 4. Each row corresponds to one of the first three total modes (out of a total of 52) describing the PCA statistical decomposition of the 156 SDT images from the normal lung database. We rendered the zeroth level set of the mean and ±2 standard deviations from the mean for each of the first three modes. This PCA statistical model constrains the whole lung segmentation algorithm to produce plausible lung shapes. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

statistics calculation) is performed using template-based priors or intensity-only considerations. All of these components can be situated within the maximum a posteriori Bayesian statistical framework. Further details regarding the theory, implementation, and application to brain segmentation can be found (17).

Similar to the three-tissue (white matter, gray matter, and cerebrospinal fluid) neuroanatomical segmentation scenario, ventilation-based segmentation of ^{3}He lung images requires a partitioning of the lung parenchyma into subregions of similar intensities. Given the set of voxels in the lung segmentation mask minus the vasculature, the solution assigns a label

(or class) to each voxel, denoted by i, with intensity y_i, to one of a set of N labels where each label corresponds to a distinct class based on ventilation. In this study we use Atropos to define classes for ventilated and unventilated regions of the lung. Specifically, we found optimal results in using two Gaussian classes to represent the poorly ventilated regions and two classes to represent the normally ventilated regions.

As mentioned previously, the intensity values only give a relative measure of ventilation. This motivates the use of approximating the intensity histogram as a sum of N Gaussians where the optimization process is used to calculate the optimal parameters of each
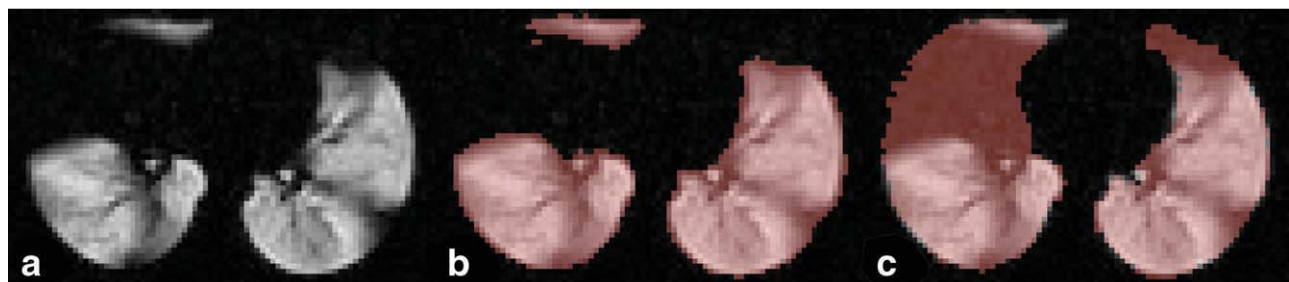

Figure 5. Whole lung segmentation. **a**: The original bias-corrected image of an individual subject exhibits a large ventilation defect. **b**: A segmentation approach that is completely data-driven misses the portion of the lung obscured by a large ventilation defect. **c**: Using the PCA statistical model incorporating a prior constraint on the lung shape, a more plausible segmentation results is obtained. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
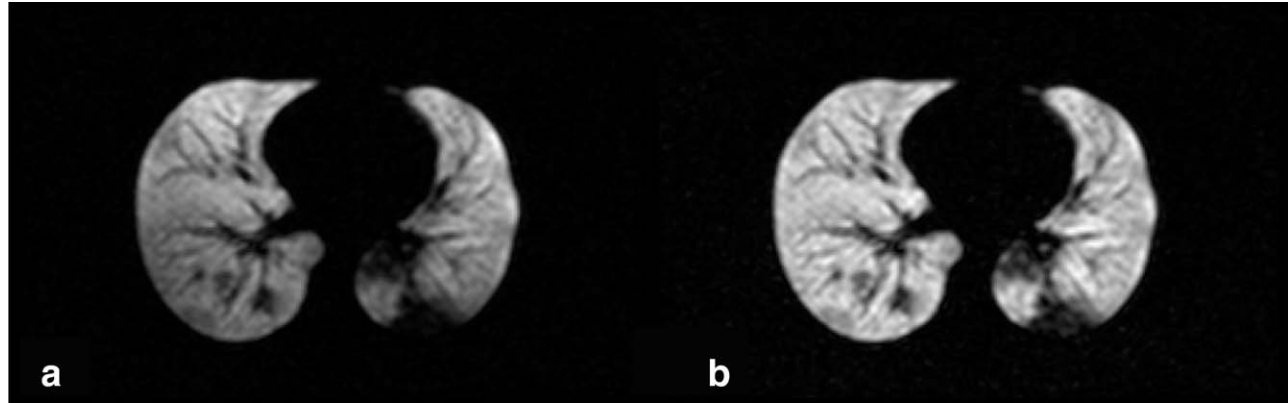
Figure 6. Sample effects of bias field correction on hyperpolarized ³He. **a**: Uncorrected image showing severe field inhomogeneity. **b**: Corrected image in which the bias effects have been significantly removed.

Gaussian (ie, the mean and variance). Each Gaussian (or sum of Gaussians) represents a distinct class characterized by its level of ventilation such that the likelihood of a single voxel of intensity y_i belonging to the nth class, c_n, parameterized by μ_n and σ_n, is given by:

$$P(y_i|x_i = c_n) = \frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(\frac{-(y_i - \mu_n)^2}{2\sigma_n^2}\right) \qquad [1]$$

where x_i is the label configuration associated with voxel i. The parameters μ_n and σ_n are updated at each iteration using the Expectation-Maximization algorithm explained in further detail below. A major drawback associated with using only finite mixture modeling for segmentation is that spatial contextual information is completely ignored. This is clearly seen
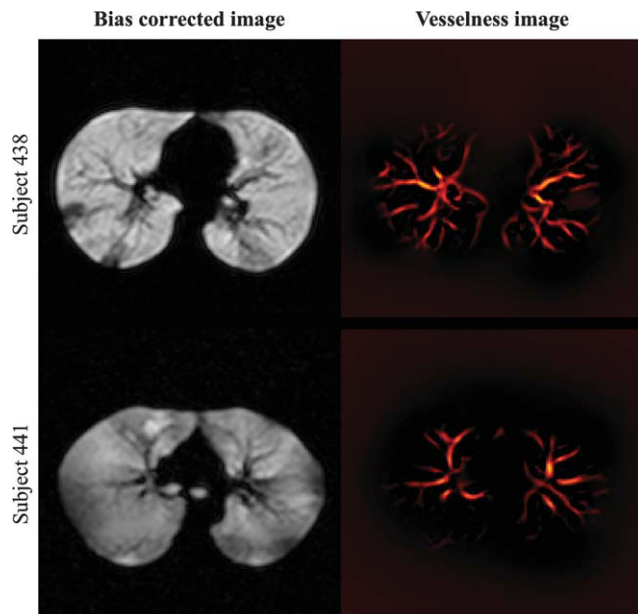

Figure 7. Sample "vesselness" images for two subjects. The left column consists of the bias corrected hyperpolarized ³He image of the two subjects. The resulting vesselness images of both subjects are located in the right column illustrating enhancement of vessel-like objects. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

from Eq. [1], where the neighbors of voxel i are not even considered in the calculation of the likelihood probability. Given this lack of a priori knowledge concerning the location of different ventilation regions, prior-based constraints using templates is not applicable as used previously (15). However, Markov random field (MRF) theory provides a general framework for modeling spatially smooth, context-dependent image analysis problems (18). Intuitively, if neighboring voxels belong to the same ventilation class, the MRF prior probability, $P(\mathbf{x})$, will be higher, which enforces a spatial clustering of voxels with similar intensities.

Combining the likelihood term, ie, Eq. [1], and the MRF prior term the solution of the segmentation problem estimates the true labeling of the set of voxels or, in other words, the correct assignment of each voxel to a single class based on the intensity. We denote this optimal configuration as $\mathbf{x}^*$, which, according to the maximum a posteriori criterion, is calculated from:

$$\mathbf{x}^* = \arg\max_{x \in X}\{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})\} \qquad [2]$$

where $\mathbf{x}$ is a particular labeling from the set of all possible labelings and $\mathbf{y}$ is the set of image intensities. Various iterative optimization approaches are possible for solving Eq. [2] for which we use a greedy approach known as iterated conditional modes (ICM) (19). ICM guarantees convergence to a local minimum in a few iterations.

The Expectation-Maximization algorithm, when applied to segmentation (16), updates labels and then parameters at each iteration. Thus, the Atropos Expectation step computes the posterior probabilities (given fixed parameters) and selects the optimal labels. We then estimate the optimal parameters, for each Gaussian, given the fixed labels (the Maximization step) from:

$$\mu_n^{(t+1)} = \frac{\sum_{i \in I} P^{(t)}(x_i = c_n|y_i)y_t}{\sum_{i \in I} P^{(t)}(x_i = c_n|y_i)} \qquad [3]$$

$$\sigma_n^{(t+1)} = \left(\frac{\sum_{i \in I} P^{(t)}(x_i = c_n|y_i)\left(y_i - \mu_n^{(t)}\right)^2}{\sum_{i \in I} P^{(t)}(x_i = c_n|y_i)}\right)^{\frac{1}{2}}. \qquad [4]$$

Table 1
Sensitivity and Specificity Numbers for Each of the Four Readers, Atropos, K-means, and Otsu Thresholding for the STAPLE Evaluation

Rater	Sensitivity	Specificity
Atropos	0.898	0.905
Radiologist 1	0.743	0.897
Radiologist 2	0.501	0.985
Radiologist 3	0.898	0.848
First author	0.600	0.984
K-means	0.973	0.735
Otsu	0.973	0.734

To initialize the algorithm, one can generate an initial partitioning of the intensity histogram using Otsu thresholding (20) or K-means clustering (21).

The final output of the algorithmic pipeline consists of N posterior probability images where the value at voxel i in the nth posterior probability image can be thought of as representing the relative membership weighting of voxel in class c_n. This soft classification can be translated into a hard segmentation by determining the maximum class membership value at each voxel and assigning that class to that voxel. From this hard classification, we can automatically identify ventilation defect regions from those regions belonging to the regions corresponding to poor ventilation.

Evaluation of the Methodology

To evaluate our methods two retrospective analyses were performed. The first evaluation employed the entire pipeline to score a subset of the data described originally (22). In that study, two independent human readers who were blinded to all clinical information counted the number of ventilation defects on previously obtained ³He ventilation MR scans. The subset analyzed by the pipeline described previously consisted of 43 subjects (8 normal and 35 diagnosed asthmatic). Based on the readers' assessment in the original study, we modified the output of our algorithm to calculate the number of distinct ventilation defects calculated on each slice. Thus, whereas our pipeline is capable of processing 3D image volumes, it can also be run on each subject on a slice-by-slice basis, as performed for this particular evaluation study.

A second analysis was used to evaluate the performance of the ventilation-based segmentation algorithm in isolation from the rest of the pipeline. Using both ³He and ¹H image data, which were acquired simultaneously, a trained radiologist (denoted "Radiologist 3" in Table 1 and Fig. 9) segmented the whole lungs for each of 18 subjects (4 normals and 14 diagnosed subjects with cystic fibrosis) using the ITK-SNAP image annotation tool (23). By simultaneously referring to both coregistered images, in which the mouse cursor was linked between the two image sets, manual segmentation was facilitated over using either modality alone. This and two other radiologists, as well as the first author, manually segmented the ventilation defects within the masked lung regions for all 18 subjects. Atropos was also used to segment the ventilation defects using four classes where the lower two classes represented the ventilation defect regions and the upper two classes represented the normal ventilation regions.

Since there is no ground truth for these data, a consensus labeling using the simultaneous truth and performance level estimation (STAPLE) algorithm (24) was used as a probabilistic estimate of the ground truth segmentation for each of the 18 subjects. The identification of the defect and normally ventilated regions produced by each of the five readers (both human and Atropos) for each of the 18 subjects resulted in 18 * 5 = 90 total segmentations. Fusing several segmentations of the same object by different raters, STAPLE iteratively estimates the performance level of each rater while simultaneously producing a probabilistic estimate of the true segmentation. In this fashion, STAPLE was used to produce 18 such probabilistic ground truth estimates, one for each subject. Since a byproduct of STAPLE is a performance estimation of each rater, performance comparison for each rater for each subject can be analyzed. Tabulating the sensitivity and specificity values of each rater over all 18 subjects on a voxel-by-voxel basis summarizes this comparison. Although not included with the STAPLE consensus estimation, we calculated the K-means and Otsu voxel classifications (using four classes) for each subject and also used the STAPLE probabilistic estimate of the ground truth to calculate the sensitivity and specificity values for these methods.

RESULTS
Automated Pipeline Comparison with Human Readers

We compared the scores produced by the two independent human readers with the results of our ventilation-based segmentation algorithmic pipeline. Bland–Altman plots illustrating interrater agreement between the algorithmic pipeline and each of the two readers is given in Fig. 8. The intraclass correlation coefficient (ICC), a measurement of reliability between different raters (25), was also calculated. There was excellent correlation between the two readers (ICC = 0.97). There was also strong correlation between Reader 1 and our algorithmic pipeline (ICC = 0.86). A strong correlation also existed between the automated and Reader 2 results (ICC = 0.85).

Given the availability of the readers' scores from the previous study (22), this first analysis was meant as a benchmark comparison for our segmentation algorithm. Even though the results are quite promising, the evaluation framework is slightly inadequate in the sense that simply counting the number of ventilation defects is not potentially as informative as other properties such as total defect volume, location, and/or shape, which is provided by our segmentation approach. These deficiencies motivated the more comprehensive second study.

STAPLE Evaluation of Atropos

The confusion ratio plots for all five raters and the K-means and Otsu segmentation strategies over all 18
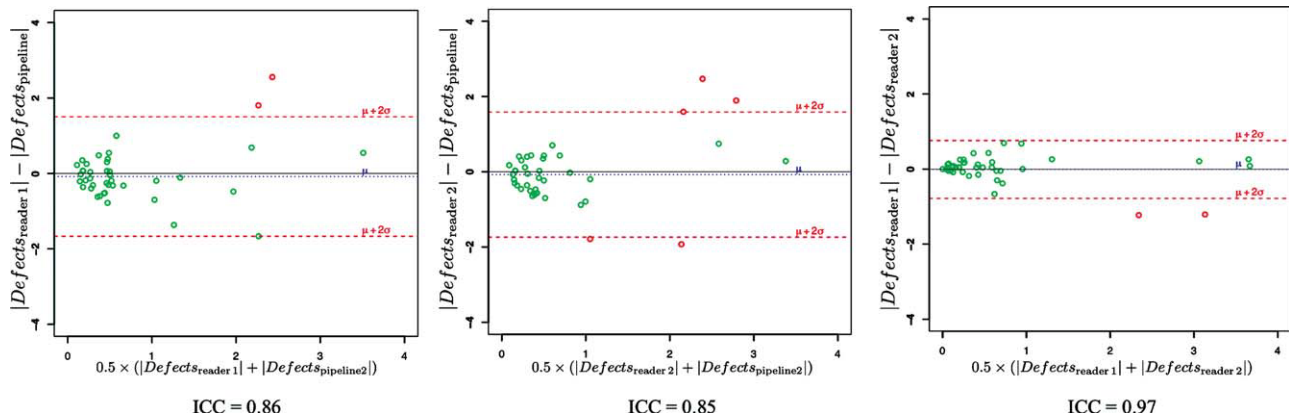
Figure 8. Bland–Altman plots for the different pairings between both readers and our segmentation pipeline. The plot on the left relates the measurements of Reader 1 vs. Reader 2 (ICC = 0.97) where the y-axis is given by the difference between the two sets of measurements and the x-axis is the average of the two. Similarly, the plot in the center relates the measurements of Reader 1 vs. our algorithmic pipeline (ICC = 0.86). Shown on the right is the plot relating Reader 2 vs. pipeline (ICC = 0.85). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

subjects are given in Fig. 9 and illustrate the true positives/negatives (upper left and lower right corners) and the false positives/negatives (upper right and lower left corners). Note that a true positive for our analysis is defined as a rater-labeled voxel specified as a ventilation defect which is also identified as a ven-

tilation defect by the STAPLE probabilistic estimate of the true labeling. One thing to note is that each rater was given both the lung mask and the ^{3}He image which could be loaded into ITK-SNAP where the ventilation defects could be manually labeled. Thus, if a rater had not labeled anything as a ventilation defect,
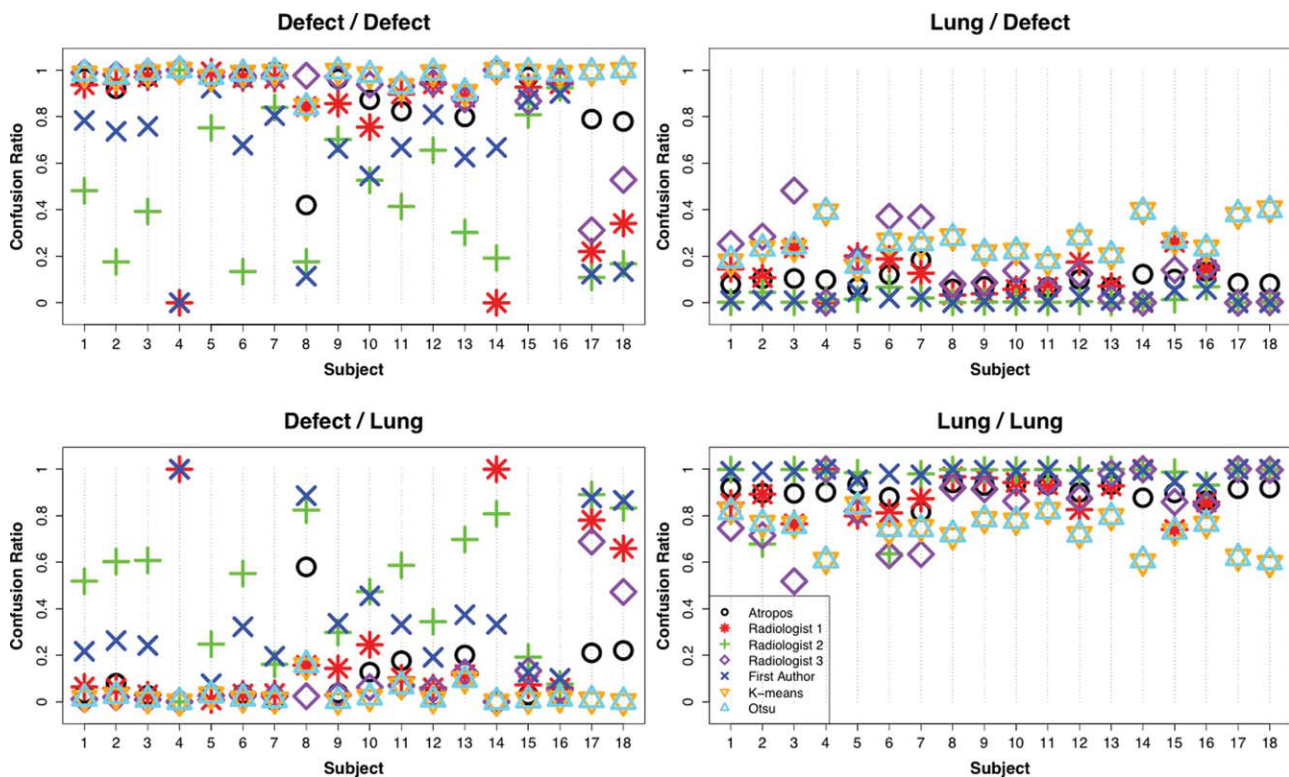

Figure 9. Confusion ratios for all 18 subjects as produced by the STAPLE analysis. The first four subjects are normals, whereas subjects 5 through 18 correspond to the subjects diagnosed with cystic fibrosis. The upper left plot gives the portion of true-positive labelings, ie, the ratio of voxels labeled as defect voxels that are classified as defect by the STAPLE consensus labeling. Similarly, the lower right plot gives the ratio of true-negative labelings, ie, the ratio of voxels labeled as lung that are classified as lung voxels by the STAPLE consensus labeling. The upper right plot depicts the false-positive ratios or the ratios of voxels labeled as defect by the raters that are actually determined to be lung voxels. The lower left plot depicts the false-negative ratios or the ratios of voxels labeled as lung by the raters that are actually determined to be defect voxels. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the corresponding plot in the lower right (true negatives) would have performance levels at 1.0 for that rater for all 18 subjects, ie, their specificity would have been perfect but their sensitivity scores would be poor.

Using the values reported in the confusion ratio plots given in Fig. 9, a summary of the performance levels for all raters are reported in Table 1 in terms of each rater's sensitivity and specificity. Our Atropos algorithm was the top performer with both sensitivity and specificity values at ≈0.9. Radiologist 3 performed almost as well with slightly lower specificity. Both the first author and Radiologist 2 had much lower sensitivity values. The K-means and Otsu approaches to ventilation segmentation performed similarly, which is not unexpected given that both algorithms seek an optimal classification based strictly on the intensity histogram without consideration of the spatial distribution of labels. These latter approaches both had relatively high sensitivity values but relatively low specificity values.

One of the issues to consider is that, whereas Atropos took <10 minutes to segment all 18 cases serially on an 8-core Mac Pro, the times taken by the human raters to segment all 18 subjects took on the order of a few hours.

DISCUSSION

Significant research has been devoted to the identification of regions of poor ventilation, ie, ventilation defects, in various image modalities. For example, previous studies used positron emission tomography (PET) to locate ventilation defects which were subsequently mapped to a 3D anatomical lung model for identification of the constricted airway(s) associated with specific ventilation defects (26). Other studies used Xe-enhanced computed tomography (CT) to study ventilation in various animal models (27,28) and human subjects (29). However, both modalities suffer from practical considerations of subject exposure to ionizing radiation.

In contrast, MRI does not rely on ionizing radiation but does suffer from technical issues in the pulmonary region such as short T1 relaxation values and significant susceptibility. However, the use of hyperpolarized [3]He as an inhaled contrast agent allows for compelling visualization of the ventilated airspaces and, concomitantly, identification of ventilation defects (30). This includes volumetric and numerical ventilation differentials established in asthmatics with increasing disease severity or with provocation such as exercise or methacholine (5,31). Other lung pathologies characterized by the presence of ventilation defects include cystic fibrosis (32,33) and chronic obstructive pulmonary disease (10,34–38).

Segmentation of these ventilation defect regions has been almost exclusively limited to manual protocols such as expert identification (38,39) or manual thresholding of the intensity histogram (40). Whereas expert segmentation suffers from potential intra- and interobserver issues (eg, repeatability), intensity-based approaches relying solely on the intensity histogram disregard the spatial distribution of the intensities within an image. Automated approaches include automatic thresholding of the intensity histogram (5) using Otsu's method, which partitions the image region of interest into a specified number of classes which exhibit minimal intraclass variance. A recent study investigated the use of imaging-based quantities to characterize the clinical diagnosis in a sample of asthmatics and normals (5). Over 1600 image features were calculated from the [3]He images such as the mean intensity and various texture-based measures. A mutual-information-based feature subset selection algorithm was used to characterize their ability to differentiate asthmatics and normals using preestablished clinical diagnoses as the gold standard. It was found that the imaging features performed very well in comparison with spirometry values (although the informational content of the latter was relatively orthogonal to the former). Motivating the work presented in this article, it was also found that quantifying the regions of poor ventilation was the top-performing feature in differentiating between clinical diagnoses of asthmatic versus normal. However, quantifying ventilation defects started with whole lung segmentation acquired by seeding a region-growing algorithm with a neighborhood in the background of the image and growing this seed point to the rest of the connected image background. Further ventilation subdivision used Otsu thresholding in the lung mask for regional ventilation-based subdivision. Such an approach is prone to failure in the presence of defects which share the same intensity signature as the background.

We presented an algorithm for ventilation-based segmentation of hyperpolarized [3]He images including the automatic delineation of ventilation defects. Although manual identification results for our evaluation study demonstrated superb intrarater reliability (as evidenced by the ICC measurement), it should be noted that the actual morphology of the ventilation defects was not specified by the two readers (unlike the capabilities of our contribution), which most likely would have demonstrated increased separation in agreement. More sophisticated histogram-based approaches have included optimal thresholding. However, any histogram-based method with exclusive reliance on intensity values neglects spatial contextual information, which is usually necessary for adequate segmentation results. We also suspect that future developments and increased study might reveal that the actual posterior probability images could perhaps be more sensitive indicators of clinical conditions.

Additionally, to our knowledge previous methods have not addressed such potential confounders as the inhomogeneity field, the intensity signature similarity between the vasculature and the hypointense regions, and whole lung segmentation difficulties caused by ventilation defects located on the boundary. A straightforward approach to whole lung [3]He segmentation is mapping the [3]He image to the ^{1}H image, which is easier to segment. This transformation typically requires a deformable component due to possible differences in lung inflation levels between the acquisitions. With COPD or other such diseases in which large ventilation defects are present, this deformable

component complicates registration, as deformable transformations permit such erroneous matching as lung boundary to inner defect boundary. In contrast, the registration used in the pipeline from an individual subject to the normal template is a linear mapping solely used to get the image in the space of the template where the statistical shape model drives the segmentation. This linear mapping is more constrained and thus more robust to the presence of ventilation defects. Given the functional aspect of hyperpolarized ^{3}He imaging, there are scenarios for which any registration and/or segmentation will be difficult and thus require other approaches such as simultaneous acquisition of ^{1}He/^{3}He images.

Of substantial significance is the relative performance between the computational algorithm and the three human readers, particularly the outperformance of the former with respect to the latter. Based on the criteria of the STAPLE solution, the optimal reader (computer or human) is going to provide the most consensus among the weighted assessments of all other readers. There are several contributing factors, all of which could contribute to explaining this outcome. Given that the manual delineations were performed by the human readers at different times under distinct conditions, one would expect disagreement between readers not only in difficult voxel-by-voxel cases but also due to consistency issues. In contrast, the computer algorithm is going to be consistent across all subjects, using a strictly quantitative decision-making formulation, and is thus not susceptible to those human performance detractors such as fatigue, distraction, task-related tedium, and performance-altering ambient factors (eg, lighting conditions). This is particularly relevant especially considering that the human raters performed the manual delineations in their spare time due to the lengthy time requirements of the task. It is also important to recognize that the computer algorithm was tuned under the guidance of two of the other authors (T.A. and J.M.) who did not participate in the reader study. Of additional practical significance is that we have offered the major components of our pipeline as open source, freely available online. The source code can be downloaded from our website (http://www.picsl.upenn.edu/ANTs). This allows interested persons to obtain the source code (along with the Insight Toolkit at http://www.itk.org), run it on their data, and even tailor the source code to their needs.

Considering the possible uses of our contribution, there are several avenues for exploration. Although our evaluation study was limited to normals and patients diagnosed with asthma and cystic fibrosis, automatic delineation of ventilation defects could perhaps be of use to the study of other lung pathologies such as COPD. Based on our experience, the parameters described in this contribution should apply equally well to these other diseases as they did for our evaluations.

REFERENCES

1. de Lange EE, Mugler III JP, Brookeman JR, et al. Lung air spaces: MR imaging evaluation with hyperpolarized 3He gas. Radiology 2011;210:851–857.

2. Altes TA, Gersbach JC, Mata JF, Mugler III JP, Brookeman JR, de Lange EE. Evaluation of the safety of hyperpolarized helium-3 gas as an inhaled contrast agent for MRI. In: Proc 15th Annual Meeting ISMRM, Berlin; 2007. Abstract number: 1305.

3. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 2011;54:2033–2044.

4. Avants BB, Yushkevich P, Pluta J, et al. The optimal template effect in hippocampus studies of diseased populations. Neuroimage 2010;49:2457–2466.

5. Tustison NJ, Altes TA, Song G, de Lange EE, Mugler III JP, Gee JC. Feature analysis of hyperpolarized helium-3 pulmonary MRI: a study of asthmatics versus nonasthmatics. Magn Reson Med 2010;63:1448–1455.

6. Leventon M, Grimson W, Faugeras O. Statistical shape influence in geodesic active contours. In: Proc IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2000; p 316–323.

7. Maurer CR, Qi R, Raghavan V. A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. IEEE Trans Pattern Anal Machine Intell 2003;25:265–270.

8. de Lange EE, Altes TA, Patrie JT, et al. The variability of regional airflow obstruction within the lungs of patients with asthma: assessment with hyperpolarized helium-3 magnetic resonance imaging. J Allergy Clin Immunol 2007;119:1072–1078.

9. Osher S, Sethian JA. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. J Comput Phys 1988;79:12–49.

10. Swift AJ, Wild JM, Fichele S, et al. Emphysematous changes and normal variation in smokers and COPD patients using diffusion 3He MRI. Eur J Radiol 2005;54:352–358.

11. Fichele S, Woodhouse N, Swift AJ, et al. MRI of helium-3 gas in healthy lungs: posture related variations of alveolar size. J Magn Reson Imaging 2004;20:331–335.

12. Tustison NJ, Avants BB, Cook P, Egan A, Chen Y, Yushkevich P, Gee JC. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 2010;29:1310–1320.

13. Frangi AF, Niessen WJ, Hoogeveen RM, van Walsum T, Viergever MA. Model-based quantitation of 3-D magnetic resonance angiographic images. IEEE Trans Med Imaging 1999;18:946–956.

14. Antiga L. Generalizing vesselness with respect to dimensionality and shape. Insight J 2007:1–14.

15. Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005;26:839–851.

16. Woodhouse N, Wild JM, Paley MNJ, et al. Combined helium-3/proton magnetic resonance imaging measurement of ventilated lung volumes in smokers compared to never-smokers. J Magn Reson Imaging 2005;21:365–369.

17. Avants BB, Tustison NJ, Wu J, Cook P, Gee JC. An open source multivariate framework for n-tissue segmentation with evaluation on public data. Neuroinformatics 2011 [Epub ahead of print].

18. Li SZ. Markov random field modeling in image analysis, 3rd ed. Berlin: Springer; 2009.

19. Besag J. On the statistical analysis of dirty pictures (with discussions). J R Stat Soc Ser B 1986;48:259–302.

20. Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 1979;9:62–66.

21. Lloyd S. Least squares quantization in PCM. IEEE Trans Inform Theory 1982;28:129–137.

22. de Lange EE, Altes TA, Patrie JT, et al. Evaluation of asthma with hyperpolarized helium-3 MRI: correlation with clinical severity and spirometry. Chest 2006;130:1055–1062.

23. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31: 1116–1128.

24. Rohlfing T, Russakoff DB, Maurer CR. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Trans Med Imaging 2004;23:983–994.

25. Turk M, Pentland A. Eigenfaces for recognition. J. Cogn Neurosci 1991;3:71–86.

26. Tgavalekos NT, Tawhai M, Harris RS, et al. Identifying airways responsible for heterogeneous ventilation and mechanical

dysfunction in asthma: an image functional modeling approach. J Appl Physiol 2005;99:2388–2397.

27. Fuld MK, Easley RB, Saba OI, et al. CT-measured regional specific volume change reflects regional ventilation in supine sheep. J Appl Physiol 2008;104:1177–1184.

28. Salito C, Aliverti A, Gierada DS, et al. Quantification of trapped gas with CT and 3 he MR imaging in a porcine model of isolated airway obstruction. Radiology 2009;253:380–389.

29. Chae EJ, Seo JB, Lee J, et al. Xenon ventilation imaging using dual-energy computed tomography in asthmatics: initial experience. Invest Radiol 2010;45:354–361.

30. Kauczor HU. Hyperpolarized helium-3 gas magnetic resonance imaging of the lung. Top Magn Reson Imaging 2003;14:223–230.

31. Zhang Y, Brady M, Smith S. Segmentation of brain MRI images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 2001;20: 45–57.

32. Koumellis P, van Beek EJR, Woodhouse N, et al. Quantitative analysis of regional airways obstruction using dynamic hyperpolarized 3He MRI-preliminary results in children with cystic fibrosis. J Magn Reson Imaging 2005;22:420–426.

33. Mentore K, Froh DK, de Lange EE, Brookeman JR, Paget-Brown AO, Altes TA. Hyperpolarized He3 MRI of the lung in cystic fibrosis: assessment at baseline and after bronchodilator and airway clearance treatment. Acad Radiol 2005;12:1423–1429.

34. Ley S, Zaporozhan J, Morbach A, et al. Functional evaluation of emphysema using diffusion-weighted 3helium-magnetic resonance imaging, high-resolution computed tomography, and lung function tests. Invest Radiol 2004;39:427–434.

35. Fain SB, Panth SR, Evans MD, et al. Early emphysematous changes in asymptomatic smokers: detection with 3He MR imaging. Radiology 2006;239:875–883.

36. Morbach AE, Gast KK, Schmiedeskamp J, et al. Diffusion-weighted MRI of the lung with hyperpolarized helium-3: a study of reproducibility. J Magn Reson Imaging 2005;21:765–774.

37. Shrout PE and Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;2:420–428.

38. Parraga G, Ouriadov A, Evans A, et al. Hyperpolarized 3He ventilation defects and apparent diffusion coefficients in chronic obstructive pulmonary disease: preliminary results at 3.0 Tesla. Invest Radiol 2007;42:384–391.

39. Mata J, Altes T, Knake J, Mugler J, Brookeman J, de Lange E. Hyperpolarized 3He MR imaging of the lung: effect of subject immobilization on the occurrence of ventilation defects. Acad Radiol 2008;15:260–264.

40. Campana L, Kenyon J, Zhalehdoust-Sani S, et al. Probing airway conditions governing ventilation defects in asthma via hyperpolarized MRI image functional modeling. J Appl Physiol 2009;106: 1293–1300.