Local SIMPLE Multi Atlas-Based Segmentation Applied to Lung Lobe Detection on Chest CT

M. Agarwal[a,b], E.A. Hendriks[b], B.C. Stoel[a], M.E. Bakker[a], J.H.C. Reiber[a] and M. Staring[a]

[a]Leiden University Medical Center, Leiden, The Netherlands
[b]Delft University of Technology, Delft, The Netherlands

ABSTRACT

For multi atlas-based segmentation approaches, a segmentation fusion scheme which considers *local* performance measures may be more accurate than a method which uses a global performance measure. We improve upon an existing segmentation fusion method called SIMPLE and extend it to be localized and suitable for multi-labeled segmentations. We demonstrate the algorithm performance on 23 CT scans of COPD patients using a leave-one-out experiment. Our algorithm performs significantly better ($p < 0.01$) than majority voting, STAPLE, and SIMPLE, with a median overlap of the fissure of 0.45, 0.48, 0.55 and 0.6 for majority voting, STAPLE, SIMPLE, and the proposed algorithm, respectively.

Keywords: multi atlas-based segmentation, SIMPLE, local, multi-label, registration.

1. INTRODUCTION

Pulmonary CT scan images have an important role in the diagnosis of several lung diseases such as lung cancer, old or new pneumonia, tuberculosis, emphysema and chronic obstructive lung diseases (COPD). The lungs are divided into lobes, separated by a thin sheets of tissue, the fissures. The left lung has one oblique fissure and thus two lobes; the right lung has an oblique and a horizontal fissure, and thus three lobes. Many of the lung diseases affect the separate lung lobes to a different extent. Sometimes a disease may even be confined to a single lobe. Therefore lobe-by-lobe quantification of the lungs becomes important for disease severity estimation and treatment planning. The fissures can be visualized using imaging, most notably with thin-slice CT. Lobe segmentation using CT is the clinical goal of this paper.

We employed an atlas-based segmentation (ABS) approach for this task, where an atlas is a set of two images: the intensity image and its segmentation. In ABS, a known segmentation of the intensity image can be used to obtain the segmentation of an unseen image. First the intensity image of the atlas is registered to the unseen image. This provides a transform from the atlas domain to the unseen image domain. Subsequently, this transform is applied on the segmentation of the atlas resulting in a segmentation of the unseen image. The accuracy of the segmentation of the unseen image thus obtained depends on a single registration. The use of multiple atlases was shown[1] to cope with the errors in a segmentation obtained from a single ABS. Furthermore, a *segmentation fusion* based multi atlas-based segmentation (MABS) approach for combining multiple atlases can outperform other single atlas-based and MABS approaches.[2] Some of the commonly used segmentation fusion approaches are: majority voting, Simultaneous Truth and Performance Level Estimation (STAPLE)[3],[4] and Selective and Iterative Method for Performance Level Estimation (SIMPLE).[5] SIMPLE was found to outperform both majority voting and STAPLE.[5] Therefore, in this paper we aim to further improve SIMPLE.

A segmentation obtained from an ABS approach may have local errors because of local failures of the image registration algorithm. A segmentation fusion strategy, which can select locally correct regions from different segmentations and combine them to make one final segmentation can perform better than a method which considers global weights of the individual segmentations.[6] Such a strategy would require a separate performance measure or weight computation for each local region, within an input segmentation. Both STAPLE and SIMPLE

Further author information: (Send correspondence to M.S.)
E-mail: m.staring@lumc.nl, Telephone: +31 71 526 2137

make use of global performance measures for individual segmentations. We propose a *local* version of SIMPLE that uses local weights for fusion of the input segmentations. Since SIMPLE is designed to work only for binary segmentations, we developed a multilabel segmentation algorithm enabling the segmentation of the right and left lung which contain 4 (0, 1, 2, and 3) and 3 labels (0, 1, and 2), respectively, 0 being background. We improved the SIMPLE algorithm on two aspects: (1) add local weight computation (2) extend the binary character to a multi-label algorithm.

In Section 2 details about SIMPLE, a description of a local and multi-label version of SIMPLE, and the registration setup is given. Section 3 contains a description of our data, experimental setup, and results. A discussion on the results and concluding remarks are given in Section 4.

2. METHODS

The algorithms presented in this paper were implemented in the C++ language using the ITK.

2.1 SIMPLE algorithm

SIMPLE is an iterative algorithm that computes an initial estimate of the segmentation by majority voting of all the input segmentations. Subsequently, the Dice overlap of each input is calculated w.r.t. this initial estimate. Inputs with a Dice value less than a certain threshold are rejected. The Dice values are then used as weights in a weighted voting of the remaining inputs. Thus, it provides a new estimate of the segmentation at the end of each iteration. Again, the overlap of each remaining input is computed w.r.t. this new estimate. The process stops when no inputs are rejected anymore. In SIMPLE, all the local regions of an input segmentation are assumed to be equally accurate and each voxel is weighted by the same global overlap weight value. See Algorithm 1. A total of N segmentations are used as input for the fusion algorithm. N_i represents the number of input segmentations remaining in the i^{th} iteration. S_n denotes the n^{th} input segmentation, where $n = 1, 2, 3 \cdots N$. μ_i^{Dice} and σ_i^{Dice} are the mean Dice and standard deviation of N_i Dice values, respectively, in the i^{th} iteration. $E(GT)_i$ represents an estimate of the ground truth segmentation at the end of the i^{th} iteration.

Algorithm 1 SIMPLE algorithm

Require: $S_j(j = 1, 2, 3 \cdots N), \alpha$
perform majority voting of S_j to obtain an initial estimate of ground truth, i.e., $E(GT)_0$
initialize $i = 1, N_0 = 0$, and $N_1 = N$
while $N_i \neq N_{i-1}$ **do**
  for $j = 1$ to N_i **do**
    compute $Dice_{i,j}$ = overlap of S_n and $E(GT)_{i-1}$
  end for
  compute μ_i^{Dice} and σ_i^{Dice} using N_i inputs
  compute $\tau_i = \mu_i^{Dice} - \alpha \times \sigma_i^{Dice}$
  for $j = 1$ to N_i **do**
    if $Dice_{i,j} < \tau_i$ **then**
      reject S_j
    end if
  end for
  $i = i + 1$
  perform weighted voting of remaining $S_j(j = 1, 2, 3 \cdots N_i)$ to obtain $E(GT)_i$. $Dice_{i,j}$ is used as weight for S_j.
end while
$E(GT)_i$ is final fusion result, i.e., $E(GT)$ (estimated ground truth segmentation)

Threshold Selection : SIMPLE computes a unique threshold value in each iteration i using: $\tau_i = \mu_i^{Dice} - \alpha \times \sigma_i^{Dice}$, where μ_i^{Dice} and σ_i^{Dice} represent the mean and standard deviation of overlap values of all the input segmentations. This threshold also depends on a scalar variable α which influences the convergence speed of SIMPLE. α controls the amount of input data rejected in an iteration. Therefore, it is important to choose the right α to get optimum performance from the algorithm.

2.2 Local Multi-label SIMPLE algorithm

The input segmentations generated by ABS may contain a local error because of local errors in the registration. An input segmentation may be locally correct in some areas (i.e. similar to the ground truth segmentation), while in in other areas it may be wrong, it is reasonable to assign different weights to different local regions within the same input segmentation.

In local multi-label SIMPLE, an input segmentation is divided into smaller overlapping cubes of pre-defined size, in other words we use a sliding window. The weight for each voxel of an input segmentation is calculated by considering the Dice overlap in its 3D neighborhood and the current estimate of the final segmentation in that neighborhood. Therefore, each voxel from an input segmentation is rejected or retained based on the overlap value for its neighborhood.The term cube size here represents the size of neighborhood considered. The process starts with an initial estimate of the segmentation by majority voting of all the input segmentations. Subsequently, the Dice overlap of a set of voxels with a common label within an overlapping cube in an input is calculated w.r.t the similar label voxels in the corresponding cube of the initial estimate of the segmentation. In this setup, we get a separate mean overlap and threshold for each label and each cube. Therefore, instead of rejecting the entire 3D input segmentation, voxels with a particular label within local cubes are rejected depending on the corresponding threshold for the label in the cube index. Weighted voting is performed for each cube index separately at the end of each iteration using only the remaining cubes corresponding to the cube index c. The process continues until no voxels are rejected from any input anymore. In this methodology, we focus on rejecting only inaccurate parts within an input segmentation and remain the accurate ones that can still participate in the weighted voting.

More formally, local multi-label SIMPLE computes the average Dice overlap $\mu_{i,c,l}^{Dice}$ and standard deviation $\sigma_{i,c,l}^{Dice}$ over N patients, associated with the label l for cube index c, in iteration i using the equation:

$$\mu_{i,c,l}^{Dice} = \frac{1}{N} \sum_{j} Dice_{i,c,l,j} \tag{1}$$

$$\sigma_{i,c,l}^{Dice} = \sqrt{\frac{1}{N-1} \sum_{j} (Dice_{i,c,l,j} - \mu_{i,c,l}^{Dice})^2} \tag{2}$$

Further, a threshold $\tau_{i,c,l}$ is computed using $\mu_{i,c,l}^{Dice}$ and $\sigma_{i,c,l}^{Dice}$ as:

$$\forall i, c, l : \tau_{i,c,l} = \mu_{i,c,l}^{Dice} - \alpha \sigma_{i,c,l}^{Dice} \tag{3}$$

In Algorithm 1, line 5 shows the computation of a global weight $Dice_{i,j}$ in i^{th} iteration, for j^{th} input segmentation. Instead, we consider local weight $Dice_{i,c,l,j}$ in local multi-label SIMPLE. Similarly μ_i^{Dice}, σ_i^{Dice}, and τ_i in the line 7 and 8 are replaced by $\mu_{i,c,l}^{Dice}$, $\sigma_{i,c,l}^{Dice}$, and $\tau_{i,c,l}$. Furthermore, $Dice_{i,c,l,j}$ is compared with $\tau_{i,c,l}$ to make a decision on cube index c from j^{th} input segmentation, as opposed to line 10 in Algorithm 1.

2.3 Image registration

All registrations in the current work were performed using the registration package `elastix`.[7] Inter-patient registration of COPD patients was complicated. The presence of vessels around the fissure and thin anatomy of fissures make it difficult to align the fissure. Therefore we enhanced the fissures in the unseen and atlas images before registration. If the intensity of a fissure distinctly differs from that of the nearby lobe region then it helps the registration to mainly align the fissure. We employed the following pre-processing steps before registration: 1) fissure enhancement, 2) vessels removal, 3) noise removal, 4) vessel suppression. The process is illustrated in Figure 1.

1. Fissure enhancement: An initial estimate of the fissure is computed using the work of Frangi *et al.*[8]

2. Vessel removal: The fissure detection not only responds to the fissure, but also to vessels. Therefore vessels that are detected based on the work of Xiao *et al.*,[9] are excluded from the result of step 1.

3. Noise removal: The image obtained after vessel removal still contains noise. Thus, we employed connected component analysis to retain only the largest component. This provides us with an initial estimate of fissure, which is further enhanced in the CT scan image.

4. Vessel suppression: We scale down the intensity of all the voxels in CT image by a constant factor, except at the fissure. This suppresses the surrounding vessels.

Lets call the intensity image in j^{th} atlas as I_j and its segmentation as Is_j. The corresponding fissure-enhanced image is represented by $Ienhanced_j$. U and $Uenhanced$ denote the original and enhanced unseen image, respectively. We employed the following three step registration strategy, which is inspired on previous research.[10] We made use of binary lung masks in the step 2 and 3.

1. Affine registration of I_j to U. We obtain the transform T_{j1} as a result of this registration. The affine registration is provides a coarse global alignment.

2. Nonrigid registration of the 'masked' I_j to the 'masked' U. The transform T_{j1} obtained in step1 is used as an initial transform for this registration. We obtain transform T_{j2} as a result of this registration.

3. Nonrigid registration of the 'masked' $Ienhanced_j$ to the 'masked' $Uenhanced$. The transform T_{j2} obtained in step2 is used as an initial transform for this registration. We obtain transform T_{j3} as a result of this registration. T_{j3} is applied to the segmentation Is_j. This gives S_j, a segmentation of the unseen image U.

See Staring *et al.*[10] for details on the registration parameters used and effect of 'masking'.

3. EXPERIMENTS AND RESULTS

3.1 Validation

The data used in this experiment contains CT scans of 23 COPD patients from a former study,[11] scanned without contrast media. On average, the resolution of one lung image was approximately $250 \times 350 \times 600$ pixels. To evaluate the proposed method, ground truth lobe segmentations were generated using an automated seeded region growing approach[12] by an experienced observer, inspected and approved by a pulmonologist. The left lung was segmented into 3 regions: upper lobe (intensity value 2), lower lobe (intensity value 1), and background (intensity value 0). Binary lung masks of the entire dataset were also prepared.

The overlap between the results of the proposed method and the ground truth was defined by the Dice similarity coefficient (DSC). To better appreciate differences, the overlap was additionally computed on the fissure area. The fissure area was defined as a three voxel wide area around the fissure, computed by a morphological gradient operation on the lobe segmentations. The DSCs are presented by box plots.

We conducted the experiment in a leave-one-out fashion. Each image was selected once as the unseen image and all remaining images as the atlas. In this paper, results are shown only for the left lung. In future we will test our algorithm on the right lung as well.

3.2 Optimization

Choice of α : Since the choice of α is important for optimal performance of the algorithm, see Section 2.1, we experimented with a set of six α values: $1, 1.25, 1.5, 1.75, 0.5/f, 1/f$, where f denotes the fraction of segmentations selected in the previous iteration. Out of these values, an α with the highest median value in the box-plot was considered best. Here we only show the result for $\alpha = 0.5/f$, which was on average the best result, although no significant difference was found with the other α values.

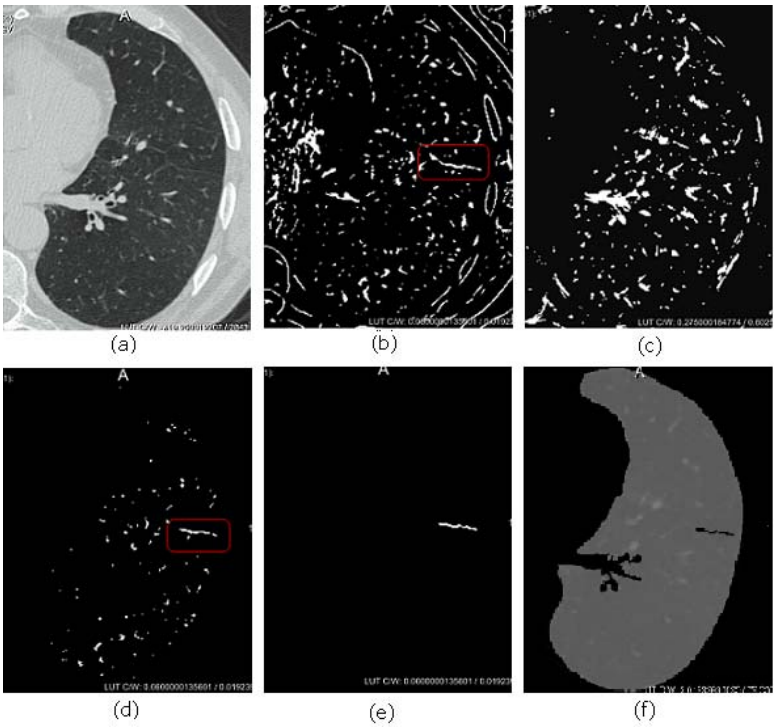
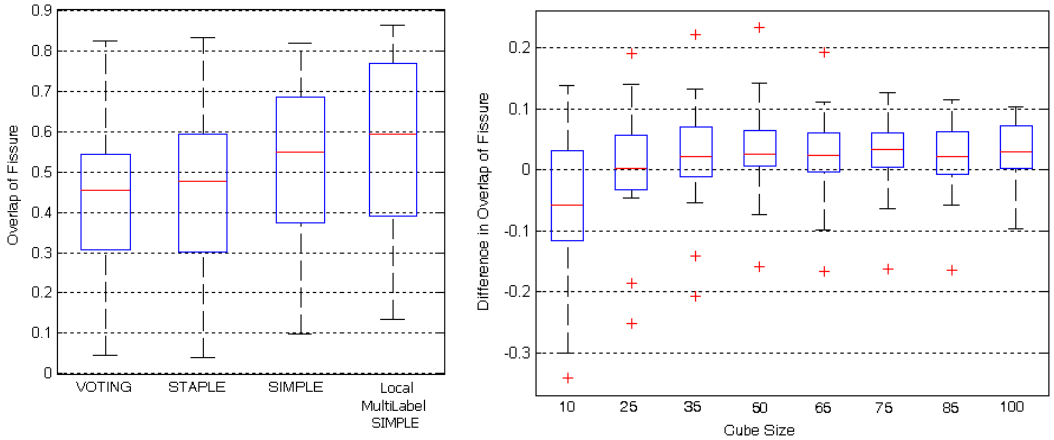


Figure 1. (a) A CT scan image of the left lung, (b) Response of the sheetness filter. The fissure (see the red ROI) and numerous vascular structures are visible, (c) Response of the vesselness filter. It does not detect all vessels that are present in the response of sheetness filter, (d) Image obtained from subtracting (b) - (c) and masked afterwards. Some of the vessels are still present in addition to the fissure, (e) Fissure obtained after connected component analysis of segmentation from step (d), (f) The *enhanced* CT image.



(a) (b)

Figure 2. a) Box-plot of the 23 overlap values obtained on comparing the left lung fissure (3 voxel wide) in segmentation fusion result and the ground truth segmentation. A result obtained on using cube size 50 is shown for local multi-label SIMPLE. b) Comparison against SIMPLE. Effect of changing cube size on difference in overlap of fissure for local multi-label SIMPLE versus SIMPLE algorithm. The results are shown for $\alpha = 0.5/f$.

3.3 Results

Result of SIMPLE compared to STAPLE and VOTING : Figure 2(a) shows performance of majority voting, STAPLE, and SIMPLE when tested on our dataset. It is evident that SIMPLE outperforms both majority voting and STAPLE, as suggested in the available literature.[5]

Result of local multi-label SIMPLE : Figure 2(b) gives an overview of performance of local multi-label SIMPLE for various cube sizes. The Wilcoxon test statistics for comparing the Dice overlap between local and global SIMPLE gave the following p-values: 0.0516, 0.4115, 0.0885, 0.0177, 0.0138, 0.0208, 0.0225, and 0.0176 for cube sizes 10, 25, 35, 50, 65, 75, 85, and 100, respectively. Considering a threshold of 0.05, local multi-label SIMPLE provides significant improvement over SIMPLE for all cube sizes greater than 35. In figure 2(b), the box-plot of fissure overlap is quite low for cube size 10. It seems that cube size 10 is too low (therefore contains too small number of voxels) to compute the Dice measure reliably.

4. DISCUSSION AND CONCLUSION

Majority voting, STAPLE, and SIMPLE are global segmentation fusion strategies. They differ from each other in the manner in which weights are assigned to participating segmentations. Whereas majority voting assigns weight 1 to each segmentation and STAPLE makes use of the confusion matrix for the weight computation, SIMPLE uses the Dice measures as weights and if the weight of a segmentation is below a certain threshold, it becomes 0. But it is possible that even a segmentation which is assigned weight 0 resembles the final segmentation in some of the local regions. A local segmentation fusion strategy can more efficiently select locally correct regions in various segmentations compared to a global strategy. Therefore, we proposed a local multi-label version of SIMPLE. The multi-label aspect is important to obtain a lobe segmentation of the lung data having 3 labels for the left and 4 labels for the right lung. The result of local multi-label SIMPLE provides a statistically significant improvement over SIMPLE for all cube sizes ≥ 50 using a Dice overlap of 3 voxel wide fissure, and over majority voting and STAPLE too. For cube size 50, a median overlap of the fissure is 0.55 and 0.6 for SIMPLE and the proposed algorithm, respectively. The reason for the variation in performance with cube size is caused by two factors: (1) Smaller cube size provides more local weights and as we increase the cube size, the weights become more and more global. Therefore we lose the advantage of locality. It suggests choosing a lower cube size for higher performance. (2) A very small cube size may not contain enough voxels to compute the Dice measure reliably. Unreliable Dice weights is the main reason for poor performance at lower cube sizes. This effect works contrary to the first one.

One limitation of multi-atlas based segmentation is that if all the input segmentations are wrong at a particular local region, their fusion output would also be wrong at that location. The low tail of box-plot in figure 2(a) is caused by such input segmentations. Therefore no fusion strategy, global or local, could detect the fissures correctly. The only way to improve this situation is to obtain better input segmentations by improving the image registration strategy.

In conclusion, the local multi-label SIMPLE outperforms majority voting, STAPLE, and SIMPLE.

REFERENCES

1. T. Rohlfing *et al.*, "Quo vadis, atlas-based segmentation?," in *The Handbook of Medical Image Analysis – Volume III: Registration Models*, J. Suri, D. L. Wilson, and S. Laxminarayan, eds., ch. 11, pp. 435–486, Kluwer Academic / Plenum Publishers, New York, NY, Aug. 2005.
2. T. Rohlfing *et al.*, "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage* **21**, pp. 1428–1442, Apr. 2004.
3. S. K. Warfield *et al.*, "Simultaneous truth and performance level estimation (staple): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imaging* **23**, pp. 903–921, 2004.
4. T. Rohlfing *et al.*, "Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation," *IEEE Trans. Med. Imaging* **23**, pp. 983–994, Aug. 2004.
5. R. Langerak *et al.*, "Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (simple)," *IEEE Trans. Med. Imaging* **29**(12), pp. 2000–2008, 2010.
6. X. Artaechevarria *et al.*, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Trans. Med. Imaging* **28**, pp. 1266–1277, Aug. 2009.
7. S. Klein and M. Staring *et al.*, "elastix: a toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging* **29**, pp. 196 – 205, January 2010.

8. R. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," pp. 130–137, Springer-Verlag, 1998.
9. C. Xiao *et al.*, "A strain energy filter for 3d vessel enhancement with application to pulmonary ct images," *Medical Image Analysis* **15**(1), pp. 112 – 124, 2011.
10. M. Staring *et al.*, "Pulmonary Image Registration With elastix Using a Standard Intensity-Based Algorithm," in *Medical Image Analysis for the Clinic: A Grand Challenge, Workshop Proceedings of MICCAI*, 2010.
11. T. Lapperre, J. Sont, A. van Schadewijk, M. Gosman, D. Postma, I. Bajema, W. Timens, T. Mauad, P. Hiemstra, and the GLUCOLD Study Group, "Smoking cessation and bronchial epithelial remodelling in copd: a cross-sectional study," *Respiratory Research* **8**(1), p. 85, 2007.
12. B. C. Stoel and J. Stolk, "Optimization and standardization of lung densitometry in the assessment of pulmonary emphysema," *Invest Radiol* **39**(11), pp. 681 – 688, 2004.