Automated segmentation of the thyroid gland on thoracic CT scans by multiatlas label fusion and random forest classification

Divya Narayanan
Jiamin Liu
Lauren Kim
Kevin W. Chang
Le Lu
Jianhua Yao
Evrin B. Turkbey
Ronald M. Summers

Automated segmentation of the thyroid gland on thoracic CT scans by multiatlas label fusion and random forest classification

Divya Narayanan, Jiamin Liu, Lauren Kim, Kevin W. Chang, Le Lu, Jianhua Yao, Evrim B. Turkbey, and Ronald M. Summers*

National Institutes of Health Clinical Center, Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Building 10, Room 1C224, MSC 1182, Bethesda, Maryland 20892-1182, United States

Abstract. The thyroid is an endocrine gland that regulates metabolism. Thyroid image analysis plays an important role in both diagnostic radiology and radiation oncology treatment planning. Low tissue contrast of the thyroid relative to surrounding anatomic structures makes manual segmentation of this organ challenging. This work proposes a fully automated system for thyroid segmentation on CT imaging. Following initial thyroid segmentation with multiatlas joint label fusion, a random forest (RF) algorithm was applied. Multiatlas label fusion transfers labels from labeled atlases and warps them to target images using deformable registration. A consensus atlas solution was formed based on optimal weighting of atlases and similarity to a given target image. Following the initial segmentation, a trained RF classifier employed voxel scanning to assign class-conditional probabilities to the voxels in the target image. Thyroid voxels were categorized with positive labels and nonthyroid voxels were categorized with negative labels. Our method was evaluated on CT scans from 66 patients, 6 of which served as atlases for multiatlas label fusion. The system with independent multiatlas label fusion method and RF classifier achieved average dice similarity coefficients of 0.72 ± 0.13 and 0.57 ± 0.14 , respectively. The system with sequential multiatlas label fusion followed by RF correction increased the dice similarity coefficient to 0.76 ± 0.11 and improved the segmentation accuracy. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.2.4.044006](https://doi.org/10.1117/1.JMI.2.4.044006)]

Keywords: thyroid segmentation; multiatlas label fusion; random forest.

Paper 15150PR received Jul. 20, 2015; accepted for publication Dec. 1, 2015; published online Dec. 30, 2015.

1 Introduction

The thyroid is a small H- or U-shaped gland that is located anterior to the trachea and inferior to the larynx. As an endocrine gland, it regulates metabolism by secreting thyroid hormone to modulate energy utilization.

Thyroid image analysis plays an important role in both diagnosis and treatment. For example, volume analysis can be used for the diagnosis of thyroid disorders. In patients with head and neck cancer, precise delineation of the thyroid is required to spare thyroid tissue from irradiation.^{1,2} Nodule detection can be applied and organ-specific radiation dose can be estimated with thyroid segmentation. Furthermore, thyroid segmentation may improve the performance of other methods to detect abnormalities, such as mediastinal and supraclavicular lymph nodes, by reducing false positives.

In the current clinical workflow, radiologists and radiation oncologists manually segment the thyroid slice by slice on CT images, a time-consuming and error-prone task. As a result, a fully automated system for segmentation would be beneficial. However, automated segmentation of the thyroid is challenging. Anatomic structures surrounding the thyroid, such as muscles, veins, and lymph nodes, have low tissue contrast that makes them difficult to distinguish from the thyroid. Figure 1 shows an example of the thyroid on a chest CT exam. Figure 1(a)

also shows that the superior portion can be excluded from the field of view (FOV) on chest CT.

Although chest CT exams may not cover the entire thyroid, our work was motivated by the reduction of radiation in patients undergoing CT imaging. Chest CT is a common imaging modality and can impart a substantial radiation dose to the thyroid. The thyroid irradiation is usually incidental to the purpose of the scan, generally to evaluate the lungs and mediastinum.

Because the thyroid is surrounded by structures with similar attenuation, utilization of standard intensity-based methods such as level-set³ or graph-cut⁴ are insufficient for segmentation.² Atlas-based segmentation, which uses registration, has proven to be a powerful and successful concept.⁵ The rationale behind atlas-based segmentation is that segmentation of a target image correlates with an atlas image by looking at spatial features and anatomic landmarks captured in an atlas. A target image can be segmented by deforming the atlas image; the segmentation of the atlas is then transferred to the target image by deformable registration. As an extension, multiatlas segmentation makes use of more than one atlas to compensate for the potential biases of individual atlases. The final segmentation is generated by joint label fusion, which combines the results produced by all atlases into a consensus solution.

Atlas-based segmentation methods have been used for thyroid segmentation in CT imaging.^{2,5} In Chen et al.,² methods using single and multiple atlases were evaluated and it was

*Address all correspondence to: Ronald M. Summers, E-mail: rms@nih.gov

Fig. 1 (a) 3-D model of the thyroid (white), (b) thyroid in two-dimensional axial CT slice with (c) ground truth segmentation (pink).

concluded that multiatlas methods outperformed single-atlas methods for thyroid segmentation.² In Asman et al.,⁵ multiatlas with nonlocal statistical label fusion was applied to segment the thyroid on CT images.

One drawback of atlas-based segmentation is that insufficient similarity between the atlases and the target image often results in segmentation errors. There are two common approaches to address this issue: optimal atlas selection^{6–8} and robust label fusion.^{5,9} Although adding atlases can improve performance, registration between atlases and target images is time-consuming, making the method unsuitable for clinical practice. Previous works on advanced label fusion have reported better performance than the traditional majority voting and similarity voting methods.^{5,9} In this work, we present a fully automated method for thyroid segmentation on CT images. Our system utilizes multiatlas label fusion (MALF) to generate an initial segmentation of the thyroid. Following MALF, the initial thyroid segmentation is corrected by supervised, learning-based voxel labeling with a random forest (RF) classification.² The major contribution of our work is the combined utilization of MALF and supervised learning in the form of an RF classifier to correct the systematic segmentation error of the multiatlas methods for more accurate thyroid segmentation.

2 Methods

Our method consisted of three major steps: (1) identification of region of interest (ROI), (2) initial thyroid estimation by MALF, and (3) segmentation correction by RF classification.

2.1 Region of Interest

Because the thyroid is a small structure in chest CT images, an ROI enclosing it was identified using segmentations of the trachea and spine to reduce the search region and to increase the segmentation efficiency. In addition, compared with using the entire chest CT image, the smaller ROI can provide an initial localization for the deformable registration in the multiatlas method that follows. The trachea was segmented using thresholding and morphological operations, and the spine was segmented using a watershed algorithm followed by a directed graph search.¹⁰ The result of the two reference segmentations was an automatically determined thyroid ROI (Fig. 2). The ventral boundary of the ROI was set to be the ventral surface of the thyroid cartilage, and the dorsal boundary of the ROI was set as the ventral margin of the vertebral body.

The width of the ROI, or the x -dimension, was determined by the width of the spinal column, and the y -dimension of the ROI was set as twice the diameter of the trachea [Fig. 2(a)]. Based on the average thyroid height of 5 cm in the US adult population,¹¹

the height of the ROI, or the dimension of the z -direction, was set to 6 cm for all patients. Due to positioning of the thyroid in a chest CT scan, the bounding box in the z -direction extended from the beginning of the scan to a slice 6 cm inferior. A buffer region of a few voxels was included in all three directions to fully capture the thyroid within the bounding box [Fig. 2(b)]. The following processing steps are only applied to these thyroid ROI subimages for efficiency.

2.2 Initial Thyroid Segmentation

In this work, the thyroid was estimated by multiatlas with joint label fusion.¹² An atlas A was defined as a pair of images $A = (I, L)$, where I was the grayscale CT image and L was the manually labeled thyroid gland. Let $U = (I_T, L_T)$ be the target to be segmented, where L_T denotes the target thyroid. $A_1 = (I_1, L_1), \dots, A_n = (I_n, L_n)$ are n atlases. A registration was required to propagate the labels L_i of atlas A_i to the target U . In our implementation, each CT image I_i in the atlas A_i was registered to the CT image I_T in target U by the fast free-form deformation algorithm developed in NiftyReg.¹³ Cubic B-splines were used to deform a source image to optimize an objective function based on the normalized mutual information and a bending energy term. Grid spacing along three axes was set as 5 mm. The weight of the bending energy term was 0.005, and normalized mutual information with 64 bins was used. The optimization was performed in three coarse-to-fine levels, and the maximal number of iterations per level was 300. More details can be found in Ref. 13. The registration resulted in n transformations that were used to warp the thyroid L_i in atlas A_i to L_i' in the space of target U . Nearest-neighbor interpolation was used because the labels are binary images. For

Fig. 2 Thyroid ROI. (a) x - and y -dimensions (red arrows) are determined by reference segmentations of the spinal column (ivory), spinal canal (teal), and trachea (dark blue), giving rise to the ROI (dashed yellow box). (b) Height of the ROI set to 6 cm (yellow arrow) around thyroid (pink).

each voxel in I_T of the target U , each atlas provided an opinion about the label. Finally, $\hat{L}_T(x)$, the probability of thyroid at voxel x in the target U , was determined by

$$\hat{L}_T(x) = \sum_{i=1}^n \omega_i(x) L_i'(x),$$

where $L_i'(x)$ is the warped i 'th thyroid atlas, and the thyroid segmentation $L_T(x)$ in target U could be computed by thresholding $\hat{L}_T(x)$. Figure 3 shows the method of MALF.

$\omega_i(x)$ is a weight assigned to the i 'th atlas with $\sum_{i=1}^n \omega_i(x) = 1$. $\omega_i(x)$ is usually estimated by similarity-weighted voting strategies,⁷ where the grayscale image in an atlas that was more similar to the target image would have been assigned higher weight. One limitation of this approach was that the voting weight for each atlas was estimated independently from other atlases, ignoring potential correlations

among the atlases. To address this issue, we employed the joint label fusion algorithm,¹² which estimated voting weights $\omega_i(x)$ by simultaneously considering pairwise atlas correlations

$$\omega_i(x) = \frac{M_x^{-1} \mathbf{1}_n}{\mathbf{1}_n^T M_x^{-1} \mathbf{1}_n},$$

where M_x is a matrix measuring the correlation between the i 'th and j 'th atlases and $\mathbf{1}_n = [1; 1; \dots; 1]$ is a vector of size n . More detail about how to capture the probability that different atlases produce the same label error at location x by a dependency matrix M_x can be found in Ref. 12. As shown in Ref. 12, joint label fusion performed better than label fusion with independent similarity-based weight estimation.

Six atlases ($n = 6$) were used in our experiment. The details about atlas selection are described in the Sec. 3. Figure 4(b) shows the probability map of thyroid using multiatlas with joint label fusion.

Fig. 3 Illustration of MALF.

2.3 Segmentation Correction by Random Forest

Segmentation errors from the multiatlas-based method could be corrected by a data-driven learning method. In this work, the initial thyroid segmentation was corrected by RF classification. An RF classifier is often considered an ensemble method, as it consists of a number of trees that generate many weak classifiers and aggregates their results in order to make decisions via a strong classifier. The RF classifier has a number of advantages: (i) training and classification are both computationally efficient, (ii) features do not require normalization, (iii) a large variety of features can be seamlessly handled, and (iv) feature selection is inherent. Hence, RF classification has been successfully applied to various organ localization and segmentation tasks in different imaging modalities.^{14–16}

Fig. 4 (a) Thyroid ROI in target image. (b) Thyroid probability map $\hat{L}_T$ from multiatlas joint label fusion. (c) Initial estimated thyroid with surrounding voxels (all voxels with $\hat{L}_T \geq 0.10$). (d) Thyroid probability map after RF correction.

Fig. 5 (a) The RF is trained with positive samples (+) from all voxels of manually labeled thyroids and (b) negative samples (-) randomly selected from a surrounding shell.

In RF training, positive samples were all voxels of the labeled thyroids. An equal number of negative samples were obtained by randomly selecting voxels within a shell surrounding the thyroid. All negative voxels within a 2-voxel distance from the boundary of the labeled thyroid, as determined by the distance transform, were not used for training due to their low confidence of being negative or positive. Negatives beyond 2 voxels but within 15 voxels were randomly sampled (Fig. 5).

During training, our RF classifier took in features including the histogram of oriented gradient within a $7 \times 7 \times 3$ patch of voxels, CT attenuation, and the thyroid probability $\hat{L}_T$ from MALF. These features were selected in order to maximize the information gain function. Fifty trees were used in our RF classifier, where each node of a given tree assessed possible sample splits on three randomly selected features. A tree branch was terminated if a split node resulted in two or fewer samples in the branch.

Once trained, the forest was then applied to the initially estimated thyroid (all voxels with $\hat{L}_T \geq 0.10$), an empirically derived threshold [Fig. 4(c)]. Each voxel was pushed through the tree starting at the root. The resulting path of a given voxel was dependent on the threshold value of a given feature, as determined during the training phase, and the response to the randomly selected features of the particular voxel. The posterior probability at each final node was averaged to determine the final likelihood of thyroid [Fig. 4(d)]. The final segmentation of thyroid was derived to be all voxels where RF probability ≥ 0.8 , an empirically derived threshold. The small isolated segmented voxels ($\leq 10 \text{ mm}^3$) were removed by connected component analysis.

3 Experiment and Results

3.1 Data Set

As a retrospective review of imaging and the electronic medical record, this study was HIPAA-compliant, approved by our institutional review board, and the need for signed informed consent was waived.

Our patient population consisted of 66 patients with large FOV, intravenous contrast-enhanced, chest CT scans that included the thyroid. Of these 66 patients, 6 served as atlases, as discussed in the atlas selection section, and the remaining 60 patients were used for segmentation evaluation. The CT slice thickness was 1 mm, with axial slice voxel spacing in the range of 0.7 to 0.9 mm. The reference standard consisted of manual tracings of the 66 thyroids by a trained research assistant under the supervision of a board-certified radiologist. The manual

Fig. 6 MALF performance as a function of the number of atlases.

tracings were completed on a touch-screen tablet using a three-dimensional (3-D) slicer¹⁷ with a soft tissue window level setting.

3.2 Atlas Selection

Atlas selection is beneficial for MALF to generate more accurate segmentations based on similarity between atlases and target images. In our experiment, we started with two randomly selected atlases. The performance of thyroid segmentation using MALF was evaluated on a random subset data set of 20 patients. Using the dice similarity coefficient (DSC)¹⁸ as a measure of segmentation accuracy, a new atlas was randomly selected, added iteratively, and plotted against the respective DSC (Fig. 6).

Table 1 Average computation time per patient.

Computation	Time (min)
ROI	1
Multiatlas registration	30
Label fusion	1
RF correction	3
Total time	35

Table 2 Segmentation accuracy (mean $\pm$ standard deviation).

	MALF	RF	MALF+RF
Dice similarity coefficient	0.72 ± 0.13	0.57 ± 0.15	0.76 ± 0.11
Mean surface distance (mm)	1.20 ± 1.59	2.41 ± 2.21	0.81 ± 1.09

Fig. 7 Box and whisker plot comparing DSC for three methods.

Figure 6 shows that the segmentation accuracy did not improve with the use of more than six atlases. The six atlases that provided the optimal performance in terms of DSC were then used in the MALF of our experiment. The remaining 60 patients were used for thyroid segmentation and evaluation.

3.3 Thyroid Segmentation Evaluation

Segmentation performance was evaluated with threefold cross validation on 60 patients. For each fold, there were on average 730k positive/negative samples, and the RF training took around 100 min. For each patient, six pairs (atlas-to-target) of registrations took around 30 min on average. The computation time for each step is summarized in Table 1.

Systems with three methods, MALF only, RF only, and MALF with RF correction, were evaluated using the DSC and mean surface distance. For the method with MALF only, six atlases were registered to each one of the 60 patients and joint label fusion was utilized to segment the thyroid. For the method with RF only, we still used the registration from the six atlases to provide spatial information for thyroid, and three-fold cross-validation was used for RF classification. There were two aspects of differences compared to the method of MALF with RF correction. First, unlike the case where all voxels with thyroid probability $\hat{L}_T \geq 0.1$ were used for RF correction, all voxels in $\bigcup_{i=1}^6 L'_i$ (the union of all six warped atlases) were pushed into the RF for classification for the RF-only method. Second, the thyroid probability $\hat{L}_T$ from MALF was missing in the feature set for the RF-only method. We set the number of trees as 50 and minimal leaf as 2 in the RF training stage. We also tried different numbers of trees (100, 200, 500) and minimal leaf (2, 3, 5). There is no significant difference for the segmentation results.

The DSC for MALF with RF correction was significantly better ($P \ll 0.01$, paired t -test) than the individual MALF and RF methods (Table 2). A box and whisker plot comparing DSC for the three methods is shown in Fig. 7. The MALF segmentation with RF correction resembled the ground truth thyroid more closely than the other two methods using both

Fig. 8 (a) Thyroid in ROI of CT scan. (b) Manually labeled thyroid (pink). (c) Segmentation from MALF (light blue). (d) Segmentation from MALF with RF correction (green). The improvements (yellow boxes) by the RF correction are highlighted in (d).

DSC and mean surface distance metrics of evaluation. The surface distance metrics were computed in terms of the distance from the manually labeled thyroid to the automated thyroid segmentation.

It is reasonable that MALF with RF correction reported the best performance, because multiple atlases and joint label fusion provided a very good initial segmentation on the object level, and the RF classifier was then trained to correct the segmentation on the voxel level. Figure 8 shows that the falsely segmented and missing voxels were corrected by the RF classifier. Figure 9(a) shows a CT slice where thyroid is not visible. However, MALF reported high thyroid probability [Fig. 9(b)], which will generate false-positive segmentation (DSC = 0.78). The segmentation with RF correction reported very low thyroid

probability [Fig. 9(c)] and corrected the false-positive segmentation error (DSC = 0.82).

More segmentation results are shown in Figs. 10 and 11. Segmentation errors in outlier cases showed that the false-positive segmentations often occurred on surrounding structures, such as the jugular vein or thyroid cartilage (Fig. 11). The oversegmentation of such structures leads to a decreased DSC.

Thyroid volume measurement is important for thyroid disease diagnosis, therapy planning, and treatment assessment. A Bland–Altman plot between the automated volume of thyroid (using MALF with RF correction) and the volume of manually labeled thyroid is shown in Fig. 12. The thyroid volume using our automated method was on average 0.81 mL less than the

Fig. 9 (a) Example of CT slice where thyroid is not visible. (b) Thyroid probability map from MALF. (c) Thyroid probability map from MALF with RF correction.

Fig. 10 Three cases (a)–(c). (a) Thyroid in ROI, (b) manually labeled GT segmentations, and (c) comparable automated thyroid segmentations.

Fig. 11 Three cases (a)–(c). (a) Thyroid in ROI, (b) manually labeled GT segmentations, and (c) poor automated thyroid segmentations.

Fig. 12 Thyroid volume Bland–Altman plot comparing the automated method (MALF with RF correction) and ground truth (GT). The mean difference in volume (red line) and a 95% confidence interval (dotted red lines).

ground truth volume. Two examples of thyroid segmentation in 3-D view are shown in Fig. 13.

4 Discussion and Conclusion

In this work, we created a system utilizing multiatlas and RF classification that automates segmentation of the thyroid. The performance of our system with a DSC of 0.76 for multiatlas with RF correction is better than the current state-of-the-art method with a DSC of 0.74.^{2,5} By extrapolation of data from the box plots presented in Ref. 5, our mean MSD values demonstrate better performance. Threefold cross-validation showed a statistically significant improvement for the method MALF with RF correction, indicating that application of the RF correction improved segmentation accuracy of the automated system.

The selection of atlases given a particular data set impacts the quality of registration of the evaluation images. If a given target

Fig. 13 Thyroid segmentation in 3-D using MALF with RF correction (green) and GT (pink), with overlay between them (blue).

image is not heavily correlated with the selected atlas images, the registration as well as the segmentations resulting from MALF and RF will not give rise to accurate depiction of the thyroid. This also impacts the volume estimation of the thyroid (Fig. 11).

The consistent presence of outliers in our data can be attributed to potential errors in registration of atlas labels to target images. The same patients who are outliers of MALF are outliers of RF and MALF with RF correction, as seen in Fig. 7. These subjects had segmentations that were dissimilar to atlas segmentations, giving rise to poor DSCs. Furthermore, the presence of surrounding cartilage, vessels, and muscle, which have similar CT attenuation to thyroid, can artificially be marked as thyroid by our system (Fig. 11). One way to ameliorate such an issue and have improved RF is to incorporate more nuanced structures into atlases in MALF with these features. The segmentation seen in Figs. 11(b) and 11(c) may not have been oversegmented if the atlases had included the jugular veins for negative training. Another way to remedy this issue would be to use a more diverse atlas set, although employing such an approach would be computationally expensive, as seen in the registration step in Table 1.

There were several limitations to our study. First, the thyroid ground truth tracings were manually completed by one observer. Having multiple observers manually segment these images could reduce the bias introduced by a single observer. Second, the first slice of CT examinations for some patients began in the middle of the thyroid, cutting off superior portions of the volume to be segmented [Fig. 1(a)]. However, this situation is inherent to dedicated chest CT examinations.

Atlas selection is critical for MALF. In our experiment, only one trial was completed for each iterative addition of an atlas. More advanced atlas selection technique could benefit the segmentation of MALF. However, our approach demonstrated the power of the RF corrective learning algorithm. This algorithm has compensated for the requirement of complex atlas selection and registration methods.

The CT scans used in our study were large FOV chest CT examinations. This is a more challenging problem than small FOV scans such as those seen in dedicated neck CT examinations because the spatial resolution of the thyroid is greater in such scans. Because the thyroid is a small organ, having images that do not magnify the thyroid further is a restriction on how the RF algorithm is able to accurately classify a given voxel as thyroid, especially given the large range of similar CT attenuations of surrounding tissues and organs. Our proposed method should be extensible to neck CT scans. Furthermore, neck CT scans with a smaller FOV may lead to more accurate segmentations of the thyroid and other surrounding organs.

In this paper, we present a fully automated thyroid segmentation system for body CT examinations using MALF with a RF correction with a DSC of 0.76, which represents state-of-the-art performance. This system has potential clinical applications and can be used as a diagnostic tool for organ-specific radiation dose estimation, as well as radiotherapy treatment planning.

Acknowledgments

This research was supported by the Intramural Research Program at the National Institutes of Health, Clinical Center. The authors thank Holger Roth for assistance with registration and Dr. Andrew Dwyer, MD, for critical review of this paper.

References

1. W. Shabana, E. Peeters, and M. De Maeseneer, "Measuring thyroid gland volume: should we change the correction factor?" *Am. J. Roentgenol.* **186**, 234–236 (2006).
2. A. Chen et al., "Evaluation of multiple-atlas-based strategies for the segmentation of the thyroid gland in head and neck CT images for IMRT," *Phys. Med. Biol.* **57**, 93–111 (2012).
3. J. A. Sethian, "Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science," in *Cambridge Monographs on Applied and Computational Mathematics*, 2nd ed., p. 378, Cambridge University Press, Cambridge, UK (1999).
4. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 1222–1239 (2001).
5. A. J. Asman and B. A. Landman, "Non-local statistical label fusion for multi-atlas segmentation," *Med. Image Anal.* **17**, 194–208 (2013).
6. 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**, 1428–1442 (2004).
7. M. Wu et al., "Optimum template selection for atlas-based segmentation," *NeuroImage* **34**, 1612–1618 (2007).
8. S. Klein et al., "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," *Med. Phys.* **35**, 1407–1417 (2008).
9. H. Wang and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion and corrective learning—an open source implementation," *Front. Neuroinform.* **7**, 27 (2013).
10. J. Yao, S. D. O'Connor, and R. M. Summers, "Automated spinal column extraction and partitioning," in *3rd IEEE Int. Symp. on Biomedical Imaging: Nano to Macro 2006*, pp. 390–393 (2006).
11. V. Chaudhary and S. Bano, "Thyroid ultrasound," *Indian J. Endocr. Metab.* **17**, 219–227 (2013).
12. H. Wang et al., "Multi-atlas segmentation with joint label fusion," *IEEE Trans. Pattern. Anal. Mach. Intell.* **35**(3), 611–623 (2012).
13. M. Modat, J. McClelland, and S. Ourselin, "Lung registration using the NiftyReg package," in *Medical Image Analysis for the Clinic-A Grand Challenge 2010*, pp. 33–42 (2010).
14. V. Lempitsky et al., "Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography," in *Functional Imaging and Modeling of the Heart, Proc.*, Vol. **5528**, pp. 447–456 (2009).
15. O. Pauly et al., "Fast multiple organ detection and localization in whole-body MR dixon sequences," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 239–247 (2011).
16. R. Cuingnet et al., "Automatic detection and segmentation of kidneys in 3D CT images using random forests," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 66–74 (2012).
17. A. Fedorov et al., "3D slicer as an image computing platform for the quantitative imaging network," *Magn. Reson. Imaging* **30**(9), 1323–1341 (2012).
18. L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology* **26**, 297–302 (1945).

Biographies for the authors are not available.