Effect of segmentation algorithms on the performance of computerized detection of lung nodules in CT

Wei Guo

School of Computer, Shenyang Aerospace University, Daoyi Development District, Shenyang, Liaoning 110136, China

Qiang Li^{a)}

Medical Imaging Information Laboratory, Shanghai Advanced Research Institute, Chinese Academy of Sciences, 99 Haik Road, Pudong District, Shanghai 201203, China and Carl E. Ravin Advanced Imaging Laboratories, Department of Radiology, Duke University Medical Center, Durham, North Carolina 27705

(Received 17 April 2013; revised 7 July 2014; accepted for publication 9 July 2014; published 18 August 2014)

Purpose: The purpose of this study is to reveal how the performance of lung nodule segmentation algorithm impacts the performance of lung nodule detection, and to provide guidelines for choosing an appropriate segmentation algorithm with appropriate parameters in a computer-aided detection (CAD) scheme.

Methods: The database consisted of 85 CT scans with 111 nodules of 3 mm or larger in diameter from the standard CT lung nodule database created by the Lung Image Database Consortium. The initial nodule candidates were identified as those with strong response to a selective nodule enhancement filter. A uniform viewpoint reformation technique was applied to a three-dimensional nodule candidate to generate 24 two-dimensional (2D) reformatted images, which would be used to effectively distinguish between true nodules and false positives. Six different algorithms were employed to segment the initial nodule candidates in the 2D reformatted images. Finally, 2D features from the segmented areas in the 24 reformatted images were determined, selected, and classified for removal of false positives. Therefore, there were six similar CAD schemes, in which only the segmentation algorithms were different. The six segmentation algorithms included the fixed thresholding (FT), Otsu thresholding (OTSU), fuzzy C-means (FCM), Gaussian mixture model (GMM), Chan and Vese model (CV), and local binary fitting (LBF). The mean Jaccard index and the mean absolute distance (Dmean) were employed to evaluate the performance of segmentation algorithms, and the number of false positives at a fixed sensitivity was employed to evaluate the performance of the CAD schemes.

Results: For the segmentation algorithms of FT, OTSU, FCM, GMM, CV, and LBF, the highest mean Jaccard index between the segmented nodule and the ground truth were 0.601, 0.586, 0.588, 0.563, 0.543, and 0.553, respectively, and the corresponding Dmean were 1.74, 1.80, 2.32, 2.80, 3.48, and 3.18 pixels, respectively. With these segmentation results of the six segmentation algorithms, the six CAD schemes reported 4.4, 8.8, 3.4, 9.2, 13.6, and 10.4 false positives per CT scan at a sensitivity of 80%.

Conclusions: When multiple algorithms are available for segmenting nodule candidates in a CAD scheme, the “optimal” segmentation algorithm did not necessarily lead to the “optimal” CAD detection performance. © 2014 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4892056>]

Key words: lung nodule, segmentation, detection, computer-aided diagnosis

1. INTRODUCTION

Lung nodule segmentation is an essential part of the computer-aided detection (CAD) scheme for lung nodule. The accuracy of nodule segmentation has large impact on the performance of a CAD scheme. Many researchers have developed nodule segmentation algorithm,^{1–5} and they wished to improve the performance of nodule detection by improving the performance of nodule segmentation.

Wang *et al.*⁶ used a spiral-scanning technique to convert a 3D image to a generalized 2D polar-coordinate image and then employed a 2D dynamic programming technique to segment the nodule in the converted image. Their segmentation algorithm was tested on two Lung Imaging Database Consor-

tium (LIDC) datasets with 23 and 73 nodules, and the mean overlap values for the two sets were 66% and 64%, respectively. Dehmshki⁷ *et al.* proposed an efficient algorithm for segmenting different types of pulmonary nodules; they performed an adaptive sphericity-oriented contrast region growing algorithm based on the fuzzy connectivity map of the object of interest. The algorithm was tested on 815 nodules, and 84% of the segmentation results were considered acceptable by radiologists. Sahiner⁸ designed a 3D region-growing method and a 3D active contour model for delineation of the boundaries of nodule candidates. They further applied the two segmentation methods to a nodule detection scheme. For nodules without pure ground-glass opacities, their detection scheme reported 2.2, 2.2, and 1.5 false positives per scan at

70% sensitivity by use of the active-contour feature, region-growing feature, and the combined feature. They found that their nodule detection performance could be improved by combining the features from the two different segmentation methods. However, they did not analyze how a segmentation method could impact the performance of a CAD scheme.

It is common knowledge that the performance of a nodule segmentation algorithm would greatly impact the performance of a nodule detection scheme. However, to our knowledge, no study has been done to reveal how a segmentation method would affect the performance of a CAD scheme. For instance, among many segmentation methods, would the “optimal” one lead to the “optimal” performance for nodule detection when other conditions remain unchanged? This is a crucial issue in the field of CAD. We in this study will analyze the relationship between the performances of lung nodule segmentation and lung nodule detection, and will provide in Sec. 5 the guidelines for choosing an appropriate segmentation algorithm with appropriate parameters in a CAD scheme.

2. MATERIAL

The CT scans used in this study were obtained from the standard CT lung nodule database provided by the LIDC.⁹ The database consisted of 85 CT scans with 111 nodules, each of which contained 0–8 nodules. The tube exposure time product was 40–422 mAs, and the tube potential was 120–140 kVp. The slice thickness was 1.25–3 mm, and the inter-slice distance was 0.75–3 mm. Each slice had a matrix size of 512×512 pixels, and the pixel size was 0.5–0.8 mm.

The centers and reference outlines of nodules were established in a two-phase reading process.⁹ First, four radiologists who participated in the LIDC independently identified the locations and outlines of nodules of 3 mm or larger in diameter in each CT scan. After the completion of the first reading, the results were sent back to all of the four radiologists. Then, for each nodule, each radiologist either maintained his or her previous judgment or made a change by consulting the results of other three radiologists. One hundred and eleven nodules were judged as nodules 3 mm or larger by at least two of the four radiologists and were used for evaluating the performance of the nodule segmentation methods and nodule detection schemes.

The inter-rater variability of the radiologists is typically large as described in Ref. 9. For reducing variability in ground truths, four expert radiologists delineated the contour of each nodule. As described by other investigators,^{10–14} we might establish ground truths from contours provided by at least one, two, three, or all four radiologists. However, the ground truths determined by only one radiologist might include many non-nodules because three radiologists think otherwise; on the other hand, those determined by all four radiologists might include only very obvious nodules. We believe that the ground truths determined by at least two or three radiologists appear to be more reasonable, in practice. Because there are more nodules selected by at least two radiologists than by at least three radiologists, we finally used those determined by at least two radiologists.

The 3D reference masks of nodules were firstly created from nodule outlines delineated by radiologists. From the 3D reference mask for each nodule, 24 2D images with 2D reference masks were created from 24 uniformly distributed viewpoints around the 3D nodule candidate, as described below. The 24 2D reference masks were then employed as the 2D reference standards for nodule segmentation. The nodule segmentation was performed and evaluated on the 2D images. The nodule detection was initially performed in the 3D images. The segmentation, feature extraction, and classification were performed in 24 2D images, and finally the classification results of nodules in the 24 2D images were fused to provide nodule detection results for 3D images.

3. METHODS

In this study, a uniform viewpoint reformation technique developed by our group recently was used to detect the lung nodules.¹⁵ The lung region of interest was first segmented from other regions using a thresholding technique, and a pixel with a CT value between −1000 HU and −400 HU was considered to be located inside the lung region. The outline of the lung region was then trimmed in order to include nodules connected to the pleura. A selective enhancement filter¹⁶ was employed to enhance nodules, and a thresholding technique using a threshold of 40 was employed to detect the initial nodule candidates in the nodule-enhanced image. The uniform viewpoint reformation technique was then utilized to convert a 3D volume of interest of a nodule candidate into a set of 24 2D reformatted images, from which nodules could be well distinguished from false positives. Six different segmentation algorithms were employed to segment the nodule candidates in the 2D reformatted images. Thirty-three 2D features were determined from each segmented candidate area based on either a single 2D reformatted image or multiple 2D reformatted images of a nodule candidate. Then, an “optimal” subset of features was selected, which maximized the separation between the nodules and false positives. We finally classified the 2D nodule candidates into nodule and non-nodule based on the “optimal” features by use of a piecewise linear classifier.¹⁷ If at least 17 (70% of 24) of the 24 2D nodule candidates of a 3D nodule candidate were classified as true nodules, the 3D nodule candidate was retained as a true nodule. Otherwise, it was discarded as a false positive. Readers are referred to Ref. 15 for details.

The six segmentation algorithms included the fixed thresholding (FT), Otsu thresholding (OTSU), fuzzy C-means (FCM), Gaussian mixture model (GMM), Chan and Vese model (CV), and local binary fitting (LBF). The six segmentation algorithms led to six similar but different CAD schemes. The methods and the results of the three pre-segmentation steps (lung segmentation, nodule enhancement, and initial nodule detection) are exactly the same for the six CAD schemes. The methods used in the three post-segmentation steps (feature determination, feature selection, classifier training, and application) are also the same for the 6 CAD schemes. However, since the six segmentation methods are

different, the *results* of these postsegmentation steps are different for the 6 CAD schemes.

We employed two metrics to assess the performance of the six segmentation methods. The first is the mean Jaccard index defined as the mean ratio of the intersection to the union between the 2D reformatted area of the 3D reference nodule and that of the automatically segmented nodule. The Jaccard index measures the overall consistency between the two nodule volumes. The value of the Jaccard index ranges from 0%, no overlap between the two volumes, to 100%, a perfect overlap. The second is the mean absolute distance (Dmean) defined as the average distance from all points on the border of the automatically segmented nodule to those of the reference nodule. Dmean assesses the dissimilarity in shape between the borders of the reference nodule and those of the segmented nodule. The smaller the Dmean, the closer the border points of automatically segmented nodule to those of the reference nodule.

Since the number of initially detected candidates was too large (219 720 2D candidates from 9155 3D nodule candidates), we used a revised leave-one-case-out method to evaluate the performance of the six nodule detection CAD schemes. As with the regular leave-one-case-out method, a case was left out as a test case, and other cases were used as the training cases. However, for each 3D nodule candidate in the training cases, four, instead of 24, 2D nodule candidates were used for training the steps of feature selection and classification, and finally the trained features and classifiers were applied to the test case. The number of false positives at three sensitivities (75%, 80%, and 85%) was used as a metric to evaluate the performance of nodule detection scheme.

The statistical significance of differences between the performance of different nodule segmentation methods and different nodule detection methods was analyzed, respectively, by use of the balanced two-way analysis of variance (ANOVA) and the McNemar's test in Sec. 3.G. The six segmentation algorithms will be described briefly in Secs. 3.A–3.F.

3.A. Fixed thresholding

The thresholding technique with a fixed threshold is a common algorithm for image segmentation. Its primary advantage is the simplicity and efficiency. When the object and background pixels have significantly different intensity levels, the thresholding technique is both effective and efficient for the object segmentation. A fixed threshold was chosen empirically for the segmentation of all 2D reformatted images. In this study, we chose different fixed thresholds to obtain different nodule segmentation performance and nodule detection performance.

3.B. Otsu thresholding

Otsu thresholding¹⁸ selects an “optimal” threshold to maximize the between-class variance in gray levels between an object and a background. The procedure is very simple and widely used in image processing society. The algorithm

assumes that the image to be thresholded contains pixels belonging to two classes (object and background), and searches for the optimal threshold leading to the maximum between-class variance (i.e., the greatest separation) between the two classes. In other words, the segmentation with the maximum between-class variance has the minimal classification error between the two classes. There is no adjustable parameter for OTSU method; for a given image, the “optimal” threshold is unique and the segmentation result is also unique.

3.C. Fuzzy C-mean algorithm

Fuzzy C-means¹⁹ segments objects from backgrounds by clustering all pixels into two clusters of object and background. The clustering criterion of FCM is to minimize a cost function based on the least-squared errors within the two clusters. The membership functions of all pixels were randomly initialized with numbers between 0 and 1, and the cluster centers and cost function were initialized accordingly. The membership functions of all pixels and the cluster centers were then iteratively updated by optimizing the cost function. The iteration procedure stops if (1) the number of the iterations was larger than 100 or (2) the absolute difference of the membership functions in two successive iterations was less than 0.1.

The degree of membership of x_j in the i th is

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left[\frac{d_{ij}}{d_{kj}} \right]^{2/(m-1)}},$$

where d_{ij} is a distance measure between object x_j and cluster center v_i , C is the number of clusters, and m is employed to adjust the degree of membership of the object to the cluster.

In this study, we selected different indices m of the membership function to achieve different levels of segmentation performance.

3.D. Gaussian mixture model

Gaussian mixture model²⁰ segments nodules by fitting the gray level distributions of object and background regions in the image with multiple Gaussian models, each of which represents an object or background region. The model parameters were estimated with a maximum likelihood estimation method. Once the model parameters were determined, the object and background regions were segmented. The initial standard deviation of Gaussian function was set with the range of gray value in the image in order to prevent fitting a very narrow Gaussian model to the gray level distribution. In this study, we selected 2 to 4 Gaussian models to achieve different levels of segmentation performance.

3.E. Chan and Vese model

CV (Ref. 21) model segments objects in a given image by minimizing an energy function with a curve evolution technique based on Mumford-Shah functional and level sets. It can segment objects whose boundaries are not necessarily

clearly defined by edge gradient. The object curve was initialized with a circle located at the center of the image ROI. During the curve evolution, finite difference method is employed to iteratively solve the numerical solution of level sets. In this study, we selected different radii of the initial circles, the coefficients of the closed curves in the energy function, and the maximum allowed numbers of iterations to achieve different levels of segmentation performance.

3.F. Local binary fitting

The local binary fitting algorithm is based on an assumption that an image region can be approximated locally by a binary image.^{22,23} The LBF method incorporates local image intensity information into a region-based active contour model by introducing a local binary fitting energy in an energy variational formulation. Its advantages include good segmentation for images with intensity inhomogeneity. The object region curve was initialized with a circle located at the center of the image ROI. In this study, the finite difference method was employed to solve the numerical solution of the curve evolution. Different radii of the initial circles and coefficients of the closed curve in the energy function were selected to achieve different levels of segmentation performance.

3.G. Statistical analysis

We employed the balanced ANOVA (Ref. 24) to test if a segmentation method is statistically more accurate than another one in terms of a segmentation metric (mean Jaccard index or Dmean). In the two-way ANOVA, the two factors indicate, respectively, the two segmentation methods (degree of freedom = 1) and the 24 measurement values of 2D segmentation accuracy for a 3D nodule (degree of freedom = 23); however, we are only interested in the factor of two different segmentation methods, rather than the 24 measurements for a nodule.

We employed McNemar's test to see if a detection CAD scheme is statistically more accurate than another one in terms of the number of false positives per scan at three fixed sensitivities of 85%, 80%, and 75%. Please see Ref. 25 for details.

Because there were 15 possible pairs between six different segmentation methods and 15 pairs between 6 CAD schemes, we employed Bonferroni correction to lower the p value threshold to 0.0033 ($=0.05/15$). A p value of 0.0033 or less implies that there is statistically significant difference between two segmentation/CAD methods.

4. RESULTS

4.A. Segmentation results by use of different segmentation algorithms

In this study, the algorithms of FT, OTSU, FCM, GMM, CV, and LBF were employed to segment the nodule candidates in the 2D reformatted images. Figure 1(a) indicates the original image of a nodule, and Figs. 1(b)–1(g) show the segmentation results by use of FT, OTSU, FCM, GMM, CV, and

LBF, respectively. Only 8 of 24 2D reformatted images of a 3D nodule are shown in Fig. 1 for clarity.

4.B. Performances of segmentation and detection by use of different segmentation algorithms

Table I shows the performance of different nodule segmentation methods and the performance of corresponding detection methods. The performance of each segmentation method shown here was its “optimal” performance we could achieve by adjusting its parameters. Figure 2 shows part of the empirical FROC curves of the six nodule detection methods based on the “optimal” performances of the six different segmentation methods. Because the number of nodules was relatively small, the FROC curves look a little less smooth.

It is apparent from Table I that when multiple segmentation algorithms are available for selection, the one with the highest segmentation performance did not necessarily lead to the highest detection performance. For instance, the fixed thresholding segmentation method achieved a higher segmentation performance than the FCM method in terms of both segmentation metrics (overlap: $p < 0.0001$; Dmean: $p = 0.0243$); however, the CAD scheme using the fixed thresholding method achieved a lower detection performance than that using the FCM method at all three detection sensitivities of 85%, 80%, and 75% (all $p < 0.0001$). Therefore, one should select a segmentation method based on the detection performance rather than the segmentation performance, because the segmentation performance alone is not adequate for predicting the detection performance.

Although there is no statistically significant difference in segmentation results in terms of Dmean, its p value of 0.0243 is still considered a small value. Overall, we cautiously conclude that there is statistically significant difference in segmentation results between the FT and FCM methods by use of the two metrics.

4.C. Performances of segmentation and detection by use of different parameters of the same segmentation algorithm

For the segmentation algorithms FT, FCM, GMM, CV, and LBF, we adjusted their parameters to obtain different segmentation performances and determined the corresponding detection performances. OTSU segmentation method does not have parameter; therefore, its segmentation performance and corresponding detection performance are fixed. Figures 3–7 show the performances of nodule detection at three segmentation performances for the five segmentation methods. When the segmentation performances were improved by adjusting parameters, the detection performances were typically improved as well. Although this statement might not be correct for some part of the FROC curves in Figs. 3–7, the general trend of this statement would be correct.

5. DISCUSSION

The performances of segmentation methods are typically evaluated by use of a set of true nodules in a database. The

FIG. 1. The segmentation results of a nodule by using different methods. (a) Eight reformatted images of a nodule and the segmentation result by using (b) fixed thresholding, (c) Otsu thresholding, (d) fuzzy C-means, (e) Gaussian mixture model, (f) Chan and Vese model, and (g) local binary fitting.

non-nodules are generally ignored and discarded because it is impractical to evaluate how well a segmentation method segments a non-nodule. In fact, it is impractical to create the reference standards of the contours of non-nodules for evaluating the segmentation method. Therefore, when a segmentation method is considered as a “good” method, one can only

confidently state that the method is good for segmenting true nodules.

However, for evaluating the performance of a detection scheme, one must use not only true nodules but also non-nodules. Obviously, the sets of objects used for evaluating the segmentation and detection methods are different. Therefore,

TABLE I. Performance of nodule segmentation and detection by use of different segmentation algorithms. The bold-face numbers indicate the highest segmentation and detection performance levels among six methods.

	Performance of segmentation		Performance of detection		
	Jaccard index	Dmean (pixels)	FPS/scan (Sensitivity = 85%)	FPS/scan (Sensitivity = 80%)	FPS/scan (Sensitivity = 75%)
FT	0.601^a	1.74^b	21.5 ^c	4.4 ^d	2.1 ^e
OTSU	0.586	1.80	18.9	8.8	2.4
FCM	0.588 ^a	2.32 ^b	9.8^c	3.4^d	1.5^e
GMM	0.563	2.80	19.9	9.2	4.9
CV	0.543	3.48	43.7	13.6	3.8
LBF	0.553	3.18	43.3	10.4	7.8

^a $p < 0.0001$ using two-way ANOVA test.

^b $p = 0.0243$ using two-way ANOVA test.

^c $p < 0.0001$ using McNemar's test.

^d $p < 0.0001$ using McNemar's test.

^e $p < 0.0001$ using McNemar's test.

FIG. 2. FROC curves for the different algorithms with different Jaccard index values.

FIG. 5. Empirical FROC curves for the GMM algorithm with different Jaccard index values.

FIG. 3. Empirical FROC curves for the threshold algorithm with different Jaccard index values.

FIG. 6. Empirical FROC curves for the CV algorithm with different Jaccard index values.

FIG. 4. Empirical FROC curves for the FCM algorithm with different Jaccard index values.

FIG. 7. Empirical FROC curves for the LBF algorithm with different Jaccard index values.

if a segmentation method is to be used in a detection scheme, it is inadequate to evaluating only the performance of segmentation method based on the true nodules; it is more important to evaluate the performance of detection scheme based on both nodules and false positives.

The above concept has large impact on the development of CAD scheme. Researchers usually select from multiple segmentation methods the “optimal” one for segmenting true nodules, and also consider it the “optimal” one for a nodule detection CAD scheme. We have shown that the “optimal” segmentation method may not be the “optimal” one for the detection CAD scheme; other sub-optimal segmentation methods may lead to better performances for a detection scheme than the “optimal” segmentation method.

We employed in this study the uniform viewpoint reformation technique to transform a 3D nodule image into 24 reformatted 2D nodule images and then analyzed the relationship between the performances of nodule segmentation and nodule detection. The primary advantage of employing 2D nodule images was that the 2D segmentation algorithms were faster and simpler. In addition, the number of true nodules was increased considerably by transforming a single 3D nodule into 24 (though correlated) 2D nodules in the reformatted 2D images. A CAD scheme trained with more samples should become more robust.

Although much of the detection schemes in this study is very similar to that in Ref. 15, for simplicity, some of the steps in Ref. 15 were not employed in this study, including the two rules used for removing some obvious false positives. Therefore, the detection performances in this study were slightly different from that in Ref. 15.

Our conclusion was drawn for nodule segmentation and detection in CT scan. It is not verified for segmentation and detection of other lesions in other medical images. However, based on our experience, we feel that the conclusion would still be applicable to other lesions in other medical images.

We used ANOVA and McNemar methods to test the significance of difference in performance between two segmentation methods and two CAD schemes, respectively. These statistical test methods require that the input data be independent. In this study, the initial 9155 nodule candidates come from 85 CT cases. Therefore, the requirement of independence among them may not be met, which is a limitation of using the two-way ANOVA and McNemar tests for statistical analysis in our study.

6. CONCLUSION

When multiple algorithms are available for segmenting nodule candidates in a CAD scheme, the “optimal” segmentation algorithm did not necessarily lead to the “optimal” CAD detection performance.

ACKNOWLEDGMENTS

This work was supported in part by USPHS Grant No. R01 CA113870, and W.G. is supported in part by Shenyang

Science and Technology Foundation No. F13-316-1-35, the Ph.D. startup Fund of National Science Foundation of Liaoning Province No. 20131086, the Ph.D. startup Fund of SAU (No. 13YB16), National Aerospace science Foundation (No. 2103ZE54025), and National Natural Science Foundation of China (Nos. 61373088 and 61300233). CAD technologies developed by Qiang Li and his colleagues have been licensed to companies including Hologic, Inc., Riverain Medical Group, Median Technology, Mitsubishi Space Software Co., General Electric Corporation, and Toshiba Corporation.

^{a)} Author to whom correspondence should be addressed. Electronic mail: liqiang@sari.ac.cn

¹ A. Reeves, A. Chan, D. Yankelevitz, C. Henschke, B. Kressler, and W. Kostis, “On measuring the change in size of pulmonary nodules,” *IEEE Trans. Med. Imaging* **25**, 435–450 (2006).

² X. Ye, M. Siddique, A. Douiri, G. Beddoe, and G. Slabaugh, “Image segmentation using joint spatial-intensity-shape features: application to CT lung nodule segmentation,” *Proc. SPIE* **7259**, 72594V.1–72594V.9 (2009).

³ T. Way, L. Hadjiiski, B. Sahiner, H. Chan, P. Cascade, E. Kazerooni, N. Bogot, and C. Zhou, “Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours,” *Med. Phys.* **33**, 2323–2337 (2006).

⁴ W. Mullally, M. Betke, J. Wang, and J. P. Ko, “Segmentation of nodules on chest computed tomography for growth assessment,” *Med. Phys.* **31**, 839–848 (2004).

⁵ W. J. Kostis, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke, “Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images,” *IEEE Trans. Med. Imaging* **22**, 1259–1274 (2003).

⁶ J. Wang, R. Engelmann, and Q. Li, “Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique,” *Med. Phys.* **34**, 4678–4689 (2007).

⁷ J. Dehmshki, H. Amin, M. Valdivieso, and X. Ye, “Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach,” *IEEE Trans. Med. Imaging* **27**, 467–480 (2008).

⁸ B. Sahiner, L. M. Hadjiiski, H. Chan, J. Shi, T. Way, P. N. Cascade, E. A. Kazerooni, C. Zhou, and J. Wei, “The effect of nodule segmentation on the accuracy of computerized lung nodule detection on CT scans: Comparison on a data set annotated by multiple radiologists,” *Proc. SPIE* **6514**, 65140L.1–65140L.7 (2007).

⁹ M. F. McNitt-Gray, S. G. Armato III, C. R. Meyer, A. P. Reeves, G. McLennan, R. C. Pais, J. Frymann, M. S. Brown, R. M. Engelmann, P. H. Bland, G. E. Laderach, C. Piker, J. Guo, Z. Towfic, D. Qiang, D. Yankelevitz, R. Aberle, E. J. Van Beek, H. Macmahon, E. A. Kazerooni, B. Y. Croft, and L. P. Clarke, “The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation,” *Acad. Radiol.* **14**, 1464–1474 (2007).

¹⁰ J. Wang, F. Li, and Q. Li, “Automated segmentation of lungs with severe interstitial lung disease in CT,” *Med. Phys.* **36**(10), 4592–4599 (2009).

¹¹ J. J. Suárez-Cuenca, W. Guo, and Q. Li, “Automated detection of pulmonary nodules in CT: False positive reduction by combining multiple classifiers,” *SPIE Med. Imaging, Comput.-Aided Diagnos.* **7963**, 796338 (2011).

¹² T. Messay, R. C. Hardie, and S. K. Rogers, “A new computationally efficient CAD system for pulmonary nodule detection in CT imagery,” *Med. Image Anal.* **14**(3), 390–406 (2010).

¹³ A. Riccardi, T. S. Petkov, G. Ferri, M. Masotti, and R. Campanini, “Computer-aided detection of lung nodules via 3D fast radial transform, scale space representation, and Zernike MIP classification,” *Med. Phys.* **38**, 1962–1971 (2011).

¹⁴ S. Ozekesand and O. Osman, “Computerized lung nodule detection using 3D feature extraction and learning based algorithms,” *J. Med. Syst.* **34**(2), 185–194 (2010).

¹⁵ W. Guo and Q. Li, “High performance lung nodule detection schemes in CT using local and global information,” *Med. Phys.* **39**(8), 5157–5168 (2012).

- ¹⁶Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans," *Med. Phys.* **30**, 2040–2051 (2003).
- ¹⁷Q. Li, F. Li, and K. Doi, "Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier," *Acad. Radiol.* **15**, 165–175 (2008).
- ¹⁸N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cyber.* **9**, 62–66 (1979).
- ¹⁹J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum Press, New York, 1981), pp. 310–350.
- ²⁰D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Proc.* **10**, 19–41 (2000).
- ²¹T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Proc.* **10**, 266–277 (2001).
- ²²C. Li, C. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Trans. Image Proc.* **17**, 1940–1949 (2008).
- ²³C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: A new variational formulation," *Comput. Ver. Pattern Recogn.* **1**, 430–436 (2005).
- ²⁴M. Lynch and B. Walsh, *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, 1998).
- ²⁵P. A. Lachenbruch and C. J. Lynch, "Assessing screening tests: Extensions of McNemar's test," *Stat. Med.* **17**(19), 2207–2217 (1998).