

Computer-aided detection of lung nodules: False positive reduction using a 3D gradient field method and 3D ellipsoid fitting

Zhanyu Ge, Berkman Sahiner,^{a)} Heang-Ping Chan, Lubomir M. Hadjiiski, Philip N. Cascade, Naama Bogot, Ella A. Kazerooni, Jun Wei, and Chuan Zhou
Department of Radiology, The University of Michigan, Ann Arbor, Michigan 48109

(Received 16 November 2004; revised 23 March 2005; accepted for publication 9 May 2005; published 12 July 2005)

We are developing a computer-aided detection system to assist radiologists in the detection of lung nodules on thoracic computed tomography (CT) images. The purpose of this study was to improve the false-positive (FP) reduction stage of our algorithm by developing features that extract three-dimensional (3D) shape information from volumes of interest identified in the prescreening stage. We formulated 3D gradient field descriptors, and derived 19 gradient field features from their statistics. Six ellipsoid features were obtained by computing the lengths and the length ratios of the principal axes of an ellipsoid fitted to a segmented object. Both the gradient field features and the ellipsoid features were designed to distinguish spherical objects such as lung nodules from elongated objects such as vessels. The FP reduction performance in this new 25-dimensional feature space was compared to the performance in a 19-dimensional space that consisted of features extracted using previously developed methods. The performance in the 44-dimensional combined feature space was also evaluated. Linear discriminant analysis with stepwise feature selection was used for classification. The parameters used for feature selection were optimized using the simplex algorithm. Training and testing were performed using a leave-one-patient-out scheme. The FP reduction performances in different feature spaces were evaluated by using the area A_z under the receiver operating characteristic curve and the number of FPs per CT section at a given sensitivity as accuracy measures. Our data set consisted of 82 CT scans (3551 axial sections) from 56 patients with section thickness ranging from 1.0 to 2.5 mm. Our prescreening algorithm detected 111 of the 116 solid nodules (nodule size: 3.0–30.6 mm) marked by experienced thoracic radiologists. The test A_z values were 0.95 ± 0.01 , 0.88 ± 0.02 , and 0.94 ± 0.01 in the new, previous, and combined feature spaces, respectively. The number of FPs per section at 80% sensitivity in these three feature spaces were 0.37, 1.61, and 0.34, respectively. The improvement in the test A_z with the 25 new features was statistically significant ($p < 0.0001$) compared to that with the previous 19 features alone. © 2005 American Association of Physicists in Medicine. [DOI: 10.1118/1.1944667]

Key words: computer-aided detection, false positive reduction, gradient field technique, lung nodule detection, ellipsoid fitting

I. INTRODUCTION

Lung cancer is the leading cause of cancer deaths for both men and women in the United States. Studies indicate that patients treated for stage I lung cancer have better survival than patients presenting with more advanced stage disease.¹ With the rapid growth of data volumes in computed tomography (CT) imaging and the potential of applying CT for lung cancer screening, the interpretation of thoracic CT scans is becoming more challenging for radiologists. The challenge arises not only because of the increasing amount of data, but also because of the complex anatomical structures (lungs, soft tissues, vessels, airways) in the thorax. Computer-aided detection (CAD) systems^{2–10} can play an important role in mitigating the burden of radiologists by alerting them to suspicious lesions. Although much effort has been devoted to it, the development of CAD systems for lung nodule detection on CT scans remains a difficult and ongoing task.

In a lung nodule CAD system, lower threshold values are usually used during the prescreening stage to achieve high

sensitivity at the expense of a large number of false positives (FPs). The FPs are then analyzed and reduced by feature extraction and classification techniques to increase the specificity. Ideally, the FP reduction stage should eliminate as many FPs as possible without affecting nodule detection sensitivity. A partial summary of the features used for FP reduction can be found in the literature.¹¹ Many of these features are variations of standard techniques that are found in the image processing literature. We are developing feature extraction and classification techniques for distinguishing true and false nodules. Lung nodules vary in shape, size, and location within the lungs. Except for spiculated and ill-defined nodules, the shapes of lung nodules that are not attached to the pleural surface are mostly spherical, while those that are attached to the pleural surface (juxta-pleural nodules) are hemi-spherical. Pulmonary blood vessels are elongated and more cylindrical. A solitary lung nodule is defined as a single round intra-parenchymal opacity, at least moderately well-margined and less than 3 cm in maximum

diameter.¹² Many studies have sought to use the shape difference between round nodules and elongated vessels to reduce FPs in CAD for lung nodule detection.

Armato *et al.*^{7,13} used an automated nodule detection method based on two- and three-dimensional [(2D) and (3D)] analysis of CT image data. Lung segmentation was performed on a 2D section-by-section basis to construct a lung volume. Lung nodule candidates were identified using multiple gray-level thresholding of the lung volume, grouping voxels according to 18-connectivity, and thresholding using a volume criterion. For each of the nodule candidates, they derived 9 features to differentiate true positives from false positives, including circularity, sphericity, and compactness. Rule-based and LDA classifiers were used to reduce false positives. Their method was applied to 43 scans totaling 1209 axial sections and containing 171 identified nodules. The automated method yielded an overall nodule detection sensitivity of 70%, with an average of 1.5 FPs per section. In a more recent study,¹⁴ they reported an improved performance of 80% sensitivity at 1.0 FPs per section. Ko *et al.*⁵ analyzed nodule candidates by location and shape to differentiate normal structures from nodules. Candidates were categorized into five regions of the lungs based on location information, and shape information was characterized by circularity of the candidate nodules. The algorithm, tested on 16 chest CT scans containing 370 nodules, achieved a sensitivity of 86% at 2.3 FPs per section. Lee *et al.*⁶ used a template matching technique to detect lung nodules on chest CT scans. A genetic algorithm was designed to determine the target position and to select a template image from the reference patterns. The four reference templates were established according to the gray-level values of 3D Gaussian distributions, with values in the z direction (the vertical direction) regulated by a factor. The algorithm worked only on three slices in the z direction to detect nodules less than 3 cm in diameter. The matching was determined by a reference volume and the normalized gray-scale correlation of the candidate region. Thirteen features were extracted to eliminate FPs. They applied their algorithms to 557 axial sections from 20 scans, obtaining a sensitivity of 72% at 1.1 FPs per section. Kanazawa *et al.*¹⁰ developed a rule-based CAD system using features such as size, circularity, contrast, convexness and roundness, and applied it to a data set of 450 chest CT scans containing 230 lung nodules. Their system detected 90% of nodules characterized as definitely malignant or suspicious by three radiologists at 8.6 FPs per scan. The same system was later applied to 249 different scans in a field test, detecting 34 of the 47 nodules (72%) characterized as suspicious for malignancy by three radiologists, and 10 of the 14 nodules (71%) characterized as benign.¹⁵ All of the above-discussed studies analyzed the shape information of the extracted nodule candidates to discriminate the lung nodules from normal structures through compactness or circularity on a stack of 2D sections. Li *et al.*¹⁶ proposed three selective enhancement filters for dots, lines, and planes, which can simultaneously enhance objects of a specific shape and suppress objects of other shapes. They blurred the CT image with a Gaussian kernel that matched the size of the nodule to

be detected before calculating the eigenvalues of the Hessian matrix that were used for selective enhancement. They used multiple scales of the Gaussian kernel to find a match with the nodule size. The prescreening algorithm was applied to 73 scans and detected 71 of 76 lung nodules with an average of 4.2 FPs per section. Brown *et al.*⁹ used a data set of 14 chest CT scans obtained using 1 mm collimation for the development of a system to detect micronodules (nodules with a diameter of less than or equal to 3 mm), and tested their system on 15 different scans. They achieved 100% sensitivity for lung nodules larger than 3 mm, and 70% sensitivity for lung nodules less than or equal to 3 mm at 15 FPs per scan.

The computer-aided detection system developed in our laboratory² consists of the following steps: lung volume segmentation, lung partitioning and sectioning, lung nodule candidate detection and segmentation, volume of interest (VOI) extraction, feature extraction, and feature classification. We refer to the first three steps as the adaptive prescreening stage, and the last three steps as the FP reduction stage. The 3D gradient field method and the ellipsoid fitting developed in this study are used to extract shape features in the FP reduction stage.

The gradient field information has been used to study the distribution of intensity fields or shape information of objects in physics, computer vision, and many other applications. Recent work on image gradient analysis has demonstrated its potential in different areas of CAD such as breast tumor detection and colonic polyp detection and FP reduction.¹⁷ These investigators extracted gradient field information from individual 2D images or from a 3D data set and detected objects by using the maximum gradient convergence within a predefined range along the radial direction. Our 3D gradient field method aims at exploiting the shape information of the identified objects in the 3D image gradient field to distinguish nodules from other tissues in the lungs, and hence reduce FPs and increase specificity for automated lung nodule detection.

Ellipsoid fitting has been used for feature extraction in a number of applications in medical imaging, visualization, and pattern recognition.^{18–20} We applied 3D ellipsoid fitting to the binary objects extracted during the prescreening stage, and used the lengths and length ratios of the principal axes as discriminant features for classification of spherical and elongated objects.

In our previous work, we used 19 features related to the shape of the detected objects and the gray-scale distribution within the objects for FP reduction.² In this study, we extracted a total of 25 new features based on gradient field analysis and ellipsoid fitting for the same purpose. We compared the FP reduction performance in the new and previous feature spaces. Linear discriminant analysis with stepwise feature selection was used for distinguishing true lung nodules from FPs. The classification performance was measured using the area A_z under the receiver operating characteristic (ROC) curve and the overall detection accuracy was evaluated by free-response receiver operating characteristic (FROC) analysis.

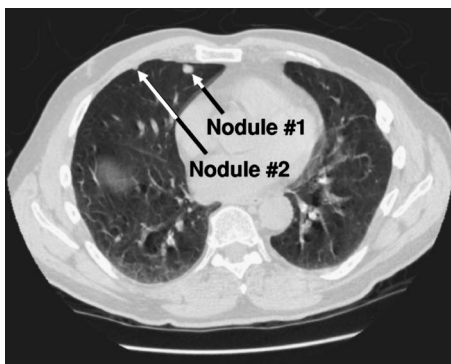


FIG. 1. A CT slice containing significant dependent atelectasis in our secondary data set. A subtle juxta-pleural nodule (nodule 2) that was detected by the CAD program is at the upper chest wall, in addition to the 8.5 mm lobulated nodule (nodule 1) located medial to it.

II. METHODS

A. Data sets

The chest CT scans used in this study were collected with approval from the Institutional Review Board. A total of 91 chest CT scans in 63 patients were selected from the patient archives. The CT scans were randomly selected except for two criteria: the slice thickness was 2.5 mm or less and the CT scan contained no more than six lung nodules. The data set was further separated into a main data set and a secondary data set. The main data set consisted of 82 chest CT scans from 56 patients with lung nodules, whose lungs were free of other parenchymal abnormality that focally or diffusely increased lung attenuation, such as interstitial or alveolar disease. The secondary data set consisted of the remaining 9 chest CT scans from 9 patients that contained areas of visible dependent atelectasis²¹ in addition to the lung nodules. These cases, referred to as “atelectasis cases” in the following discussion, were separated from the main data set based on the judgment of experienced thoracic radiologists that they contained sufficient dependent atelectasis to create conspicuous ground glass opacity on the CT images. An example of an axial section from a chest CT case with dependent atelectasis is shown in Fig. 1. This axial section contains two nodules, one of which is a subtle juxta-pleural nodule, and the other is an 8.5 mm internal nodule.

For the chest CT scans in our data set, all 2D image sections had a matrix size of 512×512 pixels. The in-plane resolution ranged from 0.546 to 0.839 mm, with an average of 0.674 mm. The section thickness ranged from 1.0 to 2.5 mm. Many of the scans in our data set only contained a partial volume of the lungs. Although the use of full lung scans is preferable, we decided not to exclude the partial scans because it is important to train and test the developed algorithms with as large a data set as possible. Since we will report the sensitivity for individual nodules and the FP rate as the number of FPs per section, the partial scan should provide similar performance statistics as a full thoracic CT scan. In this study, we used the main data set for training and testing of a classifier in a leave-one-patient-out resampling

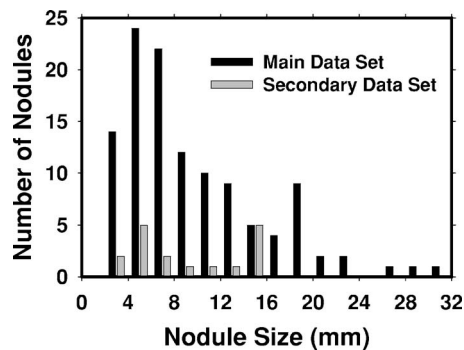


FIG. 2. Distribution of nodule sizes for the main and secondary data sets.

scheme. The trained classifier was also applied to the secondary data set for evaluation of its performance in the atelectasis cases and for comparison with that in the main data set.

Three experienced thoracic radiologists read different subsets of our CT database using a graphical user interface (GUI) specifically designed in our laboratory for reading the CT scans and collecting data of the nodule characteristics. The GUI was developed in collaboration with the radiologists such that nodule characteristics and their descriptors of clinical relevance were included. The radiologists marked the location of the lung nodules, measured the long and short-axis lengths of the nodules on the slice containing the largest axial cross section of the nodule using an electronic ruler, and rated the conspicuity of the nodules relative to those encountered in clinical practice on a 5-point scale, with 1 representing the most obvious and 5 representing the subtlest nodules. The radiologists also rated the nodule margins as smooth, lobulated, or spiculated/irregular, and identified nodules that were juxta-pleural or juxta-vascular.

The radiologists marked 116 solid nodules on the 3551 sections in the main data set; nodule size ranged from 3.0 to 30.6 mm (median=7.8 mm). The average conspicuity of the nodules was 2.4. Nineteen lung nodules in the main data set were biopsy-proven malignancy, and 90 were benign by either biopsy or two-year follow-up showing lack of change or disappearance of the nodule. For seven nodules in the main data set, neither biopsy nor two-year follow-up information was available to ascertain the nodule status as ma-

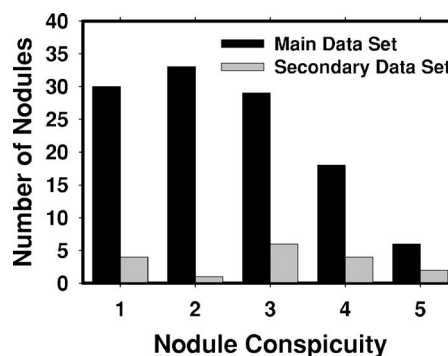


FIG. 3. Distribution of nodule conspicuity (1=Most conspicuous, 5=Least conspicuous) for the main and secondary data sets.

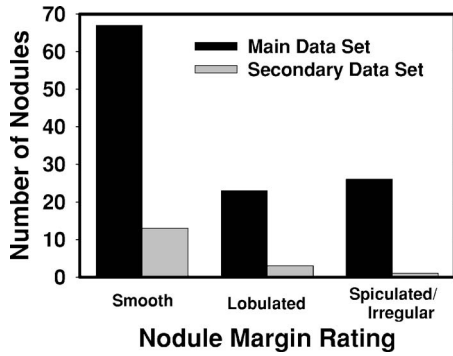


FIG. 4. Distribution of nodule margin ratings for the main and secondary data sets.

lignant or benign. In the secondary data set, the radiologists marked 17 lung nodules on the 581 sections, of which 4 were malignant, 10 were benign, and 3 had insufficient information to ascertain the nodule status. The nodule size ranged from 3.9 to 15.6 mm (median=7.8 mm). The average conspicuity of the nodules was 3.2. Figures 2–4 show the distributions of the nodule size, conspicuity ratings, and margin ratings for the main and secondary data sets.

B. 3D gradient field features

A block diagram of our CAD system, which consists of the prescreening and FP reduction stages, is illustrated in Fig. 5. The prescreening algorithm segmented suspicious objects within the lungs based on weighted *k*-means clustering. In our previous work, 19 features related to the shape of a detected object and gray-scale distribution within the object

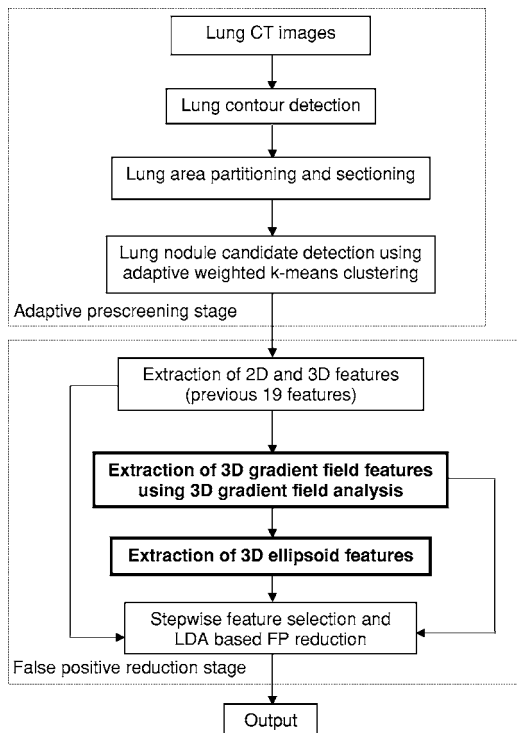


FIG. 5. Schematic illustration of our CT lung nodule CAD system.

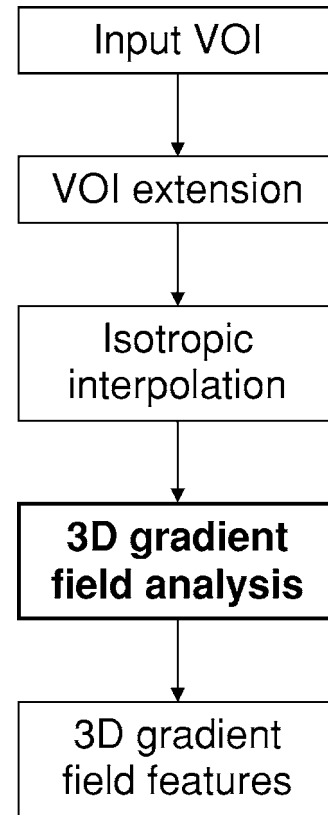


FIG. 6. Schematic illustration of our 3D gradient field analysis algorithm.

were extracted for FP reduction. The details of our prescreening algorithm and the definition of these features can be found in the literature.²

The main focuses of this study, extraction of 3D gradient field features and extraction of 3D ellipsoid features, are shown in boldface in Fig. 5. A block diagram of the 3D gradient field algorithm is shown in Fig. 6. The 3D gradient field algorithm used the VOIs extracted after the prescreening stage as its input. Since the identified VOI might not include the entire nodule, it was first extended in all directions by a fixed amount (5 mm in this study). A 3D isotropic interpolation algorithm was applied to each VOI to generate voxels with equal side-lengths in all three dimensions. The interpolation does not improve the spatial resolution in the *z* direction, but the isometric voxels facilitate the implementation of the 3D gradient field calculation and other image processing operations in the CAD system. 3D gradient field image data were obtained by filtering each of the VOIs with three $3 \times 3 \times 3$ convolution kernels, one for each of the *x*, *y*, *z* directions. The kernel for extracting the *z*-direction gradient is shown in Fig. 7, where the solid disc represents the voxel under consideration and the circles represent the kernel voxels. The 3D kernel coefficients were inversely proportional to a power *m* of the distance between the voxel of interest and the voxel at a given location on the kernel, which is a generalization of the isotropic 2D kernel suggested by Jain.²² The kernel coefficients for kernel voxels 0 to 17 were determined by

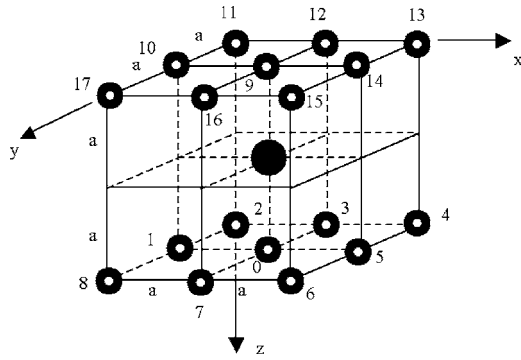


FIG. 7. 3D gradient kernel for the z-direction gradient calculation. A similar kernel was used for the x direction and y direction by rotating the two planes of nonzero weights to be perpendicular to the x axis and the y axis, respectively.

$$w_i = \frac{1}{d_i^m} / \sum_{j=0}^8 \frac{1}{d_j^m}, \tag{1}$$

$$d_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2}, \quad (i = 0, 1, \dots, 8),$$

$$w_i = -w_{i-9}, \quad (i = 9, 10, \dots, 17), \tag{2}$$

where, d_i =the distance between the i th kernel voxel (x_i, y_i, z_i) and the center voxel (x_0, y_0, z_0) , and m =power index of the distance, chosen as $m=1$ in this study. The gradient component in the z direction can be expressed as

$$G_z = \sum_{i=0}^{17} V_i w_i, \tag{3}$$

where, V_i =gray-scale value at kernel voxel i . The gradient components in the x and y directions, G_x and G_y , can be calculated similarly.

The main purpose of the gradient features is to discriminate objects whose gray-level distribution is approximately radially symmetric from other objects whose gray-level distribution are highly asymmetric. For example, consider the standard deviation of the gradient magnitude calculated over points on a spherical surface centered at the centroid of two objects, of which the first object represents an idealized nodule, and the second object approximates a vessel. If the gray levels of the first object are radially symmetric, then the standard deviation of the radial gradients would be zero. Assume that the second object is derived from the first object by rescaling its axes dramatically so that the second object looks like a stretched ellipsoid. The standard deviation for the second object would be substantially different from zero. In this idealized situation in a continuous space, the standard deviations are calculated over all the points on the surface of a sphere centered at the centroid of the object. In a 3D CT volume with discrete voxels, it is desirable to consider a large number of voxels uniformly distributed over the surface of the sphere in order to capture the possible deviation from symmetry. In this study, taking into consideration the

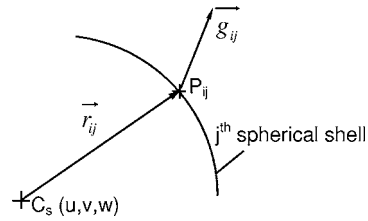


FIG. 8. i th radial vector on the j th spherical shell, and the gradient field vector.

tradeoff between computation time and adequate sampling, we chose to consider the statistics of the gradients over 26 voxels uniformly distributed over a spherical shell. The 26 directions would cover all neighboring voxels for a nodule about 3×3 voxels in size. The sampling became more sparse as the nodule size increased but it would still sample the gradient field in all directions.

To estimate the 3D gradient field vectors at voxels within the VOI segmented by our adaptive prescreening algorithm, we first calculated R_{eq} , the radius of the equivalent sphere, which has the same volume as the nodule candidate, V_{obj} , using

$$R_{\text{eq}} = \sqrt[3]{\frac{0.75V_{\text{obj}}}{\pi}} = 0.62\sqrt[3]{V_{\text{obj}}}. \tag{4}$$

To calculate the gradient field features at a voxel C_s , three spherical shells of radii R_{eq} and $R_{\text{eq}} \pm \Delta R$ were drawn centered at the voxel C_s , where ΔR =radial distance between adjacent shells. The use of three shells instead of just one shell at R_{eq} was intended to decrease the effect of the potential error in object volume estimation from the prescreening stage. The radial distance between adjacent shells was chosen as $\Delta R=0.20R_{\text{eq}}$ in this study so that it was adaptive to the nodule size. On each of the three shells, $j=0, 1, 2$, we used 26 uniformly distributed voxels P_{ij} ($i=0, 1, 2, \dots, 25$) to compute gradient statistics. An example of the i th radial vector, \vec{r}_{ij} , radiating from the center voxel, C_s , to the i th voxel on the j th shell is shown in Fig. 8. The gradient field orientation at voxel i on shell j , SD_{ij} , was defined as the cosine of the angle between the i th radial vector, \vec{r}_{ij} , and the gradient field vector, \vec{g}_{ij} , as shown in Fig. 8,

$$SD_{ij} = \frac{\vec{r}_{ij} \cdot \vec{g}_{ij}}{|\vec{r}_{ij}| |\vec{g}_{ij}|}, \quad i = 0, 1, 2, \dots, 25, \quad j = 0, 1, 2, \tag{5}$$

where the center dot denotes the inner product, and $|\vec{r}_{ij}|$ and $|\vec{g}_{ij}|$ are the magnitudes of the vectors \vec{r}_{ij} and \vec{g}_{ij} , respectively. The gradient field strength, SM_{ij} , is the magnitude of \vec{g}_{ij} , i.e.,

$$SM_{ij} = |\vec{g}_{ij}|. \tag{6}$$

A total of 19 gradient field features were extracted from the statistics of the $\{SD_{ij}\}$ and the $\{SM_{ij}\}$ on three shells and along the 26 radial directions to describe the shape information within the VOI containing a segmented object. Seven of these were extracted from the gradient field strength, and 12 were extracted from the gradient field orientation.

Among the seven gradient field strength features, six were obtained based on shells, and one along the radial direction. The six shell-based gradient field strength features were the average (GM_{av}), standard deviation (GM_{std}), coefficient of variation (GM_{cv}), maximum value (GM_{max}), minimum value (GM_{min}), and the ratio of standard deviation to median value ($GM_{std/med}$) of the gradient field strength. To define these features, we first employed the common definitions of average, standard deviation, etc., on one shell for one voxel in the VOI, and then combined the information from multiple shells and different voxels within the VOI. For example, the (GM_{std}) feature was defined as

$$GM_{std} = \min_{v \in U} \left\{ \min_{j \in J} \left\{ \sqrt{\frac{\sum_{i=0}^{K-1} (SM_{ij} - \sum_{i=0}^{K-1} SM_{ij}/K)^2}{K-1}} \right\} \right\}, \tag{7}$$

where v =the voxel under consideration; U =a collection of all the voxels within the VOI; J =a collection of the three shells centered at v ; and K =the number of voxels on the shell that were inside the lung volume, $K \leq 26$. As mentioned previously, for an ideal spherical nodule, the standard deviation of the gradient magnitude over the voxels P_{ij} is zero for any j , i.e., any radius. However, for a real nodule, the standard deviation will depend on the radius. To capture the standard deviation for the radius at which the gray-level values of the object were most similar to a radially symmetric object, we first computed the minimum over the three spherical

shells centered at voxel v , as indicated by the inner minimum operator in Eq. (7). The outer minimum operator in Eq. (7) took into account the fact that the standard deviation will be zero only if the shells are centered at the centroid of an ideal nodule. Since the centroid of the segmented object may not be the center of symmetry of an idealized object, and since true nodules may not have a center of symmetry at all, we computed the minimum over all pixels within the VOI to find the standard deviation for the voxel that most resembled a center of symmetry. Similar definitions were used to find the other five gradient field strength features that were based on shells.

In addition to these six gradient field strength features that were based on shells, we also defined a radial gradient field strength feature, $GMR_{std/av}$, along the radial direction. As implied by the subscript (std/av), the $GMR_{std/av}$ feature contained two components, the average and the standard deviation of the gradient magnitude. Each of the 26 voxels (P_{ij} , $i=0, 1, 2, \dots, 25$) on the surface of the j th shell defines a radial direction away from the center of the sphere. For this radial feature, the average was defined by first finding the maximum gradient magnitude in each radial direction among the shells, and then computing the mean over the 26 directions. Similarly, the definition of the standard deviation involved the maximum gradient magnitude in each radial direction among the shells. This radial feature was therefore defined as

$$GMR_{std/av} = \min_{v \in U} \left\{ \frac{1}{\sum_{i=0}^{K-1} \max_{j \in J} \{SM_{ij}\}} \sqrt{\frac{\sum_{i=0}^{K-1} (K \cdot \max_{j \in J} \{SM_{ij}\} - \sum_{i=0}^{K-1} \max_{j \in J} \{SM_{ij}\})^2}{K-1}} \right\}. \tag{8}$$

Twelve gradient field orientation features were defined, based on $\{SD_{ij}\}$ on the shells or in the radial directions. Two gradient field orientation features obtained in the radial directions were average maximum gradient field orientation value (GDR_{av}), and standard deviation of the maximum gradient field orientation value (GDR_{std}). Similar to the radial gradient field strength features, these radial gradient field orientation features were defined by first finding the maximum gradient field orientation in each radial direction among the shells, and then computing the appropriate statistics. The other ten gradient field orientation features were derived from $\{SD_{ij}\}$ based on shells. They were the maximum value (GD_{max}), minimum value (GD_{min}), median value (GD_{med}), average (GD_{av}), standard deviation (GD_{std}), coefficient of variation (GD_{cv}), ratio of standard deviation to median value ($GD_{std/med}$), ratio of median value to average value ($GD_{med/av}$), squared ratio of minimum value to maximum

value ($GD_{(min/max)}^2$), and squared ratio of median value to maximum value ($GD_{(med/max)}^2$) of the gradient field orientation feature. To compute these ten features, we first computed the appropriate statistics, e.g., maximum, minimum, median, average, etc., over each shell, and then combined the information from different shells and different voxels within the VOI. The definitions of these features were very similar to those of the corresponding gradient field strength features described earlier, except that SM_{ij} was replaced by SD_{ij} . For example, the standard deviation (SD_{std}) feature was defined as

$$GD_{std} = \min_{v \in U} \left\{ \min_{j \in J} \left\{ \sqrt{\frac{\sum_{i=0}^{K-1} (SD_{ij} - \sum_{i=0}^{K-1} SD_{ij}/K)^2}{K-1}} \right\} \right\} \tag{9}$$

C. 3D ellipsoid features

3D ellipsoid fitting has been widely used to approximate the shape or distribution of a set of data points. There are nine parameters for an ellipsoid, three for the center coordinates, three for the lengths of the principal axes, and three for the orientations of the axes. In this application, we are only interested in the lengths of the three principal axes, from which features are derived to characterize the shapes of the identified objects. Let q_1 , q_2 , and q_3 denote the lengths of the principal axes of the ellipsoid fitted to an object segmented in the prescreening stage of our algorithm. The six ellipsoid features were defined as

$$L_{11} = q_1, \quad L_{22} = q_2, \quad L_{33} = q_3, \quad L_{12} = \frac{q_1}{q_2},$$

$$L_{13} = \frac{q_1}{q_3}, \quad L_{23} = \frac{q_2}{q_3}.$$

For spherical nodules, the values of L_{12} , L_{13} , and L_{23} are close to unity, while for elongated objects such as vessels, one of the ellipsoid axes will be much longer than the other two. Assuming q_1 is the longest axis and q_3 the shortest, L_{13} will be much larger than 1. These features can thus help differentiate nodules from vessels.

D. Feature spaces

As discussed at the beginning of Sec. II B, we had previously developed 19 features for FP reduction.² These features were object volume, surface area, average and standard deviation of the gray-scale values, bounding box and its volume, ratio of object volume to bounding box volume, maximum 2D area, maximum perimeter, maximum circularity, maximum eccentricity, maximum fitting ellipse major and minor axes and their ratio, compactness, and skewness, and kurtosis of the gray-level histogram. To measure the effect of the new features developed in this study on the detection performance, we performed ROC and FROC analyses in three feature spaces. The first feature space contained the 25 new features (19 gradient field and 6 ellipsoid features), which is referred to as the new feature space. The second feature space included the previous 19 features, and is referred to as the previous feature space. The third feature space combined the 19 previous features and the 25 new features. This feature space, which contains 44 features, is referred to as the combined feature space.

E. Classification

In this study, we used a linear discriminant analysis (LDA) classifier with stepwise feature selection to discriminate between true and false positives. In stepwise feature selection, individual features are entered into or removed from the selected feature pool by analyzing the effect of the entry or removal on a selection criterion. Given a set of features already in the selected feature pool, the stepwise procedure inspects the significance of the change in the selection criterion that would be obtained by entering each

feature that has not been selected into the feature pool one at a time. The best feature at a given step is entered into the selected feature pool if the significance of the change is higher than a pre-selected F -to-enter threshold, F_{in} . Another threshold, F_{out} , is preselected for feature removal. The tolerance of entering a feature that has a high correlation with the features that are already in the selected feature pool is set by using a third threshold tol . For a given training data set, the thresholds F_{in} , F_{out} , and tol need to be optimized to yield an optimal or near-optimal set of features. We used a simplex algorithm^{23,24} to perform the optimization, which minimizes an error function through rolling the defined hyper-polygon toward the direction of the vertex with the minimum error value. The optimization is terminated when the improvement in the error function stagnates or when the number of iterations reaches a preset value.

Training and testing for the main data set were performed using a leave-one-patient-out scheme. The main data set of 56 patients was partitioned 56 times into 55 training cases and one test case by changing the test case in a round-robin manner. For each partition, feature selection with simplex optimization and LDA coefficient estimation were performed based on the CT scans of the training cases, and the designed classifier was applied to the CT scans of the left-out test case. The test scores obtained by this leave-one-patient-out partitioning method were used as the decision variable in the ROC and FROC analyses. The ROC curves were estimated

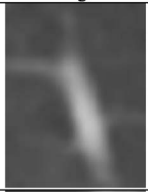

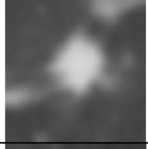


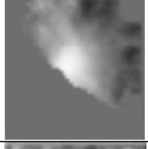
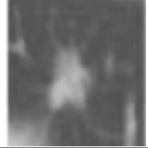
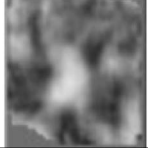
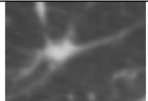

Identified objects	Original slice images	Average gradient field orientation	GD_{av}
Vessel			0.7881
Circumscribed nodule			0.9828
Juxtapleural nodule			0.9473
Spiculated nodule			0.8652
Spiculated nodule			0.7604

FIG. 9. Example of identified TP and FP objects and the spatial distribution of their average gradient field orientation that has been mapped linearly to a gray scale of 256. The feature, GD_{av} , was calculated in the 3D volume of interest. These examples show the central slice through each of the VOIs.

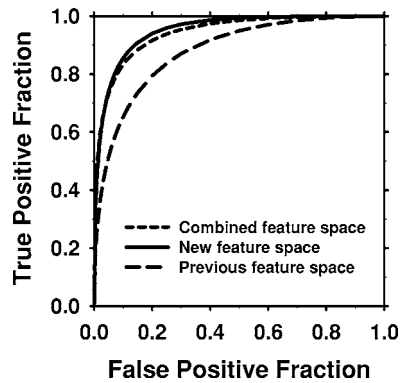


FIG. 10. Test ROC curves in different feature spaces for the main data set. The A_z values are 0.88 ± 0.02 , 0.95 ± 0.01 , and 0.94 ± 0.01 for the previous, new, and combined feature space, respectively. The differences in the A_z values between the new and the previous feature spaces and between the combined and the previous feature spaces were both statistically significant with a two-tailed p -value < 0.0001 . The difference between the new and the combined feature spaces did not achieve statistical significance.

using the LABROC software by Metz *et al.*,²⁵ which also provides the area A_z under the curve and its standard deviation. To obtain the test results for the secondary data set, the FP classifier was trained using the entire main data set, and then applied to the CT scans of the secondary data set.

III. RESULTS

Figure 9 shows examples of the central slice through the VOIs containing objects identified after prescreening. The corresponding average gradient field orientation at every pixel on the slice, and the values of the GD_{av} feature within the VOIs are also shown to illustrate the gradient field orientation of and around the objects.

A. Main data set

For the main data set, our prescreening stage detected 24 563 nodule candidates in the 82 chest CT scans, including 111 true positives and 24 452 FPs. The sensitivity of the prescreening algorithm was 96% (111/116) and the number of FPs per section was 6.92. Of the five lung nodules that were not detected at this stage, four were connected to fissures, and one was connected to blood vessels.

In the new feature space, the stepwise feature selection algorithm selected an average of 13.7 features using the

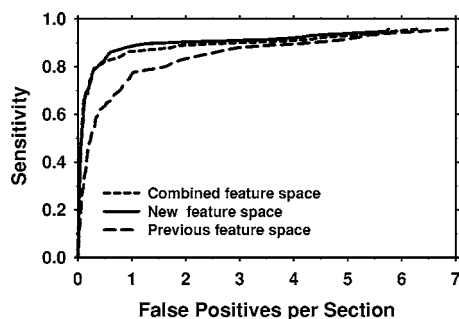


FIG. 11. Test FROC curves in different feature spaces for the main data set.

TABLE I. Test A_z values and the FP rate in the different feature spaces for the main data set.

Feature space	A_z value	FPs/section at 80% sensitivity
New feature space	0.95 ± 0.01	0.37
Previous feature space	0.88 ± 0.02	1.61
Combined feature space	0.94 ± 0.01	0.34

simplex-optimized feature selection parameters. The test A_z of the LDA classifier was 0.95 ± 0.01 with 0.37 FPs/section at 80% sensitivity, as shown in Figs. 10 and 11. In the previous feature space, stepwise feature selection selected an average of 9.4 features. The test A_z was 0.88 ± 0.02 with 1.61 FPs/section at 80% sensitivity. In the combined feature space, the average number of selected features was 24.4, and the test A_z was 0.94 ± 0.01 . At 80% sensitivity there were 0.34 FPs/section. The test results are summarized in Table I. The difference in the A_z values between the new and the previous feature spaces was statistically significant (two-tailed p -value < 0.0001). Likewise, the difference in the A_z values between the combined and the previous feature spaces was statistically significant (two-tailed p -value < 0.0001). The difference between the new and the combined feature spaces did not achieve statistical significance.

The characteristics of the false negative (FN) nodules in the main data set at a decision threshold of 0.5 FPs/section in the three feature spaces are tabulated in Table II. In the 82 chest CT scans, 27 nodules were rated as spiculated/irregular, 36 nodules as juxta-pleural, and 8 nodules as juxta-vascular by the radiologists. At 0.5 FPs/section, there were 38 FNs in the previous space (including 9 spiculated nodules, 12 juxta-pleural nodules, 5 juxta-vascular nodules). The number of FNs was reduced to 14 in the new feature space (including 4 spiculated nodules, 8 juxta-pleural nodules, and 1 juxta-vascular nodule), and 15 in the combined feature space (including 7 spiculated nodules, 7 juxta-pleural nodules, and 1 juxta-vascular nodule). The drastic reduction in the FNs, especially for the more regularly shaped (near spherical) nodules, in the new and combined feature spaces indicated the effectiveness of the new features in distinguishing true nodules and false positive objects.

TABLE II. Characteristics of the false negative (FN) nodules for the main data set at a FP rate of 0.5/section in the different feature spaces.

Nodule types	Total identified by radiologists	Previous feature space	New feature space	Combined feature space
Spiculated/irregular	27	9	4	7
Juxta-pleural	36	12	8	7
Juxta-vascular	8	5	1	1
Other	45	12	1	0
Total No. of FNs at 0.5 FPs/section		38	14	15

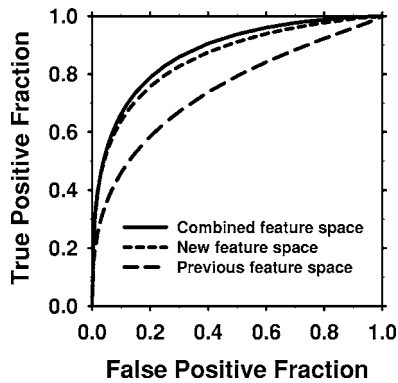


FIG. 12. Test ROC curves in different feature spaces for the secondary data set that consisted of cases containing significant atelectasis. The A_z values are 0.75 ± 0.07 , 0.86 ± 0.05 , and 0.88 ± 0.05 for the previous, new, and combined feature space, respectively. The difference in the A_z values between the combined and the previous feature spaces was statistically significant (two-tailed p -value=0.05). The difference between the new and the previous feature spaces did not achieve statistical significance

B. Secondary data set

For the secondary data set of 9 chest CT scans with dependent atelectasis, the prescreening algorithm detected 5592 nodule candidates, including 16 of the 17 true nodules and 5576 FP objects. The number of FPs/section was 9.59 (5576/581), much larger than that for the main data set. The test results are presented in Figs. 12 and 13, and Table III. The number of FPs/section at 80% sensitivity for this group was higher than that in the main data set because of the atelectasis opacities on the CT images. There were a large number of small structures with compact shapes in these areas, as shown in Fig. 14, compared to the main data set. Many of these structures could not be eliminated in the FP reduction stage and resulted in a large number of FPs for the secondary data set. The A_z values in the previous, new, and combined feature spaces were 0.75 ± 0.07 , 0.86 ± 0.05 , and 0.88 ± 0.05 , respectively. The difference in the A_z values between the combined and the previous feature spaces was statistically significant (two-tailed p -value=0.05). The difference between the new and the previous feature spaces did not achieve statistical significance.

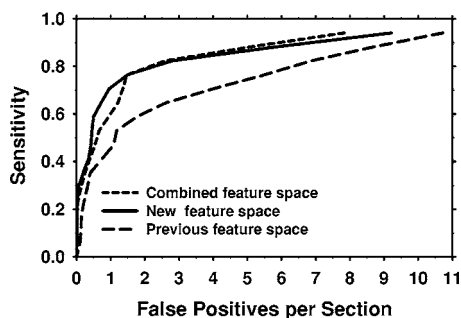


FIG. 13. Test FROC curves in different feature spaces for the secondary data set that consisted of cases containing significant atelectasis.

TABLE III. Test A_z values and the FP rate in the different feature spaces for the secondary data set.

Feature space	A_z value	FPs/section at 80% sensitivity
New feature space	0.86 ± 0.05	2.28
Previous feature space	0.75 ± 0.07	6.32
Combined feature space	0.88 ± 0.05	2.15

IV. DISCUSSION

Figure 9 shows examples of the spatial distribution of the average gradient field orientation for nodules and vessels. The gradient field orientation value was calculated in the 3D volume of interest and the examples show only the central slice through the VOI. The average gradient field orientation feature within the VOI, GD_{av} , can distinguish the vessel from the circumscribed and juxta-pleural nodules, and one of the spiculated nodules in these examples. Through a statistical analysis of all the gradient field orientation and gradient field strength features, the shape information of the candidate objects can be explored in a more thorough way. For spiculated nodules, the gradient field features may not be able to provide as high a level of discrimination from the vessels because the gradient direction around the spicules may not be radial. Gradient field features depend on the distribution of the gray-level values around the candidate objects. As a result, they are less influenced by the accuracy of the segmentation because the shape information is derived from the original gray-scale image data, not the segmented binary image data. Ellipsoid features represent the shape information in a different manner. They characterize the outline of the segmented objects, which is highly dependent upon the accuracy of object segmentation. The combination of the gradient field and the ellipsoid features provides a more detailed depiction of the objects by utilizing both types of information.

Our test results indicate that compared with our previously designed features, the newly designed gradient field and ellipsoid features significantly improved the FP reduction performance. Table I shows the improvement in performance of the LDA classifiers designed with the 25 new fea-

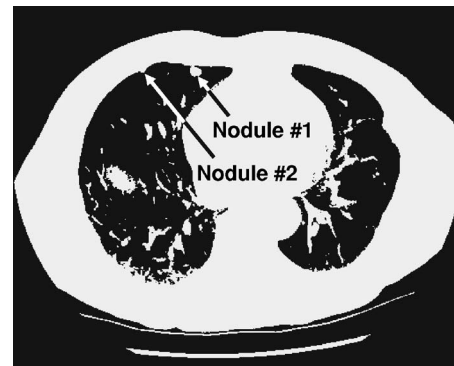


FIG. 14. Nodule candidates detected in the prescreening stage for the case with significant dependent atelectasis shown in Fig. 1.

tures over that with the previous 19 features alone. The test A_z value is significantly improved ($p < 0.0001$) from 0.88 to 0.95 with the use of the new features. The number of false positives at a sensitivity of 80% was reduced by approximately a factor of 4. The improvement is a result of more detailed 3D description of the gray-level value distribution around the nodule candidates. Table I and Figs. 10 and 11 also show that the performance in the new feature space is very similar to that in the combined feature space. The A_z value in the new feature space (95%) is slightly better than that (94%) in the combined feature space. To keep the dimensionality of the selected feature space to a minimum,^{26,27} it is prudent to use the 25 new features to replace the previous 19 features instead of combining the two feature spaces.

In this study, we separated the data set of 91 scans into a main data set that did not contain other major diseases or symptoms in addition to the lung nodules, and a secondary data set that contained significant dependent atelectasis. In this way, we were able to analyze separately the detection performance for cases that contained an additional diffuse abnormality, and for cases that did not. The improvement obtained by the new feature space in our secondary data set parallels that obtained in the main data set. However, the overall detection accuracy for the secondary data set is lower because of the FPs caused by the opacities in the atelectasis areas in the CT images. The lower test A_z values for this data set compared to the main data set indicates that our features are not as effective in distinguishing these FPs from true nodules. We are currently investigating methods for identifying areas of atelectasis on CT scans so that improved techniques can be devised to reduce the FPs in these areas.

Based on radiologists' ratings, 11 of the 116 nodules in our data set contained a ground-glass component. Two of the 11 nodules with a ground glass component had smooth boundaries, eight had spiculated/irregular boundaries, and one had a lobulated boundary. Our prescreening algorithm detected 111 nodules, including all 11 nodules with ground-glass component and 100 other nodules. We compared the scores attained by these two types of nodules detected by the prescreening program in our main data set. The average and standard deviation of the scores for the 100 nodules without a ground glass component were 4.98 and 6.67, respectively. In comparison, the corresponding numbers for the 11 nodules with a ground-glass component were 1.40 and 5.22, respectively. Therefore, on average, the nodules with a ground-glass component had lower scores compared to other nodules. However, the difference was within one standard deviation of the nodule scores, indicating that the difference did not achieve statistical significance.

In this study, three thoracic radiologists with 4 to 30 years of experience in the interpretation of chest radiographs and thoracic CT images interpreted different subsets of the data set. The radiologists used clinical criteria to detect nodules, to decide whether a nodule was juxta-pleural or juxta-vascular, and to provide the descriptors of nodule characteristics such as composition and margins. To minimize the interobserver variability, they were provided with clear

instructions on other criteria, and with tools on a user interface to assist them in marking locations, making quantitative measurements, and recording nodule characteristics. As in other areas of diagnostic interpretation, there will be interobserver variability in these interpretations. We plan to study the impact of interobserver variability on the performance evaluation of the CAD system in the future by having multiple radiologists to read the entire test data set. We also expect that the database from the Lung Image Database Consortium (LIDC) will be available with gold standards provided by multiple radiologists, which will serve as a better test set for the evaluation of our CAD system.

For all of the CT volumes in our data set, the voxel size in the axial (x - y) plane was smaller than that in the z direction. We used linear interpolation in the z direction to interpolate the VOI into isotropic voxels. The purpose of the interpolation was to facilitate the gradient calculation and other image processing operations. For CT data with relatively lower spatial resolution in the z direction, the features of an object in this dimension will be distorted and cannot be recovered by interpolation. As a result of the rapid advancement in multidetector row CT technology in recent years, it may be possible to obtain chest CT volumes such that the z direction resolution approaches that on the x - y plane in the near future. Ellipsoid, gradient, and other shape features extracted from detected objects on these thinner-slice images are expected to be more accurate and more effective for discriminating nodules and vessels. The performance of our FP reduction technique will likely be improved when applied to CT volumes with higher z -direction resolution.

We analyzed the types of missed lung nodules at an FP rate of 0.5/section in order to gain further insight into the performance of our detection system. As shown in Table II, the decrease in the numbers of false-negatives for spiculated, juxta-pleural, and juxta-vascular nodules after FP reduction indicates that the new features provide better differentiation between these nodules and FPs than the previous features even for nonspherical nodules, although the new features were not designed specifically for these types of nodules. It is also interesting to note that at 0.5 FPs/section, almost all of the missed nodules in the combined or new feature spaces are either spiculated/irregular or juxta-pleural, indicating that the new features are more effective for nodules with higher sphericity.

Based on the main data set and a leave-one-patient-out resampling scheme, an average of 13.4 features were selected in the new feature space. The frequency of the features selected in the new feature space is tabulated in Table IV. Among the selected features, 5 are ellipsoid features, 3 are gradient field strength features, and 7 are gradient field orientation features. This indicates that both gradient field strength and gradient field orientation features are useful for FP reduction. In addition to the ratio of the longest and shortest principal axes of the fitted ellipsoid (L_{13}), the lengths of the principal axes (L_{11} , L_{22} , and L_{33}) may have been selected by the feature selection algorithm because of their ability for characterizing the lesion size. The remaining 10 features in

TABLE IV. The frequency of the features selected in the 56 cycles of leave-one-patient-out resampling during classifier design.

	Number of times selected	Gradient field orientation	Number of times selected	Gradient field strength	Number of times selected
L_{11}	56	GD_{\min}	45	GM_{cv}	55
L_{22}	35	GD_{cv}	56	$GMR_{std/av}$	55
L_{33}	56	$GD_{(\min/\max)}^2$	56	GM_{std}	56
L_{13}	35	GD_{av}	56		
L_{23}	21	GDR_{av}	56		
		GDR_{std}	56		
		GD_{std}	43		

the new feature space were not selected at all, indicating that they were not as effective as the other features.

V. CONCLUSION

We formulated the gradient field orientation and gradient field strength features based on the gradient field vectors at voxels within the 3D image gradient field. Nineteen gradient field features were derived from the statistics of these gradient field vectors. In addition, six ellipsoid features were extracted based on ellipsoids fitted to the binary objects. We demonstrated that a subset of features selected from these 3D gradient field and 3D ellipsoid features can significantly reduce false positives in our computerized lung nodule detection system. The new 3D gradient field features and the 3D ellipsoid features can replace our previously developed features and improve the differentiation of true lung nodules from other structures such as vessels. With a leave-one-patient-out training and testing scheme, the test A_z values with the new features and with our previous features were estimated to be 0.95 ± 0.01 and 0.88 ± 0.02 , respectively. These classifiers reduced the FP rates to 0.37 and 1.61 per section, respectively, at 80% sensitivity for our detection system. The improvement in the test A_z value with the new features over that of the previous features was statistically significant (two-tailed p -value < 0.0001). The features in this study mainly explored the shape information using features around the boundary of the identified nodule candidates. Future investigation will explore the capability of gray level and texture information within the objects for further reduction of false positive detections.

ACKNOWLEDGMENTS

This work is supported by USPHS Grant No. CA93517. The authors are thankful to Charles E. Metz, Ph.D., for providing the LABROC program.

^{a)}Electronic mail: berki@umich.edu

¹G. L. Zorn and J. C. Nesbitt, "Surgical management of early stage lung cancer," *Semin Surg. Oncol.* **18**, 124–136 (2000).

²M. N. Gurcan, B. Sahiner, N. Petrick, H. P. Chan, E. A. Kazerooni, P. N. Cascade, and L. Hadjiiski, "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system," *Med. Phys.* **29**, 2552–2558 (2002).

³J. W. Gurney, "Determining the likelihood of malignancy in solitary pul-

monary nodules with Bayesian analysis. I. Theory," *Radiology* **186**, 405–413 (1993).

⁴J. W. Gurney, "Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. II. Application," *Radiology* **186**, 415–422 (1993).

⁵J. P. Ko and M. Betke, "Chest CT: Automated nodule detection and assessment of change over time—Preliminary experience," *Radiology* **218**, 267–273 (2001).

⁶Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Trans. Med. Imaging* **20**, 595–604 (2001).

⁷S. G. Armato, M. L. Giger, and H. MacMahon, "Automated detection of lung nodules in CT scans: Preliminary results," *Med. Phys.* **28**, 1552–1561 (2001).

⁸M. F. McNitt-Gray, E. M. Hart, N. Wyckoff, J. W. Sayre, J. G. Goldin, and D. R. Aberle, "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results," *Med. Phys.* **26**, 880–888 (1999).

⁹M. S. Brown, J. G. Goldin, R. D. Suh, M. F. McNitt-Gray, J. W. Sayre, and D. R. Aberle, "Lung micronodules: Automated method for detection at thin-section CT—Initial experience," *Radiology* **226**, 256–262 (2003).

¹⁰K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, R. Kakinuma, M. Kaneko, N. Moriyama, and K. Eguchi, "Computer-aided diagnosis for pulmonary nodules based on helical CT images," *Comput. Med. Imaging Graph.* **22**, 157–167 (1998).

¹¹B. V. Ginneken, B. T. H. Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey," *IEEE Trans. Med. Imaging* **20**, 1228–1241 (2001).

¹²J. Klein, "Solitary pulmonary nodule: Diagnostic assessment," *RSNA, Categorical Course in Diagnostic Radiology: Thoracic Imaging—Chest and Thoracic*, 2001, pp. 185–194.

¹³S. G. Armato, M. B. Altman, and J. Wilkie, "Automated lung nodule classification following automated nodule detection on CT: A serial approach," *Med. Phys.* **30**, 1188–1197 (2003).

¹⁴S. Armato, F. Li, M. Giger, H. MacMahon, S. Sone, and K. Doi, "Lung cancer: Performance of automated lung nodule detection applied to cancers missed in a CT screening program," *Radiology* **225**, 685–692 (2002).

¹⁵H. Satoh, Y. Ukai, N. Niki, K. Eguchi, K. Mori, H. Ohmatsu, R. Kakinuma, M. Kaneko, and N. Moriyama, "Computer aided diagnosis system for lung cancer based on retrospective helical CT image," *Proc. SPIE* **3661**, 1324–1335 (1999).

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).

¹⁷J. Nappi and H. Yoshida, "Automated detection of polyps with CT colonography: Evaluation of volumetric features for reduction of false-positive findings," *Acad. Radiol.* **9**, 386–397 (2002).

¹⁸S. Jaggi, W. Karl, and A. Willsky, "Estimation of dynamically evolving ellipsoids with application to medical imaging," *IEEE Trans. Med. Imaging* **14**, 249–258 (1995).

¹⁹V. Ranjan and A. Fournier, "Volume models for volumetric data," *Computer* **27**, 28–36 (1994).

²⁰K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," *Signal Process. Image Commun.* **12**, 263–281 (1998).

²¹J. Volpe, M. L. Storto, K. Lee, and W. R. Webb, "High-resolution CT of the lung: Determination of the usefulness of CT scans obtained with the patient prone based on plain radiographic findings," *AJR, Am. J. Roentgenol.* **169**, 369–374 (1997).

²²A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice-Hall, Eaglewood, Cliffs, NJ, 1989).

²³L. M. Hadjiiski, H. P. Chan, B. Sahiner, N. Petrick, and M. A. Helvie, "Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis—local affine transformation for improved localization," *Med. Phys.* **28**, 1070–1079 (2001).

²⁴S. S. Rao, *Optimization: Theory and Applications* (Wiley Eastern Limited, New York, 1979).

²⁵C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from

continuously-distributed data," *Stat. Med.* **17**, 1033–1053 (1998).

²⁶H. P. Chan, B. Sahiner, R. F. Wagner, and N. Petrick, "Classifier design for computer-aided diagnosis in mammography: Effects of finite sample size," *Med. Phys.* **24**, 1034–1035 (1997).

²⁷H. P. Chan, B. Sahiner, R. F. Wagner, and N. Petrick, "Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers," *Med. Phys.* **26**, 2654–2668 (1999).