A NOVEL SHAPE-BASED DIAGNOSTIC APPROACH FOR EARLY DIAGNOSIS OF LUNG NODULES

A. El-Baz[1], *M. Nitzken*[1], *E. Vanbogaert*[1],*G. Gimel'farb*[2], *R. Falk*[3], *and M. Abo El-Ghar*[4]

[1]BioImaging Laboratory, Bioengineering Department,University of Louisville, Louisville, KY, USA.
[2]Department of Computer Science, University of Auckland, Auckland, New Zealand.
[3]Medical Imaging Division, Department of Radiology, Jewish Hospital, Louisville, KY, USA.
[4]Urology and Nephrology Department, University of Mansoura, Mansoura, Egypt.

ABSTRACT

An alternative method of diagnosing malignant lung nodules by their shape rather than conventional growth rate is proposed. The 3D surfaces of the detected lung nodules are delineated by spherical harmonic analysis that represents a 3D surface of the lung nodule supported by the unit sphere with a linear combination of special basis functions, called spherical harmonics (SHs). The proposed 3D shape analysis is carried out in five steps: (*i*) 3D lung nodule segmentation with a deformable 3D boundary controlled by two probabilistic visual appearance models (the learned prior and the estimated current appearance one); (*ii*) 3D Delaunay triangulation to construct a 3D mesh model of the segmented lung nodule surface; (*iii*) mapping this model to the unit sphere; (*iv*) computing the SHs for the surface; and (*v*) determining the number of the SHs to delineate the lung nodule. We describe the lung nodule shape complexity with a new shape index, the estimated number of the SHs, and use it for the K-nearest classification into malignant and benign lung nodules. Preliminary experiments on 109 lung nodules (51 malignant and 58 benign) resulted in the 94.4% correct classification (for the 95% confidence interval), showing the proposed method is a promising supplement to current technologies for the early diagnosis of lung cancer.

Index Terms— Lung nodules, shape analysis, spherical harmonics.

1. INTRODUCTION

Pulmonary nodules are the most common manifestation of lung cancer, which is the principal cause of cancer-related deaths [1]. Fast and accurate classification of the nodules is of major importance for medical computer-aided diagnostic systems (CAD). A nodule is an approximately spherical volume of higher-density tissue visible in an X-ray lung image. Large malignant nodules (generally defined as greater than 1 cm in diameter) are easily detected with any traditional imaging equipment and then diagnosed by needle biopsy or bronchoscopy. However, diagnostic options for small malignant nodules are limited due to difficulties in their accessibility, especially if they are located deep in the tissue or away from the large airways. Therefore, additional imaging and CAD techniques are needed. The popular direction of detecting small cancerous nodules is to analyze their growth rate over time. This paper introduces a new approach to characterize the detected nodules based on their shape.

Related work: A great deal of work has been published regarding the usefulness of morphologic features for discriminating malignant from benign pulmonary nodules on computed tomography (CT) and to a lesser extent, chest radiographs. Several studies have shown a correlation between different nodule shape characteristics and underlying pathology. For example, Furuya et al. [2] analyzed the margin characteristics of 193 pulmonary nodules on high-resolution CT and subjectively classified them as one of several types, including round, lobulated, densely spiculated, ragged, and halo. They found a high level of malignancy among the lobulated (82%), spiculated (97%), ragged (93%), and halo nodules (100%), while 66% of the round nodules proved to be benign.

Automatically extracted features have also been shown to correlate with underlying malignancy. Kawata et al. [3] quantified the surface curvature and degree of surrounding radiating pattern in biopsy-proven benign and malignant nodules, and compared the resulting feature maps. Their results showed good separation of the feature maps between the two categories. Similarly, fractal analysis has been used to quantify the nodule margin characteristics of benign and malignant nodules. Kido et al. [4] used 2D and 3D fractal dimensions to analyze the lung-nodule interface in a series of 117 peripheral pulmonary nodules of various underlying pathology including benign hamartomas, tuberculomas, and pneumonias as well as malignant diagnoses including brochogenic carcinomas. They noted statistically significant differences between the 2D fractal dimensions of hamartomas and all other nodules, as well as differences between the 3D fractal dimensions pneumonias and tuberculomas and bronchogenic carcinomas. Although none of these studies directly assessed the accuracy of their methods in predicting a diagnosis, they support the notion that nodule shape can potentially be used by automated systems to distinguish benign from malignant nodules.

Several groups have designed computer aided diagnostic (CAD) systems with the goal of predicting a diagnosis based on features extracted from CT or chest radiographs. In general, they share common schema, first extracting features from the images, then designing and using an automatic classifier to categorize nodules based on these features, and lastly evaluating the performance of the system with receiver operating characteristics (ROC) analysis. The CAD systems differ in the specific extracted features and the type of classifier used, with linear discriminant classifiers (LDC) and neural networks (NN) being the most common. First, those systems using LDCs will be discussed followed by NN-based classifier systems.

Kawata and colleagues [5] designed a CT-based CAD system that classified pulmonary nodules based on a combination of curvature index and the relationship of the nodules to their surrounding features. Mori et al. [6] also designed a CAD system using curvedness index in combination with dynamic contrast-enhanced CT in order to evaluate temporal change as a possible discriminating fea-

ture of benign and malignant nodules. The CAD system developed by McNitt-Gray et al. [7] used a pattern classification approach to determine the malignancy of pulmonary nodules on HRCT in a series of 31 cases (17 malignant, 14 benign). They identified solitary nodules using a semi-automated contouring technique, and extracted quantitative measures of the resulting contour related to shape, size, attenuation, distribution of attenuation and texture.

One of the early neural network based CAD systems was developed by Gurney and Swensen [8]. They compared two systems, one using a neural network based classifier and one using a Bayesian classifier. Both systems used a combination of subjectively evaluated clinical and radiologic characteristics including border smoothness, spiculation and lobulation.

In summary, the aforementioned existing approaches show the following limitations:

- Most of them classify the lung nodules based on extracted 2D features (e.g., round, lobulated, ragged, and halo, etc.) and they did not take into account the 3D features of lung nodules.
- Most of them did not provide a quantitative measure that has the ability to describe the shape complexity of lung nodules.
- Most of the existing features depend on the accuracy of the used nodule segmentation algorithm which make them difficult to be used by clinical practitioners.

This work aims to address these variations and discrepancies in a way that will make evaluating small lung masses more consistent and allow for a more accurate description of the shape complexity of the detected lung nodules.

2. METHODS

In this work, we propose a novel shape-based approach for the analysis of lung nodule variability between malignant and benign nodules. The proposed analysis begins by using lung nodules that are segmented by applying our algorithm that is introduced in [9]. (For completeness, the following section will briefly describe our algorithm. In the experimental results section we will demonstrate its accuracy on the new data sets).

2.1. Lung Nodules Segmentation

Accurate lung nodules segmentation from a 3D LDCT images is a challenging problem because intensities of the lung nodules and surrounding tissues (e.g., blood vessels, chest, etc.) are not clearly distinguishable. To overcome this problem, we use a conventional 3D parametric deformable boundary [10] but control its evolution with two probabilistic visual appearance models, namely, a learned lung nodule appearance prior and a current appearance model of the image to be segmented. The prior is a 3D Markov-Gibbs random field (MGRF) model of the lung nodules intensities with translation- and rotation-invariant pairwise voxel interaction, being learned analytically from training data in accord with [11]. The current appearance is modeled by a mixed marginal distribution of the voxel intensities in both the lung nodule and surrounding tissues. To extract the voxel-wise model of the current nodule appearance, the mixture is precisely approximated with a linear combination of discrete Gaussians (LCDG) [12] and automatically separated into the lung nodule and background LCDG models. For more details about the segmentation model and for a comparison with other approaches see [9].

2.2. Spherical harmonics (SHs) shape analysis

Spectral SH analysis [13] considers 3D surface data as a linear combination of specific basis functions. In our case, the surface of the segmented lung nodule is approximated first by a triangulated 3D mesh (see Fig. 1) built with an algorithm by Fang and Boas [14]. Secondly, the lung nodule surface for each subject is mapped for the SH decomposition to the unit sphere. We propose a novel mapping approach, called "*Attraction-Repulsion*" that calls for all the mesh nodes to meet two conditions: (*i*) the unit distance of each node from the lung nodule center as shown in Fig. 2, and (*ii*) an equal distance of each node from all of its nearest neighbors as shown in Fig. 3.

Fig. 1. Generating a 3D mesh for the lung nodule surface from a stack of successive segmented 2D LDCT slices.

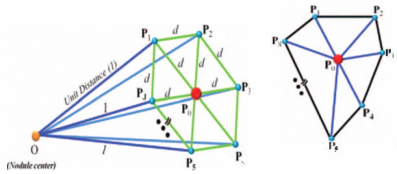

Fig. 2. 3D illustration of the unit distance from all surface nodes to the center of the lung nodule.

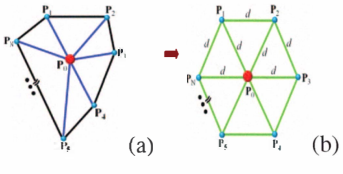

Fig. 3. 2D illustration of the neighbors rearrangement: initial (a) vs. final equidistant locations (b) in all the directions.

To detail our Attraction-Repulsion Algorithm (see its summary in Algorithm 1), let τ denote the iteration index, I be the total number of the mesh nodes (in all the experiments below $I = 4896$ nodes), and $\mathbf{P}_{\tau,i}$ be the Cartesian coordinates of the surface node i at iteration τ; $i = 1, \ldots, I$. Let J be the number of the neighbors for a mesh node (see e.g. Fig. 3) and $d_{\tau,ij}$ denote the Euclidean distance between the surface nodes i and j at iteration τ (as shown in Fig. 3(b)), where $i = 1, \ldots, I$ and $j = 1, \ldots, J$. Let $\mathbf{\Delta}_{\tau,ji} = \mathbf{P}_{\tau,j} - \mathbf{P}_{\tau,i}$ denote the displacement between the nodes j and i at iteration τ. Let $C_{\mathrm{A},1}$, $C_{\mathrm{A},2}$, C_{R} be the attraction and repulsion constants, respectively, that control the displacement of each surface node.

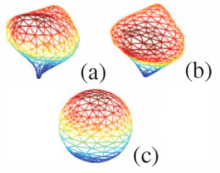

Fig. 4. Lung nodule mesh (a), its smoothed version (b), and the Attraction-Repulsion mapping to the unit sphere (c).

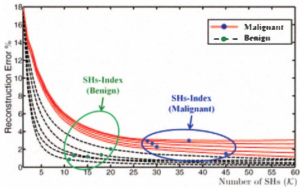

Fig. 5. Estimation of the shape index from the total nodule approximation error for malignant and benign nodules.

The starting attraction step of the proposed mapping tends to center each node $\mathbf{P}_i$; $i = 1, \ldots, I$, with respect to its neighbors by

adjusting iteratively its location:

$$\mathbf{P}'_{\tau,i} = \mathbf{P}_{\tau,i} + C_{\mathrm{A},1} \sum_{j=1;j\neq i}^{J} \boldsymbol{\Delta}_{\tau,ji} d^2_{\tau,ji} + C_{\mathrm{A},2} \frac{\boldsymbol{\Delta}_{\tau,ji}}{d_{\tau,ji}} \quad (1)$$

where the factor $C_{\mathrm{A},2}$ keeps the tightly packed nodes from collision and also pushes the adjusted nodes away from their neighbors if a certain neighbor is much closer than the others.

The subsequent repulsion step inflates the whole mesh by pushing all the nodes outwards to become evenly spaced after their final back-projection onto the unit sphere along the rays from the center of the sphere. To ensure that the nodes that have not been shifted will not collide with the altered node, the location of each node $\mathbf{P}_i$; $i = 1, \ldots, I$, is updated before the back-projection as follows:

$$\mathbf{P}^{\circ}_{\tau+1,i} = \mathbf{P}'_{\tau,i} + \frac{C_{\mathrm{R}}}{2I} \sum_{j=1;j\neq i}^{I} \left(\frac{\boldsymbol{\Delta}_{\tau,ji}}{|\boldsymbol{\Delta}_{\tau,ji}|^2} \right) \quad (2)$$

where a repulsion constant C_{R} controls the displacement of each surface node and establishes a balance between the processing time and accuracy (e.g. a smaller C_{R} values guarantees that the node faces will not become crossed during the iterations at the expense of the increased processing time). All the experiments below are obtained with $0.3 \leq C_{\mathrm{R}} \leq 0.7$.

Algorithm 1: Attraction-Repulsion Algorithm

Initialization

- Construct the 3D lung nodule mesh (Fig. 4,a).
- Smooth it by the Laplacian filtering (Fig. 4,b).
- Initialize the mapping of the smoothed mesh to the unit sphere.

Repeat

- **For** $i = 1 \longrightarrow I$
  - **Attraction:**
    - Select a node to process.
    - Update the node using Eq. (1)
  - **Repulsion:**
    - Update the node using Eq. (2).
- **End** (all nodes in the mesh are shifted and back-projected onto the unit sphere).

While changes in the node positions occur (Fig. 4,c).

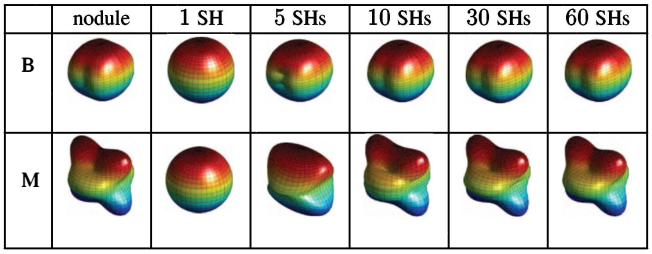

Fig. 6. Approximation of the 3D shape for malignant (M) and benign lung nodules (B).

The original lung nodule mapped to the unit sphere with the proposed Attraction-Repulsion algorithm is approximated by a linear combination of SHs, the lower-order harmonics being sufficient to represent more generic information, while the finer details requiring the higher-order ones. The SHs are generated by the solving an isotropic heat equation for the nodule surface on the unit sphere. Let $\mathcal{S} : \mathbf{M} \rightarrow \mathbf{U}$ denote the mapping of a nodule mesh $\mathbf{M}$ to the unit sphere $\mathbf{U}$. Each node $\mathbf{P} = (x, y, z) \in \mathbf{M}$ mapped to the spherical position $\mathbf{u} = S(\mathbf{P})$ is represented by the spherical coordinates $\mathbf{u} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ where $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$ are the polar and azimuth angles, respectively. The SH $Y_{\alpha\beta}$ of degree α and order β is defined as [15]:

$$Y_{\alpha\beta} = \begin{cases} c_{\alpha\beta} G_{\alpha}^{|\beta|} \cos\theta \sin(|\beta|\varphi) & -\alpha \leq \beta \leq -1 \\ \frac{c_{\alpha\beta}}{\sqrt{2}} G_{\alpha}^{|\beta|} \cos\theta & \beta = 0 \\ c_{\alpha\beta} G_{\alpha}^{|\beta|} \cos\theta \cos(|\beta|\varphi) & 1 \leq \beta \leq \alpha \end{cases} \quad (3)$$

where $c_{\alpha\beta} = \left(\frac{2\alpha+1}{2\pi} \frac{(\alpha-|\beta|)!}{(\alpha+|\beta|)!} \right)^{\frac{1}{2}}$ and $G_{\alpha}^{|\beta|}$ is the associated Legendre polynomial of degree α and order β. For the fixed α, the polynomials G_{α}^{β} are orthogonal over the range $[-1, 1]$. As shown in [15], the Legendre polynomials are effective in calculating SHs, and this is the main motivation behind their use in this work.

Finally, the lung nodule is reconstructed from the SHs of Eq. (3). In the case of the SHs expansion, the standard least-square fitting does not model accurately the 3D shape of the lung nodule and can miss some of the shape details that discriminate between the malignant and benign lung nodules. To circumvent this problem, we used the iterative residual fitting by Shen et al. [16] that accurately approximates 3D shape of malignant and benign lung nodules. As demonstrated in Fig. 6, the model accuracy does not significantly change for the benign nodule from the 15 to 60 SHs, while it continues to increase for the malignant nodule.

2.3. Quantitative lung nodule shape analysis

Our main hypothesis is that the shape of malignant nodules is more complicated (e.g., with spiculation) if it is compared with the shape of benign nodules which is simpler (smoothed shape) as in Fig. 6, so that more SHs have to be used for accurate approximation of the shape of malignant lung nodule. Therefore, the number of the SHs after which there are no significant changes in the approximations can be used as a new shape index quantifying the shape complexity of the detected lung nodules. Due to the unit sphere mapping, the original mesh for each nodule is inherently aligned with the mesh for the approximate shape, and the sum of the Euclidean distances between the corresponding nodes gives the total error between both the mesh models. As shown in Fig. 5, the total error curves for the increasing number $\mathcal{K}$ of the SHs can be used as a discriminatory feature to differentiate between the subjects.

3. EXPERIMENTAL RESULTS AND CONCLUSIONS

To justify the proposed methodology of analysis the 3D shape of both malignant and benign nodules, the above proposed shape analysis framework was pilot-tested on a database of clinical multislice 3D chest LDCT scans of 109 lung nodules (51 malignant and 58 benign). The scanned CT data sets each have $0.7 \times 0.7 \times 2.0$ mm^3 voxels, the diameters of the nodules ranging from 3 mm to 30 mm.

Segmentation results: Figure 7 illustrates results of segmenting pleural attached nodules shown by axial, sagittal, and coronal cross sections. The pixel-wise Gibbs energies in each cross section are higher for the nodules than for any other lung voxels including the

attached artery. Therefore, our approach separates accurately the pulmonary nodules from any part of the attached artery.

In total, our segmentation of the 109 nodules has an error range of 0.4% – 2.35% with a mean error of 0.96%, and a standard error deviation of 1.1%.

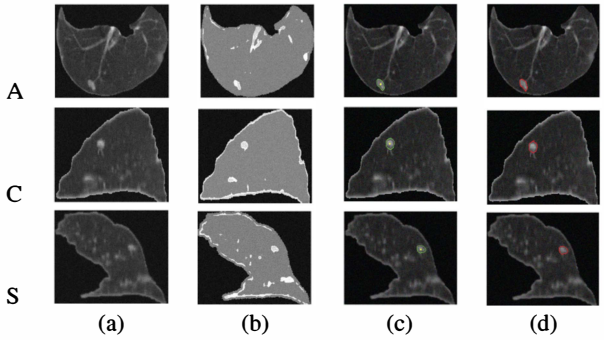

Fig. 7. 3D segmentation of pleural attached nodules; results are projected onto 2D axial (A), coronal (C), and saggital (S) planes for visualization: 2D profile of the original nodule (a), pixel-wise Gibbs energies (b), our segmentation (c), and (d) the radiologist's segmentation.

Diagnostic results: The training subset for classification (10 lung nodules shown in Fig. 5) was arbitrarily selected among all the 109 lung nodules. The accuracy of classification based on using K-nearest classifier of both the training and test subjects was evaluated using the χ^2-test at 95% confidence level. At the 95% confidence level, the correctly classified 47 out of 51 malignant nodules (a 92.1% accuracy), and 55 out of 58 control subjects (a 94.8% accuracy). So, the overall accuracy using the proposed 3D shape-based CAD system for 95% confidence level is 94.4% in **the first detection of lung nodules**. The classification based on traditional growth rate approach [17] **over one year** is 29 out of 51 malignant nodules (a 56.9% accuracy), and 49 out of 58 benign nodules (a 84.4% accuracy) at a 95% confidence level, these results highlight the advantage of the proposed shape-based diagnostic approach.

As demonstrated in this paper, the preliminary results justify further elaboration of the proposed alternative method for diagnosing malignant lung nodules. Its novelty lies in using the shape of a segmented 3D nodule instead of the more conventional growth rate as a reliable diagnostic feature. The shape is described in terms of a linear combination of spherical harmonics (SHs).

The proposed nodule shape analysis could lead to more accurate, fast, and clinically useful diagnostics of detected pulmonary nodules without the need for investigating their temporal development on the successive LDCT images of the same subject collected for a relatively long time.

4. REFERENCES

[1] H. K.Weir et al., "Annual report to the nation on the status of cancer, 1975-2000," *Journal National Cancer Institute*, vol. 95, no. 17, pp. 1276-1299, 2003.

[2] K. Furuya, S. Murayama, H. Soeda, J. Murakami, et al., "New classification of small pulmonary nodules by margin characteristics on high-resolution CT," *Acta Radiologica*, 1999, pp. 496–504.

[3] Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, et al., "Quantitative surface characterization of pulmonary nodules based on thin-section CT images," *IEEE Trans.on Nuclear Science*, vol. 45, pp. 2132–2138, 1998.

[4] S. Kido, K. Kuriyama, M. Higashiyama, T. Kasugai, and C. Kuroda, "Fractal analysis of small peripheral pulmonary nodules in thin-section CT: evaluation of the lung-nodule interfaces," *Journal of computer assisted tomography*, vol. 26, no. 4, pp. 573–578, 2002.

[5] Y. Kawata, N. Niki, H. Ohmatsu, M. Kusumoto, et al., "Computerized analysis of 3-D pulmonary nodule images in surrounding and internal structure feature spaces," *Proc. Int. Conference on Image Processing*, vol. 2, pp. 889-892, 2001.

[6] K. Mori, N. Niki, T. Kondo, Y. Kamiyama, et al., "Development of a novel computer-aided diagnosis system for automatic discrimination of malignant from benign solitary pulmonary nodules on thin-section dynamic computed tomography," *Journal of computer assisted tomography*, vol. 29, pp. 215-222.

[7] M. McNitt-Gray, E. Hart, N. Wyckoff, J. Sayre, J. Goldin, and D. Aberle, "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results," *Medical Physics*, vol. 26, pp. 880–888, 1999.

[8] W. Gurney and S. Swensen, "Solitary pulmonary nodules: determining the likelihood of malignancy with neural network analysis," *Radiology*, vol. 196, pp. 823–829, 1995.

[9] A. Farag, A. El-Baz, G. Gimelfarb, R. Falk, M. Abou El-Ghar, T. El-Diasty, and S. Elshazly, "Appearance Models for Robust Segmentation of Pulmonary Nodules in 3D LDCT Chest Images," *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06)*, Copenhagen, Denmark, 2006, vol. 1, pp. 662-670.

[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. of Comp. Vision*, vol. 1, pp. 321–331, 1987.

[11] G. Gimel'farb, *Image Textures and Gibbs Random Fields,* Kluwer Academic, 1999.

[12] A. El-Baz and G. Gimel'farb, "EM Based Approximation of Empirical Distributions with Linear Combinations of Discrete Gaussians," *Proc. of IEEE Int. Conf. on Image Processing (ICIP'07)*, San Antonio, USA, 2007, pp. 373–376.

[13] M. K. Chung, L. Shen, K. M. Dalton, et al., "Weighted Fourier series representation and its application to quantifying the amount of gray matter," *IEEE Trans. on Med. Imaging,* vol. 26, pp. 566–581, 2007.

[14] Q. Fang and D. Boas, "Tetrahedral mesh generation from volumetric binary and gray-scale images," in: *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'09), June 28 – July 1, 2009, Boston, MA, USA*, IEEE Press, pp. 1142–1145, 2009.

[15] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. II. Interscience, New York, 1953.

[16] L. Shen and M. K. Chung, "Large-scale modeling of parametric surfaces using spherical harmonics," in *Proc. 3rd Int. Symp. 3D Data Process. Visualiz. Transmission*, Chapel Hill, NC, USA June 14–16, 2006, pp. 294–301, 2006.

[17] A. Reeves, A. Chan, D. Yankelevitz, C. Henschke, B. Kressler, and W. Kostis, "Measuring the Change in Size of Pulmonary Nodules," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 435-449, April 2006.