

Realistic simulated lung nodule dataset for testing CAD detection and sizing

R.D. AMBROSINI¹ AND W.G. O'DELL^{1,2}

Departments of ¹Biomedical Engineering and ²Radiation Oncology, University of Rochester, Rochester, NY

Abstract

The development of computer-aided diagnosis (CAD) methods for the processing of CT lung scans continues to become increasingly popular due to the potential of these algorithms to reduce image reading time, errors caused by user fatigue, and user subjectivity when screening for the presence of malignant lesions. This study seeks to address the critical need for a realistic simulated lung nodule CT image dataset based on real tumor morphologies that can be used for the quantitative evaluation and comparison of these CAD algorithms. The manual contouring of 17 different lung metastases was performed and reconstruction of the full 3D surface of each tumor was achieved through the utilization of an analytical equation comprised of a spherical harmonics series. 2D nodule slice representations were then computed based on these analytical equations to produce realistic simulated nodules that can be inserted into CT datasets with well-circumscribed, vascularized, or juxtapleural borders and also be scaled to represent nodule growth. The 3D shape and intensity profile of each simulated nodule created from the spherical harmonics reconstruction was compared to the real patient CT lung metastasis from which its contour points were derived through the calculation of a correlation coefficient, producing an average value of 0.8897 (± 0.0609). This database of realistic simulated nodules can fulfill the need for a reproducible and reliable gold standard for CAD algorithms with regards to nodule detection and nodule sizing, especially given its virtually unlimited capacity for expansion to other nodule shape variants, organ systems, and imaging modalities.

Keywords: Lung nodules, CAD validation, spherical harmonics

Purpose: Recent advances in treatment, such as those seen with stereotactic radiosurgery and radiotherapy techniques, have resulted in renewed motivation for the detection and diagnosis of small lesions in the lung, especially in regards to surveillance for metastatic disease to the lung in high-risk patient populations with primary cancer of the breast, colon, bladder, and prostate. However, multi-detector CT scanners typically generate 300 to 600 image slices and the reading and interpretation of these massive amounts of image data require a tremendous level of radiologist effort, predisposing the screening process to human error. A chief concern is the detection of nodules of small size, low contrast, or those located near vessel structures, since these nodules are often missed by the unaided radiologist. Thus, computer-aided diagnosis (CAD) approaches are becoming increasingly necessary for both reducing radiologists' effort and improving detection sensitivity of lung nodules.

The Lung Image Database Consortium (LIDC) and Early Lung Cancer Action Project (ELCAP) studies sought to employ real patient CT datasets as standards for comparisons of competing CAD methods. However, both studies found a surprisingly large variation in their readers' definitions of ground truth that is likely to mask any real differences in performance of most CAD schemes, hereby negating the usefulness of these databases. Likewise, there was tremendous disagreement among the LIDC reviewers in their estimates of nodule volume, thwarting the use of these datasets for comparison of automated approaches for quantifying tumor size and growth. We propose to create a realistic simulated lung nodule image dataset that overcomes the aforementioned limitations in order to provide a basis for future comparison of CAD and human readers for both detection and sizing of nodules. Unlike prior work, our synthetic database is constructed with nodule shapes derived from real patient lung lesions, employs mathematically scaled versions of each nodule so as to enable the true assessment of nodule growth, and permits vascularized and juxtapleural nodules in proportions representative of the patterns of occurrence found in large patient datasets.

Methods: In order to provide an objective and adaptable system to validate the accuracy of lung nodule CAD detection and sizing algorithms, we have designed a synthetic nodule image dataset consisting of isolated nodules. For the purpose of constructing this dataset, a total of 17 confirmed lung metastases, ranging in diameter from approximately 8 to 20 millimeters, were contoured manually from four different patient CT datasets (slice thickness and separation of 3 mm and in-plane resolution of 0.9375 mm). This contouring process generated

roughly 150 to 500 unique surface points for each patient nodule. An analytical equation comprised of a spherical harmonic series was then used to reconstruct the full 3D surface of each tumor:

$$r = \sum_{l=0}^3 \sum_{m=-l}^l A_{l,m} Y_l^m(\theta, \phi)$$

$$Y_l^m(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) e^{im\phi}$$

where r is the nodule radius, A represents the constants determined from the extracted nodule contouring, Y is the spherical harmonic, θ is the polar (colatitudinal) coordinate, ϕ is the azimuthal (longitudinal) coordinate, and P is an associated Legendre polynomial. Through the utilization of the analytical equation, a high-resolution surface mesh was created that allowed an exact computation of each nodule's true volume. Furthermore, this surface mesh also enabled the 3D reconstruction of the appearance of each nodule, subjected to any degree of size scaling, under any set of voxel dimensions.

The appearance of each simulated nodule on a series of CT image slices was achieved through a sequence of steps. First, the eventual central coordinate of each nodule within the 3D lung space of the designated insertion CT dataset was calculated randomly, with precision equal to 0.1 of the voxel dimension in each direction. Thus, nodules were not merely centered on a given pixel and the positions of different nodules were not identically centered. The 2D nodule slice representations were computed assuming a uniform nodule density (matched to the intensity of the insertion dataset blood signal) where relative voxel intensities were based upon the partial volume of the nodule in each voxel. In the process of simulated nodule insertion, the prevention of unrealistically bright voxels was achieved by using the maximum intensity between the simulated nodule and the background features of the insertion dataset, rather than the sum of intensities, in each overlapping voxel. The point spread function of a standard clinical imaging scanner was then replicated through blurring of the simulated nodule and its neighboring voxels with a Gaussian filter. Lastly, Gaussian noise was added utilizing a standard deviation derived from the average noise inherent within the employed patient CT datasets.

During a review of their database of 200 pulmonary nodules, Kostis and his group found that almost half of all nodules were vascularized, around one-third had well-circumscribed boundaries, and approaching one-quarter of all nodules were juxtaleural (1). The fourth class, nodules with pleural tails, were found to be present in only about 1% of the nodules in the database and as such, are assumed to represent a considerable rarity among pulmonary nodules. In order to account for the three major classes of lung nodule border profiles, the complete database of simulated nodules was implemented and tested in all three different manifestations. Well-circumscribed simulated nodules were constructed as detailed above and then inserted into a region of a healthy patient CT dataset where no overlap was permitted between the voxels of the inserted nodule and blood vessel or lung wall voxels of comparable brightness. Vascularized simulated nodules involved an identical design procedure as performed with the well-circumscribed simulated nodule versions, however they were inserted deliberately at a large vessel bifurcation and allowed to share voxels with the background blood vessels to generate realistically the scenario where both structures are located within the same CT image slice. As far as juxtaleural simulated nodules, the effects of growth along the lung pleural surface on the nodule's 3D shape was incorporated into the creation process during the 3D nodule reconstruction process itself to ensure that a correct gold standard volume was still calculated for the adjusted shape. The lung slice contours for the insertion CT dataset were extracted using a semi-automatic snake implementation designed by our lab and built upon the NIH's ImageJ platform. A solid 3D lung mask was generated through a process of simple interpolation in-between extracted slices and once an insertion point within the lung had been selected, the spatial coordinates provided by the lung mask were combined with the spatial coordinates of the simulated nodules' reconstruction to produce a juxtaleural simulated nodule. Through the use of this style of juxtaleural nodule creation, a progressive increase in scaling size for a

given 3D shape led to the replication of the appearance of nodule growth adjacent to and outward from the lung wall.

Results: A critical aspect of this work was establishing the degree of similarity between our simulated nodules that were produced through the spherical harmonics-based 3D reconstructions and the real patient lung nodules that were extracted manually to provide the 3D contour points utilized as the source data for the reconstruction process. Each individual simulated nodule was compared to the corresponding real patient lung nodule that was used as the basis for its construction through the calculation of a 3D correlation coefficient. For the 17 simulated-to-real nodule comparisons, an average correlation coefficient value of 0.8897 with a standard deviation of 0.0609 was calculated (see Table 1). This high quantitative correlation signifies that the generated simulated nodules possess both realistic 3D morphologies and intensity profiles.

In order to have confidence in a lung nodule CAD algorithm, the applied method must show the capability to handle effectively vascularized and juxtapleural boundary conditions, in addition to well-circumscribed margins, as these are also common lung nodule presentations. Consequently, the other vital point of analysis for our system was its capacity to produce simulated nodules with realistic boundary presentations. The sequential CT slice appearance of a sample vascularized simulated nodule is shown in Figure 1 while the sequential CT slice appearance of a sample juxtapleural simulated nodule is displayed in Figure 2. These examples demonstrate that our approach does in fact generate nodules within CT datasets that simulate realistically the challenging presentation of vascularized and juxtapleural nodules for CAD algorithm testing.

New/Breakthrough Work: In this paper, a novel system for constructing a database of simulated nodules with realistic 3D shapes, intensity profiles, and either well-circumscribed, vascularized, or juxtapleural borders is presented. Another new aspect of this work is the quantitative assessment of the similarity between the 3D shape and intensity profile of each initially generated simulated nodule and the corresponding manually contoured real lung nodule on which it was based.

Conclusions: The results presented in this work indicate that our system of spherical harmonics-based 3D reconstructions provides simulated nodules with realistic 3D shapes, borders, and intensity profiles that can be utilized to assess the detection and sizing capabilities of any CT lung nodule CAD approach. Planned extensions of this work include the development of a web-based interface that would allow the dissemination of simulated image datasets and the uploading of user results for the automatic determination of standardized error metrics, as well as the expansion of the simulated nodule database through the incorporation of a greater number of 3D real patient nodule reconstructions.

Other Publication/Presentation Submissions: The work pertaining to the construction of realistic simulated nodules with vascularized and juxtapleural border profiles and the quantitative comparison of simulated nodule reconstructions to the original extracted real patient nodules has not been submitted for publication or presentation elsewhere.

References

1 Kostis WJ, Reeves AP, Yankelevitz DF, Henschke CI. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. *IEEE Trans Med Imag* 2003;22(10):1259-1274.

Simulated Nodule	Correlation Coefficient
1	0.8906
2	0.9237
3	0.9390
4	0.7605
5	0.9224
6	0.8854
7	0.8902
8	0.7139
9	0.9129
10	0.9323
11	0.9160
12	0.8640
13	0.9104
14	0.9075
15	0.9141
16	0.9248
17	0.9168
0.8897 (\pm 0.0609)	

Table 1: 3D correlation coefficient values calculated between 17 individual simulated nodules generated from spherical harmonics-based 3D shape reconstructions and the corresponding real patient CT lung nodules.

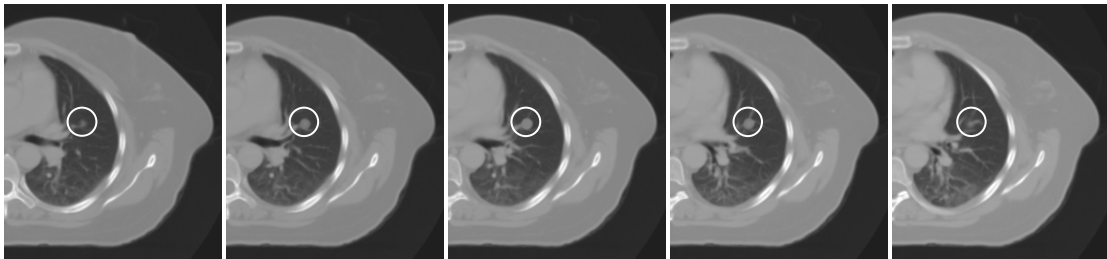


Figure 1: Five sequential CT slices illustrating the insertion of a simulated nodule (circled) with a vascularized border profile.

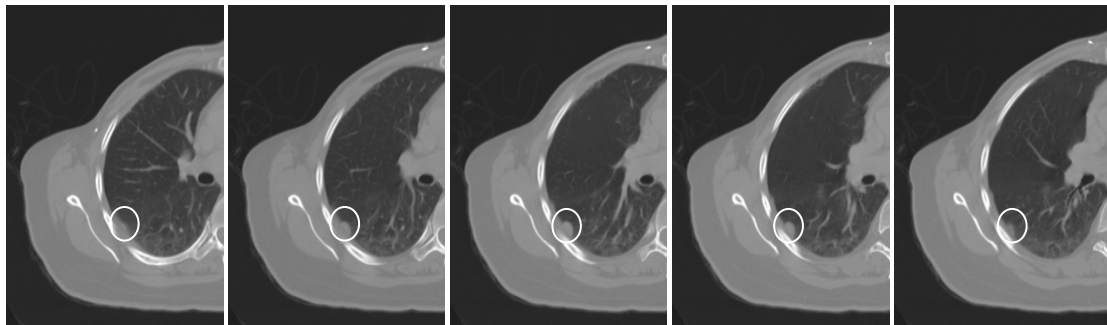


Figure 2: Five sequential CT slices illustrating the insertion of a simulated nodule (circled) with a juxtapleural border profile.