PROCEEDINGS OF SPIE

A new method for predicting CT lung nodule volume measurement performance

Ricardo S. Avila
Artit Jirapatnakul
Raja Subramaniam
David Yankelevitz

SPIE.

A New Method for Predicting CT Lung Nodule Volume Measurement Performance

Ricardo S. Avila*[a], Artit Jirapatnakul[b], Raja Subramaniam[b], David Yankelevitz[b]

[a]Accumetra, LLC, 7 Corporate Drive, Clifton Park, NY, USA 12065-8612; [b]Icahn School of Medicine at Mount Sinai, 1 Gustave L Levy Place, New York, NY USA 10029-6504

ABSTRACT

Purpose: To evaluate a new approach for predicting nodule volume measurement bias and variability when scanning with a specific CT scanner and acquisition protocol. Methods: A GE LightSpeed VCT scanner was used to scan 3 new rolls of 3M 3/4 x 1000 Inch Scotch Magic tape with a routine chest protocol (120 kVp, 100 mA, 0.4 s rotation, .98 pitch, STANDARD kernel) at three different slice thicknesses and spacings. Each tape scan was independently analyzed by fully automated image quality assessment software, producing fundamental image quality characteristics and simulated lung nodule volume measurements for a range of sphere diameters. The same VCT scanner and protocol was then used to obtain 10 repeat CT scans of an anthropomorphic chest phantom containing multiple Teflon spheres embedded in foam (diameters = 4.76mm, 6.25mm, and 7.94mm). The observed volume of the spheres in the 30 (3 reconstructions per scan) repeat scans was provided by independently developed nodule measurement software. Results: The predicted vs observed mean volume (mm^3) and CV for 3 slice thicknesses and sphere sizes was obtained. For 0.625mm slice thickness scans the predicted vs observed values were (44.3,0.91)-vs-(48.2,1.17), (110.4,0.51)-vs-(124.1,0.47), and (219.9,0.29)-vs-(250.1,0.34), for 4.76mm, 6.25mm, and 7.94mm spheres respectively. For 1.25mm slice thickness the corresponding values were (42.1,0.98)-vs-(47.6,1.35), (106.9,0.56)-vs-(123.1,0.61), and (214.8,0.32)-vs-(248.8,0.41). For 2.5mm slice thickness the corresponding values were (23.9,9.53)-vs-(36.8,12.50), (77.6,3.84)-vs-(110.5,3.20), and (173.0,1.57)-vs-(233.9,1.32). Conclusion: Volume measurement bias and variability for lung nodules based on nodule size and acquisition protocol can potentially be predicted using a new method that utilizes fundamental image characteristics and simulation.

Keywords: computed tomography, lung nodule, image quality, calibration, quantitative imaging

1. INTRODUCTION

The ability to accurately measure the size of sub-centimeter lung nodules found in low dose computed tomography (CT) scans is critical for delivering high quality lung cancer screening services. Screening sites must make sure that radiation dose delivered to patients is As Low As Reasonably Achievable (ALARA) [1] while maintaining sufficiently high CT image quality to perform nodule volume measurements with the necessary accuracy and precision. All CT scanners now provide physicians with tools to estimate the amount of radiation that will be delivered when a patient is scanned with a specific CT acquisition protocol. However, no corresponding tools are available that can estimate the image quality that will be achieved using a specific CT scanner and protocol, particularly in terms of the expected performance for a specific clinical task. There is great need for such a clinically focused tool given the fundamental tradeoff between image quality and radiation dose and the complex relationships between CT acquisition parameters and clinical task performance [2-6]. Here we explore this research area in the context of developing a set of methods that has the potential to provide physicians with an estimate of CT image quality in terms of CT lung nodule volume measurement performance.

Many research groups have studied medical image quality and have developed general image quality indicators [7] as well as metrics that correlate with clinical task performance [8]. The motivation here is to estimate medical image quality by predicting the statistical performance of a specific clinical task if it were to be performed numerous times with a specific acquisition system and analysis algorithm. Our view is that this prediction approach to task-based image quality performance assessment is more meaningful and directly useful than more general image quality indices that are designed to correlate with image quality. In addition, an image quality metric that predicts clinical task performance is more straightforward to evaluate as it allows for direct comparison of predicted and observed performance, as was done for this study.

The approach explored here is based on three high level processing steps. First, a core set of fundamental image quality characteristics is estimated for a CT scanner and acquisition protocol by scanning and analyzing a well characterized reference object. This analysis results in a first order mathematical approximation, or model, of the image acquisition system. Next, the resulting image acquisition model is combined with a set of virtual nodule models to simulate the production of numerous CT images containing the virtual nodules while allowing for variation in position, noise and other characteristics that are consistent with what would be encountered in real world scans. Finally, the nodules in the simulated images are segmented with a volume segmentation algorithm and measured. The resulting volume measurements are then analyzed to produce estimates of the expected volume measurement bias and precision of the target acquisition and measurement system. In all there are three image measurement sub-systems that are being modeled with this approach, an image acquisition model, a target object model, and a software analysis algorithm.

2. METHODS

2.1 Modeling Methods

The image acquisition model used in this study relied upon a core set of fundamental image characteristics. This included specifications for a 3D sampling rate, 3D resolution, CT linearity, and noise. The 3D sampling rate was expressed as the distances between voxels in X, Y, and Z in millimeters. 3D resolution was expressed as a 3D Gaussian Point Spread Function (PSF) with sigma values in millimeters for each 3D dimension (σ_x, σ_y, σ_z). CT linearity specified the amount of bias in the imaging system from the expected Hounsfield Units (HU) for known homogeneous materials without the influence of partial volume artifact. Image noise was represented as the standard deviation of HU densities for a homogeneous material.

The target object model used in this study was a solid lung nodule represented as a perfect sphere with a specified radius and position. In addition, the sphere model also specified the HU density of the sphere and the HU density of the surrounding material.

Simulation of the image formation process was performed by first creating a blank 3D image with the same 3D spacing as was used in the chest phantom protocol. Next the 3D PSF was convolved with the target object model representation, including expected HU nodule and background densities, to produce a simulated image with partial volume and sampling artifacts. Each time a simulated image was generated the exact position of the sphere model within the 3D voxel grid was permitted to slightly change randomly, which is also the case in real world scanning. Finally, random image noise was introduced into the image as additive white Gaussian noise.

The software analysis algorithm is the final sub-system modeled with this approach. The 3D segmentation algorithm used on the simulated nodule images was obtained by applying a constant threshold with the threshold set to half-way between the expected background and foreground HU densities. A 3D marching cubes algorithm was then used to obtain a polygonal closed surface from which the volume of the simulated nodule was calculated. Insight Segmentation and Registration Toolkit (ITK) algorithms were used for these segmentation and measurement calculations [9].

2.2 Performance Prediction Methods

A reference object was needed to estimate the image acquisition model characteristics. The approach taken here relied on CT scanning and analysis of rolls of 3M Scotch Magic™ tape, which serve as reference objects that are both low cost and ubiquitous. Most important is that the rolls of scotch tape exhibit a core set of geometric and material properties that are essential for understanding CT lung nodule volumetric performance.

Three new rolls of 3M 3/4 x 1000 Inch Scotch Magic tape were scanned on a GE LightSpeed VCT scanner with a routine chest protocol (120 kVp, 100 mA, 0.4 s rotation, .98 pitch, STANDARD kernel). Three slice thicknesses were used to reconstruct the acquired CT scotch tape data resulting in CT datasets with 0.65 mm, 1.25 mm, and 2.5mm slice thickness. Slice spacing was set to equal slice thickness for all CT datasets acquired in this study. **Figure 1** shows the placement of the scotch tape on the CT scanner table and the alignment of one roll of tape with CT scanner iso-center.

Each of the three CT tape scan datasets was independently analyzed by fully automated image quality assessment software. The tape analysis software detected the position and orientation of all scotch tape rolls in the scan data by searching the CT image data for objects with a combination of HU density, size, and shape characteristics. Next, for each roll of tape identified the analysis algorithm determined the region of the tape's core inner cylindrical ring that was not

compromised by a central plastic structure. A multiple pass optimization algorithm was then used to determine the 3D position, 3D orientation, and 3D Gaussian PSF sigmas (σ_x, σ_y, σ_z) that best fit the actual scotch tape scan data where σ_x = σ_y. The use of a single value for in-plane resolution is common in CT calibration and justified as almost all thoracic CT scans are currently acquired with a helical scanning approach. The cylindrical shape of the tape and the orientation of the tape as it lays flat on the CT table allows for the estimation of an in-plane σ_x and σ_y as well as σ_z.

The measured HU for tape and air were calculated as the mean HU within the scotch tape material as well as the air region at the center of the scotch tape core taking special care not to allow partial volume artifacts to influence the measurements (based on the 3D PSF). Thus, the bias was calculated as the mean HU minus the expected HU for air and tape materials. Image noise was represented as the standard deviation of the two materials using the same region of support as the measured mean HU values. Many other image quality properties were calculated from each tape scan, such as levels of edge enhancement, but were not used in this analysis.
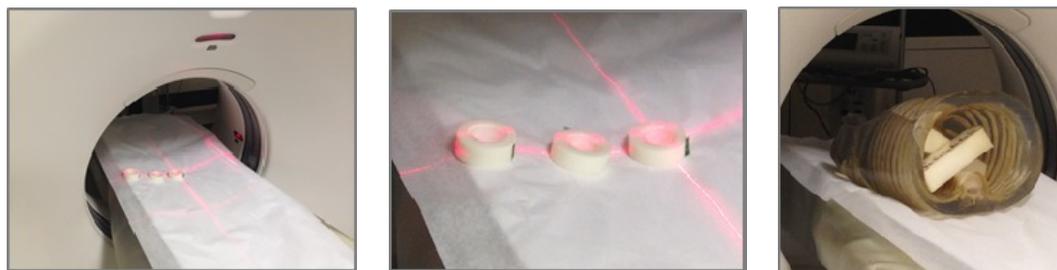
Figure 1: Three rolls of scotch tape were placed on the table of a GE LightSpeed VCT scanner and scanned with a standard chest protocol (left and middle). This scan was analyzed and used to predict volumetric sphere measurement performance with a fully automated modeling and simulation method. An anthropomorphic chest phantom containing Teflon spheres was scanned multiple times with the same chest protocol (right) and independent software measured the volume of the spheres.

Three different nodule models were used in this study spanning a range of Teflon nodule diameters (4.0mm, 6.0mm, and 8.0mm). As the goal was to predict measurement performance for Teflon spheres surrounded by low density foam, the HU sphere and background densities were all set to expected values for the Teflon (HU = 870) and foam (HU = -973) materials. Although bias values were calculated for air and tape materials, they were not used for image simulation. CT bias for all materials was assumed to be zero.

Simulated images were generated, automatically segmented, and volumetrically measured for each of three nodule model diameters and CT slice thicknesses used in the study. Mean, standard deviation, and coefficient of variation statistics were calculated using the volume measurements for each nodule size and slice thickness pair using interpolation between the simulated nodule diameter values, establishing the measurement performance prediction. Each mean and CV pair for each of the three nodule model diameters was calculated using 150 simulated images.

2.3 Observed Performance Methods

10 repeat CT scans of an anthropomorphic chest phantom containing multiple Teflon spheres embedded in foam (diameters = 4.76mm, 6.25mm, and 7.94mm) were acquired using the same VCT scanner and acquisition protocol as the scotch tape scan. An independently developed nodule measurement algorithm [10] provided by AJ and based on constant thresholding was obtained and used to measure the volume of the Teflon spheres in the 30 (3 reconstructions per scan) repeat scans. This segmentation algorithm used the same threshold as the prediction algorithm but was not based on ITK and used different algorithms for thresholding and measuring the spheres. Mean, standard deviation, and coefficient of variation statistics were calculated using these volume measurements, thus establishing the observed measurement performance.

3. RESULTS

The predicted and observed volumetric measurement statistics for three different Teflon sphere diameters and three different CT slice thicknesses are summarized in **Table 1**. For each of these a pair of values is reported representing the volumetric measurement mean and CV percentage.

Table 1. Predicted versus observed volumetric measurements. Each pair of values represents the mean volume (mm^3) and the coefficient of variation (%) of the volumetric measurements.

Slice Thickness	Sphere Diameter					
	4.76 mm		**6.25 mm**		**7.94 mm**	
	Predicted	**Observed**	**Predicted**	**Observed**	**Predicted**	**Observed**
0.625 mm	44.3, 0.91	48.2, 1.17	110.4, 0.51	124.1, 0.47	219.9, 0.29	250.1, 0.34
1.25 mm	42.1, 0.98	47.6, 1.35	106.9, 0.56	123.1, 0.61	214.8, 0.32	248.8, 0.41
2.5 mm	23.9, 9.53	36.8, 12.50	77.6, 3.84	110.5, 3.20	173.0, 1.57	233.9, 1.32

Graphs showing predicted versus observed volumetric performance are shown in the next two figures. **Figure 2** shows the mean values of the predicted and observed volume measurements.
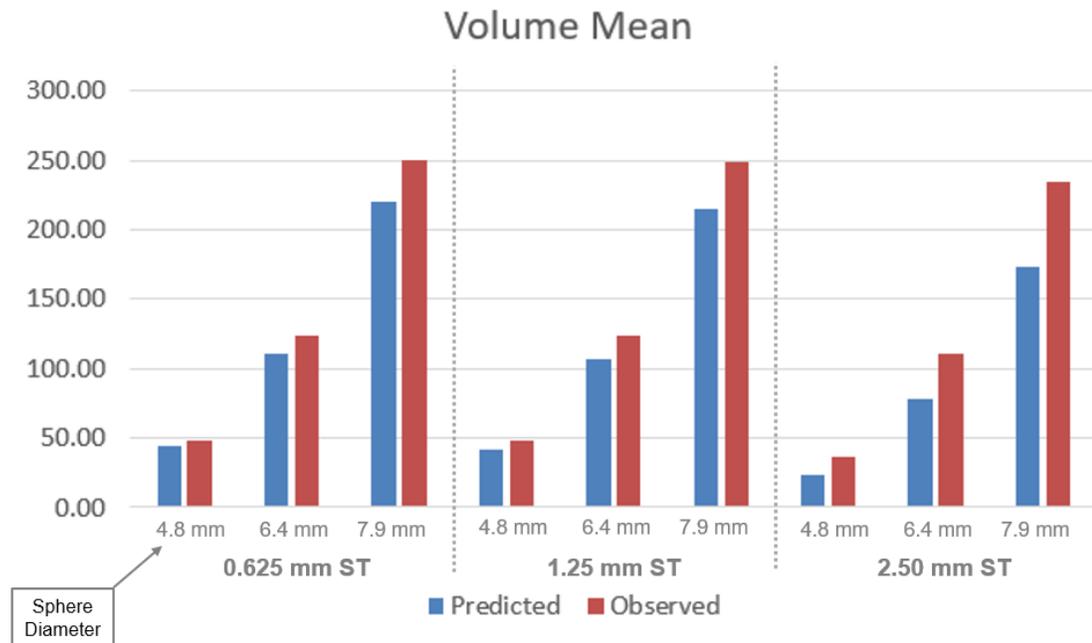
Figure 2: Comparison of predicted (blue) and observed (red) mean volumetric measurements of Teflon spheres repeat CT scanned 10 times within an anthropomorphic chest phantom. Three different sphere diameters and three different slice thicknesses were used.

Figure 3 shows the corresponding CV percentage values for the predicted and observed volume measurements.
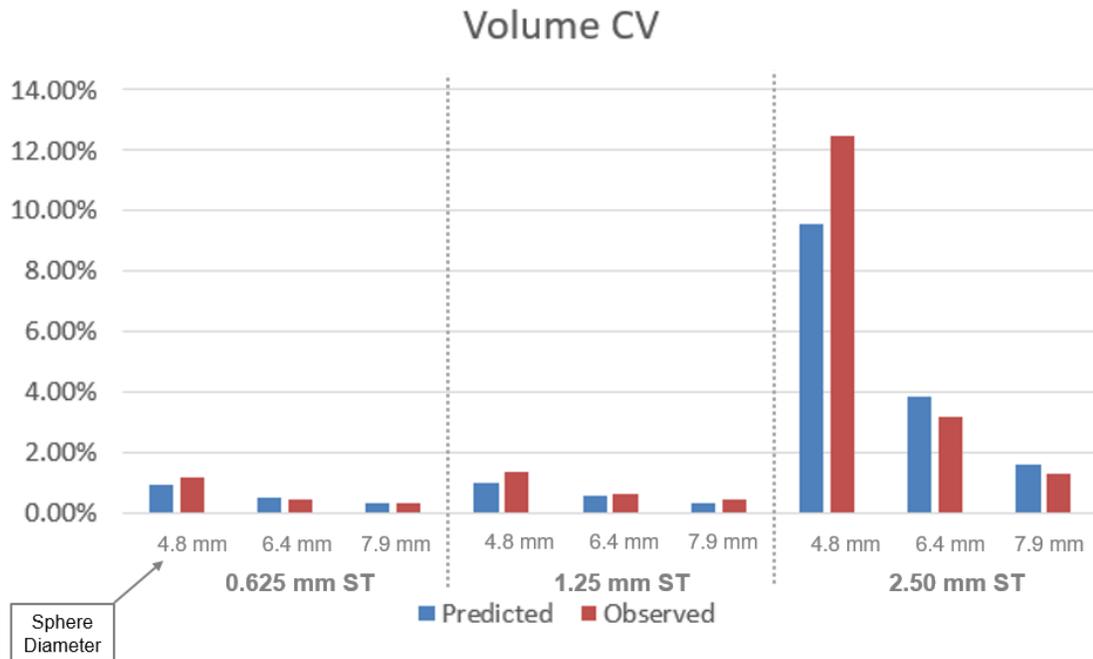
Figure 3: Comparison of predicted (blue) and observed (red) volumetric measurement Coefficient of Variation (CV) of Teflon spheres repeat CT scanned 10 times within an anthropomorphic chest phantom. Three different sphere diameters and three different slice thicknesses were used. Note that predicted vs observed remained similar despite the large range in requested slice thickness (0.625mm to 1.25mm to 2.5mm).

4. DISCUSSION

This study demonstrated several important capabilities and advancements. First, is that the estimation of a small set of fundamental image characteristics can be used with an image generation simulation engine to predict the performance statistics of a specific clinical task. Similar to radiation dose estimates provided by CT scanners, which are also based on modeling and assumptions, we obtained volume performance predictions that were consistent with the observed volume sphere measurements that were scanned in an anthropomorphic chest phantom. It is particularly important that the approach expressed predicted performance for a specific clinical task as physicians need to understand expected image quality in clinically meaningful terms. Second is that easy to obtain and inexpensive reference objects, specifically three rolls of scotch tape, can be used to characterize an image acquisition system. Overall, the data illustrates that a small but fundamental set of image quality characteristics obtained using a low cost reference object can be used to characterize the expected performance of a medical scanning and measurement system.

We observed a consistent underestimation of mean and CV volume measurement statistics, particularly for the thickest slice thickness used. This indicates that better modeling of each of our three modeling sub-systems has the potential to improve prediction performance. Use of a more sophisticated noise model and a closer match of the processing components used for the predicted and observed segmentation algorithms are some of the steps needed to obtain better task performance prediction values. Also, a more advanced method for modeling the PSF along the Z axis would improve prediction performance for thicker slice thicknesses as a Gaussian PSF model is not always representative. Although a simple sphere model was used for this study, the approach should extend to more complex nodule and patient presentations. With further study, we will determine how much more sophisticated the modeling methods will need to be to reasonably characterize a wide range of scanners, patient presentations, and analysis algorithms.

It should also be noted that better prediction performance assessment methods are also needed. The use of more than 10 repeat scans for each nodule size and slice thickness pair would have improved the ability to characterize observed

performance. However, this is challenging to achieve as the CT acquisition time needed for high levels of repeat scanning is not available at most clinical institutions.

Despite the modeling limitations, the prediction results demonstrated that fundamental performance characteristics can inform prediction performance. A naïve approach to performance prediction would have assumed that a slice thickness of 1.25 mm would have significantly higher volumetric measurement CV than a 0.625 mm slice thickness. However, the analysis algorithm found that the Z resolution for these two slice thicknesses was in actuality fairly similar. Because the simulation engine used the actual resolution estimated from the tape measurements the predicted performance was able to closely resemble the observed performance. Had our approach relied upon a Z resolution estimate based on the requested slice thickness in the DICOM header, prediction performance would very likely not have been as good as found in this study. The approach we propose here, namely to estimate and use fundamental image characteristics, has the potential to offer resilience with many other scan parameters beyond CT slice thickness.

It should also be noted that this approach can potentially support many other clinical tasks and modalities beyond the CT lung nodule measurement task demonstrated here. One of the most useful is the potential ability to support the quantitative optimization of both radiation dose and task specific image quality, a task that is currently performed subjectively even when scanning is done for advanced computational analysis methods with poorly understood image quality requirements.

5. CONCLUSION

The bias and variability of lung nodule volume measurements can potentially be predicted using a new method that utilizes fundamental image characteristics and simulation.

REFERENCES

[1] ACR, "ACR Statement on FDA Radiation Reduction Program," https://www.acr.org/About-Us/Media-Center/Position-Statements/Position-Statements-Folder/ACR-Statement-on-FDA-Radiation-Reduction-Program. Position Statement, (2011).
[2] Ravenel, J.G., Leue, W.M., Nietert, P.J., Miller, J.V., Taylor, K.K., Silvestri, G.A., "Pulmonary nodule volume: effects of reconstruction parameters on automated measurements--a phantom study," Radiology; 247:400-8, (2008).
[3] Das, M., Ley-Zaporozhan, J., Gietema, H.A., et al., "Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners," EurRadiol. ;17:1979-84, (2007).
[4] Petrou, M., Quint, L.E., Nan, B., Baker, L.H., "Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology," AJR Am J Roentgenol;188:306-12, (2007).
[5] Wielpütz, M.O., Lederlin M., Wroblewski, J., et al., "CT volumetry of artificial pulmonary nodules using an ex vivo lung phantom: influence of exposure parameters and iterative reconstruction on reproducibility," Eur J Radiol;82:1577-83, (2013).
[6] Wang, Y., de Bock, G.H., van Klaveren, R.J., et al., "Volumetric measurement of pulmonary nodules at low-dose chest CT: effect of reconstruction setting on measurement variability," EurRadiol; 20:1180-7, (2010).
[7] Goldman, L.W., "Principles of CT: radiation dose and image quality," J Nucl Med Technol. ;35(4):213-25; quiz 226-8, (2007).
[8] Christianson, O., Chen, J.J., Yang, Z., Saiprasad, G., Dima, A., Filliben, J.J., Peskin, A., Trimble, C., Siegel, E.L., Samei, E., "An Improved Index of Image Quality for Task-based Performance of CT Iterative Reconstruction across Three Commercial Implementations," Radiology.; 275(3):725-34, (2015).
[9] Yoo, T.S., Metaxas, D.N. "Open science—combining open data and open source software: medical image analysis with the Insight Toolkit," Med Image Anal. 9 (6): 503–6. doi:10.1016/j.media.2005.04.008. PMID 16169766, (2005).
[10] Reeves, A.P., Chan, A.B., Yankelevitz, D.F., Henschke, C.I., Kressler, B., Kostis, W.J., "On measuring the change in size of pulmonary nodules," IEEE Trans Medical Imaging, 25: 435-50, (2006).