The Lung Image Database Consortium (LIDC): Pulmonary nodule measurements, the variation and the difference between different size metrics

Anthony P. Reeves,[a] Alberto M. Biancardi,[a] Tatiyana V. Apanasovich,[b] Charles R. Meyer,[c] Heber MacMahon,[d] Edwin J.R. van Beek,[e] Ella A. Kazerooni,[c] David Yankelevitz,[f] Michael F. McNitt-Gray,[g] Geoffrey McLennan,[h] Samuel G. Armato III,[d] Denise R. Aberle,[g] Claudia I. Henschke,[f] Eric A. Hoffman,[e] Barbara Y. Croft,[i] and Laurence P. Clarke[i]

[a]Electrical and Computer Engr., Cornell University, Ithaca, NY, USA;
[b]Op Research and Industrial Engr., Cornell University, Ithaca, NY, USA;
[c]Department of Radiology, The University of Michigan, Ann Arbor, MI, USA;
[d]Department of Radiology, The University of Chicago, Chicago, IL, USA;
[e]Department of Radiology, University of Iowa, Iowa City, IA, USA;
[f]Weill Medical College, Cornell University, New York, NY, USA;
[g]Department of Radiological Sciences, UCLA, Los Angeles, CA, USA;
[h]Medicine and Biomedical Engr., University of Iowa, Iowa City, IA, USA;
[i]Cancer Imaging Program, National Cancer Institute, Bethesda, MD, USA

ABSTRACT

Size is an important metric for pulmonary nodule characterization. Furthermore, it is an important parameter in measuring the performance of computer aided detection systems since they are always qualified with respect to a given size range of nodules. The first 120 whole-lung CT scans documented by the Lung Image Database Consortium using their protocol for nodule evaluation were used in this study. For documentation, each inspected lesion was reviewed independently by four expert radiologists and, when a lesion was considered to be a nodule larger than 3 mm, the radiologist provided boundary markings in each image in which the nodule was contained. Three size metrics were considered: a uni-dimensional and a bi-dimensional measure on a single image slice and a volumetric measurement based on all the image slices. In this study we analyzed the boundary markings of these nodules in the context of these three size metrics to characterize the inter-radiologist variation and to examine the difference between these metrics. A data set of 63 nodules each having four observations was analyzed for inter-observer variation and an extended set of 252 nodules each having at least one observation was analyzed for the difference between the metrics. A very high inter-observer variation was observed for all these metrics and also a very large difference among the metrics was observed.

Keywords: Quantitative image analysis, X-ray CT, Detection, Nodule measurement, Size metrics

1. INTRODUCTION

Accurate and reliable measurement of pulmonary nodule size from CT scans has an important role in computer assisted evaluation of lung lesions. It is a key factor in the diagnosis of lung cancer as the estimation of nodule growth rates serves as a predictor of malignancy; size change can also be used to assess the efficacy of a therapeutic treatment. It is also a critical aspect of computer assisted diagnosis (CAD) systems since those systems, and in particular their detection sub-systems, are always qualified with respect to a given size range of nodules. The overall approach involves using a subset of a documented image database with respect to a starting operational size.

In the context of spatial extent for a three-dimensional object without restrictions on shape, the size is best expressed by the volume occupied by that object. Other important considerations in choosing a method for

Send correspondence to A.P.Reeves e-mail:reeves@ece.cornell.edu, phone:+1.607.255.2342, fax:+1.607.255.9072

Medical Imaging 2007: Computer-Aided Diagnosis, edited by Maryellen L. Giger, Nico Karssemeijer,
Proc. of SPIE Vol. 6514, 65140J, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.713672

estimation of size also include the imaging modality and the time available to the physician. Image modalities may be only two-dimensional or highly anisotropic with regards to the third dimension. Manually measuring the lesion volume involves inspecting all images that involve the lesion and is very time consuming; the current clinical practice is to estimate lesion size by making a single linear measurement in just one image of the lesion.

There is active interest in developing computer assisted methods that will aid the physician in measuring the size of lesions using volumetric methods ;[1–6] the challenge here is how to calibrate and validate such methods. For images of real lesions the only accepted method to establish their size is based on annotations performed by expert radiologists. In this paper we explore the variation between expert radiologists for the volumetric measurement task and consider the relationship between the three-dimensional volumetric measurements and standard measurement methods currently used by physicians that involve a single two-dimensional image.

To provide a standard method for lesion size measurement, the World Health Organization proposed in 1979 the use of the product of the maximal diameter and its longest perpendicular,[7] while the Response Evaluation Criteria in Solid Tumors (RECIST) working group in 1998–2000 proposed the use of the (uni-dimensional) maximal diameter as a more efficient standard estimator of lesion volume.[8]

In 2000 the National Institutes of Health launched a cooperative effort, known as the Lung Image Database Consortium,[9] to construct a set of annotated CT scans for the development and the evaluation of different CAD approaches. The LIDC developed a pulmonary nodule documentation process where expert radiologists marked the visible lesion boundary belonging to each lesion in all the relevant axial images.

The LIDC documentation process did not require the expert radiologist to provide either uni- or bi-dimensional measurement, a technique commonly used in clinical practice. However, given the full boundary of the lesion as marked by the radiologist we used computer algorithms to apply the rules from RECIST (uni-dimensional) and WHO (bi-dimensional) to provide estimates for these measurements. We also computed volumetric measurements based on the boundary documentation. A previous study[10] compared five different measurement methods for change in lesion size between scans on CT images of liver metastases. The authors concluded that the three-dimensional methods were a viable alternative to two dimensional methods. Other studies on size change for treatment-response assessment in lung cancer compared uni- and automated three-dimensional measures,[11] uni-, bi- and three-dimensional measures,[12,13] and manual bi-dimensional measures with an automated contour technique.[14] The studies that involved volume measurements concluded that there was poor agreement between three-dimensional measures and single-image based methods. In this paper we are primarily interested in studying the absolute size of nodules, especially given its relevance to the CAD community, we analyze the variation between the expert markings and we compare the estimated RECIST and WHO measurement methods with the 3D volume measurements.

2. MATERIALS AND METHODS

The evaluation of the impact of different size metrics was performed on whole-lung CT scans that were documented by the Lung Image Database Consortium (LIDC).[15] As per the LIDC process model, each scan was assessed by 4 board-certified thoracic radiologists. The radiologists were presented with detailed instructions to either mark the central location of a small nodule (less than 3mm in diameter) or, for larger nodules, to mark the entire boundary, in all the relevant axial scan images. The boundary was chosen to be those pixels that were just outside the region of the nodule. All the markings are stored as boundary pixels lying on the axial image planes. This marking is illustrated in Figure 1. The marked image is shown in 1 b with the marker's boundary points shown in white and the region designating the the lesion marked in black. The National Cancer Imaging Archive (NCIA) repository of the National Cancer Institute[16] makes available the CT scans that have been fully documented by the LIDC together with XML documentation files that contain the boundary points chosen by all of the expert radiologists.

All the scans and the XML files that were currently available were imported and parsed to extract the outline information and to determine the unique Regions of Interest (ROIs) of each physical nodule. As each radiologist marking is described by a set of boundaries, every set was mapped to a 3D image and filled according to the LIDC process definitions (the black inner region in figure 1 b) to get a 3D model of each radiologist-outlined nodule.

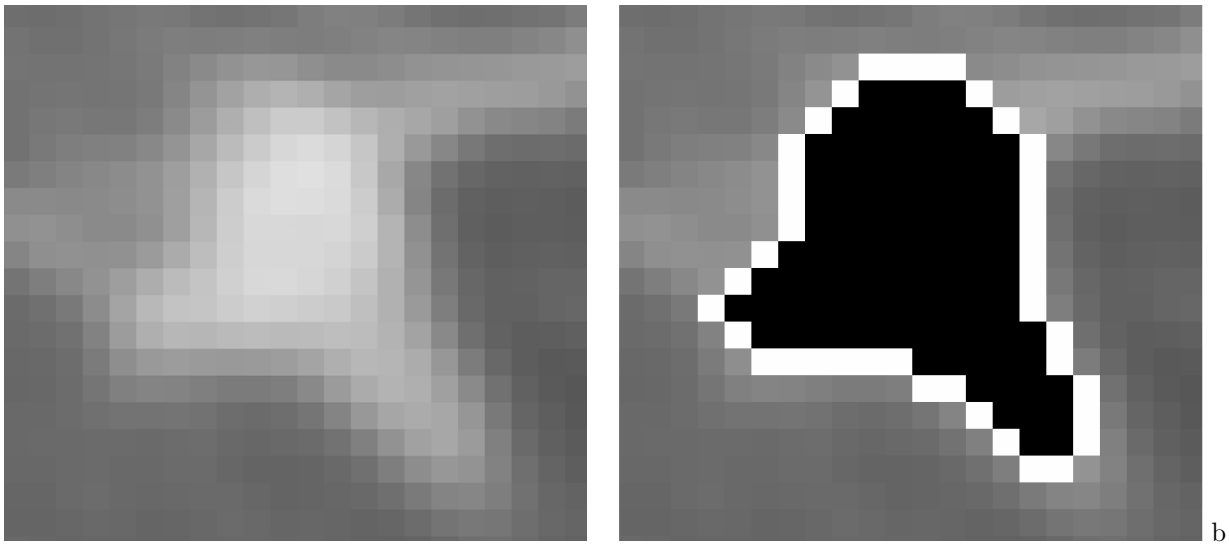


Figure 1. An example of the LIDC rules in documenting nodules. On the right, the raw scan data is presented. On the left, the white boundary shows the actual boundary drawn by the radiologist that encloses the black inner region belonging to the nodule.

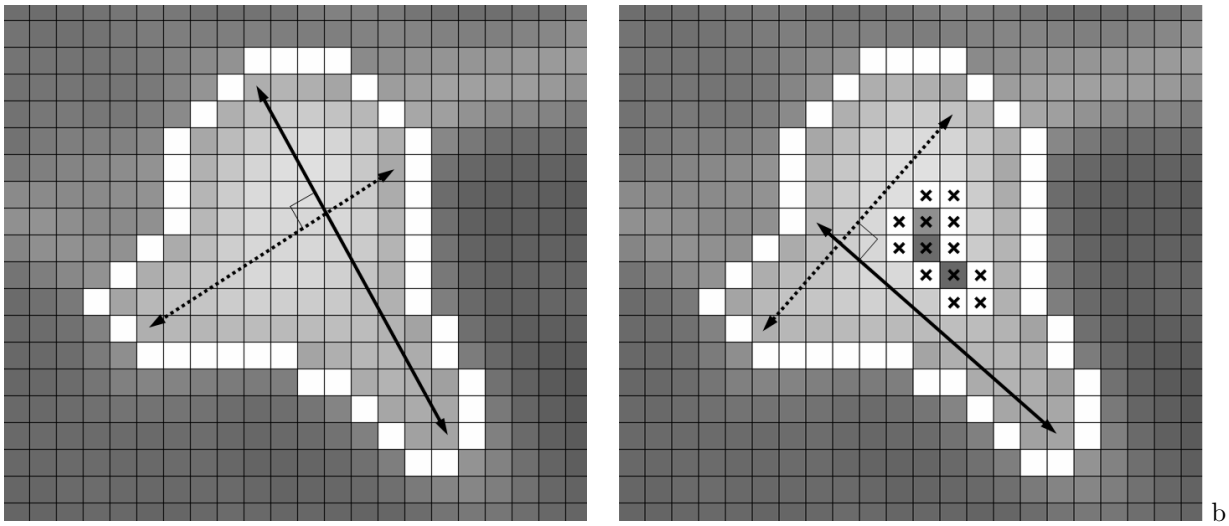


Figure 2. This figure, on the right, describes graphically how the diameter and its longest perpendicular are computed as surrogates of radiologist actions. On the left, if the region with the crossed pixel were to be hypothetically removed from the actual nodule region, then the previous diameter would not be valid any longer and the new diameter with the relative longest perpendicular would have to be determined.

Next the 3D images were used to compute the values of the metrics under investigation: nodule volume, largest diameter and the product of the largest diameter with its longest perpendicular. The total lesion volume, as in many CAD tools, is estimated by counting the number of nodule pixels in each of the image slices and then multiplying their sum by the unit volume.[17] The largest diameter is determined as the maximum diameter, i.e. the longest rectilinear segment that completely lies within the nodule region (see figure 2), among all the axial-planar subsets of the nodule; this measure is similar and, except for 8 out of 252 nodules, meets the guidelines (especially with respect to Appendix I) of RECIST.[8] The 8 nodules were technically too small with respect to slice thickness to meet the RECIST guidelines for measurement.

The computer estimation of the RECIST measurement is illustrated in figure 2 a. The solid line is the longest diameter that can be placed in any axial image within the marked boundary. For the WHO measurement the longest perpendicular is also computed. Some lesions have cavities or holes in them and these are also marked in the LIDC database. As a hypothetical example, if the radiologist had marked the pixels within the nodule shown white with an x in figure 2 b, this would imply a hole for that region. For the RECIST criterion the diameter must be within the lesion and not include the cavity; therefore, the dimension shown by the solid line would be the newly computer-determined RECIST measurement for that lesion. For the WHO criterion, a product is obtained by multiplying the largest diameter with longest perpendicular.

Finally, for our statistical evaluation, each measure is made equivalent and directly comparable to the others by expressing its value in terms of the diameter of an equivalent sphere, i.e. for any given nodule volume estimate the diameter of a sphere with the same volume is calculated using $2\sqrt[3]{\frac{3v}{4\pi}}$, where the volume of the equivalent sphere is $v = \frac{4}{3}\pi(\frac{d}{2})^3$, while the bi-dimensional metric is equal to a diameter of $2\sqrt{\frac{s}{\pi}}$, where s is the product of length and width.

3. THE DATA-SET

The processing steps described in the previous section were carried out on 120 documented whole-lung CT scans. All the scans were acquired from multi-detector row CT scanners with pixel resolutions ranging from 0.54 to 0.74 mm and an axial resolution ranging from 0.75 to 3.00 mm. The tube current ranged mainly from 40 to 260 mA except for 6 scans where it was between 264 and 486 mA; tube voltage range was either 120 or 140 kVp.

The whole dataset contained 252 lesions for which boundaries were marked. Of these 252 lesions only 63 were considered to be nodules greater than 3 mm by all four radiologists and, therefore, had four sets of boundary markings available. The full set of 252 lesions was used for the comparison between metrics while only the 63 cases with all four markings were used for our analysis of between reader variation. When multiple markings were available for a nodule, the median value of each size metric from these markings was used to represent the size for that nodule. Two nodules with diameters derived from three-dimensional measures were greater than 30 mm, 36 and 50 mm respectively, and were not considered in the analysis.

4. RESULTS

The between reader variation was analyzed for the 63 nodules with four readings each. We computed means and standard deviations based on those measurements. Figure 3 shows computed smoothed estimators of the variation as a function of the measurement for the three metrics. It is clear that the uni-dimensional and bi-dimensional size metrics have a slightly higher inter-observer variation than the three-dimensional one. For the three-dimensional 95% of estimated standard deviations were within the following range [0.4684, 1.1230]. For the uni-dimensional and bi-dimensional size metric most of the standard deviations (95%) were within the range [0.7405, 1.9130] and [0.9023, 1.7064] respectively.

For the three-dimensional derived diameters, all were within two standard deviations while six nodules (10%) and eleven nodules (18%) had a standard deviation greater than two for the uni-dimensional and bi-dimensional metric respectively. As an example, for the three-, uni- and bi-dimensional size measurements of 10mm, the estimated standard deviations were 1.01, 1.08 and 1.34 respectively.

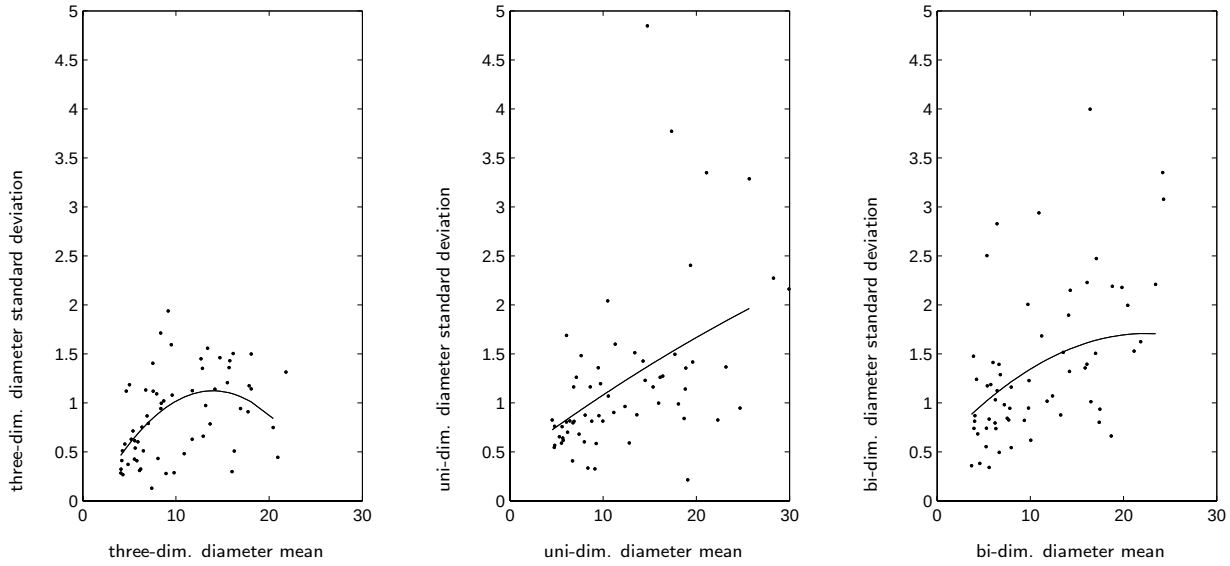


Figure 3. Scatter plot of the standard deviation versus means of four experts' measurements along with a nonparametric regression curve for volume(left), uni-dimensional size(center), and bi-dimensional(right) estimates

One reason for this high variability is that, when the nodule has a complex shape, each radiologist may mark some boundary pixels as belonging or not belonging to the nodule region. This in turn is reflected into diameters and perpendicular that can show significant differences in length as shown in Figure 4.

For the metric comparison we used the kernel estimate of the conditional distribution of the volume measurement given the uni-dimensional and bi-dimensional sizes.[18,19] From estimates of the conditional distribution, estimates of conditional highest density regions (HDR) were computed. HDRs are the most appropriate subset to use to summarize a multi-modal and non-symmetrical probability distribution. Figure 5 shows HDRs for volume with 95 and 99 percent coverage conditional on the uni-dimensional and bi-dimensional size measurements. The conditional medians are marked with dots. For example, if a radiologist reports a uni-dimensional size measurement of 10mm, then the regions of probability coverage 0.95 and 0.99 with the smallest extent for the volume measurement are the intervals [3.48, 11.96] and [2.32, 13.26] respectively. If the 10mm measure was coming from a bi-dimensional measurement, then the region intervals would be [4.88, 12.25] and [4.00, 13.06] respectively. Note that these intervals are wider than expected just due to the measurement error and not centered at 10.

Poor agreement between the three-dimensional and two-dimensional methods is commonly seen when the nodule does not conform to the approximately spherical assumption. Examples of this are shown in Figures 6

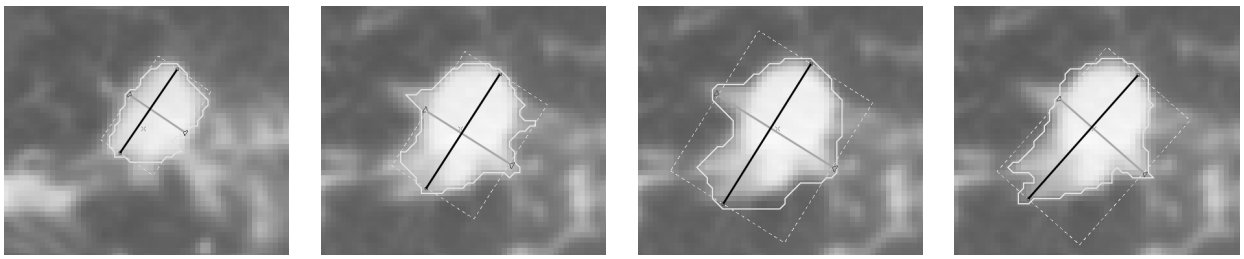
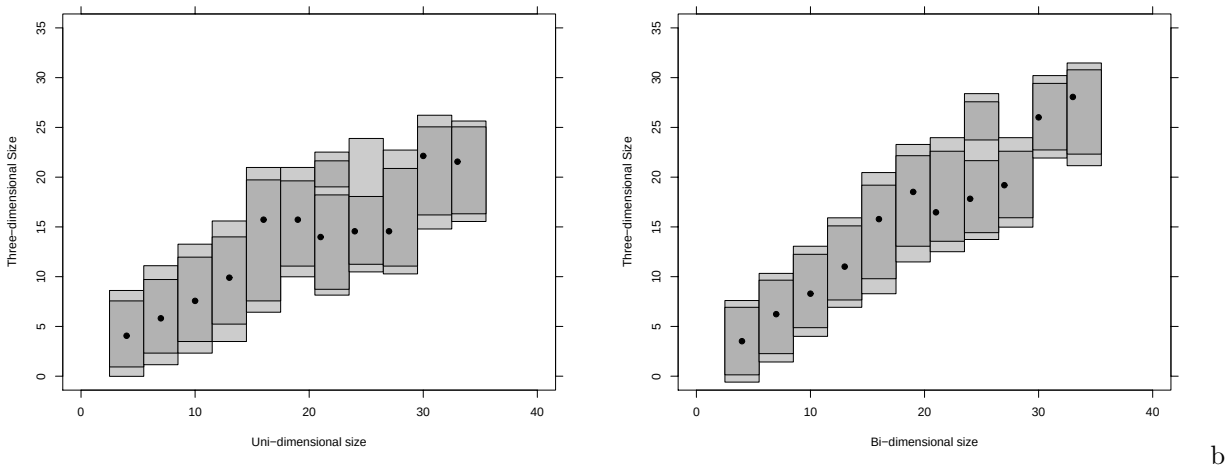


Figure 4. An example of variability among radiologists. Each image shows the slice where the largest diameter (dark line) and longest perpendicular (gray line) were determined according to the markings provided by each radiologist. The image on the left is on a different slice than the other three.



a b

Figure 5. 95% and 99% HDRs for the volume size estimate conditional on the uni-dimensional metric (a) and on the bi-dimensional metric (b)

and 7. In Figure 6 the extent of the lesion is greatest aligned with the axial dimension, hence the two dimensional measurement under-estimates the three-dimensional derived diameter, while in Figure 7 the greatest extent of the nodule is perpendicular to the axial dimension, hence the two-dimensional measurement over-estimates the three-dimensional derived diameter. In general, for the uni-dimensional metric we anticipated an over estimation of the three-dimensional value since it is a measure of maximal extent rather than mean extent in two dimensions.

5. CONCLUSIONS

The preliminary analysis preformed on a set of 120 documented whole-lung scans focusing on the metric evaluation of radiologist reading highlighted two main findings. The first result, in concordance with literature,[11, 20, 21] is that reader subjectivity on boundary locations propagates into a very large inter-observer variation of the size estimates for pulmonary nodules (in the order of 5 to 20% relative to the measured size) for uni-dimensional, bi-dimensional and volumetric size measurement methods.

The second result is that the difference between the uni-dimensional and three-dimensional measurements and between the two-dimensional and the three-dimensional measurements are also very large. While this difference may play a lesser role in size change with time estimation of nodules, it has important implications for the

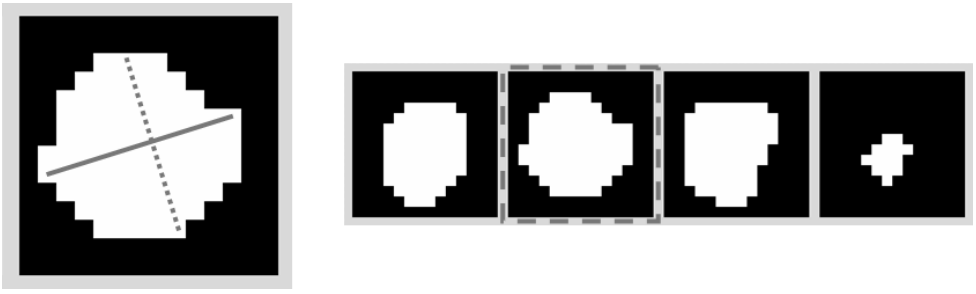


Figure 6. A selected case where the three-dimensional size (10.0mm) is greater than the uni-dimensional (8.3mm) and bi-dimensional (9.0mm) sizes. The frame with dashed boundary is enlarged on the left hand of the figure to show the largest diameter (solid line) and its longest perpendicular (dotted line).

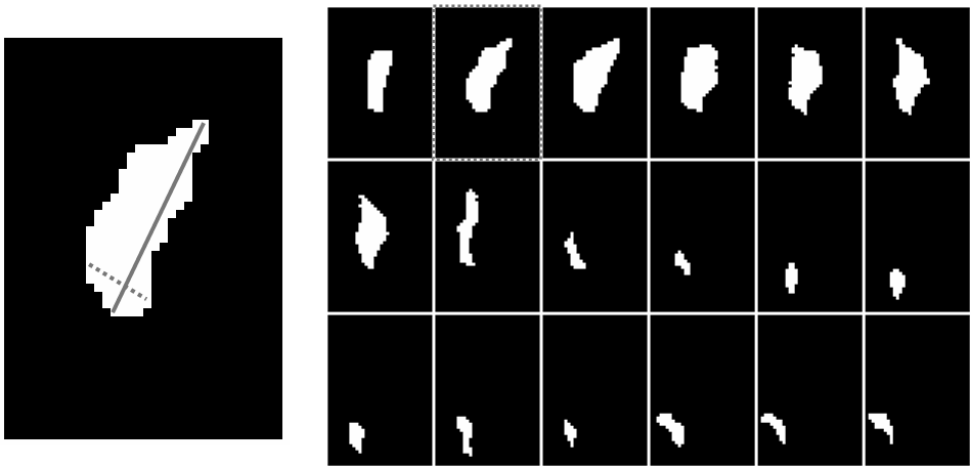


Figure 7. A selected case where the three-dimensional size (10.6 mm) is smaller than the uni-dimensional (21.7 mm) and bi-dimensional (15.9 mm) sizes. The frame with dotted boundary is enlarged on the left hand of the figure to show the largest diameter (solid line) and its longest perpendicular (dotted line).

selection of nodules from a database. This impacts the performance evaluation of the detection stage of a CAD system since using a different size metric will result in a significantly different set of nodules that are considered to be the ground truth for detection.

REFERENCES

1. J. P. Ko, H. Rusinek, E. L. Jacobs, J. S. Babb, M. Betke, G. McGuinness, and D. P. Naidich, "Small pulmonary nodules: Volume measurement at chest CT – phantom study," *Radiology* **228**, pp. 864–870, September 2003.
2. J.-M. Kuhnigk, V. Dicken, L. Bornemann, D. Wormanns, S. Krass, and H.-O. Peitgen, "Fast automated segmentation and reproducible volumetry of pulmonary metastases in CT-scans for therapy monitoring," in *Lecture Notes in Computer Science*, **3217**, pp. 933–941, Medical Image Computing and Computer-Assisted Intervention, Springer-Verlag GmbH, 2004.
3. K. Okada, D. Comaniciu, and A. Krishnan, "Robust anisotropic gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT," *IEEE Transactions on Medical Imaging* **24**, pp. 409–423, March 2005.
4. L. R. Goodman, M. Gulsun, L. Washington, P. G. Nagy, and K. L. Piacsek, "Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements," *American Journal of Roentgenology* **186**, pp. 989–994, April 2006.
5. M.-P. Revel, A. Merlin, S. Peyrard, R. Triki, S. Couchon, G. Chatellier, and G. Frija, "Software volumetric evaluation of doubling times for differentiating benign versus malignant pulmonary nodules," *American Journal of Roentgenology* **187**, pp. 135–142, July 2006.
6. A. Reeves, A. Chan, D. Yankelevitz, C. Henschke, B. Kressler, and W. Kostis, "On measuring the change in size of pulmonary nodules," *IEEE Transactions on Medical Imaging* **25**, pp. 435–450, April 2006.
7. "WHO handbook for reporting results of cancer treatment." World Healt Organization Offset Publication No. 48, 1979.
8. P. Therasse, S. Arbuck, E. Eisenhauer, J. Wanders, R. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A. van Oosterom, M. Christian, and S. Gwyther, "New guidelines to evaluate the response to treatment in solid tumors," *J. Natl. Cancer Inst.* **92**, pp. 205–216, February 2000.
9. National Institute of Healt, "Lung image database resource for imaging research." http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-01-001.html, April 2000. Accessed Jan 22nd, 2007.

10. L. Van Hoe, E. Van Cutsem, I. Vergote, A. L. Baert, E. Bellon, and P. Dupont, "Size quantification of liver metastases in patients undergoing cancer treatment: reproducibility of one-, two-, and three-dimensional measurements determined with spiral CT," *Radiology* **202**, pp. 671–675, 1997.
11. K. Marten, F. Auer, S. Schmidt, G. Kohl, E. J. Rummeny, and C. Engelke, "Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria," *European Radiology* **16**, pp. 781–790, April 2006.
12. L. N. Tran, M. S. Brown, J. G. Goldin, X. Yan, R. C. Pais, M. F. McNitt-Gray, D. Gjertson, S. R. Rogers, and D. R. Aberle, "Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography," *Acad. Radiol.* **11**(12), pp. 1355–1360, 2004.
13. S. G. Jennings, H. T. Winer-Muram, R. D. Tarver, and M. O. Farber, "Lung tumor growth: Assessment with CT – comparison of diameter and cross-sectional area with volume measurements," *Radiology* **231**, pp. 866–871, June 2004.
14. L. H. Schwartz, M. S. Ginsberg, D. DeCorato, L. N. Rothenberg, S. Einstein, P. Kijewski, and D. M. Panicek, "Evaluation of tumor measurements in oncology: Use of film-based and electronic techniques," *Journal of Clinical Oncology* **18**, pp. 2179–2184, May 2000.
15. S. G. Armato, G. McLennan, M. F. McNitt-Gray, C. R. Meyer, D. Yankelevitz, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. Kazerooni, H. MacMahon, A. P. Reeves, B. Y. Croft, L. P. Clarke, and the Lung Image Database Consortium Research Group, "Lung image database consortium: developing a resource for the medical imaging research community," *Radiology* **232**(3), pp. 739–748, 2004.
16. National Cancer Institute, "National cancer imaging archive." https://imaging.nci.nih.gov/ncia/. Accessed Jan 22nd, 2007.
17. R. S. Breiman, J. W. Beck, M. Korobkin, R. Glenny, O. E. Akwari, D. K. Heaston, A. V. Moore, and P. C. Ram, "Volume determinations using computed tomography," *American Journal of Roentgenology* **138**(2), pp. 329–333, 1982.
18. R. J. Hyndman, D. M. Bashtannyk, and G. K. Grunwald, "Estimating and visualizing conditional densities," *Journal of Computational and Graphical Statistics* **5**, pp. 315–336, December 1996.
19. R. J. Hyndman, "Computing and graphing highest density regions," *The American Statistician* **50**, pp. 120–126, May 1996.
20. N. R. Bogot, E. A. Kazerooni, A. M. Kelly, L. E. Quint, B. Desjardins, and B. Nan, "Interobserver and intraobserver variability in the assessment of pulmonary nodule size on CT using film and computer display methods," *Academic Radiology* **12**, pp. 948–956, August 2005.
21. C. R. Meyer, T. D. Johnson, G. McLennan, D. R. Aberle, E. A. Kazerooni, H. MacMahon, B. F. Mullan, D. F. Yankelevitz, E. J. R. van Beek, S. G. Armato III, M. F. McNitt-Gray, A. P. Reeves, D. Gur, C. I. Henschke, E. A. Hoffman, P. H. Bland, G. Laderach, R. Pais, D. Qing, C. Piker, J. Guo, A. Starkey, D. Max, B. Y. Croft, and L. P. Clarke, "Evaluation of lung MDCT nodule annotation across radiologists and methods," *Academic Radiology* **13**, pp. 1254–1265, 2006.