An Automated Method for Locating Phantom Nodules in Anthropomorphic Thoracic Phantom CT Studies

Adele P. Peskin[1], **Alden A. Dima**[2], **and Ganesh Saiprasad**[23]
[1]NIST, Boulder, CO, USA
[2]NIST, Gaithersburg, MD, USA
[3]University of Maryland Medical Center, Baltimore, MD, USA

Abstract— *The Cancer Imaging Archive (TCIA) has a publicly available FDA database consisting of just over one thousand CT scans intended for facilitating the assessment of lung nodule size estimation methodologies, the development of image analysis software, as well as a wide range of different analyses. The use of these scans would be greatly facilitated by the availability of phantom nodule location data that could be input into methods that require them. This paper outlines a new image processing method to locate the phantom nodules in these CT scans in order to supplant manual location prior to their use. We present a method for extracting the phantom nodules, that involves phantom lung wall removal and separation of the phantom nodules from surrounding phantom blood vessels. Nodule locations are described by rectangular boxes bounding their positions in the scans and volume estimations.*

Keywords: Anthropomorphic thoracic phantom, image processing, phantom nodule

1. Introduction

Nodules in the lung are classified according to their measured growth in CT scans taken over a period of time. Many prognostic techniques for cancerous nodules depend on accurate volumetric measurement techniques. Fully automated volumetric techniques must be able to handle the complex task of separating nodules from the blood vessels and structures around them, and often attached to them. Many different techniques have been developed both fully and semi-automated, and comparative studies have been done and are still underway [1]-[5]. The Radiological Society of North America's QIBA (Quantitative Imaging Biomarkers Alliance) Volumetric CT Study 3A is in the process of testing the variation and performance of algorithms for computing nodule volumes in CT lung nodule data sets (http://qibawiki.rsna.org). The study is currently being performed on phantom lung nodule data. The phantom data, generated at the FDA, is available via the TCIA for lung nodule size estimation assessment and image analysis software development (https://wiki.cancerimagingarchive.net/display/Public/ Phantom+FDA). However, the locations of the phantom nodules in these sets are not known well enough to be used as inputs to methods that require them. Studies using phantom nodules are one way to make direct quantitative comparisons of volumetric techniques. However, manually collecting the location data for the phantom lung nodules is a time-consuming process. As a result, the FDA phantom CT database would be more easily utilized if the centroid locations of the nodules imaged in them are known well enough for methods requiring locations as input. In this paper we describe a technique to extract the location of phantom nodules from these phantom lung CT data sets. We developed the method by working with a set of 15 CT scans, and then tested the method by use of another 40 scans. The results of these tests are given.

2. Extraction Method

There are several main challenges to separating the phantom nodules from the rest of the lung data. Pixel intensities that represent the nodules in these scans are in the same range as those representing the anthropomorphic phantom's lung walls and the blood vessels. The first two major steps in this method are to define the boundaries of the lung walls in each CT slice, and then to separate the nodules from the blood vessels in the areas inside each lung. Histograms of pixel intensities for a set of CT slices containing a phantom nodule, and for the nodule alone are shown in Figure 1. The phantom nodules each have slightly different pixel distributions. Nodules with different densities are located in the anthropomorphic thoracic phantom, with mean values for the nodule pixels falling in two ranges, approximately -600 and 0 Hounsfield units. The pixel intensities that represent the air in the lungs are approximately -1000 Hounsfield units, and the lung walls approximately 0 HU.

2.1 Lung Wall Definition

It is important to define the boundaries of the lung wall in each CT slice, so that we limit our search to the regions where the nodules occur. Clusters of pixels that represent the lung walls themselves, therefore will not be part of our search. To do this we take advantage of the large difference in pixel intensity between pixels representing the lung wall and pixels representing the air inside of the lung. Figure 2 shows a section of a sample slice of both clinical and

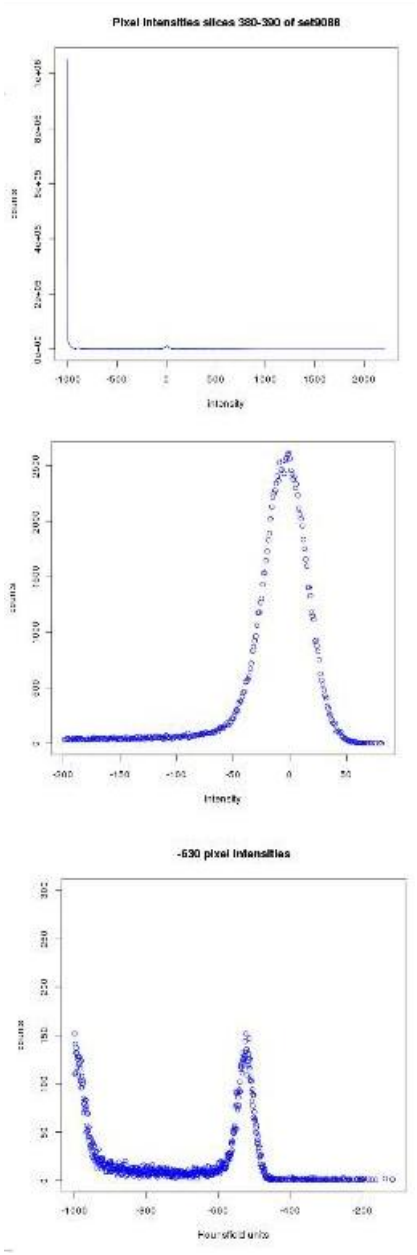

Fig. 1: Histograms of: 1.) pixels intensities in a set of slices containing a phantom nodule with a mean pixel intensity around 0; 2.) pixel intensities in the extracted tumor itself, illustrating the large difference in pixel intensities between the nodules and air inside the lung; 3.) pixel intensities from a nodule with lower mean pixel value.

phantom CT data that contains a nodule, color coded to show the large differences in pixels intensities between the wall and the lung region. We use these large pixel intensity differences to locate the lung walls in the data sets. In each slice of data, we find each segment of the outer and inner walls of the lung by searching for sets of pixels that maintain a high Hounsfield value and then shift to a lower Hounsfield value. Figure 3 shows the results for 2 data sets: the first, a 119 slice 2.5 mm slice thickness set, the second a 749 slice 0.4 mm slice thickness set. From each slice we extract values for both the inner and outer boundaries in that slice.

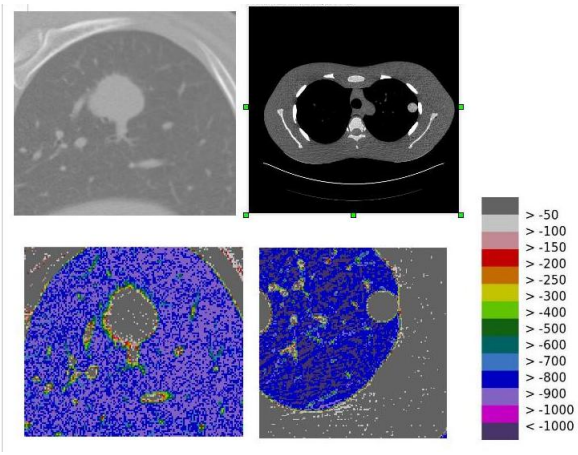

Fig. 2: Section of a sample slice of both clinical (left) and phantom (right) lung tumor data, and below them the color coded corresponding pixels.

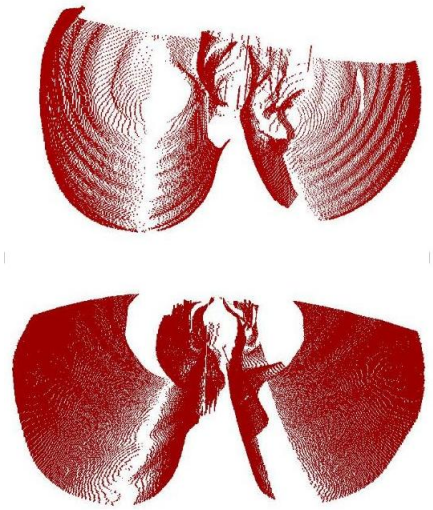

Fig. 3: Lung wall from a data set of 119 2.5 mm. and a set of 749 0.4 mm slices.

2.2 Lung Wall Removal

Once the inner and outer boundaries of the lung wall are determined within each slice, the space inside the lungs is scanned for nodules. In general, the nodules will be represented in the data by clusters of pixels that are inside the lung and have a relatively high pixel intensity. If clusters of pixels whose intensities are greater than -100 Hounsfield units are selected from these regions, there are 2 factors that complicate the process of selecting out those clusters that represent nodules. The first is the fact that blood vessels and nodules have overlapping pixel intensity ranges. The second is the fact that some of the nodules are attached to the lung wall, and therefore those nodules do not show up as isolated pixel clusters. An example of an attached nodule is shown in Figure 4.

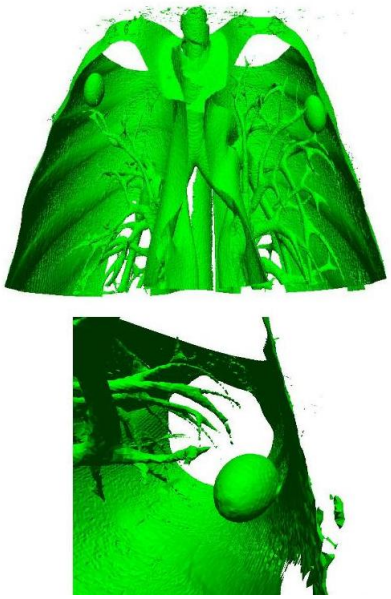

Fig. 4: Top: An isosurface from a phantom tumor data set at -100 Hounsfield units. Bottom: a close up of the attached tumor from this set.

Separation of nodules from blood vessels is described in the next section. Here we describe a technique to separate nodules attached to the lung wall. Because these sets of data are all images of the same anthropomorphic thoracic lung, the lung wall shows up consistently in every set. Looking in the slice direction at the 3-D set of pixel intensities, the lung wall consists of high intensity clusters of pixels that span most of the slice direction. Specifically, if we look at sets of pixels that are 10 pixels wide in the x direction (10 $\times$ y dimension $\times$ z dimension), there exist clusters of high intensity lung wall pixels that span the entire slice direction (see Figure 5). Our method scans these sets of high intensity pixel clusters, and removes those that span the set in the slice direction. In this manner, 2 outcomes are helpful for our nodule search: Both outer and inner walls are removed, and

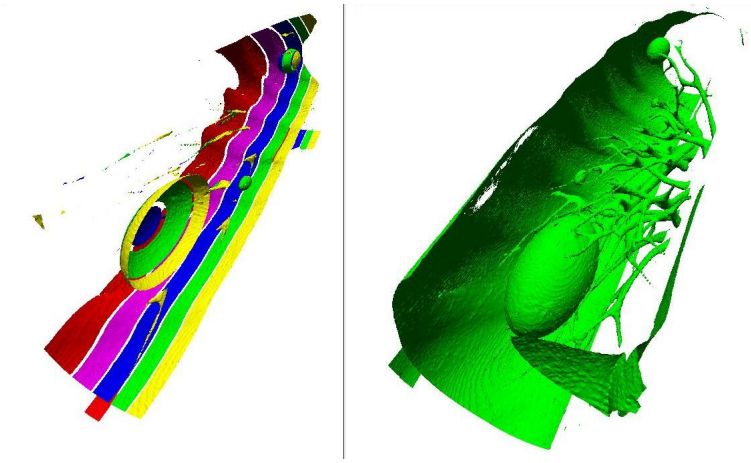

Fig. 5: Clusters of high intensity pixels are seen in isosurface at -100 HU. On the right is a section of the lung, and on the left some 10 pixel sections of that isosurface in the slice direction, many of which span the slice direction and will be removed with the lung wall removal algorithm.

attached nodules are separated from the wall. That is, where nodules are attached, the 10-pixel columns of high intensity pixels will leave clusters that represent the nodules, since these do not span the whole slice direction, while removing clusters that make up the wall. Figure 6 shows the same attached nodule as in Figure 4, and then sets of high intensity pixels that remain after elimination of the lung wall. In some sets of data the entire wall is removed by this process (see Figure 7); in others, parts of the wall remain and are later removed by another processing step. In every case we have run so far, a good part of the lung wall is removed and the attached nodules are separated.

2.3 Separation of Tumors from Blood Vessels

At this point we distinguish between the nodules with mean intensities around 0 and around -600 Hounsfield units. We search for clusters of pixels inside the lung in one of our two pixel intensity ranges. These clusters of pixels inside the lung often contain both nodule and blood vessel pixels. We separate out all of the smaller (less than 300 pixel) clusters, and perform an erosion on the rest of the sets of pixel clusters to separate pieces of a single cluster, i.e., for each pixel cluster, a pixel that does not have a neighbor immediately to the right, to the left, up, down, and in the previous and next slice direction is eliminated. The smaller clusters are not put through this process, so that the small nodules are not lost. Pixel clusters representing the blood vessels appear not as whole blood vessels, because the pixel distribution of blood vessel pixels as a whole is different from that of the nodules and the distribution is more spread. We see incomplete vessels in these pixel clusters, which are easily separated into pieces, removing them from the nodules (see Figure 8). Because we are using data sets from the phantom

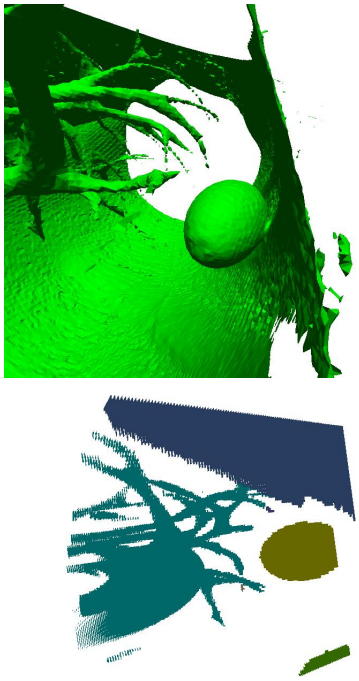

Fig. 6: Clusters of high intensity pixels after wall removal in the region of the data set of Figure 3 where the tumor is attached to the lung wall.

lung, the pixel distribution of the blood vessels will differ from those of clinical data. However, from our past work, we have seen that the blood vessel data have a fairly wide distribution in clinical data, and one that is centered around mean values lower than those of clinical nodules [6]-[7].

2.4 Final Tumor Selection

At this point, we are left with sets of pixel clusters (either in the range of greater than -100 or -630 to -450), and must make decisions about whether or not they represent the phantom nodules. For each cluster we at look a variety of factors: 1. its position in the data set (is it within the boundaries set by our lung wall calculations?); 2. aspect ratio; 3. ellipsoidal shape: the ratio of the number of pixels to the corresponding volume of an ellipsoid that would fit in the bounding box of the nodule; 4. percent of the bounding box of the nodule that is filled by the pixel cluster. Nodule clusters are often ellipsoidal, have aspect ratios that are not exaggerated in a particular direction, and fill their bounding boxes more than the spindly clusters resulting from blood vessel do. Clusters representing the lung wall that were not removed completely by the process above are eliminated in the step by their aspect ratio. We evaluated the nodules in the test set of 15 data sets and set limits on the geometric factors based on the results of the 15 sets. We then tested

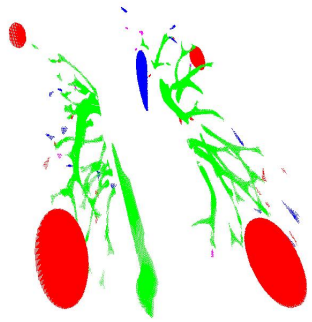

Fig. 7: Large pixel clusters remaining after wall removal on another set of 749 slices.

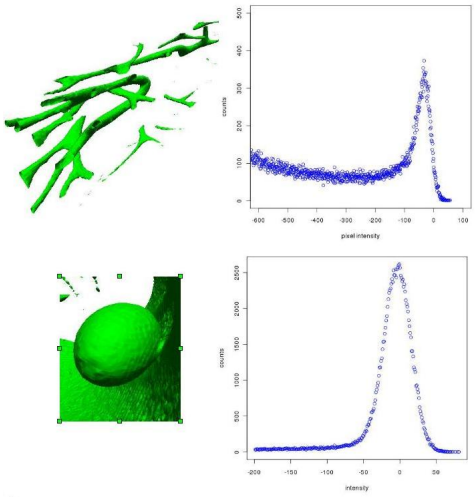

Fig. 8: Comparative histograms of pixels in the region of blood vessels and the pixels representing the phantom nodule from Figure 4 in an isosurface at -100 Hounsfield units.

those limits on an additional 40 sets. Some of these limits were size dependent; i.e. Smaller nodules are expected to be rounder than larger nodules. Table 1 has the limits for these factors over different size ranges of the high intensity pixel clusters. Nodule shapes in these sets include spherical, irregular, elliptical, lobulated, and spiculated. Results are output as both the bounding boxes for each nodule and their volume estimated by the number of cluster pixels.

3. Results

For the initial 15 sets we worked with to develop the software, and for the 40 test data sets, manually selected ground truth estimates were generated. Groups of these sets had the same phantom lung nodule layouts, and results for these sets should look the same. Some of the sets consisted of 119 2.5 mm slices and the rest of 749 0.4 mm slices. Pixel-pixel distances in plane were 0.78125 mm. Table 2

Table 1: Selection factors

Screening order	Pixel clump size	Aspect ratio	Ellipsoidal ratio	Fraction of box filled
1	all			≥ 0.1
1	all	≤ 10.0		
1	all	≤ 5.0 and	≤ 3.0	
1	all	≤ 4.0 and	≤ 2.0 and	≥ 0.2
2	≤ 500	≤ 2.5 and	≤ 2.0	
3	≤ 500		≤ 1.5 or	≥ 0.5
2	500-1000	≤ 5.5 and	≤ 1.7 and	≥ 0.2
2	≥ 1000	≤ 5.0 and	≤ 5.0 and	≥ 0.1

outlines the different overlays of phantom nodules in the 40 sets. Table 3 gives the results generated for the 40 sets. For each set, we report the number of nodules found, comparing the computed boundary box and estimated volume with the manually generated bounding box and the known volume of the phantom. A false positive is a cluster of grid points not associated with a reference nodule. In general, we were able to locate medium and large nodules that were located in the data sets. In the sets with thicker slice distances, we are not able to locate the smaller nodules that are represented in only a very small number of slices. In the thin sliced sets, we are able to locate 96% of all phantoms. Figure 9 visually shows some the results from both a 119 slice set and the corresponding 749 slice set from identical phantom nodule layouts.

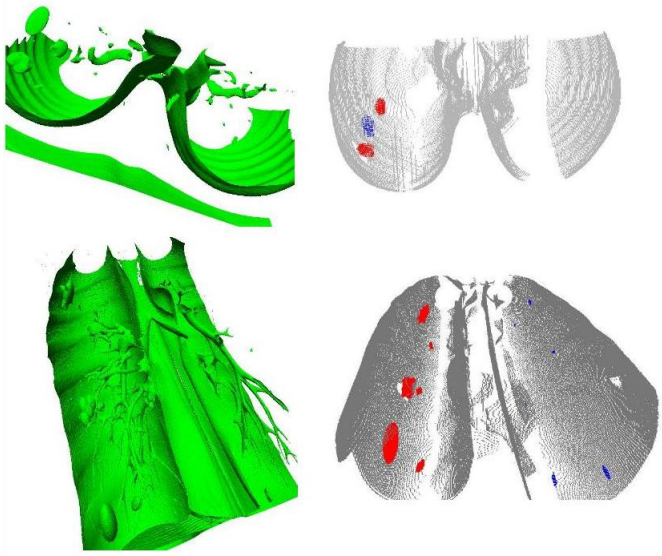

Fig. 9: Isosurfaces at -100 Hounsfield units of a 119 slice set and a corresponding 749 slice set, and the results of our software for each. The results are shown with the lung wall, so that the position of nodules found is clear. More accurate nodule detection is associated with thinner slice distances.

Table 2: Phantom lung nodules used in the 40 test sets

Nodule	size mm^3	shape	Density (HU)	Layout
1	511.13	spherical	100	2
17	676.	irregular	100	2
18	263.	iregular	20	2
30	63.72	spherical	100	2
31	254.71	spherical	100	2
19	253.	irregular	-300	2
3	526.64	lobulated	100	4
5	528.67	spiculated	100	4
7	4207.83	elliptical	100	4
20	4920.35	elliptical	-630	4
21	471.3	elliptical	100	4
22	679.31	elliptical	-630	4
23	4350.42	lobulated	100	4
24	5062.15	lobulated	-630	4
25	569.17	lobulated	-630	4
26	5279.07	spiculated	-630	4
27	706.75	spiculated	-630	4
8	4232.05	spherical	100	6
9	4286.76	spherical	-10	6
6	533.69	spherical	-10	6
29	33781.46	spherical	100	6
33	4215.1	spherical	-630	6
36	34389.21	spherical	-10	6
61	69.24	spherical	-10	6
62	282.35	spherical	-10	6
4	527.42	lobulated	-10	7
2	524.67	spiculated	-10	7
10	4315.84	elliptical	-10	7
39	280.45	elliptical	-10	7
40	283.36	lobulated	-10	7
41	524.63	elliptical	-10	7
42	67.65	lobulated	-10	7
43	4234.41	lobulated	-10	7
44	69.8	elliptical	-10	7
46	68.71	spiculated	-10	7
47	283.91	spiculated	-10	7
63	4398.	spiculated	-10	7

Table 3: Results for 40 data sets

Layout	Data set	Thick or thin slice	0 HU found	-600 HU found	False detections
2	7749,7755	thick	2/5,2/5	0/1	0,0
4	7495,7501,	thick	3/5,4/5,	5/6,5/6,	0,0,
	7507,7513,		3/5,3/5,	6/6,5/6,	0,0,
	7519,7525		4/5,4/5	6/6,6/6	0,0
6	9511,9517,	thick	4/7,4/7,	1/1,1/1,	0,1
	9523,9529,		5/7,5/7,	1/1,1/1	1,0,
	9535,9541		5/7,5/7	1/1,1/1	0,1
7	490,496,	thick	2/12,2/12,	0/0,0/0,	0,0,
	502,508,		2/12,3/12,	0/0,0/0,	0,0,
	514,520		2/12,3/12	0/0,0/0	1,0
2	7001,7004	thin	5/5,5/5	1/1	1,1
4	1071,1077,	thin	5/5,5/5,	6/6,6/6,	0,0,
	1083,1089,		5/5.5/5,	6/6,6/6,	0,0,
	1095,1101		5/5,5/5	6/6,6/6	0,0
6	9040,9046,	thin	7/7,7/7,	1/1,1/1,	0,1
	9052,9058,		7/7,7/7,	1/1,1/1,	1,0,
	9064,9070		7/7,7/7	1/1,1/1	0,1
7	3,9,	thin	11/12,11/12,	0/0,0/0,	0,0,
	15,21,		11/12,10/12,	0/0,0/0,	0,0,
	27,33		11/12,11/12	0/0,0/0	0,0

4. Conclusions and future work

We have developed a method to capture the location of individual phantom nodules in lung phantom data sets. The method includes an algorithm to locate the lung wall in the data, and an algorithm to remove the walls from the search for the nodules within them. This method has been successfully tested with 40 sets of data for which we have manually located the phantom nodules. Future work includes applying our method to the full set of publicly available FDA phantom CT scans available via The Cancer Imaging Archive.

5. Acknowledgements

We acknowledge Nicholas Petrick, Deputy Director, Division of Imaging and Applied Math, CDRH, at the Food and Drug Administration, for his helpful comments with the manuscript.

References

[1] Jirapatnakul,A.C., Reeves,A.P., Apanasovich,T.V., Biancardi,A.M.,Yankelevitz,D.F.,Henschke,C.I.: Pulmonary Nodule Classification: Size Distribution Issues. ISBI, 1248-1251 (2007).

[2] Kostis, W.J., Reeves, A.P., Yankelevitz, D.F., Henschke, C.I.: Three-Dimensional Segmentation and Growth-Rate Estimation of Small Pulmonary Nodules in Helical CT Images. IEEE Trans. on Medical Imaging, 22(10), (October 2003).

[3] Bonnet,N., Cutrona,J., Herbin,M., A no-threshold histogram-based image segmentation method. Pattern Recognition 35 : 2319-2322 (2002).

[4] 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. Med. Imaging, 25(4), 435-450 (2006).

[5] Gavrielides, M.A., Kinnard, L.M., Myers, K.J., Peregoy, J., Pritchard, W.F., Zeng, R., Esparza, J., Karanian, J., Petrick, N.: A resource for the assessment of lung nodule size estimation methods: database of thoracic CT scans of an anthropomorphic phantom, Optics Express, 18(14), 15244-15255 (2010).

[6] Peskin, A.P., Kafadar, K., Santos,A.M., Haemer,G.G.: Robust Volume Calculations of Tumors of Various Sizes. The 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition, 43-47 (July 2009).

[7] Peskin,A., Dima,A.: Modeling Clinical Tumors to Create Reference Data for Tumor Volume Measurement, 6th International Symposium on Visual Computing 2010. (G. Bebis et al. (Eds.): ISVC, Part II, LNCS 6454, 736-746. Springer, Heidelberg (2010).