PET TUMOR SEGMENTATION: MULTI-OBSERVER VALIDATION OF A GRADIENT-BASED METHOD USING A NSCLC PET PHANTOM

A. D. Nelson[1], M. Werner-Wasik[2], W. Choi[3], Y. Arai[4], P. F. Faulhaber[5,6], N. Ohri[2], J. W. Piper[1,7], K. D. Brockway[1], A. S. Nelson[1]

[1]MIMvista Corp, Cleveland, OH, [2]Thomas Jefferson University Hospital, Philadelphia, PA, [3]Beth Israel Medical Center, New York, NY, [4]University of Pittsburgh Medical Center Health Systems, Pittsburgh, PA, [5]University Hospitals Case Medical Center, Cleveland, OH, [6]Case Western Reserve University, Cleveland, OH, [7]Dept of Computer Science, Wake Forest University, Winston-Salem, NC

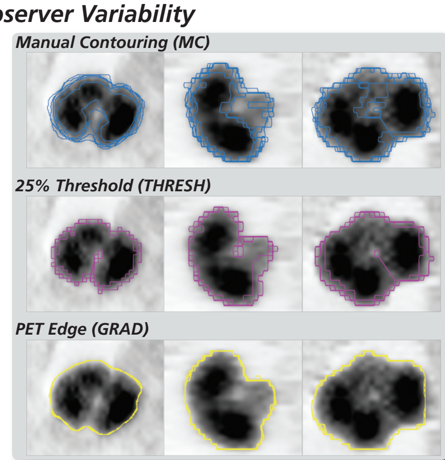
Purpose

Consistent and accurate methods for PET tumor segmentation are needed in Radiation Therapy with the growing role of PET for target definition, prognosis, and therapy response assessment. Previously we demonstrated the superior accuracy of a gradient-based PET segmentation method, PET Edge (GRAD), to constant threshold (THRESH) methods using constant activity spherical phantoms[1] which were obtained to emulate clinical activity, acquisition, and reconstruction on five different PET cameras. In a follow-up single observer study we demonstrated the superior accuracy of GRAD compared to constant threshold methods[2] using published Monte Carlo simulated PET scans of the thorax with "tumors" in commonly encountered locations for lung cancer patients treated in Radiation Therapy[3]. Our goal in the present study is to determine for multiple observers the accuracy and consistency of GRAD, THRESH and manual contouring (MC) using the same Monte Carlo simulated PET scans of the thorax.

Materials/Methods

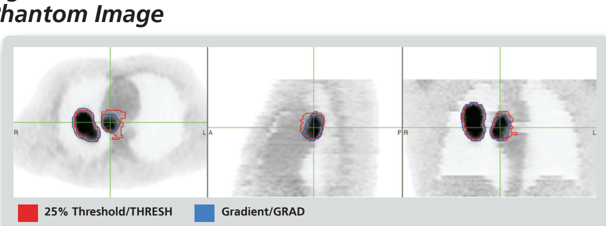
Twenty-five realistic digital PET phantoms of the thorax were obtained with 31 simulated tumors of varying size, shape and location. Tumors ranged in size from 7 ml to 264 ml with some tumors surrounded by lung while other tumors were adjacent to the mediastinum or chest wall. Five observers (three Radiation Oncologists, one Radiation Oncologist with Radiology training, and one PET trained Radiologist) segmented each tumor with GRAD, THRESH, and MC. THRESH was performed using thresholds of 25-50% of maximum counts at 5% increments. Tumor volumes for each method were compared to known volumes prior to simulation. Accuracy was measured by calculating the mean absolute % difference for the volume (absVdiff%) and mean % difference for the volume (Vdiff%) for each group using all methods.

Figure 1
Inter-observer Variability



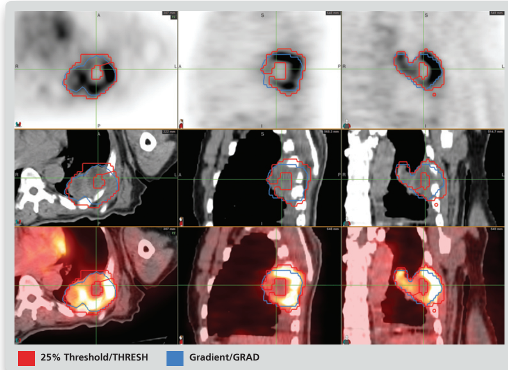
Surrounding normal tissues and heterogeneous tumor activity can result in substantial variability in tumor segmentation. This example demonstrates the robust, consistent results generated using PET Edge compared with threshold and manual contouring.

Figure 2
Phantom Image



For cases where the tumor is in the center of lung (i.e. high source-to-background) and fairly homogeneous, 25% THRESH performs fairly well, however, in the mediastinum with lower source-to-background 25% THRESH performs poorly. GRAD produces more accurate segmentations in both of these scenarios.

Figure 3
Patient Image



Case illustrates limitations of constant threshold in cases of heterogeneous tumor metabolism and areas of decreased source to background ratio (adjacent to chest wall) while GRAD produces more accurate segmentations in both of these scenarios.

Results

Across all observers, GRAD was the most accurate segmentation technique (see table 1) with absVdiff% of 11.3% (12.2% SD). The next most accurate, 25% THRESH, with absVdiff% of 15.1% (16.7%), and MC, with absVdiff% 18.9% (17.3%), were statistically significantly less accurate ($p < 0.01$). GRAD also had the smallest amount of systematic bias (see table 2) with volume differences Vdiff% of -0.45% (16.7%) compared with -7.0% (21.4%) and -15.7% (20.3%) for 25% THRESH and MC, respectively ($p < 0.01$). Finally, GRAD was more consistent between observers than either MC or 25% THRESH in 25/31 and 21/31 lesions, respectively (See Figure 1). Using Levene's test for equality of variances, inter-observer variability was statistically significantly reduced using GRAD compared to MC ($p < 0.01$) but GRAD vs 25% THRESH was non-significant.

Table 1
Volumetric Accuracy

Technique	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5	Overall (Average)
PET Edge	10.979	11.017	13.106	10.783	10.799	**11.337**
Manual	13.603	25.175	36.541	8.241	10.993	**18.911**
25%	13.433	11.242	12.776	21.585	16.368	**15.081**
30%	19.007	17.091	17.198	26.390	21.399	**20.217**
35%	25.392	23.274	24.389	31.939	25.739	**26.147**
40%	32.538	31.987	31.849	38.370	34.711	**33.891**
50%	50.405	49.622	50.332	47.863	51.644	**49.973**

Mean Absolute Percent Differences in Volume (Vdiff%)

Table 2
Systematic Bias

Technique	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5	Overall (Average)
PET Edge	-0.354	-0.508	-3.870	1.547	0.934	**-0.450**
Manual	-7.280	-23.373	-35.279	-4.581	-7.964	**-15.695**
25%	-5.536	-6.389	-4.210	-19.281	0.176	**-7.048**
30%	-17.156	-12.194	-13.880	-26.248	-12.910	**-16.478**
35%	-23.680	-21.792	-22.780	-31.939	-24.304	**-24.899**
40%	-32.538	-29.764	-31.147	-38.370	-34.711	**-33.306**
50%	-50.405	-49.622	-50.332	-47.862	-51.644	**-49.973**

Mean Percent Differences in Volume

Conclusions

The gradient-based PET segmentation technique, PET Edge (GRAD) was the most accurate and consistent PET tumor segmentation technique across observers. GRAD has the potential to play an important role in Radiation Therapy through more accurate and consistent PET Tumor segmentation.

References

[1] Shen G, Nelson D, Adler L. PET Tumor Segmentation: Comparison of Gradient-Based Algorithm to Constant Threshold Algorithm. Med Phys. 2007;34:2395.

[2] Nelson D, et al. PET Tumor Segmentation: validation of a gradient-based method using a NSCLC PET Phantom. J Nucl Med. 2009; 50 (Supplement 2):1659.

[3] Aristophanous M, Penney BC, Pelizzari CA. The development and testing of a digital PET phantom for the evaluation of tumor volume segmentation techniques. Med Phys. 2008;35:3331-3342.