Radiotherapy and Oncology 123 (2017) 257–262

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Quantitative imaging in lung

Quantitative imaging: Correlating image features with the segmentation accuracy of PET based tumor contours in the lung

Perry B. Johnson [a], Lori A. Young [b], Narottam Lamichhane [c], Vivek Patel [d], Felix M. Chinea [d], Fei Yang [d,*]

[a] Radiation Oncology/Biomedical Engineering, University of Miami, Miami, FL; [b] Radiation Oncology, University of Washington, Seattle, WA; [c] Radiation Oncology, University of Maryland, Baltimore, MD; and [d] Radiation Oncology, University of Miami, Miami, FL, USA

ARTICLE INFO

Article history:
Received 3 October 2016
Received in revised form 2 March 2017
Accepted 12 March 2017
Available online 20 April 2017

Keywords:
Quantitative imaging
Positron emission tomography
Target delineation
Image texture

ABSTRACT

The purpose of this study was to investigate the correlation between image features extracted from PET images and the accuracy of manually drawn lesion contours in the lung. Such correlations are interesting in that they could potentially be used in predictive models to help guide physician contouring. In this work, 26 synthetic PET datasets were created using an anthropomorphic phantom and Monte Carlo simulation. Manual contours of simulated lesions were provided by 10 physicians. Contour accuracy was quantified using five commonly used similarity metrics which were then correlated with several features extracted from the images. Features were sub-divided into three groups using intensity, geometry, and texture as categorical descriptors. When averaged among the participants, the results showed relatively strong correlations with complexity and contrast$_I$ ($r \geq 0.65$, $p < 0.001$), and moderate correlations with several other image features ($r \geq 0.5$, $p < 0.01$). The predictive nature of these correlations was improved through stepwise regression and the creation of multi-feature models. Imaging features were also correlated with the standard deviation of contouring error in order to investigate inter-observer variability. Several features were consistently identified as influential including integral of mean curvature and complexity. These relationships further the understanding as to what causes variation in the contouring of PET positive lesions.

 Radiotherapy and Oncology 123 (2017) 257–262

Defined as the extraction of quantifiable features from medical images for the purpose of medical assessment [1], quantitative imaging is an emerging domain within radiation oncology. This is primarily due to the mainstream integration of multimodality imaging within the clinical workflow, and the ease in which advance metrics can be calculated from digital images [2–4]. One of the largest challenges for quantitative imaging is the translation of complex variables into tangible data that connects with how clinicians assess, plan, and ultimately treat patients. A strong candidate for this type of effort is the contouring of positron emission tomography (PET) based lesions in the lung. In both the conventional and hypo-fractionated settings, the treatment of lung cancer often begins with PET imaging which is utilized during the assessment and planning stages of the radiotherapy treatment process [5]. While the accuracy of the delineated tumor volume is integral to the quality of the therapy, contouring can be highly variable due to a number of factors. These include the low spatial resolution and high noise characteristics of PET imaging, the uneven uptake of ^{18}F-FDG within lung tumors, and the sensitivity in which the brightness and contrast affect the apparent tumor volume on a digital display [6]. Reducing this variability is an open topic in medical physics and a variety of automated methods have been proposed often involving thresholding, variational approaches, learning methods, and stochastic models [7–9]. Each approach brings its own set of advantages and limitations where the tradeoffs typically represent a balance between simplicity, efficiency, sensitivity, specificity, and accuracy. While significant progress has been made in this regard, widespread adoption of these tools has yet to occur and remains a promise for the near future.

Barriers to reaching such a future include both technical and cultural factors. As noted above, noise, low spatial resolution, and tumor heterogeneity all present challenges for computational algorithms, which are often designed to detect edges, gradients, or cluster voxels with similar gray scale values. Additionally, these techniques do not necessarily mimic the way clinicians perceive and analyze medical images. Beyond a robust ability to distinguish edge detail and discontinuities via unique phenomena such as edge enhancement, see Mach band effect [10], pattern recognition plays a dominant role in allowing humans to connect textural features with meaning. In this way an expert eye not only views but also interprets an image. Culturally, physician drawn contours are widely considered the gold standard, and there exists a strong

* Corresponding author at: University of Miami, 1475 NW 12th Ave, Miami, FL 33136, USA.
E-mail address: fei@miami.edu (F. Yang).

http://dx.doi.org/10.1016/j.radonc.2017.03.008
0167-8140/

historical inertia for physician-based segmentation built around training, experience, and a desire for autonomy.

As a result of these limitations, a number of authors have suggested semi-automated, rather than fully automated techniques as a way to bridge the gap between computer and human delineation [11,12]. One example of semi-automation is the Contouring Co-pilot™ available in the MIM Maestro platform (MIM Software, Cleveland, OH), which suggests paths to the user based on deformable image registration. This technique can be helpful but requires practice in order to familiarize the user with the predictive features of the tool. Another interesting form of semi-automation yet to be fully explored is the use of quantitative imaging to provide prospective information to the user. Such information could be actionable if framed appropriately, such as providing guidance as to when a user is likely to over- or under-contour a PET-positive lesion. For the success of such guidance, correlations must exist between quantifiable features of the image and tangible metrics that relate to contour accuracy. It is reasonable to hypothesize that such correlations exist considering that the human perception mechanism relies in part upon pattern recognition based on image features. These features, including intensity, textural, and morphological descriptors, can be quantified in medical images.

In the current study, this hypothesis was tested in the lung by collecting 260 manual drawn target contours (10 physicians × 26 datasets) and correlating the accuracy of those contours with several different image features. Segmentation accuracy was assessed using a Monte Carlo-derived ground truth model, and the image features were broadly categorized as related to intensity, geometry, or texture. The correlations were analyzed individually and collectively and the results are discussed within the context of an enhanced contouring workflow.

Methods and materials

Ground truth model

In order to test the hypothesis, the study required multiple PET images with ground truth knowledge of lesion geometry and location. This was achieved by creating 26 synthetic datasets using the anthropomorphic Zubal phantom [13] and the Monte Carlo based SimSET (a Simulation System for Emission Tomography) computational package [14]. This method was previously validated in comparison with clinical data and has been utilized in a variety of segmentation-oriented studies [15–17]. Briefly, the simulation allows the user to place concentrations of ^{18}F at different locations within the digital phantom to approximate ^{18}F-FDG uptake within metabolically active tumors and normal tissues. The activity indices can take values from 0 to 255 where each index is proportional to an activity concentration in $\mu Ci/cm^3$. The indices ranged from 30 to 195 for lesions and from 0 to 38 for normal tissues. A Monte Carlo technique is used to simulate β^+ decay, and annihilation photons are subsequently detected by a continuous cylindrical BGO detector (40 cm radius) also modeled by the software. Sinograms (128 × 128) are collected using single slice re-binning (SSRB), and images are reconstructed using the ordered subset expectation maximization (OSEM) algorithm (8 iterations/4 subsets). A post-reconstruction 3D Gaussian filter (FWHM = 5 mm) is applied to smooth the images. Attenuation correction is done with each organ using a tissue index that maps to an attenuation coefficient used in the SimSet package. The various aspects of simulation specifics related to detector property, sonogram acquisition, and reconstruction algorithm were chosen to match those of the PET scanner at the local facility. In so doing, quality of the simulated image data can be assessed with the clinical data. Furthermore, the inter-rater adaptation variability in visual perception and recognition of PET imaged tumors with respect to the variation of various imaging parameters would be minimized.

Target contouring and evaluation

Target contours were provided by 10 physicians using the software MIM Maestro v6.5.5 (MIM Software, Cleveland, OH). Each physician possessed clinical experience in contouring PET based lesions as part of the radiation therapy treatment planning process. No specific instructions were given regarding window/level or threshold settings as neither is standardized at the local institution. Also, only simple contouring tools were used within the MIM Maestro program, *i.e.* no atlas-based or semi-automated contouring techniques. The study was conducted in a double-blind fashion to guard against bias. Segmentation accuracy was evaluated using six commonly used volumetric and distance-based similarity metrics. These included:

Dice index: Measures extent of spatial overlap between two contours.

$$Dice = 2 \times \frac{|A \cap B|}{|A| + |B|} \tag{1}$$

Jaccard's index: Differs from Dice in that equal weight is given to matches and non-matches.

$$Jaccard = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

False positive and false negative Dice indices (FPD and FND): A bounding box is used to provide complements for both A and B. False positive dice provides a measure for over-segmentation (Type I error) and false negative dice provides a measure for under-segmentation (Type II error).

$$Dice_{false+} = 2 \times \frac{|A \cap \bar{B}|}{|A| + |B|} \tag{3}$$

$$Dice_{false-} = 2 \times \frac{|\bar{A} \cap B|}{|A| + |B|} \tag{4}$$

Symmetric mean absolute surface distance (SMASD): Estimates the average extent to which the surfaces of A and B differ. Variables n_A and n_B are the number of surface voxels on A and B; d^{AB} is the distance to the closest voxel of B for the ith surface voxel of A, and d^{BA} is the distance to the closest voxel of A for the jth surface voxel of B.

$$SMASD = \frac{1}{n_A + n_B}\left(\sum_{i=1}^{n_A}\left|d_i^{AB}\right| + \sum_{j=1}^{n_B}\left|d_j^{BA}\right|\right) \tag{5}$$

Absolute volumetric difference (AVD): Percent difference between the volumes of A and B where B is designated as the gold standard.

$$AVD = \frac{|V_A - V_B|}{V_B} \times 100 \tag{6}$$

Image feature extraction

For each dataset, image features were extracted using voxels located within two distinct volumes of interest (VOIs). The simulated lesion was used as the ground truth VOI. The outline of this VOI was used as the ground truth contour. A 2 cm isotropic margin around the ground truth VOI was used as the local background VOI. The purpose of this structure was to provide contextual image information for a given lesion.

Features were sub-divided into three groups using intensity, geometry, and texture as categorical descriptors (Table 1). Starting with intensity, only one feature was included in this category,