Disease Quantification on PET/CT Images without Object Delineation

Yubing Tong[1], Jayaram K. Udupa[1], Dewey Odhner[1], Caiyun Wu[1],
Danielle Fitzpatrick[2], Nicole Winchell[2], Stephen J. Schuster[2], Drew A. Torigian[1]
[1]Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, 19104, United States. [2]Abramson Cancer Center, Perelman Center for Advanced Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States.

ABSTRACT

The derivation of quantitative information from images to make quantitative radiology (QR) clinically practical continues to face a major image analysis hurdle because of image segmentation challenges. This paper presents a novel approach to disease quantification (DQ) via positron emission tomography/computed tomography (PET/CT) images that explores how to decouple DQ methods from explicit dependence on object segmentation through the use of only object recognition results to quantify disease burden. The concept of an object-dependent *disease map* is introduced to express disease severity without performing explicit delineation and partial volume correction of either objects or lesions. The parameters of the disease map are estimated from a set of training image data sets. The idea is illustrated on 20 lung lesions and 20 liver lesions derived from ^{18}F-2-fluoro-2-deoxy-D-glucose (FDG)-PET/CT scans of patients with various types of cancers and also on 20 NEMA PET/CT phantom data sets. Our preliminary results show that, on phantom data sets, "disease burden" can be estimated to within 2% of known absolute true activity. Notwithstanding the difficulty in establishing true quantification on patient PET images, our results achieve 8% deviation from "true" estimates, with slightly larger deviations for small and diffuse lesions where establishing ground truth becomes really questionable, and smaller deviations for larger lesions where ground truth set up becomes more reliable. We are currently exploring extensions of the approach to include fully automated body-wide DQ, extensions to just CT or magnetic resonance imaging (MRI) alone, to PET/CT performed with radiotracers other than FDG, and other functional forms of disease maps.

Keywords: Disease quantification, cancer, object recognition, PET/CT

1. INTRODUCTION

It is now generally believed that quantitative radiology (QR), when brought to routine clinical practice, will bring about significant enhancement of the role of radiology in the medical milieu. The derivation of quantitative information from images, however, continues to face a major image analysis hurdle, namely the identification and delineation of "objects" of interest in the image. The "object" may be an anatomic organ, a sub-organ, a tissue region, a pathological region, or an anatomic zone such as a well-defined lymph node station. Called image segmentation, this process has a rich and long history spanning nearly 5 decades [1, 2] in the general area of image processing. Image segmentation has, however, remained the toughest challenge in image analysis and analytics and an essential roadblock to the practical implementation of QR. The purpose of this paper is to present a novel approach that explores how to decouple disease quantification (DQ) methods from explicit dependence on image segmentation. The idea is to express disease severity by designing disease mapping functions without performing partial volume correction and delineation of either objects or lesions. The parameters of the disease map are estimated from a set of training image data sets.

The main contribution of this paper is to propose a novel disease quantification approach for PET/CT images without delineation of objects. In the demonstrated manual mode of recognition, whether at the object level or even at the lesion level, the method is least intrusive on the lesion itself in terms of human interaction needed and the actual quantification strategy.

Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging
edited by Andrzej Krol, Barjor Gimi, Proc. of SPIE Vol. 10137, 101370V · © 2017 SPIE
CCC code: 1605-7422/17/$18 · doi: 10.1117/12.2254907

2. MATERIALS & METHODS

2.1 Image data

Twenty phantom PET/CT scans were obtained based on a NEMA [7, 8] NU-2 IQ phantom (manufactured by Data Spectrum, Durham NC), but with the central 5 cm diameter "lung" cylinder of the IQ phantom removed and the initial background activity level set to be equivalent to 15 mCi in a 70 kg patient, and the 271-day half-life of ^{68}Ge after 6 months the activity of about 9.5 mCi. Body-wide ^{18}F-2-fluoro-2-deoxy-D-glucose (FDG)-PET/CT scans are utilized in this study, as PET/CT is the most commonly used modality for molecular imaging of patients with cancer and provides image data that are quantifiable prior to and following treatment, allowing for individualized regional and global disease assessment of patients with cancer [9]. All patient PET/CT images were selected from our health system patient image database by a board certified radiologist (Torigian) following approval from the Institutional Review Board at the Hospital of the University of Pennsylvania along with a Health Insurance Portability and Accountability Act waiver. PET/CT scans with 20 lung lesions and 20 liver lesions were used to illustrate the disease quantification approach. The leave-one-out strategy is used for evaluation.

2.2 Methods

Recognition and delineation

As formulated in all of our previous segmentation work, we think of image segmentation as consisting of two related processes – object *recognition* and object *delineation*. *Recognition* is the high-level process of determining the whereabouts of the object or locating the object in an image. *Delineation* is the low-level process of precisely demarcating the boundary of or the region occupied by the object in the image. In a segmentation method, each process can be implemented to perform manually or automatically [3] or at different levels of automation and by using different strategies, and the degree of coupling and integration between the processes can also vary. Generally, when the processes are properly decoupled, the recognition process can be automated to perform much more robustly than delineation. On many occasions, delineation becomes ill-defined due to artifacts such as noise, blur, image non-uniformity, intensity non-standardness, the presence of pathology and its variations, and simply the lack of adequate intensity information in the image, even though recognition itself can be performed quite effectively. The aim of this study is to explore the possibility of using only recognition results to perform disease quantification via whole-body PET/CT acquisitions.

We will denote the low-dose unenhanced CT image by I_C and the matching PET image by I_P for all PET/CT images we employ. We will also assume that they are in alignment. Since our focus here is not automatic object recognition, we will assume that the object of interest O (which may be a solid organ, a tissue region, or a lymph node zone) has already been recognized or localized by some method, and that a rough (or fuzzy) mask (or model) FM(O) is available for O. For the purpose of illustration of the DQ ideas, our method of recognition here can be manual or automatic; that is, we will indicate FM(O) by a rectangular box on (I_C, I_P) just enclosing the object or FM(O) as the recognition results. Note that manual "recognition" can happen at the object level or lesion level; disease quantification in both cases is fully automatic with no interaction of any kind such as adjusting parameters, etc. needed.

Disease map and quantification

For quantifying disease burden in a body region b, we perform disease quantification for each object O of b by making use of FM(O), the standardized uptake value (SUV) image I_S derived from the PET image I_P, and an object-specific *disease map*, denoted by $d_O(x)$, which indicates disease severity within O at every voxel as a function of the voxel's SUV value x. $d_O(x)$ is modeled as $d_O(x) = g_d(x) - g_n(x)$, where $g_d(x)$ and $g_n(x)$ are half and full Gaussians with parameters (μ_d, σ_d) and (μ_n, σ_n), respectively. Our intent is that $g_n(x)$ describes the SUV distribution within the *normal* tissues of object O, and $g_d(x)$ expresses SUV-to-degree-of-disease relationship for O.

$$g_d(x) = \begin{cases} \exp[-(x-\mu_d)^2 / 2\sigma_d^2], & \text{if } x < \mu_d \\ 1, & \text{if } x \geq \mu_d \end{cases}, \qquad g_n(x) = \exp[-(x-\mu_n)^2 / 2\sigma_n^2]. \tag{1}$$

The disease map $d_O(x)$ removes any contributions from normal tissue to the "degree of disease". Assume for now that parameters (μ_d, σ_d) and (μ_n, σ_n) have been determined (see below for the estimation method). We describe disease burden within O by: SUVmean(O), SUVmax(O), and total lesion glycolysis TLG(O). SUVmean(O) and SUVmax(O) can be determined in a straightforward manner from FM(O) and I_S. Denoting the membership within FM(O) at voxel v by $fm_O(v)$ and volume of voxel v by |v|, TLG(O) is computed from

$$TLG(O) = |v| \sum_v d_O(I_S(v))\, I_S(v)\, fm_O(v). \tag{2}$$

In words, TLG(O) (expressed in cc) is a weighted sum of the SUV values of voxels within object mask FM(O) multiplied by the voxel volume |v| (expressed in cc), assuming all voxels are of the same size. There are two weights for each voxel – mask weight $fm_O(v)$ and disease weight $d_O(x)$. Mask weight $fm_O(v)$ is binary when FM(O) is specified by a manual recognition method, or a fuzzy membership value when O is recognized by an automated method such as the Automatic Anatomy Recognition (AAR) methodology of [3]. The fuzzy treatment in disease quantification enables us to handle both the segmentation issue of deciding whether or not a voxel belongs to a lesion and determining the SUV measurement at each voxel without explicit partial volume correction and committing a binary segmentation. Consequently, if done properly, we believe the process becomes more stable and repeatable.

To estimate the parameters of $g_n(x)$, we use normal data sets and the ground truth delineations of normal samples of objects. Optimal estimations of parameters (μ_d, σ_d), denoted (m_d, s_d), are determined through an optimization process of minimizing the difference between "true" estimation of total lesion glycolysis, denoted $TLG_t(O)$, which we assume to be available, and our fuzzy estimation TLG(O) in (2). Since different objects can differ in their SUV distributions in normal and disease conditions, this optimization must be performed separately for each object (and also disease type). For object O, the optimum parameters are given by

$$(m_d, s_d) \in \underset{\mu_d, \sigma_d}{\operatorname{argmin}} \left[\sum_J (TLG(O) - TLG_t(O))^2 \right], \tag{3}$$

where J denotes the set of training images used for optimization. The determination of true disease burden $TLG_t(O)$ is really challenging. Although commercial clinical image analysis software systems are available [4, 5], which can perform expert guided disease measurement, their performance is not stable for lesions that are not focal, large, discrete, and sufficiently radiotracer-avid with respect to the background. They also require a region of interest (ROI) to be specified very tightly with respect to each lesion to be measured. Owing to these reasons, we created true measurements by individually thresholding each lesion on the PET image to produce visually optimal results under the guidance of an expert radiologist (Torigian) who has over 12 years of experience in making such measurements clinically. We iterated this process over all lesions considered in this study repeatedly until we did not find the need to change the segmentations. Because of the difficulty of establishing true measurements, we also conducted a phantom study where truth is known absolutely and does not depend on image acquisition or any image operations. We will denote these "true" measurements for object O by $SUVmean_t(O)$, $SUVmax_t(O)$, and $TLG_t(O)$.

In our approach, apart from computing total disease burden within object O, each "lesion" within O can also be detected as individual fuzzy connected components by using the machinery of fuzzy connectedness [6]. Once recognized in this fashion, the disease burden in terms of SUVmean, SUVmax, and TLG can be computed for each lesion of O as well. Recall that O refers to a solid organ, a tissue region, or a lymph node zone. For a lymph node zone, individual lesions correspond to lymph nodes with abnormal radiotracer uptake.

In summary, the disease quantification process consists of a one-time training step to optimally estimate the parameters of the disease map for each object O, recognizing object O in a given image I_C, and computing disease burden of O within O's fuzzy mask at the object level and lesion level.

3. RESULTS

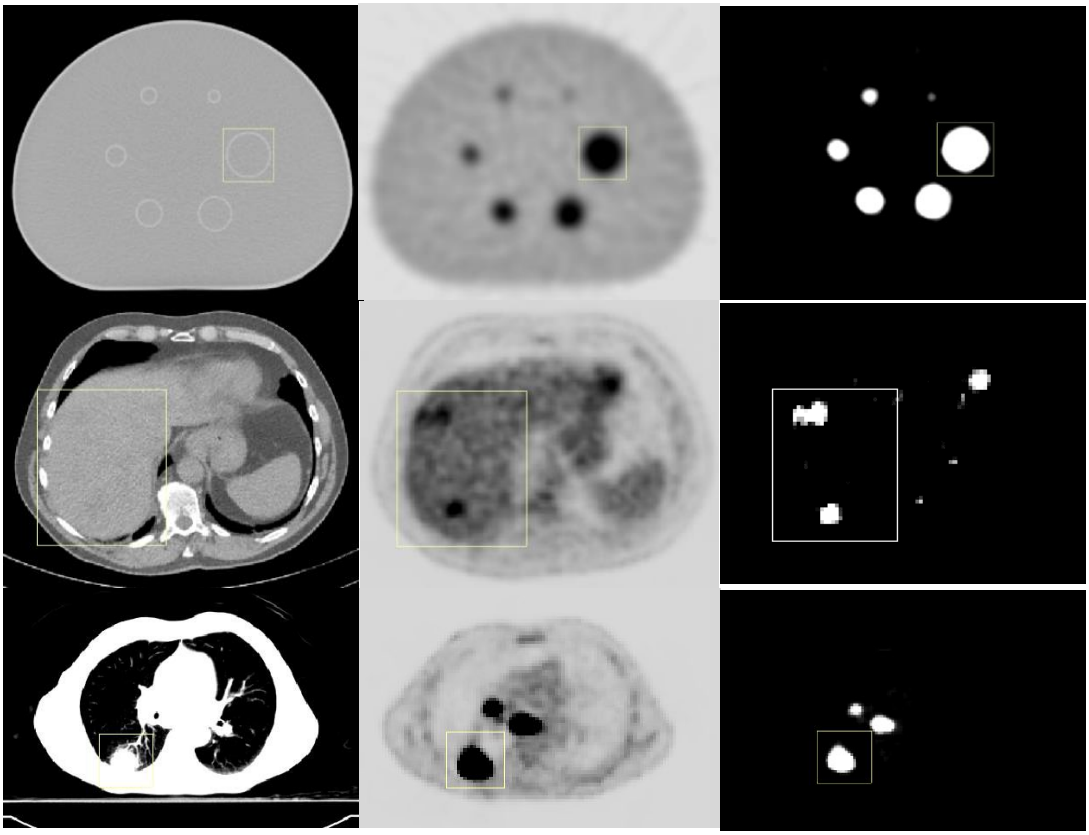

Figure 1. Phantom (1st row), liver lesion (2nd row), and lung lesion (3rd row) examples. CT image slice with a box (left), PET image slice with the same box (middle), and the derived disease map (right). In this example, the liver contains two lesions and the right lung contains one.

Both manual and automatic anatomy recognition strategies are used in this manuscript. Figure 1 and Table 1 show the disease quantification results from the manual recognition strategy. AAR-based object recognition [3] and AAR-DQ are shown in Figure 2 and Table 2. An ROI, FM(O), a binary mask, is specified by a rectangular box to keep manual recognition simple and efficient. We denote manual methods by MO and ML, respectively; in MO the organ is specified by a rectangular box just enclosing the organ, and in ML the box specified just encloses each lesion. Disease quantification results from ML for lung and liver lesions as well as for the simulated lesions in phantom data sets are shown in Table 1 below. Results for the ML method are very similar or slightly better. Figure 1 displays one slice from each of the phantom data set, a patient liver study, and a patient lung study, where the CT and PET slices along with the disease map as gray-scale images are displayed.

Table 1. Mean (sd) of the %deviation of TLG(O) from $TLG_t(O)$ for phantom and patient lesions.

Phantom	Liver lesions		Lung lesions		Mean
All lesions	Large	Small	Large	Small	
1.93 (1.24)	5.51 (4.46)	7.74 (3.4)	8.52 (4.56)	15.13 (3.28)	7.77 (3.33)

Table 1 lists the %deviation of TLG(O) from $TLG_t(O)$ for phantom, liver, and lung lesions. The phantom results demonstrate that when ground truth is known precisely, the DQ-MO approach measures total disease burden very accurately. The larger deviations in measuring smaller lesions in patients also highlight the difficulty in establishing

ground truth for patient images in a reliable and consistent manner, especially for lesions that are not focal, discrete, large, or radiotracer-avid with respect to the background.

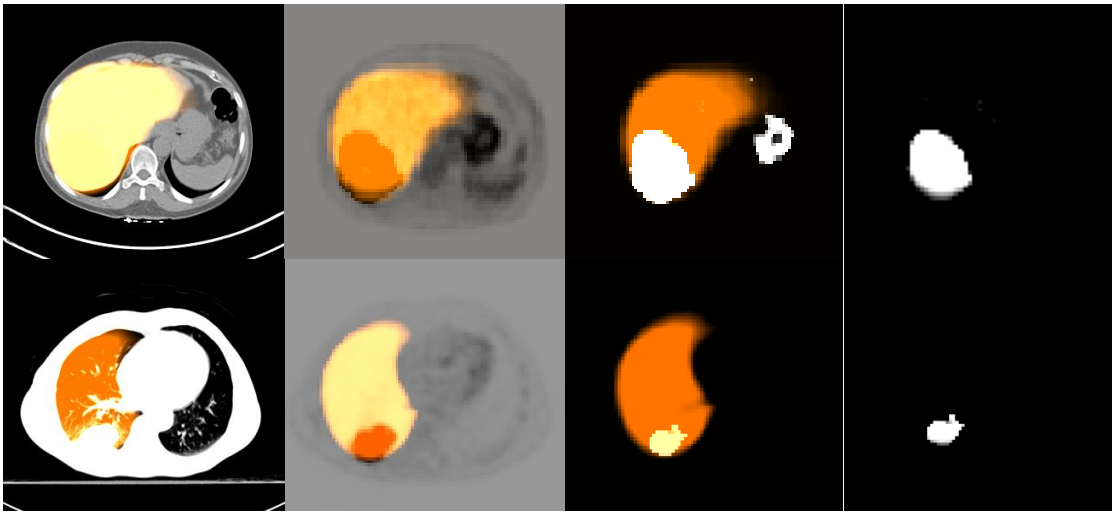

Figure 2. DQ-AAR on liver (top row) and lung (bottom row). Left column is the recognition results (in yellow) of liver and lung on the CT images, and the second column is the recognition results from CT being propagated to PET image. Third column is the recognition results being overplayed onto the disease map of the whole image and last column shows the final disease map.

Table 2. Mean (sd) of the %deviation of TLG(O) from $TLG_t(O)$ for patient lesions with ROI from manual delineation (DQ-MO) and AAR-recognition (AAR-DQ).

Liver lesions		Lung lesions		Mean
DQ-MO	AAR-DQ	DQ-MO	AAR-DQ	
11.30 (7.22)	12.27 (8.85)	9.55 (4.05)	14.67 (9.24)	11.97 (10.85)

Figure 2 shows the disease maps for a liver lesion and for a lung lesion based on AAR recognition results. The regions marked in yellow are the recognized liver and right lung objects. Firstly, we build fuzzy object models for liver and lung with a specific structure [3], and then use the AAR approach to automatically recognize the liver and lung. The above operations are performed on CT images as shown in the first column of Figure 2, and the recognition results are propagated to the PET images as shown in the second column. The third column shows the recognition results being overlaid onto the disease map of the whole image, and the last column shows the final disease map inside of the liver and right lung. Table 2 shows the mean (sd) of the %deviation of TLG(O) from $TLG_t(O)$ for patient lesions with DQ-MO and AAR-DQ approaches. Generally, DQ-MO can achieve disease quantification accuracy at almost 90% and over 85% for AAR-DQ.

4. CONCLUSIONS

In this paper, we demonstrate a practical way to decouple disease quantification on PET/CT images from explicit dependence on image segmentation (meaning delineation). The approach enables disease quantification based only on object recognition results without calling for explicit partial volume correction or delineation. The method shows the potential of the approach for body-wide disease quantification based on PET/CT images when combined with body-wide anatomy recognition methods [3]. We are currently exploring these extensions as well as extensions to just CT or MRI alone, to PET/CT images acquired using radiotracers other than FDG, and disease maps other than Gaussian.

Acknowledgement

The authors acknowledge philanthropic support from James and Frances Maguire, Margarita Louis-Dreyfus, and Sharon Berman for the Lymphoma Program at the Abramson Cancer Center of the University of Pennsylvania.

References

[1] Doyle W., "Operations Useful for Similarity-Invariant Pattern Recognition," Journal of the ACM, 9, 259-267 (1962).
[2] Narasimhan R, Fornango JP, "Some Further Experiments in the Parallel Processing of Pictures," IEEE Transaction on Electronic Computers EC-13, 748-750 (1963).
[3] Udupa JK, Odhner D, Zhao LM, Tong YB, Matsumoto M.M.S, Ciesielski KC, Falcao AX, Ciesielski PV, Syedmehrdad B., Sin S, Arens R, Torigian DA. "Body-Wide Hierarchical Fuzzy Modeling, Recognition and Delineation of Anatomy in Medical Images," Medical Image Analysis 18, 752-771 (2014).
[4] Hofheinz F, Poetzsch C, van den Hoff J, 2007. Quantitative 3D ROI delineation in PET: algorithm and validation. *J Nucl Med* 48 (Supplement 2), 407P.
[5] Hofheinz F, Dittrich S, Potzsch C, Hoff J, 2010. Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold. *Phys Med Biol* 55(4), 1099-1113.
[6] Udupa JK, Samarasekera S. Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation, *Graphical Models and Image Processing*, 58:246-261, 1996.
[7] NEMA, I.B.P.S., "International Standard: Radionuclide Imaging Devices Characteristics and Test Conditions Part 1: Positron Emission Tomographs," Technical Report. International Electrotechnical Commission (IEC) (1998).
[8] Smith K, Clark K, Bennett W, Nolan T, Kirby J, Wolfsberger M, Moulton J, Vendt B, Freymann J., "Data From RIDER_PHANTOM_PET-CT," (2015).
[9] Kwee TC, Torigian DA, Alavi A, "Overview of positron emission tomography, hybrid positron emission tomography instrumentation, and positron emission tomography quantification," J Thorac Imaging. 28(1), 4-10 (2013).