Medical Physics
The International Journal of Medical Physics Research and Practice

Automatic anatomy recognition in whole-body PET/CT images

Huiqian Wang, Jayaram K. Udupa, Dewey Odhner, Yubing Tong, Liming Zhao, and Drew A. Torigian

Citation: Medical Physics **43**, 613 (2016); doi: 10.1118/1.4939127
View online: http://dx.doi.org/10.1118/1.4939127
View Table of Contents: http://scitation.aip.org/content/aapm/journal/medphys/43/1?ver=pdfcov
Published by the American Association of Physicists in Medicine

Articles you may be interested in

Local respiratory motion correction for PET/CT imaging: Application to lung cancer
Med. Phys. **42**, 5903 (2015); 10.1118/1.4930251

Evaluation of radiation dose and image quality of CT scan for whole-body pediatric PET/CT: A phantom study
Med. Phys. **41**, 092505 (2014); 10.1118/1.4893273

Description and assessment of a registration-based approach to include bones for attenuation correction of whole-body PET/MRI
Med. Phys. **40**, 082509 (2013); 10.1118/1.4816301

Improved quantitation for PET/CT image reconstruction with system modeling and anatomical priors
Med. Phys. **33**, 4095 (2006); 10.1118/1.2358198

Four-dimensional (4D) PET/CT imaging of the thorax
Med. Phys. **31**, 3179 (2004); 10.1118/1.1809778

SUN NUCLEAR corporation
AUTOMATE YOUR QA

MACHINE
SNC Machine™
Increased Objectivity for a Smarter Workflow

PATIENT
PerFRACTION™ 3D
Actionable Insights for Enhanced Safety

Learn More

Automatic anatomy recognition in whole-body PET/CT images

Huiqian Wang
College of Optoelectronic Engineering, Chongqing University, Chongqing 400044, China and Medical Image Processing Group Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Jayaram K. Udupa,[a)] Dewey Odhner, and Yubing Tong
Medical Image Processing Group Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Liming Zhao
Medical Image Processing Group Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104 and Research Center of Intelligent System and Robotics, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Drew A. Torigian
Medical Image Processing Group Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 26 July 2015; revised 12 December 2015; accepted for publication 16 December 2015; published 7 January 2016)

Purpose: Whole-body positron emission tomography/computed tomography (PET/CT) has become a standard method of imaging patients with various disease conditions, especially cancer. Body-wide accurate quantification of disease burden in PET/CT images is important for characterizing lesions, staging disease, prognosticating patient outcome, planning treatment, and evaluating disease response to therapeutic interventions. However, body-wide anatomy recognition in PET/CT is a critical first step for accurately and automatically quantifying disease body-wide, body-region-wise, and organwise. This latter process, however, has remained a challenge due to the lower quality of the anatomic information portrayed in the CT component of this imaging modality and the paucity of anatomic details in the PET component. In this paper, the authors demonstrate the adaptation of a recently developed automatic anatomy recognition (AAR) methodology [Udupa *et al.*, "Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images," Med. Image Anal. **18**, 752–771 (2014)] to PET/CT images. Their goal was to test what level of object localization accuracy can be achieved on PET/CT compared to that achieved on diagnostic CT images.

Methods: The authors advance the AAR approach in this work in three fronts: (i) from body-region-wise treatment in the work of Udupa *et al.* to whole body; (ii) from the use of image intensity in optimal object recognition in the work of Udupa *et al.* to intensity plus object-specific texture properties, and (iii) from the intramodality model-building-recognition strategy to the intermodality approach. The whole-body approach allows consideration of relationships among objects in different body regions, which was previously not possible. Consideration of object texture allows generalizing the previous optimal threshold-based fuzzy model recognition method from intensity images to any derived fuzzy membership image, and in the process, to bring performance to the level achieved on diagnostic CT and MR images in body-region-wise approaches. The intermodality approach fosters the use of already existing fuzzy models, previously created from diagnostic CT images, on PET/CT and other derived images, thus truly separating the modality-independent object assembly anatomy from modality-specific tissue property portrayal in the image.

Results: Key ways of combining the above three basic ideas lead them to 15 different strategies for recognizing objects in PET/CT images. Utilizing 50 diagnostic CT image data sets from the thoracic and abdominal body regions and 16 whole-body PET/CT image data sets, the authors compare the recognition performance among these 15 strategies on 18 objects from the thorax, abdomen, and pelvis in object localization error and size estimation error. Particularly on texture membership images, object localization is within three voxels on whole-body low-dose CT images and 2 voxels on body-region-wise low-dose images of known true locations. Surprisingly, even on direct body-region-wise PET images, localization error within 3 voxels seems possible.

Conclusions: The previous body-region-wise approach can be extended to whole-body torso with similar object localization performance. Combined use of image texture and intensity property yields the best object localization accuracy. In both body-region-wise and whole-body approaches, recognition performance on low-dose CT images reaches levels previously achieved on diagnostic CT images. The best object recognition strategy varies among objects; the proposed framework

however allows employing a strategy that is optimal for each object. © *2016 American Association of Physicists in Medicine.* [http://dx.doi.org/10.1118/1.4939127]

Key words: image segmentation, PET/CT imaging, object recognition, texture, fuzzy models

1. INTRODUCTION

Whole-body positron emission tomography (PET)/computed tomography (CT) imaging provides coregistered molecular and anatomic images of patients in one single procedure. This modality has recently become the standard method for clinical molecular imaging assessment of patients with various disease conditions, including cancer.[2–4] In clinical research, whole-body PET/CT is also frequently utilized as a robust means of noninvasively providing quantitative information about diseases of interest and the effects of experimental therapeutic interventions upon lesions and other tissues in the body. With the rapid growth of PET/CT-based medical applications, anatomy recognition in whole-body PET/CT images is critical for quantifying body-wide disease burden. This, however, is challenging due to the low spatial resolution of PET image with its unclear anatomy reference and due to the low contrast resolution of the associated unenhanced low-dose CT image. To illustrate this point, in Fig. 1, we display a radiologically normal sample slice of a contrast-enhanced diagnostic CT image, an approximately matching unenhanced low-dose CT image from a different subject, and the PET image matching the low-dose CT image following intravenous administration of ^{18}F-fluorodeoxyglucose (FDG).

Many image segmentation methods have been investigated and applied to PET/CT images, but mostly for segmenting pathological regions. These include thresholding,[5–9] gradient-directed,[10] region growing,[11] clustering,[12] deformable model driven techniques,[13] and graph-based approaches.[14] Thresholding based on PET standardized uptake values (SUVs) is the most common method employed in PET segmentation which is comparable in popularity to manual delineation in clinical practice. There are also thresholding methods based on a percentage of the maximum SUV,[6,7] the signal to background ratio or SBR,[5] and other variants.[8] Besides fixed thresholding, many adaptive thresholding methods have also been proposed[9] in which an optimal threshold level is determined automatically. Despite their simplicity and ease of use, the threshold methods are sensitive to noise and always require prior knowledge of the tumor volume of interest before segmentation.[7] Gradient-based methods[10] are not well suited for noisy and low resolution PET images and need denoising and deblurring operations in advance. Region growing methods require seed points selected from manual segmentation drawn or maximum intensity pixels and may fail to segment heterogeneous regions due to nonuniform uptake distributions.[11] Clustering methods consider the fuzzy nature of lesion boundary, but their performance is suspect for small or complex lesions.[12] Deformable models are topologically adaptive which permits smooth segmentation, but they require sufficient homogenous and clear regions of interest for effectiveness. Graph-based methods such as graph cuts and fuzzy connectedness are feasible for complex and fuzzy images. However, they require adequate seed sets and may have a leakage issue.[15] Recently, several studies advanced cosegmentation ideas which incorporate both molecular information from PET and associated anatomical information from CT simultaneously to achieve good lesion delineation performance on clinical data sets.[14,16,17]

In many body-wide applications, anatomy recognition of objects in low-dose CT is a precursor and essential first step to further delineation and quantification of disease and tissue composition body-wide, by body region, by organ system, or by organ. This is a fundamental and challenging problem in all body-wide applications. As seen from the above literature review, most published papers on PET/CT image analysis have focused only on pathological region recognition and not organ anatomy recognition. References 18–24 studied anatomy recognition on PET and low-dose CT images, but mostly one specific object without considering multiple objects in the whole-body. Recently, there have been studies toward multiorgan segmentation. References 25–28 studied multiorgan segmentation on diagnostic CT images, but mostly their applications are constrained to a body region and not extended to low-dose CT or PET images and body-wide which is more challenging. Reference 29 proposed a regression approach to

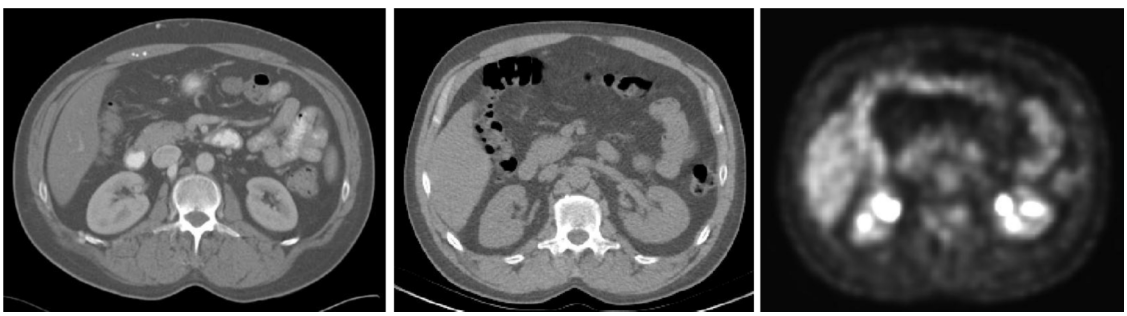

FIG. 1. L to R: One slice of contrast-enhanced diagnostic CT image, a corresponding unenhanced low-dose CT image from a different patient at approximately the same anatomic level, and the associated FDG-PET image.

ocalize multiple organs on diagnostic CT images in a rectangular box form and demonstrated on 26 organs. Reference 30 used information theory to multiorgan localization on PET/CT images and demonstrated on six organs. Such efforts will become increasingly important for quantification and characterization of disease body-wide. Therefore, the development of a robust object localization/recognition method that works on low-dose CT or PET images body-wide would constitute an advancement of the state-of-the-art in PET/CT quantitative image analysis.

Motivated by body-wide applications and generalizability of methods, we have recently developed an automatic anatomy recognition (AAR) methodology[1] and demonstrated its operability in three different body regions on over 35 objects in contrast-enhanced diagnostic CT and MR images. The AAR methodology consists of two related processes: recognition and delineation. Recognition is a high-level process of determining the whereabouts of the object in the image. Delineation is a low-level process of determining the precise spatial extent of the object in the image. It has been demonstrated in Ref. 1 that with acceptable recognition, good delineation results can be obtained. In the AAR approach, a fuzzy anatomy model of the body region with all its major objects arranged in a hierarchy is first built by utilizing existing clinical image data sets. Subsequently, the objects are recognized in any given image following the hierarchical order and exploiting object relationships. Subsequent to recognition, fuzzy connectedness algorithms tightly integrated with the fuzzy models are used to delineate the objects in the image, also following the hierarchical order. The generalizability and performance of the AAR methodology led us to study its applicability to whole-body PET/CT images.

The aim of the present work was thus to investigate strategies for adapting the previous AAR system to PET/CT images and also advance its approach in several key ways. First, instead of the body-region-wise treatment of Ref. 1, the whole body (hereafter, by whole body, we mean the body torso) is considered. In the process, exploitation and consideration of the relationships among objects in different body regions have become possible for performing AAR. This also obviates the need for breaking up the whole-body image into images corresponding to individual body regions (either manually or in a reliable automatic manner) and truly facilitates dealing with whole-body PET/CT images. Second, instead of the use of just image intensity for object recognition in the previous process, we demonstrate the utility of object textural properties together with intensity in improving recognition. In the process, the previous optimal threshold recognition strategy is generalized to any input image, original or derived. Third, we study the process of recognition on PET images alone, CT images alone, and the two together, and demonstrate that good recognition accuracy can be obtained even on just PET images. Fourth, we include a new body region, namely, the pelvis, which we did not previously consider in AAR. These advances are described in Sec. 2. In Sec. 3, we explain the experiments and evaluate results for various key strategies involving combinations of approaches for model building (from diagnostic CT or low-dose CT), images tested for recognition (low-dose CT, PET, PET, and CT combined, or texture-derived images), and body region consideration (body-region-wise or whole body). In Sec. 4, we summarize our findings and point out the key new outcomes, hurdles and limitations encountered, and future work. A very preliminary version of this work was presented at the SPIE Medical Imaging Conference held in February 2015 in Orlando.[31]

2. WHOLE-BODY AAR

In this paper, we focus only on the *recognition* process of AAR and investigate it closely in order to achieve good performance on PET/CT images. We note that several papers have been published in the literature on the process of just object recognition or localization in the form of locating a rectangular box enclosing the object,[20,23,29,32–38] albeit on diagnostic CT images. Thus, advancing just object recognition strategies without focusing on subsequent delineation is important in its own right. We also note that AAR recognition is different from these approaches in that it actually finds the optimal pose of a fuzzy model of the object. Also, as we demonstrated in Ref. 1, good recognition strategies can lead to good model-based delineation of objects subsequently.

We will follow the terminology of Ref. 1 but introduce new notations as well. G is the population group for which the models are built. B is the body region of focus, with $B \in \{\text{Thx}, \text{Abd}, \text{Plv}, \text{BT}\}$ where the elements represent, respectively, body regions thorax, abdomen, pelvis, and the body torso, the latter being considered to be the union of the other three body regions. $O_1, \ldots, O_L$ is L objects or organs of B. $I^m = \{I_1^m, \ldots, I_N^m\}$ is the set of images in modality m of a body region B for G from N subjects. $I^b = \{I_{n,l}: 1 \le n \le N \text{ and } 1 \le l \le L\}$ is the set of all binary images used for model building, $I_{n,l}$ being the binary image representing O_l in image I_n^m. Note that I^b may be derived from modality $m1$ for building the models but the models may be deployed on images from another modality $m2$ for AAR. FM(O_l): Fuzzy model of object O_l derived from the set of all binary images $I_l^b = \{I_{n,l}: 1 \le n \le N\}$ of O_l. FAM(B, G): Fuzzy anatomy model of the whole-object assembly in B. FM^t(O_l) is transformed (adjusted) FM(O_l) corresponding to the state when O_l is recognized in a given patient image I of B. lCT is low-dose unenhanced CT of whole-body PET/CT acquisition. dCT is diagnostic CT which is usually performed for a body region and not whole body. In this paper, dCT images are used only for model building. PET&CT are a fused image derived from CT and PET images of the same subject. tcCT, trCT, teCT, and tbCT are texture images derived from lCT corresponding to contrast, correlation, energy, and ball-scale properties, respectively. tcPET, trPET, tePET, and tbPET are texture images derived from PET corresponding to contrast, correlation, energy, and ball-scale properties, respectively. tmCT is a texture-based membership image derived from lCT and teCT images. tmPET is a texture membership image derived from PET and tcPET images. Details of how the fused image, the texture images and the texture membership images are created are described later in this section.

We will present different AAR strategies involving three main variables: the body region $B \in \{\text{Thx}, \text{Abd}, \text{Plv}, \text{BT}\}$,

TABLE I. Recognition strategies presented in the paper and their denotation.

Recognition approach	Body region	Source images for model building	Image modality for recognition
*B-x-s-d*CT-*m-l*CT	*x* = Thx or Abd	*d*CT	Low-dose CT
*B-x-s-d*CT-*m*-PET	*x* = Thx or Abd	*d*CT	PET
*B-x-s-d*CT-*m*-PET&CT	*x* = Thx or Abd	*d*CT	Combined PET and CT
*B-x-s-d*CT-*m-tc*CT	*x* = Thx or Abd	*d*CT	Texture (contrast) image of low-dose CT
*B-x-s-d*CT-*m-tr*CT	*x* = Thx or Abd	*d*CT	Texture (correlation) image of low-dose CT
*B-x-s-d*CT-*m-te*CT	*x* = Thx or Abd	*d*CT	Texture (energy) image of low-dose CT
*B-x-s-d*CT-*m-tb*CT	*x* = Thx or Abd	*d*CT	Texture (b-scale) image of low-dose CT
*B-x-s-d*CT-*m-tc*PET	*x* = Thx or Abd	*d*CT	Texture (contrast) image of PET
*B-x-s-d*CT-*m-tr*PET	*x* = Thx or Abd	*d*CT	Texture (correlation) image of PET
*B-x-s-d*CT-*m-te*PET	*x* = Thx or Abd	*d*CT	Texture (energy) image of PET
*B-x-s-d*CT-*m-tb*PET	*x* = Thx or Abd	*d*CT	Texture (b-scale) image of PET
*B-x-s-d*CT-*m-tm*CT	*x* = Thx or Abd	*d*CT	Texture membership image for low-dose CT
*B-x-s-d*CT-*m-tm*PET	*x* = Thx or Abd	*d*CT	Texture membership image for PET
B-BT-*s-l*CT-*m-l*CT	Whole-body torso	*l*CT	Low-dose CT
B-BT-*s-l*CT-*m-tm*CT	Whole-body torso	*l*CT	Texture membership image for low-dose CT

source image modality used for model building, $s \in \{d\text{CT}, l\text{CT}\}$, and the image modality of the test images on which recognition is performed, $m \in \{l\text{CT}, \text{PET}, \text{PET\&CT}, tc\text{CT}, tr\text{CT}, te\text{CT}, tb\text{CT}, tc\text{PET}, tr\text{PET}, te\text{PET}, tb\text{PET}, tm\text{CT}, tm\text{PET}\}$. The different AAR strategies proposed in this paper will be compactly denoted by *B-x-s-y-m-z* by using the values assumed by the three variables as mnemonics. As an example, *B*-Thx-*s-d*CT-*m-l*CT denotes a recognition strategy for the thoracic body region involving models created from *d*CT and recognition performed on *l*CT images. Note that not all possible combinations of the values assumed by the variables denote plausible strategies. For example, *B*-BT-*s-d*CT-*m-l*CT is not feasible since *d*CT images are typically not available for whole-body torso. Several among the feasible combinations have not been tested in this work, such as *B*-BT-*s-l*CT-*m-PE*T, based on our initial experience with the approaches. Table I lists the 15 processes which we thought were promising and implemented and tested in this paper.

Our proposed advances on the original AAR approach are illustrated in Fig. 2. These will be discussed step-by-step in Secs. 2.A and 2.B.

2.A. Model building

2.A.1. Setting up image database

This retrospective study was conducted 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. As in Ref. 1, we make use of existing patient images which are near normal for the body region under consideration. For $B \in \{\text{Thx}, \text{Abd}\}$, a board certified radiologist (co-author DAT) selected all *d*CT image data sets from our health system patient image database in such a manner that the images appeared radiologically normal for the body region considered, with exception of minimal incidental focal abnormalities such as cysts and small pulmonary

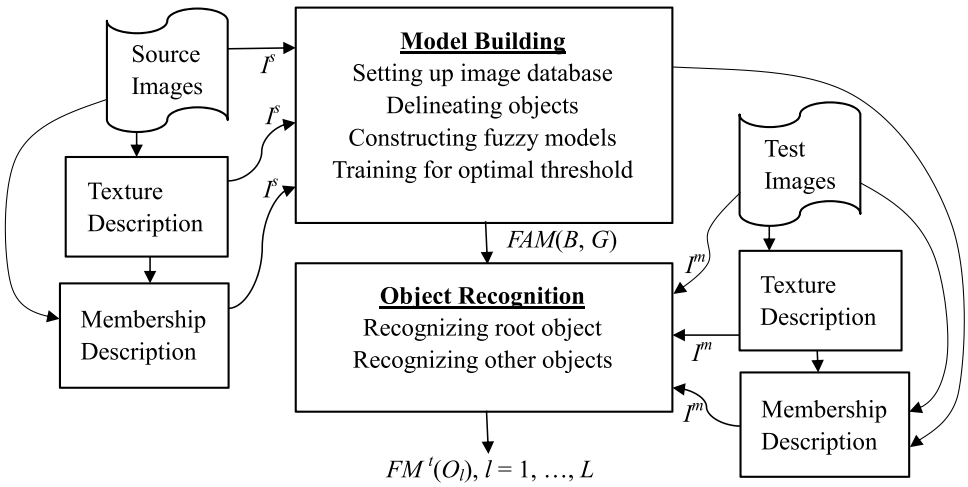

FIG. 2. A schematic depiction of the proposed recognition strategies of the AAR approach.

nodules. For these two body regions, the population groups considered have an age range of approximately 50–60 yr. For $B \in \{\text{Plv,BT}\}$, the same radiologist selected whole-body FDG PET/CT images of patients in the age range of 31–71 yr. The separate CT and PET images of these data sets formed the *l*CT and PET modalities.

2.A.2. Delineating objects

Following the AAR methodology, we define precisely each body region $B \in \{\text{Thx, Abd, Plv, BT}\}$ and each object considered in each body region, see Table II. For Thx and Abd, we use the definitions from Ref. 1: Thx extends axially from 5 mm below the base of the lungs to 15 mm above their apex. Abd extends inferiorly from the bifurcation point of the abdominal aorta into common iliac arteries to superiorly the superior aspect of the liver. The Plv region is defined to extend inferiorly from the inferior aspect of the ischial tuberosities of the pelvis to the inferior boundary of the Abd region. Since BT is defined to be the union of Thx, Abd, and Plv regions, it extends from the inferior aspect of Plv to the superior aspect of Thx. All source and test image sets are trimmed as per these definitions for each $B \in \{\text{Thx, Abd, Plv, BT}\}$. In this first effort on adapting AAR to PET/CT images, we focused on the 18 objects listed in Table II. We call objects that constitute unions of basic objects *composite objects*. The latter are useful for improving accuracy of object recognition by localizing them first more globally and then honing in on their component objects relative to them taken as parent reference objects.

All objects are delineated by strictly following their definition and using a combination of automatic and interactive tools to minimize human labor and maximize accuracy. All tracings were examined for accuracy by several checks—3D surface renditions of objects from each subject in each body region in various object combinations as well as a subsequent slice-by-slice verification of the delineations overlaid on the gray images. As per our notation, the set of binary images generated in this step for all objects for a fixed B and G is $I^b = \{I_{n,l}: 1 \le n \le N \text{ and } 1 \le l \le L\}$, and the set just for object O_l is $I_l^b = \{I_{n,l}: 1 \le n \le N\}$.

2.A.3. Constructing fuzzy models

For a body region B of subject group G, the fuzzy anatomy model,[1] $\text{FAM}(B,G)$, is denoted by $\text{FAM}(B,G) = (H,M,\rho,\lambda,\eta)$. H is a hierarchy of objects in B, represented as a tree. This tree structure permits imposing an order among objects, rather than treating them as an amorphous collection and allows encoding nonlinear and very detailed anatomic information about group G into the model. M is a set of fuzzy models, one model for each of the L objects in B, $M = \{\text{FM}(O_k): k = 1,\ldots,L\}$. ρ describes the parent-to-offspring relationship in H over G. λ is a family of scale factor ranges. η denotes a set of measurements pertaining to the object assembly in B including intensity properties and all learned parameters that are used in recognition and delineation.

To build $\text{FAM}(B, G)$ for body torso BT, we chose the anatomical hierarchy depicted in Fig. 3. For thoracic and abdominal body regions, we chose a hierarchy similar to that described in Ref. 1 in its order but simplified to the fewer objects considered in this paper, also depicted in Fig. 3. We also added a new body region, namely, the pelvis.

The fuzzy model $\text{FM}(O_l)$ of each object O_l encodes the variations in the object's form over G and is independent of

TABLE II. Definition of body regions and objects.

Abbreviation	Description
Thx	Thoracic region extending from 5 mm below the base of the lungs to 15 mm above their apex
Abd	Abdominal region extending from the point of bifurcation of the abdominal aorta into common iliac arteries to the superior aspect of the liver
Plv	Pelvic region extending from the inferior aspect of the ischial tuberosities of the pelvis to the inferior boundary of the Abd region
BT	Body torso defined to extend from the inferior aspect of the Plv region to the superior aspect of the Thx region
BTSkn	The outer boundary of the body torso skin. The interior region constitutes the entire BT body region
TSkn	The outer boundary of the thoracic skin. The interior region constitutes Thx
TO	A composite object called thoracic objects made up of PS and PC
PS	A composite object called pleural space made up of LPS and RPS
LPS	Left pleural space—the outer boundary of the left lung along the left pleura
RPS	Right pleural space—the outer boundary of the right lung along the right pleura
PC	Region within the boundary of pericardial sac. The superior aspect is defined by the branching of the main pulmonary artery
ASkn	The outer boundary of the abdominal skin. The interior region constitutes Abd
AO	A composite object called abdominal objects made up of Lvr, Kd, and Spl
Lvr	The outer boundary of the liver. The intrahepatic portal veins and hepatic arteries are included in this region
Kd	A composite object made up of LKd and RKd
LKd	The outer boundary of the left kidney. All external blood vessels are excluded
RKd	The outer boundary of the right kidney. All external blood vessels are excluded
Spl	The outer boundary of the spleen. All external blood vessels are excluded
PO	A composite object called pelvic objects made up of Bld, PSAT, and PVAT
Bld	The outer boundary of the bladder. Ureters and urethra are excluded
PSAT	Adipose tissue in the subcutaneous region in the pelvis
PVAT	Adipose tissue internal to the pelvic body wall musculature, pelvic floor musculature, and pelvic bones

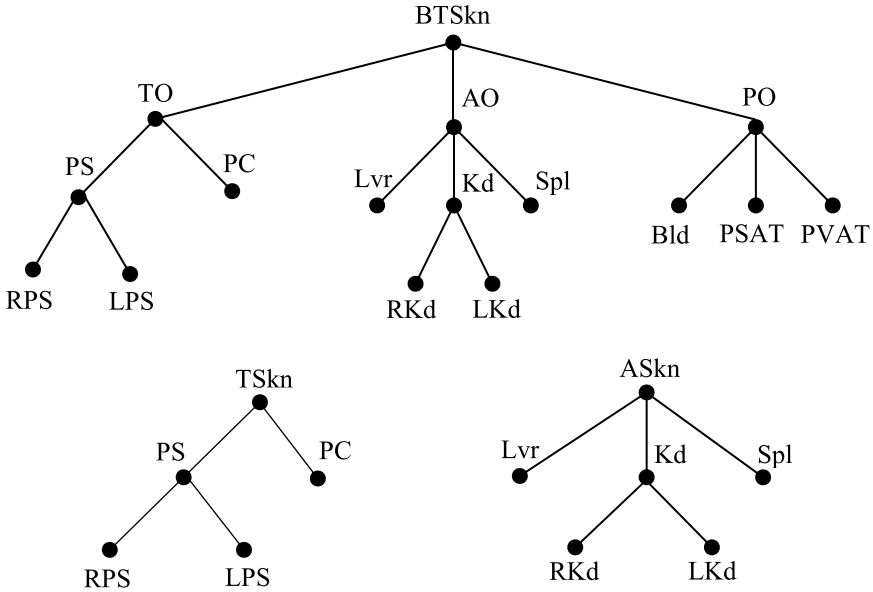

FIG. 3. Top: Hierarchy chosen for whole-body torso. Object abbreviations are described in Table II. Bottom: Hierarchy for Thx and Abd.

image intensity and modality. As a result, the fuzzy model set M can be used in different modalities and on other derived images as depicted by the modality indicator $m \in \{l\text{CT}, \text{PET}, \text{PET\&CT}, tc\text{CT}, tr\text{CT}, te\text{CT}, tb\text{CT}, tc\text{PET}, tr\text{PET}, te\text{PET}, tb\text{PET}, tm\text{CT}, tm\text{PET}\}$. FM($O_l$) is built from training binary images in the set I_l^b as described in Ref. 1. Briefly, this process consists of estimating the mean shape length and geometric center of O_l over G, repositioning all samples of O_l to this mean position and rescaling them to mean shape length. Subsequently a distance transform is applied to each resulting sample to propagate the shape form inward and outward from its boundary; the distance values are then averaged, and the average distance is transformed to a fuzzy object membership value. From the repositioned and resized samples, the parent-to-offspring relationship ρ_l of O_l with respect to its unique parent in the hierarchy is estimated over G. Similarly, the size variation bounds $\lambda = \{\lambda_l = [\lambda^b{}_l, \lambda^h{}_l]\colon 1 \le l \le L\}$ over G are estimated from the same samples using the shape length of each O_l.

The only component that needs change when using FAM(B, G) in different image modalities for a given B and G is the fifth element η of FAM(B, G). This entity is a place holder for a variety of measurements pertaining to the body region B which include normative descriptions of anatomic, physiological, and functional properties of the object assembly in B as well as image intensity statistics. From the perspective of object recognition, there is a parameter called *optimal threshold interval* stored in η which is associated with each object and which is image type/modality-dependent. This parameter is estimated for any given image type/modality as described below.

2.A.4. Training for optimal threshold

The basic object recognition engine used in this paper is the thresholded optimal search method described in Ref. 1 for dCT images of Thx and Abd. It requires determining an optimal threshold interval Th_l (in dCT) for each object O_l in B. To understand the process of estimating Th_l, it is necessary to understand overall the recognition process. Therefore, we will first briefly outline the earlier recognition process and then describe how Th_l is estimated. In Sec. 2.B, we will delineate how this method is generalized to the scenarios depicted in Table I. So assume for now that Th_l has already been estimated. To recognize O_l in any given test dCT image I, the initial pose of FM(O_l) in I is first determined with respect to the parent of O_l (assuming that the parent has already been recognized in the hierarchical order) from knowledge of parent-to-offspring relationship ρ_l. Then, a search for optimal pose for FM(O_l) is made within a region of the pose space defined by the variation observed in ρ_l. The optimal pose, determined by exhaustive sampled search within this region, is defined by the pose that yields minimal mismatch between the binary image resulting from thresholding I at Th_l and the pose-adjusted FM(O_l).

In the previous AAR approach,[1] the same training image sets were used for model building (meaning constructing M, ρ, λ, and η) as well as estimating Th_l. In view of the expanded scope of the proposed recognition process (as per Table I), the method of estimating Th_l in this paper is different from the earlier approach. We will first briefly outline the previous approach and then present the new strategy. The previous method to estimate Th_l involved a rehearsal of the recognition process carried out on the training image set. Since we do not know Th_l but have the true segmentations, the idea was to test recognition efficacy for each of a number of threshold intervals t and then select the interval Th_l that yielded the best match of the model with the known true segmentations for O_l. Let I_n^m be the training gray image from which the true binary delineation $I_{n,l}$ of O_l is obtained. If $J_n(t)$ is the binary image resulting from thresholding I_n^m at t, then

$$Th_l \in \arg\min_{p,t} \sum_n |(J_n(t) \times \text{FM}^p(O_l)) - I_{n,l}| + |I_{n,l} - (J_n(t) \times \text{FM}^p(O_l))|.$$

In words, the optimal threshold Th_l is found by searching over the pose space over all training data sets and all thresholds the best match between the true segmentation of O_l with the result of thresholding I_n^m restricted to the model. Since this process involves optimization over the pose space and threshold intervals, it was computationally expensive. To keep computation manageable, in Ref. 1 we confined the above search to 81 different t intervals (resulting from 9 different settings at the lower end of t times 9 at its upper end).

The new approach, called AAR-OTE (for optimal threshold estimation), presented below, works not only on the original gray images used for model building but also on any derived image of type $m \in \{l\text{CT}, \text{PET}, \text{PET\&CT}, tc\text{CT}, tr\text{CT}, te\text{CT}, tb\text{CT}, tc\text{PET}, tr\text{PET}, te\text{PET}, tb\text{PET}, tm\text{CT}, tm\text{PET}\}$. The speed of AAR-OTE ensues from the fact that the search here is in the 1D intensity space rather than the 5D pose and intensity space. Consequently, there is also no chance of missing the global optimal which can happen in the earlier approach because of limited search subspace. Let $h(X)$ denote the histogram of image X.

Procedure AAR-OTE

Input: Training gray image set $I^m = \{I_1^m, \ldots, I_N^m\}$ for some modality $m \in \{l\text{CT}, \text{PET}, \text{PET\&CT}, tc\text{CT}, tr\text{CT}, te\text{CT}, tb\text{CT}, tc\text{PET}, tr\text{PET}, te\text{PET}, tb\text{PET}, tm\text{CT}, tm\text{PET}\}$, training binary image set $I_l^b = \{I_{n,l}: 1 \le n \le N\}$ for object O_l, all for the same body region B and group G.

Output: Optimal threshold Th_l for O_l for modality m for B and G.

Begin

1. Compute super mask Q for O_l over all training samples, $Q = I_{1,l} \cup \cdots \cup I_{N,l}$;
2. *For* $n = 1$ to N do
3. Compute histograms $h_n^o(I_{n,l} \times I_n)$ and $h_n^b([Q - I_{n,l}] \times I_n)$.
4. *End*;
5. Determine cumulative histograms $h^o(x) = h_1^o(x) + \cdots + h_N^o(x)$ and $h^b(x) = h_1^b(x) + \cdots + h_N^b(x)$.
7. For any threshold interval Th and any histogram $h(x)$, let $A(h(x), Th)$ be the area of $h(x)$ defined by Th. Find interval $Th_l \in \arg\max_{Th} |A(h^o(x), Th) - A(h^b(x), Th)|$.
8. Output Th_l;

End

Super mask Q represents the union of all training binary image samples for O_l. $h_n^o(x)$ denotes the histogram of the gray values of O_l in modality m in the nth sample.[48] Similarly, $h_n^b(x)$ represents the histogram of the gray values outside O_l but inside Q in the nth sample. Th_l then corresponds to a threshold interval for m that maximally separates the object from the surrounding background over all training samples.

At the end of the model building stage, we have FAM(B, G) complete with fuzzy models of all objects in B and other associated information (ρ and λ) for a given hierarchy, all constructed from one modality $m1$, plus optimal threshold information for each object which may have been derived from a different modality $m2$. All these items of information play a role in the subsequent step of object recognition.

2.B. Object recognition

The *mixed-modality recognition* process (*AAR-MMR*) which is a modification of the original *AAR-R* procedure from Ref. 1 is listed below. We will highlight only the differences but otherwise give a brief but complete description of the process.

Procedure *AAR-MMR*

Input: FAM(B, G), an image I^m of B of some modality $m \in \{l\text{CT}, \text{PET}, \text{PET\&CT}, tc\text{CT}, tr\text{CT}, te\text{CT}, tb\text{CT}, tc\text{PET}, tr\text{PET}, te\text{PET}, tb\text{PET}, tm\text{CT}, tm\text{PET}\}$ for group G such that FAM(B, G) has encoded in it the trained optimal threshold for m for all objects of B.

Output: Optimally pose-adjusted fuzzy models $FM^t(O_l)$, $l = 1, \ldots, L$.

Begin

1. Call *MMR-ROOT* to recognize the root object of H in I^m;
2. Repeat
3. Find the next offspring object O_k in H to *recognize* in I^m;
4. Knowing $\text{FM}^t(O_l)$, ρ_k, and λ_k, call *MMR-OBJECT* to recognize O_k in I^m;
5. *Until* all objects are covered in H;
6. Output $\text{FM}^t(O_l)$, $l = 1, \ldots, L$;

End

To make the explanation of *AAR-MMR* more tangible, we will take strategy *B-x-s-d*CT-*m*-PET from Table I with $x = \text{Thx}$ as an example. In this case, the fuzzy models $\text{FM}(O_l)$, $l = 1, \ldots, L$, are built from dCT and the binary images of the objects of the thorax, and so also object position relationship information ρ and size variation information λ are all associated with dCT. However, since recognition is to be performed on m = PET images, the optimal threshold information Th_l for every object is obtained from PET images. *AAR-MMR* proceeds hierarchically following H in a breadth-first manner. The root object O_1 (which is TSkn—the

TABLE III. Image data sets used in the experiments.

Group	Number of subjects	Image modality	Imaging protocol details	Image information
31–71 male	16 normal	PET/CT	Unenhanced, axial	PET: $144 \times 144 \times$ 338–443, $4 \times 4 \times 4$ mm^3 CT: $512 \times 512 \times 338$–443, $1.2 \times 1.2 \times 4$ mm^3
50–60 male	25 normal	Diagnostic CT	Contrast-enhanced, axial, breath-hold	$512 \times 512 \times 51$–69, $0.9 \times 0.9 \times 5$ mm^3
50–60 male	25 normal	Diagnostic CT	Contrast-enhanced, axial, breath-hold	$512 \times 512 \times 38$–55, $0.9 \times 0.9 \times 5$ mm^3

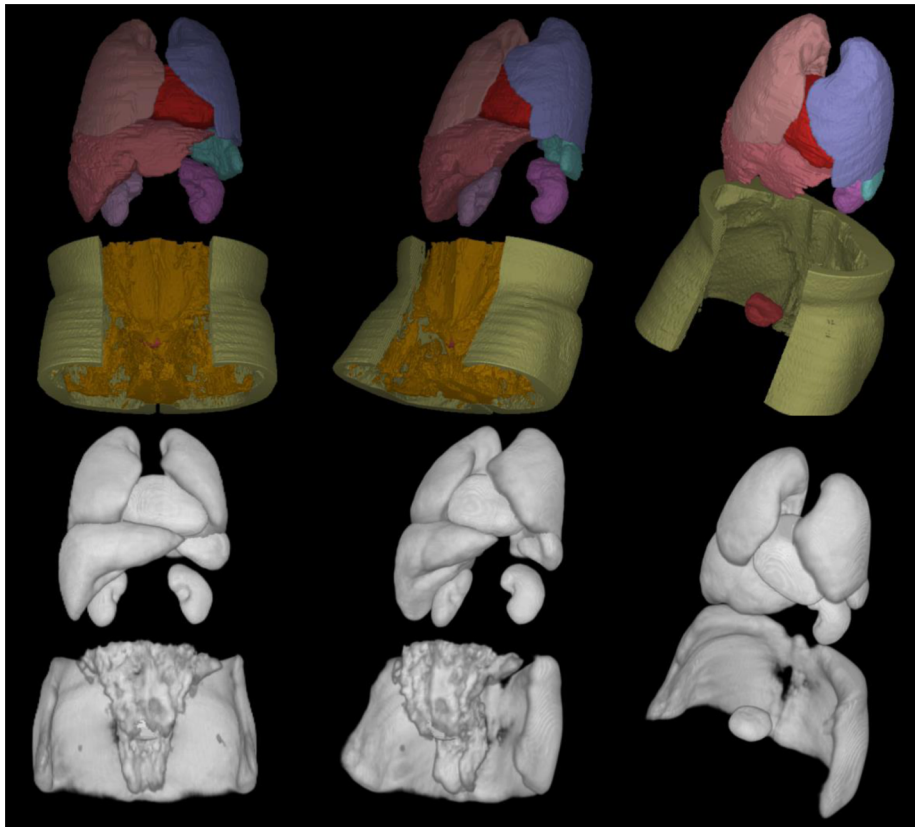

FIG. 4. Surface renditions of sample objects derived from *l*CT from one subject used for model building (top) and volume renditions of the object fuzzy models (bottom).

thoracic skin object) is first recognized in the test PET image I^m by using the optimal threshold interval value Th_1 of O_1 for PET. Note that the optimal threshold value for O_1 from dCT would not be appropriate for I^m. The rest of the procedure for recognizing O_1 is the same as described in Ref. 1.

After the root object O_1 is localized, other objects are recognized following the hierarchy by using the procedure *MMR-OBJECT*. Assume that for some object O_k, its parent O_l has already been recognized. The parent-to-offspring relationship ρ_k (coming from dCT) is used to find an initial pose for the model $\mathrm{FM}(O_k)$ of O_k. This pose is then refined by searching in the pose space in a region around the initial pose. This region is determined from knowledge of ρ_k, its variation, and the scale factor range λ_k (all of this information coming from dCT). Suppose $\mathrm{FM}^p(O_k)$ is expressed as an image and denotes the fuzzy model of O_k at pose p. Let J denote the binary image resulting from thresholding the PET image I at the optimal threshold Th_k for PET for object O_k. Then, the optimum pose p^* is found by

$$p^* \in \arg\min_p (|\mathrm{FM}^p(O_k) - J| + |J - \mathrm{FM}^p(O_k)|).$$

TABLE IV. Location and size error (mean and SD) for recognition on PET/CT images Thx and Abd.

	Strategy	TSkn	PS	RPS	LPS	PC	ASkn	Lvr	Kd	RKd	LKd	Spl	Mean
Location error (mm)	*B-x-s-d*CT-*m-l*CT	4.1	5.9	7.1	5.5	8.8	3.2	8.9	14.7	16.1	17.8	19.2	10.1
		0.9	2.4	2.0	1.5	4.8	1.7	3.6	10.2	9.3	12.0	18.7	6.1
	*B-x-s-d*CT-*m*-PET	4.7	14.2	21.7	19.0	12.7	2.9	11.2	12.7	20.5	16.0	22.0	14.3
		2.0	4.6	5.6	6.1	5.9	1.4	5.5	11.8	17.1	15.3	21.3	8.8
	*B-x-s-d*CT-*m*-PET&CT	2.8	6.4	6.0	6.3	16.2	2.4	11.9	13.3	21.0	17.3	21.0	11.3
		1.2	2.7	3.9	3.1	14.8	1.4	5.3	12.2	17.5	14.3	20.1	8.8
Size error	*B-x-s-d*CT-*m-l*CT	1.00	0.98	1.00	1.00	1.03	1.01	1.09	1.04	1.20	1.19	1.14	1.06
		0.01	0.02	0.03	0.03	0.06	0.01	0.08	0.05	0.15	0.12	0.24	0.07
	*B-x-s-d*CT-*m*-PET	1.00	1.10	1.27	1.28	1.03	1.00	1.02	1.00	1.06	1.03	1.03	1.07
		0.02	0.05	0.09	0.09	0.06	0.00	0.08	0.10	0.16	0.14	0.24	0.09
	*B-x-s-d*CT-*m*-PET&CT	1.00	0.99	0.98	0.97	1.05	1.00	1.03	1.00	1.07	1.06	0.99	1.01
		0.01	0.02	0.03	0.04	0.07	0.00	0.07	0.10	0.15	0.13	0.24	0.08

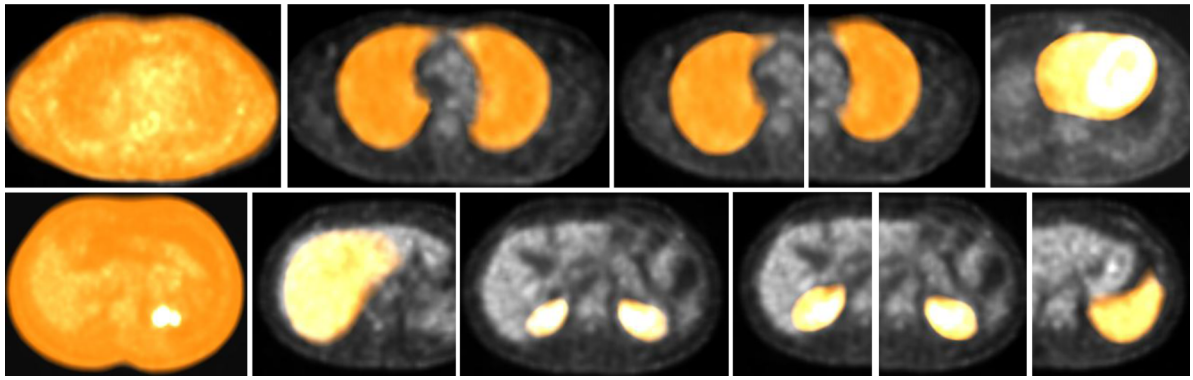

FIG. 5. Sample recognition results for thorax and abdomen in PET images for the strategy where models were built body region-wise from *d*CT and deployed on PET images. The model slices are shown overlaid in color on the test image slices. Left to right: TSkn, PS, RPS, LPS, PC, ASkn, Lvr, Kd, RKd, LKd, and Spl.

Here, $|x|$ denotes the fuzzy cardinality of x and image subtraction is done in the sense of fuzzy logic. Thus, in our example, the geographic and geometric prior information for the objects come from *d*CT and the modality-specific information comes from PET. As we will demonstrate in Sec. 3, because of this synergistic use of very detailed prior information, good recognition results can be achieved even in PET images where there is little anatomic tissue detail.

As mentioned earlier, the modality m of images employed in this paper comes from the set {*l*CT, PET, PET&CT, *tc*CT, *tr*CT, *te*CT, *tb*CT, *tc*PET, *tr*PET, *te*PET, *tb*PET, *tm*CT, *tm*PET}. Among its members, the first two constitute acquired original images. Others are derived from the acquired images as described below.

2.B.1. Combined modality PET&CT

The combined image I^m is obtained by first scaling the intensities of *l*CT and PET images to the same scale and then by weighted addition with equal weights.

2.B.2. Texture images for CT and PET

Image texture properties within an object region represent the interior structure of the object and are used extensively in image processing and analysis, particularly segmentation.[39] The use of texture in the literature is mostly for detecting pathologies and characterizing and classifying them[17,40–43] in CT and PET images and not for anatomy recognition *per se*. The potential of texture inspired us for use in the AAR approach for the possibility of discriminatively enhancing each object by an appropriate texture property expressed at each voxel with the hope of improving recognition performance over the original acquired image (*l*CT or PET). A great variety of texture description methods have been developed. Since visual conspicuity is a clue for the effectiveness of recognition, based on our initial visual observation of their ability to enhance object regions, we decided to examine four texture properties in this paper—correlation, contrast, and energy, all derived from gray level cooccurrence (GLC) matrix pertaining to the original image, and ball-scale or b-scale, derived directly from the original image.[44] Although initial selection is based on visual conspicuity, the best among the properties is selected based on their recognition performance as described under Results.

Given an image I^m, $m \in \{l\text{CT}, \text{PET}\}$, our goal is to derive a new image I^x, for each $x \in \{tc\text{CT}, tr\text{CT}, te\text{CT}, tb\text{CT}, tc\text{PET}, tr\text{PET}, te\text{PET}, tb\text{PET}\}$, where a texture property is estimated at each voxel v of I^m by considering a window of size $w \times w$ positioned at v in the 2D plane of the natural slice[49] of I^m. The GLC idea involves determining how frequently each possible pair (a, b) of intensities occurs within this window by considering all possible pairs of voxels which are separated by a distance d at an angle ϕ. From the GLC matrix, $A(a,b)$ thus determined at v, texture contrast, correlation, and energy values are estimated at v from $A(a,b)$ by using the following expressions shown for *l*CT. Expressions for PET are similar. Here, (μ_1, σ_1) and (μ_2, σ_2) denote mean and standard deviation of the variables a and b, respectively [in our case, $(\mu_1, \sigma_1) = (\mu_2, \sigma_2)$]. In our implementation, we fixed the variables at $w = 15$, $d = 1$, and $\varphi = 0°$, 45°, 90°, and 135°,

TABLE V. Location and size error (mean and SD) for recognition on PET/CT images in body torso.

	BTSkn	TO	PS	RPS	LPS	PC	AO	Lvr	Kd	RKd	LKd	Spl	PO	PVAT	PSAT	Bld	Mean
Location error (mm)	4.4	6.8	5.8	7.2	6.1	12.7	11.5	15.1	11.4	8.3	12.9	22.3	12.7	15.3	13.4	9.9	11.0
	2.4	5.5	4.6	4.4	3.2	8.9	4.5	7.5	10.4	9.3	15.2	24.2	7.6	8.0	7.2	5.7	8.0
Size error	1.0	1.0	1.0	1.0	1.0	1.1	1.1	1.1	1.0	1.1	1.1	1.0	1.0	1.0	1.0	1.1	1.0
	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.0	0.1	0.0	0.1	0.1

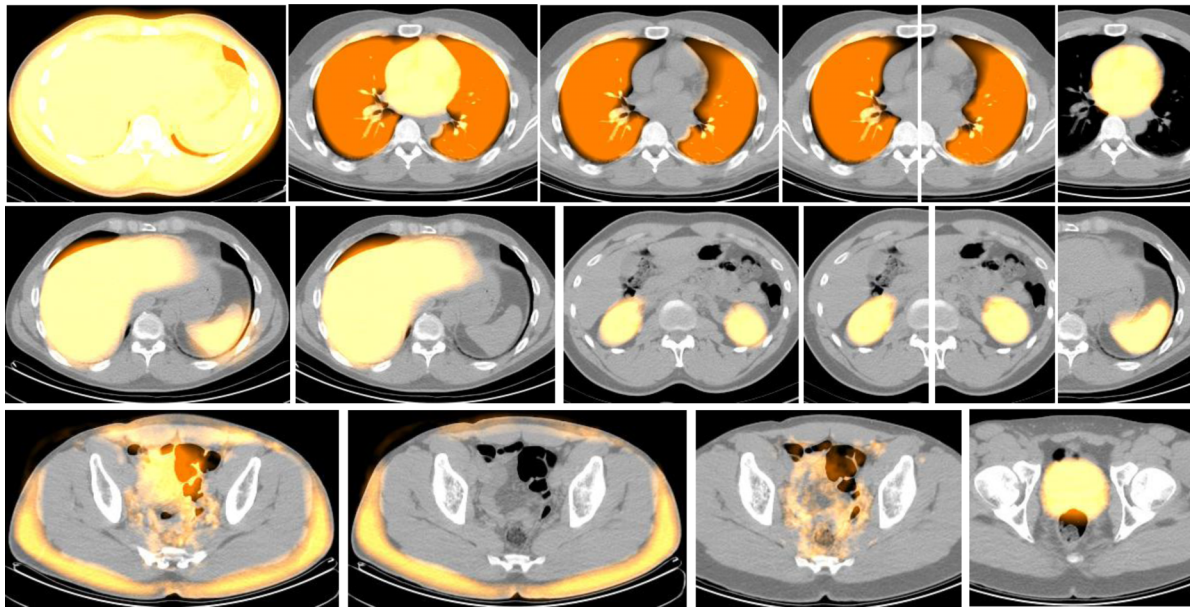

FIG. 6. Sample recognition results for thorax, abdomen, and pelvis in *l*CT images for the strategy where models were built for full body torso from *l*CT and deployed on *l*CT images. Model slices are shown overlaid in color on test image slices. Left to right: BTSkn, TO, PS, RPS, LPS, PC, AO, Lvr, Kd, RKd, LKd, Spl, PO, PSAT, PVAT, and Bld.

$$I^{tc\mathrm{CT}}(v) = \sum_{a,b} |a-b|^2 A(a,b),$$

$$I^{tr\mathrm{CT}}(v) = \frac{\sum_{a,b} abA(a,b) - \mu_1\mu_2}{\sigma_1\sigma_2},$$

$$I^{te\mathrm{CT}}(v) = \sum_{a,b} A^2(a,b).$$

Ball scale or b-scale[44] is a concept of using the largest homogeneous ball centered at each voxel in any given image, instead of the voxel itself, as the basic unit in image processing and analysis. Its advantages in image filtering, segmentation, interpolation, registration, and object recognition have been previously demonstrated in the literature. We compute a b-scale image I^x, $x \in \{tb\mathrm{CT}, tb\mathrm{PET}\}$, corresponding to a given *l*CT or PET image I^m following the methods of Ref. 44, where each voxel v in I^x is assigned the radius of the largest homogeneous ball centered at v. For reasons mentioned above, we use a 2D disc, instead of a ball, in the plane of the natural slice of I^m for defining the b-scale value at v. There is one scale parameter that defines "homogeneity," which we estimate automatically from I^m following the method described in Ref. 44.

2.B.3. Texture-based membership images tmCT and tmPET

In this modality, the value of a voxel v of I^x for $x = tm\mathrm{CT}$ is determined as an indicator of "objectness" at v from the values assigned to v in its original acquired image I^{m1}, $m1 = l\mathrm{CT}$, and texture energy image I^{m2}, $m2 = te\mathrm{CT}$. The 2D histogram $h_{m1,m2}(a,b)$ of *l*CT and texture energy of *l*CT is taken to be the objectness function. The derived modality I^x for $x = tm\mathrm{PET}$ is similarly defined from the respective PET images, $m1 = PET$ and $m2 = tc\mathrm{PET}$. For enhancing object information in the membership image, texture energy was found to be better for *l*CT than other texture properties. Similarly, for PET, texture contrast was found to be preferable. The objectness

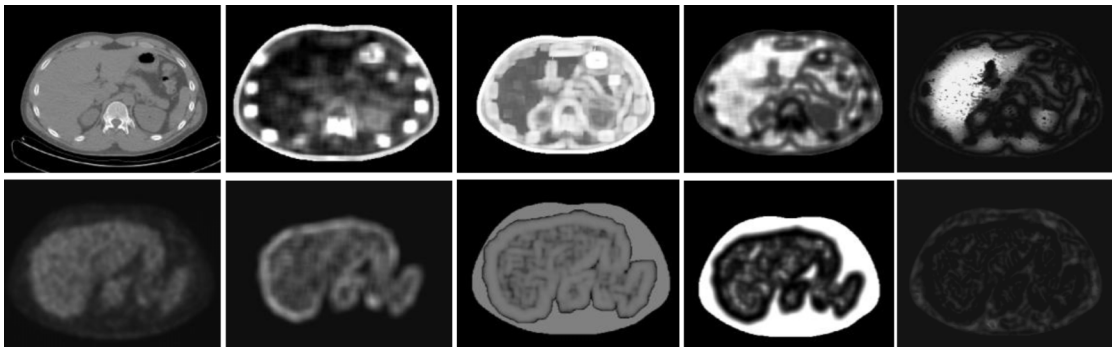

FIG. 7. Texture images derived from one sample *l*CT and PET image slice. Top row: *l*CT, *tc*CT, *tr*CT, *te*CT, and *tb*CT. Bottom row: PET, *tc*PET, *tr*PET, *te*PET, and *tb*PET.

TABLE VI. Location and size error (mean and SD) over all objects in Thx and Abd for recognition on texture images.

	*tc*CT	*tr*CT	*te*CT	*tb*CT	*tc*PET	*tr*PET	*te*PET	*tb*PET
Location error (mm)	11.5	14.3	10.3	13.7	12.7	13.0	13.0	14.5
	7.0	7.3	5.5	9.8	7.1	6.8	7.8	6.5
Size error	1.05	1.04	1.07	1.04	1.02	1.03	1.03	1.10
	0.09	0.08	0.07	0.08	0.09	0.08	0.08	0.08

function is estimated as a cumulative histogram over the training images. Given a test image I^{m1}, then its texture (energy or contrast) image I^{m2} is first computed, and subsequently, by using the objectness function estimated at the model building stage, the objectness or texture membership image is computed from I^{m1} and I^{m2}. Since this objectness function is stored in FAM(B, G), in Fig. 2 there is an arrow connecting the "Model Building" block to the "Membership Description" block on the right corresponding to the test images.

3. EXPERIMENTAL METHODS AND RESULTS

3.A. Image data sets

This retrospective study was conducted following approval from the Institutional Review Board at the Hospital of the University of Pennsylvania along with a Health Insurance Portability and Accountability Act (HIPAA) waiver. All image data sets are selected from the existing patient image database of our health system. We employ 16 whole-body PET/CT images for testing our methods and 25 *d*CT data sets employed in Ref. 1 for each of two body regions—thorax and abdomen—to build the respective fuzzy anatomy models for BT, Thx, and Abd. Table III lists the image details. All whole-body FDG-PET/CT image data sets utilized in the current study had previously been acquired as part of a separate prospective research study (PI co-author DAT) that enrolled asymptomatic healthy normal male volunteers who provided written consent for use of their scans for other future research studies. For all 16 *l*CT and PET image data sets employed in the current study, the images appeared radiologically normal with exception of minimal incidental focal abnormalities. All of these scans had been obtained on a 16 multidetector row LYSO whole-body PET/CT scanner with time-of-flight capabilities (Gemini TF, Philips Medical Systems, Bothell, WA). 3D PET data had been acquired from the skull vertex to the toes ~60 min after intravenous administration of ~555 MBq of FDG for 3 min per bed position. Image reconstruction had been performed at 4 mm nominal slice thickness in the axial plane using a list-mode maximum-likelihood expectation–maximization (ML-EM) algorithm with 33 ordered subsets and 3 iterations, and the system model included time-of-flight as well as normalization, attenuation, randoms, and scatter corrections, where rescaled low-dose CT images were utilized for attenuation correction of PET images. The 25 *d*CT image data sets of the thorax and the 25 *d*CT image data sets of the abdomen had previously been acquired in patients for various clinical indications on 16 or 64 multidetector row CT scanners (Siemens Medical Solutions, Malvern, PA) during full inspiratory breath-hold during the venous phase of enhancement following intravenous administration of iodinated contrast material (Iopamidol (Isovue-370), Bracco Diagnostics, Monroe Township, NJ). A kVp of 120, an average effective tube current-time product of 150–200 mAs, with tube current modulation on, and a gantry rotation time of 0.5 s had been utilized during image acquisition. Images were reconstructed at a nominal slice thickness of 5 mm with an interval of 5 mm in the axial plane, a 512×512 matrix, and a B30f reconstruction kernel. For all *d*CT image data sets employed in the current study, the images appeared radiologically normal with exception of minimal incidental focal abnormalities. All 18 objects were delineated in $25 + 25 + 16$ 3D image sets following strictly their definition using manual painting/tracing/editing, iterative live wire, live wire, and thresholding methods.

3.B. Evaluation strategies

As depicted in Table I, we have conducted 28 recognition experiments by utilizing the data sets listed in Table III to perform the following comparisons: for *l*CT and PET between using models from *d*CT versus models from *l*CT, between using different texture properties versus original acquired images, and between body-region-wise approach versus whole-body torso approach. For all strategies denoted *B*-*x*-*s*-*d*CT-*m*-*y*, all 25 *d*CT data sets were used for model

TABLE VII. Location and size error (mean and SD) for recognition on *te*CT and *tc*PET images in Thx and Abd.

	Strategy	TSkn	PS	RPS	LPS	PC	ASkn	Lvr	Kd	RKd	LKd	Spl	Mean
Location error (mm)	*B*-*x*-*s*-*d*CT-*m*-*te*CT	4.3	12.3	16.7	16.1	6.9	2.3	12.4	6.8	7.1	7.0	21.4	10.3
		1.5	4.8	5.5	6.9	4.4	1.5	7.1	4.0	7.4	3.5	13.8	5.5
	B-*x*-*s*-*d*CT-*m*-*tc*PET	3.3	15.9	18.8	19.7	21.5	3.3	13.1	6.3	6.8	5.9	24.9	12.7
		1.3	6.4	9.5	9.2	13.5	1.6	4.8	2.5	3.8	3.3	21.7	7.1
Size error	*B*-*x*-*s*-*d*CT-*m*-*te*CT	0.98	1.09	1.26	1.27	1.00	1.00	1.01	0.98	1.00	1.02	1.18	1.07
		0.01	0.05	0.09	0.09	0.06	0.00	0.08	0.05	0.10	0.10	0.19	0.07
	B-*x*-*s*-*d*CT-*m*-*tc*PET	1.01	1.00	1.04	0.97	1.05	1.00	1.12	0.98	1.00	1.00	1.04	1.02
		0.01	0.05	0.10	0.11	0.07	0.00	0.09	0.06	0.10	0.10	0.26	0.09

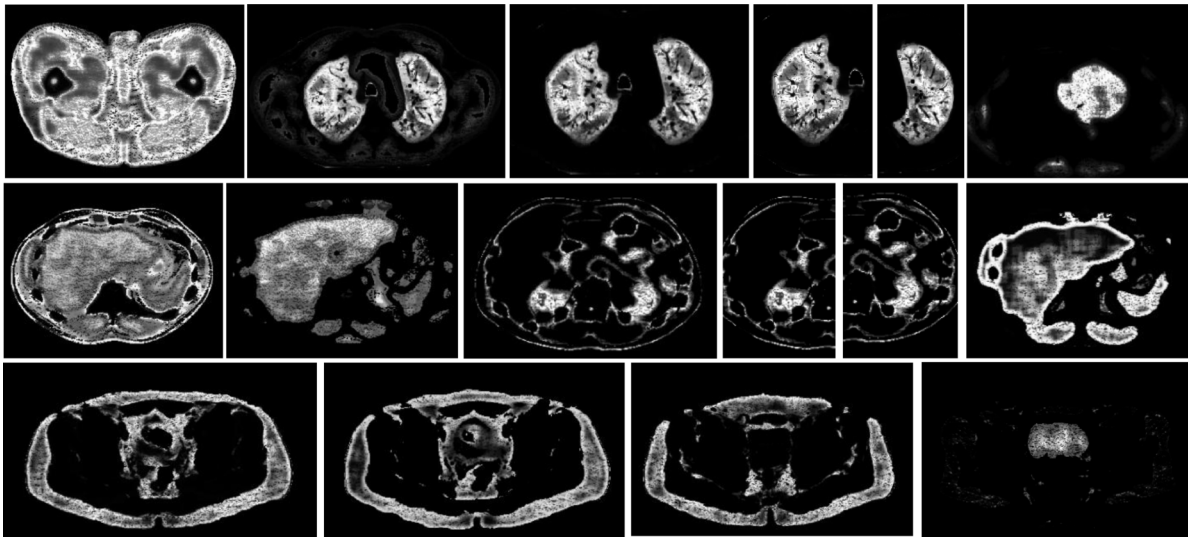

FIG. 8. Texture-based membership images derived from *l*CT and *te*CT images for different objects. L to R: BTSkn, TO, PS, RPS, LPS, PC, AO, Lvr, Kd, RKd, Lkd, Spl, PO, PVAT, PSAT, and Bld.

building. Optimal threshold intervals were estimated from 6 of 16 PET/CT data sets, testing was done on the remaining 10 data sets, and the optimal-threshold-training-testing was repeated 3 times on different partitioning of the 16 data sets. For *B*-*x*-*s*-*l*CT-*m*-*l*CT, $x \in \{\text{Thx}, \text{Abd}, \text{Plv}\}$, the whole-body images were divided into body region images for Thx, Abd, and Plv, models were built from 8 of the 16 *l*CT data sets, testing was done on the remaining 8 data sets, and the model-building-testing process was repeated 3 times on different partitioning of the 16 data sets. For optimal threshold estimation, the eight training data sets were used. For *B*-BT-*s*-*l*CT-*m*-*l*CT and *B*-BT-*s*-*l*CT-*m*-*tm*CT also, the model-building-testing process was similar except that the whole-body images were trimmed as per body torso definition although no body region subdivision was needed.

We describe the accuracy of various recognition strategies in terms of *location error* and *scale* (or *size*) error. Location error is expressed as the distance between the centers of the object model output by AAR-MMR and the true object. Scale error is expressed as a ratio of the estimated size of the output object model and the known true object size. The ideal values for the two error metrics are thus 0 and 1, respectively. In Secs. 3.C–3.E, we will present results for recognition on original PET/CT, texture, and texture-based membership images, respectively.

3.C. Recognition in PET/CT images

In this section, we will examine results on images I^x, for $x \in \{l\text{CT}, \text{PET}, \text{PET\&CT}\}$. Some exemplary combinations of objects are displayed in Fig. 4 from the body torso, both surface renditions of sample objects used for model building (top) and volume renditions of the object fuzzy models that were created (bottom). The SAT object is partially removed to facilitate visualization. The recognition strategies compared are *B*-*x*-*s*-*d*CT-*m*-*l*CT, *B*-*x*-*s*-*d*CT-*m*-PET, and *B*-*x*-*s*-*d*CT-*m*-PET&CT, for $x \in \{\text{Thx}, \text{Abd}\}$. The results are summarized in Table IV.

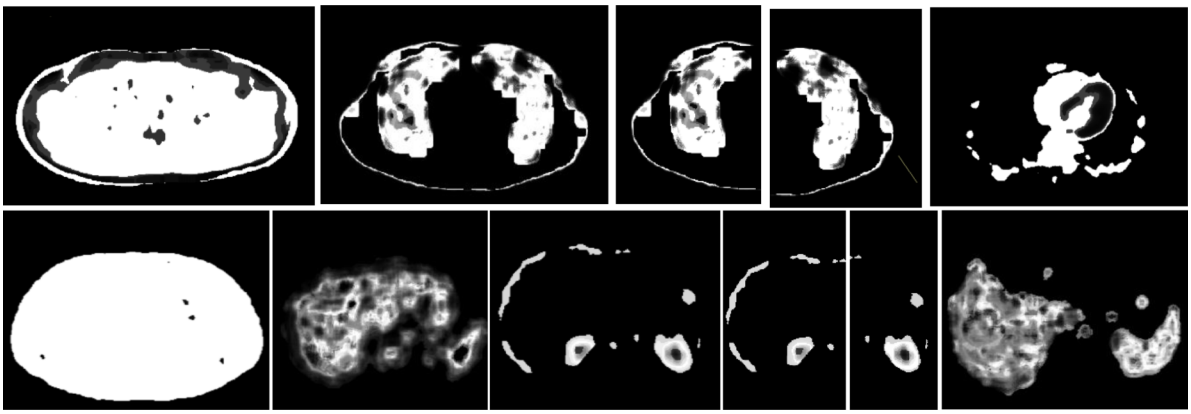

FIG. 9. Texture-based membership images derived from PET and *tc*PET images for different objects. L to R: TSkn, PS, RPS, LPS, PC, ASkn, Lvr, Kd, RKd, Lkd, and Spl.

TABLE VIII. Location and size error (mean and SD) for recognition on texture membership images in Thx and Abd.

	Strategy	TSkn	PS	RPS	LPS	PC	ASkn	Lvr	Kd	RKd	LKd	Spl	Mean
Location error (mm)	*B*-*x*-*s*-*d*CT-*m*-*tm*CT	4.3	6.0	7.4	4.9	6.1	3.9	7.3	13.5	4.3	9.4	8.5	6.9
		1.5	2.3	2.2	1.8	3.5	1.9	3.8	9.1	2.5	7.9	3.5	3.6
	B-*x*-*s*-*d*CT-*m*-*tm*PET	4.5	13.4	16.8	20.4	12.6	2.6	13.4	11.6	4.6	5.9	15.8	11.1
		1.6	5.4	6.1	8.0	6.2	0.9	6.1	8.8	2.9	5.6	12.9	5.9
Size error	*B*-*x*-*s*-*d*CT-*m*-*tm*CT	0.98	0.97	1.00	1.00	1.00	0.99	0.98	1.00	0.97	1.02	0.81	0.97
		0.01	0.03	0.03	0.03	0.06	0.02	0.05	0.03	0.07	0.11	0.10	0.05
	B-*x*-*s*-*d*CT-*m*-*tm*PET	1.00	0.99	1.02	1.01	1.03	1.01	1.04	0.96	0.91	0.90	1.08	1.00
		0.01	0.07	0.14	0.15	0.06	0.01	0.09	0.06	0.11	0.09	0.22	0.09

From Table IV, we note that the average position error for recognition in *l*CT for thoracic and abdominal objects is close to 10 mm (<3 voxels) and size error is always close to 1 when models are created from body region *d*CT. Surprisingly, with models from *d*CT, recognition on PET with ∼3 voxels of localization error becomes feasible even with the poor but faint hint of anatomy seen in PET images. This is mainly because of the rich anatomic knowledge encoded in FAM(*B*, *G*). For the lungs, the location error in PET is a bit high because of the very low radiotracer uptake in these organs. AAR on combined PET&CT images has better performance compared to PET images. Among the strategies illustrated in Table IV, *l*CT yields the best performance, which is interestingly quite comparable to the recognition results on *d*CT images demonstrated in Ref. 1 for the same objects in thorax and abdomen. It should be emphasized that the boundaries of many objects including PC, Lvr, RKd, LKd, Spl, and Bld are quite fuzzy in *l*CT images compared to those in *d*CT. We display in Fig. 5 some sample recognition results for PET by overlaying cross sections of the models at recognition over PET image slices for the strategy *B*-*x*-*s*-*d*CT-*m*-PET.

Recognition results for *l*CT using the whole-body torso strategy *B*-BT-*s*-*l*CT-*m*-*l*CT (models from *l*CT, recognition on *l*CT, both over full BT) are summarized in Table V. Figure 6 shows a sample recognition result for each object in BT with the model slices overlaid on the test image slices. Overall, the mean position error is 11.0 mm (∼3 voxels) and size error is close to 1. Generally speaking, AAR shows consistent and good recognition performance on whole-body torso *l*CT images.

3.D. Recognition on texture images

Figure 7 shows one sample texture image corresponding to *l*CT and PET for each of the four texture properties considered in this paper. Instead of presenting all results for all objects for all texture properties, we first summarize the mean results over all objects in Table VI and then present the detailed results for the best among these properties. Table VI lists mean location and size errors over all objects for body-region-wise recognition approaches for the four texture description methods for both *l*CT and PET. Since, by far, texture energy for *l*CT and texture contrast for PET outperformed other attributes, we list their detailed results in Table VII.

Compared to *l*CT, its texture modality *te*CT shows some improvement for abdominal objects but not thoracic objects. For *tc*PET, the improvement for abdominal objects over PET is more significant. For the same reason explained earlier for PET, thoracic objects do not show any advantages in considering texture modality for improving recognition.

3.E. Recognition on texture-based membership images

Some sample texture membership images for different objects derived from *l*CT and PET modalities are displayed in Figs. 8 and 9, respectively. In Table VIII, we present body-region-wise recognition results for texture membership images for *l*CT and PET. In Table IX, whole-body torso results for texture membership images derived from *l*CT are presented.

Results in texture membership modalities *tm*CT and *tm*PET are overall significantly better than those in original images *l*CT and PET. For the kidneys in particular, the improvement is substantial, even in PET images. The strategy of body-region-wise modeling from *d*CT and recognition in texture membership images produces the best results among all strategies tested, bringing object localization error within 2 voxels. If we exclude spleen, the location error for the whole-body torso strategy on *tm*CT images also approaches 2 voxels. We

TABLE IX. Location and size error (mean and SD) for recognition on texture membership images derived from *l*CT in body torso.

	BTSkn	TO	PS	RPS	LPS	PC	AO	Lvr	Kd	RKd	LKd	Spl	PO	PVAT	PSAT	Bld	Mean
Location error (mm)	4.2	7.7	6.1	7.6	6.4	7.5	9.1	12.6	8.4	3.7	5.8	24.2	13.6	12.2	13.6	14.0	9.8
	2.5	5.2	4.1	4.0	2.7	3.6	4.0	7.4	5.9	2.6	5.1	26.5	7.9	7.5	7.4	16.1	7.0
Size error	1.00	0.97	0.98	1.00	0.99	1.00	0.96	1.01	0.98	0.99	0.98	0.80	0.98	0.93	0.98	0.86	0.96
	0.01	0.02	0.02	0.04	0.03	0.04	0.07	0.05	0.05	0.06	0.09	0.13	0.02	0.07	0.03	0.11	0.05

TABLE X. Comparison of body-region-wise and whole-body torso approaches.

Strategy	Location error in mm (mean, SD)	Size error (mean, SD)
Body-region-wise in *l*CT	11.5	1.07
	7.2	0.09
Whole-body torso in *l*CT	11.3	1.04
	9.7	0.08
Body-region-wise in *te*CT	11.8	1.09
	6.4	0.09
Body-region-wise in *tm*CT	7.5	0.97
	4.1	0.06
Whole-body torso in *tm*CT	9.1	0.97
	6.9	0.06
Body-region-wise in PET	16.7	1.09
	10.4	0.11
Body-region-wise in *tc*PET	14.8	1.02
	8.3	0.1
Body-region-wise in *tm*PET	12.7	0.99
	6.9	0.11

note that the number of training samples for all strategies where models are built from *l*CT images is much smaller (8) compared to number (25) in strategies where models are built from *d*CT. The whole-body strategies may therefore approach or even surpass body-region-wise approaches with a sufficient number of samples.

To explore this dichotomy between body-region-wise and body-wide approaches further, we list in Table X location and size errors over all objects that were common to both strategies. Note that pelvic objects did not participate in body-region-wise strategies since we did not have *d*CT derived models for them in our previous study.[1] The involved objects are PS, RPS, LPS, PC, Lvr, Kd, RKd, LKd, and Spl. Comparison is made among the top contenders only. While comparing based on location error, *t*-tests showed the differences to be statistically significant between the following pairs: BT in *l*CT vs body-region-wise in PET ($P = 0.03$); body-region-wise in *tm*CT vs body-region-wise in PET ($P = 0.002$); body-region-wise in *tm*CT vs body-region-wise in *tc*PET ($P = 0.03$); BT in *tm*CT vs body-region-wise in PET ($P = 0.008$); body-region-wise in *tm*CT vs body-region-wise in *tm*PET ($P = 0.04$).

TABLE XI. Computational times in seconds for the key steps.

Operation	Body torso	Thorax	Abdomen
Optimal threshold estimation	1.3	0.7	1.1
Model building	1.2	1.2	0.6
Texture image	51.3	21.7	18.4
Texture membership image	1.1	0.3	0.4
Object recognition	32.2	5.9	24.3

Body-region-wise and whole-body torso recognition on *tm*CT images seem to be the best strategies overall. Since there is no statistically significant difference between their performances, the whole-body approach is to be preferred since it does not call for an extra step of body region subdivision.

3.F. Computational considerations

The computational times are estimated on a Dell computer with the following specifications: 4-core Intel Xeon 3.3 GHz base to 3.7 GHz max turbo CPU with 8 GB RAM and running the GNU/Linux 3.11.10-25-desktop operating system. Mean computational times for the AAR steps are listed in Table XI. Note that the texture image is computed only once for a given test image irrespective of the number of objects to recognize but the texture membership image is computed separately for each object. Accordingly, the time reported in the table for the membership image is per object per subject, as are the times listed for OTE and model building. For the best strategies involving *tm*CT images, thus, the total times for locating all considered objects in one image for body torso, thorax, and abdomen are 584.1, 52.7, and 166.6 s, and per object recognition times are roughly 36.5, 10.5, and 28.8 s, respectively.

In Table XII, we compare the recognition performance of the previous optimal threshold method to OTE for two top strategies found in this paper, namely, *B-x-s-d*CT-*m-l*CT (Table IV) and *B-x-s-d*CT-*m-tm*CT (Table VIII). From these tables, we conclude that, while the performance on *l*CT images is similar for the two methods, the previous restricted search method fails on texture images to find optimal thresholds that are best suited for the different objects. If we expand the search range of the previous method, its computational cost would increase rapidly because of the 5D search space.

TABLE XII. Recognition results (mean and SD) on *l*CT and *tm*CT images by using the optimal threshold method of Ref. 1.

	Strategy	TSkn	PS	RPS	LPS	PC	ASkn	Lvr	Kd	RKd	LKd	Spl	Mean
Location error (mm)	*B-x-s-d*CT-*m-l*CT	4.1	6.3	7.7	5.9	7.9	2.9	9.2	14.3	9.8	14.0	25.1	9.7
		1.0	2.6	3.3	2.9	4.6	1.5	4.0	10.2	6.8	11.8	21.9	6.4
	*B-x-s-d*CT-*m-tm*CT	4.0	9.9	15.6	7.0	37.7	7.5	34.2	46.0	50.7	53.6	60.6	29.7
		1.0	7.2	12.0	4.2	7.1	2.5	9.4	10.3	10.5	9.3	12.9	7.9
Size error	*B-x-s-d*CT-*m-l*CT	1.00	0.97	0.98	0.98	1.02	1.01	1.09	1.04	1.15	1.15	1.19	1.05
		0.01	0.03	0.03	0.04	0.06	0.01	0.07	0.05	0.16	0.14	0.23	0.07
	*B-x-s-d*CT-*m-tm*CT	1.00	1.02	1.10	1.03	1.01	1.02	1.17	1.10	1.36	1.33	1.22	1.12
		0.01	0.07	0.14	0.07	0.06	0.00	0.10	0.07	0.10	0.08	0.19	0.08

4. DISCUSSION AND CONCLUDING REMARKS

4.A. Summary of advances

Considering the fact that there are numerous clinical and research applications of whole-body PET/CT, our goal in this work was to test how feasible it is to perform object localization on low-dose CT (*l*CT) of PET/CT acquisitions and also on PET images. Building on the previous AAR approach,[1] we set out to determine how AAR can be further advanced toward this goal. In the process, we accomplished the following *four key advances*.

(1) Given the low spatial and contrast resolution of *l*CT and the lack of definitive anatomic details in PET images, we sought to determine if image texture would facilitate object recognition. Although texture itself helped somewhat, the best strategy turned out to be the use of what we called a membership (or fuzzy) image resulting from the joint use of voxel intensity and its texture property, leading to mean object localization within 1–2 voxels. Amazingly, even in PET with only a faint hint of anatomy, AAR's richly coded anatomic prior information was able to make up for the lack of detail in the image itself. Object texture jointly with its intensity is definitely useful in body-wide object localization.

(2) We demonstrated the use of fuzzy models created in our previous work[1] from diagnostic CT (even from different patient groups) in recognizing objects in *l*CT, PET, and the derived texture membership images, suggesting the potential for rapid prototyping in body-wide PET/CT applications. In this connection, we also arrived at a new optimal threshold retraining technique, which, compared to our previous algorithm,[1] is applicable to any (original or derived) image without the need for approximate knowledge of the threshold needed to highlight each object, improves recognition, and is computationally efficient.

(3) We extended the previous body-region-wise AAR to whole-body torso approach which required consideration of object relationships across body regions. As seen from Table X, the whole-body torso approach achieves localization accuracies comparable to body-region-wise AAR even with a much smaller (1/3) training set, with the advantage that the former obviates the need for body region subdivision required in the latter method, affording greater automation and direct body-wide application.

(4) In Table XIII, we catalog the best localization strategy found for each of the 18 objects studied. Interestingly, the best strategy varies among objects. It is plausible to make such a catalog part of the model itself so the best strategy can be called upon when recognizing each object. For illustration, consider kidneys. The table suggests that, to localize kidneys with the best accuracy, the composite object Kd (the union of LKd and RKd as a single object) should be first recognized in the PET texture contrast image. Subsequently, the component objects LKd and RKd should be localized on the texture membership images derived from *l*CT and *te*CT. Such object-specific strategy implementation is certainly feasible within the AAR framework as the fuzzy anatomy model FAM(B, G) is currently set up.

4.B. Comparison with other methods

Most segmentation methods for PET/CT images described in the literature[18–24] have confined to mostly one object in one body region and did not consider a whole body region or the whole-body torso. Their goal was different, namely, delineation of a particular object of interest but not object localization. A method of multiorgan detection based on information theory was proposed in Ref. 30. The central idea was to schedule tasks in an order so that each operation achieves maximum expected information gain. The validation was carried out on multiorgan detection in whole-body *d*CT images and liver segmentation in PET/CT images. The average localization error for kidney was reported as 8.97 mm s and maximum error as 19 mm s on *d*CT images. Liver was represented by a deformable model which was defined with seven anatomical landmarks on the liver including liver center. Landmarks were detected by using the feature selection and learning-based approach of Ref. 45. The average localization error for the liver center was reported as 15 mm s on PET/CT scans.[30] Object localization on *d*CT images with and without contrast via rectangular boxes was reported in Ref. 29, where the average localization error was 17 mm s with maximum being 34 mm s. The localization accuracies of the proposed approach (even on PET images) are

TABLE XIII. Object-by-object best recognition strategy. See Table I for strategy nomenclature.

Object	Strategy	Object	Strategy
BTSkn	*B*-BT-*s*-*l*CT-*m*-*tm*CT	Lvr	*B*-Abd-*s*-*d*CT-*m*-*tm*CT
TSkn	*B*-Thx-*s*-*d*CT-*m*-*tr*CT	Kd	*B*-Abd-*s*-*d*CT-*m*-*tc*PET
TO	*B*-BT-*s*-*l*CT-*m*-*l*CT	LKd	*B*-BT-*s*-*l*CT-*m*-*tm*CT
PS	*B*-BT-*s*-*l*CT-*m*-*l*CT	RKd	*B*-BT-*s*-*l*CT-*m*-*tm*CT
LPS	*B*-Thx-*s*-*d*CT-*m*-*tm*CT	Spl	*B*-Abd-*s*-*d*CT-*m*-*tm*CT
RPS	*B*-Thx-*s*-*d*CT-*m*-PET&CT	PO	*B*-BT-*s*-*l*CT-*m*-*l*CT
PC	*B*-Thx-*s*-*d*CT-*m*-*tm*CT	Bld	*B*-BT-*s*-*l*CT-*m*-*l*CT
ASkn	*B*-Abd-*s*-*d*CT-*m*-*te*CT	PSAT	*B*-BT-*s*-*l*CT-*m*-*l*CT
AO	*B*-BT-*s*-*l*CT-*m*-*tm*CT	PVAT	*B*-BT-*s*-*l*CT-*m*-*tm*CT

comparable to or better than the above results in the literature. An additional advantage of the AAR approach is that it actually fits a model to the object to be found which is of the same form as the object rather than a rectangular box. We note that although many different PET image reconstruction methods are available now, major scanner vendors all offer very similar image quality. No special reconstruction method was used for the images utilized in the paper and they all constitute routine clinical scans.

4.C. Future opportunities

There are several further avenues we are considering for the body-wide PET/CT AAR approach. In this paper, we studied a few simple texture properties. Other advanced descriptors particularly fractals, Gabor filters, Minkowski texture set, *g*-scale, and tensor scale may offer representations that are more specific to each object.

The main motivations for focusing on object recognition in the AAR methodology are twofold. First, effective object localization is essential for successful object delineation. If a model (whatever form it takes such as an atlas, shape model, fuzzy model, or any other embodiment of prior information) cannot be brought to fit closely with the object in the image, delineation will likely fail. Second, disease quantification can potentially be done without explicit delineation of objects but from the result after object recognition. We are exploring this latter idea in body-wide AAR applications in different systemic disease conditions.

Table XIII suggests that it may be advantageous to tailor object recognition methods separately for each object based on the strategy that is best suited for the object. This requires a mode of thinking that is different from a straightjacket approach where a uniform method is applied to all objects. Object-specific strategies allow for highly nonlinear phenomena to be handled effectively. It is feasible to implement the strategies implied by Table XIII within the current AAR approach.

As demonstrated by the texture membership image idea, although we focused on FDG as the radiotracer for PET in this paper, the proposed approach is general and applicable to PET images obtained with other radiotracers including 18F-fluoride, 18F-fluorothymidine, 18F-fluorocholine, 18F-fluoroestradiol, and 68Ga-DOTATOC amongst many others. In the same vein, extension of these approaches to PET/MRI offers exciting possibilities and applications.

In body-wide applications, when dealing with, say, of the order of 30 objects, the proposed method would require about 20 min of computational time per patient study. For applications which require interactive speed of response, the current speed of AAR is certainly not acceptable. However, for applications where offline computing fits the clinical workflow, the time is certainly not unacceptable. Upgrading the computing platform itself may improve the speed by a factor of two. So, optimized and more efficient implementation of the AAR methodology is a worthwhile future undertaking.

Intrascanner and interscanner variations, especially for PET images, can affect object intensity distributions as well as some texture properties. This phenomenon is not addressed in this paper. Perhaps accuracy can be improved if this issue is handled in a manner akin to MRI intensity standardization in the future.

4.D. Limitations

In this paper, we have not considered many objects that may be important from the viewpoint of body-wide applications, especially nonblob-like less space-filling objects, which are very difficult to segment (delineate) even in higher quality *d*CT images. Our rationale was to investigate if it is feasible to localize better defined objects in these low-quality images first and then proceed to more difficult sparse objects. We have recently developed sparse object AAR techniques to recognize and delineate such objects in *d*CT images.[46] We are studying ways to adapt these techniques to PET/CT images.

Another limitation of this work is the small set of PET/CT images considered. Conversely, considering the small number of data sets used for model building and approximately the same number of independent data sets used for testing (this should be compared against the leave-one-out method that is commonly used in which the model is over fitted), the results seem to strongly validate the potential of the modified AAR approach for automatic PET/CT object localization. We also note that the use of data sets comparable in size to ours is quite common in the literature.[19,25,47]

ACKNOWLEDGMENTS

The training of Ms. Huiqian Wang in the Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, for the duration of one year was supported by China Scholarship Council.

[a]Author to whom correspondence should be addressed. Electronic mail: jay@mail.med.upenn.edu

[1]J. K. Udupa, D. Odhner, L. Zhao, Y. Tong, M. Matsumoto, K. C. Ciesielski, A. X. Falcao, P. Vaideeswaran, V. Ciesielski, and B. Saboury, "Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images," Med. Image Anal. **18**, 752–771 (2014).
[2]T. C. Kwee, D. A. Torigian, and A. Alavi, "Oncological applications of positron emission tomography for evaluation of the thorax," J. Thorac. Imaging **28**, 11–24 (2013).
[3]T. C. Kwee, D. A. Torigian, and A. Alavi, "Nononcological applications of positron emission tomography for evaluation of the thorax," J. Thorac. Imaging **28**, 25–39 (2013).
[4]D. Townsend, "Multimodality imaging of structure and function," Phys. Med. Biol. **53**, R1–R39 (2008).
[5]J.-F. Daisne, M. Sibomana, A. Bol, T. Doumont, M. Lonneux, and V. Grégoire, "Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: Influence of reconstruction algorithms," Radiother. Oncol. **69**, 247–250 (2003).
[6]C. Scarfone, W. C. Lavely, A. J. Cmelak, D. Delbeke, W. H. Martin, D. Billheimer, and D. E. Hallahan, "Prospective feasibility trial of radiotherapy target definition for head and neck cancer using 3-dimensional PET and CT imaging," J. Nucl. Med. **45**, 543–552 (2004).
[7]E. C. Ford, P. E. Kinahan, L. Hanlon, A. Alessio, J. Rajendran, D. L. Schwartz, and M. Phillips, "Tumor delineation using PET in head and neck cancers: Threshold contouring and lesion volumes," Med. Phys. **33**, 4280–4288 (2006).
[8]L. Drever, D. M. Robinson, A. McEwan, and W. Roa, "A local contrast based approach to threshold segmentation for PET target volume delineation," Med. Phys. **33**, 1583–1594 (2006).

[9]L. Drever, W. Roa, A. McEwan, and D. Robinson, "Iterative threshold segmentation for PET target volume delineation," Med. Phys. **34**, 1253–1265 (2007).
[10]X. Geets, J. A. Lee, A. Bol, M. Lonneux, and V. Grégoire, "A gradient-based method for segmenting FDG-PET images: Methodology and validation," Eur. J. Nucl. Med. Mol. Imaging **34**, 1427–1438 (2007).
[11]E. Day, J. Betler, D. Parda, B. Reitz, A. Kirichenko, S. Mohammadi, and M. Miften, "A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients," Med. Phys. **36**, 4349–4358 (2009).
[12]M. Hatt, C. Cheze le Rest, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET," IEEE Trans. Med. Imaging **28**, 881–893 (2009).
[13]M. Abdoli, R. A. J. O. Dierckx, and H. Zaidi, "Contourlet-based active contour model for PET image segmentation," Med. Phys. **40**, 082507 (12pp.) (2013).
[14]U. Bagci, J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, and D. J. Mollura, "Joint segmentation of anatomical and functional Images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images," Med. Image Anal. **17**, 929–945 (2013).
[15]K. C. Ciesielski, P. Miranda, A. Falcão, and J. K. Udupa, "Joint graph cut and relative fuzzy connectedness image segmentation algorithm," Med. Image Anal. **17**, 1046–1057 (2013).
[16]D. Han, J. Bayouth, Q. Song, A. Taurani, M. Sonka, J. Buatti, and X. Wu, "Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method," in *22nd International Conference on Information Processing in Medical Imaging* (Springer Verlag, Kloster Irsee, Germany, 2011), Vol. 6801, pp. 245–256.
[17]H. Yu, C. Caldwell, K. Mah, and D. Mozeg, "Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning," IEEE Trans. Med. Imaging **28**, 374–383 (2009).
[18]A. M. Ali and A. A. Farag, "Automatic lung segmentation of volumetric low-dose CT scans using graph cuts," in *Advances in Visual Computing* (Springer Berlin Heidelberg, Berlin, Germany, 2008), pp. 258–267.
[19]H. Lamecker, S. Zachow, A. Wittmers, B. Weber, H. Hege, B. Isholtz, and M. Stiller, "Automatic segmentation of mandibles in low-dose CT-data," Int. J. Comput. Assisted Radiol. Surg. **1**, 393–397 (2006).
[20]Y. Xie, M. Liang, D. F. Yankelevitz, C. I. Henschke, and A. P. Reeves, "Automated segmentation of cardiac visceral fat in low-dose non-contrast chest CT images," Proc. SPIE **9414**, 94140G-1–94140G-8 (2015).
[21]S. Liu, M. Salvatore, D. F. Yankelevitz, C. I. Henschke, and A. P. Reeves, "Segmentation of the whole breast from low-dose chest CT Images," Proc. SPIE **9414**, 94140I-1–94140I-11 (2015).
[22]Y. Xie, M. Liang, D. F. Yankelevitz, C. I. Henschke, and A. P. Reeves, "Automated measurement of pulmonary artery in low-dose non-contrast chest CT images," Proc. SPIE **9414**, 94141G-1–94141G-9 (2015).
[23]S. Liu, Y. Xie, and A. P. Reeves, "Segmentation of the sternum from low-dose chest CT images," Proc. SPIE **9414**, 941403-1–941403-10 (2015).
[24]C.-Y. Hsu, C.-Y. Liu, and C.-M. Chen, "Automatic segmentation of liver PET images," Comput. Med. Imaging Graphics **32**, 601–610 (2008).
[25]M. G. Linguraru, J. A. Pura, V. Pamulapati, and R. M. Summers, "Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT," Med. Image Anal. **16**, 904–914 (2012).
[26]A. Shimizu, R. Ohno, T. Ikegami, H. Kobatake, S. Nawano, and D. Smutek, "Segmentation of multiple organs in non-contrast 3D abdominal CT images," Int. J. Comput. Assisted Radiol. Surg. **2**, 135–142 (2007).
[27]R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch, "Multi-organ localization with cascaded global-to-local regression and shape prior," Med. Image Anal. **23**, 70–83 (2015).
[28]B. Glocker, O. Pauly, E. Konukoglu, and A. Criminisi, "Joint classification-regression forests for spatially structured multi-object segmentation," in *Computer Vision–ECCV 2012* (Springer-Verlag Berlin, Berlin, 2012), pp. 870–881.
[29]A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui, "Regression forests for efficient anatomy detection and localization in computed tomography scans," Med. Image Anal. **17**, 1293–1303 (2013).
[30]Y. Zhan, X. S. Zhou, Z. Peng, and A. Krishnan, "Active scheduling of organ detection and segmentation in whole-body medical images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008* (Springer Berlin Heidelberg, Berlin, Germany, 2008), pp. 313–321.
[31]H. Wang, J. K. Udupa, D. Odhner, Y. Tong, L. Zhao, and D. A. Torigian, "Body-wide anatomy recognition in PET/CT images," Proc. SPIE **9415**, 941518-1–941518-7 (2015).
[32]J. Liu, K. Chang, L. Kim, E. Turkbey, L. Lu, J. Yao, and R. Summers, "Automated segmentation of thyroid gland on CT images with multi-atlas label fusion and random classification forest," Proc. SPIE **9414**, 941413-1–941413-7 (2015).
[33]X. Zhou, S. Morita, X. Zhou, H. Chen, T. Hara, R. Yokoyama, M. Kanematsu, H. Hoshi, and H. Fujita, "Automatic anatomy partitioning of the torso region on CT images by using multiple organ localizations with a group-wise calibration technique," Proc. SPIE **9414**, 94143K-1–94143K-6 (2015).
[34]H. Frimmel, J. Näppi, and H. Yoshida, "Centerline-based colon segmentation for CT colonography," Med. Phys. **32**, 2665–2672 (2005).
[35]S. L. Lee and D. H. Cho, "Solitary pulmonary nodule detection on thoracic CT images through object continuity analyses," Appl. Mech. Mater. **284**, 1681–1685 (2013).
[36]X. Zhou, S. Wang, H. Chen, T. Hara, R. Yokoyama, M. Kanematsu, and H. Fujita, "Automatic localization of solid organs on 3D CT images by a collaborative majority voting decision based on ensemble learning," Comput. Med. Imaging Graphics **36**, 304–313 (2012).
[37]P. Campadelli, E. Casiraghi, and A. Esposito, "Liver segmentation from computed tomography scans: A survey and a new algorithm," Artif. Intell. Med. **45**, 185–196 (2009).
[38]R. Cuingnet, R. Prevost, D. Lesage, L. Cohen, B. Mory, and R. Ardon, "Automatic detection and segmentation of kidneys in 3D CT images using random forests," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012* (Springer Berlin Heidelberg, Berlin, Germany, 2012), pp. 66–74.
[39]M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision* (Cengage Learning, Stamford, CT, 2008).
[40]Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman, "MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies," IEEE Trans. Med. Imaging **25**, 464–475 (2006).
[41]N. Székely, N. Tóth, and B. Pataki, "A hybrid system for detecting masses in mammographic images," IEEE Trans. Instrum. Meas. **55**, 944–952 (2006).
[42]R. Bellotti, F. De Carlo, S. Tangaro, G. Gargano, G. Maggipinto, M. Castellano, R. Massafra, D. Cascio, F. Fauci, and R. Magro, "A completely automated CAD system for mass detection in a large mammographic database," Med. Phys. **33**, 3066–3075 (2006).
[43]S. Mohamed, M. Salama, M. Kamel, E. El-Saadany, K. Rizkalla, and J. Chin, "Prostate cancer multi-feature analysis using trans-rectal ultrasound images," Phys. Med. Biol. **50**, N175–N185 (2005).
[44]P. K. Saha, J. K. Udupa, and D. Odhner, "Scale-based fuzzy connected image segmentation: Theory, algorithms, and validation," Comput. Vision Image Understanding **77**, 145–174 (2000).
[45]T. Messay, R. C. Hardie, and T. R. Tuinstra, "Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung Image Database consortium and Image Database Resource Initiative dataset," Med. Image Anal. **22**, 48–62 (2015).
[46]L. Zhao, J. K. Udupa, D. Odhner, H. Wang, Y. Tong, and D. A. Torigian, "Automatic anatomy recognition of sparse objects," Proc. SPIE **9413**, 94133N-1–94133N-8 (2015).
[47]T. Okada, K. Yokota, M. Hori, M. Nakamoto, H. Nakamura, and Y. Sato, "Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008* (Springer Berlin Heidelberg, Berlin, Germany, 2008), pp. 502–509.
[48]Since the voxel size in different images may not be the same in the training samples, interpolation and resizing of the image domain are done first for all samples so that it makes sense to perform the union operation for estimating Q.
[49]Since the acquired images are not isotropic for ICT, we decided to describe texture in the 2D slice for both ICT and PET.