

Body region localization in whole-body low-dose CT images of PET/CT scans using virtual landmarks

Peirui Bai#

College of Electronics, Communication and Physics, Shandong University of Science and Technology, Qingdao, Shandong 266590, China

Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA

Jayaram K. Udupa^{*a)} and Yubing Tong[#]

Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA

ShiPeng Xie

Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China

Drew A. Torigian

Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA

(Received 7 October 2018; revised 11 December 2018; accepted for publication 31 December 2018; published 24 January 2019)

Purpose: Radiological imaging and image interpretation for clinical decision making are mostly specific to each body region such as head and neck, thorax, abdomen, pelvis, and extremities. In this study, we present a new solution to trim automatically the given axial image stack into image volumes satisfying the given body region definition.

Methods: The proposed approach consists of the following steps. First, a set of reference objects is selected and roughly segmented. Virtual landmarks (VLs) for the objects are then identified by using principal component analysis and recursive subdivision of the object via the principal axes system. The VLs can be defined based on just the binary objects or objects with gray values also considered. The VLs may lie anywhere with respect to the object, inside or outside, and rarely on the object surface, and are tethered to the object. Second, a classic neural network regressor is configured to learn the geometric mapping relationship between the VLs and the boundary locations of each body region. The trained network is then used to predict the locations of the body region boundaries. In this study, we focus on three body regions — thorax, abdomen, and pelvis, and predict their superior and inferior axial locations denoted by TS(I), TI(I), AS(I), AI(I), PS(I), and PI(I), respectively, for any given volume image *I*. Two kinds of reference objects — the skeleton and the lungs and airways, are employed to test the localization performance of the proposed approach.

Results: Our method is tested by using low-dose unenhanced computed tomography (CT) images of 180 near whole-body ¹⁸F-fluorodeoxyglucose-positron emission tomography/computed tomography (FDG-PET/CT) scans (including 34 whole-body scans) which are randomly divided into training and testing sets with a ratio of 85%:15%. The procedure is repeated six times and three times for the case of lungs and skeleton, respectively, with different divisions of the entire data set at this proportion. For the case of using skeleton as a reference object, the overall mean localization error for the six locations expressed as number of slices (*nS*) and distance (*dS*) in mm, is found to be *nS*: 3.4, 4.7, 4.1, 5.2, 5.2, and 3.9; *dS*: 13.4, 18.9, 16.5, 20.8, 20.8, and 15.5 mm for binary objects; *nS*: 4.1, 5.7, 4.3, 5.9, 5.9, and 4.0; *dS*: 16.2, 22.7, 17.2, 23.7, 23.7, and 16.1 mm for gray objects, respectively. For the case of using lungs and airways as a reference object, the corresponding results are, *nS*: 4.0, 5.3, 4.1, 6.9, 6.9, and 7.4; *dS*: 15.0, 19.7, 15.3, 26.2, 26.2, and 27.9 mm for binary objects; *nS*: 3.9, 5.4, 3.6, 7.2, 7.2, and 7.6; *dS*: 14.6, 20.1, 13.7, 27.3, 27.3, and 28.6 mm for gray objects, respectively.

Conclusions: Precise body region identification automatically in whole-body or body region tomographic images is vital for numerous medical image analysis and analytics applications. Despite its importance, this issue has received very little attention in the literature. We present a solution to this problem in this study using the concept of virtual landmarks. The method achieves localization accuracy within 2–3 slices, which is roughly comparable to the variation found in localization by experts. As long as the reference objects can be roughly segmented, the method with its learned VLs-toboundary location relationship and predictive ability is transferable from one image modality to another. © 2019 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.13376] Key words: body region identification, computed tomography (CT), neural network learning, principal component analysis, virtual landmarks

1. INTRODUCTION

1.A. Background

To fully harness the power of Quantitative Radiology in numerous applications, body-wide localization and delineation of objects is becoming increasingly important. An "object" here may denote an organ, a lymph node zone, a tissue mass or region (such as intrathoracic adipose tissue), or a body region (such as thorax). For developing generalizable methods that operate body-wide, for meaningful use of quantitative information, and for standardized clinical operation, standardized definitions of these objects become essential. For example, without a precise definition of the boundaries of the thoracic body region and intrathoracic adipose tissue region, standardized quantification of intrathoracic fat becomes impossible.¹⁻³ Standardized object definitions can also facilitate enriching and sharpening prior knowledge that is encoded into and utilized in methods for localizing objects body-wide^{4,5} and distinguishing different patient groups.^{6,7} In this spirit, body region definition becomes just as important as or even more important than objects contained in the body region. Some objects that cross body regions depend directly on precise body region definition for their accurate specification. For example, the superior and inferior boundaries of thoracic esophagus and thoracic spinal cord are decided by the superior-inferior boundaries of the thoracic body region. In our previous work on body-wide Automatic Anatomy Recognition (AAR),^{4,5} standardized definitions were employed for objects and body regions. However, body regions were located manually in the given computed tomography (CT)/magnetic resonance imaging (MRI)/positron emission tomography/computed tomography (PET/CT) data sets by specifying their superior and inferior axial boundaries. Subsequently, the AAR algorithms localized objects contained in the body region automatically. In this paper, we address the first problem of automatically determining the body regions — thorax, abdomen, and pelvis — in given whole-body image data sets. The solution trivially extends also from whole-body images to acquired image data sets of specific body regions.

1.B. Related works

Published works directly addressing the above problem are quite sparse.^{8–10} In fact, we did not come across any publication that directly dealt with the specific problem of identifying body regions as addressed in this paper. As to specific body region localization, in order to detect lymphoma regions automatically in the thresholded whole-body PET/CT images, Bi et al.⁸ used an adaptive thresholding method to estimate the section of lungs, and then partition roughly PET/ CT images into three sections — above lungs, lungs, and below lungs to reduce the search space. More recently, Bai et al.⁹ proposed an automatic thoracic body region localization method using a neural network regression learning technique, following the body region definition formulated in the AAR system. Perhaps the most relevant paper relating to the problem we tackle in this work is Ref. [10] where the authors' goal is segmentation and quantification of intrathoracic adipose tissue based on PET/CT scans. The authors employ the definition of the thoracic body region formulated in Ref. [4] and propose a method to automatically localize thusly defined top and bottom slices of abdominal and thoracic regions based on one shot learning. Deep learning (Convolutional Neural Network features) was employed to avoid confusion between similarly appearing slices.¹⁰

Investigations that are peripherally related to the specific problem we address are those¹¹⁻¹⁷ whose aim is to localize an anatomic organ such as liver, spleen, vertebral bodies, etc., or an anatomic feature such as iliac crest. For example, Rohr et al.^{11,12} introduced a deformable model and subsequent three-dimensional (3D) parametric intensity models to estimate position of the 3D point landmarks of the human head. The models can detect tip-like, saddlelike, and sphere-like anatomical structures efficiently by considering more global image information. It is demonstrated that the approach significantly improves the localization accuracy of the ventricular horns, the zygomatic bone, and the eyes. Yao et al.¹³ proposed an approach for simultaneously localizing rectangular boxes bounding eleven organs such as liver, spleen, etc., in abdominal CT images. They use a probabilistic atlas of organs to guide the selection of candidate organ locations by matching the local eigen-organ spaces and the global eigen-space which are constructed using principal component analysis (PCA) from training samples. Criminisi et al.14,15 introduced random decision forests and random regression forests for automatic detection and localization of anatomical structures from 3D CT volume scans. This kind of discriminative classification approach is shown to be better suited to multiclass problems. Multiple anatomical organs such as head, heart, eves, kidney, lung, and liver could be localized simultaneously with attractive accuracy. Chu et al.¹⁶ proposed a similar solution for automatic localization of vertebral body from 3D CT/MR images. They first obtain a probability map of the vertebral body center using random forest regression, and then define a target region of interest from regularized operation of the probability map using Hidden Markov Model (HMM). For this supervised classification technique, a manual annotation to prepare the labeled ground-truth database and exemplars (e.g., a 3D bounding box centered on each organ) is required. Potesil et al.¹⁷ proposed a parts-based graphical model to localize 22 specific anatomical landmarks in the human upper body in 3D CT scans. The method adopts dense matching of the parts-based graphical models to accurately and reliably localize standard anatomical landmarks.

While some of the above methods, particularly those on finding bounding boxes encasing specific anatomic features, may be adapted to the problem formulated in this paper, it is not obvious how this generalization can be done or what level of accuracy can be expected.

2. MATERIALS AND METHODS

We will follow the schematic in Fig. 1 to describe our approach.

Our approach to the problem of parcellation of body into body regions on whole-body low-dose CT images of PET/CT acquisitions, as explained in Section 2, consists of two stages — a training stage and a testing stage each further consisting of two steps. See Fig. 1 for a schematic illustration. We assume that a fixed definition of each body region of focus in this paper - Thorax, Abdomen, and Pelvis - is available in the form of their superior and inferior anatomic axial boundaries in the craniocaudal direction, and denote these axial locations on an image I by TS(I), TI(I), AS(I), AI(I), PS(I), and PI(I), respectively. We then determine a set of easily (roughly) identifiable reference objects such as lung space and skeletal structures in the CT image of the CT data set. In the training stage, first a set of Virtual Landmarks (VLs) is computed for the reference objects in each training data set. Roughly speaking, VLs are points in the anatomic space that are tethered to the reference objects and may lie anywhere with respect to the objects — inside, outside,



Fig. 1. A schematic illustration of the proposed approach for predicting body region boundaries TS(I), TI(I), AS(I), AI(I), PS(I), PI(I).

or on the boundary. Just the binary images constituting the reference objects or the binary images together with the gray values may also be used to determine the VLs. Then, a neural network is trained to regress the relationship between the VLs and the known true locations of TS(I), TI(I), AS(I), AI(I), PS(I), and PI(I) over all training images I. In the testing stage, given a PET/CT data set, first the reference objects are roughly identified on the CT image and their VLs are computed. Subsequently, the trained neural network is employed to predict the locations

An early version of this work was presented at the SPIE Medical Imaging Conference in 2017.⁹ The paper published in the proceedings of that conference differs in major ways from this work as follows. (a) That paper focused only on the thoracic body region. Here, we generalize the approach to include multiple body regions. (b) The conference report dealt with binary images only for computing VLs. In this work, we generalize from binary only to binary and gray images. (c) We consider a much larger number of data sets in this work and a more comprehensive evaluation strategy than the earlier work. (d) Here, we present a full review of related research which was not included in the SPIE conference paper.

of the six body region boundaries in the CT image. We

evaluate our approach utilizing 180 PET/CT data sets as described in Section 3 and summarize our conclusions

2.A. Data sets and notations

from this work in Section 4.

In this study, we use ¹⁸F-fluorodeoxyglucose (FDG)-PET/ CT scans from 180 patients already existing in our health system patient image database. We obtained approval for data usage from the Institutional Review Board at the Hospital of the University of Pennsylvania along with a Health Insurance Portability and Accountability Act waiver. Subjects include near-normal cases and patients with different types of disease conditions where all scans were administered for clinical reasons only. Of these 180 scans, 34 were scans covering the entire body from head to feet (typically comprising of 465 axial slices) and the remaining 146 were near whole-body scans extending from neck to feet (each comprising of close to 300 axial slices). At present, we use only the low-dose CT portion of these data sets (see Discussion for further comments). The voxel size in these CT data sets is roughly $1 \times 1 \times 4$ mm³; the slice spacing varied from 3 to 5 mm: 139 studies with 4 mm, 37 with 3 mm, and 4 with 5 mm, their weighted average being 3.8 mm. It is important to keep in mind this clinical slice spacing (~4 mm) in understanding the accuracy of our results.

We will use the following notations throughout. \mathcal{I} : Our collection of CT image data sets. \mathcal{I}_T : The subset of \mathcal{I} used for training our methods. \mathcal{I}_t : The subset of \mathcal{I} used for testing our approach. *I*: a given image data set of a patient. *TS*(*I*), *TI*(*I*), *AS*(*I*), *AI*(*I*), *PS*(*I*), *PI*(*I*): Known true superior and inferior boundary locations of the thorax, abdomen, and pelvis, respectively, in image *I*. *ts*(*I*), *ti*(*I*), *as*(*I*), *ai*(*I*), *ps*(*I*), and *pi*(*I*):

Superior and inferior boundary locations of the thorax, abdomen, and pelvis, respectively, in image I predicted by our approach.

2.B. Definition of body regions

In medical practice, the human body is divided into several regions in the cranio-caudal direction: head, neck, upper extremities, thorax, abdomen, pelvis, and lower extremities. In this study, we focus on three body regions — thorax, abdomen, and pelvis, and reuse in this paper their definitions formulated in our previous work.^{4,5} Table I summarizes the definitions. We define a body region by two axial slices: one denotes the superior axial limit or boundary and the other denotes the inferior axial boundary. Given a scan or image I, we denote the location of the superior axial slice of the thorax in I by TS(I) and the location of its inferior axial slice by TI(I)as defined in Table I. Similarly, we denote the superior and inferior axial locations of the abdominal and pelvic regions by AS(I), AI(I), PS(I), and PI(I), respectively, per Table I. Locations in all images are specified with reference to a fixed scanner coordinate system.

Note that, per our definition, AI(I) = PS(I). Note also how the abdominal and thoracic regions overlap. This is inevitable since the boundaries are defined through axial planes, and some of the axial planes passing through the thorax contain abdominal tissue regions. In Fig. 2, we show a close-up pictorial view of the definitions by displaying slices containing the features (encircled in the figure) that are used to determine true slice boundaries. Note how the distinguishing feature for AI is very subtle. Of course, our method does not look for these features but is based only on the relationship between VLs and the true boundary slice locations.

In Fig. 3, we display exemplar boundary slices from two subjects. There exist substantial differences in the appearance of slices at the same boundary location among different subjects. Depending on the thickness and spacing of the axial slices, there is some "digital ambiguity" as to which precise slice is to be selected to denote a specific boundary location. For example, for AI(I), where to call the exact slice where the abdominal aorta bifurcates into common iliac arteries is ambiguous by one or two slice. Thus, even when experts identify boundary slice locations manually, there can be a variation by about two slices. The difference in appearance of

boundary slices among subjects also suggests that it may be difficult to automatically locate boundary slices based only on intensity information. For all data sets in our collection \mathcal{I} and for all body regions, we have identified the true body region boundary locations TS(I), TI(I), AS(I), AI(I)/PS(I), and PI(I) manually via slice visualization under the guidance of the radiologist in our team (Torigian). These locations will be used as true locations for training our methods and for testing the accuracy of the locations predicted by our approach.

Our problem is: Given any PET/CT image I of a whole body, to find automatically the predicted locations ts(I), ti(I), as(I), ai(I), ps(I), and pi(I) of respectively TS(I), TI(I), AS(I), AI(I), PS(I), and PI(I) that are close to these true locations. We assume that the slices of I are organized axially and the region of the body it covers *properly includes* the body regions to be identified in I. For whole-body PET/CT imagery, this condition is always met. If this is not the case, the location predicted by our method will extend beyond the body region covered by I and will be in correct relationship with that data set and the subject although the corresponding slice may not be found in I.

2.C. Training: binary/gray-valued objects and their virtual landmarks

In this section, we will first explain how the reference objects are obtained and then the concept of virtual landmarks (VLs) for binary objects and its generalization to objects with gray values.

2.C.1. Reference objects and their segmentation

An object selected as a reference object should satisfy three key conditions: (C1) It should be segmentable reliably fully automatically and simply. (C2) It should not be confined within some small space in the body. (C3) It is manifested with roughly the same form and shape in all subject data sets when derived by the segmentation strategy satisfying C1. In our experience, objects that satisfy these conditions are: skeleton and lungs including trachea and bronchi. These objects can be segmented by thresholding (in CT images) owing to their distinct Hounsfield Unit (HU) ranges, the values we used being: [176, 3071] and [-894, -424] in HU, respectively. We will show results for these objects and

TABLE I. Definition of body regions and their boundary locations.

| Body region | Boundaries | Description | Definition |
|----------------|------------|--|--|
| Thorax | TS | Thoracic superior axial boundary location | 15 mm above the apex of the lungs |
| | TI | Thoracic inferior axial boundary location | 5 mm below the base of the lungs |
| Abdomen | AS | Abdominal superior axial boundary location | Superior-most aspect of the liver |
| | AI | Abdominal inferior axial boundary location | Point of bifurcation of the abdominal aorta into common iliac arteries |
| Pelvis | PS | Pelvic superior axial boundary location | Inferior boundary of the abdominal region |
| | PI | Pelvic inferior axial boundary location | Inferior-most aspect of the ischial tuberosities of the pelvis |



Fig. 2. Illustration of features (circled) used for defining region boundary planes. The slices displayed do not necessarily correspond to boundary planes per se. (a) A slice at the superior-most aspect of the lung (circled). TS is located 15 mm above the level shown. (b) A slice at the superior-most aspect of the liver (circled) = location of AS. (c) A slice at the inferior-most aspect of the lung (circled). TI is located 5 mm below this level. (d) A slice showing just when the abdominal aorta bifurcates into common iliac arteries (see magnified inset showing two roughly circular cross sections) = location of AI. (e) A slice showing the inferior-most aspect of the ischial tuberosities (circled) = location of PI. [Color figure can be viewed at wileyonlinelibrary.com]

discuss their pros and cons. We emphasize that after the thresholding operation, we perform an automated operation to remove any isolated voxels. Thus, the segmentation will correspond to the main reference object bulk. Otherwise, the segmentation does not have to be perfect as long as it is similar in all patients (condition C3).

2.C.2. Virtual landmarks

The idea of VLs is illustrated with a two-dimensional (binary) example in Fig. 4. Given a binary image representing the object, PCA of the entire binary object is first carried out to find the four principal axes directions, denoted in the figure in green by A_{1,1}, A_{1,2}, A_{1,3}, A_{1,4}, emanating from the geometric centroid of the object indicated by $P_{1,1,0}$ (small circle). Along these axes, we find points P_{1,1,1}, P_{1,1,2}, P_{1,1,3}, and P_{1,1,4} that indicate the extent of the object in those directions. These five points form the first level landmarks, where the first subscript denotes level number, second denotes quadrant number, and the third indicates point number. These points and the axes subdivide the shape into four pieces in the four quadrants. For each piece, we perform PCA again and find the 20 s level landmarks denoted P_{2,1,0}, P_{2,1,1}, ..., P_{2,1,4}, P_{2,2,0}, ..., $P_{2,4,4}$. The five points $P_{2,4,0}$, ..., $P_{2,4,4}$ obtained for the 4th quadrant are shown in the figure for illustration (2nd level principal axes are shown in red). The process continues up to a specified level. Since the points are ordered, each point has a unique label. This allows us to specify the VLs we need by their label for representing a given shape. For example, we may use just the 8 points P_{1,1,0}, P_{1,1,1}, P_{1,1,2}, P_{1,1,3}, P_{1,1,4}, P_{2,4,1}, P_{2,4,2}, and P_{2,4,4} (which already denote the shape roughly). Note how the points tend to move closer to the

object surface at higher levels. Points at early levels capture overall form and add details at later levels.

The total number N(x) of VLs for a *d*-dimensional object derived from *x* levels will be $N(x) = \sum_{n=1}^{x} (2d+1)2^{d(n-1)}$. If we consider only the geometric centroids (points identified by 0 value for their 3rd subscript index such as P_{2,2,0}), the total number of points will be $\sum_{n=1}^{x} 2^{d(n-1)}$. The approach readily generalizes to multiple objects — either by finding VLs for each object separately and pooling the VLs together, or by first pooling the objects into one object and finding its VLs. The sets of resulting VLs in the two cases may not be the same. For objects with gray CT values, the gray pixel values in the shape are used as a weight factor in PCA computation. Now, the shape as well as the gray values of the object influence the location of the VLs relative to the object. Obviously gray-value-based VLs also generalize straightforwardly from single object to multiple objects and even to vectorvalued images.

Note that the VLs may lie anywhere with respect to the object. In fact, in our experience with several anatomic objects, VLs are rarely located exactly on the object boundary. The number of VLs rises rapidly with the number of levels. For example, for a 3D object, N(1) = 7, N(2) = 63, and N(3) = 511. If we consider only geometric centers, N(1) = 1, N(2) = 9, and N(3) = 73. Unlike methods of finding landmarks on object boundaries, the concept of VLs generalizes to spaces of any finite dimension directly and easily. Usually, 50–100 VLs are sufficient to describe a large object. For an early report on the concept of VLs and their general properties, see Ref. [19]. In this paper, we will not study these matters and focus only on the application of VLs for localizing body regions. After the reference objects are segmented,



Fig. 3. Illustration of the boundary locations of the three body regions and examples of axial slices at those locations selected from two patient computed tomography data sets. Illustration on left reproduced with permission from Ref. [18]. [Color figure can be viewed at wileyonlinelibrary.com]



Fig. 4. Illustration of the process of defining virtual landmarks for a twodimensional binary shape. [Color figure can be viewed at wileyonlinelibra ry.com]

their VLs are computed automatically following the above recursive subdivision algorithm. We will test the prediction performance of our approach using different reference objects and both binary and gray-valued versions of the objects.

2.D. Training: Learning the relationship between VLs and body region boundaries

For this stage, input is the set of VLs of the chosen reference objects in the training image set \mathcal{I}_T and the set of true boundary locations TS(I), TI(I), AS(I), AI(I), PS(I), and PI(I)for each image I in \mathcal{I}_T . The outcome of this stage is a trained neural network. Input-vector-output-vector pairs (u(I), v(I))used for network training are as follows: [u(I), v(I)], for all Iin the training image data set \mathcal{I}_T , where $u(I) = [P_{1,1,0}(I),$ $P_{1,1,1}(I)$, $P_{1,1,2}(I)$, ..., $P_{L,8,0}(I)$, ..., $P_{L,8,6}(I)]^t$ and v(I) = [TS(I), TI(I), AS(I), AI(I), PS(I), $PI(I)]^t$. All locations here are expressed in terms of coordinates with respect to the scanner coordinate system associated with image I. Note that each VL is described by its three coordinates and each boundary location TS(I), TI(I), ..., PI(I) is described by one coordinate, namely the coordinate in the cranio-caudal direction. We employed a neural network regressor²⁰ (Neural Network Toolbox, Version 9.0, of MATLAB, Version R2016a) to learn the relationship between VLs and boundary locations. This toolbox provides a convenient platform to design an application-oriented neural network. Relevant details pertaining to our application are as follows.

2.D.1. Choice of neural network architecture and configuration

As we want to solve a nonlinear mapping problem, a multilayer architecture with a single hidden layer would be sufficient. Here, we follow the layer designation of MATLAB's Neural Network Toolbox. As shown in Fig. 5 a layer of neurons includes the weights, multiplication, and summing operations. It is common for the number of inputs to a layer to be different from the number of neurons. In our application, the input is presented as a set of vectors $\{u(I), I \in \mathcal{I}_T\}$. The number of elements in each input vector is $N \times 3$ where N is the number of VLs employed. The number of neurons, denoted as S in the hidden layer, is adjusted to minimize the localization errors. S is not necessarily equal to N. The number of elements in the output vector is M = 6 (strictly speaking, 5).

2.D.2. Choice of training parameters

For each neuron, there are three operations — the weight function (matrix multiplication), the net input function (summation), and the transfer function. Here, we adopt the "Hyperbolic Tangent Sigmoid" transfer function in the hidden layers, and the "Linear" transfer function in the output layer (which are denoted as "tansig" and "purelin" in MATLAB's toolbox, respectively). The "Bayesian Regularization" training algorithm is employed to prevent overfitting.^{21,22} We employ two stopping criteria, namely, a fixed number of iterations and the gradient of the performance index to control the iterative procedure. The training

performance index is selected as Mean Squared Error (which is denoted as "MSE" in the toolbox).

2.E. Testing: predicting body region boundary locations

In this stage, given an image I, first the reference objects employed in the previous stage are identified in I by using the same segmentation strategy as employed in the previous stage. Then, the same specific set of VLs of the reference objects as utilized for training the network is computed. These VLs are fed to the neural network whose output variables correspond to the six predicted boundary locations of the three body regions.

2.F. Evaluation

To evaluate the performance of this approach, we compare the predicted boundary locations to the "true" expert determined locations and express the deviation in terms of number of slices, nS, and the distance, dS (in mm), between the two locations. Our evaluations will involve several different divisions of the data set \mathcal{I} into training and testing subsets \mathcal{I}_T and \mathcal{I}_t and multiple folds (different repetitions of this division) and use four reference objects - skeleton and lungs in both binary and gray forms. We have also experimented with different numbers of VLs - 9, 17, 25, and 73. All VLs considered here are geometric centers only and are confined to the first three levels. In our notation, these points are identified by $P_{i,i,0}$ (see Fig. 4). The set with 9 points corresponds to all VLs of this type from the first two levels (1 from 1st level + 8 from 2nd level). The set with 73 points corresponds to all VLs of this type from the first three levels (1 + 8+64). Sets with 17 and 25 VLs are formed by selecting all VLs from the first two levels and different subsets from the third level. One other variable involved in our experiments is the number of coordinates considered for the VLs - all three coordinates (x, y, z) and only the third coordinate (z) in the cranio-caudal



Fig. 5. Illustration of the neural network architecture, employed from MATLAB's Neural Network Toolbox, used as a regressor in our method. The architecture shows one hidden layer.

direction. The idea for a single coordinate stems from the consideration that we are interested in predicting only the *z*-level of the boundary slices.

3. RESULTS

In Fig. 6, we depict via 3D renditions the (set of 73) VLs derived from binary and gray versions of sample reference objects obtained from one subject. Notably many VLs lie outside the object. VLs that are interior to the object are not visible in the display. The animations with translucent surface displays available at the link in Ref. [23] depict more vividly the spatial distribution of VLs that are interior and exterior to the objects, where the virtual landmarks from both binary image and gray images for the two reference objects are shown.

Sample results (good and poor) of identified slices for the region boundaries are displayed in Fig. 7, where the true slices are also shown. Tables II and III summarize mean prediction errors resulting when Skeleton and Lungs, respectively, are used as reference objects. The tables include results for different settings — binary and gray objects, different numbers of VLs, different selections of coordinates, and different numbers of hidden layers employed in the network.

We randomly divided our data samples into training and testing data sets with the ratio 0.85:0.15 and repeated the experiments six times for the case of using lungs as the reference object on the 146 near whole-body scans and three times (at the same ratio) for the case of using skeleton as the reference object on the 34 whole-body scans. Figure 8 shows scatter plots of the mean prediction errors (nS) listed in Tables II and III using Skeleton and Lungs as reference objects and employing (x, y, z) and (z) coordinates. To study the difference in prediction accuracy among different scenarios, we performed t tests pairwise between different scenarios. P values from such comparisons using binary vs gray images are summarized in Table IV. To understand the accuracy of our method, we conducted an experiment to study the variability in body region boundary localization by knowledgeable operators. Table V summarizes the variability found between two operators in labeling all six boundary locations in all 180 CT data sets.

4. DISCUSSIONS

We make the following inferences from Tables II–V.

(a)For a given region boundary, prediction accuracy varies with gray/binary objects utilized for deriving VLs, the



FIG. 6. Three-dimensional renditions of the reference objects. (a) Skeleton, (b) Lungs, along with the associated VLs derived from binary (left) and gray (right) objects. [Color figure can be viewed at wileyonlinelibrary.com]



Fig. 7. Exemplar true and predicted region boundary slices for a good case (top two rows; $nS \le 2$) and a poor case (bottom two rows; nS of 3–5 slices). Please see Table I for body region boundary definitions.

number of VLs used, coordinate selection, number of hidden layers in the network, and the actual reference object. For example, in Table II, for virtual landmarks from binary mask and gray image, with the same number of hidden layers (2) and virtual landmarks (25), the average prediction error is 4.2 and 16.8 for nS and dS, receptively for binary mask and 4.8 and 19.1 for gray image. We can also observe a similar difference in Table III. Figure 8 shows scatter plots of the mean prediction errors for nS listed in Tables II and III, where the different experimental scenarios are also indicated. For example, "Binary-xyz-HL2-VLs9" denotes the situation of using binary object, (x, y, z) coordinates, two hidden layers, and nine virtual landmark points. Two observations can be made from the scatter plots. The spread of error for localizing AS(I)seems to be the smallest among all boundaries. Also, although the errors themselves are larger, the spread of the errors seems to be the smallest for the case of using Lungs with *z*-coordinate only.

(b)Our observations from Table IV (and other similar comparisons) can be summarized as follows. (a) When using the skeleton as the reference object with smaller number of hidden layers (\leq 5) and VLs (\leq 17) the mean prediction error ($nS \sim 4$) is lower (P < 0.02) than when using larger number of hidden layers and VLs ($nS \sim 5$). With lung as the reference object (binary or gray), in almost all scenarios with the same number of hidden layers statistically significantly (P < 0.001) greater when using smaller number of VLs. (c) The difference in accuracy between (x, y, z) vs z over all was not statistically

significant (not shown in Table IV). (d) Comparing between the two reference objects, the best accuracy achieved with skeleton (bold in Table II) is better than the best accuracy achieved with lungs (bold in Table III) but not with statistical significance (P = 0.10).

(c)From Table V, we observe that expert variability is the smallest for TS(I) and AS(I), intermediate for TI(I) and PI(I), and notably the largest for AI(I) = PS(I). The best accuracy achieved (shown in bold in Tables II and III) is variable among the different region boundaries and follows the trend in variability in expert localization of the boundaries. That is, when expert localization variability is greater so is the error in automatic localization. AI(I) is the most challenging to localize for our method as well as for experts. The best overall mean localization accuracy (error) for our method is nS = 4.1 and dS = 15.5 mm for the case of using lungs as the reference object, and nS = 3.2 and dS = 12.7 mm for the case of using skeleton as the reference object. There is only one paper¹⁰ we came across that addressed the problem of body region localization as formulated in this paper. The error reported in that study is ~47 mm for localizing the thoracic body region following the definition shown in Table I. The focus of that study was to detect white and brown adipose tissues automatically from PET/CT scans. A rough initial boundary estimation seems enough for that application and it does not require very precise body region localization. Furthermore, their application, data sets, and study scope were different from ours.

(d)Some region boundaries are more challenging than others for accurate localization (recall Fig. 2). As seen from

TABLE II. Mean and SD (second entry in each cell) of prediction errors *nS* and *dS* (in mm) over all tested data sets for the different region boundaries when using *Skeleton* as the reference object. The column Mean shows mean error over all region boundaries over all tested data sets. Bold entries indicate the setting with the best result.

| | Train/test/ | (<i>x</i> , <i>y</i> , <i>z</i>) or (<i>z</i>) | (x, y, z) | Hiddon | | t: | s(I) | t | i(I) | а | s(I) | ai(I) | = ps(I) | р | i(I) | Μ | lean |
|--------|-------------|--|-----------|--------|-----|------|------|------|------|------|------|-------|---------|------|------|------|------|
| Object | folds | | layers | VLs | nS | dS | nS | dS | nS | dS | nS | dS | nS | dS | nS | dS | |
| Binary | 29/5/3 | (x, y, z) | 2 | 9 | 3.0 | 12.0 | 4.2 | 16.6 | 3.7 | 14.9 | 5.5 | 22.0 | 3.8 | 15.0 | 4.0 | 16.1 | |
| | | | | | 2.5 | 10.0 | 3.1 | 12.5 | 2.7 | 10.7 | 2.9 | 11.6 | 2.8 | 11.1 | | | |
| | | | 1 | 17 | 3.0 | 12.2 | 4.4 | 17.5 | 3.6 | 14.3 | 5.8 | 23.4 | 4.8 | 19.1 | 4.3 | 17.3 | |
| | | | | | 1.9 | 7.7 | 3.0 | 12.0 | 2.9 | 11.7 | 4.3 | 17.0 | 4.0 | 15.8 | | | |
| | | | 2 | 25 | 2.4 | 9.7 | 4.3 | 17.0 | 4.2 | 16.8 | 6.0 | 23.9 | 4.2 | 16.6 | 4.2 | 16.8 | |
| | | | | | 2.2 | 8.9 | 2.9 | 11.6 | 3.0 | 11.9 | 3.4 | 13.5 | 2.1 | 8.3 | | | |
| | | | 3 | 73 | 2.6 | 10.5 | 5.2 | 20.8 | 3.5 | 14.0 | 5.2 | 20.7 | 3.1 | 12.4 | 3.9 | 15.7 | |
| | | | | | 1.5 | 6.1 | 3.1 | 12.3 | 2.3 | 9.2 | 3.1 | 12.4 | 2.2 | 8.6 | | | |
| | | (<i>z</i>) | 2 | 9 | 4.7 | 18.7 | 5.4 | 21.5 | 4.6 | 18.5 | 4.4 | 17.5 | 3.4 | 13.7 | 4.5 | 18.0 | |
| | | | | | 2.9 | 11.7 | 3.5 | 14.2 | 3.3 | 13.2 | 3.2 | 12.7 | 1.8 | 7.2 | | | |
| | | | 1 | 17 | 3.5 | 13.9 | 4.9 | 19.7 | 4.3 | 17.3 | 5.2 | 20.8 | 4.5 | 18.0 | 4.5 | 17.9 | |
| | | | | | 2.6 | 10.5 | 3.0 | 12.0 | 2.5 | 10.1 | 4.4 | 17.4 | 4.1 | 16.3 | | | |
| | | | 1 | 25 | 3.8 | 15.0 | 5.2 | 20.8 | 5.0 | 19.9 | 5.0 | 20.1 | 4.1 | 16.4 | 4.6 | 18.4 | |
| | | | | | 2.8 | 11.0 | 3.2 | 13.0 | 2.7 | 10.8 | 4.4 | 17.6 | 3.1 | 12.5 | | | |
| | | | 3 | 73 | 3.8 | 15.4 | 4.3 | 17.3 | 4.0 | 16.1 | 4.6 | 18.3 | 3.1 | 12.5 | 4.0 | 15.9 | |
| | | | | | 3.4 | 13.5 | 3.5 | 13.9 | 1.5 | 6.1 | 3.0 | 11.8 | 2.3 | 9.2 | | | |
| Gray | 29/5/3 | (x, y, z) | 1 | 9 | 3.8 | 15.2 | 5.5 | 22.0 | 3.9 | 15.6 | 6.0 | 24.0 | 5.2 | 20.8 | 4.9 | 19.5 | |
| | | | | | 3.0 | 12.1 | 4.3 | 17.3 | 2.2 | 8.7 | 4.4 | 17.7 | 3.5 | 13.8 | | | |
| | | | 5 | 17 | 4.7 | 18.8 | 7.0 | 28.1 | 3.0 | 12.2 | 7.2 | 28.6 | 4.1 | 16.2 | 5.2 | 20.8 | |
| | | | | | 2.9 | 11.5 | 4.5 | 18.0 | 2.0 | 8.1 | 4.1 | 16.3 | 2.2 | 8.9 | | | |
| | | | 2 | 25 | 2.9 | 11.8 | 5.5 | 21.9 | 4.7 | 18.8 | 6.1 | 24.4 | 4.7 | 18.7 | 4.8 | 19.1 | |
| | | | | | 2.3 | 9.3 | 3.3 | 13.1 | 3.0 | 12.1 | 3.8 | 15.4 | 2.3 | 9.2 | | | |
| | | | 3 | 73 | 2.6 | 10.6 | 5.4 | 21.6 | 3.3 | 13.0 | 5.0 | 19.9 | 3.1 | 12.4 | 3.9 | 15.5 | |
| | | | | | 1.5 | 5.9 | 3.3 | 13.1 | 2.2 | 8.9 | 3.4 | 13.7 | 2.8 | 11.4 | | | |
| | | (z) | 2 | 9 | 6.1 | 24.5 | 6.5 | 26.1 | 5.2 | 20.6 | 5.7 | 22.7 | 4.2 | 16.9 | 5.5 | 22.2 | |
| | | | | | 4.5 | 18.0 | 4.8 | 19.1 | 3.3 | 13.0 | 5.6 | 22.4 | 4.2 | 17.0 | | | |
| | | | 5 | 17 | 5.6 | 22.3 | 7.7 | 30.9 | 5.6 | 22.5 | 8.0 | 31.8 | 3.8 | 15.1 | 6.1 | 24.5 | |
| | | | | | 2.5 | 10.1 | 5.3 | 21.1 | 4.7 | 18.8 | 4.3 | 17.3 | 3.1 | 12.4 | | | |
| | | | 1 | 25 | 4.0 | 15.8 | 4.6 | 18.5 | 5.1 | 20.3 | 5.5 | 22.1 | 4.6 | 18.3 | 4.8 | 19.0 | |
| | | | | | 2.0 | 7.8 | 4.2 | 16.7 | 2.9 | 11.4 | 4.3 | 17.1 | 3.7 | 14.7 | | | |
| | | | 4 | 73 | 2.7 | 10.9 | 3.0 | 12.1 | 3.7 | 14.7 | 3.9 | 15.8 | 2.5 | 10.2 | 3.2 | 12.7 | |
| | | | | | 1.8 | 7.1 | 3.0 | 12.0 | 1.9 | 7.7 | 3.2 | 12.7 | 1.8 | 7.4 | | | |

Tables II and III, TI, AI = PS, and PI seem to have less accuracy than the locations in the superior portion of the thorax. This may be due to large motion of the inferior thorax and abdominopelvic regions compared to the upper parts of the thorax.

(e)Dependence of accuracy on reference object size and shape is hard to decode. Objects that have larger spatial coverage seem to fare better than spatially confined objects. What seems to matter most is the precision of the relationship between the chosen VLs and region boundaries. Generally, the skeleton as the reference object seems to fare better than lungs, notably for PI, AI = PS, and TI.

(f)In principle, gray values bring in an additional subtlety to VL definition from image intensity pattern details. However, similar localization accuracies are obtained for VLs from

gray objects and binary versions as graphically illustrated in Fig. 9 which displays errors in nS for the different cases of objects and gray/binary versions of VLs. The vertical intervals indicate one standard deviation on either side of the mean.

(g)VLs with just their z coordinates seem to regress their relationship to region boundaries just as well as or better than all three coordinates taken together. Figure 10 graphically illustrates the differences. This suggests that the location of the VLs on the slice plane is less important than their slice location in the cranio-caudal direction. As mentioned in (d) above, the precision of the relationship to region boundaries is perhaps better for z alone vs all coordinates collectively.

(h)Variability in localization error (expressed by SD values in the tables) seems to be generally smaller for gray-value-based

TABLE III. Mean and SD (second entry in each cell) of prediction errors nS and dS (in mm) over all tested data sets for the different region boundaries when using *Lungs* as the reference object. The column Mean shows mean error over all region boundaries over all tested data sets. Bold entries indicate the setting with the best result.

| | | | | | t. | s(I) | ti | (I) | a | s(I) | ai(I) = | = ps(I) | pi | i(I) | М | lean |
|--------|------------------|--------------------|---------------|-----|-----|------|------|------|-----|------|---------|---------|------|------|-----|------|
| Object | Train/test/folds | (x, y, z) or (z) | Hidden layers | VLs | nS | dS | nS | dS | nS | dS | nS | dS | nS | dS | nS | dS |
| Binary | 153/27/6 | (x, y, z) | 4 | 9 | 4.0 | 15.0 | 5.0 | 18.6 | 3.8 | 14.3 | 6.2 | 23.5 | 6.6 | 25.1 | 5.1 | 19.3 |
| | | | | | 3.5 | 13.3 | 5.3 | 19.5 | 3.7 | 14.0 | 5.2 | 20.0 | 5.6 | 21.2 | | |
| | | | 1 | 17 | 4.5 | 17.2 | 4.6 | 17.1 | 4.9 | 18.4 | 5.6 | 21.5 | 6.5 | 24.8 | 5.2 | 19.8 |
| | | | | | 3.7 | 14.2 | 5.3 | 19.7 | 3.8 | 14.6 | 5.0 | 18.9 | 4.8 | 18.2 | | |
| | | | 10 | 25 | 4.6 | 17.3 | 7.9 | 29.7 | 5.2 | 19.6 | 8.5 | 32.2 | 8.1 | 31.0 | 6.9 | 26.0 |
| | | | | | 4.2 | 15.8 | 7.7 | 28.5 | 5.0 | 18.9 | 6.9 | 26.3 | 6.8 | 25.6 | | |
| | | | 10 | 73 | 6.8 | 25.3 | 8.3 | 31.1 | 6.7 | 25.0 | 12.0 | 45.1 | 11.8 | 44.2 | 9.1 | 34.1 |
| | | | | | 7.2 | 26.8 | 8.1 | 30.1 | 6.3 | 23.8 | 10.9 | 41.1 | 10.2 | 38.0 | | |
| | | (<i>z</i>) | 4 | 9 | 2.6 | 9.7 | 3.9 | 14.6 | 3.0 | 11.2 | 5.7 | 21.7 | 6.5 | 24.8 | 4.3 | 16.4 |
| | | | | | 2.7 | 10.5 | 4.4 | 16.3 | 2.9 | 10.9 | 5.0 | 19.0 | 5.0 | 19.2 | | |
| | | | 5 | 17 | 2.7 | 10.3 | 4.0 | 15.0 | 2.8 | 10.8 | 5.7 | 21.9 | 6.5 | 24.6 | 4.3 | 16.5 |
| | | | | | 2.7 | 10.3 | 4.6 | 16.8 | 2.8 | 10.6 | 5.0 | 19.1 | 4.9 | 18.9 | | |
| | | | 8 | 25 | 2.7 | 10.3 | 3.9 | 14.7 | 2.8 | 10.4 | 5.5 | 21.1 | 5.9 | 22.3 | 4.2 | 15.7 |
| | | | | | 2.8 | 10.7 | 4.5 | 16.6 | 2.7 | 10.4 | 4.8 | 18.3 | 4.6 | 17.7 | | |
| | | | 3 | 73 | 4.0 | 15.2 | 4.5 | 16.8 | 3.4 | 12.9 | 6.0 | 22.9 | 6.9 | 26.0 | 5.0 | 18.8 |
| | | | | | 3.5 | 13.5 | 5.1 | 18.9 | 3.2 | 12.1 | 5.4 | 20.4 | 5.1 | 19.3 | | |
| Gray | 153/27/6 | (x, y, z) | 4 | 9 | 3.7 | 13.8 | 4.4 | 16.7 | 3.2 | 12.2 | 6.3 | 24.0 | 6.7 | 25.5 | 4.9 | 18.4 |
| | | | | | 3.5 | 13.1 | 5.1 | 18.9 | 3.1 | 12.0 | 5.2 | 19.9 | 5.5 | 20.9 | | |
| | | | 2 | 17 | 4.6 | 17.3 | 4.8 | 18.0 | 3.9 | 14.8 | 6.3 | 23.8 | 6.9 | 26.3 | 5.3 | 20.0 |
| | | | | | 4.1 | 15.7 | 5.2 | 19.0 | 3.4 | 12.8 | 5.2 | 20.1 | 5.3 | 20.4 | | |
| | | | 7 | 25 | 4.6 | 17.3 | 7.7 | 28.8 | 4.6 | 17.2 | 9.8 | 36.9 | 9.6 | 36.1 | 7.3 | 27.3 |
| | | | | | 4.5 | 16.8 | 8.4 | 31.5 | 4.5 | 17.0 | 8.4 | 31.8 | 8.5 | 32.1 | | |
| | | | 20 | 73 | 7.7 | 28.9 | 10.3 | 38.7 | 6.7 | 24.9 | 11.9 | 44.8 | 11.1 | 41.9 | 9.5 | 35.8 |
| | | | | | 6.0 | 22.6 | 9.6 | 36.1 | 6.1 | 22.8 | 10.9 | 41.4 | 9.9 | 37.3 | | |
| | | (<i>z</i>) | 4 | 9 | 2.3 | 8.8 | 3.8 | 14.1 | 2.8 | 10.7 | 5.8 | 22.0 | 6.7 | 25.3 | 4.3 | 16.2 |
| | | | | | 2.6 | 9.8 | 4.4 | 16.2 | 2.8 | 10.7 | 5.0 | 19.0 | 5.2 | 20.0 | | |
| | | | 5 | 17 | 2.3 | 8.8 | 3.8 | 14.3 | 2.4 | 9.1 | 5.5 | 21.1 | 6.4 | 24.3 | 4.1 | 15.5 |
| | | | | | 2.5 | 9.6 | 4.4 | 16.0 | 2.4 | 9.3 | 4.9 | 18.8 | 5.0 | 19.2 | | |
| | | | 7 | 25 | 2.5 | 9.4 | 3.9 | 14.7 | 2.5 | 9.4 | 6.0 | 22.8 | 6.6 | 24.9 | 4.3 | 16.2 |
| | | | | | 2.7 | 10.2 | 4.5 | 16.4 | 2.4 | 9.0 | 5.3 | 20.0 | 5.5 | 20.7 | | |
| | | | 6 | 73 | 3.3 | 12.5 | 4.2 | 15.8 | 3.0 | 11.3 | 5.9 | 22.6 | 6.5 | 24.8 | 4.6 | 17.4 |
| | | | | | 3.3 | 12.6 | 4.9 | 18.2 | 2.9 | 11.2 | 5.1 | 19.3 | 4.8 | 18.5 | | |

VLs than for their binary counterpart. A similar trend can be observed for the locations which are affected by the larger motion of the thoraco-abdominal junction.

4.A. Computational considerations

Time for VL computation on an Intel(R) Core(TM) i7-7700K Processor (4-cores) computer with 64 GB RAM under Ubuntu 16.04 OS is in the range of 150–230 s per data set. When a large number of VLs with all (x, y, z) coordinates are employed, more neurons in the hidden layer will be needed and consequently, the training time goes up. For example, when dealing with 73 lung VLs with (x, y, z) coordinates, the training time is 4169 s, while the prediction time per data set for any configuration is ~0.3 s.

5. CONCLUDING REMARKS

Automatic localization of human body regions in clinical images can facilitate image reading, reporting, and quantification. More importantly, this becomes essential in systems designed for automatic recognition of anatomic organs and zones such as lymph node stations when such systems employ precise definitions of body regions, organs, and zones. In this paper, we presented a novel strategy for accomplishing this task based on the concept of virtual landmarks and learning their relationship to boundary locations of body regions via a neural network. The method can be utilized for carving out precisely (within about 3 slices) a specific body region or multiple body regions from a whole-body scan. This approach is generalizable provided reliable segmentation can be obtained across subjects, regardless of the



FIG. 8. Scatter plots of the mean prediction errors (nS) listed in Tables II and III using Skeleton (Row 1) and Lungs (Row 2) as reference objects and employing (x, y, z; left column) and (z; right column) coordinates. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE IV. P values from t tests comparing prediction errors over all boundary locations by using VLs from binary mask vs VLs from gray region.

| Skeleton as ref | erence object | | | Lungs as reference object | | | | | | |
|------------------|--------------------------------|----------------------|--|---------------------------|---------------------|--------------------------------|----------------------|--|---------|--|
| (x, y, z) or (z) | Hidden layers (binary/gray) | VLs (binary/gray) | Mean prediction error (<i>nS</i>) (binary/gray) | P value | (x, y, z) or (z) | Hidden layers (binary/gray) | VLs (binary/gray) | Mean prediction error (<i>nS</i>) (binary/gray) | P value | |
| (x,y,z) | 2/1 | 9/9 | 4.0/4.9 | 0.0122 | (x,y,z) | 4/4 | 9/9 | 5.1/4.9 | 0.0069 | |
| | 1/5 | 17/17 | 4.3/5.2 | 0.0518 | | 1/2 | 17/17 | 5.2/5.3 | 0.5634 | |
| | 2/2 | 25/25 | 4.2/4.8 | 0.0062 | | 10/7 | 25/25 | 6.9/7.3 | 0.1498 | |
| | 3/3 | 73/73 | 3.9/3.9 | 0.7845 | | 10/20 | 73/73 | 9.1/9.5 | 0.2865 | |
| (z) | 2/2 | 9/9 | 4.5/5.5 | 0.0211 | (z) | 4/4 | 9/9 | 4.3/4.3 | 0.1465 | |
| | 1/5 | 17/17 | 4.5/6.1 | 0.0027 | | 5/5 | 17/17 | 4.3/4.1 | 0.0002 | |
| | 1/1 | 25/25 | 4.6/4.8 | 0.4623 | | 8/7 | 25/25 | 4.2/4.3 | 0.1257 | |
| | 3/4 | 73/73 | 4.0/3.2 | 0.0030 | | 3/6 | 73/73 | 5.0/4.6 | 0.0058 | |

TABLE V. Mean (and SD) of the variability in nS and dS observed in boundary localization by two operators.

| | TS(I) | TI(I) | AS(I) | AI(I) = PS(I) | PI(I) |
|---------|-------|-------|-------|---------------|-------|
| nS | 0.10 | 1.04 | 0.14 | 3.19 | 0.76 |
| | 0.29 | 0.81 | 0.42 | 1.99 | 0.64 |
| dS (mm) | 0.4 | 4.16 | 0.56 | 12.76 | 3.04 |
| | 1.16 | 3.24 | 1.68 | 7.96 | 2.56 |
| | | | | | |

modality, although in this paper we demonstrated its performance on low-dose CT images of PET/CT acquisitions.

There are several potential applications of virtual landmarks in image and object analytics. One avenue we are currently pursuing is in organ localization (recognition). We are also examining the use of multiple objects simultaneously, instead of individually as investigated in this paper, for potentially improving accuracy. We believe that not all VLs behave the same way, and so weeding out those VLs that show large

Medical Physics, 46 (3), March 2019

subject-to-subject variation in their relationship (particularly in *z* value) to boundary locations may improve performance. On diagnostic CT scans, which are typically performed on individual body regions and not the whole body, performance may be better than on the low-dose CT scans we studied in this paper, specifically when gray-valued objects are used for computing VLs. A separate investigation is needed to study this task of individual body region localization on individual body region scans. We are also developing deep learning methods to localize region boundaries without the use of VLs. As we pointed out earlier, higher level landmarks correspond to fine structures and tend to be patient dependent. The issue of how to capture fine granularity via higher level VLs and describe differences in objects at that level over a population is another interesting topic of study for the future.

There are some potential limitations of this approach related to the need for a reference object and the use of PCA. The method will not work if the field of view of the scan does not include the full body region on slices (as it may often happen in MRI acquisitions) since such honed-in acquisitions may also affect reference objects due to their partial coverage.



FIG. 9. Graphical comparison of mean prediction error (nS) for all region boundaries for different reference objects. [Color figure can be viewed at wileyonline library.com]



Fig. 10. Graphical comparison of mean prediction error (nS) over all region boundaries for different reference objects: Use of all three coordinates vs only z. [Color figure can be viewed at wileyonlinelibrary.com]

However, if the skin outer boundary falls fully within the field of view, then rough segmentation of the body region (the interior of the skin boundary as a solid mask) is not difficult and then skin may be used as a reference object. Skeleton is easy to be achieved from CT images and can be used as a good reference object for computing VLs. From our observations on 34 whole-body PET/CT scans where skeleton is used as a reference object for VLs, we find that the effects of metallic artifacts or positive contrast agents may still share similar attenuation values with skeletal structures.

Another issue may arise from stray voxels which are isolated and/or fall outside the body region. They need to be removed; otherwise PCA may be influenced and the resulting VLs may not be reliable. For skeleton, this is usually not an issue. However, for lung object, in case of extreme pathology, automatic rough segmentation may not be acceptable. In our approach, after the reference object is roughly segmented, we detect and automatically remove such extraneous voxels by connectivity and morphological operations.

ACKNOWLEDGMENTS

This work is supported in part by an NIH grant 1R41CA199735-01A1. The internship of Peirui Bai at the Medical Image Processing Group was supported by the National Natural Science Foundation of China under Grant

No. 61471225. The authors are grateful to Mr. Dewey Odhner for help with the use and operations of the CAVASS software system.

CONFLICT OF INTERESTS

The authors have no conflicts to disclose.

[#]Co-first authorship and contributed equally.

*Corresponding author who conceived the whole idea and did most of the writing.

^{a)}Author to whom correspondence should be addressed. Electronic mail: jay@pennmedicine.upenn.edu; Telephone: 215-746-8627; Fax: 215-349-8426.

REFERENCES

- Tong Y, Udupa JK, Torigian DA, et al. Chest fat quantification via CT based on standardized anatomy space in adult lung transplant candidates. *PLoS ONE*. 2017;12:e0168932.
- Tong Y, Udupa JK, Torigian DA. Optimization of abdominal fat quantification on CT imaging through use of standardized anatomic space-a novel approach. *Med Phys.* 2014;41:063501–063511.
- Tong YB, Udupa JK, Wu CY, et al. Fat segmentation on chest CT images via fuzzy models. In: Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling, Proceedings of SPIE, 9786, 978609-1:6; 2016.
- Udupa JK, Odhner D, Zhao L, et al. Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images. *Med Image Anal.* 2014;18:752–771.

- Tong Y, Udupa JK, Sin S, et al. MR image analytics to characterize the upper airway structure in obese children with obstructive sleep apnea syndrome. *PLoS ONE*. 2016;11:e0159327.
- Matsumoto MM, Udupa JK, Tong Y, Saboury B, Torigian DA. Quantitative normal thoracic anatomy at CT. *Comput Med Imaging Graph*. 2016;51:1–10.
- Bi L, Kim J, Feng D, Fulham M. Multi-stage thresholded region classification for whole-body PET-CT lymphoma studies. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, eds. *Medical Image Computing and Computer-Assisted Intervention-MICCAI*. Cham: Springer; 2014:569–576.
- Bai P, Udupa JK, Tong Y, Xie S, Torigian DA. Automatic thoracic body region localization. *Proc SPIE*. 2017;10134:101343X.
- Hussein S, Green A, Watane A, et al. Automatic segmentation and quantification of white and brown adipose tissues from PET/CT scans. *IEEE Trans Med Imaging*. 2017;36:734–744.
- Frantz S, Rohr K, Stiehl HS. Localization of 3D anatomical point landmarks in 3D tomographic images using deformable models. In: Delp SL, DiGoia AM, Jaramaz B, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000.* LNCS. Vol. 1935. Berlin: Springer; 2000:492–501.
- Wörz S, Rohr K. Localization of anatomical point landmarks in 3D medical images by fitting 3D parametric intensity models. *Med Image Anal.* 2006;10:41–58.
- 13. Yao C, Wada T, Shimizu A, Kobatake H, Nawano S. Simultaneous location detection of multi-organ by atlas-guided eigen-organ method

in volumetric medical images. Int J Comp Assist Radiol Surg. 2006;1:42-45.

- Criminisi A, Shotton J, Bucciarelli S. Decision forests with long-range spatial context for organ localization in CT volumes. MICCAI Workshop on probabilistic models for medical image analysis; 2009: 69–80.
- Criminisi A, Robertson D, Konukoglu E, et al. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Med Image Anal.* 2013;17:1293–1303.
- Chu C, Belavý DL, Armbrecht G, Bansmann M, Felsenberg D, Zheng G. Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method. *PLoS ONE*. 2015;10:1–22.
- Potesil V, Kadir T, Platsch G, Brady SM. Personalized graphical models for anatomical landmark localization in whole-body medical images. *Int J Comp Vis.* 2015;111:29–49.
- Gao S, Yu P. Atlas of Human Anatomy, 3rd edn. Shang Hai: Shanghai Science and Technology Press; 2000.
- Tong Y, Udupa JK, Odhner D, Bai P, Torigian DA. Virtual landmarks. Proc SPIE. 2017;10135:1013521.
- Hagan MT, Demuth HB, Beale MH, Jesus OD. *Neural Network Design*, 2nd ed. India: Thomson Press. http://hagan.okstate.edu/nnd.html. 23 August, 2016.
- Foresee FD, Hagan MT. Gauss-Newton approximation to Bayesian learning. *Int Conf Neural Netw.* 1997;3:1930–1935.
- 22. Beale MH, Hagan MT, Demuth HB. *Neural Network Toolbox[™] User's Guide- MATLAB R2015a*. Natick, MA: The MathWorks, Inc.; 2015.
- Yubing T, Jayaram KU. Virtual landmarks. 2019. http://www.mipg. upenn.edu/Vnews/VirtualLDMK/vldmkAnimaton.pptx. Accessed January 8, 2019.