Automatic thoracic body region localization

PeiRui Bai[a,b], Jayaram K Udupa[b], YuBing Tong[b], ShiPeng Xie[b,c], Drew A Torigian[b]
[a]College of Electronics, Communication and Physics, Shandong University of Science and Technology, Qingdao 266590, China;[b]Medical Image Processing Group, Goddard Building – 6th floor, 3710 Hamilton Walk, Department of Radiology, University of Pennsylvania, Philadelphia, PA , USA 19104; [c]College of Telecommunications & Information Engineering,Nanjing University of Posts andTelecommunications, Nanjing, Jiangsu 210003, China

ABSTRACT

Radiological imaging and image interpretation for clinical decision making are mostly specific to each body region such as head & neck, thorax, abdomen, pelvis, and extremities. For automating image analysis and consistency of results, standardizing definitions of body regions and the various anatomic objects, tissue regions, and zones in them becomes essential. Assuming that a standardized definition of body regions is available, a fundamental early step needed in automated image and object analytics is to automatically trim the given image stack into image volumes exactly satisfying the body region definition. This paper presents a solution to this problem based on the concept of virtual landmarks and evaluates it on whole-body positron emission tomography/computed tomography (PET/CT) scans. The method first selects a (set of) reference object(s), segments it (them) roughly, and identifies virtual landmarks for the object(s). The geometric relationship between these landmarks and the boundary locations of body regions in the cranio-caudal direction is then learned through a neural network regressor, and the locations are predicted. Based on low-dose unenhanced CT images of 180 near whole-body PET/CT scans (which includes 34 whole-body PET/CT scans), the mean localization error for the boundaries of superior of thorax (*TS*) and inferior of thorax (*TI*), expressed as number of slices (slice spacing ≈ 4mm)), and using either the skeleton or the pleural spaces as reference objects, is found to be 3,2 (using skeleton) and 3, 5 (using pleural spaces) respectively, or in mm 13, 10 mm (using skeleton) and 10.5, 20 mm (using pleural spaces), respectively. Improvements of this performance via optimal selection of objects and virtual landmarks and other object analytics applications are currently being pursued.

and the skeleton and pleural spaces used as a reference objects

Keywords: body region identification, virtual landmarks, principal component analysis, positron emission tomography (PET)/computed tomography (CT), neural network learning regression.

1. INTRODUCTION

Radiological imaging and image interpretation and the utilization of images for clinical decision making are mostly specific to each body region such as head, neck, thorax, abdomen, pelvis, and extremities. Although imaging is often performed body-region-wise, the acquired images do not usually conform to any standardized definition of body regions. For automating image analysis and consistency of results, standardizing definitions of body regions and the various anatomic objects, tissue regions, and zones in them becomes essential [1]. Even when imaging is performed body-wide, like in whole-body PET/CT images, it becomes necessary to perform analysis body-region-wise in numerous applications, particularly those involving systemic diseases, for disease staging, treatment planning, response assessment, and restaging. Although methods have been reported in the literature for automatically locating regions of interest based on physical anatomic landmarks observable on images, the regions of interest considered have been mainly whole organs. We are not aware of any published papers that focused on the problem of automatically defining entire body regions. This paper presents a solution to this problem, employing not physical landmarks but conceptual or virtual landmarks that may not correspond to any physically observable anatomic feature on the image.

Assuming that a standardized definition of body regions is available, one of the fundamental early steps needed in automated image analysis, image analytics, and object analytics is to automatically trim the given tomographic image stack of a body region into an image volume exactly satisfying the body region definition. If a whole-body tomographic image stack is given, the task is to precisely divide the stack into body regions as per the exact definition of each body

Medical Imaging 2017: Computer-Aided Diagnosis, edited by Samuel G. Armato III, Nicholas A. Petrick,
Proc. of SPIE Vol. 10134, 101343X · © 2017 SPIE · CCC code: 1605-7422/17/$18 · doi: 10.1117/12.2254862

region. Although many approaches to automatic identification of objects, landmarks, and specific anatomic features are available in the literature [1-6], this important and fundamental problem does not seem to have been addressed. This paper presents a solution to this problem and evaluates it on whole-body or near whole-body PET/CT images. The main contributions of this work can be summarized as follows: (1) An automated whole-body region localization method. (2) The idea of a virtual landmark can take into account both object shape and intensity appearance. (3) The method can be generalized to other imaging modalities and reference objects easily.

2. METHODS

For initial demonstration, we will focus on one body region in this paper, namely the thorax, and follow its definition formulated in [1,2]. We define the thoracic body region as extending 5 mm superior to the lung apices to 15 mm inferior to the lung bases. In any given image, I, we will denote the location of the superior axial slice of the thorax by $TS(I)$ and the location of the inferior axial slice of the thorax by $TI(I)$. Locations in all images are specified with reference to some fixed coordinate system, such as the scanner coordinate system. Our problem is: Given any image I, find automatically locations $TS(I)$ and $TI(I)$ in I. We assume that the slices of I are organized axially and that the body region it encompasses properly includes the body regions to be identified in I. In this work, we deal with whole-body or near whole-body PET/CT imagery, and so, 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, but will be in correct relationship with that data set and the subject in the scanner coordinate system.

The proposed method for body-region localization works overall as follows. It uses one or more objects that can be easily segmented in I as reference objects and finds in them *virtual landmarks*. It then trains a neural network to learn the relationship between these landmarks and the known true locations for $TS(I)$ and $TI(I)$ which are obtained from a training set of images. Subsequently the trained network is used to predict the body region boundary locations in any given image. These found locations are then used to subdivide the given image into body regions. These steps are further described below.

(a) *Image data*:We selected whole-body CT image datasets from existing PET/CT scans of subjects from our health system patient image database 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. The subjects whose image data are selected included near-normal cases and patients with different types of disease conditions. Among these, 34 were whole-body scans (head to foot) and 180 were near whole-body (neck to foot). They all included fully the body region of interest, namely the thorax. In this study, we used only the low-dose unenhanced CT portion of the PET/CT data sets. However, our approach is applicable to just PET alone and both CT and PET used simultaneously. These investigations will be considered in the future. The voxel size in the CT data sets is roughly spread from 1mm×1mm×3 to 1mm×1mm×5 mm for CT-scans of different subjects.

(b) *Reference objects*:In this work, we selected the left and right pleural spaces together as one reference object and the skeletal structure of the entire body as another object of reference. Segmentation of the objects was performed using the algorithms implemented in the CAVASS software system [7]. The segmentation results are stored as binary images. The requirements for a reference object are: it should be fully contained within the data set, it should be easily segmented automatically, and it should not be clustered within some small space in the body. Thus, for the skeleton reference object, only 34 data sets were available, since the other data sets do not satisfy the first condition. However, for the pleural spaces reference object, all 180 data sets were utilized.

(c) *Extraction of virtual landmarks*: A separate paper is presented at this conference that describes the concept and techniques underlying the idea of virtual landmarks[8]. Briefly, virtual landmarks associated with an anatomic object are reference points which are tethered to the object. The points may lie anywhere within the body with respect to the object – inside, outside, or on its boundary, and most likely they do not lie exactly on the boundary. They can be defined on the binary image representing the object or using both object shape and gray value appearance from an underlying image at voxels that belong to the object. The landmarks are obtained through a process of recursive subdivision of the object guided by principal component analysis (PCA). At the highest level, the geometric center of the object is the only landmark produced. The eigenvectors associated with the object define a principal axis system and divide the object into 8 octants. The part of the object in each of these octants is again subjected to PCA which yields a geometric center and a principal axis system. At the second level, thus, 8 landmarks are generated. In the third level, continuing this process of subdivision, 64 landmarks are generated, and so on. The method guarantees correspondence among homologous points in different samples of the object from different subjects automatically, as is clear from the process of generating the points. It is also clear that the method allows selecting any desired virtual landmarks and any number of them since each landmark has a unique identifier associated with it in the process of subdivision. In this work, we used different numbers of landmarks as described below. The total number of virtual landmarks generated in *n* levels for a 3D object is $\sum_{i=1}^{i=n} 2^{3(i-1)}$. For *n*=2, 3, and 4 levels, the number of virtual landmarks generated is 9, 73, and 585, respectively.

(d) *Labeling true locations of body region boundaries*: The actual locations in the cranio-caudal direction of the boundaries of the body region, or the values of *TS*(*I*) and *TI*(*I*), for all images in our data sets, are obtained by a visual examination of each data set, identifying the slices that define the boundaries of each body region, and recording the physical location of the slices in the scanner coordinate system using the CAVASS software. Note that there is potential subjective variation in this specification by readers which is due to how the body region boundary definition is interpreted by each reader on a specific image. Typically this variation is within one slice.

(e) *Learning the relationship between virtual landmarks and true locations of body-region boundaries*: A neural network is configured as a regressor by feeding virtual landmarks and true locations as input and target output data sets, respectively, and the relationship is learned through network training, validation, and testing procedures. Here we adopt a simple architecture of a multiple-layer perception with one hidden layer. The number of neurons in the input layer is the same as the dimension of virtual landmarks, and the number of neurons in the output layer is the same as the size of target locations vector. In addition, the numbers of neurons in the hidden layers are determined by the complexities of the nonlinear mapping function that is being approximated. The details of the neural network configuring and training will be presented in the next section. Then, the trained network is used to predict locations *TS*(*I*) and *TI*(*I*) for any test image *I*. Since true locations have been recorded for all data sets, we can evaluate the prediction error by comparing the predicted with the true locations.

3. EXPERIMENTS AND RESULTS

To demonstrate the geometric relationship between the reference object and its virtual landmarks, in Figures 1 and 2 we display the 3D surfaces of the skeleton and the pleural space derived from whole-body PET/CT datasets of two patients along with the set of 73 virtual landmarks derived for this object from three levels (*n*=3). Notice that the 73 virtual landmarks spread quite uniformly in and around the object. It should be noted that some landmarks fall outside the object in the case of the skeleton reference object.

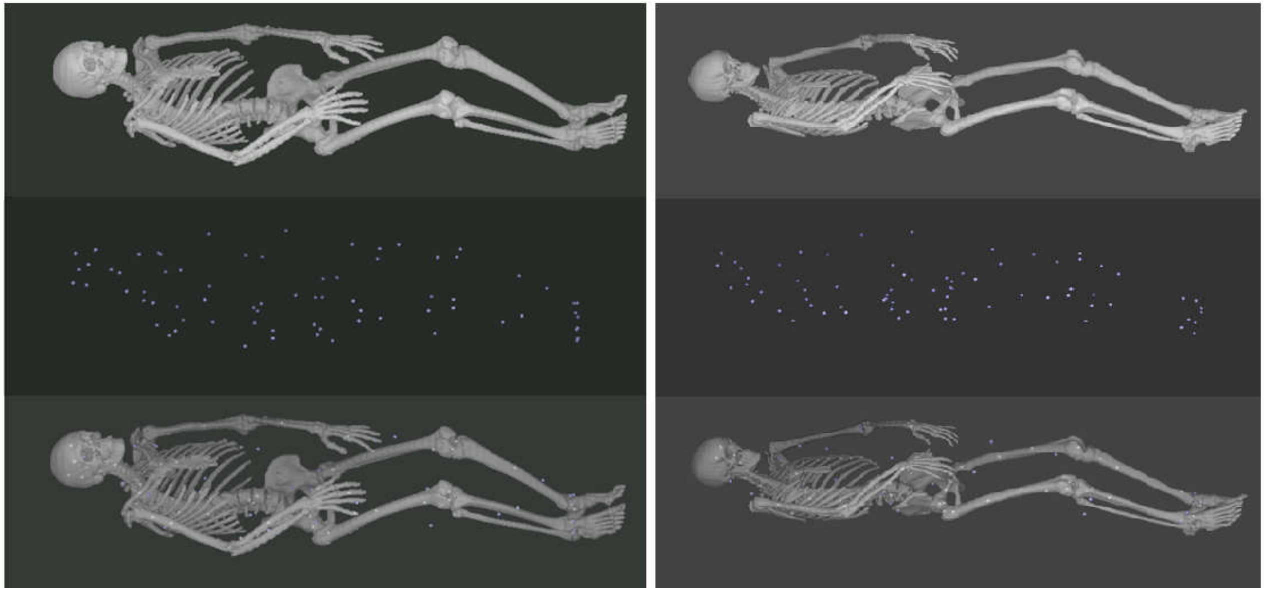

Figure 1. 3D renditions of the skeleton representing one reference object derived from two patient data sets along with the derived virtual landmarks are shown in different forms: The object by itself, only the landmarks, and the two together.

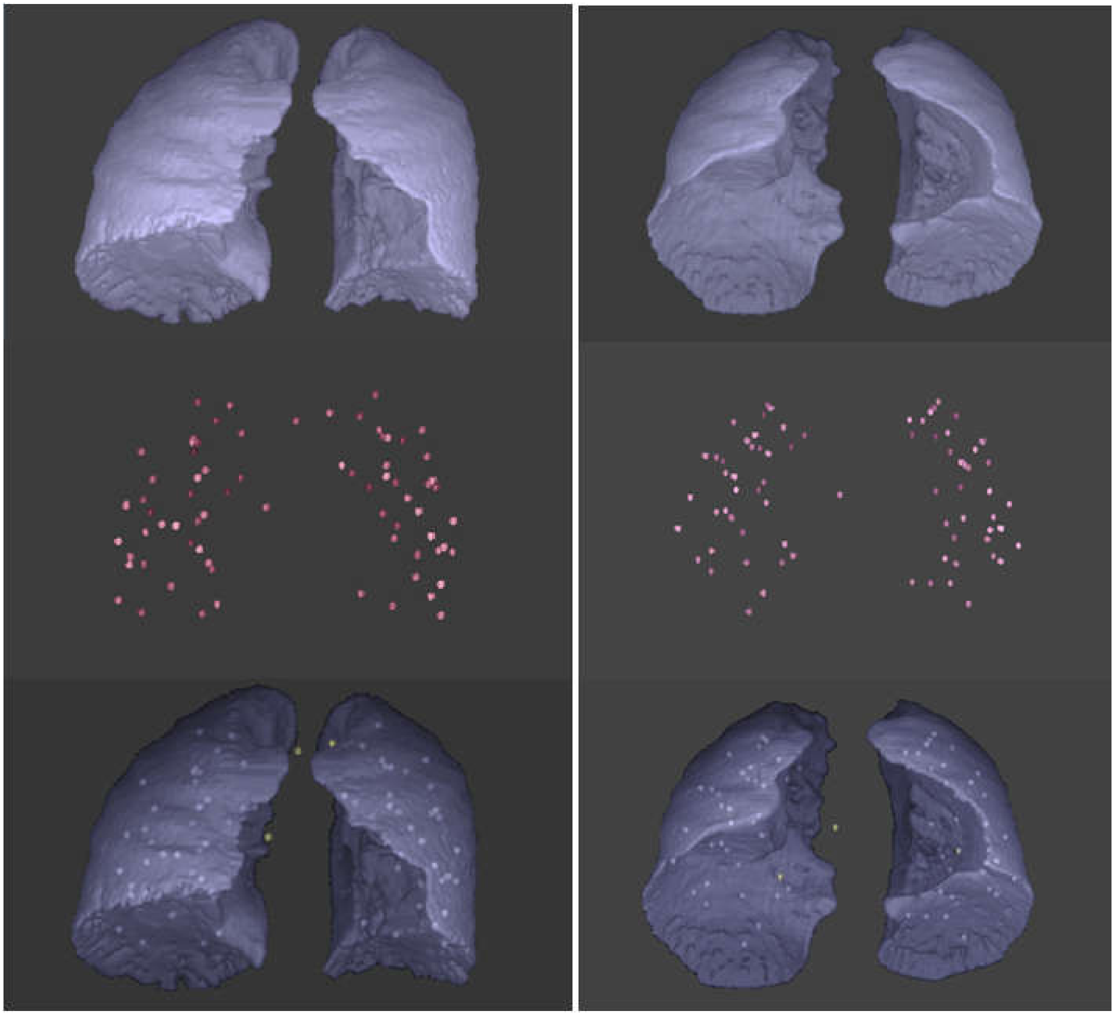

Figure 2. 3D renditions of the pleural space representing one reference object derived from two patient data sets along with the derived virtual landmarks are shown in different forms: The object by itself, only the landmarks, and the two together.

The landmark data with dimension as $N \times D \times S$ are set as input data to the neural network, where N represents the number of landmarks, D represents 3D spatial coordinates (x, y, z) of the landmark points, and S represents number of subjects. Thus, if all data were to be utilized, the dimensionality of input data will be 73×3×34 for the skeleton reference object and 73×3×180 for pleural spaces reference object. The dimensionality of the target output data is $L \times S$, where L represents the number of body-region boundary locations, in our case $L = 2$ (it corresponds to *TS* and *TI*).

Here, we design the dimensionality of input data in terms of the coordinates of a subset of the virtual landmarks generated for a given number of levels (we chose $L = 3$). For the skeleton reference object, due to the small number of data sets available, we used fewer landmarks instead of all possible landmarks generated for 3 levels. For example, to form a subset of 9 landmark points, we select 1 point from the first level and all 8 points from the second level. For a subset which contains 25 points, we select 1 point from the first level, all 8 points from the second level, and 16 points from the third level by choosing 16 designated octants out of the total 64 octants. In addition to varying L, we also tested a strategy, where, instead of using the (x,y,z) coordinates of points, only the z coordinate is considered for determining the input variables.

After configuring the input data and target data of the neural network regressor, three things such as the data dividing, the network architecture, and the training algorithm should be considered carefully. Thanks for the neural network toolbox of MATLAB (it provides a powerful framework), it is convenient to implement the above operation[9]. For both case of using skeleton as reference object and using pleural spaces as reference object, we choose the Bayesian

Regularization algorithm (BR) to implement the training process because it can prevent overfitting and provide better performance than the Levenberg-Marquardt algorithm. As the BR training algorithm is used, the validation set can be merged with the training set. So, we only need to divide the number of subjects (*S*) into two sets i.e. training set and testing set. Here, we adopt the default setting of the toolbox which means that the training set make up approximately 85% of the full data set, and the testing set make up 15% of the full data set. That is, the ratio of the training set and the testing set are 29/5 and 153/27 for the two cases. A multilayer perception with a single hidden layer is chosen to be the neural network architecture, and different neuron numbers of hidden layers are selected in terms of empirical tests (which are listed in Table 1 for different coordinates of landmarks and dimension of input data) for the two cases. The performance index is mean square error (MSE) .

Table 1 presents the localization error for *TS* and *TI* in units of Number of Slices (NoS) and millimeters (mm). The mean localization error in NoS for *TS* and *TI* is 3 and 2, respectively, when the skeleton is used as a reference object, and 3 and 5, respectively, when the pleural spaces is used as a reference object. The corresponding location errors in mm are 13 and 10, respectively, for skeleton and 10.5 and 20, respectively, for pleural spaces. Note that the total number of slices in a whole-body PET/CT scan is roughly 400 to 500 slices. A localization error of less than 3 in NoS is almost perfect keeping in mind the applications for which body region identification is automated and also considering the fact that the variation in human identification of the region boundary locations can be 1-2 slices. A location error of less than 5 NoS seems acceptable. It is noted that the localization error for thorax presented in this paper are more attractive when comparing to the more recent report in reference [10] with body region localization error of 47.01mm.

Table 1. Localization error for *TS* and *TI* using skeleton and pleural spaces as reference objects.

Reference object	Coordinates of landmarks	Dimensionality of input data	Neuron number of hidden layers	Localization error (NoS, mm)	
				TS	*TI*
Skeleton	(x,y,z)	$25\times3\times S$	3	3, 12	1, 12
		$9\times3\times S$	2	3, 12	3, 12
	z	$25\times S$	1	4, 16	2, 8
		$9\times S$	2	3, 12	2, 8
Pleural spaces	(x,y,z)	$73\times3\times S$	10	4, 15.3	9, 34.4
		$25\times3\times S$	10	2, 7.6	5, 19.1
	z	$73\times S$	3	3, 11.5	4, 15.3
		$25\times S$	8	2, 7.6	3, 11.5

From Table1, we can learn several interesting aspects of the present method. Firstly, different localization accuracy can be obtained by employing different reference objects. The localizing performance by using the skeleton as a reference object is better than that by using the pleural spaces. This may be due to the higher complexity of the shape of the skeleton and its body-wide spatial extent, which may yield more powerful virtual landmarks. Second, the localization accuracy using only *z* coordinates seems better than that of using (x,y,z) coordinates, particularly for *TI* when using pleural spaces as the reference object. Third, a larger number of virtual landmarks does not necessarily imply higher accuracy in localization. Thus, finding landmarks which have higher information content may be important.

4. CONCLUSIONS

In this work, a new method to localize the thoracic body region from whole-body or near whole-body PET/CT scans is presented. The key idea is to find a nonlinear mapping relationship between the 3D virtual landmarks and the ground-truth of body-region boundaries. By searching key points along the shape of reference objects with a multiple levels style, the obtained virtual landmarks can represent a large object effectively. The relationship between the virtual landmarks and ground-truth positions can then be learned by employing a simple neural network regressor. Experimental results confirm that the present method can localize body-region boundaries automatically and reliably with promising accuracy. It is interesting that the accuracy of localizing specific body-region boundaries seems to be dependent on the reference object(s) and the particular virtual landmarks employed. In the future, we will investigate the performance of virtual landmarks taking into account both object shape and appearance in their definition.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (under Grant No. 61471225). We are grateful to Dewey Odhner for his help with the CAVASS software system and the programs for identifying virtual landmarks. This work is supported by an NIH grant 1R41CA199735-01A1.

REFERENCE

[1] Udupa J K, Odhner D, Liming Z, *et al*. "Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images". Medical Image Analysis, 18(5),752-771(2014).

[2] Wang H, Udupa J K, Odhner D, *et al*. "Automatic anatomy recognition in whole-body PET/CT images". Medical Physics,43(1),613-629(2016).

[3] Stefan W, Karl R. "Localization of anatomical point landmarks in 3D medical images by fitting 3D parametric intensity models". Medical Image Analysis, 10(1), 41–58(2006).

[4] Vaclav P, Timor K, Günther P, *et al*. "Personalized Graphical Models for Anatomical Landmark Localization in Whole-Body Medical Images", International Journal of Computer Vision, 111(1),29-49(2015).

[5] Criminisi A, Shotton J, Bucciarelli S. "Decision forests with long-range spatial context for organ localization in CT volumes". In MICCAI Workshop on Probabilistic Models for Medical Image Analysis(2009).

[6] Criminisi A, Shotton J, Robertson D, *et al*. "Regression forests for efficient anatomy detection and localization in CT studies". In Medical Computer Vision. Recognition Technique and Applications in Medical Imaging, 106–117(2011).

[7] Grevera, G., Udupa, J.K.,Odhner, D., Zhuge, Y., Souza, A., Iwanaga, T. and Mishra, S. "CAVASS: A computer-assisted visualization and analysis software system", Journal of Digital Imaging, 20(Supplement 1),101–118(2007).

[8] Yubing Tong, Jayaram K Udupa, Dewey Odhner, Peirui Bai, Drew A. Torigian. "Iterative PCA based virtual landmark: a novel automatic landmark identification approach", SPIE Medical Imaging (2017)(Accepted).

[9] Martin T. Hagan, Howard B. Demuth, Mark Hudson Beale, *et al*. [Neural Network Design], 2nd Edition, eBook.(2014).

[10] Sarfaraz Hussein, Aileen Green, Arjun Watane, David Reiter, Xinjian Chen, *et al*. "Automatic segmentation and quantification of white and brown adipose tissues from PET/CT scans", IEEE Transactions on Medical Imaging(2016)(to be published).