Hierarchical Pictorial Structures for Simultaneously Localizing Multiple Organs in Volumetric Pre-Scan CT

Albert Montillo[1*], Qi Song[1*], Bipul Das[2], Zhye Yin[1]

GE Global Research, [1]Niskayuna, NY, USA and [2] Bangalore, India
[*]authors contributed equally

ABSTRACT

Parsing volumetric computed tomography (CT) into 10 or more salient organs simultaneously is a challenging task with many applications such as personalized scan planning and dose reporting. In the clinic, pre-scan data can come in the form of very low dose volumes acquired just prior to the primary scan or from an existing primary scan. To localize organs in such diverse data, we propose a new learning based framework that we call hierarchical pictorial structures (HPS) which builds multiple levels of models in a tree-like hierarchy that mirrors the natural decomposition of human anatomy from gross structures to finer structures. Each node of our hierarchical model learns (1) the local appearance and shape of structures, and (2) a generative global model that learns probabilistic, structural arrangement. Our main contribution is two fold. First we embed the pictorial structures approach in a hierarchical framework which reduces test time image interpretation and allows for the incorporation of additional geometric constraints that robustly guide model fitting in the presence of noise. Second we guide our HPS framework with the probabilistic cost maps extracted using random decision forests using volumetric 3D HOG features which makes our model fast to train and fast to apply to novel test data and posses a high degree of invariance to shape distortion and imaging artifacts. All steps require approximate 3 mins to compute and all organs are located with suitably high accuracy for our clinical applications such as personalized scan planning for radiation dose reduction. We assess our method using a database of volumetric CT scans from 81 subjects with widely varying age and pathology and with simulated ultra low dose cadaver pre-scan data.

Keywords: hierarchical pictorial structures, probabilistic decision forest, 3D HOG image descriptor

1. INTRODUCTION

Extracting bounding boxes of the salient organs in a patient's radiological scan provides important information that can be used to guide many subsequent applications. The localization information can also be used to guide finer scale segmentation of each organs boundary surface by constraining the segmentation to each localized ROI. In large volumetric scans, bounding box information can aid in the visualization of just the anatomical structures of interest which can simplify workflow and interpretation by a radiologist. In addition, and a focus of ours, in the case of computed tomography (CT), organ localization information can be used to guide personalized scan planning and dose reporting. Other methods have been developed for organ localization using 2-D scout images such as X-Ray radiographs,[1,2] however 3D image data is necessary to disambiguate the extents of abdomen organs (liver, spleen, kidneys, etc). In the clinic, the volumetric pre-scan data can come in the form of very low dose volumes acquired just prior to the primary scan or from an existing primary scan.

Some 3-D automated organ localization methods have been developed in recent years. Wolz et al.[3] proposed an atlas-based method for abdominal multi-organ segmentation in CT. Both multi-atlas registration and patch-based segmentation techniques were combined in a single framework. Criminisi *et al.* proposed regression forest based methods for anatomical la dmark detecit on using local context features.[4] The methods were validated

Further author information: (Send correspondence to Z.Y.)
A.M.: E-mail: montillo@emailtreo.com
Q.S.: E-mail: songq03@gmail.com
B.D.: E-mail: bipul.das@ge.com
Z.Y.: E-mail: yin@research.ge.com

Medical Imaging 2015: Image Processing, edited by Sébastien Ourselin, Martin A. Styner

on 100 CT scans. Erdt *et al.*[5] developed a learning-based region detection approach using HOG features for 3-D organ localization in standard-dose CT. The detected results were further refined using a PCA-based shape model. None of above methods have been applied on low dose CT scan.

In this work, we propose a new learning based framework that we call hierarchical pictorial structures (HPS) , which builds multiple levels of models in a tree-like hierarchy. We apply it to localize multiple (11) organs simultaneously and accurately in a database of 51 volumetric CT scans and demonstrate its suitability to very low dose pre-scan acquisition technology which we are developing separately.

2. METHODS

Our organ localization approach consists of two parts. First, single rectangular region of each organ is detected independently using 3-D HOG based features, which results in a probability map of candidate locations for the associated organ. Second, a tree-shaped pictorial structure model encoding context information between certain pairs of organs is constructed. The localization of multiple organs is then performed in a coarse to fine strategy using dynamic programming approach. The following sections describe each step in details.

2.1 3-D HOG features

HOG features are frequently used in computer vision area with applications including object detection,[5,6] video tracking,[7] image retrieval,[8] etc. The main idea is to count distribution of gradient orientation in local patches of the image, which describes both local appearance and shape information of target object.[6] Compared to other feature descriptors, HOG features are invariant to local deformation and scales, and are therefore well suited for organ detection tasks where substantial inter-patient variances can exist. Note that HOG features are not rotation invariant beyond the bin angle width. However, the general scan orientation of patients are usually known in medical imaging. Thus we can roughly rotationally align the data before computing HOG features.

To compute 3-D HOG features, a 3-D patch of the target area is obtained from input volumetric images. Then the patch is divided into $3 \times 3 \times 3$ equally sized sub-cells. In each sub-cell a weighted 2-D histograms of gradient orientation is computed. Specifically, a 3-D gradient vector $v = (r, \theta, \varphi)$ is computed for each voxel inside the cell, where r is the magnitude of the gradient; θ and φ are the polar angle and the azimuthal angle of the gradient vector, respectively. Then each voxel casts a magnitude weighted vote in the 2-D orientation-based histogram $H(\theta, \varphi)$. Fig. 2a illustrates our 3-D HOG computation. For time efficiency, an integral histogram image is computed in a Cartesian space as described in,[9] which greatly helps to reduce HOG computation time.

The computed features are then used to train individual classifiers for single organ detection. A 3-D ground truth bounding-box is manually labeled for each target organ on the training image, from which 3-D HOG features are computed as positive samples. Negative samples are generated by randomly sampling over the image space with the constraint that it does not overlap the manually-labeled positive bounding-box. A decision forest[10] classifier is then trained for each of target organs individually. The detection is performed on novel images by scanning over the volume at candidate locations $l(x, y, z)$ with different scales of the image patches to capture the scale variance of target organs. The computed HOG features are then input to the classifier, which outputs a matching score $m_i(l, s)$ for organ o_i normalized in the range of [0-1], representing the probability that the given image patches centered at $l(x, y, z)$ with scale s matches the training models for organ o_i. Fig. 2a shows examples of the computed score with optimal scale for left lung, heart and right kidney, respectively. The classifier generated a strong response in the center area of the target organs.

2.2 Pictorial structure model

Our proposed approach is inspired by the pictorial structure model for object recognition.[11] The basic idea is to describe an object as a collection of parts with a connection between selected pairs of parts. It can be expressed as an undirected graph $G = (V, E)$, where set of vertices $V = \{v_1, \ldots, v_N\}$ corresponds to the N parts of the object, while each edge $(v_i, v_j) \in E$ defines the connectivity relationship between part v_i and v_j. If we let

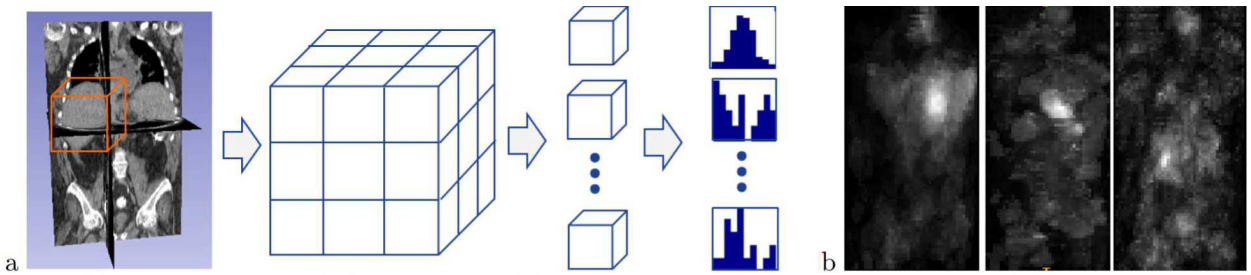

Figure 1. (a) The steps of 3-D HOG computation. (b) Example of computed matching score of HOG features with optimal scale on one coronal slice for left lung (left panel), heart (middle panel) and right kidney (right panel). The scores are normalized in the range of [0-1], with white representing 1 and black representing 0. Note that the classifier generated a strong response in the center area of the target organs.

$L = (l_1, \ldots, l_N)$ define the locations of N parts in a given image I, then the object recognition problem can be formulated as an optimization problem:

$$L^* = \arg\min_L \left(\sum_{i=1}^{n} c_i(l_i) + \sum_{(v_i,v_j)\in E} d_{ij}(l_i, l_j)\right) \tag{1}$$

where $c_i(l_i)$ is an image matching cost inversely related to the likelihood that part i is located at l_i. $d_{ij}(l_i, l_j)$ encodes the pairwise geometric contextual relationship between part i and j. Specifically, it measures the deformation cost of the model when placing part i at location l_i and placing j at location l_j. As described in,[11] a globally optimal solution can be achieved in linear time to the number of possible discrete space when the model is restricted to a tree structure.

In our proposed approach, we make use of similar idea for organ localization. Each organ is viewed as a graph node in the pictorial structures model. The context information between pairs of organs can be captured by adding weighted edges between nodes. Fig. 2a is illustrative of the structure of our context model. A statistical model is used for both matching cost and the deformation cost. Here $c_i(l_i)$ at location l_i is designed based on the matching score obtained from the 3-D Hog Feature. To enforce context constraints between pairs of organs, we identify one of the organs as the root organ o_r. Then a probabilistic model of the relative location of the other organs to the root organ is built from the training data. Here we assume a Gaussian distribution model $N_{ij}(s_{ij}, \Sigma_{ij})$ on the distance vector between o_i and o_j, where s_{ij} is the mean distance vector and Σ_{ij} is the associated covariance matrix, then the deformation cost is given by $d_{ij}(l_i, l_j) = N_{ij}(l_i - l_j, s_{ij}, \Sigma_{ij})$, measuring the deformation of current distance vector $l_i - l_j$ with respect to the trained context model. A globally optimal solution is achieved using a generalized distance transform under a Mahalanobis distance with diagonal covariance matrix, which leads to a linear time solution.[11,12] Once we compute the best location of l_i^* for organ o_i, the bounding box can be further obtained by centering the image patch at l_i^* with optimal scale obtained from 3-D HOG feature computation.

2.3 Hierarchical Organ Localization

We were inspired to build a model composed of submodels because human anatomy has a natural hierarchical decomposition. In HPS, each submodel has one root and one or more leaves. The first level has one model while subsequent levels have one or more submodels. Fig. 2b illustrates our two-level model. Level 1 uses spine as root and has head, pelvis, and chest/abdominal organs as leaves. The head is also the root for a level 2 submodel which has the eyes (dual) as a leaf, while the pelvis is the root for a level 2 submodel which has the testes as a leaf (gender is generally available in the dicom header). Organs within a level provide context for one another, and root organs in level 1 provide context for the leaves of level 2. To train appearance models for level 2, negative examples are drawn from search windows (yellow boxes in Fig 2c) around root organs of level 1. Each submodel's local organ appearances and global context model is *trained* separately using these search window specifications and ground truth annotations. To apply the trained model, the model fitting is iterated successively from the top level to the lower level. Here we first compute the probability maps for the organs at

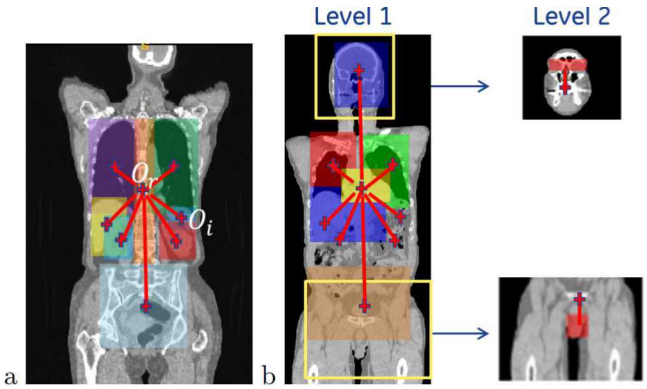

Figure 2. (a) Pictorial structure model. The red line indicates the connectivity relationship between root organ o_r and remaining organs o_i. (b) Multi level hierarchical model. Level 1 organs include spine as root organ and head, lungs, heart, liver, spleen, kidneys, pelvis as leaves. Level 2 models include (top) head as root with dual eye model as leaf, and (bottom) pelvis as root with testes as leaf.

level 1 and fit its context model, then we use the organ localization results from level 1 to form the root organ locations of the submodels in level 2 and fit them accordingly. Note that here we only need to search in a much smaller region around the detected root organ locations from level 1, which significantly reduces the computation time of model fitting. Also, multi-resolution organ localization is allowed since we can use larger resolution for large organs (lung, liver, etc) in level 1 and smaller resolution for eyes and testes in level 2.

3. EXPERIMENTS AND RESULTS

In our first experiment, volumetric CT scans from a database of 81 subjects acquired from PET-CT scan were used to simulate the clinical situation where subjects have a previous (longitudinal) acquisition that can be used to help guide subsequent scan planning. The image resolution ranges from $0.98 \times 0.98 \times 3.27\ mm^3$ to $1.37 \times 1.37 \times 3.75\ mm^3$. Note that CT images in wholebody PET-CT scan usually have large slice thickness and low contrast. Hence the organ localization in PET-CT is more challenging than ordinary CT. Ground-truth bounding-boxes for lungs, liver, heart, spleen, kidneys, head, pelvis, eyes and testes were manually identified on all 81 subjects. We randomly selected 30 subject for training and the remaining 51 were used for testing. Fig. 3a shows examples of the accurate localization that is attained using our method for larger organs, while Fig. 3b,c shows the highly accurate localization for smaller organs such as the eyes and testes. For quantitative evaluation, the unsigned distance between the computed box sides and the ground-truth box sides were computed. The mean errors for each organ across the 51 test subjects are shown in Fig. 3d. All organs have fairly low errors (less than 20mm) and meet our clinical purpose of personalized scan plan optimization for a CT scanner.

In our second experiment, we test the applicability of our method trained on the 30 subjects in the previous experiment to generalize to an extremely low volumetric pre-scan, with dose equivalent to a traditional 2D x-ray radiograph. Due to ethical (e.g. IRB) concerns, we simulate such a volumetric pre-scan by: (1) scanning a human cadaver at $CTDI_{vol}$ of 0.9mGy which is less than 10% of the dose of a standard volumetric CT, and then reduce the dose another 10x to a $CTDI_{vol}$ of 0.09mGy by using only every 10th view when performing a sparse filtered backprojection. To reduce noise we apply noise reduction in projection domain, such as detector blur as described by Yin et al,[13] and Gaussian smoothing in the image domain. *Despite the fact that the simulated pre-scan volume has a dose of just 0.09mGy*, and the resulting volume is of lower quality than the standard dose data, our method still achieves high quality localization of the cadaver's organs. Fig. 3d shows the visually accurate organ localizations for the lungs heart and pelvis (left panel, axial), and spine, liver (middle panel) and the kidneys (right panel). This ability to generalize well to never seen before data is promising for our future clinical applications of our approach.

In addition to high accuracy and generalization, our approach also enjoys high computational efficiency. Our current C++, ITK based implementation requires roughly 3 min to run on a standard 4 core workstation. We achieve efficiency in several ways including the use of integral histogram image which greatly helps reduce

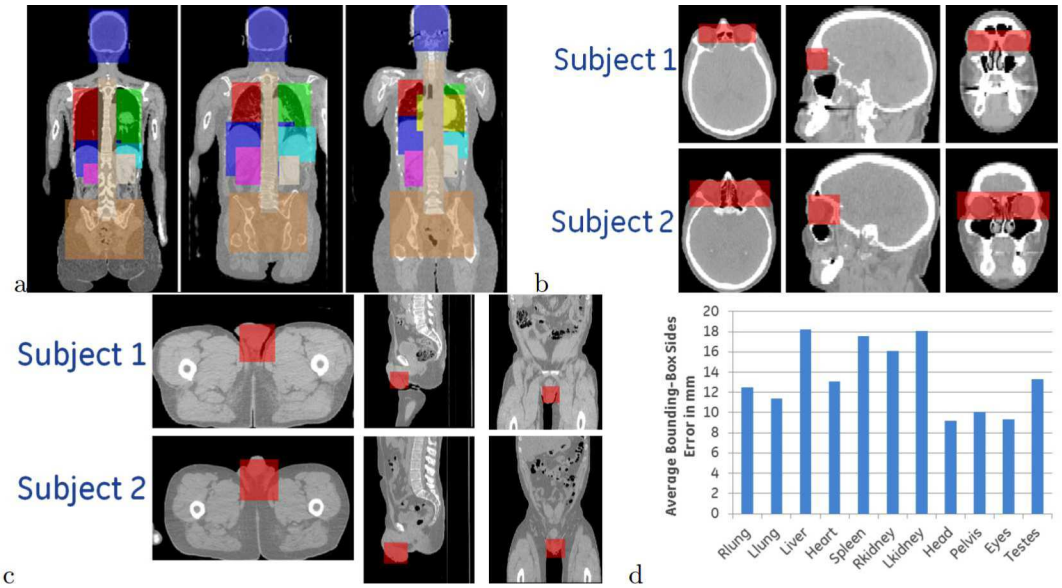

Figure 3. Automatically localized organs on standard dose volumetric CT using 2-level model. (a) Level 1 organ localization in a young subject (left), middle aged male (middle) and female (right). Level 2 localization results including (b) eye detection results on two additional subjects, and (c) testes for two more subjects. (d) Average mm error of bounding box sides for each of the 11 organs across 51 test subjects.

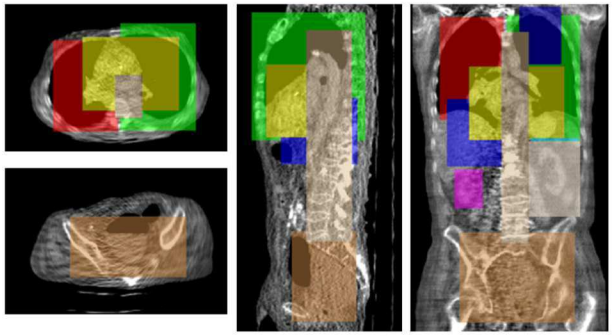

Figure 4. Automatically localized organs on extremely low dose (1%) volumetric pre-scan scout acquisition. Left panels axial views showing lungs heart (top) and pelvis (bottom). Middle panel sagittal view also shows spine (tan)) and liver (blue), while right panel shows kidneys (purple, beige).

HOG computation time. Currently most remaining computation time is spent in the application of the organ probability map computation, however even this can be fully parallelized in future implementation, reducing total run time to roughly 20 seconds.

4. DISCUSSION AND CONCLUSIONS

Hierarchical pictorial structures afford three primary advantages. (1) The HPS model captures both local appearance and shape information of organs as well as global context constraints between organs in a coarse to fine manner. (2) The contextual model can have tighter probabilistic distributions than single level model because the hierarchical model allows the root to be more proximal to leaves. For example the eyes have low variance distance with respect to head center compared to spine center due to head orientation. (3) The hierarchical model allows for an additional form of geometry based constraints (relative search windows) that can present fewer false positives in cost maps to the model fitting step.

We have also shown that HPS simultaneously localizes a set of 11 organs across a large database of 51 subjects and promising results on extremely low dose (0.09mGy dose) cadaver pre-scan simulations. HPS is fully automated, runs in tens of seconds and the accuracy we achieve is suitable for important clinical applications

such as radition dose reduction through personalized scan planning, particularly when used on low dose pre-scan data acquired immediately prior to the actual CT scan. We note that to reuse a longitudinal scan a fairly straight forward, rigid registration would also be required, however consistent patient positioning on the scanner's table (e.g. with a physical guide) could be used to suppress much of this variation, leaving just the organ localization challenge solved by our method. We look forward to integrating HPS with personalized scan planning and multi-modal quantification (e.g. PET/CT).

REFERENCES

[1] Q. Song, V. Srikrishnan, B. Das, and R. Bhagalia, "Cardiac localization in topograms using hierarchical models," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pp. 105–108, IEEE, 2013.

[2] Q. Song, A. Montillo, R. Bhagalia, and V. Srikrishnan, "Organ localization using joint ap/lat view landmark consensus detection and hierarchical active appearance models," in *Medical Computer Vision. Large Data in Medical Imaging*, pp. 138–147, Springer International Publishing, 2014.

[3] R. Wolz, C. Chu, K. Misawa, K. Mori, and D. Rueckert, "Multi-organ abdominal ct segmentation using hierarchically weighted subject-specific atlases," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 10–17, Springer, 2012.

[4] A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu, "Regression forests for efficient anatomy detection and localization in ct studies," in *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, pp. 106–117, Springer, 2011.

[5] M. Erdt, O. Knapp, K. Drechsler, and S. Wesarg, "Region detection in medical images using hog classifiers and a body landmark network," in *SPIE Medical Imaging*, pp. 867004–867004, International Society for Optics and Photonics, 2013.

[6] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, **1**, pp. 886–893, IEEE, 2005.

[7] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Robust Tracking-by-detection Using a Detector Confidence Particle Filter," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 1515–1522, IEEE, 2009.

[8] T. Kobayashi, "BFO Meets HOG: Feature Extraction Based on Histograms of Oriented p.d.f. Gradients for Image Classification," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 747–754, IEEE, 2013.

[9] F. Porikli, "Integral histogram: A fast way to extract histograms in cartesian spaces," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, **1**, pp. 829–836, IEEE, 2005.

[10] L. Breiman, "Random forests," *Machine Learning* **45**(1), pp. 5–32, 2001.

[11] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision* **61**(1), pp. 55–79, 2005.

[12] P. Felzenszwalb and D. Huttenlocher, "Distance Transforms of Sampled Functions," tech. rep., Cornell University, 2004.

[13] Z. Yin, Y. Yao, A. Montillo, P. M. Edic, and B. De Man, "Feasibility study on ultra-low dose 3D scout of organ based CT scan planning," in *The Third International Conference on Image Formation in X-Ray Computed Tomography*, pp. 52–55, 2014.