Automatic organ localizations on 3D CT images by using majority-voting of multiple 2D detections based on local binary patterns and Haar-like features

Xiangrong Zhou*[a], Shoutarou Yamaguchi[a], Xinin Zhou[b], Huayue Chen[c] , Takeshi Hara[a], Ryujiro Yokoyama[d], Masayuki Kanematsu[d,e], and Hiroshi Fujita[a]

[a] Department of Intelligent Image Information, Division of Regeneration and Advanced Medical Sciences, Graduate School of Medicine, Gifu University, Gifu-shi, 501-1194 Japan
[b] School of Information Culture, Nagoya Bunri University, 365 Maeda, Inazawa-cho, Inazawa-shi, 492-8520 Japan
[c] Department of Anatomy, Division of Disease Control, Graduate School of Medicine, Gifu University, Gifu-shi, 501-1194 Japan
[d] Radiology Services, Gifu University Hospital, Gifu-shi, 501-1194 Japan
[e] Department of Radiology, Gifu University Hospital, Gifu-shi, 501-1194 Japan

ABSTRACT

This paper describes an approach to accomplish the fast and automatic localization of the different inner organ regions on 3D CT scans. The proposed approach combines object detections and the majority voting technique to achieve the robust and quick organ localization. The basic idea of proposed method is to detect a number of 2D partial appearances of a 3D target region on CT images from multiple body directions, on multiple image scales, by using multiple feature spaces, and vote all the 2D detecting results back to the 3D image space to statistically decide one 3D bounding rectangle of the target organ. Ensemble learning was used to train the multiple 2D detectors based on template matching on local binary patterns and Haar-like feature spaces. A collaborative voting was used to decide the corner coordinates of the 3D bounding rectangle of the target organ region based on the coordinate histograms from detection results in three body directions. Since the architecture of the proposed method (multiple independent detections connected to a majority voting) naturally fits the parallel computing paradigm and multi-core CPU hardware, the proposed algorithm was easy to achieve a high computational efficiently for the organ localizations on a whole body CT scan by using general-purpose computers. We applied this approach to localization of 12 kinds of major organ regions independently on 1,300 torso CT scans. In our experiments, we randomly selected 300 CT scans (with human indicated organ and tissue locations) for training, and then, applied the proposed approach with the training results to localize each of the target regions on the other 1,000 CT scans for the performance testing. The experimental results showed the possibility of the proposed approach to automatically locate different kinds of organs on the whole body CT scans.

Keywords: Torso CT images, Organ localization, Object detection, Majority voting, Haar-like feature, Local binary patterns.

1. INTRODUCTION

CT imaging has been widely used in the clinical medicine to support diagnosis, surgery and therapy. Effective image analysis algorithms and software tools can substantially help doctors to increase efficiency and accuracy, reduce tedium and oversights during the CT image interpretations. Accurately and efficiently detecting the location of an object of interest (an organ, a lesion, etc) plays an important role in the automated CT image analysis. The detected object location is usually in the form of a closed surface that bounds the object of interest. If we require the closed surface to be exactly aligned with the object boundary, the problem of object location detection is reduced to the challenging problem of image segmentation. In this paper, we focus on the object location detection by finding the 3D minimum bounding rectangle (with six faces parallel to the *x*-*y*, *y*-*z*, and *z*-*x* planes respectively) that tightly covers the object of interest. Such a 3D rectangle not only describes certain geometry properties of the object of interest, but also reduces the search space and difficulty for further image segmentation.

Medical Imaging 2013: Computer-Aided Diagnosis, edited by Carol L. Novak, Stephen Aylward, Proc. of SPIE Vol. 8670, 86703A · © 2013 SPIE · CCC code: 1605-7422/13/$18 · doi: 10.1117/12.2007466

Ensemble learning, such as AdaBoosting, has been successfully used for solving object detection problems in many computer-vision applications [1,2]. It has also been used for 3D CT image analysis, including heart structure recognition [3], liver segmentation [4], and anatomical landmark detection [5]. Recently, decision forests have been used successfully for detecting the inner organs in CT images [6]. All of those works reported good performance and demonstrated potentials of using ensemble learning for organ segmentation and localization in CT images. However, classical ensemble learning requires a large number of samples for training and testing. Especially, 3D CT images have a high feature dimension and we need a large number of training images to avoid the over-learning problem. In practice, it is difficult to collect a number of labeled 3D CT scans for satisfying this requirement. For example, as far as we know, few previous works reported performances on more than 1,000 CT scans and showed the possibility of the proposed algorithm for the localization of all the major organs in human torso from 3D CT images.

In this paper, we propose an approach to detect inner organ locations on 3D CT images based on ensemble-learning with a majority voting method. This approach is aimed to solve the different kinds of inner organ detection problems generally and can handle real clinical images generated by the different CT scanners, including non-contrast and contrast-enhanced, normal and abnormal cases. Our purpose is to achieve robust and automatic organ localization by using only a small size of training data.

2. METHODS

2.1 Outline

The process flow of the proposed approach is shown in Fig.1. In this paper, we handle the location detection of different inner organs separately and independently. Our method is to treat 3D organ localization in a 3D CT volume as detecting several independent 2D objects in a series of 2D image slices [7,8]. Obviously, this solution can reduce the feature dimension (3D to 2D) and increase the number of training samples (one 3D training sample consists of a large number of 2D training samples) during the ensemble learning. This can increase the detection robustness according to Occam's razor. For an unseen 3D CT scan, our method applies different 2D detectors to each voxel independently to detect a number of 2D candidates of a target and votes those 2D candidates back to the 3D space. Finally, our algorithm judges the existence of the target by checking the mutual consent of the responses from all 2D detectors; selecting the location-majority of the related 2D candidates in the 3D voting space as the final target location.

The location of an inner organ is defined by a ground-truth 3D minimum bounding rectangle (MBR) that covers all the voxels in the target organ region, where the MBR is aligned with the x, y and z-axes, i.e., its six faces are parallel to x-y, y-z and z-x planes, respectively. The 3D MBR of an organ can be uniquely described by two corners $P_{min}=(x_{min},y_{min},z_{min})^t$ and $P_{max}=(x_{max},y_{max},z_{max})^t$. The x_{max}, y_{max}, z_{max}, x_{min}, y_{min}, z_{min} are the maximum and the minimum coordinates of all the

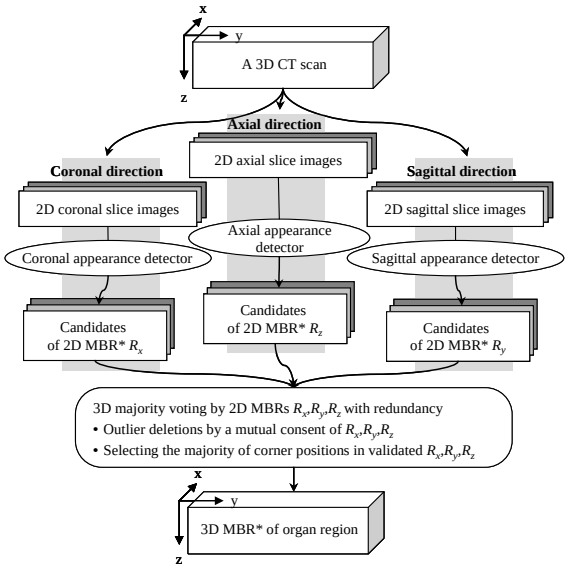


Fig.1. Process flow for localizing a target organ on a 3D CT scan (*MBR: minimum bounding rectangle).

voxels in the organ region along the sagittal, coronal and axial body directions [7]. In this way, the problem of detecting the location of an inner organ is reduced to a problem of finding the two MBR corners P_{min} and P_{max}.

Instead of directly finding P_{min} and P_{max} for the 3D MBR, we try to find three 2D MBRs, which are the projections of the 3D MBR onto *x*-*y*, *y*-*z* and *z*-*x* planes, respectively [8]. The 2D MBR R_z on the *x*-*y* plane, defined by two corners $P^z_{min}=(x_{min},y_{min})^t$ and $P^z_{max}=(x_{max},y_{max})^t$; the 2D MBR R_x on the *y*-*z* plane, defined by two corners $P^x_{min}=(y_{min},z_{min})^t$ and $P^x_{max}=(y_{max},z_{max})^t$; and the 2D MBR R_y on the *z*-*x* plane, defined by two corners $P^y_{min}=(z_{min},x_{min})^t$ and $P^y_{max}=(z_{max},x_{max})^t$. As shown in Fig.2, we train three 2D location detectors for finding a number of candidates of 2D MBRs R_x, R_y, R_z independently on each sagittal-, coronal-, and axial-direction slice of a 3D CT scan. Finally, the corners of the detected 2D bounding rectangles are back-projected to the 3D space and voted for estimating the underlying 3D MDR

2.2 Training 2D detectors by ensemble learning

An inner organ region in a 3D CT scan is constructed by a series of consecutive 2D slices along a given direction (*x*, *y* or *z*). The appearance of an organ in each 2D slice is highly correlated and similar to its neighbor slices. Our basic assumption is that the appearances of an inner organ on 2D slices along the same direction are similar and could be recognized by a single 2D detector. Here, we only require a "weak" 2D detector that can output a part of 2D slices of the target organ region (low true positive rate) and allow a lot number of false positive (FP). This is exactly the strength of traditional ensemble learning approach. Because the 2D detection results have a very high redundancy to show the 3D organ locations, the later majority voting can further reduce FP rate without the accuracy-loss during the final decision.

We take the 2D slices from the 3D training CT scans (with the manually labeled ground-truth 3D MBRs) for training the 2D organ-location detectors. Specifically, the slices along the sagittal, coronal, and axial directions are used for training the detectors for finding the candidates of 2D MBRs R_x, R_y, R_z, respectively as shown in Fig.2. Without loss of generality, in the following we focus on describing the training algorithm for finding the candidates of 2D MBR R_z. We collect the slices of the 3D training images along the axial body direction. If a slice intersects the ground-truth 3D MBR, we further check the 2D bounding rectangle resulting from this intersection. If the corseted target-organ in this slice is representative, i.e., the target-organ pixels count for a high percentage of the area of the 2D bounding rectangle, we crop this slice by this 2D bounding rectangle and then take the cropped slice as a positive 2D training sample. We randomly select a set of training slices cropped by rectangles that has no overlap with the ground-truth MBR as the negative 2D training samples. We then apply a cascaded AdaBoosting algorithm [9], using 2D Haar-like features [2] and local binary patters[10], to train the 2D target-organ location detector that can be applied to other axial-direction CT slices for finding the candidates of 2D MBR R_z. In the same way, we can train 2D detectors for finding the candidates of 2D MBRs R_x and R_y using the slices along the coronal and sagittal directions.

2.3 Localizing a target organ in a 3D CT scan by majority voting

Given an unseen 3D CT scan, we first apply the trained three 2D location detectors on all the slices along the three directions respectively to find out the possible existence of a target organ by using a sliding-window. Each location in the CT scan will be checked three times, i.e., along axial, coronal and sagittal directions, respectively, and only the locations that passed three examinations are regarded as the candidates of the target location. Ideally, 2D bounding rectangles detected from different slices provide consistent values of x_{min}, y_{min}, z_{min}, x_{max}, y_{max}, z_{max}, from which we can derive the 3D MBR. In practice, however, detected 2D rectangles may not lead to consistent values of x_{min}, y_{min}, z_{min}, x_{max}, y_{max}, z_{max} because of various noise and detection errors. We propose to use a majority-voting technique [8] to achieve an optimal estimate of the values of x_{min}, y_{min}, z_{min}, x_{max}, y_{max}, z_{max} (Fig.2). The detailed algorithm consists of the following steps:

1. Given an input test 3D CT scan *A*, we construct an all-zero output 3D image *B*, which is of the same size as *A*.

2. For each slice along the sagittal, coronal, or axial direction in *A*, we use the corresponding 2D target-location detector to detect 2D bounding rectangles. Note that no limit is set for the number of 2D rectangles detected on each slice.

3. Loop over all the 2D rectangles detected in Step 2. Whenever a voxel in *A* is covered by a rectangle, we increment the intensity value (votes) of the corresponding voxel in the output image *B* by 1.

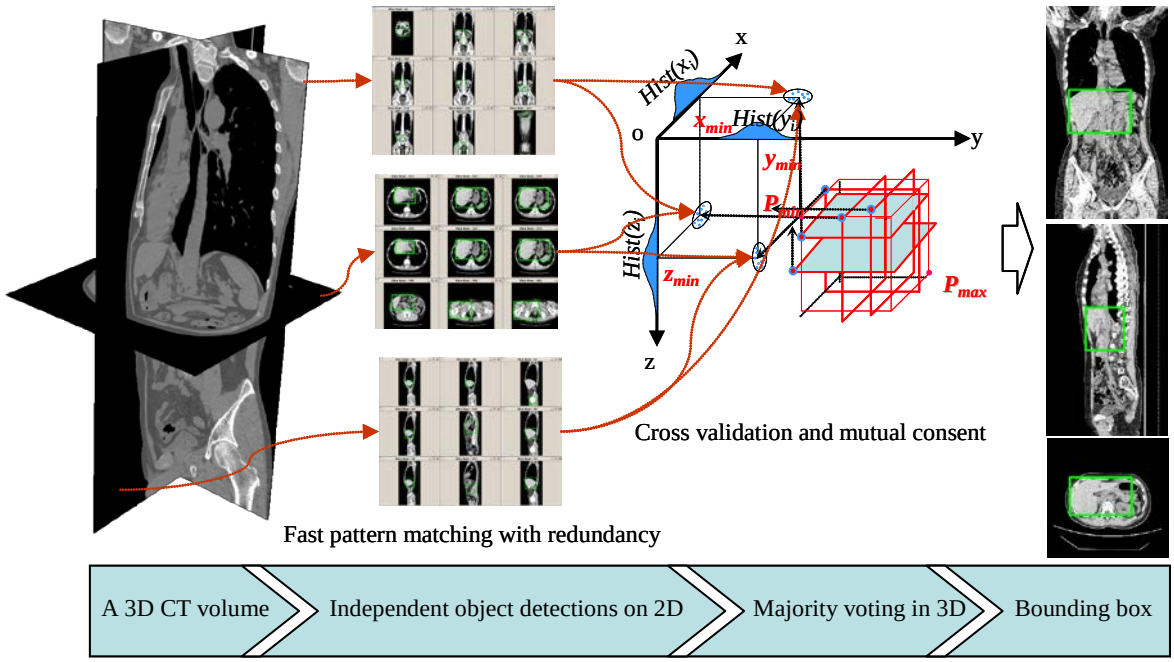


Fig.2. Majority voting based on *n*-slice-detection redundancy for 3D organ localization.

4. The voxel intensity in the output image *B* takes a vote value of 0, 1, 2, or 3. A voxel intensity of 3 at *B* indicates that this voxel is located in the detected 2D rectangles along both sagittal, coronal, and axial directions and the corresponding voxel in *A* shows a strong likeliness to be covered by the desirable 3D MBR. We find the 3D connected components based only on the voxels with intensity 3 in *B* and then select the 3D connected components with the largest volume as the candidates for target. If no voxel in *B* has intensity 3, algorithm is terminated with no target detected, i.e., the target is not involved in this 3D CT scan or may have been removed by surgery.

5. For the selected 3D connected component, we find its gravity center by averaging all the involved voxels (with intensity 3 in image *B*). Find the voxel *Vc* closest to this gravity center. Now we use all the voxels with nonzero intensity in *B* to find the connected component that contains the voxel *Vc*.

6. The voxels with nonzero intensity in *B*, but not included in the connected component found in Step 5, are treated as outliers. All the 2D rectangles detected in Step 2 that cover the outlier voxels are discarded.

7. For the remaining 2D rectangles, we check their x_{min}, y_{min}, z_{min}, x_{max}, y_{max}, z_{max} coordinates and construct a histogram for each one of these six coordinates (Fig.2). For example, the histogram *Hist(x)* for x_{min} describes the number of occurrence of each possible x_{min} used in the remaining 2D rectangle (R_z and R_y along two directions).

8. Smooth these six histogram functions by triangular windowing and then pick the optimal x_{min}, y_{min}, z_{min}, x_{max}, y_{max}, z_{max} which maximize their respective histogram functions. These optimal x_{min}, y_{min}, z_{min}, x_{max}, y_{max}, z_{max} defines a 3D rectangle, which we use as the final estimate of the underlying 3D MBR and provides a location detection of the target organ in this 3D CT scan, as illustrated in Fig. 2.

3. EXPERIMENT AND RESULTS

A dataset including 1,300 cases of 3D volumetric CT scans was used in this experiment. These CT scans were collected in Gifu University Hospital by two kinds of multi-slice CT scanners (LightSpeed Ultra16 of GE Healthcare and Brilliance 64 of Philips Medical Systems). Each CT scan used a common protocol (120 kV/Auto mA) and covered the entire human torso region. Each 3D CT scan has approximately 800-1200 axial CT slices by an isotopic spatial resolution of approximately 0.6-0.7 mm and a density (CT number) resolution of 12 bits. A part of CT cases were

scanned with the contrast enhancements. The age of these patients ranges from 25 to 92 years old. All of these CT images are taken for the patients with certain real or suspicious abnormalities.

Twelve major organs including heart, liver, spleen, pancreas, stomach, bladder, left and right lungs, left and right kidneys, left and right femur-heads were selected as the detection targets for evaluating the performance of the proposed approach. The 3D MBRs (P_{min}, P_{max}) of those targets in each 3D CT scan were manually marked by the authors as the ground-truth. 300 non-contrast 3D CT scans were randomly selected for training. This leads to about 4,000-10,000 positive 2D training samples and 40,000-100,000 negative 2D training samples that were used to train the 2D location detectors for those organs along each of the sagittal, coronal, and axial directions. The detectors for each organ were trained separately and independently. Each 2D location detector consisted of 15-20 cascades and each stage in the cascades was a classifier by combining (boosting) 5-100 weak classifiers.

The proposed approach was applied to localize 12 target organs in 1,000 3D CT scans in the dataset independently. Those CT scans were test samples and not used for training. Accuracy of the organ localization was first carried out by a subjective evaluation by authors (including an anatomist and a radiologist), and then, we randomly selected a number of samples for quantitative evaluations. The ground truth 3D MBRs of each target organ were manually labeled by authors, and the volume-center-distance *Distc* and volume-variance-ratio *Volvr* that defined as the value of (1-(volume of detected MBR)/(volume of ground-truth MBR)) between the detected 3D rectangle and ground-truth MBR are used as the evaluation measures. An example of the detected organs in a 3D CT scan is shown in Fig.3, The histograms of *Distc* and *Volvr* for these target organs are shown in Fig.4. The computing time for detecting a target organ location was less than 15 seconds per CT scan by using a general purpose computer equipped with an Intel Core2Due 2.23 GHz CPU.

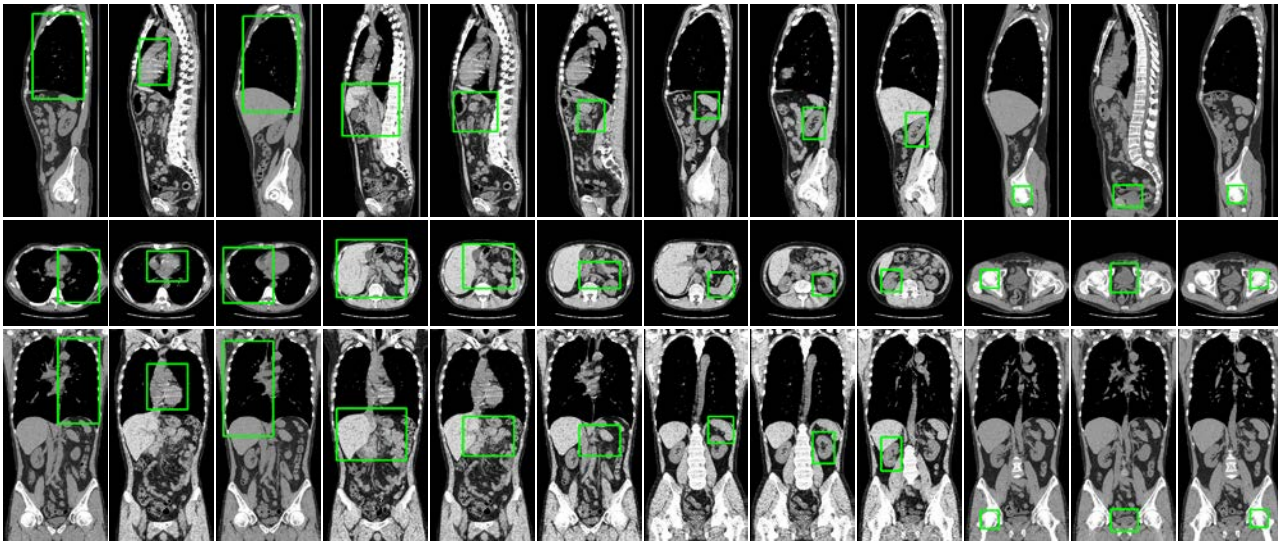
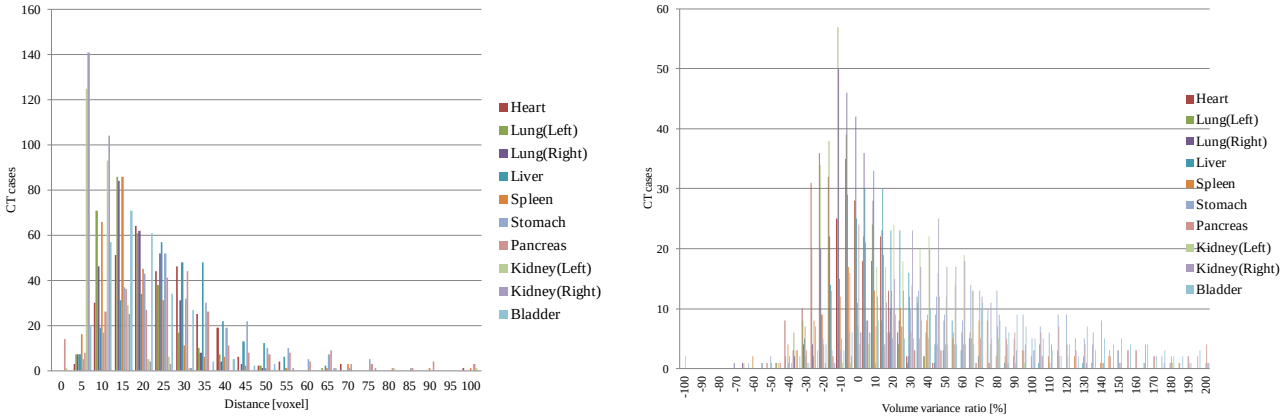


Fig.3. An example of the localization results for 12 kinds of major organs in a 3D CT scan. Three slices that pass through the detected center position of the target organ are shown. The rectangle indicates the detected organ location (bounding rectangle of the left-lung, heart, right-lung, liver, stomach, pancreas, spleen, left-kidney, right-kidney, right-femur-head, bladder, and left-femur-head from the left to right sides).

4. DISCUSSION AND CONCLUSION

In this study, the detected location was considered to be correct if the majority parts of detected 3D rectangle and the ground-truth MBR overlap with each other. Our subjective evaluation based on 1,000 test CT cases showed that the heart location in 983 cases, liver location in 978 cases, stomach location in 952 cases, pancreas location in 943 cases, spleen location in 932 cases, left kidney localization in 963 cases, right kidney localization in 969 cases, left lung location in 992 cases, right lung location in 988 cases, bladder location in 984 cases, right-femur-head location in 578 cases, and left-femur-head location in 551 cases were correct. Those results showed that the proposed approach can accomplish the localization tasks in over 94% CT cases except for the femur-heads. The failure of the femur-head locations was due to the poor performance of the trained 2D detectors. This problem was caused by the smaller sample number and less texture information of the femur-head region in contrast to the other organs during the training stage.



(a) Histogram of volume-center-distances values (*Distc*) (b) Histogram of the volume-variance-ratio (*Volvr*)

Fig.4. Accuracy evaluations of organ localization results by comparing with the human manual inputs (ground-truth) using 300 CT scans.

For the quantitative evaluation as shown in Fig.4, *Distcs* of the target organs were less than 50 voxels in most CT cases with the modes ranged from 10 to 20 voxels. The histograms of *Volvr* of the most organs were distributed mostly between -20% and 20% with the modes around zero except for the bladder and liver. This evaluation result showed that the proposed approach can find the center position approximately and determine the size of the bounding rectangle of the target organ regions. However, in the cases that the target organ has a large variance in the volume (bladder) or unsymmetrical shape (liver), the accuracy of the detected bounding rectangle needs to be improved.

The major contribution of this paper is the use of the ensemble learning for the 2D location detection along three different directions, and then the integration of the 2D detection results to estimate the desirable 3D organ location. The proposed approach only requires the 2D detectors to be "weak" detectors. Majority voting of multiple 2D candidates from three independent directions will combine all 2D weak detectors into a 3D "strong" detector. AdaBoosting algorithm based on Haar-like features and local binary patterns has been used for face detection with very high accuracy. However, in the proposed approach, each individual 2D location detector leads to very low detection accuracy. Based on our experiments on the 300 3D CT scans that were used for training, more than 50% of the 2D rectangles detected in the Step 2 of the algorithm described in Section 2.3 were outliers. On the other hand, less than 50% of the ground-truth 2D MBRs were not found by the 2D location detectors on the corresponding slices. This is completely reasonable, because the inner organ regions may show inconsistent intensity and appearance in different 3D CT scans, and the contrast between the organ region and the surrounding background is usually very poor. For such organ detection problems, our approach takes advantage of the redundancy among the 2D detection results drawn from different directions and applies a majority voting technique to remove the outliers in the 2D detection results.

In conclusion, we proposed a universal approach that can be used to localize the major organs automatically on 3D CT scans. This approach was applied to 12 kinds of organ regions and its efficiency and accuracy were validated by using 1,300 clinical CT scans.

ACKNOWLEDGEMENTS

Authors thank the members of Fujita Laboratory, who assisted to establish the dataset. This research work was funded in part by a Grant-in-Aid for Scientific Research on Innovative Areas (21103004), and in part by Grant-in-Aid for Scientific Research (C23500118), MEXT, Japan.

REFERENCES

1. P. Viola and M. J. Jones: Rapid Object Detection using a Boosted Cascade of Simple Features, IEEE CVPR, 2001.
2. R. Lienhart and J. Maydt: An Extended Set of Haar-like Features for Rapid Object Detection, IEEE ICIP 2002, vol. 1, pp. 900-903, 2002.

3. Y. Zheng, B. Adrian, G. Bogdan, S. Michael, C. Dorin: Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features, IEEE TMI, vol.27, pp.1668-1681, 2008.
4. H. Ling, S. K. Zhou, Y. Zheng, B. Georgescu, M. Suehling, and C. Dorin: Hierarchical, Learning-based Automatic Liver Segmentation, IEEE CVPR, 2008.
5. M. Dikmen, Y. Zhan, and X. S. Zhou: Joint Detection and Localization of Multiple Anatomical Landmarks Through Learning, SPIE Medical Imaging 2008, vol.6915, 2008.
6. A. Criminisi, J. Shotton, and S. Bucciarelli: Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes, in MICCAI workshop on Probabilistic Models for Medical Image Analysis (MICCAI-PMMIA), 2009.
7. X. Zhou, and H. Fujita: Automatic Organ Localization on X-ray CT Images by Using Ensemble Learning Techniques, in Machine Learning in Computer-aided Diagnosis: Medical Imaging Intelligence and Analysis, ed. by K. Suzuki, IGI Global, USA, pp. 403-418, 2012.
8. X. Zhou, S. Wang, H. Chen, T. Hara, R. Yokoyama, M. Kanematsu, H. Hoshi, and H. Fujita: Automatic Localization of Solid Organs on 3D CT Images by A Collaborative Majority Voting Decision Based on Ensemble Learning, Computerized Medical Imaging and Graphics, vol.36, 4, pp. 304-313, 2012 .
9. Y. Freund, and R.E.Schapire: A Decision-theoretic Generalization of On-line Learning and an Application to Boosting. In Computational Learning Theory: Eurocolt 95, Springer-Verlag, pp. 23–37, 1995.
10. T. Ojala, M. Pietikäinen, and D. Harwood: Performance Evaluation of Texture Measures with Classification based on Kullback Discrimination of Distributions, Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994), vol. 1, pp. 582 - 585, 1994.