Automated Localization of Solid Organs in 3D CT Images: A Majority Voting Algorithm Based on Ensemble Learning

Xiangrong Zhou [1], Syunichi Yoshimoto [1], Song Wang [2], Huayue Chen [1], Takeshi Hara [1], Ryujiro Yokoyama [3], Masayuki, Kanematsu [3], and Hiroshi Fujita [1]

[1] Graduate School of Medicine, Gifu University, Gifu 501-1194, Japan
{zxr,hara,ryujiro,fujita}@fjt.info.gifu-u.ac.jp
[2] Department of Computer Science and Engineering, University of South Carolina, USA
[3] Gifu University Hospital, Gifu 501-1194, Japan

Abstract. This paper describes a new approach to automatically find out the location of a target solid organ in 3D CT scans. Specifically, our goal is to detect a 3D rectangular box for the target organ in a way that this rectangle bounds the organ region tightly and accurately. The proposed approach combines the ensemble learning and the majority voting techniques to achieve a robust detection by using a small number of CT scans for training. A database including 3,329 torso CT scans is used in experiments. Among them, we manually label the heart and the left/right kidneys from nearly 100 3D CT scans as training samples, and use the proposed approach to localize those organs in the other CT scans. Experimental results show that detection rates are 99% for the heart, 85%-87% for the right and left kidney, with a computation time less than 15 seconds per CT scan on a general PC.

Keywords: CT images, solid organ localization, ensemble learning, majority voting.

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 rectangular box (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 rectangular box not only describes certain geometry properties of the object of interest, but also reduces the search space and difficulty for further image segmentation.

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, no previous works reported performances on more than 1,000 CT scans.

In this paper, we propose a new approach to detect inner organ locations in 3D CT images based on ensemble-learning with a majority voting method. This approach is aimed to solve the different solid organ detection problems generally and can handle different kinds of real clinical CT images, 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 number (about 100) of CT scans for training.

2 Methods

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. 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, we judge the existence of the target by checking the mutual consent of the responses from all 2D detectors; selecting the majority of the related 2D candidates in the 3D voting space as the target location.

2.1 Overview

The location of an inner organ is defined by a ground-truth 3D minimum bounding rectangular box (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 voxels in the organ region along the sagittal, coronal and axial body directions. This way, the problem of

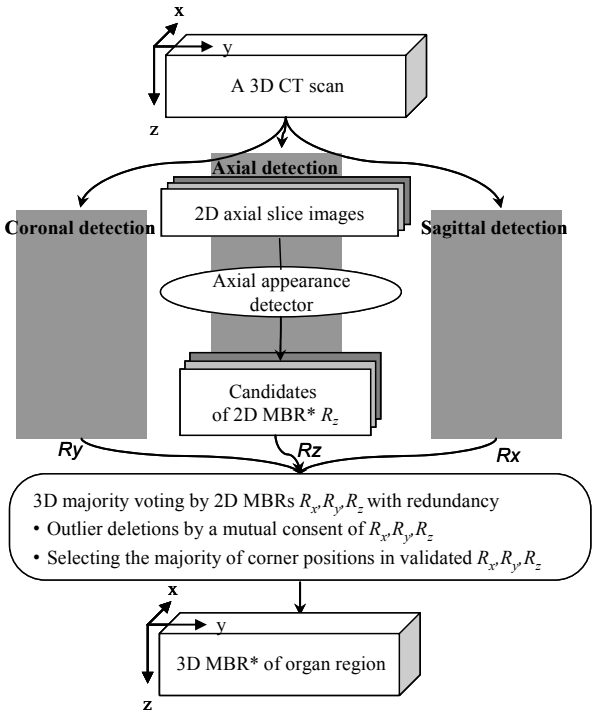


Fig.1. Process flow for localizing a solid organ in a 3D CT scan. (*MBR: minimum bounding rectangle)

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. 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 summarized in Fig.1, 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

A solid 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 a solid 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 which may have optimal balance between the false positive (FP) and true positive rates, enhancing both efficiency and quality. This is exactly the strength of traditional ensemble learning approach. The later majority voting step would further reduce FP rate and make a correct 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. 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 [1], using 2D Haar-like features [2], 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. 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 to achieve an optimal estimate of the values of x_{min},y_{min},z_{min}, x_{max},y_{max},z_{max}. 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.

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. 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 rectangular box, 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.

3 Experiment and Results

A database includes 3,329 cases of 3D volumetric CT scans were 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 isotropic spatial resolution of approximately 0.625 mm and a density (CT number) resolution of 12 bits. Contrast media was used for enhancement in 272 CT scans and the other 3,057 scans are non-contrast CT images. The age of these patients is from 25 to 92. All of these CT images are taken for the patients with certain real or suspicious abnormalities. Furthermore, the left or right kidney in some cases has been physically removed by surgery.

Heart, left and right kidneys are selected as the detection targets for evaluating the proposed approach. The 3D MBRs (P_{min} ,P_{max}) of those targets in each 3D CT scan are manually marked by the authors. 100 non-contrast 3D CT scans are randomly selected for training. Considering that one kidney is removed in some cases, actually, 91 3D CT scan are used for training the left kidney detectors and 97 3D CT scans are used for training the right kidney detectors. This leads to 700-1800 positive 2D

training samples and 10,000-25,000 negative 2D training samples that are used to train the 2D location detectors for each organ along each of the sagittal, coronal, and axial directions. The detectors for heart, left and right kidneys are trained separately and independently. Each 2D location detector consists of 10-15 cascades and each stage in the cascades is a classifier by combining (boosting) 10-30 weak classifiers.

The proposed approach is applied to localize the heart, left and right kidneys in 3,228 CT scans in the database independently. Those CT scans are 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 559 CT scans from test samples for quantitative evaluations. The ground truth 3D MBRs of each target organ were manually labeled by authors, and the volume overlap between the detected 3D rectangular box A and ground-truth MBR B (*JSC*: Jaccard similarity coefficient=$(A \cap B)/(A \cup B)$), Euclidean distance between the centers of these two 3D rectangular boxes (*Distc*) are used as the evaluation measures. An example of the detected heart, left and right kidneys in a 3D CT scan is shown in Fig.2, The histograms of *JSC* and *Distc* for these target organs are shown in Fig.3. The computing time for detecting a target location is less than 15 seconds per CT scan by using a computer equipped with an Intel Core2Due 2.23 GHz CPU.

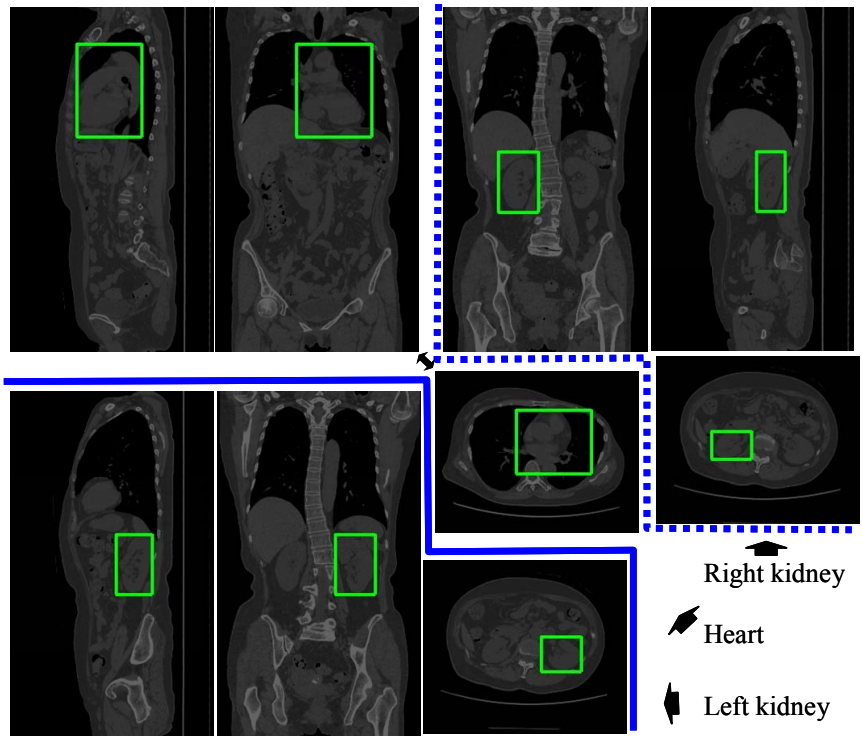
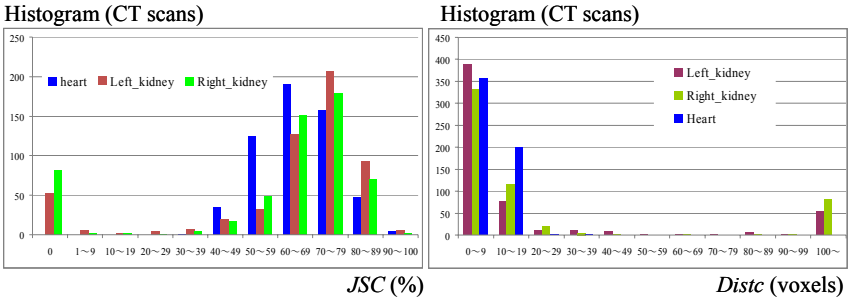


Fig.2. An example of the localization results for heart, left and right kidneys in a 3D CT scan. Three slices that pass through the detected center position of the target organ are shown. Green box indicates the detected organ location (bounding rectangle of the heart, left and right kidney regions).

4 Discussion and Conclusion

In this study, the detected location is considered to be correct if the majority parts (about 70% volume) of detected 3D rectangular box and the ground-truth MBR overlap with each other. Our subjective evaluation shows that the heart localization in 3,201(99%) cases, left kidney localization in 2,817 (87%) cases, and right kidney localization in 2,728 (85%) cases were correct. As shown in Fig.3, for the quantitative evaluation on the randomly selected CT scans, *JSC* histograms of left and right kidney are mostly distributed from 40% to 100% and centered at 75%, and localizations in 12% of 559 test CT scans completely fail. In the case of heart localization, the histogram of *JSC* is distributed from 40% to 90% and centered at 65%, without any failure in 559 test CT scans (Fig.3(a)). The histograms of *Distc* for heart, left and right kidney localizations are distributed mostly within 20 voxels as shown in Fig.3(b). Except the failure cases, most of kidney or heart localizations show a *Distc* error that is less than 15 voxels. Those results show that the proposed approach can accomplish the heart center detection and estimate its approximate extent. For kidney detections, our approach fails on 12% of the test CT scans and in the other 88% of CT scans, the performance (*JSC* and *Distc*) of kidney detections are actually higher than that for the heart. The performance for left kidney is slightly better than that for the right kidney. The failures are mainly from the non-contrast CT images, where the kidney appearance is unclear, and insufficiency of training samples for representing variety of kidney appearances.

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. We only require 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 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 559 3D CT scans that are not used for training, 55% and 37% (for the left and right kidneys respectively) of the 2D rectangles detected in the Step 2 of



(a) (b)

Fig.3. Accuracy evaluations of organ localizations using a number of test CT scans. (a) Histogram of *JSC* values; (b) Histogram of distance of center shift (*Distc*).

the algorithm described in Section 2.3 are outliers. On the other hand, 65% and 79% (for the left and right kidneys respectively) of the ground-truth 2D MBRs are not found by the 2D location detectors on the corresponding slices. This is completely reasonable because, the kidney may show inconsistent intensity and appearance in different 3D CT scans, and the contrast between kidney 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.

For the related works, Probabilistic Boosting Tree with 3D Haar-like features has been used to detect heart locations as the pre-processing for heart structures segmentation [3]. This method was trained and tested successfully using 323 CT scans from 137 patient cases by a four-fold cross validation. Our approach accomplished same detection task in a lower-dimensional feature space (2D Haar-like feature). In addition, our approach uses a smaller number of training samples (100) and is validated by a much larger number (3,229) of unseen CT scans. Recently, an approach based on decision forests with long-range spatial context has been used for organ localization [6]. This approach was trained and tested for 9-organ localization based on 39 CT scans, resulting in an average localization error (*Distc*) of 21.32 mm for heart, 25.42 mm for left kidney, and 44.52 mm for right kidney. Comparing to [6], our approach is much simpler and only needs to use local features within extent of the target organ. Additionally, our approach produces a smaller localization errors (*Disc*) of heart, kidney localizations in most CT scans.

As the conclusion, we proposed a simple approach that can be used to localize the solid organ automatically in 3D CT scans. This approach was applied to the heart, left and right kidney detections and its efficiency and accuracy were validated on a large number of clinical CT scans.

References

1. Paul Viola and Michael J. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, IEEE CVPR, 2001.
2. Rainer Lienhart and Jochen Maydt, “An Extended Set of Haar-like Features for Rapid Object Detection”, IEEE ICIP 2002, vol. 1, pp. 900-903, 2002.
3. Yefeng Zheng, Barbu Adrian, Georgescu Bogdan, Scheuering Michael, Comaniciu 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. Haibin Ling, S. Kevin Zhou, Yefeng Zheng, Bogdan Georgescu, Michael Suehling, and Dorin Comaniciu, “Hierarchical, Learning-based Automatic Liver Segmentation”, IEEE CVPR, 2008.
5. Mert Dikmen, Yiqiang Zhan, and Xiang Sean Zhou, “Joint Detection and Localization of Multiple Anatomical Landmarks Through Learning”, SPIE Medical Imaging 2008, vol.6915, 2008.
6. Antonio Criminisi, Jamie Shotton, and Stefano 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.