Computerized Medical Imaging and Graphics 36 (2012) 304–313

Contents lists available at SciVerse ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Automatic localization of solid organs on 3D CT images by a collaborative majority voting decision based on ensemble learning

Xiangrong Zhou [a,*], Song Wang [b], Huayue Chen [c], Takeshi Hara [a], Ryujiro Yokoyama [d], Masayuki Kanematsu [e], Hiroshi Fujita [a]

[a] Department of Intelligent Image Information, Division of Regeneration and Advanced Medical Sciences, Graduate School of Medicine, Gifu University, Japan
[b] Department of Computer Science and Engineering, University of South Carolina, United States
[c] Department of Anatomy, Graduate School of Medicine, Gifu University, Japan
[d] Department of Radiology Services, Gifu University Hospital, Japan
[e] Department of Radiology, Gifu University Hospital, Japan

ARTICLE INFO

Article history:
Received 4 November 2010
Received in revised form 18 May 2011
Accepted 19 December 2011

Keywords:
3D CT torso images
Inner organ localization
Ensemble learning
Collaborative majority voting

ABSTRACT

Purpose: Organ segmentation is an essential step in the development of computer-aided diagnosis/surgery systems based on computed tomography (CT) images. A universal segmentation approach/scheme that can adapt to different organ segmentations can substantially increase the efficiency and robustness of such computer-aided systems. However, this is a very challenging problem. An initial determination of the approximate position and range of a target organ in CT images is prerequisite for precise organ segmentations. In this study, we have proposed a universal approach that enables automatic localization of the approximate position and range of different solid organs in the torso region on three-dimensional (3D) CT scans.

Methods: The location of a target organ in a 3D CT scan is presented as a 3D rectangle that bounds the organ region tightly and accurately. Our goal was to automatically and effectively detect such a target organ-specific 3D rectangle. In our proposed approach, multiple 2D detectors are trained using ensemble learning and their outputs are combined using a collaborative majority voting in 3D to accomplish the robust organ localizations.

Results: We applied this approach to localize the heart, liver, spleen, left-kidney, and right-kidney regions independently using a CT image database that includes 660 torso CT scans. In the experiment, we manually labeled the abovementioned target organs from 101 3D CT scans as training samples and used our proposed approach to localize the 5 kinds of target organs separately on the remaining 559 torso CT scans. The localization results of each organ were evaluated quantitatively by comparing with the corresponding ground truths obtained from the target organs that were manually labeled by human operators. Experimental results showed that success rates of such organ localizations were distributed from 99% to 75% of the 559 test CT scans. We compared the performance of our approach with an atlas-based approach. The errors of the detected organ-center-positions in the successful CT scans by our approach had a mean value of 5.14 voxels, and those errors were much smaller than the results (mean value about 25 voxels) from the atlas-based approach. The potential usefulness of the proposed organ localization was also shown in a preliminary investigation of left kidney segmentation in non-contrast CT images.

Conclusions: We proposed an approach to accomplish automatic localizations of major solid organs on torso CT scans. The accuracy of localizations, flexibility of localizations of different organs, robustness to contrast and non-contrast CT images, and normal and abnormal patient cases, and computing efficiency were validated on the basis of a large number of torso CT scans.

1. Introduction

Three-dimensional (3D) images of anatomical structures of a patient are critical factors that support diagnosis, surgery, and therapy in clinical medicine. A modern radiographic computed tomography (CT) scan can generate a 3D volumetric image that provides detailed anatomical information of a human torso within

* Corresponding author. Tel.: +81 582306510; fax: +81 582306514.
E-mail address: zxr@fjt.info.gifu-u.ac.jp (X. Zhou).

0895-6111/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2011.12.004

10 s. However, a long time and experience are required to interpret such volumetric CT images by the traditional method to identify suspicious regions. Therefore, development of computer-based image analysis algorithms and visualization tools is expected to enable doctors increase efficiency and accuracy of interpretation, and reduce tedium and oversights during CT image interpretation.

A fundamental component of such a computer system is accurate and efficient organ segmentation. Separating a target organ region from the background by identifying the contour (surface in 3D) of the organ region is the major goal of organ segmentation. However, it may not be feasible to search for all voxels in a CT scan to determine the correct contour of a target organ. Detecting the center position and deciding the range (a 3D bounding rectangle) of the target organ in a CT scan is a practical pre-processing step that can be very useful for facilitating accurate organ segmentation. Such a 3D rectangle not only increases the accuracy and reduces the difficulty for further image segmentation, but also describes local geometry and density properties that can be used for diagnosis. In practice, a torso CT scan always shows the complex anatomical structures of all inner organs. Thus, automatic localization of different organ regions in torso CT scans is a very difficult problem that remains to be addressed.

Object localization is a major topic in the research field of computer vision, and ensemble learning such as AdaBoosting is an efficient way to solve face detection problems from pictures and videos [1,2]. Recently, ensemble learning has also been used for 3D CT image analysis, including heart structure recognition [3], liver segmentation [4], and anatomical landmark detection [5]. Those studies have directly extended the 3D-Haar features and probabilistic boosting-tree [6] that are used for face or nature object detection and applied them to 3D organ localizations in CT images. In another approach, decision forests based on long-range spatial context [7] were successfully used for the localization of solid organs such as heart, lung, and liver in CT images. All of these works reported good performance and demonstrated the potential of using ensemble learning for organ segmentation and localization in CT images. However, such approaches had a drawback that classical ensemble learning requires a large number of training samples so that the usefulness and robustness of the trained detector can be maintained for unseen samples. In particular, 3D CT images have a high feature dimension and to avoid the over-learning problem, a large number of training samples are required. However, in practice, it is difficult to collect a large number of labeled 3D CT scans to fulfill this requirement. For example, as far as we know, no previous work reported the performance using more than 500 CT scans. How to train a useful detector for organ localization on the basis of a small number of CT scans and validate its performance are still challenging issues.

In this paper, we propose a new approach based on ensemble learning to localize major solid organs in 3D CT scans. The localization scheme generated by the proposed approach is applicable for different solid organ localizations, adaptive to both non-contrast and contrast-enhanced 3D CT images, and is robust to the unknown irregularity in abnormal CT scans. Due to the practical nature of clinical medicine, two additional requirements were also considered: (1) using only a small number (about 100) of CT scans for training and (2) accomplishing the organ localization quickly (within 1 min) on the basis of a general computer.

To satisfy these two requirements, we introduce two special components into ensemble learning. First, we carry out organ detection on a 2D slice unit instead of a 3D scan. In this case, each 2D CT slice from the same patient acts as an individual training sample. This way, even a small number of manually labeled 3D CT scans can provide a large number of 2D slices for training. Second, on the basis of the fact that the organ appearances on adjacent 2D slices along the same direction are similar to each other, we train three 2D location detectors to generate multiple candidates for organ location along the three orthogonal (sagittal, coronal, and axial) body directions, respectively. We then use majority voting of those candidates to achieve a highly accurate 3D organ localization. By adopting a cascade structure [1] and parallel computing technique, the proposed organ localization is computationally very efficient.

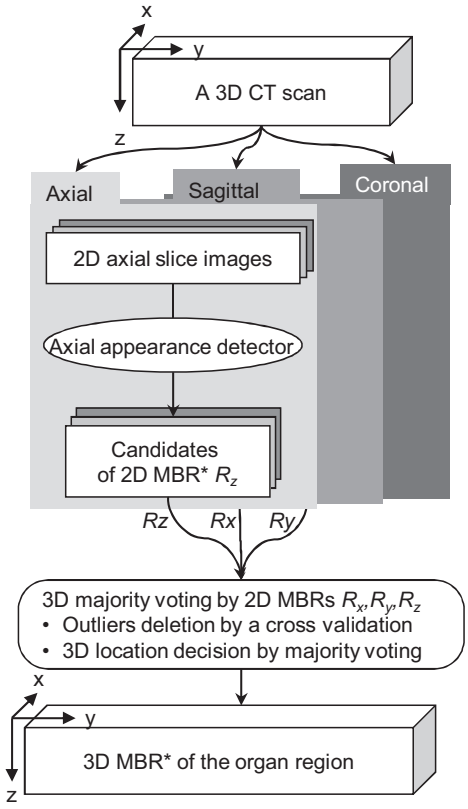

Fig. 1. Processing flow for detecting organ locations in a three-dimensional computed tomography (3D CT) scan. (*MBR, minimum bounding rectangle.)

This paper is organized as follows. Section 2 gives a technical overview of the proposed approach and a detailed description of the 2D detectors and 3D majority voting. The details of the experiment for localizing 5 kinds of solid organs on the basis of the approach proposed above are described in Section 3 and the results are shown in Section 4. We have discussed the performance of the proposed method in Section 5 and have given a conclusion in Section 6.

2. Methods

The process flow of the proposed approach is shown in Fig. 1. In this paper, we handled the location detection of different inner organs separately and independently. Our method was to treat 3D organ localization in a 3D CT scan 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 ensemble learning. This should increase the robustness of the trained detector for unknown samples according to Occam's razor. For an unknown 3D CT scan, our method applies different 2D detectors to each voxel independently to detect a number of 2D candidates of a target along three orthogonal directions and votes those 2D candidates back to the 3D space. Finally, we decide the existence and approximate center position of the target by checking the mutual consent of the responses from

all 2D detectors and select the majority of the range of the related 2D candidates in the 3D voting space as the target location.

2.1. Overview of the proposed approach

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 the 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} and x_{min}, y_{min}, z_{min} are the maximum and the minimum coordinates, respectively, of all the voxels in the organ region along the sagittal, coronal, and axial body directions. This way, the problem of detecting the location of an inner organ is reduced to the 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 the x–y, y–z, and z–x planes: the 2D MBR R_z on the x–y plane, defined by the two corners $P_{zmin} = (x_{min}, y_{min})^t$ and $P_{zmax} = (x_{max}, y_{max})^t$; the 2D MBR R_x on the y–z plane, defined by the two corners $P_{xmin} = (y_{min}, z_{min})^t$ and $P_{xmax} = (y_{max}, z_{max})^t$; and the 2D MBR R_y on the z-x plane, defined by the two corners $P_{ymin} = (z_{min}, x_{min})^t$ and $P_{ymax} = (z_{max}, x_{max})^t$. As summarized in Fig. 1, we train three 2D location detectors to identify a number of 2D MBRs R_x, R_y, R_z candidates 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 of the 2D location detectors

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 an 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 ensure the correct detection.

We took the 2D slices from the 3D training CT scans (with the manually labeled ground-truth 3D MBRs) for training the 2D organ-location detectors as shown in Fig. 2. Specifically, the slices along the sagittal, coronal, and axial directions are used for training the detectors for finding 2D MBRs R_x, R_y, R_z candidates, respectively. Without loss of generality, in the following we focus on describing the training algorithm for finding 2D MBR R_z candidates. We collected slices from 3D training images along the axial body. If a slice intersected the ground-truth 3D MBR, we further checked the 2D bounding rectangle resulting from this intersection. If the corseted target-organ in this slice was representative (the target-organ pixels count was approximately over 50 percent of the 2D bounding rectangle area), we cropped this slice by the 2D bounding rectangle and then took the cropped slice as a positive 2D training sample. All the positive training samples along the axial body were arranged to a 2D rectangle with a size S_a by image interpolation; S_a was empirically decided as one-fourth of the average size of positive training samples in this study. We randomly selected a set of training slices cropped by rectangles with size same as that of S_a that did not overlap with the ground-truth MBR as negative 2D training samples. A set of simple stump classifiers based on 2D Haar-like features (see Appendix A and Fig. 3) was used for detection. A typical detecting window had a large number of Haar-like features;

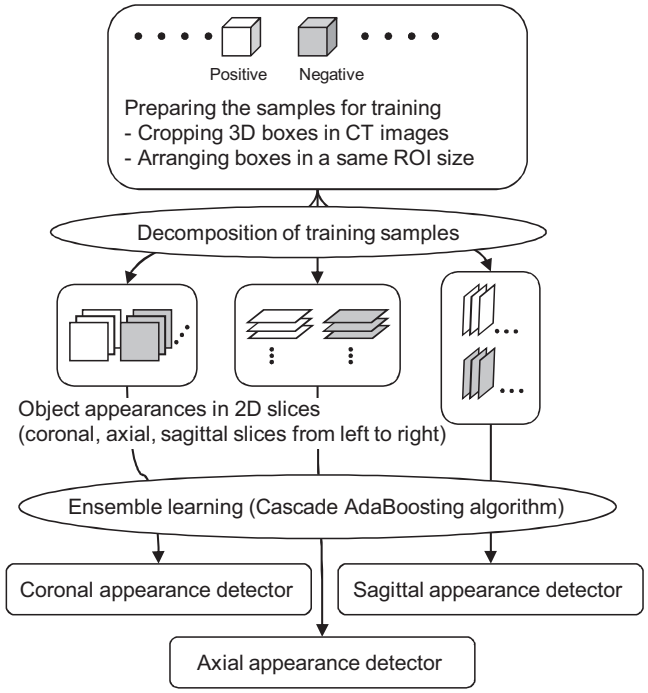

Fig. 2. Training three 2D detectors for detecting the organ appearances in coronal, axial, and sagittal 2D slices.

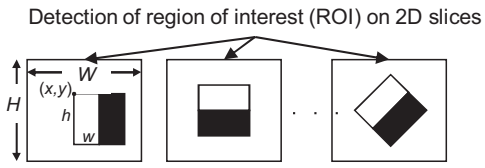

Fig. 3. Examples of 2D Haar-like features inside a detecting window with W, H pixels. The features are calculated from a set of sums of the *black* and *white* rectangles specified by the left-upper position x, y, size w, h and a rotation angle (see Appendix A).

the details of these features have been explained in the study [2]. A cascaded AdaBoosting algorithm [1] was applied to select a number (100–200) of useful features and combine those stump classifiers to a cascaded structure, as shown in Fig. 4. This cascaded structure acts as a detector that can be applied to other axial-direction CT slices for the identification of 2D MBR R_z candidates. Likewise, we trained 2D detectors to identify 2D MBRs R_x and R_y candidates using the slices along the coronal and sagittal directions. The technical details of the AdaBoosting algorithm are given in Appendix B.

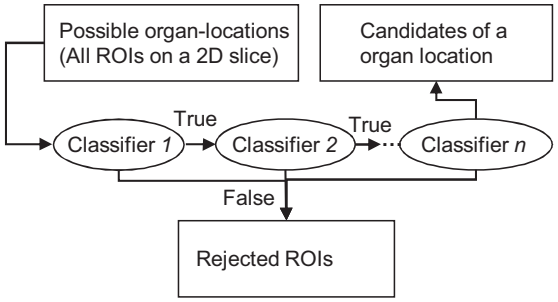

Fig. 4. Cascade structure based on AdaBoost classifiers for detecting candidates of target organ sections on 2D slices.

2.3. Majority voting for final organ location estimation

For an unknown test 3D CT scan, we first applied the three trained 2D organ-location detectors on all the slices along the three orthogonal directions, respectively. Scanning was repeated and the size of the detecting window was increased by 10% between the subsequent scans. By using variable sizes of detecting window, the detector can find all the candidates who may show spatial resolutions and physical sizes different from those of the training samples on test CT images. Clearly, 2D rectangles detected from different slices will not lead to exactly identical values for x_{min}, y_{min}, z_{min}, and x_{max}, y_{max}, z_{max} because of various noise and location detection errors. Based on the detected 2D rectangles, we can apply a majority voting technique to achieve an optimal estimate of the x_{min}, y_{min}, z_{min}, and x_{max}, y_{max}, z_{max} values. This proposed approach consists of the following steps:

Step 1. Given an input test 3D CT scan A, we constructed an all-zero output 3D volumetric image B, which was the same size as A.

Step 2. For each slice along the sagittal, coronal, or axial direction in A, we used the corresponding 2D organ-location detector to detect 2D rectangles. Note that there was no constraint for the number of 2D rectangles detected per slice.

Step 3. We looped over all of the 2D rectangles detected in Step 2. Whenever a voxel in A was covered by a rectangle, we increased the intensity value of the corresponding voxel in the output image B by 1.

Step 4. The voxel intensity in the output image B took a value of 0, 1, 2, or 3. A voxel intensity of 3 indicated that this voxel was located in the detected 2D rectangles along both sets of sagittal, coronal, and axial directions and that the corresponding voxel in A showed a strong likeliness to be covered by the desired 3D MBR. We found that the 3D connected components based only on voxels with an intensity of 3 in B and then selected a 3D connected component with the largest volume as the candidates for the target organ. If no voxel in B has an intensity of 3, the algorithm terminated with no target detected, i.e., the organ is not involved in this 3D CT scan or may have been removed by surgery.

Step 5. For the selected 3D connected component, we found its gravity center by averaging all of the involved voxels (with an intensity of 3 in image B) to find the voxel V_c closest to this gravity center. Then we used all the voxels with a nonzero intensity in B to find the connected component that contained voxel V_c.

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

Step 7. For the remaining 2D rectangles, we checked their x_{min}, y_{min}, z_{min} and x_{max}, y_{max}, z_{max} coordinates and constructed a histogram for each one of these six coordinates. For example, the histogram for x_{min} describes the number of occurrences of each possible x_{min} used in the remaining 2D rectangles along the y and z directions (refer to Fig. 5).

Step 8. These six histogram functions were smoothened by triangular windowing and then the optimal x_{min}, y_{min}, z_{min} and x_{max}, y_{max}, z_{max} were chosen by maximizing their respective histogram functions. For example, $x_{min} = argmax(Hist(x_i))$, as shown in Fig. 5. These optimal x_{min}, y_{min}, z_{min} and x_{max}, y_{max}, z_{max} defined a 3D rectangle, which we used as the final estimate of the underlying 3D MBR, and provided a detection location for the target organ in this 3D CT scan, as illustrated in Fig. 5.

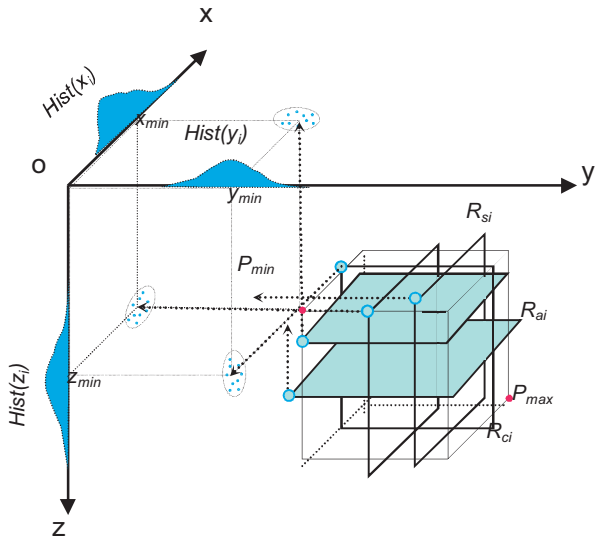

Fig. 5. Decision of 3D organ location $P_{min} = (x_{min}, y_{min}, z_{min})$ by majority voting with multi-slice (R_i), triple-direction (sagittal, axial, coronal) redundancy. Another point P_{max} can be decided by the same method.

3. Experiments

A database containing 660 patient cases (male, 344; female, 316; age range, 29–92 years) of 3D volumetric CT scans were used in our experiment. These 3D CT scans were generated in Gifu University Hospital by two different multi-slice CT scanners (LightSpeed Ultra16, GE Healthcare, and Brilliance 64, Philips Medical Systems). All CT scans were obtained using a common protocol (120 kV/Auto mA) and covered the entire human torso region. Each 3D CT scan had approximately 800–1200 axial CT slices with an isotropic spatial resolution of approximately 0.625 mm and a density (CT number) resolution of 12 bits. Among these 660 3D CT scans, 272 cases used contrast media for enhancement. All of these CT scans were taken for patients with certain real or suspicious abnormalities. Furthermore, the left kidney in 16 patients, the right kidney in 6 patients, and the spleen in 10 patients had been surgically removed.

The inner organs such as the lungs, heart, liver, spleen, pancreas, kidney, and colon are major targets during CT image interpretation. Besides the lungs and some parenchymal organs such as the colon, which can be identified by pre-filled air regions, most of the solid organs had similar intensities and indistinct contours that cannot be easily localized and segmented. In this paper, we selected such solid organs to validate the performance of the proposed approach. In fact, we selected the heart, liver, spleen, left kidney, and right kidney as localization targets in this experiment.

The 3D MBRs (P_{min}, P_{max}) of each target organ in each 3D CT scan were manually marked by an anatomist, author H.C. One hundred and one non-contrast 3D CT scans were randomly selected for training. Considering that one kidney was removed in some images, 91 3D CT scans were used for training the left kidney detectors and 97 3D CT scans for training the right kidney detectors. This led to 700–1800 positive 2D training samples and 10,000–25,000 negative 2D training samples that were used to train three 2D location detectors, each along the sagittal, coronal, and axial directions.

As mentioned above, a cascaded AdaBoosting algorithm [1] using Haar-like features [2] was applied for training the three 2D location detectors (see Fig. 2) using the positive and negative 2D training samples along the three different directions. The detectors for each organ were trained separately and independently. Each 2D location detector consisted of 10–15 cascades and each stage in the cascade was a "strong" classifier by combining (boosting) 10–30

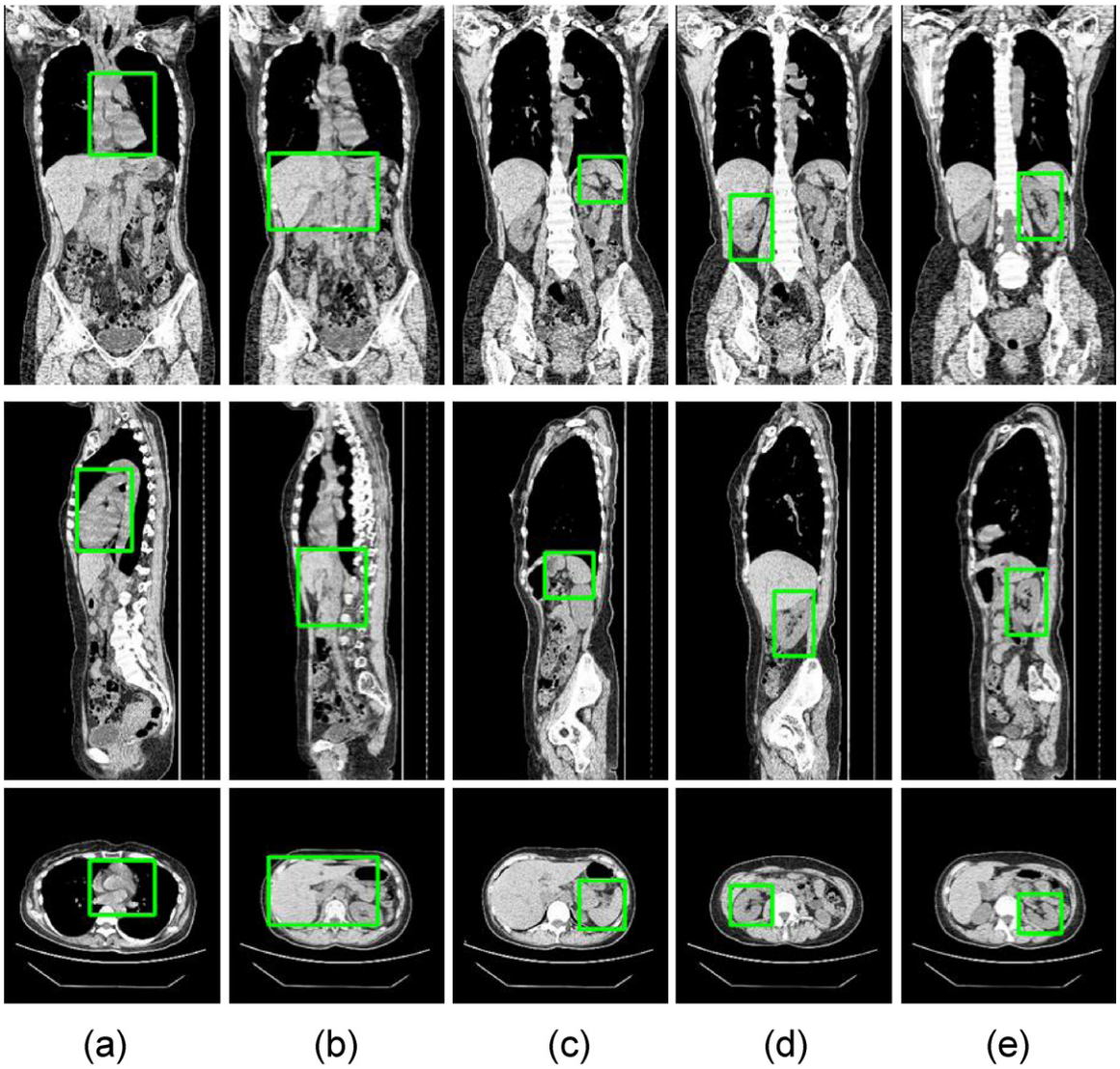

Fig. 6. An example of organ localization results. Three slices that pass through the detected center position of the (a) heart, (b) liver, (c) spleen, (d) right kidney, (e) left kidney in a CT scan are shown. Green box indicates the detected organ location (minimum bounding rectangle of the target organ). (For interpretation of the references to color in figure caption, the reader is referred to the web version of the article.)

weak classifiers. The window sizes of the 2D detectors in the x, y, and z directions were initialized to (42, 36, 52), (70, 50, 70), (50, 50, 50), and (46, 46, 76) voxels for scanning the heart, liver, spleen, and right/left kidney, respectively. In the voting process described in Step 8, a function returns a 5-point triangular window that was used for smoothing the histograms in three directions.

4. Results

The proposed approach was applied to detect the heart, liver, spleen, left kidney, and right kidney locations in 559 3D CT scans in the database that were not used for training. An example of the detection results in a 3D CT scan is shown in Fig. 6(a)–(e). The computing time for detecting each organ location was less than 15 s/3D CT scan using a computer equipped with an Intel Due2Core 2.23 GHz CPU. Accuracy of the organ localization was evaluated quantitatively by comparing the detected 3D rectangle with the ground-truth 3D MBR. The volume coinciding between these two 3D rectangles, the detected rectangle A and the ground-truth B (JSC: Jaccard similarity coefficient = $A \cap B / A \cup B$), and the Euclidean distance between the centers of these two 3D rectangles ($Dist_c$) were used as evaluation measures. The evaluation results using these two measures are shown in Figs. 7 and 8, respectively.

In this study, the detected location was considered to be correct if the detected 3D rectangle and the ground-truth MBR had a $JSC > 50\%$. Note that the criterion of $JSC > 0.5$ does not mean that we only require 50% of the volumes of the two involved rectangles to overlap. Fig. 9 shows two examples of heart and liver localizations that have the smallest JSC values in the experiment. We can see that some unsuccessful results with a $JSC < 50\%$ can also provide approximate locations of the target organs. The histogram of the JSC in Fig. 9 indicates that the heart locations in 99% (556/559), liver locations in 97% (540/559), right kidney locations in 91% (498/557), left kidney locations in 87% (466/553), and spleen locations in 75% (421/559) of the CT scans have been detected correctly. Organ detection completely failed ($JSC = 0$) in 5% (28/559) of the CT scans for spleen localization, 5% (28/559) of the CT scans for right-kidney localization, and in 6.6% (37/559) of the CT scans for left-kidney detection. The average and the standard deviation of JSC values for 559 test CT scans are shown in Table 1. Fig. 8 shows the histogram of $Dist_c$ values. We confirmed that the $Dist_c$ values of the heart in 97% (542/559) of CT scans, the liver in 74% (413/559) of CT scans, the

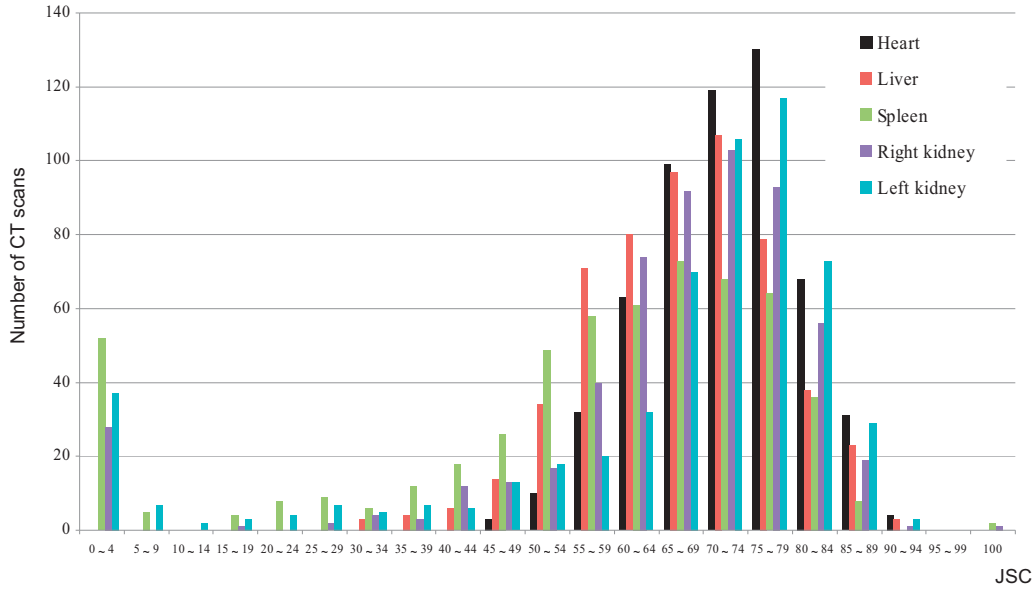

Fig. 7. Histogram of Jaccard similarity coefficient (*JSC*) between detected MBR (minimum bounding rectangle) and manual inputted MBR of the heart, liver, spleen, and left and right kidney regions using 559 test CT scans.

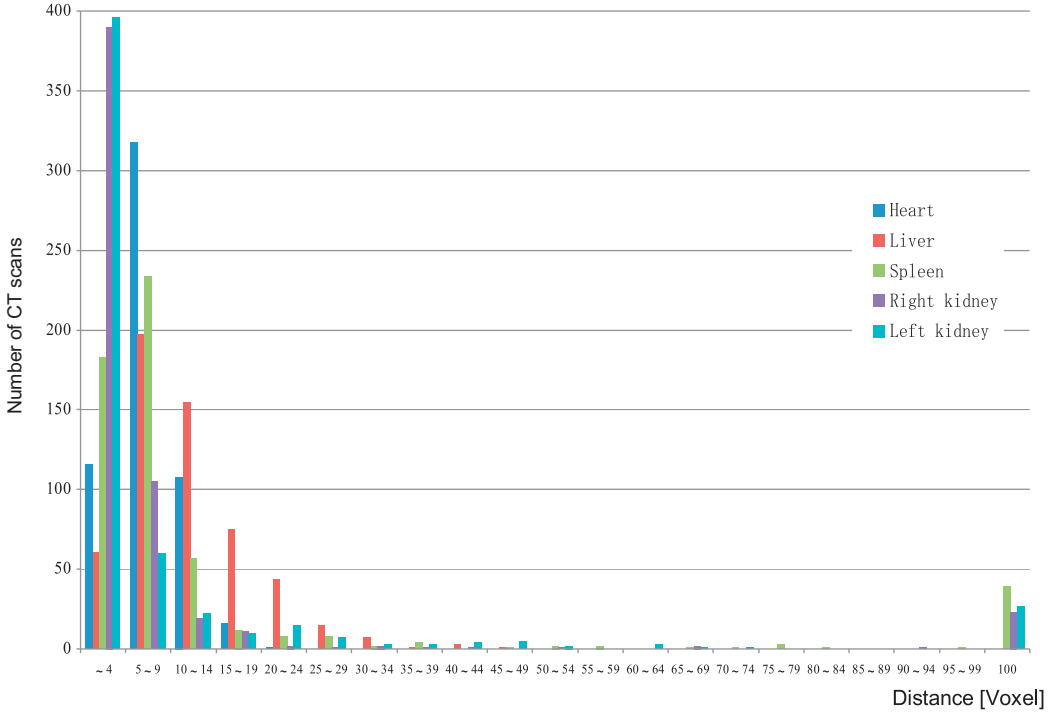

Fig. 8. Histogram of 3D Euclidean distances between the center position of detected MBR (minimum bounding rectangle) and manual inputted MBR of the 5 target organs using 559 test CT cases.

Table 1
Statistic summary for the *JSC* value (%) between the localization result and ground truth in 559 testing CT cases.

	Maximum	Minimum	Total 559 testing CT cases		Only the CT cases with *JSC* > 50%	
			Average	Standard deviation	Average	Standard deviation
Heart	94.2	49.1	72.5	8.4	72.6	8.2
Liver	96.3	29.9	72.2	10.9	73.2	8.7
Spleen	100	0	57.1	23.6	68.4	9.5
Left kidney	93.4	0	64.1	22.9	73.0	8.7
Right kidney	90.3	0	65.9	17.6	70.8	8.5

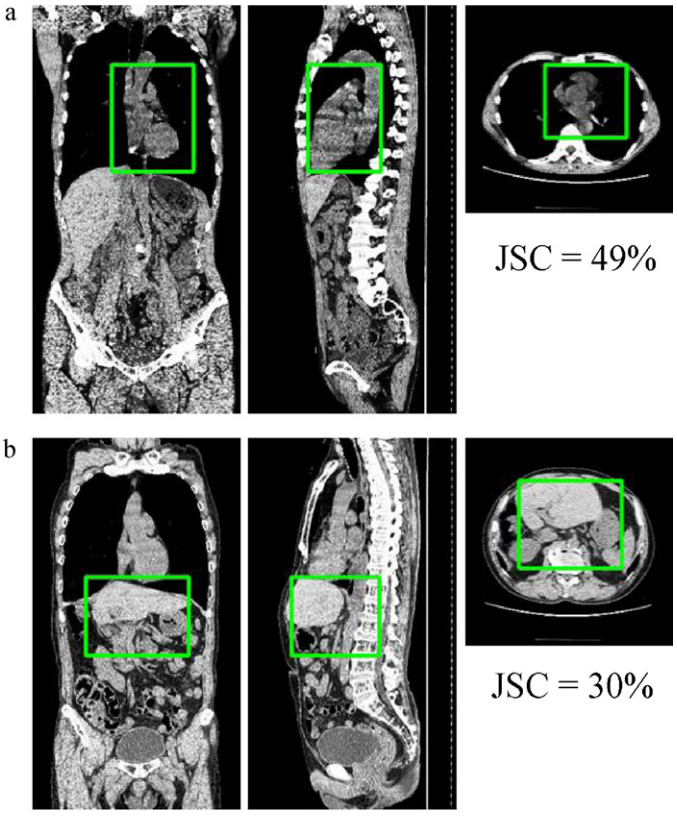

Fig. 9. Localization results of the heart (a) and liver (b) that showed the worst accuracies (*JSC*) in the experiments. Green box indicates the detected organ location (minimum bounding rectangle of the target organ). (For interpretation of the references to color in figure caption, the reader is referred to the web version of the article.)

spleen in 85% (474/559) of CT scans, the left kidney in 86% (478/553) of CT scans, and the right kidney in 92% (514/557) of CT scans were distributed within 15 voxels. We also applied the organ localization algorithm to the 101 training samples, and found that the performance (*JSC*) was slightly better than those results obtained on the test samples. A histogram of *JSCs* for the left and right kidney localizations on the training samples is shown in Fig. 10. Considering the large variation of the solid organs, this performance should be acceptable to solve organ localization problems in CT scans. Specifically, we found that most solid organs with a normal appearance in our database were localized successfully.

According to the assumption that the anatomical structures in different patients are almost identical and that patients are always centered in the images during the CT scan, there are some arguments that organ location may be simply estimated by using a predefined atlas (average of organ locations in a number of patients). Therefore, we compared the localization results of our approach with the ones from an atlas-based approach. The atlas used in our experiments was generated by averaging the center positions of the ground truths (MBRs of five target organs) in 559 test CT scans after normalization based on a rigid registration (manually rescaling and translating the MBR of torso region in different patients into a standard 3D-box that starts from the apex of the lung region in an axial direction). By manually re-scaling and translating this standard 3D-box to fit the torso region in an unknown CT scan based on the same rigid registration method, the center positions of the heart, liver, spleen, and left and right kidneys can be predicted by the atlas inside the 3D box. We applied this atlas-based approach to the same 559 CT scans used in our experiments. The histogram of $Dist_c$ values of heart, liver, spleen, and left and right kidney localizations are shown in Fig. 11. We found that the errors of estimated center positions of each target organ ranged from 0 to 55 voxels with an average of about 25 voxels. This result shows that the organ locations in CT scans cannot be accurately predicted by a predefined atlas. Using the same measure in as Fig. 8, we found that the localizations resulting from the proposed approach are clearly better than the results from the atlas-based approach.

In this experiment, we only used non-contrast CT scans for training. We found that the obtained organ location detectors also worked well for contrast-enhanced CT scans. There was no apparent deterioration between the detection performance on non-contrast and contrast-enhanced CT scans. We believe this is partially indebted to the adopted weak classifier based on Haar-like features, which is insensitive to the contrast increase between the target organ and the surrounding background.

The time required for organ localization was less than 15 s for each torso CT scan, and this time can be further reduced to about 6 s by parallel computing using the 2D detectors all on a PC with a quad-core CPU (Intel Core i7 975, 3.33 GHz) and 4 GB memory. The algorithms were developed by using C++ under the CentOS 5.0.

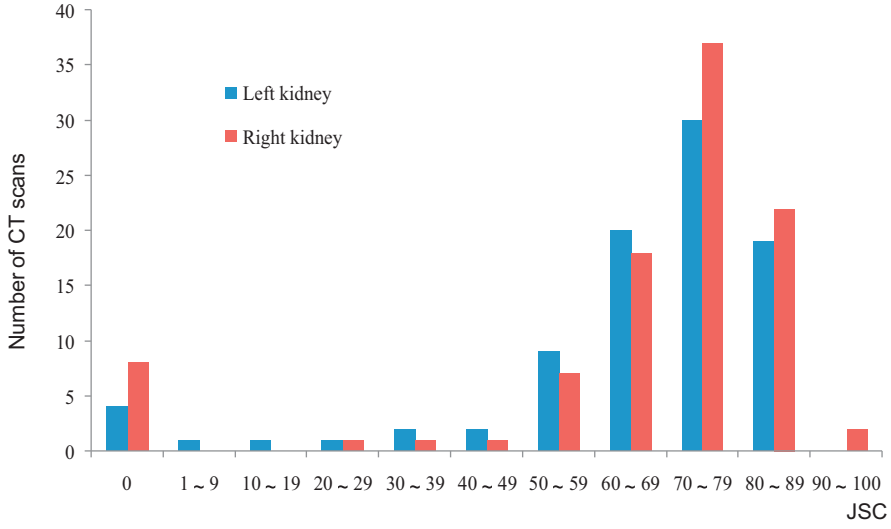

Fig. 10. Histogram of Jaccard similarity coefficient (*JSC*) between detected MBR (minimum bounding rectangle) and manual inputted MBR of the left and right kidney regions using 101 training CT scans.

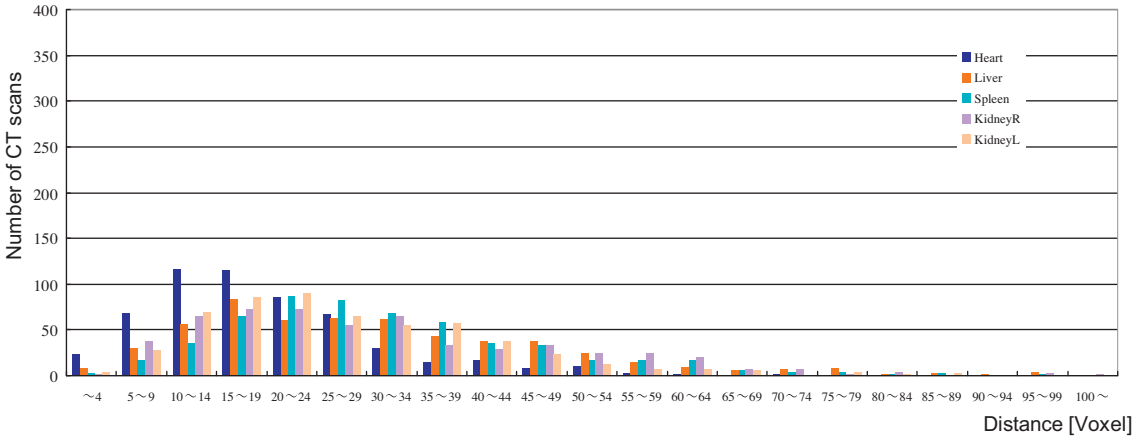

Fig. 11. Histogram of 3D Euclidean distances between the center position of estimated MBR (minimum bounding rectangle) using an atlas and manual inputted MBR of the 5 target organs using 559 test CT cases.

5. Discussion

The major contribution of this paper is the use of ensemble learning for 2D location detection along three different directions, followed by the integration of the 2D detection results to estimate the 3D organ location. We only required the 2D detectors to be "weak" detectors. Majority voting of multiple 2D candidates from three independent directions can combine all weak 2D detectors into a "strong" 3D 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 had a very low detection accuracy. Based on our experiments on the 559 3D CT scans that were not used for training, 55% and 37% (for the left and right kidneys, respectively) of the 2D rectangles detected in Step 2 of the algorithm described in Section 2.3 are outliers. On the other hand, the 2D location detectors on the corresponding slices did not find 65% and 79% (for left and right kidneys, respectively) of the ground-truth 2D MBRs. This is completely reasonable since solid organs such as kidneys may show an inconsistent intensity and appearance in different 3D CT scans, and the contrast between the solid organ and 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.

The motivation of using multiple 2D detectors instead of a 3D detector was due to the large number of 3D training samples required for direct 3D detection, which are difficult to obtain. Particularly, the organs (e.g., kidney) in clinical CT images may be abnormal, resulting in an irregular 3D appearance in part of this organ (refer to Fig. 12). Therefore, we would need an even greater number of training samples to cover all possible appearance irregularities in 3D. Our 2D approach addresses this problem by (1) having a large number of 2D training samples, and (2) in abnormal cases, detecting a partially normal appearance in 2D sections (good parts in "rotten apple" as shown in Fig. 12((b)–(d)) and then deriving the 3D bounding box (Fig. 12(a)) using such normal 2D sections. Another benefit of 2D detections is that independent 2D detections are more flexible than 3D CT scans with different resolutions along the axial body direction. By adjusting the resolution of the 2D detecting windows, we found that our approach could be applied to CT scans with 2–5 (mm) slice spacing in the axial direction to identify organ locations.

In related work, probabilistic boosting tree with 3D Haar-like features has been used to detect heart location as the preprocessing step for heart structure segmentation [3]. This method was successfully trained and tested using 457 CT scans from 186 patient cases by a four-fold cross validation. To ensure that the robustness and accuracy of detection can be maintained, this study outputted 100 potential locations of a target organ as the final result in a CT case. Our approach accomplishes the same detection task in a lower-dimensional feature space (2D Haar-like feature). In addition, our approach uses a smaller number of training samples (101) and was validated on a larger number (559) of untrained CT scans under more stringent requirements that our approach indicated the organ location itself exactly in a CT case. Recently, an approach based on decision forests with a 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 a mean localization error ($Dist_c$) of 21.32 mm for the heart, 22.68 mm for the liver, 25.42 mm for the left kidney, and 44.52 mm for the right kidney. Compared to [6], our approach is much simpler and only needs to use local features within the extent of the target organ. Considering that CT images do not always cover an extensive area of the

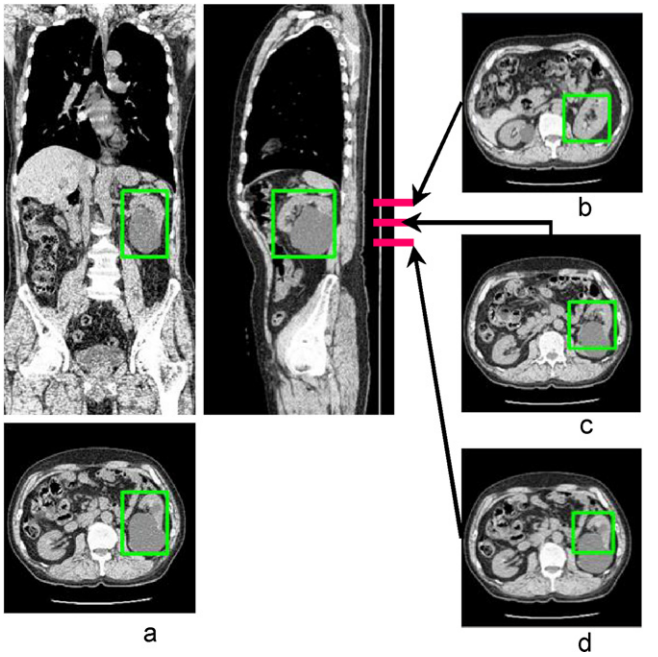

Fig. 12. An example of left kidney localization result in an abnormal CT scan that did not appear in training samples. (a) Three slices that pass through the final detected center position. (b)–(d) A part of 2D detection results in axial slices.

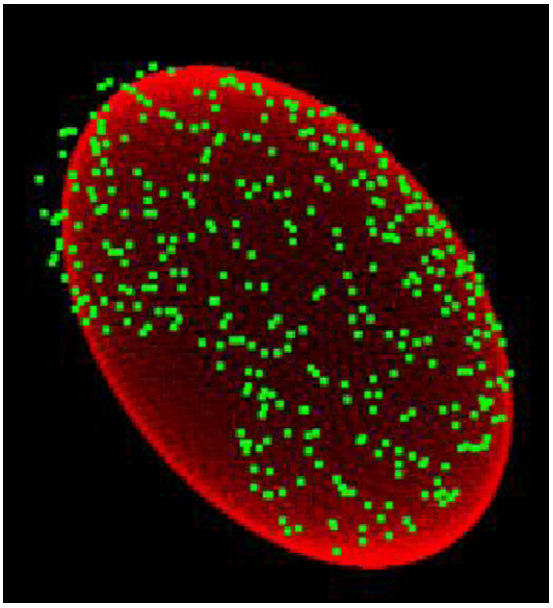

Fig. 13. A 3D view of a left kidney segmentation result in a CT scan. Red: a 3D ellipse that automatically arranged on the basis of kidney localization result. Green: sample points of a left kidney surface that manually inputted by a medical expert. (For interpretation of the references to color in figure caption, the reader is referred to the web version of the article.)

human body, the use of our approach in clinical practice is more realistic. Additionally, our approach produces smaller localization errors ($Dist_c$) for such organ localizations in most CT scans.

The major application of organ localization is to help organ segmentation in CT images. We conducted a preliminary investigation to demonstrate the usefulness of the obtained localization results for the respective organ segmentation. We selected the left kidney as the segmentation target, and put a 3D inscribed ellipse with a fixed orientation inside the detected 3D rectangle as an estimate of the desired left kidney boundary. We manually identify approximately 200–400 sample points on the boundaries of the left kidney in 90 non-contrast CT scans ($JSC > 50\%$) to evaluate accuracy. Fig. 13 shows a 3D view of an inscribed ellipse and sample points on a left kidney boundary in a CT scan. The minimum Euclidean distance $Dist_s$ from each sample point to the ellipse were measured. We found that the average $Dist_s$ value of all sample points were distributed from 6 to 23 voxels as shown in Fig. 14. Obviously, simple ellipses cannot cover the variations in kidney orientation and shape from different patients; however, these results indicate the potential usefulness of the organ localization for supporting organ segmentation based on CT images. Further development in the universal organ segmentation method will be done on the basis of our localization approach in the future.

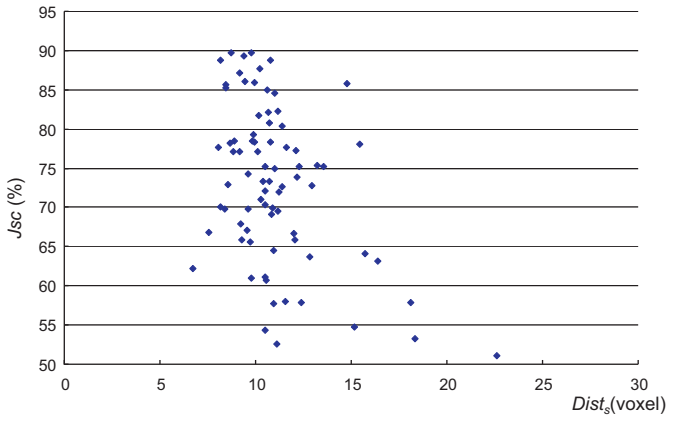

Fig. 14. Relation between localization accuracy Jaccard similarity coefficient (*JSC*) values and segmentation accuracy *Dists* (average value of the minimum Euclidean distance from a left kidney surface to a pre-arranged ellipse) values in 90 non-contrast CT scans.

6. Conclusion

We proposed a universal approach for automatic localization of different solid organs on 3D CT scans. In this approach, we used a collaborative majority voting decision based on 2D ensemble learning for quickly localizing the organs. This approach was applied to a database that included 660 CT scans to localize the heart, liver, spleen, and the left and right kidneys. We performed experiments to validate the efficiency and accuracy of this method. In the future, we plan to combine the proposed approach with a probabilistic atlas-based approach [8] to provide a universal tool for segmentation of all the solid organs and for recognition of the anatomical structures in 3D CT images.

Acknowledgments

The authors would like to thank the other members of Fujita Lab for their valuable discussion. This research work was funded in part by a Grant-in-Aid for Scientific Research on Innovative Areas, MEXT, Japan, in part by Grant-in-Aid for Scientific Research (C23500118), MEXT, Japan, and in part by the Ministry of Health, Labour, and Welfare under a Grant-In-Aid for Cancer Research, Japan.

Appendix A. 2D Haar-like features [2]

We assume the basic unit for detecting an object is a window of $W \times H$ pixels. A rectangle inside of the window is specified as $r = (x, y, w, h, a)$ with $0 < x, x + w < W$, $0 < y, y + h < H$, $x, y > 0$, $w, h > 0$, and $a = \{0, 45^\circ\}$, where the x, y are the coordinates of the left-upper position, w, h is the size, and a is the rotation angle of the rectangle (Fig. 3). Let us use $RecSum(r)$ to denote the pixel sum of r. The 2D Haar-like feature is defined as the set of all possible features of the form

$$feature_I = \sum_{i \in I = \{1, \cdots N\}} w_i \times RecSum(r_i),$$

where the weights $w_i \in \Re$, the rectangles r_i, and N are arbitrarily chosen.

The feature set is almost infinitely large. For practical reasons, it is reduced by some constraints (for example, only two rectangles were used, the weights have opposite signs, and weight values were calculated based on the area of two rectangles) and condensed into 14 feature prototypes to show the edge, line, and center-surround features that are closed to the human visual pathway. Further techniques for generating the 2D Haar-like features can be found in Ref. [2].

Appendix B. AdaBoosting algorithm [1]

We show the pseudo codes of AdaBoosting algorithm [1] that used in this work for training a strong classifier. Each round of boosting selects one feature from a feature pool.

1. Input training sample images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for indicating negative (background) and positive examples (target organ), respectively.

2. Initialize weights $w_{1,i} = 1/(2m)$, $1/(2l)$ for $y_i = 0$, 1, respectively, where m and l are the number of negatives and positives, respectively.
3. For $t = 1,\ldots, T$:
   - Normalize the weights, $w_{t,i} \leftarrow w_{t,i}/\sum_{j=1}^{n} w_{t,j}$.
   - For each feature, j, train a classifier h_j which is restricted to using a single feature.
   - The error is evaluated with respect to w_t, $\varepsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
   - Select the classifier, h_t, with the smaller error.
   - Update the weights: $w_{t+1,i} = w_{t,i}\beta_t^{1-e_i}$, where $e_i = 0$ if example x_i is classified correctly, otherwise $e_i = 1$, and $\beta_t = \varepsilon_t/(1-\varepsilon_t)$.
4. The final strong classifier is: $h(x) = \begin{cases} 1 & \sum_{t=1}^{T}\alpha_t h_t(x) \geq \frac{1}{2}\sum_{t=1}^{T}\alpha_t \\ 0 & \text{otherwise} \end{cases}$, where $\alpha_t = \log(1/\beta_t)$.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.compmedimag.2011.12.004.

References

[1] Viola P, Jones MJ. Rapid object detection using a boosted cascade of simple features. Proc Intl Conf on Computer Vision and Pattern Recognition (CVPR) 2001:511–8.
[2] Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection. Proc Intl Conf on Image Processing (ICIP) 2002;1:900–3.
[3] Zheng Y, Adrian B, Bogdan G, Michael S, Dorin C. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans Med Imaging 2008;27:1668–81.
[4] Ling H, Zhou SK, Zheng Y, Georgescu B, Suehling M, Comaniciu D. Hierarchical, learning-based automatic liver segmentation. Proc Intl Conf on Computer Vision and Pattern Recognition (CVPR) 2008:1–8.
[5] Dikmen M, Zhan Y, Zhou XS. Joint detection and localization of multiple anatomical landmarks through learning. Proc SPIE Medical Imaging 2008;6915:6915381–9.
[6] Tu Z. Probabilistic boosting-tree: learning discriminative methods for classification, recognition, and clustering. Proc Intl Conf on International Conference on Computer Vision (ICCV) 2005:1589–96.
[7] Criminisi A, Shotton J, Bucciarelli S. Decision forests with long-range spatial context for organ localization in CT volumes. In: Proc on intl conf on medical image computing and computer assisted intervention workshop on probabilistic models for medical image analysis (MICCAI-PMMIA). 2009.
[8] Zhou X, Kitagawa T, Hara T, Fujita H, Zhang X, Yokoyama R, et al. Constructing a probabilistic model for automated liver region segmentation using non-contrast X-ray torso CT images. Proc Intl Conf on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2006;4191:856–63.