

# Ensemble segmentation using AdaBoost with application to liver lesion extraction from a CT volume

Akinobu SHIMIZU\* Takuya NARIHIRA\* Daisuke FURUKAWA\*  
Hidefumi KOBATAKE\* Shigeru NAWANO\*\* and Kenji SHINOZAKI\*\*\*

\*Tokyo University of Agriculture and Technology

\*\*Center for Radiological Sciences, International University of Health and Welfare

\*\*\*National Kyusyu Cancer Center

E-mail : simiz@cc.tuat.ac.jp

**Abstract.** This paper describes an ensemble segmentation trained by the AdaBoost algorithm, which finds a sequence of weak hypotheses, each of which is appropriate for the distribution on training example, and combines the weak hypotheses by a weighted majority vote. In our study, a weak hypothesis corresponds to a weak segmentation process. This paper shows a procedure for generating an ensemble segmentation algorithm using AdaBoost, and applies it to a liver lesion extraction problem from a contrast enhanced abdominal CT volume. A leave-one-patient-out validation test using 16 CT volumes demonstrated the effectiveness of the generated ensemble segmentation algorithm. In addition, we evaluated the performance by applying the algorithm to unknown test data provided by the “3D Liver Tumor Segmentation Challenge 2008”.

**Keywords:** ensemble segmentation, AdaBoost, CT volume, liver, lesion extraction, metastasis

## 1 Introduction

Extraction of a lesion, such as those seen in liver metastasis, from a given image is an essential process for computer-aided diagnosis and computer-aided surgery. Although many algorithms for liver lesion extraction have been proposed [1, 2], they are not accurate, because the characteristics of lesions are often similar to those of the surrounding normal tissues. For example, the CT value of metastases in liver is sometimes identical to that of the gallbladder, and the tumor necrosis shows CT values similar to those of fat, resulting in false positives. One effective approach is extracting the organ before lesion extraction. However accuracy of the lesion extraction then depends on that of organ segmentation. If a part of an organ affected by a lesion is misidentified as another organ, because of the difference in CT values and shapes from that of a normal liver (failure examples can be found in [3]), the lesion cannot be extracted by a subsequent process. Hence, a lesion extraction algorithm that can extract lesions directly from a given image is needed.

In addition to the problem caused by surrounding tissues, there is a wide variation in lesion characteristics, which complicates the lesion extraction task. To cope with the variation of the targets, some researchers have focused on combinations of multiple segmentations, each of which can deal with a certain type of target. Thus far, combinations of multiple segmentations, using vote rule decision fusion, has been found to reduce random errors in the segmentation, resulting in high segmentation accuracy. Such techniques have been applied successfully for segmentation of bee brains in confocal microscopy images [4, 5], and human brains in MR images [6, 7].

In the field of pattern recognition, boosting is an effective method of producing a very accurate classification rule. AdaBoost [8] is a well-know boosting algorithm that calls a weak classifier repeatedly in the series of learning rounds using a training dataset and constructs a strong ensemble classifier as a linear combination of the weak classifiers. It can be applicable to the segmentation problem, but less attention has been paid to the use of the AdaBoost algorithm to generate an ensemble segmentation process.

This paper aims to present an ensemble segmentation trained by AdaBoost. The main contribution of the paper is twofold. First, this paper shows a procedure for generating an ensemble segmentation algorithm using AdaBoost. Second, we applied the proposed procedure to the liver lesion extraction problem and generated an ensemble segmentation algorithm. We performed a leave-one-patient-out test on contrast enhanced 3D CT volumes taken from 16 patients to validate the generated ensemble segmentation algorithm. Moreover, an ensemble segmentation algorithm was trained using 16 CT volumes as well as 4 volumes provided by the “Liver Tumor Segmentation Challenge 2008”, and the generated segmentation algorithm was then applied to the unknown test dataset consisting of 6 volumes given by the competition.

## 2 Method

### 2.1 Ensemble segmentation by AdaBoost

Consider a segmentation that distinguishes two different classes in a label set  $\Lambda = \{-1, 1\}$ . The segmentation process of a three dimensional image includes mapping from the coordinates  $\mathbf{x}$  to the labels  $h_t: R^3 \rightarrow \Lambda$ . Pseudocode for AdaBoost is provided in figure 1. AdaBoost finds an optimum weak segmentation process  $h_t$  repeatedly in a series of training rounds given by  $t = 1, \dots, T$ , which minimizes a weighted error  $\varepsilon_t$  computed from the training dataset. Finally the sequence of the weak segmentation process is integrated into  $H(\mathbf{x})$  with weight  $\alpha_t$ , which intuitively measures the importance assigned to  $h_t(\mathbf{x})$ , and  $\alpha_t$  increases as error  $\varepsilon_t$  of  $h_t(\mathbf{x})$  decreases. The advantages of the AdaBoost algorithm are that it is very simple to implement and can maximize the margin, resulting in a high generalization performance. The weak segmentation processes employed in this study will be explained in the next subsection.

Given  $(\mathbf{x}_1, \lambda_1), \dots, (\mathbf{x}_m, \lambda_m)$  where  $\mathbf{x}_i \in \mathbf{X}$ ,  $\lambda_i \in \Lambda = \{-1, 1\}$

Initialize  $D_1(i) = 1/m$

For  $t = 1, \dots, T$

1. Train weak segmentation processes using distribution  $D_t$ .
2. Select optimum weak segmentation  $h_t$  with minimum error

$$\varepsilon_t = \Pr_{i \sim D_t} \{ h_t(\mathbf{x}_i) \neq \lambda_i \}$$

3. Choose  $\alpha_t = 0.5 \ln \{ (1 - \varepsilon_t) / \varepsilon_t \}$

4. Update distribution  $D_t$ .

$$D_{t+1}(i) = D_t(i) \exp \{ -\alpha_t \lambda_i h_t(\mathbf{x}_i) \} / Z_t$$

where  $Z_t$  is a normalization factor.

Output the final segmentation  $H(\mathbf{x})$ :

$$H(\mathbf{x}) = \text{sign} \{ \sum_{t=1, \dots, T} \alpha_t h_t(\mathbf{x}) \}$$

**Fig. 1.** A boosting algorithm for ensemble segmentation based on AdaBoost.

## 2.2 Weak segmentation process

Performance of a generated ensemble segmentation depends on weak segmentation process considered in the AdaBoost learning process. In this study, we focused on a liver lesion segmentation problem (Figure 2) and prepared 54 weak segmentation processes, each of which uses a feature listed in Table 1.



**Fig. 2.** An example of lesions in a liver CT volume. Arrows show a metastasis in the liver.

According to the following equation, the weak segmentation process determines if a voxel of interest belongs to a target lesion.

$$\begin{aligned} f > Th &\Rightarrow \textit{lesion} \\ \textit{else} &\Rightarrow \textit{others} \end{aligned} \quad (1)$$

where  $f$  is a value of a feature listed in Table 1. Note that if segmentation error  $\varepsilon_t$  is greater than 0.5,  $\alpha_t$  becomes negative, which reverses the inequality sign of the equation. In each training round of figure 1, the threshold of weak segmentation

process was trained so that the error computed on the training dataset would be minimum, and an optimum weak segmentation  $h_i$  with minimum error was selected in step 2.

**Table 1.** Features used in the weak segmentation processes.

	Features	Parameters
CT value based features computed from a local region	Average, variance, skewness and kurtosis of CT values	Size of local region: $3 \times 3 \times 3, 5 \times 5 \times 5, 7 \times 7 \times 7$
	Minimum and maximum of CT values	Size of local region: $7 \times 7 \times 7, 11 \times 11 \times 11$
	Output of sequential filter; Minimum filter $\Rightarrow$ Maximum filter	Size of the filter : $11 \times 11 \times 11$
	Average of CT values in spherical region	Radius of the region: Estimated size of lesion by convergence index filters
Features based on 3D convergence index filter (C.I.)	C.I.	$d = 3, r = 5 \sim 33, M = 62^*$
	C.I. + diffusion filter	$d = 3, r = 5 \sim 33, M = 104^*$
	C.I. with absolute operation	
	C.I. with absolute operation + diffusion filter	
Other features	Norm of sobel filter	
	Output of matched filter	Size of the filter: Estimated size of lesion by convergence index filters

\* $d$  and  $r$  will be explained in figure 3, and  $M$  indicates the number of directions to evaluate gradient vectors in the support region.

The features are divided into three groups. The first group contains CT value based features, measured in a region in the neighborhood of the voxel of interest, such as the average and variance. All features in the first group are derived from an original CT image as well as a normalized CT image. The normalization is performed as follows.

$$f' = (f - \mu) / \sigma \quad (2)$$

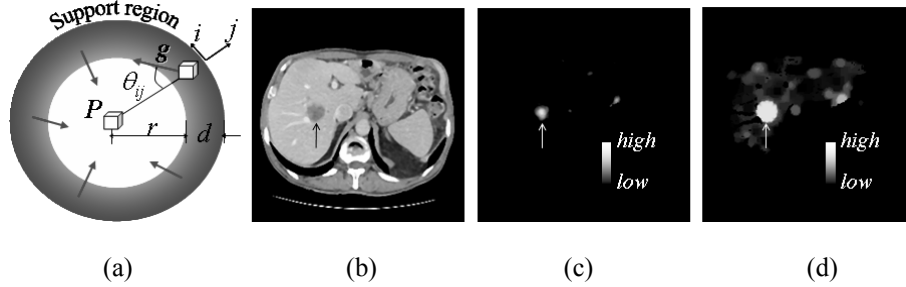
where  $f'$  is the normalized value of  $f$ .  $\mu$  and  $\sigma$  are the average and standard deviations of CT values in the liver, estimated for each image by an EM algorithm without segmentation. This normalization can reduce the variance of CT values among patients, resulting in increased separability between the lesion and other tissues.

Features in the second group are based on the output of a 3D point convergence index filter developed for enhancing circumscribed lesions [9]. This filter calculates

the convergence index (see equation (3)) of gradient vector  $\mathbf{g}$  in a spherical region of figure 3, called the support region. It outputs the index to the center  $P$ .

$$[\text{Convergence Index}] = \max_r \{ \sum_{i=0}^{M-1} \sum_{j=r+1}^{r+d} (-\cos \theta_{ij}) \} \quad (3)$$

This filter is robust with respect to changes in contrast, and effective in enhancing cancers with low contrast, which are difficult to detect by a threshold-based method. An additional advantage is that, during computation of the convergence index, it estimates the size of the lesion, which can be used by other processes, as shown in Table 1 and figure 3.



**Fig. 3.** Illustrations for the point convergence index filter. Here, (b) is an original CT image with a metastasis indicated by an arrow, (c) shows the filter output, the bright spots indicate a higher convergence index. (d) is a result of the filter followed by a modified diffusion filter that spreads the outputs of the convergence index filter within a limited area, whose size matches the estimated lesion size using the convergence index filter.

Although the convergence index filter outputs high values around the center of the lesion, as shown in figure 3 (c), the output rapidly decreases near the periphery of the lesion. Thus, a weak segmentation process based on thresholding is prone to underextract the peripheral region. To avoid this, we combined the filter with an isotropic diffusion filter that spreads the outputs of the convergence index filter within a limited area, whose size is equal to the convergence index filter's estimate of the lesion size. As shown in figure 3 (d), the diffusion process enhances the entire domain. The convergence index filter with absolute operation (Table 1) computes the absolute value of the cosine function in equation (3) to enhance the lesion neighboring the lung, air in necrosis, or visceral fat area, whose CT value is lower than the lesion.

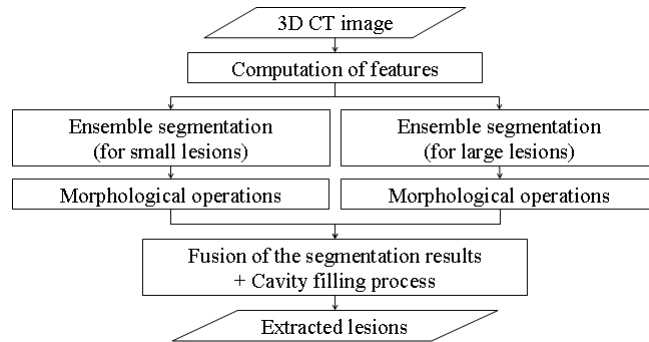
The last group consists of other features, such as the output of a sobel filter and a matched filter, computed from an original and a normalized CT image using equation (2). The matched filter was designed in terms of CT value profile of the lesion.

Total number of features is 54, and all features are converted to binary values, according to equation (1) to extract lesions during the AdaBoost training round.

### 2.3 Liver lesion extraction algorithm based on ensemble segmentations

Figure 4 shows a flowchart of the proposed lesion extraction algorithm from a given CT volume. First, it computes the features of Table 1 in the area where the prior probability of the liver is greater than 0. Note that a prior probability was estimated from a training dataset using an atlas based method [10], and the area used is sufficiently large to avoid underextraction of a liver with lesions. Second, the ensemble segmentations are performed. In this study, two ensemble segmentation algorithms were created to treat lesions of different sizes. We divided the training dataset into two sets according to lesion size (empirically less than 30,000 voxel or else), and the two ensemble segmentation processes were independently trained by the procedure proposed in 2.1. Note that there are differences in selected weak segmentation processes between the two ensemble segmentation algorithms.

For post-processing, the segmentation results were refined by morphological operations such as opening and closing, and integrated into one by a logical sum operation, followed by a cavity filling process. It should be noted that all processes are automated.



**Fig. 4.** A Liver lesion extraction algorithm. Features were measured from a given image and forwarded to the ensemble segmentation processes. Results of the ensemble segmentations were refined by morphological operations and integrated.

## 3 Experiments

### 3.1 Materials and a quantitative metric for validation test

In the first experiment, we used the data distributed at MICCAI2007 workshop [3]. The datasets includes 16 CT volumes with 34 lesions, volumes of which are distributed from 0.4 to 400 *mL*. The images were obtained using multi-detector row CT scanners, and provided several hundreds of  $512 \times 512$  pixel slice images, with 12-

bit accuracy. The voxel size of the CT volume in an axial plane is 0.543~0.865 mm, and 0.5~5 mm in the z direction.

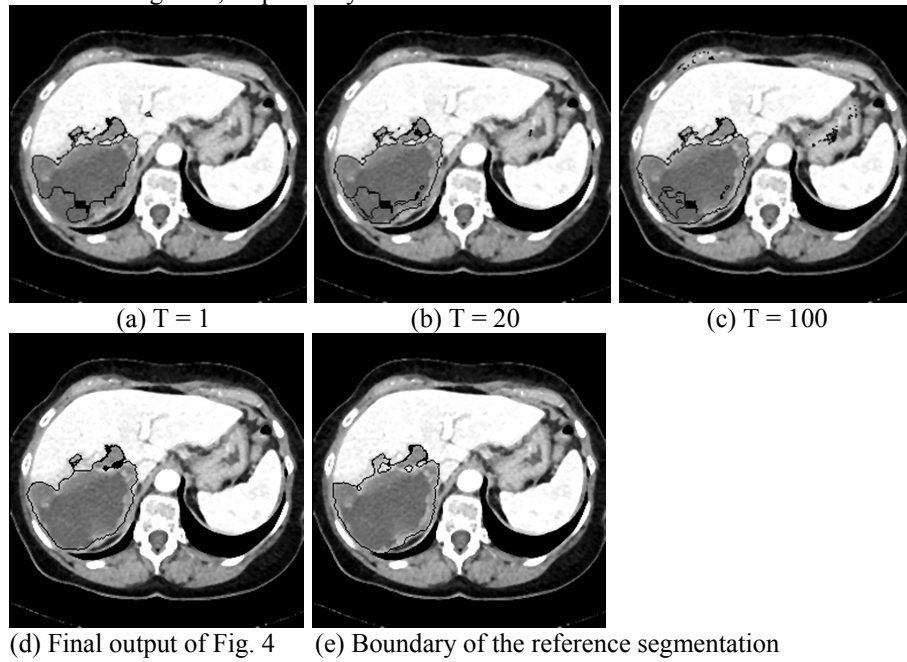
The Jaccard index between an extracted tumor region and its corresponding true region, manually segmented by the authors and cited as reference segmentation below, was adopted to evaluate the segmentation performance.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

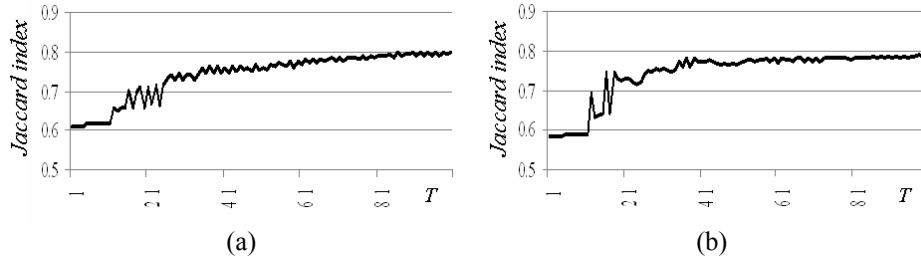
where A is a set of extracted voxels, and B is a set of voxels in the reference segmentation.

### 3.2 Validation by leave-one-patient-out test using 16 CT volumes

Leave-one-patient-out test using the 16 volumes was performed to validate the ensemble segmentation. Extracted lesions from test data with different numbers of weak segmentation processes are shown in figure 5. Figure 6 shows the relationship between the Jaccard index and the number of weak segmentation processes. The left and right graphs show the results for a training dataset of 15 patients, and those for the test data of figure 5, respectively.



**Fig. 5.** Results for validation (test) data. The boundaries are black. Figures (a), (b), and (c) show the results for  $T = 1$ , 20, and 100, respectively. (d) is the final result, which was very similar to the true boundary shown in (e). The small false positives and cavity in figure (c) were eliminated by the morphological operations followed by the cavity filling process.



**Fig. 6.** Jaccard index versus the number of weak segmentation processes in the ensemble segmentation algorithm. Here, (a) shows the average index for a training dataset from 15 patients, and (b) is the result for the validation data of figure 5.

### 3.3 Validation using testing data and quantitative metrics of the “Liver Tumor Segmentation Challenge 2008”

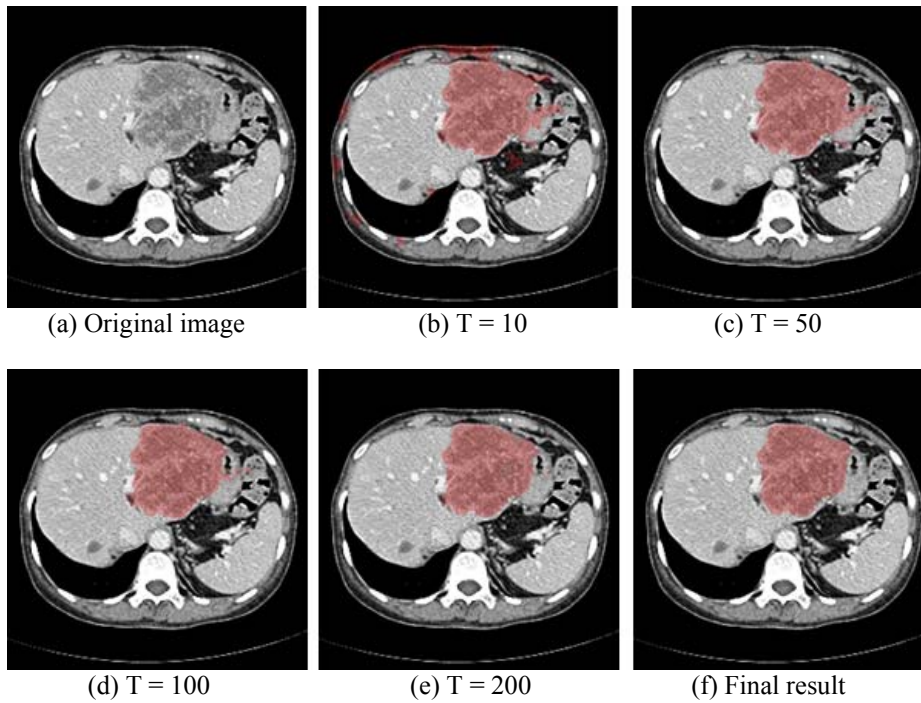
To evaluate the performance for unknown data, we trained an algorithm using the 16 CT volumes as well as 10 tumors in 4 CT volumes provided by an organizer of “Liver Tumor Segmentation Challenge 2008”, and then applied to unknown 4 testing CT volumes including 10 tumors given by the organizer.

Note that there are some modifications from the above experiment to deal with problems found when using the new training data. For example some of the CT value based features were improved to extract tumors in the training dataset precisely. The error  $\varepsilon_i$  in the figure 1 was computed from the neighboring region of a tumor instead of whole liver region, because the competition focuses on the accuracy of extracted boundaries but not false positives extracted separately from the tumor. The maximum number of weak segmentation processes was altered to 200 aiming to reduce the error of the training data. In addition, to detect not only metastases but also Hepatocellular Carcinoma (HCC) and Hemangioma (HEM) that are enhanced by contrast medium, we designed an additional ensemble segmentation algorithm dedicated to such contrast enhanced tumors using our own dataset including HCCs and HEMs [2]. The output was automatically integrated with that of figure 4 by a logical sum operation. Note that no tumor for test was used in the training phase and the whole segmentation processes are automated.

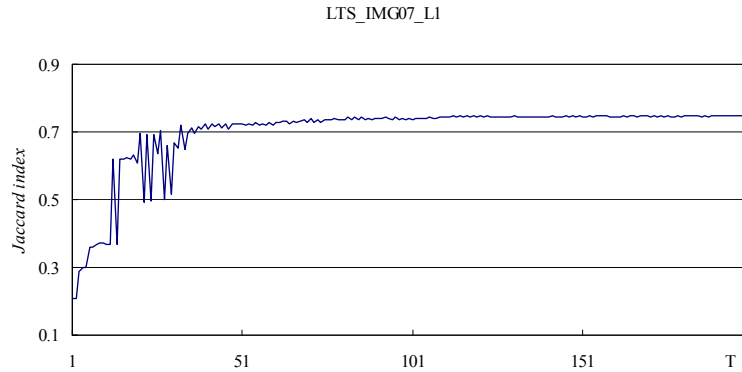
The results were evaluated quantitatively by the organizer of the competition with respect to the following five criteria, (1) volumetric overlap, (2) relative absolute volume difference, (3) average symmetric absolute surface distance, (4) symmetric RMS surface distance, and (5) maximum symmetric absolute surface distance. For more details for these criteria, refer to [11]. Table 2 summarized the performances for all testing data. Figure 7 shows a sequence of the extracted tumor regions corresponding to the tumor IMG07\_L1, in which the regions were improved as the number of weak segmentation processes  $T$  increased. Figure 8 indicates the change of the Jaccard index ( $= 1 -$  “volumetric overlap error in Table 2”) of the extracted tumor regions with the reference segmentation that was manually delineated by the authors after the evaluation process by the organizer.

**Table 2.** Results of the comparison metrics and scores for all ten tumors.

Tumor	Overlap Error		Volume Difference		Ave. Surf. Dist.		RMS Surf. Dist.		Max. Surf. Dist.		Total Score
	(%)	Score	(%)	Score	(mm)	Score	(mm)	Score	(mm)	Score	
IMG05_L1	34.89	73	4.83	95	3.17	20	3.82	47	10.58	74	62
IMG05_L2	32.46	75	19.89	79	1.18	70	1.51	79	4.51	89	78
IMG05_L3	30.30	77	30.77	68	1.36	66	2.18	70	8.14	80	72
IMG06_L1	39.18	70	22.72	76	1.22	69	1.57	78	4.55	89	76
IMG06_L2		23		0		0		0		0	5
IMG07_L1	20.67	84	8.27	91	2.52	36	3.25	55	13.13	67	67
IMG07_L2	29.10	78	35.65	63	1.54	61	1.90	73	5.46	86	72
IMG08_L1	29.58	77	28.46	70	3.54	11	4.34	39	13.28	67	53
IMG09_L1	31.42	76	13.37	86	1.10	72	1.62	77	7.28	82	79
IMG10_L1	13.20	90	0.60	99	0.64	94	0.92	87	3.06	92	90
Average	28.98	72	18.29	73	1.81	49	2.35	61	7.78	73	65



**Fig. 7.** Upper left figure is an original image with the tumor IMG07\_L1 and lower right figure is its corresponding final result. Figures from (b) to (e) show results with different number T of weak segmentation processes, in which the regions were improved as the T increased.



**Fig. 8.** The number of weak segmentation processes versus Jaccard index (= 1 - “volumetric overlap error”) of the automatically extracted tumor region with the reference segmentation of IMG07\_01 that was manually defined by the authors.

#### 4 Discussion and Conclusions

As shown in figure 5, the extracted regions improved as the number of weak segmentation processes  $T$  increased. This was also confirmed in figure 6. The Jaccard index for the test data of figure 5 was 58.3% at  $T = 1$ , and 78.8% at  $T = 100$ , resulting in a 20.5 points advantage for ensemble segmentation. In addition, the difference between the training curve and the test one as shown in figure 6 was small, indicating that the generated segmentation provided a high generalization performance.

The leave-one-patient-out test repeated the above validation, such that each data in datasets was used once as validation data. The difference between  $T = 1$  and 100 was about 12 points on average for 16 CT volumes, and the statistical “paired  $t$  test” indicated that the difference was significant ( $p < 0.005$ ).

The validation results using the testing data provided by the competition’s organizer presented some important findings. Table 2 shows the results of the comparison metrics and corresponding scores for all ten test cases. A tumor of IMG06\_L2 was missed due to the differences in distribution of CT value of the tumor as well as its shape from those of the training data. Other tumors of the testing dataset were segmented successfully. For example figure 7 shows the extracted regions corresponding to the tumor IMG07\_L1 with different number of weak segmentation processes. The proposed ensemble segmentation algorithm over-extracted the surrounding tissues, such as muscle and stomach wall, when the number  $T$  of weak segmentation processes was small. However it succeeded to reduce such false positives without increasing false negatives as the number  $T$  increased. The Jaccard index reached to 74.9% when  $T = 200$  (see figure 8), which was improved by the post-process of figure 4, resulting in 83.8%. Note that the value is different from that (= 100 - “Overlap Error” = 79.3%) in the Table 2, because the boundary defined by the

authors can be different from that of the organizer which is not yet open to the public. The number of false positive voxels was also evaluated because it is an important indicator for potential clinical use. The average false positive rate for testing data was 0.7%, which was small enough. Here the rate was obtained by dividing the number of false positive voxels extracted from a 3D volume data by the number of voxels in normal area with non-zero prior probability of liver.

In conclusion, our results suggest that ensemble segmentation is effective in segmentation of liver lesions, and increasing the number of weak segmentation processes can boost performance. In future work, we plan to assess the performance using a large dataset to evaluate generalization performance of the algorithm that includes a number of parameters in the feature measurement process, which means it might be sensitive to size and characteristics of training data. Moreover we will apply the proposed procedure for generating an ensemble segmentation algorithm by AdaBoost to other problems such as lesion extraction from multi-phase CT volumes or MR images.

## References

1. Soler, L., Delingette, H., Malandin, G. et al.: Fully automatic anatomical, pathological and functional segmentation from CT scans for hepatic surgery. *Computer-Aided Surgery* 6 3 (2001) 131-142
2. Shimizu, A., Kawamura, T., and Kobatake, H.: Proposal of computer-aided detection system for three dimensional CT images of liver cancer. *Proc. of Computer Assisted Radiology and Surgery*, (2005) 1157-1162
3. Heiman, T. Styner, M and Ginneken, B. (eds): 3D Segmentation in the Clinic: a grand challenge. *Proceeding of MICCAI2007 workshop* (2007) <http://mbi.dkfz-heidelberg.de/grand-challenge2007/>
4. Rohlfing, T., Brandt, R., Menzel, R., et al.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *Neuroimage* 21(4) (2004) 1428–1442
5. Rohlfing, T., Russakoff, D.B., Maurer, C.R.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Trans Med Imaging* 23(8) (2004) 983–994
6. Klein, A., Mensh, B., Ghosh, S., et al.: Automated brain labeling with multiple atlases. *BMC Med Imaging* 5(1) (2005)
7. Heckemann, R.A., Hajnal, J. V., Aljabar, Paul, et al.: Multiclassifier Fusion in Human Brain MR Segmentation: Modelling Convergence, *MICCAI2006 LNCS4191* (2006) 815-822
8. Freund, Y. and Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. *J. of Computer and System Sciences*, 55 (1997)
9. Kobatake, H., Jun, W., Yoshinaga, Y., et al. : Nonlinear Adaptive Convergence Index Filters and Their applications. *Proc. of ICPR*, 3 (2000) 526-529
10. Park, H., Bland, P.H., and Meyer, C.R.: Construction of an Abdominal Probabilistic Atlas its Application in Segmentation. *IEEE Trans. on Medical Imaging*, 22(4) (2003) 483-492
11. Gerig, G., Jomier, M., Chakos, M.: Valmet: A new validation tool for assessing and improving 3D object segmentation, *Proc. of MICCAI, LNCS 2208*, (2001) 516–523