Preliminary report of the CAD system competition in liver cancer extraction from 3D CT images and fusion of the algorithms

A. Shimizu\textsuperscript{a}, T. Kawamura\textsuperscript{a}, Y. Mekada\textsuperscript{b}, Y. Hayashi\textsuperscript{c}, D. Deguchi\textsuperscript{c}, S. Nawano\textsuperscript{d}

\textsuperscript{a}Tokyo University of Agriculture and Technology
\textsuperscript{b}Chukyo University
\textsuperscript{c}Nagoya University
\textsuperscript{d}National Cancer Center Hospital East

1. Purpose

We report here on the competition held at the annual conference of the Japan Society of Computer Aided Diagnosis of Medical Images. The purpose of the competition is to boost the development of computer-aided diagnosis systems for three dimensional (3D) CT images of liver in an effort to improve the detection of hepatic cellular carcinoma (HCC). The competition has been held annually since 2002. In the first three years, from 2002 to 2004 [1], the purpose of the competition was liver region extraction from 3D CT images. Now, researchers compete in liver cancer extraction from the extracted liver regions. This paper reports on the 2005 competition and shows the results of fusing the top three algorithms [2].

2. Methods

Prior to the 2005 competition, a total of 24 sets of four phase images, namely pre-contrast phase, arterial phase, portal venous phase and equilibrium phase images (total 96 images) were provided as training material to the competitors. The assignment was to develop a liver cancer extraction algorithm from 3D CT images.

The regulations announced in advance by the competition committee were as follows:

1. In addition to the 3D CT images, competitors would be allowed to input contrast conditions, size of the image, spatial resolution, and the image position of the first axial slice of the input images.
2. The border of candidate regions in the resultant binary image would be extracted and overlapped onto arterial phase images, which would then be used for evaluation by radiologists. The CT image would be converted to an 8 bit image for display, with a minimum -110 H.U. and maximum 190 H.U.

On the day of the competition, the input data was any combination of images chosen by the competitors from the four phase images taken from three HCC cases. These images had not been included in the training material. On the other hand, the output had to be an image corresponding to the arterial phase, which would be evaluated by radiologists.

The competition ran from November 19 to 20 at Chiba city, near Tokyo, and six competitors participated. At 9 am the competitors started to apply their algorithms to the test sets without further
modification of their algorithm.

A brief explanation of the top three algorithms is presented below. Notably, all of the top three algorithms used an arterial phase and equilibrium phase pair of images (Fig. 1), which is known to be effective in detecting HCC, and extracted liver regions in their first process (not explained here due to space limitations).

![Fig. 1 Examples of arterial phase (left) and equilibrium phase (right) images for training. The HCC (arrow) is enhanced in the arterial phase and is dark in the equilibrium phase, which is typical of HCC](image)

<Algorithm 1>

First, the non-rigid registration algorithm based on mutual information was carried out to align the arterial and equilibrium phase CT images. Then the algorithm extracted candidates using an adaptive neighbor type difference filter from both the early phase and equilibrium phase, separately. Subsequently, morphological operations (opening and closing) were applied to the candidate region to eliminate protrusions, holes and cavities. Finally, the candidates from the two phase images were integrated by logical product operation, which eliminated false positives.

<Algorithm 2>

This algorithm had similar processes to algorithm 1, as it was developed by the same research group. First, candidate regions were extracted from the equilibrium phase using the adaptive neighbor type difference filter. Next, the algorithm refined the boundaries of the candidates using a modified region forming partial borders algorithm. Subsequently, it evaluated the changes in CT values between the arterial phase and equilibrium phase to decrease the incidence of false positives. In order to define correspondence between the two phase images, a non-rigid registration algorithm with normalized mutual information was employed. After the registration process, the algorithm
evaluated a joint histogram of two phase CT values and classified the SR into one of two groups, namely cancer or normal tissue.

<Algorithm 3> [3]

This algorithm enhanced HCCs on the two phase images using the adaptive convergence index filter and fused the two enhanced images into one by logical sum operation after mutual information based non-rigid registration. The algorithm then extracted candidates from the fused image using a region growing algorithm and a level set method. Finally, it classified each of the candidates into one of the two groups (cancer or normal tissue) according to output of the Mahalanobis distance based classifier combined with a feature selection strategy.

<Fusion of the three algorithms>

We integrated the outputs of the above three algorithms into one based on the majority vote at each voxel. The output of each algorithm was a binary image in which the gray values of the background were assigned zero and the gray values of candidate regions were assigned one. The integration operator computed the majority of the three outputs and a labeling operation was performed on the resultant image in order to count false positives.

3. Results

Two radiologists began evaluating the resultant images for the three cases on November 20th, without prior knowledge of the algorithms developed by the competitors. Each resultant image was rated from zero to ten. After the evaluation, it was found that both Algorithm 2 and 3 achieved the highest score. The details of the evaluation are shown below.

Case 1

All three algorithms failed to detect the cancer, because the contrast was very low (see an arrow in Fig. 2), and thus all scored zero.

Fig. 2 Slice image of the missed cancer located at S8 and in the neighborhood of the IVC. Since the contrast was very weak, all of the algorithms failed to detect the cancer.
Case 2

Resultant images for case 2 are shown in Fig. 3. Algorithm 1 failed to extract the cancer at S3. Algorithm 2 succeeded in extracting the boundaries of both cancers. Although algorithm 3 identified the rough location of both cancers, it failed to define the exact boundary of the cancer in the right lobe. Consequently, the scores for algorithms 1, 2 and 3 were 5, 9 and 7, respectively.

Case 3

Figure 4 shows the resultant images produced for case 3. All algorithms succeeded in extracting the cancers with small underestimation and overestimation. However, algorithm 2 extracted one false positive at the porta hepatica, as shown in Fig. 5. Consequently, the scores for algorithms 1, 2 and 3 were 9, 7 and 9, respectively.

The total scores for algorithms 1, 2 and 3 were 14, 16 and 16, respectively, and thus the developers of algorithms 2 and 3 won the competition and received awards from the committee.

Fig. 3 Extracted results of liver cancers of case 2. One cancer was located in S3 while the other occupied a large portion of right lobe of the liver. The red lines show the boundaries of the extracted cancers.
Fig. 4 Extracted boundaries of two liver cancers in case 3, located at S8-4. All algorithms succeeded in extracting these boundaries, with a small amount of underestimation and overestimation of the cancer regions.

Fig. 5 A false positive extracted by algorithm 2.

We also fused the resultant images of these three algorithms using a majority vote for each voxel. The results for the 24 training data sets and three test data sets are summarized below.

<table>
<thead>
<tr>
<th>Table 1 Results of fusion of the three CADs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Training (24 cases)</strong></td>
</tr>
<tr>
<td>Algorithm 1</td>
</tr>
<tr>
<td>False Pos</td>
</tr>
<tr>
<td>False Neg</td>
</tr>
</tbody>
</table>

Fusion of the images appears to be more accurate as the number of false positives is smaller than that of the algorithm with the highest performance in the test data set. However, the difference is too small to statistically conclude that fusion is more effective, and so we plan to increase the number of images to determine the statistical significance of this finding. In addition, other arbitration rules of
outputs of these three algorithms, such as a linear combination of the degrees of malignancy, will be applied to achieve better performance.

4. Conclusion

We have reported on the CAD system competition for liver cancer extraction from three dimensional abdominal CT images held at the annual conference of the Japan Society of Computer Aided Diagnosis of Medical Images in 2005. Details of the top three algorithms and results of fusing these algorithms were presented. From the results of competition, it might be concluded that algorithm 2 and 3 were superior to other algorithms and fusion of the top three algorithms performed best. However, it should be noticed that number of test cases was too small to find statistical difference between the performances. So, in future we will increase the number of images for test. The competition for other organs, such as pancreas and kidneys, and other disease, such as cysts and benign lesions, are also challenging works to be done in order to develop a multi-organ, multi-disease CAD system [4].

Acknowledgements

This study was supported in part by the Grant-in-Aid for Scientific Research from Ministry of Education, Culture, Sports, Science and Technology, Japan, and the Grant-in-Aid for Cancer Research from Ministry of Health, Labour and Welfare, Japan.

References

Further information is available in the following website.  
http://www.tuat.ac.jp/~simizlab/CADM/PRELIMINARY_REPORT.html


This paper will appear on the following website.  