

Editorial: 3D Segmentation in the Clinic: A Grand Challenge II - Liver Tumor Segmentation

Xiang Deng and Guangwei Du

Center for Medical Imaging Validation, Siemens Corporate Technology

Abstract. In this paper, we present the organization of a competition of liver tumor segmentation techniques. The liver tumor segmentation competition is part of the workshop "3D Segmentation in the Clinic: A Grand Challenge II" at Medical Image Computing and Computer Assisted Intervention 2008 conference. The goal of this contest is to compare the performance of different algorithms for segmenting liver tumor from contrast enhanced CT images. Several organizing topics are described, including the motivation for organizing this competition, training and testing data and evaluation methods.

1 Introduction

Liver cancer is one of the leading causes of death world wide. It is especially common in developing countries. Measuring tumor size is very important in cancer diagnosis and evaluation of treatment response. Tumor volume has been shown to provide more accurate estimate of lesion size than routinely used 1D and 2D measurements . Recently, a number of image segmentation techniques have been developed to compute liver tumor volume from CT images. Comprehensive quantitative validation of tumor segmentation technique on large scale clinical data sets is essential for clinical use of tumor volumetry.

The goal of the liver tumor segmentation competition is to compare the performance of different liver tumor segmentation techniques under "real world" condition. The second major goal of the competition is to foster in-depth discussions and technical interchange among researchers in the field of medical image analysis from both academia and industry.

The liver tumor segmentation competition is part of the workshop "3D Segmentation in the Clinic: A Grand Challenge II" at Medical Image Computing and Computer Assisted Intervention (MICCAI) 2008 conference. This workshop is a continuation of the successful MICCAI 2007 workshop "3D Segmentation in the Clinic: A Grand Challenge".

2 Data

All liver tumor CT images were acquired on one 64-slice and two 40-slice CT scanners using a standard four-phase contrast enhanced imaging protocol with slice thickness of 1mm or 1.5mm and in-plane resolution of 0.6-0.9mm.

CT data from 30 liver tumors are used for this competition, which covers a range of patients and pathology (hepato-cellular carcinoma , hemangioma and metastasis). The 30 tumors are randomly divided into three groups: 10 for training, 10 for testing and 10 for the on-line contest. Teams registered with letter of intent can download both the training and testing datasets from the workshop website.

All the tumors were manually segmented by an experienced radiologist, and confirmed by another radiologist. The manual segmentation is used as reference for evaluation purposes. To make the reference segmentation as accurate as possible, only tumors with reference segmentation confirmed by both radiologists are used in the three sets of data. Therefore, not all the tumors identified in the three sets of data by the radiologists are used in this competition.

The training datasets are provided for participants to optimize their algorithms, which have both images and reference segmentation. The testing datasets do not include reference segmentation. Participants need to send in their segmentation results on the testing datasets for evaluation by a given deadline. The evaluation results are then sent back to the participants for preparation of their workshop papers. The on-line test datasets will be released before the start of the on-line competition on September 6, 2008. Segmentation results on these datasets are required to be submitted in three hours for evaluation.

All submitted segmentation results will be assessed using an evaluation software, which can also be downloaded from the workshop website for teams registered with letter of intent. The evaluation methods will be described in the next section.

3 Evaluation Methods

Three types of liver tumor segmentation techniques are accepted for this competition, including automatic, semi-automatic and interactive techniques¹. The performance of the segmentation algorithm is evaluated using a set of measures in a way similar to that for the MICCAI 2007 workshop "3D Segmentation in the Clinic: A Grand Challenge".

3.1 Evaluation Metrics

Several metrics have been commonly used to evaluate the quality of segmentation, including volume difference, volume overlap and surface distance. Each of these metrics, however, only captures certain aspect of the discrepancy between the segmentation and the reference, such as difference in volume and relative position. In order to have a comprehensive evaluation, we use a set of five metrics

¹ An automatic segmentation algorithm does not require any user intervention. A semi-automatic algorithm needs minimal amount of input from user, e.g., a seed point to initialize the segmentation. In comparison, an interactive algorithm requires manual editing of the final results.

to measure the accuracy of the tumor segmentation, which are described below. These metrics have been used in last year’s MICCAI workshop.

a) Volume overlap (m_1)

$$m_1 = \left(1 - \frac{\text{Vol}_{\text{seg}} \cap \text{Vol}_{\text{ref}}}{\text{Vol}_{\text{seg}} \cup \text{Vol}_{\text{ref}}}\right) \times 100\% \quad (1)$$

where Vol_{seg} denotes segmented volume. Vol_{ref} denotes reference volume.

b) Relative absolute volume difference (m_2)

$$m_2 = \frac{|\text{Vol}_{\text{seg}} - \text{Vol}_{\text{ref}}|}{\text{Vol}_{\text{ref}}} \times 100\% \quad (2)$$

c) Average symmetric surface distance (m_3)

$$m_3 = \frac{\sum_{a \in A} [\min_{b \in B} \{\text{dist}(a, b)\}] + \sum_{b \in B} [\min_{a \in A} \{\text{dist}(a, b)\}]}{N_A + N_B} \quad (3)$$

where A and B denote the surfaces of segmented and reference volumes. a and b are mesh points on A and B respectively. $\text{dist}(a, b)$ denotes the distance between a and b . N_A and N_B are the number of points on A and B .

d) RMS symmetric surface distance (m_4)

$$m_4 = \sqrt{\frac{\sum_{a \in A} [\min_{b \in B} \{\text{dist}(a, b)\}]^2 + \sum_{b \in B} [\min_{a \in A} \{\text{dist}(a, b)\}]^2}{N_A + N_B}} \quad (4)$$

e) Maximum symmetric surface distance (m_5)

$$m_5 = \max\left\{\max_{a \in A} \left\{\min_{b \in B} \{\text{dist}(a, b)\}\right\}, \max_{b \in B} \left\{\min_{a \in A} \{\text{dist}(a, b)\}\right\}\right\} \quad (5)$$

Note that all five metrics would have value 0 for perfect segmentation.

3.2 Scoring system

The five metrics described in the previous section can provide complementary information regarding the segmentation quality. For a competition, however, it would be desirable to have a single evaluation criterion. We use a scoring system similar to that in last year’s MICCAI workshop to compute an overall score by combining those from all the five metrics.

The design of the scoring system is described as follows. For perfect segmentation, each of the five metrics (values are all 0) will get a score of 100. A reference point with score 90 for each metric is determined from segmentation performed by independent users. It represents the score a human observer can get by manual segmentation. The values of the five metrics for the reference point are shown in the table below.

Score of each metric for a segmentation can be obtained using linear interpolation or extrapolation between the two points specified above. Note that 0 is the minimum score one segmentation will get, and there is no negative score.

Table 1. Values of the five metrics for the reference point

Metric	Value
Volume overlap (m_1)	12.94 %
Relative absolute volume difference (m_2)	9.64 %
Average surface distance (m_3)	0.40 mm
RMS surface distance (m_4)	0.72 mm
Maximum surface distance (m_5)	4.0 mm

4 Outlook

Since the first call for participation in early March, 41 teams from both academia and industry have registered for the competition with letter of intent, and downloaded the data. Among the 41 teams, 19 were from Europe, 12 from North America, 9 from Asia, and 1 from Africa. Such a large number of registrations has demonstrated world wide interests to this contest within the community of medical image processing and analysis.

We plan to make the datasets and evaluation results in this proceeding publicly available upon completion of the workshop.

5 Acknowledgement

We would like to thank Lan Song and Huadan Xue at Peking Union Medical College Hospital, China, Reto Merges and Jiuhong Chen at Medical Solutions, Siemens Ltd. China, Michael Suehling at Siemens Medical Solutions, Germany for making data available for this competition. We would also like to thank Chenyang Xu at Siemens Corporate Research, USA and Xiaodong Xu at Corporate Technology, Siemens Ltd. China for their support. We would like to specially thank the workshop sponsor, CT Division, Siemens Medical Solutions, Germany. Finally, we would like to thank all participants for their great effort on joining this contest.