LIVER SEGMENTATION FROM ABDOMEN CT IMAGES WITH BAYESIAN MODEL

NIDAA ALDEEK, RAJA S. ALOMARI, M B AL-ZOUBI, HAZEM HIARY
King Abdullah II School for Information Technology, The University of Jordan, Amman, Jordan
E-mail: ralomari@ju.edu.jo, mba@ju.edu.jo, hazemh@ju.edu.jo

ABSTRACT

Liver segmentation from CT volumes has been a challenging problem due to the high inter-organs intensity similarity, the intra-liver intensity variability, and the partial volume effect. In this paper, we perform an extensive review of the liver segmentation literature from CT and MRI. Furthermore, we propose a Bayesian model for a robust and reproducible semi-automatic technique for liver segmentation from CT volumes. We train our model and validate it using 44 clinical volumes for patients with various types of liver abnormality including tumor. Our segmentation results show a robust and clinically acceptable liver volume for all the 44 clinical cases we have with an average area overlap accuracy over 87%. Our method is superior to all state of the art methods that has only been validated on less number of subjects as we show during the literature survey.

Keywords: Liver Segmentation, Bayesian Model, Computed Tomography, Abdomen.

1. INTRODUCTION

Liver tumor is one of the most common types of tumors that cause death in the world. In 2010, over 18000 people died because of liver cancer in United States [1]. Surgical intervention is one of the ultimate treatment options. Surgery planning is an important step of successful tumor removal. Planning includes quantification of the tumor size, its location, and trajectories during surgery, and the effect of tumor removal on other neighboring organs such as the gall bladder and the percentage of the surgery success.

Liver segmentation from abdominal Computed Tomography (CT) volumes is an important step in many diagnostic and surgical procedures. It is also useful in building many related computer aided diagnosis, computer guided surgery systems, building anatomical atlases for the abdominal area, and many other applications. Various challenges face liver segmentation from abdominal area including: high inter-organs intensity similarity in the abdominal area, partial volume effect where organs’ boundaries are ill-defined, multi-segment of the liver especially in the lower portion, the location of the gallbladder and its great intensity similarity with the liver, the confusing internal structure of the liver and many other challenges as illustrated in Figure 1.

Many researchers work on liver segmentation from abdominal area. However, each technique has its own strength and weakness. Campadelli et al [2] presented a thorough survey of these techniques with a comparative study.

In our previous paper [3], we presented a fully automated liver segmentation from abdominal area based on a Markov Random Field for liver detection and a Gradient Vector Flow (GVF) snake model for boundary refinement. However, we only had 13 clinical cases at that time. In this paper, we validate our work on a significantly larger dataset including 44 cases.

On the other hand, quantitative segmentation validation methods are among the open issues in the literature. Many validation methods including area...
overlap, Dice measure, and Hausdorff distances are usually used for validation. However, visual validation of the cases by an expert is still a necessary measure for the validity of the segmentation method as discussed in our previous work [4]. In this work, we validate our method using both careful visual examination by an expert radiologist and using one quantitative measure for error and accuracy with one ground truth.

The rest of this paper is organized as follows: Section 2 presents an exhaustive literature survey for liver segmentation. Section 3 presents our proposed technique and our classifier. Section 4 presents our data and experimental validation methods and then we conclude in section 5.

2. LITERATURE SURVEY

Liver segmentation has been tackled by many researchers for various computer aided diagnosis and surgery applications. Many surveys exist in the literature for exhaustive review of the literature such [2]. Here we mention some recent related techniques.

Liu et al. [5],[6] used gradient vector flow (GVF) field as an external force. Active contours are obtained by defining curves whose shape and location are guided by two forces: internal forces that are derived from pixels on the curve to impose desirable properties on the curve shape and external forces that are derived from image characteristics to limit the curve to certain locations in the image. Zhou et al. [7] proposed a probabilistic model for liver segmentation from CT which was evaluated on 152 subjects. Lim et al. [8] proposed liver segmentation for volume estimation from CT. They analyzed the intensity distribution to obtain a priori model to determine the coherent regions of liver. They utilized morphological filters and a labeling-based search algorithm to obtain the liver contour, which then leads to liver volume estimation.

Ciecholewski and Debski [9] utilized the lumbar section of the spine as an initial reference to perform liver segmentation from CT. Then they constructed a finite number of joint poly-lines that resulted in two polygons representing the liver segments. Rikxoort et al. [10] started with pre-processing steps to determine the vertical scan range of the liver and to rotate the scan so that the subject is in supine position. Then they performed voxel labeling with K-nearest-neighbor. A final smoothing filter is then applied to obtain a fine segmentation. They only evaluated the system on 10 test cases. Saddi et al. [11] estimated the pose and global shape properties with a statistical shape model to learn shape features of the liver. Then they used a template to recover local deformations. They only validated their method on 10 cases. Slagmolen et al. [12] built an atlas based on 20 cases using non-rigid registration. Susomboon et al. [13] utilized intensity and texture information to generate probability images that aid in segmenting the liver for new cases. They tested their method with 10 cases. Ma and Yang [14] proposed a two-step scheme: estimation of the statistical parameter vector of a mixture Gaussian distribution by the Expectation Maximization (EM) algorithm, and then using morphological filters to remove foreign components and apply image hole-filling.

Liu et al. [15] proposed an adaptive method that utilizes a bi-class Support Vector Machine (SVM) after applying adaptive thresholding and kmeans clustering. Size of data was limited. Freiman et al. [16] presented a new algorithm for nearly automatic liver segmentation and volume estimation from abdominal CT images and its validation. They used a multi-resolution iterative scheme that repeatedly applied smoothed Bayesian classification to identify the liver and other organs. They evaluated their method with two retrospective studies on 56 validated CT images. They obtained 0.98 and 0.99 correlation for liver volume estimation, with mean volume differences of 5.36 and 2.68% with respect to manual ground truth estimation, and mean volume variability for different initial seeds of 0.54 and 0.004%, respectively.

Platero et al. [17] segmented liver from MR images based on 3D anisotropic diffusion processing. They obtained an initial segmentation using edge detection techniques (Canny edge detector), histogram analysis and binary morphological post-processing. Then they applied an active contour to refine the segmentation. They computed the undirected partial Hausdorff distance between the boundary of the computed segmentation and the boundary of the manual segmentation and obtained error of 2.3 mm for 95% percentile and 2.8 mm for 99%. Selver et al. [18] pre-processed the images by removal of the fat tissues, spine, ribs, and right kidney. Then they used k-means and MLP classifier for liver segmentation followed by smoothing operators. On 20 subjects, they obtained a success rate of 94% for segmentation. Cheng et al. [19] proposed using a level set approach for liver segmentation from MRI images. They combine shape prior knowledge with the improved Chan-Vese’s model, which may overcome the leakage and over-segmentation problems. Foruzan et al. [20] estimated the initial liver boundary from CT images with an
Expectation Maximization (EM) algorithm, elimination of the ribs and heart, applying double thresholding, and then a final labeling step. They also evaluated their method on 10 CT cases and obtained an average overlap error of 15.3%.

Luo et al. [21] proposed a three-step automated technique: texture filters, then support vector machines (SVM) to classify based on texture analysis, and then post-processing morphological operators. Zhao et al. [22] proposed a method by initially removing the ribs and spine based on thresholds, and then utilizing a neural network to perform the segmentation and the morphological post-processing. Our proposed method is unique because it deals with clinical data and seamlessly incorporates within the radiologist or technician routine to perform a quick and highly robust liver segmentation.

Zhang et al. [23] segmented the liver from CT scans with a Statistical Shape Model (SSM) integrated with an optimal-surface-detection strategy. They built an average liver shape model with a training set. Then they used subspace initialization of the SSM and shape model deformation to adapt to the liver contour through an optimal-surface-detection. Zayane et al. [24] built a priori knowledge model of location and shape of the liver from a training set of CT images. Then they applied the Greatest Connected Component (GCC) algorithm to detect the largest connected component, which is assumed to be the liver. Few morphological operators are then applied for image filling.

Masoummi et al. [25] used mathematical morphology to enhance the image through preserving edges and reducing noise. They trained the MLP neural networks to extract the features of the liver region to avoid over-segmentation. These extracted features were used to monitor the quality of the segmentation using the watershed transform and adjust the required parameters automatically. The average accuracy, based on Jaccard coefficients, was 0.94.

Few research efforts have been proposed by directly working on 3D volumes rather than slice-by-slice fashion such as Okada et al. [26], who developed an atlas-based liver segmentation method from 3D CT volumes. They performed voxel-based segmentation with a Probabilistic Atlas (PA) to an initial region for subsequent Statistical Shape Model (SSM) fitting to 3D CT images. They utilized a Multi-Level SSM (ML-SSM) to improve the reconstruction accuracy especially for largely deformed liver. The whole shape was divided into patches. For each patch, the principle Component Analysis was applied. They introduced a new constraint (adhesiveness constraint) for overlap regions among patches. Through this method, they demonstrated that segmentation accuracy improved by using PA and ML-SSM. However, the Jaccard similarity measure and average distance were 0.86 (± 0.05) and 2.15 (± 0.62) mm on average, respectively. Dawant et al. [27] used a level-set approach and a dynamically adapted speed function. They trained an anatomic atlas to reduce leakage at the liver-rib interface.

Wimmer et al. [28] proposed a two-stage liver segmentation from CT scans. A manual delineation of cross-sections of the anatomical structure in 2D multi-planer reconstruction views is constructed. Then an initial 3D surface was reconstructed using radial basis functions. Then they applied a level set algorithm incorporating a new combination of image information and shape information. They validated their method on only 10 CT scans. Kainmüller et al. [29] presented a combination of a constrained free-form and statistical deformable model for liver segmentation. Furukawa et al. [30] trained a maximum a posterior probability estimation. Then they used a combination of the probability density function of a Gaussian mixture distribution and a prior probability derived from a probabilistic atlas of the liver. Then they used level set for better final segmentation.

Seghers et al. [31] modeled an object as a set of landmarks augmented with local appearance models to perform 3D liver segmentation on 10 cases. Yussof and Burkhardt [32] used anisotropic diffusion to filter the original liver volume from 3D CT. Then they applied thresholding to preserve all tissues that have the same intensities for liver. They also performed morphological operators to ultimately obtain a smooth fine liver region with a graph-cut technique. They also evaluated their method on 10 cases. Lu et al. [33] initialized a deformable model with a manually created simple mesh model for the liver. They formulated the deformation of the shape to adapt to the boundary by minimizing a local cost function associated with each model vertex. The experimental results demonstrated the effectiveness of the proposed algorithm.

3. PROPOSED METHOD

Our method utilizes a Bayesian classifier to segment the liver from the surrounding structures. We ask the user for some interaction to guarantee a robust and clinically acceptable segmentation result. Figure 2 shows the workflow of our method.
3.1 Preprocessing

Our clinical CT volumes contain a set of images that span the abdominal area from the chest down to the pelvis in axial views. We initially perform a clinically approved preprocessing step known as window/level enhancement.

Window/Level enhancement is based on the physical properties of the tissue being scanned by the CT. This enhancement is achieved by converting the intensity values of the image into Hounsfield units [34]. Each material has physically-approved Hounsfield values that are based on the physical intensity of that material. HU values are quantitative measure of the transparency of a material to X-rays. Table 1 shows some of the known material with the corresponding Hounsfield values. Window/level enhancement aims at enhancing specific Hounsfield values (HU) over the rest of the values in the signal. Figure 3 shows the window/level application filter. We convert to Hounsfield values (HU) from DICOM intensity values (I):

\[ HU(x, y) = I(x, y) \times \text{Re scaleSlope} + \text{Intercept} \]

where \( I(x, y) \) is the DICOM intensity values at the 2D image location x, y. \( HU(x, y) \) is the corresponding Hounsfield value. Both Re scaleSlope and Intercept are two parameters supplied from the DICOM header for each slice within the abdominal volume.

<table>
<thead>
<tr>
<th>Material</th>
<th>Hounsfield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air</td>
<td>-1000</td>
</tr>
<tr>
<td>Fat</td>
<td>-120</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
</tr>
<tr>
<td>Muscle</td>
<td>+40</td>
</tr>
<tr>
<td>Bone</td>
<td>+1000</td>
</tr>
</tbody>
</table>

Table 1. Hounsfield values for some material.

Once the volume is enhanced upon the clinical standard, the radiologist uses our viewer to browse over the volume. Because our data is clinical, there are many slices that do not have the liver in them. In some cases, as much as half the abdominal volume does not have the liver because clinical abdominal standard is to acquire the whole area from the chest down to the pelvis. Thus, the radiologist (or the technician) browses over the cases from top where the heart appears down to the last slice that shows the liver. In each slice, she draws an oval that contains the liver (and some other structures). We then apply our Bayesian Model for segmentation.

3.2 Bayesian Model

We design a Bayesian classifier that learns its parameters from a set of training data. Our classifier maximizes the posterior probability \( p(c | I) \) for each pixel in the input image \( I \) (oval selected by the radiologist):

\[ L = \arg \max_c p(c | I) \]

where \( L \) is the outcome of our classifier at each pixel in the input image \( I \), \( c \) is a binary random variable where \( c = 1 \) for a liver pixel and \( c = 0 \) for a non-liver pixel. Using Bayesian:

\[ p(c | I) = \frac{p(I | c) \times p(c)}{p(I)} \]

where \( p(I | c) \) is the likelihood of a pixel to be in each class (liver/non-liver), \( p(c) \) is the prior
probability for liver/non-liver classes, and \( p(I) \) is the marginal probability over the classes (liver/non-liver).

As a standard practice, to maximize the posterior probability, we maximize the logarithm of the posterior:

\[
L = \arg \max_i \ln[ p(c \mid I) ]
\]

(4)

where

\[
\ln[ p(c \mid I) ] = \ln\left[ \frac{p(I \mid c) x p(c)}{p(I)} \right] = \ln\left[ \frac{p(I \mid c) x p(c)}{p(I)} \right]
\]

(5)

However, \( \ln[ p(I) ] \) is a constant and thus can be ignored. Thus,

\[
\ln[ p(c \mid I) ] \propto \ln[ p(I \mid c) x p(c) ]
\]

(6)

Reducing this, we conclude:

\[
L \propto \arg \max_i (\ln[ p(I \mid c) ] + \ln[ p(c) ])
\]

(7)

We then model both the likelihood \( p(I \mid c) \) as Gaussian distributions where:

\[
p(I \mid c) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(I-\mu)^2}{2\sigma^2}}
\]

(8)

where \( \mu \) and \( \sigma^2 \) are the mean and variance of the classes \( c = \{ \text{liver}, \text{non-liver} \} \). Both parameters are learned from a set of data points (1500 points (750 points of each class) that were manually selected from a set of liver images.

3.3 Post processing

Two post processing steps were performed. A 2D median filter to eliminate the salt and pepper noise resulting from some pixels with similar intensity levels for structures surrounding the liver (happen to be inside the selected oval). Then we apply an image filling operation [35] to close the holes inside the liver resulting from the internal structure of the liver.

4. DATA AND RESULTS

We collect our data from our collaborating radiology center. We obtain 44 clinical cases that have various abnormality types with some normal cases. Our data is anonymized before removal from the radiology center location. Patient consents were taken for each of the cases. Each case consists of a full volume of the abdominal area starting from the chest down to the pelvis. Some of the cases contain up to 120 slides while other might include as few as 50 slides. Images are in DICOM format with 512 x 512 x 5 mm per voxel.

We obtain the ground truth from one radiologist and currently working on having three more radiologists to manually segment each case. Our ground truth was performed as a contour surrounding the liver by software that we develop upon the convenience of the radiologist.

We run our method on all the 44 cases and obtained the full volume of each case. Figure 4 shows a sample a middle slice at each step in our proposed workflow (Figure 2).

One strength aspect of our method is that it works on all slices whether the liver is small (lower slices) or big (middle slices) which makes it clinically suitable because radiologist are concerned with accuracy more than fully automating the workflow. Figure 4 shows a set of sample slices from various abdominal levels.

We validate our proposed method in two ways: qualitative and quantitative. Segmentation accuracy quantitative techniques have positives and negatives and there is no one technique that
provides optimum sense of the meaning of accuracy [36].

For qualitative validation, we ask our radiologist to thoroughly view each slice in each of the 44 cases and record the result of the segmentation as one of two choices: clinically acceptable or not. By this decision, there is a rational factor to what is clinically acceptable or not. However, the radiologist can determine that. Our radiologist approved all cases as clinically acceptable cases.

Table 2 shows the summary of two quantitative measures that we call accuracy and error:

$$\text{Accuracy} = \frac{R \cap G}{G} \quad (9)$$

$$\text{Error} = \frac{R \cup G - R \cap G}{G} \quad (10)$$

where R and G are areas of our method’s and the gold standard segmentation results, respectively.

We randomly selected ten cases to give a sense of the quantitative accuracy. However, full results and thorough statistical analysis will appear in an extended version of this paper due to size limitation. We achieved an average of 87% area overlap.

5. CONCLUSION

In this paper, we proposed a robust, reliable, and clinically acceptable method for liver segmentation from CT volumes. We proposed a Bayesian-based classifier to model the posterior probability distribution of the image intensity. Then we assign a binary class for each pixel of being a liver or not. Our experimental validation on 44 cases shows a perfect clinical satisfaction from our collaborating radiologist who manually validated our segmentation for each case. We also provide two quantitative measures for the segmentation compared to a manual segmentation by an expert radiologist and achieved an average area overlap over 87%. Furthermore, we provided an extensive literature survey for liver segmentation from abdomen radiology images.

We work on obtaining three more manual segmentations to provide reliable ground truth that takes into consideration the inter-observer variability in liver segmentation. Moreover, we currently prepare for a thorough clinical study on the statistical significance of the inter-observer variability and reliability of liver segmentation from CT volumes.

Table 2. Quantitative Accuracy Measures (Eq. 9, 10) For A Set Of Randomly Selected Sets

<table>
<thead>
<tr>
<th>Cases</th>
<th>Error Mean for each case</th>
<th>Accuracy Mean for each case</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1400</td>
<td>0.8897</td>
</tr>
<tr>
<td>2</td>
<td>0.1374</td>
<td>0.8866</td>
</tr>
<tr>
<td>3</td>
<td>0.1263</td>
<td>0.8899</td>
</tr>
<tr>
<td>4</td>
<td>0.1638</td>
<td>0.8456</td>
</tr>
<tr>
<td>5</td>
<td>0.1317</td>
<td>0.8830</td>
</tr>
<tr>
<td>6</td>
<td>0.1798</td>
<td>0.8421</td>
</tr>
<tr>
<td>7</td>
<td>0.1681</td>
<td>0.8466</td>
</tr>
<tr>
<td>8</td>
<td>0.1123</td>
<td>0.9099</td>
</tr>
<tr>
<td>9</td>
<td>0.1601</td>
<td>0.8687</td>
</tr>
<tr>
<td>10</td>
<td>0.1573</td>
<td>0.8544</td>
</tr>
<tr>
<td>AVG.</td>
<td>0.14768</td>
<td>0.87165</td>
</tr>
</tbody>
</table>

REFERENCES:


