A Hybrid Approach for Liver Segmentation

Ruchaneewan Susomboon, Daniela Stan Raicu, and Jacob Furst

Intelligent Multimedia Processing Laboratory
School of Computer Science, Telecommunications and Information Systems
DePaul University, Chicago, USA
rsusombo@students.depaul.edu, draicu@cs.depaul.edu, jfurst@cs.depaul.edu

Abstract. In this paper, we propose a hybrid approach for fully automatic liver segmentation in Computed Tomography (CT) data. The approach consists of four stages: first, an *intensity-based partition* is applied to obtain the soft-tissue regions. Second, a *region-based texture classification* is used to classify all the soft-tissue regions based on texture features and generate probability images. Third, an initial region of the liver for each patient is determined from the probability images. In the final fourth step, a 95% confidence interval determined from the intensities of the initial regions is used to detect the liver region in each one of the consecutive slices in which the liver appears.

1 Introduction

Automatic analysis of images from various medical imaging modalities is necessary to increase the productivity of radiologists when interpreting and diagnosing hundreds of images every day. Image segmentation is an important first step in analyzing medical data and serves as a vital preliminary step in many imaging applications such as diagnosis, anatomical structure studies, treatment planning, and computer integrated surgery [1].

In general, the segmentation approaches can be divided into eight categories [2]: 1) thresholding, 2) region growing, 3) clustering, 4) Markov random field models, 5) artificial neural networks, 6) classifiers, 7) deformable models, and 8) atlas-guided approaches. Most of these approaches are based on gray-level intensities which make the liver segmentation a difficult task given that several organs adjacent to the liver (for instance, stomach, kidney, and heart) may have intensities similar to that of the liver. To deal with weak boundaries between the liver and adjacent organs, texture information can be integrated in the segmentation algorithms based on the hypothesis that soft tissues can be discriminated using texture [3]. In our previous work [4], we showed that using only pixel-level texture-based classification, it is possible to accurately perform 2D segmentation of the liver from the rest of the anatomical structures in the abdomen. In this paper, the algorithm from [4] was extended to segment the liver volume in a slice-by-slice fashion and also to take into account the computationally-expensive task of calculating the texture at the pixel-level for an entire CT slice. Our proposed hybrid approach combines image processing (quad-tree decomposition, region

growing, thresholding) and data mining techniques (expectation maximization, decision trees) to produce a fully automatic liver segmentation algorithm across a sequence of CT slices for the same patient as well as across multiple patients.

The rest of this paper is organized as follows. Our methodology is described in Section 2, the preliminary results along with their discussion are presented in Sections 3 and 4, and the conclusion and future work are described in Section 5.

2 Methodology

The main steps of our approach are shown in Figure 1; first, an intensity-based partition is applied to obtain the soft-tissue regions. Second, a region-based texture classification is used to classify the soft-tissue regions based on texture features and generate probability images. Third, based on the probability images, an initial region for the liver is determined for each patient. In the final fourth step, a 95% confidence interval determined from the intensities of the initial region is used to detect the liver regions in each one of the consecutive slices in which the liver appears.

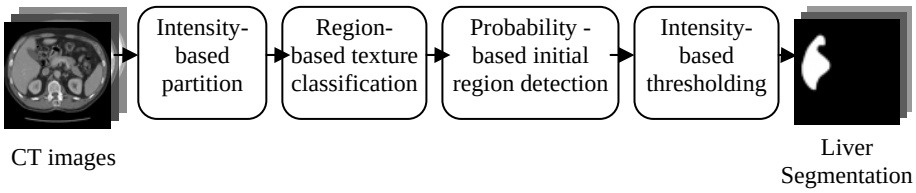

Fig. 1. Diagram of the proposed approach

2.1 Intensity-based partition

The human body is mainly composed of air, fat, soft-tissue, and bone, which on the Hounsfield Unit (HU) scale (-1024 to +3071) are represented such that the air has the lowest HU value, followed by fat and soft tissues, respectively, and bone with the highest HU value [5]. Therefore, we can assume that the intensity distribution of each patient is a mixture of four Gaussians, with the third distribution representing the soft tissues. To find the individual distributions, we apply the Expectation-Maximization (EM) algorithm [6–8] that allows finding the parameters (average μ_{ij} and standard deviation σ_{ij}) of each of the four distributions ($j = 1...4$) for each patient i (in our case, the training data consists of $i = 20$ patients); Figure 2 illustrates the process of estimating automatically the range for soft-tissue in human body.

By applying the Expectation Maximization algorithm for every patient's CT scan (that is, on the intensity distribution of all slices for that scan), a 95%

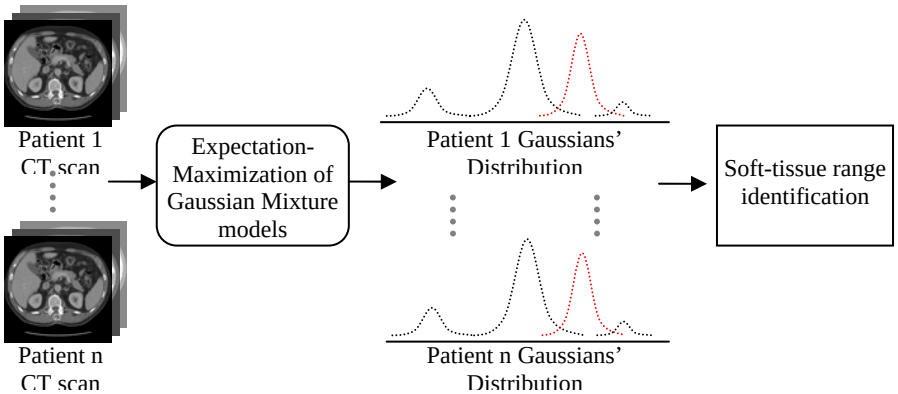

Fig. 2. Soft-tissue intensity identification diagram

confidence interval (a_i, b_i) is produced for the range of soft tissues per patient i. Since these ranges might vary across patients and the approach is expected to work across all patients, the overall range of soft tissues will be given by the minimum value of all confidence intervals' lower bounds a_i and the maximum value of all confidence intervals' upper bounds b_i:

$$(A, B) = (\min_{i=1...n}(a_i), \max_{i=1...n}(b_i)) \tag{1}$$

where $n = 20$ is the number of patients. The beauty of this approach is that the overall range of soft tissues can be easily changed when new patients are added to the data set.

Once the overall range is determined, a 2D Quad-Tree (QT) decomposition [9, 10] is applied to segment each individual slice into as pure as possible soft tissue regions and non-soft tissues regions. The criterion for splitting is chosen based on the average standard deviation of all standard deviations obtained for the distributions of soft tissues using the EM algorithm. If the standard deviation of the intensities values within a region is larger than the average EM standard deviation, then that region is split further; regions with size smaller or equal to 8 by 8 will not be split further into 4 by 4 regions. The size of the region for the stopping criterion is chosen based on the fact that regions with sizes smaller than 8 by 8 will generate results that are not statistically significant given the small number of pixel data.

At the end of the first stage, the regions with at least 85% pixels belonging to the overall range (A, B) for soft tissues will become candidates for the soft tissues regions. Both the 85% criterion and the slice-by-slice QT segmentation fashion are based on our previous 2D work [4, 11].

2.2 Region-based texture classification

Since most of the soft tissues have overlapping gray-levels, texture plays an important role to differentiate among tissues. There are several texture models, including structural, transform methods, and statistical models [12]. Among all of these models, in our previous work [13] we showed that the second-order statistical model (co-occurrence matrix) produces better classification accuracy over the transform-based approach and structural method for liver texture-based segmentation. The common technique used to extract texture from the co-occurrence matrix model is based on the Haralick texture descriptors [14].

For each soft-tissue candidate region obtained in the previous step, a normalized co-occurrence matrix is calculated along with nine Haralick texture descriptors (entropy, energy, contrast, sum average, variance, correlation, maximum probability, inverse different moment, and cluster tendency) for two distances (1, 4) and four directions (0°, 45°, 90°, 135°). Therefore, each region will be characterized by a p-dimensional feature vector $[f_1(d, \theta; \gamma), ..., f_p(d, \theta; \gamma)]$, where p indicates the number of texture features, d and θ represent the displacement and the angle, respectively, and γ indicates the corresponding Haralick feature. We use these two displacements (1 and 4) based on our previous work [15] where we showed that the displacements capturing the 'near' and 'far' patterns are enough for the liver texture co-occurrence model.

After the texture features are calculated for the soft tissue candidate regions, they are used to generate classification rules for liver versus non liver regions. In our approach, a binary Classification and Regression Tree (C&RT) model is applied because of several reasons: there are no assumptions about the distribution of the data, the model is computational efficient and easy to interpret, and it has good classification performance [16]. Once the optimal decision tree is built, each terminal node will produce a decision rule expressed in the IF THEN form. For each rule, a probability can be associated with based on the number of correct classified cases at that node. For instance, a rule as

IF $[f_1(d, \theta; \gamma) < 0.34$ AND $0.65 < f_5(d, \theta; \gamma) \leq 0.8]$THEN Liver with PROBABILITY=.9,

will denote a rule obtained from a terminal node in which 90% of the pixels found at that node were indeed liver pixels.

At the end of the classification process, the decision rules will be applied to each candidate region within the slice to be segmented producing a probability image. Each region within the image receives a probability of being liver based on the probability of the rule classifying that corresponding region (with a probability equal to zero for the non soft tissues regions).

2.3 Initial liver region detection

From the probability images, an initial region for the liver is determined based on the following algorithm (Figure 3). First, all pixels from the candidate regions (determined in Section 2.1) will receive the probability (determined in Section

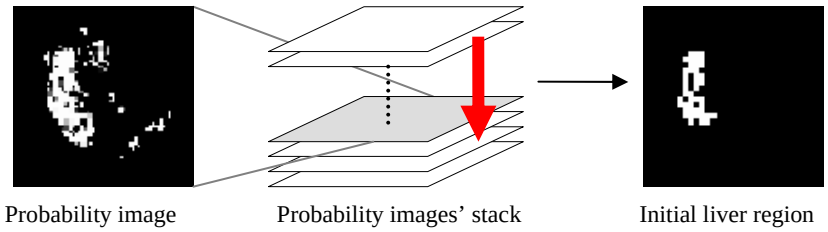

Fig. 3. The probability image shown on the left side corresponds to the gray image in the stack of slices for a certain patient and contains the largest connected component (shown on the right side) of the stack.

2.2) of the region they belong to. Second, all 4-neighborhood connected components of pixels with at least .85 probability are found for each slice in the stack of consecutive slices for each patient. Third, the connected component with the largest probability (sum of all pixel probabilities) across all slices will become the initial region for that corresponding patient. In other words, at the end of this process, there will be only one initial liver region in only one slice (named initial slice) for each one of the patients. This initial region and initial slice will be used further to find the liver regions in the stack of slices in which the liver appears.

2.4 Intensity-based segmentation

First, using the fitted normal distribution for the intensities of the initial liver region (Figure 4.a), a 95% confidence interval (CI) is determined to find the liver regions in the adjacent slices to the initial slice using a thresholding-based approach. That is, all connected-components whose pixels have values within the 95% CI will form candidate liver regions. Second, an overlapping algorithm is applied between these adjacent slices with the detected candidate liver regions and the initial slice containing the grown initial liver region. The growing of the initial liver region is performed by adding pixels for which their 9 by 9 neighborhood has an average intensity within the 95% CI. Third, a candidate liver region becomes a final liver region in an adjacent slice if there is at least an 85% overlap with the grown region in the initial slice.

The process then moves upward and downward to find liver in the next slices (Figure 4.b) and repeats until no candidate regions are found to overlap with the liver region in the initial slice.

3 Preliminary Results

Our preliminary results are based on data extracted from CT studies which are enhanced with contrast agent and scanned in the central venous phase on a

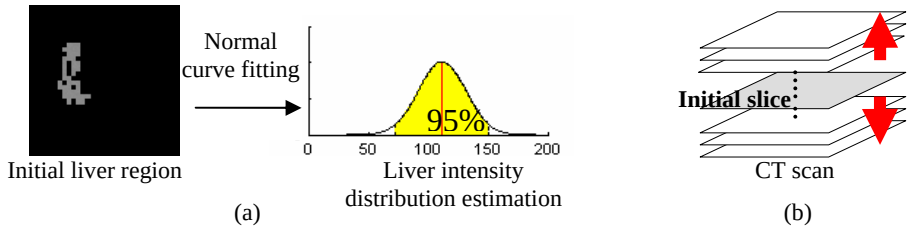

Fig. 4. (a) represents a 95% CI of the initial liver region, and (b) represents the direction of finding the liver in the stack of images using the 95% CI

variety of scanners (different manufacturers, 4, 16 and 64 detector rows). The data is acquired in transversal direction and the pixel spacing varies between 0.55 and 0.8mm while the inter-slice distance varies from 1 to 3mm. The studies are mostly pathologic including tumors, metastasis and cysts in different sizes. There are 30 CT studies; 20 training sets are supplied with liver ground truth assessed by radiologists and 10 testing sets. In order to obtain the prior knowledge for C&RT, we randomly divided all regions (86503) in the training set (consisting of 20 patients) resulted from the quad-tree decomposition into training (66%) and testing (34%) needed to construct and validate the classification approach, respectively. The optimal C&RT had 122 terminal nodes with an approximate 83% overall sensitivity and 86% overall specificity on the testing set in classifying liver versus non-liver tissues.

Figure 5 illustrates examples of segmentation results for an easy case, an average case, and a difficult case. The overall results of our proposed approach for all patients are presented in Table 1.

The score of the testing set is based on the five criteria defined by Gerig et al. [17]; 1) *Volumetric overlap*: number of voxels in the intersection of segmentation and reference, divided by the number of voxels in the union of segmentation and reference; 2) *Relative absolute volume difference*: total volume of the segmentation is divided by the total volume of the reference; 3) *Average symmetric absolute surface distance* (millimeters): average of all distance of absolute surface between the border voxels of segmentation and reference; 4) *Symmetric RMS surface distance*, (millimeters): average of all the squared distances between the border of segmentation and the reference; and 5) *Maximum symmetric absolute surface distance*, (millimeters): maximum of all border voxel distances.

While the average volumetric overlap is about 74%, patients 3 and 4, have the overlap below 50% which indicates several problems with the current approach. These problems will be discussed in detail in the next section.

4 Discussion

While the algorithm correctly detects the initial liver regions for all patients, the growing of the initial region often overestimates the liver region in the initial slice.

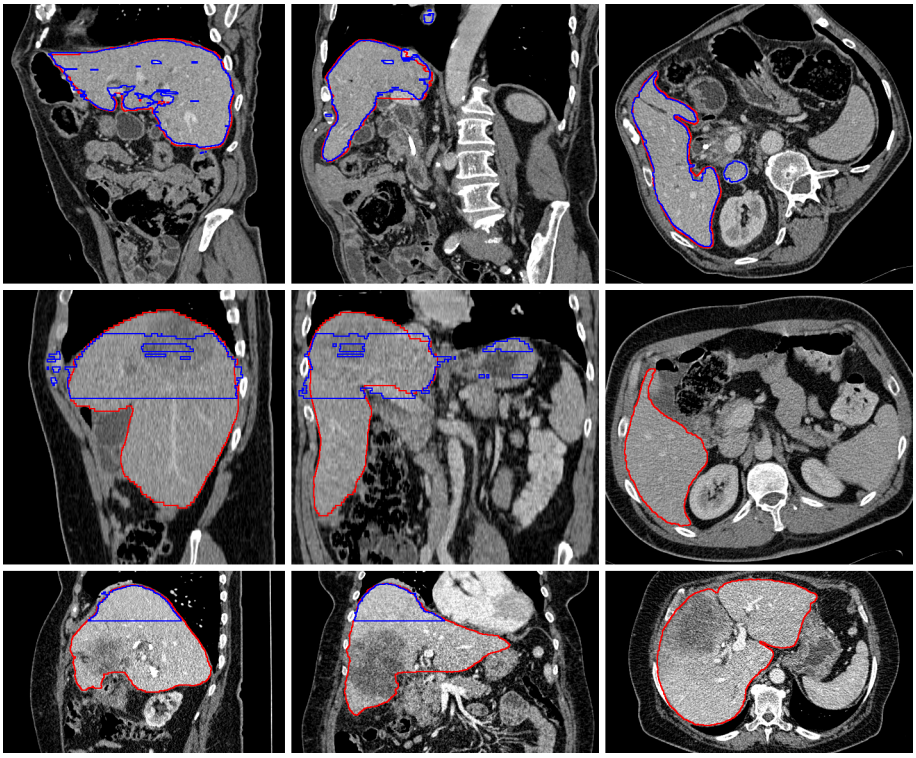

Fig. 5. From left to right, a sagittal, coronal and transversal slice from a relatively easy case (1, top), an average case (4, middle), and a relatively difficult case (3, bottom). The outline of the reference standard segmentation is in red, the outline of the segmentation of the method described in this paper is in blue. Slices are displayed with a window of 400 and a level of 70.

Based on our previous work [4, 11], we hypothesize that we will eliminate this problem by replacing the intensity-based region growing approach with a texture-based approach. Given the time constraints, for the MICCAI competition we chose the intensity-based approach because the texture-based approach is very computationally expensive requiring texture calculation for every single pixel. The intensity-based approach allowed us to segment the test data within four hours.

Furthermore, the 85% threshold for the overlap proved to be too high in order to be able to detect the liver regions in all consecutive slices as illustrated in Figure 5. In the future, we plan to lower the threshold at the expense of detecting other regions as being liver (Figure 6). This problem can be alleviated if the liver segmentation is eventually intended to be just one of a series of filters for organ segmentation. In other words, once individual organ segmentations have been generated, they can be combined to generate the final segmentation

Dataset	Overlap Error		Volume Diff.		Avg. Dist.		RMS Dist.		Max. Dist.		Total
	[%]	Score	[%]	Score	[mm]	Score	[mm]	Score	[mm]	Score	Score
1	31.2	0	22.4	0	11.4	0	22.1	0	95.4	0	0
2	23.9	7	16.9	10	7.7	0	16.0	0	73.0	4	4
3	87.0	0	-87.0	0	46.8	0	62.6	0	148.0	0	0
4	46.4	0	-36.4	0	17.9	0	33.2	0	138.0	0	0
5	14.2	44	2.2	88	4.1	0	9.5	0	73.5	3	27
6	13.2	48	-2.1	89	3.3	18	6.5	10	51.1	33	40
7	13.7	46	-7.7	59	3.0	24	5.9	19	38.0	50	40
8	10.5	59	-8.8	53	2.1	47	4.4	39	39.9	48	49
9	9.3	64	-7.0	63	1.6	61	3.3	54	28.0	63	61
10	14.5	43	-8.0	58	4.0	1	7.9	0	55.6	27	26
Average	26.4	31	-11.5	42	10.2	15	17.1	12	74.0	23	25

Table 1. Results of the comparison metrics and scores for all ten test cases.

showing all the segmented organs in the same study. The combination of the single organ filters will be based on our previous work [11]: in the case of a region which has been only labeled in one instance of single-organ segmentation, that region retains its label in the final segmentation. However, when a region has multiple labels, the algorithms will make a decision on a single label for that pixel based on information about the region to which the pixel belongs.

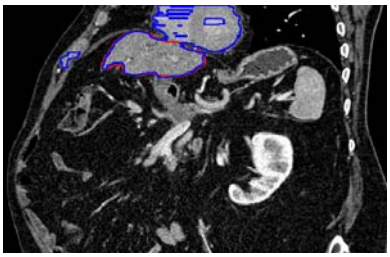

Fig. 6. Coronal slice from case 2. The outline of the reference standard segmentation is in red, the outline of our approach is in blue. The liver is leaking into the heart; we expect to fix this using combined single organ filters.

In addition to finding the liver region, we will be considered moving in the traversed directions. The current implementation does not allow us to find other regions of the liver in the same slice as it considers only upward and downward directions. Finally, an extension of the proposed approach will be to volumetric texture and 3D segmentation algorithms instead of the presented slice-by slice texture-based segmentation.

5 Conclusion and Future Work

This research work is an adaptation of our previous classification-based liver segmentation using pixel-level texture features [4, 11]. The limitation with our previous approach was that it is very computationally expensive given the calculation of the texture features for each individual pixel. In order to reduce the computation time, we modified parts of the algorithm to 1) consider regions instead of just pixels and 2) perform region growing, thresholding, and overlapping based on intensities instead of texture. The entire segmentation process of the ten patients was done in approximately four hours. However, the drawback is that we have not had enough time (before the MICCAI liver segmentation contest) to explore ways to calculate the values of the thresholds instead of using predefined ones. Nonetheless, our preliminary results show that, by using our approach, we are able to correctly detect the initial region of the liver. Much work still remains to be done with respect to the region growing within one slice and across the sequence of slices within the same patient. Refining the proposed approach for single organ segmentation and tuning the parameters to make the approach work for other organs will allow us develop a scheme for combined organ segmentation. It is expected that this combined multiple organ segmentation approach will produce better results than each one of the individual filters.

References

1. Olabarriaga, S.D., Smeulders, A.W.M.: Interaction in the segmentation in medical images: A survey. Medical Image Analysis, Vol. 5, No. 2, (2001) 127-142.
2. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. Annual Review of Biomedical Engineering, Vol. 2, (2000) 315-337
3. Channin, D., Raicu, D., Furst, J., Xu, D.H., Lilly, L., Limpsangsri, C.: Classification of Tissues in Computed Tomography using Decision Trees. The 90th Sci. Assembly and Annual Meeting of Radiology Society of North America,(2004)
4. Susomboon, R., Raicu, D.S., Furst, J.D.: Automatic Single-Organ Segmentation in Computed Tomography Images. IEEE Int. Conf. on Data Mining, December (2006)
5. Jackson, S., Thomas, R.: Cross-Sectional Imaging Made Easy. Elsevier Inc, (2004)
6. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society., Vol. 39,(1977)
7. Yamazaki, T.: Introduction of EM algorithm into color image segmentation. Proceedings of ICIRS'98, (1998) 368-371
8. McLachlan, C.J., Krishnan, T.: The EM Algorithm and Extensions. Wiley, NY (1997)
9. Martnez-Us, A., Pla, F., Garca-Sevilla, P.: A Quadtree-based Unsupervised Segmentation Algorithm for Fruit Visual Inspection. 1st Iberian Conference on Pattern Recognition and Image Analysis, (2003) 510-517
10. Smith, J.M., Chang, S.F.: Quad-tree segmentation for texture-based image query. ACM Multimedia, (1994) 279-286
11. Furst, J.D, Susomboon, R., Raicu, D.S.: Single Organ Segmentation Filters for Multiple Organ Segmentation. IEEE 2006 Int. Conf. of the Engineering in Medicine and Biology Society, (2006)

12. Chang, K.I., Bowyer, K.W., Sivagurunath, M.: Evaluation of texture segmentation algorithms. IEEE Conf. on Computer Vision and Pattern Recognition, (1999) 294-299
13. Pham, M., Susomboon, R., Disney,T., Raicu, D., Furst, J.: A Comparison of Texture Models for Automatic Liver Segmentation. SPIE Medical Imaging Conf., San Diego, CA (2007)
14. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. IEEE Trans. on Systems, Man, and Cybernetics, (1973) 610-621
15. Susomboon, R., Raicu, D.S., Furst, J.: Co-occurrence Matrix Invariance to Patient Size in Computed Tomography. CTI Research Symposium, DePaul University, (2007)
16. Varma, M., Zisserman, A.: Texture classification: are filter banks necessary?. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, (2003)
17. Gerig, M., Chakos, M.: Valmet: a new validation tool for assessing and improving 3D object segmentation. MICCAI 2001, Springer, Berlin, (2001) 516-523.