

Liver tumor segmentation in CT images.

Lauren K. Haaitzma

Abstract

Liver tumor segmentation and volumetry can help medical experts determine the rate of tumor growth and the effectiveness of cancer treatment. This thesis compares the algorithms and results of 9 key publications and will attempt to determine which algorithms performed best. Furthermore, suggestions to improve performance benchmarking are given. Due to the different goals and evaluation metrics used, not all publications could be compared with each other. Of the algorithms that were compared, one was found to be best suited for small tumors (diameter < 5 cm), one is most generally applicable and one automatic method is easiest to use.

1. Introduction

Cancer is one of the most common causes of death in the modern world, liver cancer ranking among the three most deadly [WHO, 2012]. Segmentation of liver tumors can help oncologists determine changes in tumor size. This information can then be used to evaluate the patient's response to treatment and, if necessary, adapt the therapy. While one- and two dimensional measurements are often used in clinical practice, three dimensional measurements provide a more accurate indication of tumor size [Smeets *et al.*, 2010; Yim *et al.*, 2006, Zhou *et al.*, 2010]. Currently, computed tomography (CT) is one of the most widely used imaging modalities for tumor detection and diagnosis, due to its high spatial resolution, fast imaging speed, wide availability and relatively low cost compared to MRI [Li *et al.*, 2012]. For these reasons, there has been a lot of interest in developing a good algorithm for liver tumor segmentation and volumetry [Deng and Du, 2008; Smeets *et al.*, 2010; Li *et al.*, 2012; Zhou *et al.*, 2010; Ray *et al.*, 2012; Häme *et al.*, 2012; Yim *et al.*, 2006; Kumar *et al.*, 2011; Gunasundari *et al.*, 2012; Massoptier *et al.*, 2008]. A good algorithm should work for most types of liver tumors and give reliable and reproducible results. The resulting segmentation should be an accurate delineation of the tumor volume.

Liver tumor segmentation is a challenging problem due to a significant variation in location, shape, intensity and texture (Figure 1 shows a selection of different liver tumors). This makes it difficult to develop a general algorithm that works for all cases [Li *et al.*, 2012].

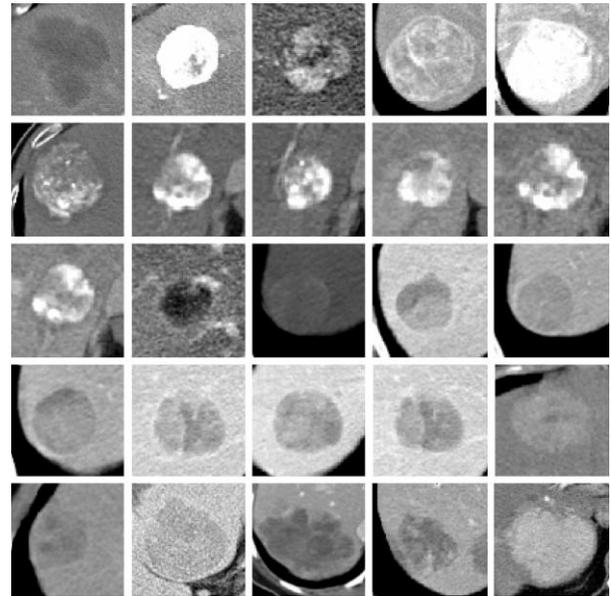


Figure 1. A selection of different liver tumors to illustrate the large variation in location, shape, intensity and texture [Li *et al.*, 2012]

This thesis compares algorithms for liver tumor segmentation in CT datasets according to the above criteria.

2. ROI Selection

Regions of interest (ROIs) are often used to reduce the computer's workload by reducing the amount of pixels to consider, thus reducing the processing time. It can also be used to reduce the amount of different tissue types to consider; when looking for liver tumors, the dataset to consider usually consist of the whole abdomen (Figure 2 shows an abdominal scan). If the region of interest is then determined to only encompass liver and tumor tissue, the other tissues in the abdomen won't need to be considered, thus simplifying the segmentation. Most images in Figure 1 are examples of a ROI with only liver and tumor tissue.

The region of interest can be determined manually, by indicating points that belong to liver tissue and others that belong to tumor tissue [Zhou *et al.*, 2010; Ray *et al.*, 2012], by indicating the approximate center and radius of the tumor [Häme *et al.*, 2012; Smeets *et al.*, 2010; Li *et al.*, 2012] or by drawing a region around the tumor [Zhou *et al.*, 2010; Yim *et al.*, 2006].

Given that the whole liver can also be considered the region of interest for liver tumor segmentation, some researchers opt for automatically segmenting the liver and searching



Figure 2. Abdominal CT scan. The liver is delineated by a blue contour, the tumor is indicated by a yellow contour. [Kumar *et al.*, 2011]

the resulting region for tumors [Kumar *et al.*, 2011; Gunasundari *et al.*, 2012; Massoptier *et al.*, 2008]. These three publications use the fact that the liver is the biggest organ present in abdominal scans and that it is homogeneous with regard to image intensity. Additionally, these algorithms are 2D techniques, working on a slice-by-slice basis.

Kumar *et al.* [2011] and Gunasundari *et al.* [2012] both use histogram analysis to select the pixels around the highest peak, excluding background and bone values. Figure 3 shows an intensity histogram of an abdominal scan. The red oval indicates the intensities that are likely to be part of the liver. All pixels in this range of intensities are then extracted to form a simplified image consisting of most of the liver pixels and some smaller structures with similar intensity values. These smaller structures are removed by means of morphological operations such as erosion [Kumar *et al.*, 2011] or closing followed by opening [Gunasundari *et al.*, 2012]. Kumar *et al.* [2011] finish the liver segmentation with region growing. The seed point required for region growing is automatically found by taking the centroid of the biggest connected area in the simplified and eroded image.

Massoptier *et al.* [2008] take a slightly different approach, still using image intensity information. First each slice is subdivided into 64 squared regions, from which the mean intensity and standard deviation is calculated. All regions with minimal standard deviation (which is defined as less than 1% of the peak value without background) are then sorted by ascending

intensity value. Because the liver is the biggest organ with relatively homogeneous image intensity, the median of all selected regions will typically belong to the liver. This region's mean pixel intensity and standard deviation are used to cluster all pixels belonging to the liver, by checking if the pixel's intensity value is within two standard deviations of the mean. Lastly, morphological operations are used to get rid of smaller non-liver structures and to fill holes in the liver itself.

To further smooth the liver surface, a Gradient Vector Flow (GVF) snake [Xu and Prince, 1998] is used. A GVF snake is an active contour that evolves towards an image boundary, influenced by internal and external forces. An example internal force is curvature. An external force in the case of GVF snakes is the gradient vector flow force, which is derived from the image's intensity information.

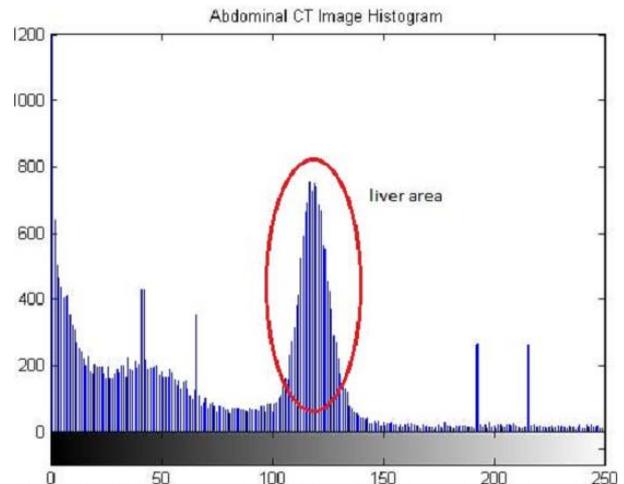


Figure 3. Histogram analysis. Liver intensity values selected for simplified image are indicated by the red oval. [Kumar *et al.* 2011]

3. Tumor Segmentation

This section will briefly explain methods used for liver tumor segmentation and their applications.

3.1 Clustering methods

One way to segment an image is through the use of clustering. Clustering divides data into clusters, based on its properties such as intensity value. Elements within a cluster should be as similar as possible while differences between clusters should be maximized. There are two types of clustering: hard clustering and fuzzy (or soft) clustering. Hard clustering divides data into distinct clusters, and each data element belongs to exactly one cluster. Fuzzy clustering indicates for each data element how strong the association with each cluster is [Suetens, 2002].

Clustering methods can be either supervised or unsupervised. In the case of a supervised method, the centroids of the clusters are calculated based on points indicated by the user. Smeets *et al.* [2010] indicates a point inside the tumor and one at approximately twice the tumor radius from the tumor center. These points are then used to estimate the mean and standard deviation for the liver and tumor clusters. This information is used for fuzzy clustering [Suetens, 2002].

Unsupervised methods use image information and optimization functions to calculate the optimal number of centroids and their locations.

K-means clustering

K-means clustering is an example of a hard clustering method. Massoptier *et al.* [2008] use this method to cluster pixels as belonging to either the liver or tumor tissue. They use the intensity histogram of the segmented liver, decomposed with a Haar wavelet [Stollnitz *et al.*, 2012] to find the k most represented coefficients. k in this case is two, for liver and tumor tissue. These two coefficients are then used as the centroids of two Gaussian functions used to estimate the intensity histogram. An expectation maximization algorithm [Dempster *et al.*, 1977] is used to assign the pixels to the corresponding cluster.

Supervised machine learning-based voxel classification

Zhou *et al.* [2008] start by training a support vector machine (SVM)-based classifier with manually selected training samples of tumor tissue and non-tumor tissue (liver and otherwise). A ROI is manually defined in a slice approximately dividing the tumor in two. The SVM classifier is used for voxel classification within the ROI to segment the tumor in this first slice.

After that, the tumor contour is both eroded and dilated by 2 to 3 pixels to create two contours. These two contours are then mapped to the neighboring slices. The larger contour is used to indicate the new ROI. The smaller contour is used as a new sample for SVM classifier training. Subsequently, the newly trained SVM classifier is used for tumor segmentation in the two new ROIs.

This propagation step continues until there are no more tumor bearing slices.

Fuzzy clustering

A widely used type of fuzzy clustering is fuzzy c-means (FCM). In FCM, a dataset is divided into a number of clusters. Every data point can belong to multiple clusters, with a membership function ranging between zero and one indicating how

much it belongs to each cluster. The sum of all memberships for each data point should be one [Bezdek *et al.*, 1984].

The FCM algorithm was first used for liver tumor segmentation by Yim *et al.* [2006]. They compared manual segmentation with semi-automatic segmentation using FCM. The algorithm starts with Gaussian smoothing and the manual delineation of a mask, about 0.5 cm outside of the tumor boundary. Then the FCM algorithm is applied with two clusters: liver tissue and tumor tissue.

Kumar *et al.* [2011] propose an iterative variant of the FCM algorithm, called alternative fuzzy c-means (AFCM), which they say is less sensitive to noise and outlying points, as well as to dissimilar cluster shape and size. Instead of using the Euclidian distance $d^2(x, y) = ||x - y||^2$ they use a new distance function $d^2(x, y) = 1 - \exp(-\beta ||x - y||^2)$, where β is a parameter greater than zero, that is estimated based on the inverse of the variance. This distance function is used to update the membership functions and cluster centers.

3.2 Region growing

Region growing uses a seed point, from which the segmented region grows by adding all pixels with similar properties.

Region growing with knowledge based constraints

Wong *et al.* [2008] use 2D region growing with knowledge based constraints on all tumor bearing slices. The constraints keep the size and shape of the segmented region within specified bounds.

Bayesian rule-based 3D region growing

Qi *et al.* [2008] assume that the intensity distribution of a tumor can be modeled by combining multiple Gaussians. Multiple seed points are placed within the tumor. The mean and standard deviation of the Gaussians are estimated within a cube around the seed voxels. One seed point in the liver is used to estimate the liver's Gaussian. Subsequently, during region growing, the Bayesian decision rule checks for every voxel if its intensity has a greater chance of belonging to liver or tumor tissue - based on the Gaussians estimated earlier - and classifies it into the most likely class.

Iterative watershed

The watershed method considers the grayscale image to be analogous to a topological surface with valleys and mountains. Region growing in this analogy is like flooding catchment basins, which are the local minima of the image. Flooding the basins creates ridge lines around them. These ridge lines separate the different regions [Ray *et al.*, 2008]

Because medical images often have high overall noise variance, there is a risk of oversegmentation due to a high density of local minima. To overcome this problem, many watershed methods use marker points, placed inside and outside the tumor by the user, to indicate where to start flooding. The watershed method is then applied to the GVF field transformation of the image. This gives an initial segmentation called level 0. Figure 4 shows the results of the first three iterations of the iterative watershed method.

The iterative watershed method uses the resulting boundary as an additional set of markers and the watershed method is applied again. This results in a segmentation called level 1, consisting of one region consisting of pixels inside the tumor, one fuzzy region and one region outside the tumor.

Next, the boundaries are again indicated as additional sets of markers and the watershed method is applied one last time, resulting in level 2. Level 2 consist of one region inside the tumor, three fuzzy regions and one region outside the tumor. The fuzzy regions are, going from outside to inside, increasingly likely to belong to the tumor.

The second most outer boundary in level 2 was found to be closest to the reference boundary. [Ray *et al.*, 2008]

3.3 Level set method

To segment a structure in an image, an initial closed curve can be expanded and contracted based on for example image intensity, gradient or texture information, until it converges on the boundary of the structure. If, however, the structure consist of multiple disconnected regions, splitting the closed curve is not straightforward. By embedding the initial position of the front as the zero level set of a time dependent function Φ one dimension higher than the front, changes in topology are possible, because the higher dimensional level set function Φ gives a smooth representation across the whole image domain [Sethian, 1999].

Level sets evolve influenced by a speed function. This function can be calculated using image information, such as image intensity, texture information, or other properties [Cremers *et al.*, 2007]. By taking the zero level set $\Phi = 0$ (See Figure 5) at a certain time, the front at this time can be retrieved.

The level set method can be used for segmentation by placing an initial contour inside the structure of interest and letting it evolve based on the speed function.

Smeets *et al.* [2010] calculate a speed image by applying a filter to the image they retrieved with fuzzy clustering (see section 3.1). The initial level set is obtained by taking the minimal cost path in a spiral scanned image (Figure 6) and converting it back to 3D. The level set method then evolves the curve towards the tumor boundary.

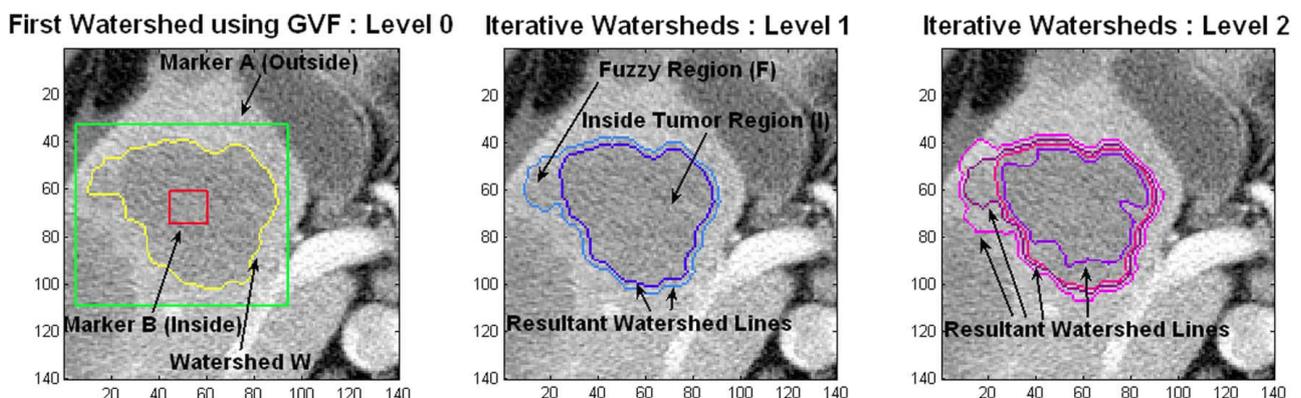


Figure 4. The first three iterations of the iterative watershed method. Level 0 gives the initial segmentation. Level 1 gives one fuzzy region, one region consisting of pixels inside the tumor and one region outside the tumor. Level 2 gives three fuzzy regions, one inside region and one outside region. Going from the outer fuzzy region to the inner one, the likelihood of those pixels belonging to the tumor increases. The second most outer boundary in level two was found to best delineate the tumor [Ray *et al.*, 2008].

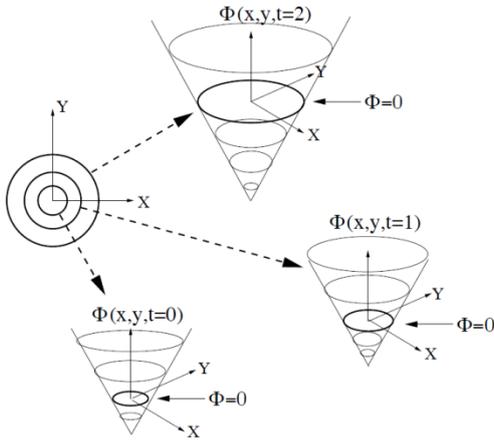


Figure 5. Evolution of a level set over time. The front in this case is a circle, growing with constant speed. This can be represented by a cone shaped embedding function, growing over time (clockwise from the bottom). By taking the zero level set $\Phi = 0$, the front can be retrieved. [Sethian 1999]

Li *et al.* [2012] argue that both edge-based as well as region-based level set methods aren't optimal for liver tumor segmentation. Edge based LSM relies on the premise that liver tumors have distinct boundaries, which is often not the case, leading to leakage of the boundary. A region based LSM can work well in the absence of distinct boundaries, but fails in the case of low contrast, such as between liver and tumor in CT images.

They propose a unified level set method, combining an enhanced edge function, bi-directional speed function and probabilistic region competition. The enhanced edge function can use either edge information, prior knowledge or a combination thereof. The bidirectional speed function uses a signed balloon force that can either push or pull, based on FCM information. FCM is also used for the probabilistic estimation of liver tumors.

3.4 Hidden Markov Measurement Field

The Hidden Markov Measurement Field method segments an image by partitioning it into a number of non-overlapping regions with either constant properties (such as intensity, texture or color) or properties that follow a simple model [Marroquin *et al.*, 2003]. The challenge is finding a label field that tells for each pixel to what region it belongs (See Häme *et al.* [2012] for a detailed description on how to obtain the label field).

4. Results

There are many evaluation metrics to quantitatively measure the performance of a segmentation algorithm. Which are used depends on the focus of the research and whether 2D or 3D segmentation is used, among other factors. This does however potentially complicate comparing the performances of different algorithms. Since the Grand Challenge II workshop on liver tumor segmentation, there has fortunately been some standardization with regards to the evaluation metrics used: to compare the workshop's results, Volumetric Overlap Error (VOE), Relative Absolute Volume Difference (RAVD), Average Symmetric Surface Distance (ASSD) and Maximum Symmetric Surface Distance (MSSD) were used [Deng and Du, 2008]. In most papers since, these have been used to evaluate the algorithms [Smeets *et al.*, 2010; Zhou *et al.*, 2010; Häme *et al.*, 2012; Li *et al.*, 2012].

Deng and Du [2008] define these metrics as:

$$VOE (\%) = \left(1 - \frac{Seg \cap Ref}{Seg \cup Ref}\right) \times 100\%$$

$$RAVD (\%) = \frac{|Seg - Ref|}{Ref} \times 100\%$$

$$ASSD (mm) = \frac{\sum_{a \in A} [\min_{b \in B} \{dist(a, b)\}] + \sum_{b \in B} [\min_{a \in A} \{dist(a, b)\}]}{N_A + N_B}$$

$$MSSD (mm) = \max \{ \max_{a \in A} [\min_{b \in B} \{dist(a, b)\}], \max_{b \in B} [\min_{a \in A} \{dist(a, b)\}] \}$$

with Seg and Ref indicating the volume of the segmentation algorithm and the ground truth respectively. A and B denote the surfaces of segmentation and reference and a and b are mesh points on the corresponding surfaces. $dist(a, b)$ is the distance between a and b and N_A and N_B are the number of points on A and B. For all metrics lower values indicate better segmentation.

In table 1 the performance of the algorithms from Smeets *et al.* [2010], Zhou *et al.* [2010] Häme *et al.* [2012] and Li *et al.* [2012] is compared, based on the above evaluation metrics. The numbers denote the average performance over all tested cases. All four publications were evaluated on the publicly available LTSC08 dataset.

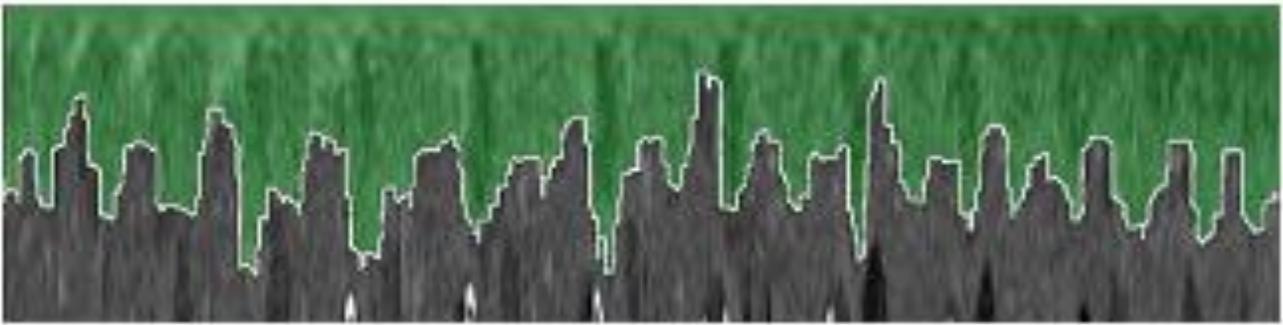


Figure 6. A spiral scanned image. Spiral scanning converts a 3D image into a 2D projection by using scan lines in a spiral like pattern. The tumor center is on top. The green area is the tumor, the boundary is determined by the minimal cost path algorithm.[Smeets *et al.*, 2010]

Table 1 Comparison of different semi-automatic methods. All results are averages of all test cases per publication. Dashes indicate metrics which were not used or which differed from those used in the other publications.

	VOE (%)	RAVD (%)	ASSD (mm)	MSSD(mm)
Smeets	32,6	17,91	1,97	10,1
Li	26,3	-	1,06	8,6
Häme	30,3	23,50	1,90	8,1
Wong	30,5	17,93	2,05	-
Zhou	25,7	17,92	1,57	-
Qi	40,0	34,74	4,12	-

Kumar *et al.* [2011] and Massoptier *et al.* [2008] can be compared on the basis of the Dice Similarity Coefficient (DSC), defined as:

$$DSC = \frac{2 \times [Ref \cap Seg]}{Ref + Seg} \times 100\%$$

Kumar *et al.* [2011] call it Spatial Overlap η , but the definition is the same, except for the conversion to percents. Table 2 shows the comparison. A higher score signifies better performance. The automatic method from Gunasundari *et al.* [2012] focuses on classifying the type of tumor, so the paper does not use any evaluation metrics regarding segmentation, hence it cannot be quantitatively compared with other algorithms.

Table 2 Comparison of two automatic methods. The Dice Similarity Coefficient (DSC) gives the amount of overlap between the reference image and the segmented image. A higher score signifies a better segmentation.

	DSC (%)
Kumar	91,7
Massoptier	88,9

The publications of Yim *et al.* [2006] and Ray *et al.* [2008] do not have metrics that can be used to compare their performance with other algorithms.

Massoptier *et al.* [2008], Zhou *et al.* [2012] and Häme *et al.* [2012] attempt to measure the processing time. Massoptier *et al.* [2008] find an average processing time of 11,4 seconds per 512x512 slice. Zhou *et al.* [2012] only estimate the processing time, estimating 5 – 30 minutes per lesion. Lastly, Häme *et al.* [2012] find processing times between less than 30 seconds and 15 minutes per tumor. Important to note is the fact that processing time highly depends on tumor size and computer configuration, among other factors [Häme *et al.*, 2012]. Häme *et al.* [2012] focus on tumors smaller than 5 cm in diameter, because their target application is radio frequency ablation (RFA) treatment planning. When testing the algorithm on preoperative RFA tumors, the average processing time goes down to 33 seconds per tumor, and the algorithm's performance increases, to an average VOE of 29,6, a RAVD of 17,75, an ASSD of 0,89 and a MSSD of 5,1. This indicates the algorithm's specialty is small tumors. The average time taken to do manual contouring of these small tumors was found to be 254 seconds per tumor [Häme *et al.*, 2012].

5. Discussion

Because every publication uses different evaluation metrics or has a different focus, it can be quite hard to quantitatively compare the different papers. Some were more focused on exactly segmenting the tumor [Smeets *et al.*, 2010; Li *et al.*, 2012; Zhou *et al.*, 2010; Ray *et al.*, 2012; Häme *et al.*, 2012; Kumar *et al.*, 2011; 2012; Massoptier *et al.*, 2008. Others were more interested in whether the volume was correct. [Yim *et al.*, 2006]. Others still were not really interested in an exact segmentation, but wanted to use the segmentation to automatically classify the type of tumor [Gunasundari *et al.*, 2012]. For this reason quantitative comparison could only be done between publications with similar metrics and goals.

For future studies, the LTSC08 dataset should be included in the evaluation and the same metrics

should be used, to facilitate a more straightforward benchmarking. Expanding the dataset to include data acquired on different phases and with tumors with different histological type, imaging properties and appearance would greatly improve the dataset and would allow for a more broad evaluation of the algorithms [Zhou *et al.*, 2010].

Most algorithms were only tested on datasets with hypodense tumors. Hypodense tumors show up in CT scans as darker than the surrounding liver, as opposed to hyperdense tumors. Only a few algorithms were tested on datasets containing both hyperdense and hypodense tumors [Smeets *et al.*, 2010; Häme *et al.*, 2012; Li *et al.*, 2012]. Other algorithms might also work on hyperdense tumors, but because this wasn't tested, this cannot be confirmed.

As can be seen in table 1, some metrics are not filled in. This is because some publications used different metrics or chose not to use some of them. Instead of the relative absolute volume difference, Li *et al.* [2012] used the relative volume difference, in which negative numbers indicate that the segmented volume was smaller than the reference volume. Unfortunately, comparing absolute and non-absolute means does not give a correct comparison, so Li *et al.*'s Relative Volume Difference could not be compared with the others. Zhou *et al.* [2010] decided not to use the Maximum Symmetric Surface Distance, arguing that it is similar to the Average Symmetric Surface Distance.

Looking at table 1, either the algorithm from Zhou *et al.* [2008] or the Li *et al.* [2012] algorithm performed best on average. Comparing the mean VOEs and ASSDs of both best scoring algorithms using a t-test, however, showed that the difference was not statistically significant (P value of 0.8895 and 0.3171 respectively). This means that both performed similarly well. The Zhou *et al.* [2008] algorithm is less generally applicable, failing in the case of blurred boundaries and low contrast, as opposed to the Li *et al.* [2012] algorithm.

An ANOVA test showed that only the Qi *et al.* [2008] algorithm had a statistically different mean, in the sense that it was the worst performing algorithm. The other algorithms did not differ significantly in their average performance.

Comparing the DSCs of Kumar *et al.* [2011] and Massoptier *et al.* [2008] using a t-test indicated that this difference was not statistically significant either (P = 0.0711).

The DSC is very similar to the VOE, making it possible to compare the VOE of table 1 with the DSC in table 2 by applying $DSC = 100 - 0.5 \times VOE$ (See Table 3).

It is however important to note that this compares 3 dimensional overlap with 2 dimensional overlap, which could lead to an unfair advantage of one over the other.

Table 3 Comparison of the DSC of different semi-automatic methods. Converted from the VOE in table 1. The first two were evaluated from 2D slices, the rest was based on 3D segmentation.

	DSC(%)
Kumar	(91,7)
Massoptier	(88,9)
Smeets	83,7
Li	86,9
Häme	84,9
Wong	84,8
Zhou	87,2
Qi	80,0

Of the 3 dimensional algorithms, the Zhou *et al.* [2008] DSC comes close to the Massoptier *et al.* [2008] DSC, especially considering the overall lower scores for evaluation in 3D.

Kumar *et al.* [2012] also use another metric, called the Coefficient of Similarity ϵ , defined as:

$$\epsilon = 1 - \frac{|Ref-Seg|}{Ref}$$

This is very similar to the RAVD defined earlier, since both look at the absolute difference between reference and computed segmentation relative to the reference segmentation. The Coefficient of Similarity can be converted to RAVD by applying $RAVD = (1 - \epsilon) \times 100\%$. This gives a mean RAVD of 10,17%, which is considerably lower than the RAVDs in table 1. Worth noting is the fact that Kumar *et al.* [2011] look at segmentation in 2D slices, as opposed to volumes for the publications in table 1. This might have an effect on the RAVD score, which is why it wasn't included in table 1.

Another factor to consider when comparing different algorithms is the use of different datasets. For example, if two algorithms perform best for hypodense tumors and one algorithm is evaluated using a dataset with many hypodense tumors, while the other has mostly hyperdense tumors, the algorithm with the more favorable dataset will perform better. This may be the case for the comparison between Li *et al.* [2012] and

Zhou *et al.* [2008], where Zhou *et al.* [2008] used only the LTSC08 dataset, while Li *et al.* [2012] used both the LTSC08 dataset and an additional dataset from the National University Hospital of Singapore, containing tumors of different densities. The fact that both algorithms still perform similarly well, indicates that the Li *et al.* [2012] algorithm works well with all kinds of liver tumors and hence is the more robust algorithm.

The following section will discuss the expected performance of the methods discussed under sub-optimal conditions.

Three challenging cases will be considered:

- Tumors with low contrast with the surrounding tissue.
- Tumors with an inhomogeneous intensity distribution.
- Non-convex tumors.

Low contrast

In the case of low contrast, the boundary between tumor and the surrounding tissue can be quite ambiguous. This is challenging even for manual segmentation, but it's especially challenging for intensity based methods.

Clustering methods like Yim *et al.* [2006], Massoptier *et al.* [2008], Zhou *et al.* [2008], Kumar *et al.* [2011] and Gunasundari *et al.* [2012] for example will likely have trouble accurately segmenting the tumor because the intensities of the different clusters are very similar, making the correct division quite ambiguous. Zhou *et al.* [2008] attempt to reduce the ambiguity by penalizing misclassified training points, but clustering remains a problem because of the amount of penalized training points in low contrast datasets. Yim *et al.* [2006] mention in their discussion that the algorithm performs poorly in the case of a low contrast to noise ratio, this is likely because the FCM algorithm that is used has trouble distinguishing between liver and tumor tissue. The Gunasundari *et al.* [2012] algorithm isn't very well suited for accurately segmenting liver tumors, even when there is enough contrast, because their focus is on classification of tumor type. This means that challenging tumors will be segmented even less accurately.

Region growing [Wong *et al.*, 2008; Qi *et al.*, 2008] and watershed [Ray *et al.*, 2008] methods will suffer from leakage due to the ambiguous boundary between liver and tumor tissue. The Qi *et al.* [2008] algorithm uses Gaussian fitting for its

region growing, which will perform poorly if there is too little difference in intensity between liver and tumor tissue. Leakage in the Wong [2008] algorithm may be reduced by the embedded constraints, though to what extent depends on the exact constraints. The constraint used in the paper is that the initial segmented region should cover at least half the size of the ROI to prevent under-segmentation. This means that the only factor preventing leakage is the edge of the ROI. The Ray *et al.* [2008] algorithm uses edge information in the form of a Gradient vector flow (GVF) field transformation. However, edges won't be very clear due to the low contrast.

Smeets *et al.* [2010] mention that their method performs best with high contrast tumors and that accuracy is insufficient in most low-contrast cases. This may be due to the ambiguous boundary in the spiral scanned image, making it more difficult to find the correct minimal cost path.

The Häme *et al.* [2012] paper states that the algorithm performed very well even for tumors with extremely low contrast and ambiguous borders, as long as they are small to medium sized. Method performance decreases with increase in tumor size.

The best performing method in the case of low contrast tumors is the Li *et al.* [2012] method. By combining region-based and edge-based level set methods (LSMs), it is much better suited for finding low contrast tumors than pure region based LSMs.

Non-homogeneous intensity distribution

In the case of tumors with inhomogeneous intensity distribution, all methods that assume that liver and tumor intensity distribution can be estimated by two Gaussian curves [Massoptier *et al.*, 2008; Qi *et al.*, 2008] will perform poorly, because inhomogeneous intensity distribution cannot be estimated by a simple curve. Intensity-based clustering methods [Yim *et al.*, 2006; Smeets *et al.*, 2010; Kumar *et al.*, 2011; Gunasundari *et al.*, 2012] will also have a problem because they cluster together pixels with similar properties. Some tumor intensity values may be the same as the liver, leading to misclassification. Kumar *et al.* [2011] for example assume three clusters: background, tumor and liver, from low to high intensity values. If the tumor intensity distribution is inhomogeneous, the lighter parts of the tumor may be misclassified as liver tissue.

Smeets *et al.* [2010] mention underperformance when the intensity distribution

has a large standard deviation. This is likely due to the fact that the algorithm uses Fuzzy C Means for their speed image.

In the case of region growing, inhomogeneous tumors may cause some under-segmentation if parts of the tumor have different intensity from the region growing part. The region growing algorithm may dismiss these differing parts as non-tumor on account of their different intensity [Wong *et al.*, 2008; Qi *et al.*, 2008].

Because the Ray *et al.* [2008] algorithm uses the GVF field transformation of the image, inhomogeneity shouldn't be an issue, because it's based on edge information, not on intensity. As long as there is good contrast with the surrounding tissue, inhomogeneous tumors should be segmented reasonably well in at least one of the levels.

The Häme *et al.* [2012] paper states that this method's performance deteriorates with high levels of inhomogeneity in tumors because these tumors may not always be sufficiently represented by the training data.

The algorithm that is best suited for segmenting inhomogeneous tumors is the Li *et al.* [2012] algorithm. This is shown by its high performance when testing the algorithm on inhomogeneous tumors.

Non-convex tumor shape

Most of the methods shouldn't have much trouble handling non-convex tumors, because no prior knowledge about shape is assumed. If prior shape knowledge was assumed, the segmentation of tumors greatly deviating from this shape may be limited in its reach. The Häme *et al.* [2012] algorithm is one of the few that assumes a roughly spherical tumor shape, which is used mainly in the ROI determination and the post-processing step. The ROI determination can be altered to account for non-spherical tumors relatively easily. The post processing step however removes extremities, which could belong to the tumor if the tumor is non-convex.

In the Smeets *et al.* [2010] algorithm it is assumed that the tumor is roughly spherical when placing the user specified points for initialization. If the tumor is very long in one direction, the sphere encompassing the tumor will also increase in size, which means that more non-tumor tissue is included within the spiral scanning sphere. This may affect the spiral scanning efficiency. Furthermore, finding the lowest cost path in the spiral scanned image may prove difficult if the tumor has many concave corners and small

extremities. On the other hand, the level set method is meant to fine tune the lowest cost path and it should be able to find the boundary of non-convex tumors as well.

Most clustering methods ([Yim *et al.*, 2006; Massoptier *et al.*, 2008; Zhou *et al.*, 2008; Kumar *et al.*, 2011; Gunasundari *et al.*, 2012]) shouldn't have trouble segmenting non-convex tumors, because no shape restrictions are used. Out of these, the Gunasundari *et al.* [2012] algorithm will likely perform poorest because it uses quite a rough post processing step that could remove parts of the non-convex tumor. The Kumar *et al.* [2011] algorithm will likely perform best, assuming tumor shape doesn't affect clustering performance, because it performed best for convex tumors.

Like the clustering methods, the region growing methods ([Wong *et al.*, 2008; Qi *et al.*, 2008]) don't use prior shape knowledge, and hence shouldn't have much trouble segmenting non-convex tumors.

The Ray *et al.* [2008] algorithm uses the GVF field transformation of the image. This is based on edge information, which should preserve any non-convex tumor shapes.

For non-convex tumors, the Li *et al.* [2012] algorithm is again expected to perform best. The variable balloon force used in their unified level set method helps the contour move towards non-convex shapes with great accuracy.

Overall the Li *et al.* [2012] algorithm will perform best for all three challenging types of tumors. The versatility of the unified level set method is shown on a number of challenging tumors, which are segmented quite accurately overall (See table 1).

6. Conclusion

Based on the metrics, it isn't possible to say with certainty which algorithm performed best. On average, the Zhou *et al.* [2008] algorithm and the Li *et al.* [2012] algorithm performed slightly better than the rest. The Li *et al.* [2012] algorithm being the more robust algorithm.

Of the automatic methods, the algorithm from Kumar *et al.* [2011] performed slightly better than the one from Massoptier *et al.* [2008]. Unfortunately, the processing time for the Kumar *et al.* [2011] algorithm wasn't measured, so it is unknown whether the algorithm is faster or slower than the Massoptier *et al.* [2008] algorithm, or any other algorithm for that matter. Considering however that automatic methods are more

computationally expensive than semi-automatic methods [Häme *et al.*, 2012] and considering that the Massoptier *et al.* [2008] algorithm is quite similar, the processing time will likely be similar to the Massoptier *et al.* [2008] algorithm and slower than most semi-automatic algorithms. If ease of use and accuracy are more important than speed, the Kumar *et al.* [2011] algorithm is the algorithm of choice.

The Häme *et al.* [2012] algorithm is best suited for small tumors (diameter < 5 cm) of all types and densities and may be used for radio frequency ablation treatment planning. It is a fast and accurate method when it comes to small tumor segmentation.

The Li *et al.* [2012] algorithm works for all types of tumor and contrast, by changing the balance between edge based and region based level set method. Furthermore, to initialize, it requires just two input points to indicate the approximate tumor center and the radius of a sphere around the tumor.

If future liver tumor segmentation studies were to use the same dataset and the same evaluation metrics, future performance benchmarking would be made much easier. My suggestion would be to use the dataset and metrics used for the Liver Tumor Segmentation Challenge 2008 [Deng and Du, 2008] as these give a good indication of segmentation accuracy and because they have been used by most papers on liver tumor segmentation since 2008. The dataset itself would benefit greatly from the addition of data acquired at different phases, containing tumors with different imaging properties, to more broadly test the algorithms.

References

Key references have been made bold in the reference list

Bezdek J. C., Ehrlich R., Full W. (1984) FCM: The fuzzy c-means clustering algorithm, *Computers & Geosciences*, Volume 10, Issues 2–3, pp: 191-203

Cremers, D., Rousson, M. and Deriche, R. (2007) A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. *International journal of computer vision*, Volume: 72, Issue: 2 pp: 195-215

Dempster A. P., Laird N. M., Rubin D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B*, Volume 39, Issue 1, pp: 1–38

Deng, X., Du, G. (2008) Editorial. In: 3D Segmentation In The Clinic: A Grand Challenge II Workshop at MICCAI 2008.

Dubus L., Vilgrain V., Gayet M., Zappa M., Abaleo L., Cooman A., Orioux G. (2011) Comparison of semi-automated and manual methods to measure the volume of liver tumours on MDCT images, *European radiology*, Volume: 21, Issue: 5, pp: 996-1003

Gunasundari S., Suganya Ananthi M. (2012) Comparison and Evaluation of Methods for Liver Tumor Classification from CT Datasets, *International journal of computer applications*, Volume: 39, Issue: 18, pp: 46-51

Häme Y., Pollari M. (2012) Semi-automatic liver tumor segmentation with hidden Markov measure field model and non-parametric distribution estimation, *Medical Image Analysis*, Volume: 16, Issue: 1, pp: 140–149

Kumar S. S., Moni R. S., Rajeesh J. (2011) Automatic Segmentation of Liver and Tumor for CAD of Liver, *Journal of advances in information technology*, Volume: 2, Issue: 1, pp: 63-70

Li B. N., Chui C. K., Chang S., Ong S. H. (2012) A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images, *Expert systems with applications*, Volume: 39, Issue: 10, pp: 9661-9668

Marroquin J. L., Santana, E., Botello, S. (2003) Hidden Markov measure field models for image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 25, Issue 11, pp: 1380–1387.

Massoptier L., Casciaro S. (2008) A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from CT scans, *European radiology*, Volume 18, Number 8, 2008, pp: 1658-1665

Qi Y., Xiong W., Leow W. K. *et al.* (2008) Semi-automatic segmentation of liver tumors from CT scans using Bayesian rule-based 3D region growing. In: 3D Segmentation In The Clinic: A Grand Challenge II Workshop at MICCAI 2008,

Ray S., Hagge R., Gillen M., Cerejo M., Shakeri S., Beckett L., Greasby T., Badawi R. D. (2008) Comparison of two-dimensional and three-dimensional iterative watershed segmentation methods in hepatic tumor volumetrics, *Medical Physics*, Volume: 35, Issue: 12, pp: 5869-5881.

Sethian, J. A. (1999). Level Set Methods and Fast Marching Methods. New York: Cambridge University Press.

Smeets, D., Loeckx, D., Stijnen, B., De Dobbelaer, B., Vandermeulen, D., Suetens, P. (2010) Semi-automatic level set segmentation of liver tumors combining a spiral-scanning technique with supervised fuzzy pixel classification, *Medical image analysis*, Volume: 14, Issue: 1, pp: 13-20

Stollnitz, E. J., DeRose, T. D., Salesin, D. H., Wavelets for Computer Graphics: A Primer, *Computer Graphics and Applications*, Volume: 15, Issue: 4, pp: 75 - 85

Suetens, P. (2002). Fundamentals of Medical Imaging. Cambridge University Press, New York.

Wong D. W., Liu J., Yin F. *et al.* (2008) A semi-automated method for liver tumor segmentation based on 2D region growing with knowledge-based constraints. In: 3D Segmentation In The Clinic: A Grand Challenge II Workshop at MICCAI 2008

World Health Organization Media Center, Fact sheet N°297, retrieved July 7, 2012, from <http://www.who.int/mediacentre/factsheets/fs297/en/>

Xu C, Prince J. L. (1998) Snakes, shapes, and gradient vector flow, *IEEE Transactions on Image Processing*, Volume: 7, Issue: 3 pp: 359–369

Yim P. J., Vora A. V., Raghavan D., Prasad R., McAulliffe M., Ohman-Strickland P., Noshier J.L., Volumetric analysis of liver metastases in computed tomography with the fuzzy C-means algorithm. *Journal of Computer Assisted Tomography*, Volume: 30, Issue: 2, 2006, pp: 212-220

Zhou J., Xiong W., Tian Q. *et al.* (2008) Semi-automatic segmentation of 3D liver tumors from CT scans using voxel classification and propagational learning. In: 3D Segmentation In The Clinic: A Grand Challenge II Workshop at MICCAI 2008.

Zhou, J., Leow, W., Wong, D.W.K., Ding, F., Venkatesh, S.K., Tian, Q., Qi, Y., Xiong, W., Liu, J.J., Liver tumour segmentation using contrast-enhanced multi-detector CT data: performance benchmarking of three semiautomated methods, *European radiology*, Volume: 20, Issue: 7, 2010, pp: 1738-1748