

# CAD for Hepatic Tumor Detection in CT Images

Hetvi Pasad

Student

K J Somaiya College of Engineering  
Department of Computer Engineering

Himani Shetty

Student

K J Somaiya College of Engineering  
Department of Computer Engineering

Ayushi Malde

Student

K J Somaiya College of Engineering  
Department of Computer Engineering

Poonam Bhogale

Professor

K J Somaiya College of Engineering  
Department of Computer Engineering

## ABSTRACT

In the abdominal CT scan, the liver region is not clearly discerned from the adjacent organs such as muscle, spleen, and pancreas. The objective of the proposed system is to devise a novel method for tumor identification which helps the medical experts for further diagnosis.

The region of interest, namely the liver, is first separated by combining ROIpoly and thresholding methods. On obtaining the liver region, the tumor if present, is extracted using Gray Level Co-occurrence Matrix (GLCM) and Fuzzy C Means (FCM). Further, we have also compared the results obtained from both the methods.

## General Terms

Tumor Detection, CAD

## Keywords

Extraction ,ROI, Segmentation

## 1. INTRODUCTION

The primary hepatic cancer starts in the liver which is located below the right lung. It is one of the largest and a vital organ in the human body. Liver cancer is found to affect around 30 people per 100,000. However the symptoms do not appear until the cancer is well advanced. To prevent further delay in the reporting, the computer aided drafting (CAD) for tumor detection has been proposed for tumor identification. The system incorporates image processing and computer vision techniques which can be used by the oncologists to determine changes in the tumor size.

A computerized tomography (CT) scan combines a series of X-ray images taken from different angles and uses computer processing to create cross-sectional images, or slices, of the bones, blood vessels and soft tissues inside your body. CT scan images provide more detailed information than plain X-rays do. [1] However, the scope of interpreting results from the image is limited for a human being. A computer system facilitates this purpose. The proposed methods have been implemented on a dataset consisting of CT images of the abdomen using MATLAB as the platform. The first step to be carried out is the separation of the region of interest and then applying the segmentation algorithms. Thus the flow diagram can be illustrated as follows.

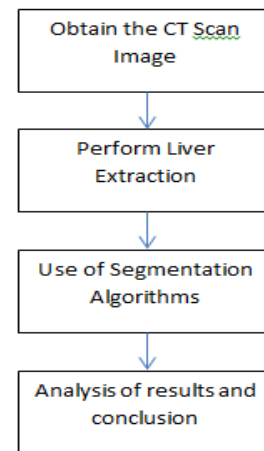


Figure 1

## 2. LIVER EXTRACTION AND SEGMENTATION

Automatic liver segmentation is essential in order to reduce the human intervention in the system and also to improve the accuracy of the results obtained from the testing. In addition, liver segmentation is a very challenging task due to countless anatomical variations and technical difficulties. Many methods have been designed to overcome these challenges.

### 2.1 Extraction

The input image is first converted to a gray scale image as a preprocessing step followed by resizing of the image. We have observed that the liver is positioned towards the left top side of the image. Thus the coordinates of this region are passed as parameters to the ROIpoly function. This reduces the presence of the unnecessary organs in the image. However, the output from this step is not accurate since the CT scans considers different angles and sizes. This causes variations in the shape and size of the liver. To obtain a better result, we have considered the intensity values of the image. They range from 0-255. Choosing an appropriate threshold, we perform thresholding. Further, we apply the code to find the largest connected area in the image. This is based on the observation that the liver is the largest organ in the CT scan of the abdomen. The results obtained have been shown in Figure 2 below.

The watershed algorithm has also been implemented but this method leads to over-segmentation.

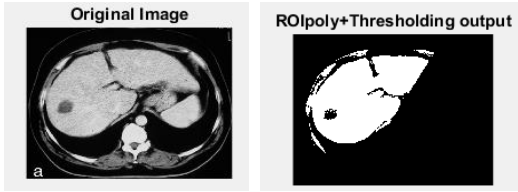


Figure 2.a

Figure 2.b

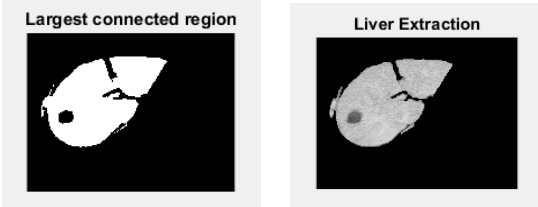


Figure 2.c

Figure 2.d

## 2.2 Segmentation using Fuzzy C Means

FCM is a classic clustering based method used for liver segmentation. The basic idea of clustering is that in an  $n$ -dimensional feature space, the distance between samples from the same class will be minimal and the similarity between samples from the same class would be higher or we can say that dissimilarity between samples from same class would be minimal. Clustering can be done in two ways, that is, hard clustering and soft clustering. In hard clustering, each sample point can belong to exactly one cluster and the membership value is assigned to each sample point that defines the degree to which that particular sample point belongs to a cluster in soft clustering. Soft clustering is also known as fuzzy clustering where each data point can belong to many clusters.

FCM is one of the popular methods used for fuzzy clustering when it comes to image processing and pattern recognition.[2]

FCM divides a set of  $n$  objects  $x$  to  $\{x_1, x_2, \dots, x_n\}$  in  $R^d$  dimensional space into fuzzy clusters, say  $c$  ( $1 < c < n$ ), with centroids or cluster centers  $y = \{y_1, y_2, \dots, y_c\}$ . The fuzzy clustering of objects is described by a fuzzy matrix containing the membership values, say  $\mu$ , with  $n$  rows and  $c$  columns, where  $n$  is the number of data objects and  $c$  is the number of clusters. The membership function or degree of association of  $i$ th data object with  $j$ th cluster is represented by  $\mu_{ij}$ . FCM algorithm consists of an objective function that aims to minimize following equation. [2]

$$J_m = \sum_{j=1}^c \sum_{i=1}^n \mu_{ij}^m d_{ij} \quad (1)$$

Where

$$d_{ij} = \|x_i - y_j\| \quad (2)$$

The scalar term  $m$  is used as weighting exponent which controls the fuzziness of the resulting clusters. Euclidian distance from object  $x_i$  to cluster centroid  $y_j$  is given by  $d_{ij}$ .

Center of  $j^{\text{th}}$  cluster,  $y_j$ , is obtained by: [2]

$$y_j = \frac{\sum_{i=1}^n \mu_{ij}^m x_i}{\sum_{i=1}^n \mu_{ij}^m} \quad (3)$$

FCM algorithm is iterative and can be stated as follow:

1. Choose the number of clusters  $c$  and the weighting exponent  $m$  ( $m > 1$ ).

2. Initialise all  $\mu_{ij}$ , membership function values randomly, where  $I = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, c$ .
3. At each step  $p$ , compute the centroid of each cluster,  $y_j$ , where  $j = 1, 2, \dots, c$ ; using equation(2).
4. Compute new membership values,  $\mu_{ij}$ , using

$$\mu_{ij} = \frac{1}{\sum_{p=1}^c \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (4)$$

5. Update the membership function matrix
6. Repeat the steps 3-5 until change of membership values is very small, typically 0.01. [2]

The final tumour obtained using the FCM algorithm is shown in Figure 3.b.

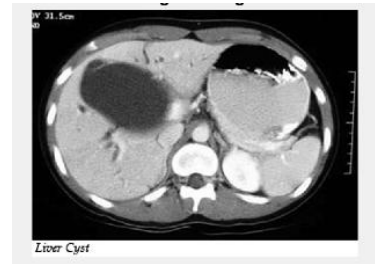


Figure 3.a. Original Image

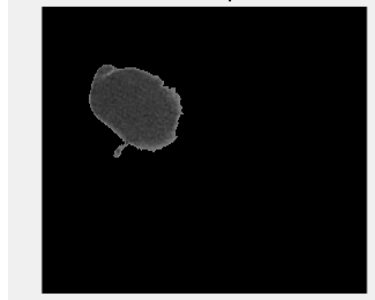


Figure 3.b FCM output

## 2.3 Gray Level Co-occurrence Matrix (GLCM)

The Gray Level Co-occurrence matrix gives a method to analyze and extract second order statistical features of a given image. The input image has intensities ranging from 0 to 255. (See Figure 2.a) The image intensity is scaled down with respect to the number of intensities specified in the function graycomatrix. Considering the number of intensities to be 8 levels, the scaled image (SI) is obtained. (See Figure 2.b) The values in SI are then used to create the GLCM.

	1	2	3	4	5
1	3	40	61	46	10
2	60	181	185	159	36
3	103	209	134	140	58
4	52	109	115	62	36
5	18	39	76	54	12

Figure 2.a Input Image Matrix of size 5x5

	1	2	3	4	5
1	1	2	3	2	
2	3	7	8	7	
3	4	8	6	6	
4	2	5	5	3	
5	1	2	3	2	

Figure 2.b Scaled Image Matrix of size 5x5

j →

	1	2	3	4	5	6	7	8
i ↓	1	0	2	0	0	0	0	0
	2	0	2	0	1	0	0	0
	3	0	3	0	0	0	0	1
	4	0	0	0	0	0	0	1
	5	0	0	1	0	1	0	0
	6	0	0	1	0	0	1	0
	7	0	1	0	0	0	0	1
	8	0	0	0	0	0	1	0

Figure 2.c GLCM Output

To illustrate, the following shows how gray co-matrix calculates the first three values of the second column in a GLCM. In the output GLCM, element (1,2), marked in red in the Figure 2.c, contains the value 2 because there are only two instances in the input image where two horizontally adjacent pixels have the values 1 and 2, respectively which have been highlighted in red in the Figure 2.b. (2,2) contains the value 0 since there are no such instances where two horizontally adjacent pixels have the values 2 and 2. Element (2,3) has the value 2, highlighted in orange, because there are three instances of two horizontally adjacent pixels with the values 2 and 3. Gray co-matrix continues processing the input image, scanning the image for other pixel pairs (i, j) and recording the sums in the corresponding elements of the GLCM. [3].

On applying GLCM to the image in Figure 3a, we obtain the following result.

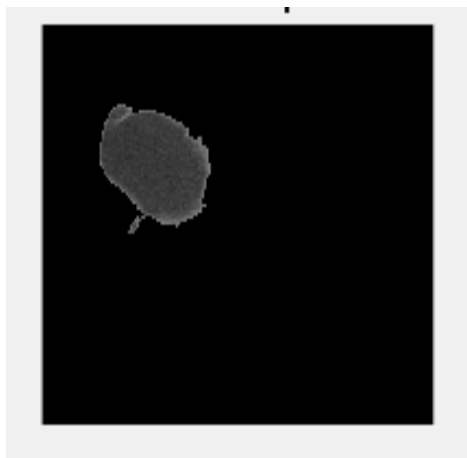


Figure 4 GLCM output

### 3. ANALYSIS OF RESULTS

The analysis was carried out on a data set of grayscale CT images of different patients. The statistical values obtained after applying all the said procedures and algorithms have been tabularized in the following tables.

Table 1

Image no	GLCM Area	GLCM Perimeter	FCM Area	FCM Perimeter	Actual Area	Actual Perimeter
1	27589	800.289	28246	771.3	28365	657.55
2	4539	396.596	7746	1723	3801	242.41
3	60	40.2590	96	49.25	16316	553.07
4	13283	177.25	14458	177.2	20794	733.53
5	4507	451.626	4991	403.8	7870	593.47
6	1422	161.707	1749	170.1	1845	156.27
7	842	247.444	60657	315.4	2287	178.53
8	13517	518.897	14319	521.1	13601	459.05
9	697	326.960	4864	994.8	11302	393.10
10	4539	396.596	7746	172.3	3780	249.51
11	174	99.2090	2677	111.1	4031	331.02
12	23036	687.156	23491	696.4	22797	581.68
13	5839	898.366	9517	1987.03	1233	132.10

Table 2

Image no	GLCM MSE	FCM MSE
1	95.6846	88.8901
2	32.4373	367.8350
3	1569.9458	1572.8152
4	346.5730	339.3623
5	192.7475	178.6499
6	21.5299	16.6061
7	33.7424	2238.8252
8	42.6496	49.6177
9	507.0104	333.5766
10	32.7955	367.1324
11	199.5123	297.5593
12	70.7496	80.7556
13	232.6618	378.7329

The actual tumor values for area and perimeter shown in Table 1 is obtained by using a manual tool. The tumor boundaries in the original image are traced using this tool. The area and perimeter of the final tumor extracted using the FCM and GLCM algorithm is also found, without manual techniques.

It is observed that the value of the area and perimeter obtained using FCM is closer to the actual values as compared to GLCM. In statistics, the mean squared error (MSE) or mean squared deviation (MSD) of an estimator measures the average of the squares of the errors or deviation, that is, the difference between the estimator and what is estimated. [4]. If  $\hat{Y}$  is an array of  $n$  predictions, and  $Y$  is an array of  $n$  observed values corresponding to the inputs to the function which generated the predictions, then the MSE can be estimated by following formula:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_1 - Y_2)^2 \quad (5)$$

i.e., the MSE is the mean  $\left(\frac{1}{n} \sum_{i=1}^n (Y_1 - Y_2)^2\right)$  of the square of the errors  $(Y_1 - Y_2)^2$ . [4] Lower the value of MSE for an algorithm, higher is its efficiency. The comparison of the mean square errors is shown in Table 2.

Further, there is, the Jaccard similarity index, which compares members for two sets to see which members are shared and which are distinct. It's a measure of similarity for the two sets of data, with a range from 0% to 100%. Higher the percentage, more is the similarity in the two populations. [5] It is thus defined as the size of the intersection divided by the size of the union of the sample sets [6]:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (6)$$

The percentage of similarity is found for the given set of images using the output of GLCM and FCM algorithms. The similarity between the tumor obtained using FCM and the actual tumor is found to be greater in most images.

The Dice index is known by several other names, especially the Sorensen index or Dice's coefficient (DC). The coefficient may be defined as twice the shared information (intersection) over the sum of cardinalities [7]

$$DC = \frac{2|A \cap B|}{|A|^2 + |B|^2} \quad (7)$$

A false positive gives the probability that the given input has no tumor but the system detects the presence. While false negative indicates the probability that the system output shows the absence of a tumor when in the actual case, a tumor is present. The ratio of diseased cells to total cells gives the false positive value. The ratio of healthy cells to total cells gives the false negative value.

In order to perform a convenient comparison of the above described parameters, the Jaccard Coefficient, Dice Coefficient, probability of a False positive and a False negative have been tabularized based on the output of the GLCM algorithm in Table 3.

Similarly, based on the output obtained from FCM algorithm, Table 4 shows the values of the above parameters.

**Table 3**

Image no.	Jaccard	Dice	False Positive	False Negative
1.	92.44	96.07	0.040	0.03
2	85.13	91.96	0.16	0.01
3	79.26	88.43	0.15	0.08
4	70.19	82.48	0.15	0.18
5	80.18	89.00	0.24	0.005
6	87.51	93.33	0.13	0.008
7	76.96	86.98	0.25	0.03
8	84.70	91.72	0.15	0.02

9	75.36	85.95	0.32	0.001
10	72.92	84.34	0.36	0.003

**Table 4**

Image no.	Jaccard	Dice	False Positive	False Negative
1	92.51	96.11	0.049	0.028
2	86.18	92.57	0.154	0.005
3	77.90	87.57	0.240	0.033
4	71.37	83.29	0.158	0.173
5	89.64	88.66	0.246	0.007
6	85.56	92.21	0.153	0.012
7	78.00	0.00	3.057	1.000
8	86.05	63.06	1.156	0.006
9	66.53	72.23	0.762	0.003
10	73.72	82.84	0.410	0.002

Some more standardized methods with respect to evaluation metrics to compare the results of algorithms include Volumetric Overlap Error (VOE), Relative Absolute Volume Difference (RAVD) and Maximum Symmetric Surface Distance (MSSD). The formulae for the same are given below [8]:

$$VOE(\%) = \left(1 - \frac{Seg \cap Ref}{Seg \cup Ref}\right) \times 100\% \quad (8)$$

$$RAVD(\%) = \left(\frac{|Seg - Ref|}{Ref}\right) \times 100\% \quad (9)$$

$$MSSD(mm) = \max\left\{\max_{a \in A}\{\min_{b \in B}\{dist(a, b)\}\}, \max_{b \in B}\{\min_{a \in A}\{dist(a, b)\}\}\right\} \quad (10)$$

Table 5 includes the respective values based on GLCM outputs while Table 6 consists of the values based on FCM outputs. The average percentage of VOE is found to be 19.93%. The average RAVD of the system is 15.38%.

**Table 5**

Image no.	VOE	RAVD	MSSD
1	7.55	0.20	3675.09
2	14.86	14.99	3833.54
3	20.73	7.84	224.19
4	29.80	3.04	96263.21
5	19.81	23.43	8626.49
6	12.48	12.47	1065.13
7	23.03	21.55	3358.07
8	15.292	13.61	4657.77
9	24.63	32.33	58198.13
10	27.07	36.29	14491.29

**Table 6**

Image no.	VOE	RAVD	MSSD
1	7.48	2.12	137.52
2	13.81	14.90	1986.22

3	22.09	20.77	333.25
4	28.62	1.53	48181.26
5	20.35	23.86	7080.30
6	14.43	14.13	1157.31
7	10.00	20.76	8860.00
8	23.94	11.97	42103.50
9	33.46	5.96	64275.93
10	29.27	10.74	19080.56

#### 4. CONCLUSION

The presence of multiple organs in the CT scan of abdomen makes it mandatory for the liver region to be extracted in order to detect a liver tumor. The proposed system makes use of ROI coupled with multi-thresholding for this purpose. It is found to give appropriate results for most images. For segmenting the tumor from liver, the two methods compared are Fuzzy C Means and Gray Level Co occurrence Matrix. Various parameters have been compared to perform the analysis. After the analysis, it has been observed that FCM gives better results than GLCM. The system can be further worked upon in order to classify the tumor that is obtained. The identification depends on whether the tumor is benign or malignant.

#### 5. REFERENCES

- [1] <http://www.mayoclinic.org/tests-procedures/ct-scan/basics/definition/prc-20014610>
- [2] Mahesh Yambal, Hitesh Gupta, "Image segmentation using Fuzzy C Means Clustering," *IJARCCCE*, vol.2, issue 7, pp. July 2013.
- [3] <http://in.mathworks.com/help/images/create-a-gray-level-co-occurrence-matrix.html>
- [4] [https://en.wikipedia.org/wiki/Mean\\_squared\\_error](https://en.wikipedia.org/wiki/Mean_squared_error)
- [5] <http://www.statisticshowto.com/jaccard-index/>
- [6] [https://en.wikipedia.org/wiki/Jaccard\\_index](https://en.wikipedia.org/wiki/Jaccard_index)
- [7] [https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%9393Dice\\_coefficient](https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%9393Dice_coefficient)
- [8] Lauren K Haaitzma, "Liver tumor segmentation in CT Images", published in Data archiving and Network Services(DANS).
- [9] Ria Benny, Dr. Tessamma Thomas, "Automatic Detection and Classification of CT-Scan Images", presented at 2015Fifth International Conference on Advance in Computing and Communications, IEEE.