

# Computer Aided Diagnosis of Liver Cancer using K-Means Cluster Analysis

Chethan K S<sup>1</sup>, Dr. Nataraj K R<sup>2</sup>, Roopa K Swamy<sup>3</sup> and Dr. Choodarathnakara A L<sup>4</sup>

<sup>1</sup>PES University, Bangalore-560085, Karnataka State, INDIA

<sup>2,3</sup>SJBIT, Bangalore-560060, Karnataka State, INDIA

<sup>4</sup>GEC, Kushalnagar-571234, Karnataka State, INDIA

(E-mail: [chethan.k04@gmail.com](mailto:chethan.k04@gmail.com), [nataraj.sjbit@gmail.com](mailto:nataraj.sjbit@gmail.com), [rupaks7@gmail.com](mailto:rupaks7@gmail.com))

**Abstract**—Medical image processing and the diagnostics that is aided by computer focuses mainly on organ or disease based application. Computer Aided Design is used for the purpose of identifying and locating tumors, also helps in finding the its intensity and taking a proper decision about the disease. Automatic segmentation of tumor from CT images poses difficulty due to size, shape, position and presence of other objects with the same intensity present in the image. Therefore, for a accurate segmentation of the liver, a prior segmentation of the liver is necessary. In this paper, liver and tumor segmentation from CT images that can be utilized for computer aided diagnosis is proposed. The method utilizes region growing with optimized threshold algorithm. Segmentation of liver is done by using region growing method that begins from a seed point and close around tumors and vessels. K-Mean cluster analysis demonstrates accuracy of 78.75% with correlation 0.8626 for ground truth value of 57946.

**Keywords**—Liver Tumor, CT-Image, K-Mean

## I. INTRODUCTION

Medical image processing is used to identify various malfunctions in a living being and the tumors are one among them (malignant tumors are known to be cancerous tumors). Rapid development in imaging modalities made it possible to capture CT datasets with a high resolution. These high resolution and good quality images are helpful for physicians to detect tumors of various sizes. Due to a large number of images in a medical dataset, it is difficult for a physician to go through all of them and to make a diagnostic decision. Moreover, the diagnostics is mainly based on physician's subjective evaluation and his/her experience in that field. However, Computer Aided Detection (CAD) systems can reduce the risk of physician by providing objective tumor detection.

M V Sudhamani and G T Raju [2] proposed a method of segmenting CT liver images that helps to analyze the occurrence of hepatic tumor and classify them from images. Authors have used region growing technique in order to decide either the neighbor of a pixel should be added to the region or not. It is an iterative process and the seed point is selected in suspected region interactively. They used watershed segmentation method for segmenting the contour and is

generated by using the region growing. Grey Level Co-occurrence Matrix (GLCM) is used to extract the texture features for segmented region. The extracted features are then used to classify the tumor as benign or malignant by using Support Vector Machine (SVM) approach. The results indicate a reduction in the liver segmentation errors and a better classification of all segmented tumors as benign or malignant.

Ali et al., proposed a method for cancer disease management using ultrasound (US) image data of various stages of liver cancer. The method uses an intelligent CDS framework and provides an automation of image enhancement, segmentation, classification of disease and progression in real time. This will enable the early diagnosis of cancer in an efficient manner. Specifically, the framework consists of several stages from acquisition until the detection of cancer. The method is tested on 200 patients and SVM produced an accuracy of 95.29%.

In segmentation, the selection of region of interest (ROI) from the suspicious region is done using region growing algorithm. Here, in order to a seed point in a suspected region, human intervention is necessary. Automatic detection of liver cancer from CT images using K-Mean cluster analysis is the primary objective of this experiment.

This paper is organized as follows: Section II provides the liver segmentation methods such as gray level based, structure based and texture based methods. In section III the proposed methodology was described. In section IV, the experimental results are presented to analyze accurate detection of liver cancer through K-mean cluster analysis. Section V presents some concluding remarks.

## II. LIVER SEGMENTATION

Liver segmentation can be broadly classified in to three classes. It includes gray level based method, structure based method and texture based method.

### A. Gray Based Method

**Merits:** They are fast and does not require prior knowledge.

**Demerits:** An image consists of several boundaries and only few of them are real liver boundaries. So, these methods tend to converge easily to stronger false boundaries and causes either over or under segmentation. The refinement of result requires human intervention.

### B. Structured Based Method

*Merits:* It uses the prior knowledge to overcome the problem of unclear boundary.

*Demerits:* It demands huge training data set to involve all possible conditions of the liver. A non-standard shape of the liver still poses a problem.

### C. Texture Based Method

*Merits:* In this, several features can be considered together and the results are much nearer to that of manual segmentation.

*Demerits:* Along with training data set, this method demands texture description. Although there have been many descriptors, selecting a suitable descriptor that is equivalent to human description is a problem.

## III. PROPOSED METHODOLOGY

The proposed method uses K-means Clustering which has three main steps. Pre-processing is the first step which aims at image smoothing and removing vessels. In the second step, tumor candidate regions are extracted by employing intensity information. In the third step, we use shape information to filter the candidate regions. Details of the method are described in the following subsection namely, Pre-processing, Extracting the tumor candidates by using intensity information and Filtering the candidates by using shape information.

### A. Pre-processing

Usually CT images are associated with some sort of noise and it needs to be reduced in preprocessing stage. An abnormal liver dataset includes three types of tissues; liver, tumours, and vessels (Fig 1). The intensity range of vessel tissues is above the intensity range of the other two tissues. We first remove vessels from liver volumes, as proposed in [5]. Authors assume a Gaussian mixture to model intensity distribution of the tissues and use EM algorithm to both find model parameters and classify image pixels. EM algorithm consists of two steps; E-step and M-step. In E-step, a probability measure (posterior probability) is assigned to each pixel which indicates the membership of the pixel in a class. In M-step, posterior probability is used to estimate parameters of the model. These two steps are iterated until the estimated parameters are converged. The class with the highest average intensity corresponds to vessels which have to be removed.

As segmentation result by EM is strongly dependent on the initialization of parameters, employ k-means algorithm to find initial values for the model parameters. Since k-means algorithm depends on initial parameters (cluster centers), use the lowest intensity value as the initial center of tumor, the intensity with the largest probability in the histogram as the initial center of liver and the maximum intensity as the initial center of vessel. To make the segmentation of liver tissue easier, at the pre-processing stage, the original image is changed to gray scale image and color image in to either RGB or HSV models.

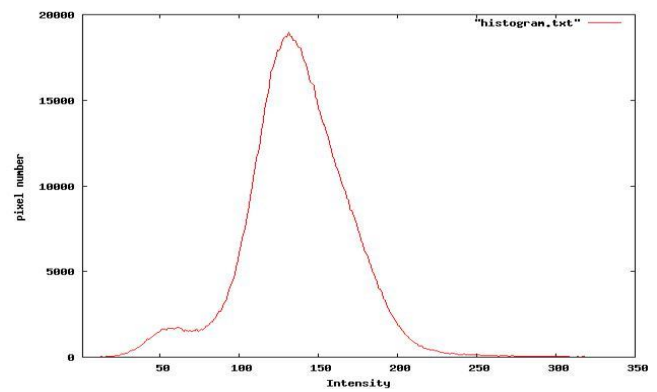


Fig 1. Typical intensity distribution of an abnormal liver in a CT image

### B. Extract the tumor candidate by using Intensity Information

To extract tumour candidate regions, EM/MPM algorithm can be utilized [8]. It uses Bayesian framework and considers Gaussian to estimate model parameters. In MPM, misclassified pixels are minimized through cost function. Thereby maximizes the posteriori probability of the label fields [9]. After this, use E-step and M-step iteratively to get the converged results.

### C. Filtering the candidate by using Shape Information

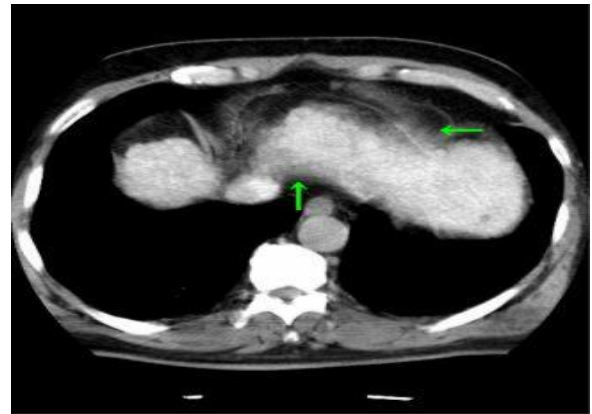


Fig. 2. Partial volume effect in CT images (Arrow indicate the effect)

Partial Volume Effect (PVE), in CT images, makes boundary of regions to be fuzzy. PVE is caused by body movements or movements of internal organs; e.g. due to breathing. It makes the intensity of a pixel to be a mixture of intensities of more than one tissue. This is especially true for the boundary regions of liver which makes them to be detected as tumor regions. Then employ a shape filter to reduce this effect. The shape filter has three components; circularity measure, proximity to liver boundaries and relationship between slices. First, check for the circularity of regions. Assume a spherical shape and circular cross-sections for tumors. If a candidate region fulfills this assumption, the probability that it belongs to a tumor goes up. Circularity of an area is defined in Equation (1). Where,  $C$  is the degree of circularity,  $S$  is the superficial measure, and  $R$  is the radius.

$$C = 4S/R^2 \quad (1)$$

In case of a tumor with a few numbers of pixels, Equation (1) does not lead to correct result. So, propose a new method to measure circularity of a cross-section. For the cross-section of tumor candidates in each slice, find the centroid and 12 boundary points; 4 points are in contact with the bounding box of the region (circle in Fig. 3) and 8 points which are located 45 degrees intervals on the candidate's boundary (diamond in Fig. 3). Using these 12 points, find their distances to the central point and employ Equation (1) to accept/reject a circular region. If the Equation is not fulfilled, then reject the candidate with value set 4.0 for  $Th$ .

The second component of the shape filter measures proximity of a tumor to liver boundaries. If the candidate region is very close to liver boundary, the probability of the partial volume is high. So we reject all regions which the distance of their centroid to liver boundary is less than 3 pixels.

The third component of the shape filter is to use the relationship between the slices. We use the idea of directed graph search [14]. The overlap region size of candidates between neighboring slices is used to identify the candidate is false positive or not. If both overlap regions of the current slice with the upper slice and the lower slice are smaller than the thresholds, the candidate is considered as false positive and is removed from candidate list. In this work, the thresholds are set as 0.1 and 0.8 for upper slice and lower slice, respectively.

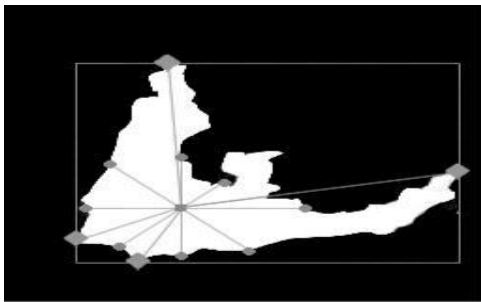


Fig. 3. Proposed method to measure circularity in a cross-section.

#### D. K-Means Clustering Algorithm

The algorithm for segmentation through K-Means clustering is explained below:

**Step 1:** The captured image is stored in the matrix format.

**Step 2:** Image size is identified and stored.

**Step 3:** Reshaping of the image for linear representation.

**Step 4:** K-means algorithm is applied to find clusters the image and determines the centroids.

**Step 5:** User defines the cluster size. The increase in the cluster size the resultant will be much clear.

**Step 6:** Much of the centroids are chosen as true value and remaining as false value.

**Step 7:** The resultant image provides a representation of dead cells of the liver and all the defected regions are also shown in the output image.

#### IV. EXPERIMENTAL RESULTS

This section discusses the experimental results of K-mean cluster analysis.

##### Step 1: Input Image

Given a Input Abdomen CT Image as shown in Fig. 4.



Fig. 4. Input Abdomen CT Image.

##### Step 2: Filtered Image

The Fig. 5 shows image after Median filtering. This filter simply sorts all values within a 3X3 Window, find the Median value. It is commonly used for removal of Salt and Pepper noise.



Fig. 5. Filtered Image.

*Step 3: Edge Image*

The Fig. 6 shows the Edge Image here the filtered image is passed through high pass filter to detect edges.

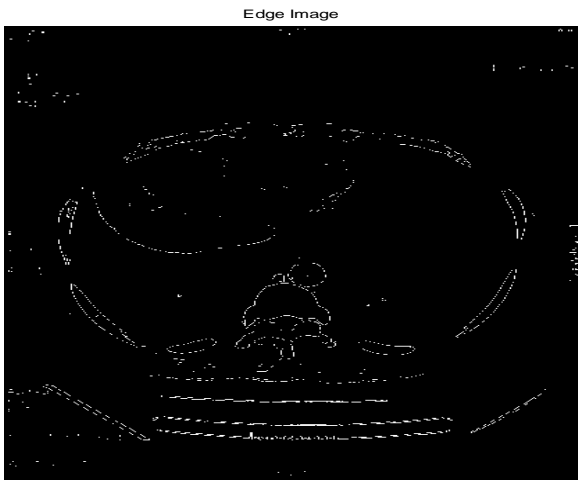


Fig. 6. Edge Image.

*Step 5: Binary Image*

Image is expressed in 1's and 0's. 1 is represents White and 0 for Black is shown in the Fig. 8.

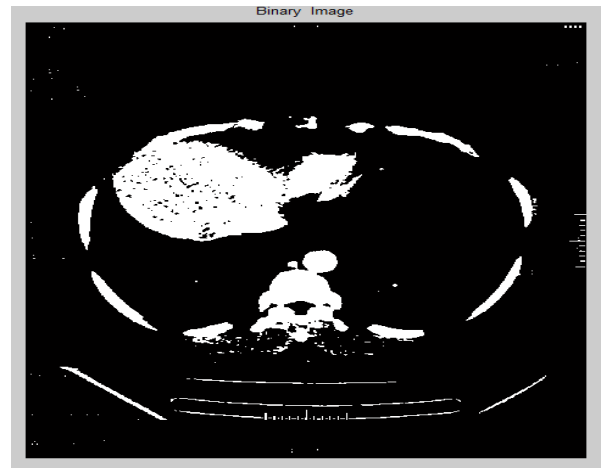


Fig. 8. Binary Image

*Step 4: Enhanced Image*

The Fig. 7 shows the Enhanced Image. The edge detected image is added to the original image in order to obtain the Enhanced Image.



Fig. 7. Enhanced Image.

*Step 6: Morphological Eroded Image*

Applying Morphological operation that is Erosion process it shrinks the thin objects in a binary image is shown in Fig. 9.

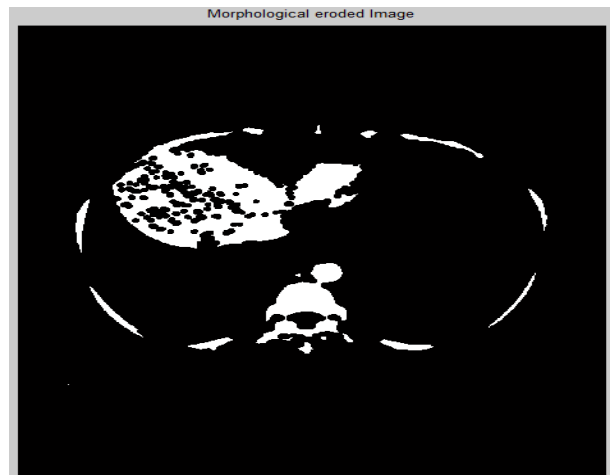


Fig. 9. Morphological Eroded Image.

*Step 7: Border Removal Image*

Here the border of the Image is removed as shown in Fig. 10.

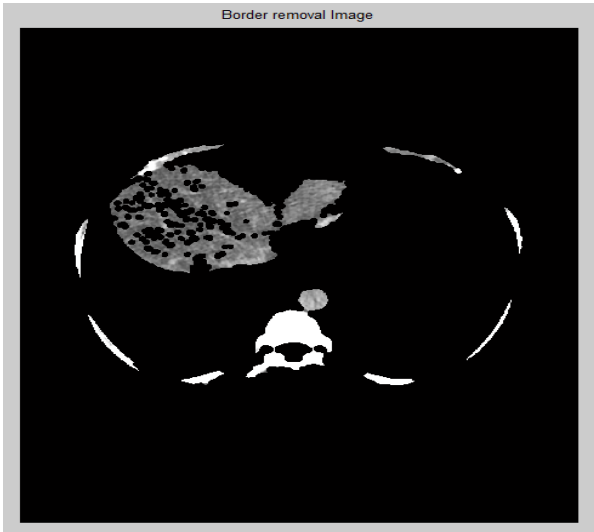


Fig. 10. Border Removal Image

*Step 9: Segmented Image*

Fig. 12 shows the segmented image. Here it separates the image foreground from its background.

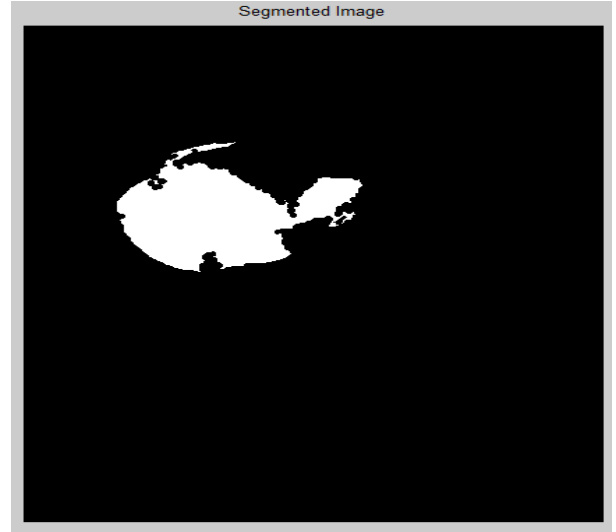


Fig. 12. Segmented Image.

*Step 8: Labelled Image*

The Fig. 11 Shows the Labelled image assigning colors for easily identify the liver cancer region. Blue, Red and Green are the assigned colors.

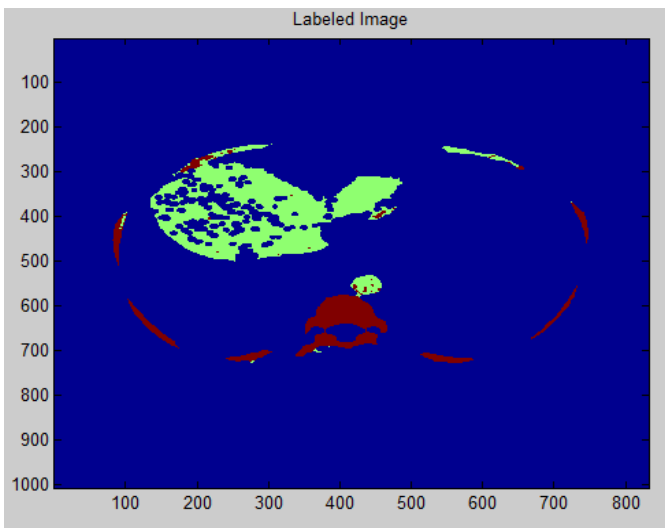


Fig. 11. Labelled Image.

The Fig. 13 Shows the Bar graph for number of cluster probability (1) represents Abdomen image (2) represents Liver image and (3) represent Cancer region in the graph.

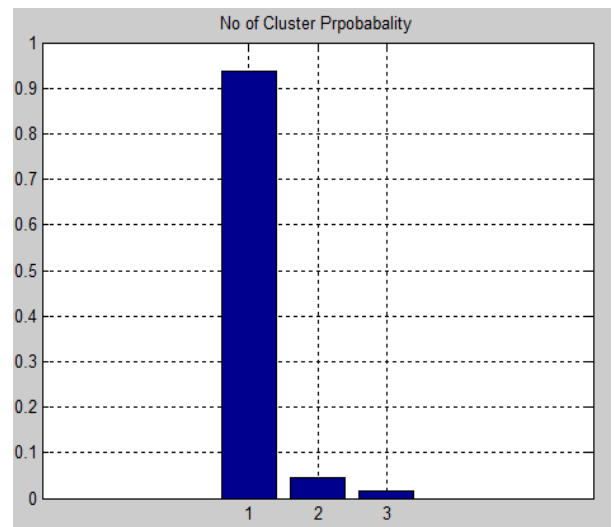


Fig. 13. Number of cluster probability.

TABLE I. AREA IN PIXELS, ACCURACY &amp; CORRELATION OF K-MEAN CLUSTER ANALYSIS FOR GROUND TRUTH = 57946

Technique Used	Area in Pixels	Accuracy	Correlation
K-Mean	45636	78.75%	0.8626

Table I show the area in pixels of tumour region in liver CT image using K-mean cluster analysis. The accuracy obtained for K-mean is 78.75% with correlation 0.8826 for ground truth value of 57946.

#### V. CONCLUSION

The proposed method to detect cancer candidate regions in CT images automatically using K-means algorithm is successful with moderate accuracy of 78.75%. The K-Means cluster analysis yields a correlation of 0.8626 for ground truth value of 57946. The K-means cluster analysis can able to detect cancer regions of both large and small sizes with relatively higher accuracy. As the challenges for the future, improvement of the detection rate and large morphology variation of cancer may be considered.

#### REFERENCES

- [1] Center for Cancer Control and Information Services, National Cancer Center, Japan, <http://ganjoho.jp/public/statistics/pub/statistics01.html>
- [2] M. V. Sudhamani, G. T. Raju, "Segmentation and Classification of Tumor in Computed Tomography Liver Images for Detection, Analysis and Preoperative Planning", International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-4, Number-1, Issue-14, March-2014.
- [3] L. Ali, A. Hussain, J. Li, U. Zakir, X. Yan, A. Shah, U. Sudhakar and B. Luo, "Intelligent Image Processing Techniques for Cancer Progression Detection, Recognition and Prediction in the Human Liver", 978-1-4799-5375-2/14 © IEEE, 2014.
- [4] Yu Masuda, Amir Hossein Foruzan, Tomoko Tateyama and Yen Wei Chen, "Automatic Liver Tumor Detection using EM/MPM Algorithm and Shape Information", 13<sup>th</sup> International Conference on Control, Automation, Robotics & Vision Marina Bay Sands, Singapore, 10-12th December 2014 (ICARCV 2014).
- [5] El Masry et al. "Automatic Liver CT Image Clustering based on Invasive Weed Optimization Algorithm", Engineering and Technology, 1-5, IEEE 2014.
- [6] Abdalla Zidan, "Level Set based CT Liver Image Segmentation with Watershed and Artificial Neural Networks", International Conference on Hybrid Intelligence Systems, 96-102, IEEE 2012.
- [7] Pedro Rodrigues, Jaime Fonseca and João L. Vilaça, "An Image Processing Application for Liver Tumor Segmentation", 1st Portuguese Meeting in Bioengineering, February 2011.
- [8] D. Smeets, D. Loeckx, B. Stijnen, B. De. Dobbelaer, D. Vandermeulen and P. Suetens, "Semi-automatic Level Set Segmentation of Liver Tumors Combining a Spiral Scanning Technique with Supervised Fuzzy Pixel Classification", Medical Image Analysis, Vol. 14, No. 1, pp. 13-20, February 2010.
- [9] Häme, Y. Alhonnoro and T. Pollari, "Image Analysis for Liver Tumor Ablation Treatment Planning, Hands-on Image Processing", Robotiker-Tecnalia, 2009.
- [10] K Mala, V. Sadasivam and S. Alagappan, "Neural Network Based Texture Analysis of Liver Tumor from Computed Tomography Images", International Journal of Biomedical Sciences 2, 33-40, 2006.
- [11] Seung-Jin Park, Kyung-Sik Seo and Jong-An Park, "Automatic Hepatic Tumor Segmentation using Statistical Optimal Threshold", Computational Science - ICCS 2005, Springer Berlin / Heidelberg, Volume 3514, pp. 934-940, 2005.
- [12] Osher and S. Fedkiw, "Level Set Methods and Dynamic Implicit Surfaces", No. 153 in Applied Mathematical Sciences. Springer-Verlag New York, Inc. ISBN: 0-387-95482-1. 2003.
- [13] Yu Masuda, Amir H. Foruzan, Tomoko Tateyama and Yen Wei Chen, "Automatic Liver Tumor Detection using EM Algorithm and 3DROI", Kamsao-section Joint Convention of Institutes of Electrical Engineering, G310, 2009.
- [14] Jianhua Yao, Stacy D. O'Connor, Ronald M. and Summers, "Automated Spinal Column Extraction and Partitioning", ISBI 2006: 390-393, 2006.