

# Colorectal Cancer MRI Image Segmentation Using Image Processing Techniques

Arjun Nelikanti

Department of Computer Science and Engineering  
VCE, Hyderabad, India  
anelikanti@gmail.com

Narasimha Prasad L V

Department of Computer Science and Engineering  
VCE, Hyderabad, India  
lvnprasad@yahoo.com

Naresh Goud M

Department of Computer Science and Engineering  
VCE, Hyderabad, India  
nareshgoudcse@gmail.com

**Abstract** — Colorectal cancer is the third most commonly diagnosed cancer and the second leading cause of cancer death in men and women. Magnetic resonance imaging (MRI) established itself as the primary method for detection and staging in patients with colorectal cancer. In this paper MRI images of Colorectal cancer are used to detect the area and mean values of tumor area and distance from tumor area to other parts for staging the cancer. This paper describes algorithms for preprocessing, clustering and post processing of MRI images. The paper shows the implementation of algorithm for preprocessing using image enhancement techniques, clustering using adaptive k-means algorithm and post processing using image processing techniques in MATLAB.

**Keywords-** Colorectal Cancer; MRI; Clustering; Adaptive K-means; MATLAB.

## I. INTRODUCTION

Colorectal cancer is a cancer from uncontrolled cell growth in the colon or rectum and is the third most commonly diagnosed cancer in the world. Colorectal cancer also know as colon cancer, rectal cancer, bowel caner or colorectal adenocarcinoma. Most colorectal cancer occurs due to lifestyle and increasing age both in men and women. Therefore early detection of cancer is important. High-resolution MRI scanning plays an important role in assessment of cancer [1]. For the primary rectal cancer staging axial T2-weighted images are useful.

In medicine, the digital image processing techniques are used to enhance the contrast or transform the intensity levels into color for easier interpretation of bio-medical images. For any digital image processing applications there are certain steps to be followed as, preprocessing, segmentation and feature extraction etc.. In preprocessing median filter [2] is used to remove noise in an image which is a nonlinear digital filtering technique and it preserves edges while removing noise. The contrast along these edges are increased by sharpening the image. A quick transition from black to white looks sharp where gradual transition from black to gray to white looks blurry. Therefore sharpening technique in image processing is used. To equalize the contrast in image contrast limited adaptive histogram equalization [3] is used.

Clustering techniques are also used on these bio-medical images so as to differentiate cancer tissues from the other textures. Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. The adaptive K-means clustering algorithm [4] is used in this paper. And in post processing the image processing [5] techniques are used for further extraction of features. This paper is implemented in MATLAB(R2013a).

## II. METHODOLOGY

The overview of proposed method is shown in the flowchart below Fig.(1). The method starts by reading an image, followed by preprocessing. And the image is segmented using clustering technique. Finally the clustered sub image is selected and processed further. The detailed version of methods as follows.

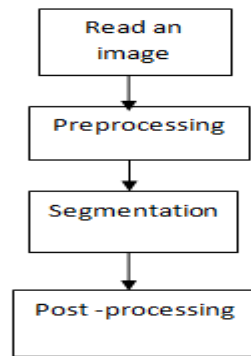


Figure 1. Flowchart for proposed method.

## 2.1 Preprocessing

The image is preprocessed for enhancement this is done by first converting it into gray image and applying two dimensional median filter [2]. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. Median filter replaces all the image pixels in the same time with the median of image pixel values in predefined (3-by-3) neighborhood of the given pixel. The Eq.(1) below represents a generalized function for any neighborhood.

$$f'(m, n) = \text{med}' \{ (-k \leq u, v \leq k) \{ f(m+u, n+v) \} \}. \quad (1)$$

where we used  $k=1$ , the median is calculated over 3-by-3 window, that is the size of filter. The noise in resulting image  $f'(m, n)$  is reduced.

Now the filtered image is sharpened by using unsharp masking. The unsharp masking technique comes from a publishing industry process in which an image is sharpened by subtracting a blurred version of the image from itself. The function `imsharpen` in matlab returns an enhanced version of the grayscale input image A, where the image features, such as edges, have been sharpened. The Eq.(2) below represents unsharp masking function.

$$g(m, n) = f(m, n) - f_{\text{smooth}}(m, n). \quad (2)$$

where  $g(m, n)$  is enhanced image,  $f_{\text{smooth}}(m, n)$  is smoothed image of  $f(m, n)$ .

The final step in image preprocessing is equalizing the contrast of the image. Using Contrast Limited Adaptive Histogram Equalization (CLAHE) [3] equalization of the contrast and clarity of the sharpened image is obtained. In CLAHE the part of the histogram that exceeds the clip limit is not discarded but it is redistributed equally among all histogram bins. The redistribution pushes some bins over the clip limit again (region shaded green in the Fig.(2)), resulting in an effective clip limit that is larger than the prescribed limit and the exact value of which depends on the image.

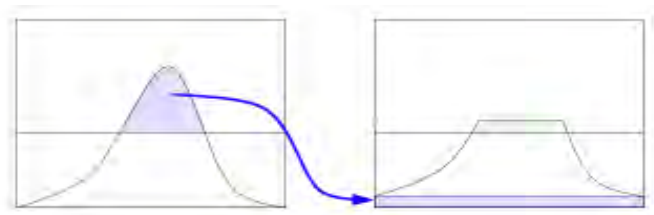


Figure 2. CLAHE distribution.

Now the input MRI image is converted into grayscale image as shown in Fig.(3) and preprocessed to enhance the image.

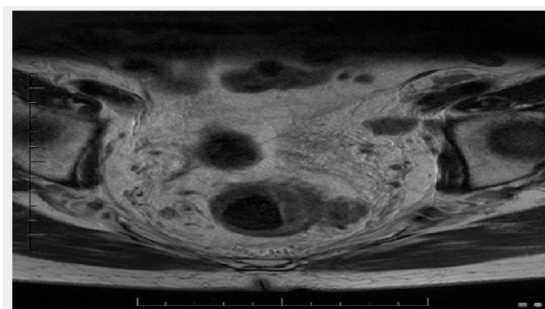


Figure 3. Original image.

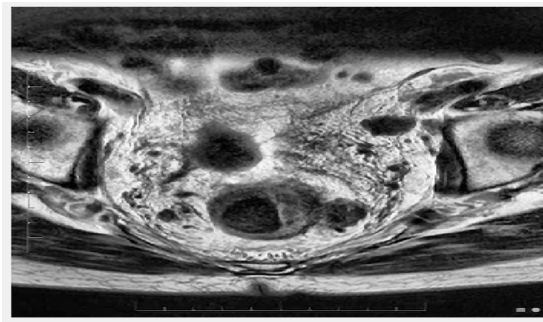


Figure 4. Enhanced image.

The above Fig. (4) is obtained after preprocessing.

## 2.2 Clustering algorithm

A popular technique for clustering is based on K-means. However, it suffers from the four main disadvantages. Firstly, it is slow and scales poorly on the time. Secondly, fixed number of clusters can make it difficult to predict what K should be. Thirdly, it may find worse local optima. Lastly, different initial partitions can result in different final clusters. To overcome with these problems an algorithm is used to perform K-means clustering that is adaptive in nature.

The image after enhancement is used to find clusters using Adaptive K-Means Clustering. The algorithm starts with random selection of initial seeds from the image, these seed properties also form the properties of cluster. The function uses Euclidean distance from centroid to quantify how close two objects are to each other. Based on the distance clustering is done. Merging of two clusters are done when the distance is less than the threshold. When all elements in the input are clustered, algorithm stops. Adaptive k-means algorithm returns an cluster centre corresponding to a cluster and label every pixel in the image with its cluster index as shown in Fig. (5).

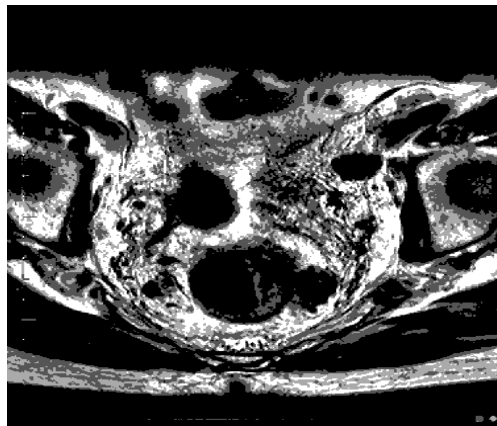


Figure 5. Labelled image.

Using cluster index we can separate the objects as show in Fig.(6) below.



Figure 6. Objects in clusters.

This algorithm is a fully automatic way to cluster an input image using k-means principle, but here we do not need to specify number of clusters or any initial seed value to start iteration, this algorithm automatically finds number of cluster and cluster center iteratively. It is a very fast implementation of clustering an image without knowing number of clusters.

### 2.3 Post Processing

In this section the clustered image that is close to the required data is selected and processed further. The objects are selected from the clustered image by adding coordinates using mouse pointer. The binary image containing objects are returned by the function bwselect and boundaries are drawn for the selected objects as shown in Fig.(7) below. And the properties such as area and mean values are calculated for the objects selected using region properties.



Figure 7. Selected object.

Area is calculated by the following Eq.(3)

$$\text{Area} = ((pr)^{1/2}) * 0.264. \tag{3}$$

where pr is number of pixels in region.

Using distance line function we draw a line by default on the figure and used to calculate the distance between cancer boundary and the other parts in the image. In T2 weighted axial rectal MRI the connective tissue and fat around the rectum is referred as mesorectum and visceral fascia surrounding it is the mesorectal fascia.

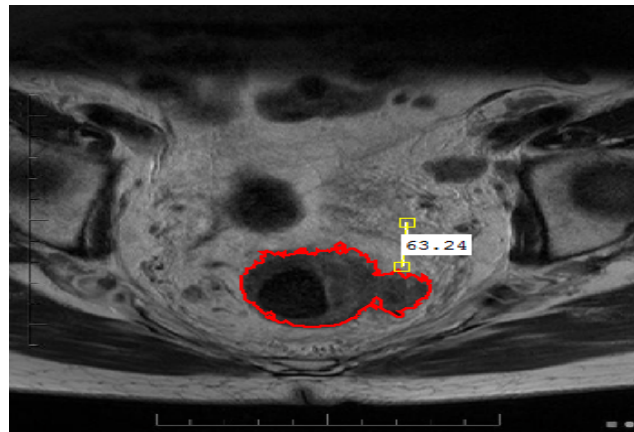


Figure 8. Showing distance line.

The distance line in the above Fig. (8) measures the distance between the two endpoints of the line. The distance in a text label superimposed over the line specifies the distance in pixels. The distance in unit pixels is converted into centimeters by the following Eq.(4).

$$\text{dcm} = \text{dpix} * 0.02645. \tag{4}$$

where dpix is distance in pixel and dcm is distance in centimeters.

### III. RESULTS AND DISCUSSIONS

For the analysis of the steps proposed for preprocessing histogram of the original and the image after preprocessing is shown in the Fig.(9). It clearly shows how CLAHE redistributed the contrast equally among all the bins in histogram.

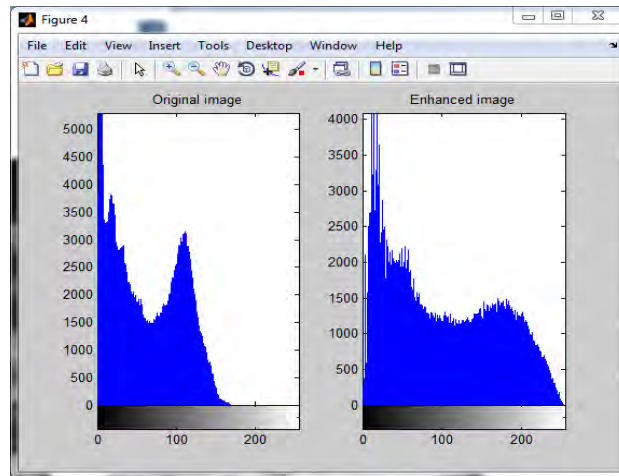


Figure 9. Histogram of original and enhanced image.

And the Table.(1) shows the values obtained for

1) Absolute mean brightness error (AMBE) which is the difference between original and enhanced image and is given as Eq.(5)

$$AMBE = | E(x) - E(y) | \tag{5}$$

where E(x) is average intensity of original image and E(y) is average intensity of enhanced image.

2) Peak-signal-to-noise-ratio (PSNR) which is evaluation standard of reconstructed image quality. The units of PSNR is decibels (dB) and is given by Eq.(6)

$$PSNR = 10 \log \left( \frac{255^2}{MSE} \right) \tag{6}$$

where MSE is mean square error.

TABLE I. ENHANCEMENT ANALYSIS

Image #	AMBE	PSNR
Image1	37.4788	14.1999
Image2	19.7099	18.4530
Image3	39.1854	14.0689
Image4	46.0584	13.0673
Image5	37.6231	14.2407

The mean values of both original and enhanced images are shown in the Fig.(10)

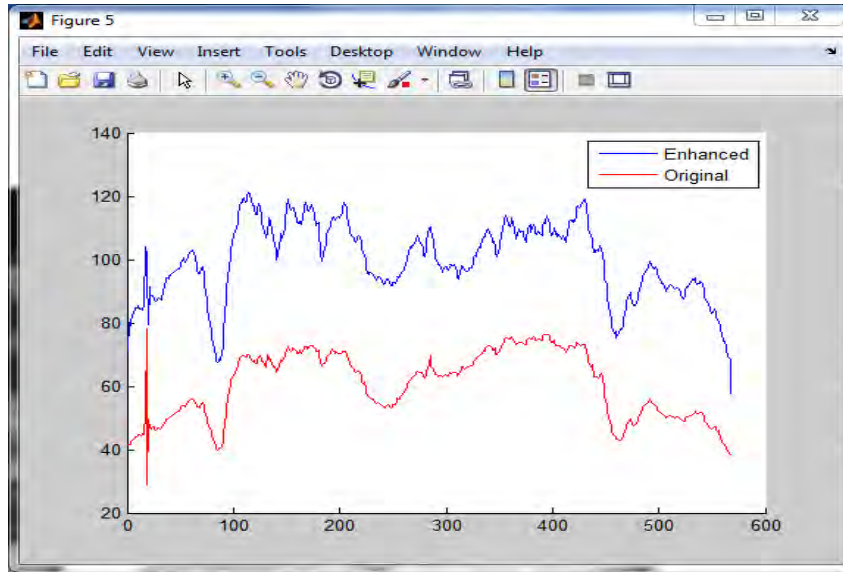


Figure 10. Mean values plots of images.

After clustering the sub image is selected for post processing. We proposed a framework where the area and mean of the tumor are calculated. These values are useful for future analysis. The Table.(2) below shows the values for the input images.

TABLE II. TUMOR DETAILS

Image #	Area	Mean value
Image1	33.185	55.591
Image2	33.076	80.171
Image3	29.649	40.964
Image4	16.205	25.314
Image5	31.188	51.801

The distance between the cancer growth and mesorectal fascia is traced for staging the cancer. TNM is the model followed for staging the rectal cancer. T is tumor, N is node and M is distance metastasis. In this paper T-staging is done and it is only for adenocarcinoma and not for other tumors. T-staging [6] is based on the distance between tumor and other parts of body. The distance to the mesorectal fascia is the single most important local prognostic factor therefore a distance line tool is used to find the distance between the tumor and the mesorectal fascia as it surrounding the rectum. The distance tool is draggable, resizable line.

Finding T3 stage is a crucial one as cancer may spread to other organs of the body. When the tumor enters the mesorectal fat the shortest distance between the tumor and the mesorectal fascia is calculated in the proposed framework. When the tumor reaches the mesorectal fascia the chances of tumor spreading to other organs are more.

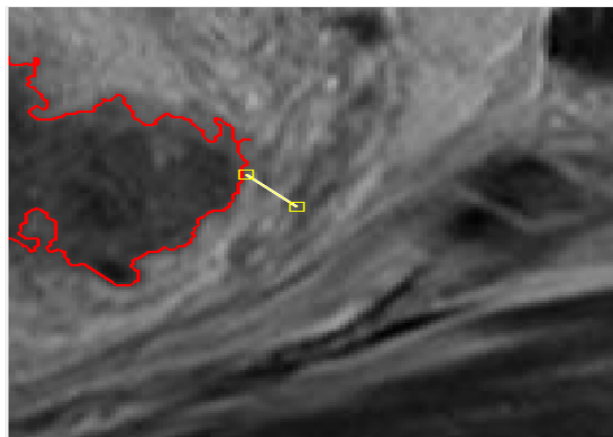


Figure 11. Distance calculation.

The Fig.(11) is zoomed to show the distance line and the distance is calculated using Eq.(4).

Distance by the line (yellow) showed above is 0.48 cm.

#### IV. CONCLUSIONS

In this paper we proposed a framework to enhance the input MRI image by image processing techniques and segmentation is done by using adaptive k-means clustering algorithm for the analysis of the colorectal cancer. We have calculated the area, mean of the tumor and the minimum distance from tumor to other parts. This helps the radiologist in staging the cancer.

#### REFERENCES

- [1] Usama Ghieda, Omar Hassanen, and Mohamed A. EltomeyA, "MRI of rectal carcinoma: Preoperative staging and planning of sphincter-sparing surgery", *The Egyptian Journal of Radiology and Nuclear Medicine* (2014) 45, 1–5.
- [2] Lim, Jae S., "Two-Dimensional Signal and Image Processing", Englewood Cliffs, NJ, Prentice Hall, 1990, pp. 469-476.
- [3] Zuiderveld, Karel. "Contrast Limited Adaptive Histogram Equalization." *Graphic Gems IV*. San Diego: Academic Press Professional, 1994. 474–485.
- [4] Sanjiv K. Bhatia, "Adaptive K-Means Clustering" In *Proceedings of Florida Artificial Intelligence Research Symposium*, 2004.
- [5] Gonzalez and Woods, 2008. "Digital Image processing" 3 Edition, Prentice Hall.
- [6] Michael R. Torkzad, Lars Pählman and Bengt Glimelius," Magnetic resonance imaging (MRI) in rectal cancer:a comprehensive review", *European Society of Radiology, Insights Imaging* (2010) 1:245–267.