

Colorectal Cancer Detection Using Image Processing Techniques: A Knowledge Transfer Perspective

Sindhu.C¹, Subhashini.S², Swathi.T³ and Gnana Saravanan S⁴

^{1,2,3}UG Scholar, Department of Electronics and Communication Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore.

⁴Assistant Professor (Senior), Department of Electronics and Communication Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore.

Article Received: 24 December 2017

Article Accepted: 27 February 2018

Article Published: 02 April 2018

ABSTRACT

Colorectal cancer (also named as colon cancer/rectal cancer/bowel cancer) is the growth of cancer in the rectum or colon (parts of large intestine). 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 known as colon cancer, rectal cancer, bowel cancer 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. 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. The overview of proposed method is developed to segment and detect colon cancer from CT images. 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.

Keywords: Digital Image Processing, Enhancement and Matlab.

1. INTRODUCTION

Digital image processing is one of the fastest growing technologies around the world. It has impacted almost all the areas and has emerged as a subject of interdisciplinary study. Digital Image Processing is useful in two prime application areas first being the improvement of pictorial information for human interpretation and the other is processing of image data for storage, transmission, and representation for autonomous machine perception.

In the medical field, 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 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 is 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 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 weighted K-means clustering algorithm is used in this project. And in post processing the image processing techniques are used for further extraction of features. This project will be implemented in MATLAB (R2014b).

Automatic detection of colon cancer has two major directions: segmentation and classification. In segmentation, heterogeneous colon samples are segregated into homogenous regions based on spatial distribution of tissues in the images. Next, normal and malignant labels are assigned to the regions based on certain features. In the literature, several approaches exist for medical image segmentation such as pixel based, region based, and graph based.

2. LITERATURE REVIEW

1. Sengar, Namita, Neeraj Mishra, Malay Kishore Dutta, Jiri Prinosil, and Radim Burget. "Grading of colorectal cancer using histology images." In *Telecommunications and Signal Processing (TSP), 2016 39th International Conference on*, pp. 529-532. IEEE, 2016.

This paper proposed [1] an automated system for grading of colorectal cancer using image processing method. Almost, half a million people die every year due to colon cancer. Histopathological tissue analysis is a common method for its detection, which needs an expert pathologist. Screening for this cancer is effective for prevention as well as early detection. The method proposed segment the glands automatically by using intensity based thresholding and organizational properties for classification. In existing literature, the majority of studies based on gland segmentation in healthy or benign samples, but rarely on intermediate or high grade cancer. Unlike most of the existing methods this system is fully automated and grades the images as benign healthy, benign adenomatous, moderately differentiated malignant and poorly differentiated malignant. The proposed method achieves overall accuracy of 81% when tested on 165 histology images.

2. Hantha Lakshmi, G., and B. Mohamed Faize. "Colorectal cancer detection using image processing based on IGVF model." *IOSR J Comput Eng* 16, no. 2 (2014): 39-43.

In this paper [2], an improved gradient vector flow (IGVF) is considered as an essential method to segment an image appropriately. This new algorithm, IGVF can improve GVF snake model's ability to capture thin boundary indentation like the boundary of cancer images. The segmented candidates are typically characterized by features describing like the polyp shape and its internal intensity distribution. Such features will serve as input for the classification system. Classification is performed by Support Vector Machine (SVM). Finally, measure the overlap between the manual and the algorithm segmentation to test the accuracy of frame work. It is expected to obtain greater accuracy and to produce good segmentation results of colonic polyps.

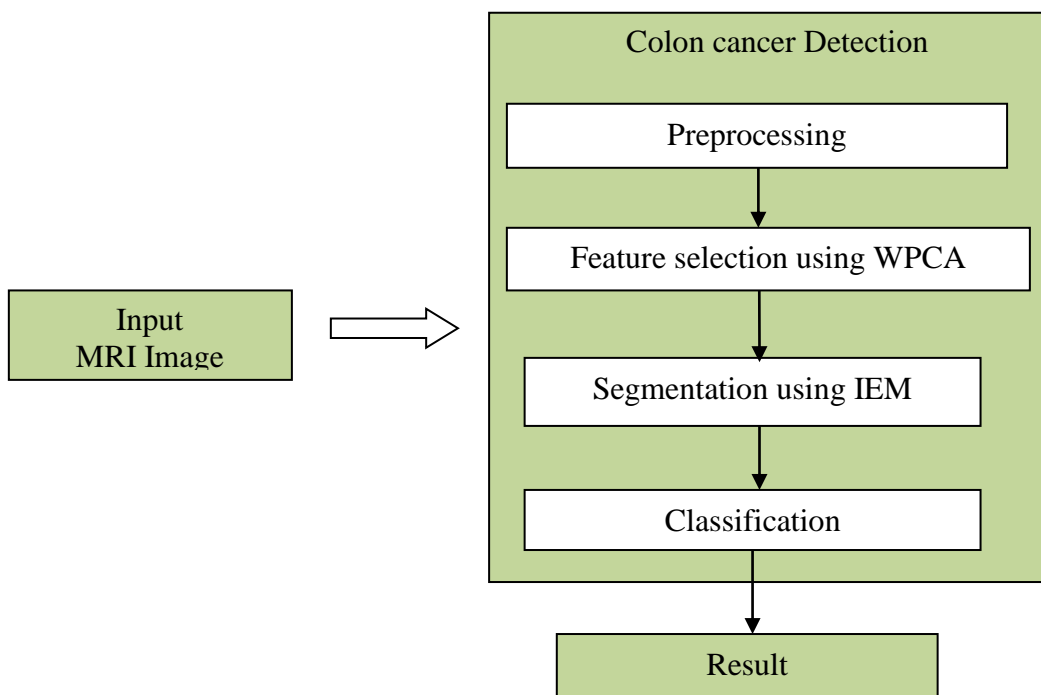
3. Ahmad, Mohd Yamin, Azlinah Mohamed, Yasmin Anum Mohd Yusof, and Siti Aishah Md Ali. "Colorectal cancer image classification using image pre-processing and multilayer Perceptron." In *Computer & Information Science (ICCIS), 2012 International Conference on*, vol. 1, pp. 275-280. IEEE, 2012.

In this paper [3], authors have proposed a method of automatic image pre-processing to extract important feature of colorectal tissue images. Images captured under microscope may vary in color brightness due to different staining concentration and the size of biopsy tissue. In this paper they proposed a method using HSV color to remove element outside the area of nucleus. In order to extract the gland shape, a gland tracking boundary and segmentation is proposed. By using the result of gland tracking, nucleus size that forms the glands are measured. Multilayer

Perceptron is being used to detect the shape of glands. By combining result of gland shape and nucleus size, they perform the image classification. The result shows that classification achieves 94% accuracy by using the proposed methods.

4. Zhang, Ruikai, Yali Zheng, Tony Wing Chung Mak, Ruoxi Yu, Sunny H. Wong, James YW Lau, and Carmen CY Poon. "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain." *IEEE journal of biomedical and health informatics* 21, no. 1 (2017): 41-47.

This paper [4] aims at developing a fully automatic algorithm to detect and classify hyperplastic and adenomatous colorectal polyps. Adenomatous polyps should be removed, whereas distal diminutive hyperplastic polyps are considered clinically insignificant and may be left in situ. A novel transfer learning application is proposed utilizing features learned from big nonmedical datasets with 1.4–2.5 million images using deep convolutional neural network. The endoscopic images authors collected for experiment were taken under random lighting conditions, zooming and optical magnification, including 1104 endoscopic nonpolyp images taken under both white-light and narrowband imaging (NBI) endoscopy and 826 NBI endoscopic polyp images, of which 263 images were hyperplasia and 563 were adenoma as confirmed by histology. The proposed method identified polyp images from nonpolyp images in the beginning followed by predicting the polyp histology. When compared with visual inspection by endoscopists, the results of this study show that the proposed method has similar precision (87.3% versus 86.4%) but a higher recall rate (87.6% versus 77.0%) and a higher accuracy (85.9% versus 74.3%). In conclusion, automatic algorithms can assist endoscopists in identifying polyps that are adenomatous but have been incorrectly judged as hyperplasia.



3. METHODOLOGY

The CT images cannot be segmented directly without any preprocessing steps. The overview of proposed method is developed to segment and detect colon cancer from CT images. 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.

This is indispensable to perform pre-processing on the input image, so that the image gets transformed to be relevant for further processing. Segmentation of image from MRI is a very important task. An image enhancement method is concerned about improving the visual appearance of images from magnetic resonance image. The proposed method consist of three main steps, image selection and pre-processing, feature selection, segmentation and classification.

Proposed Methodology

3.1 IMAGE SELECTION

In this study, abdomen CT images acquired from hospital to segment and classify the colon cancer. The images have collected from internet also. Grayscale or intensity images are displayed of default size of 512 x 512. A grayscale image can be specified by giving a large matrix. The abdomen MR images are stored in the folder in JPEG format. Figure 3.1 shows the image collected for the experiment.

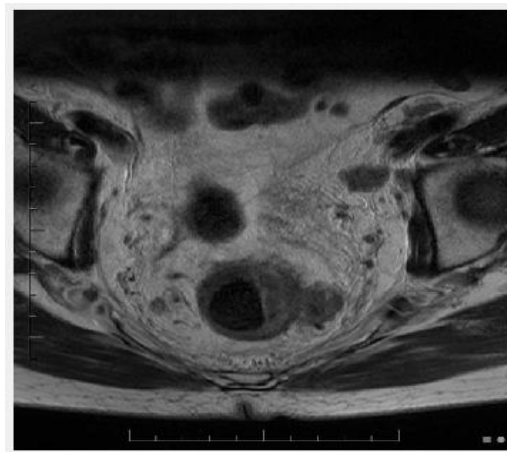


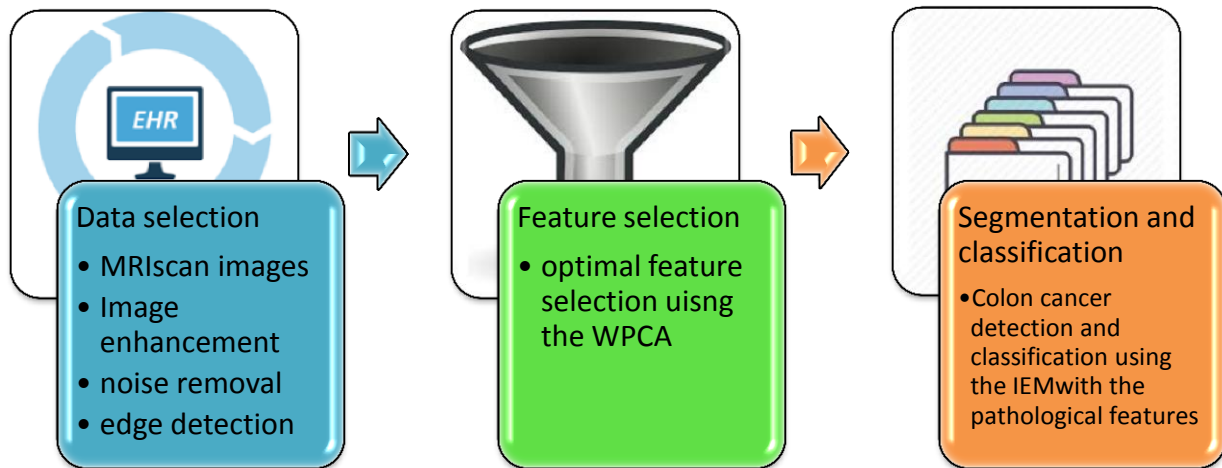
Fig 3.1 colon cancer sample training image

3.2 PREPROCESSING

The image is preprocessed for enhancement this is done by first converting it into gray image and applying two dimensional median filter. 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.(3.1) 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. Preliminary diagnosing of MRI images from the hospital cannot be relied on because of the chances of occurrence of artifacts resulting in degraded quality of image, while others may be confused with segmentation. Obtained MRI image usually contains limited quality. It becomes complex one for doctors in analyzing them. By increasing the contrast of an image, it is easy to analyze. In order to find the cancer part efficiently MRI image should be enhanced properly.



Proposed system architecture

3.3 FEATURE SELECTION

The proposed system effectively finds the features to segment and classify the cancer images separately. The feature selection process utilizes the weighted PCA algorithm for effective feature detection.

3.4 SEGMENTATION 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 Improved Expectation maximization Clustering. The algorithm starts with high priority feature selection from the results of WPCA, these seed properties also form the properties of cluster. The function uses MLE (maximum likelihood estimation) 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. IEM returns a cluster center corresponding to a cluster and label every pixel in the image with its cluster index.

4. SOFTWARE ENVIRONMENT DESCRIPTION

4.1 Matlab

MATLAB stands for Matrix Laboratory. According to The Math works, its producer, it is a "technical computing environment". We will take the more mundane view that it is a programming language. This section covers much of the language, but by no means all. We aspire to at the least to promote a reasonable proficiency in reading procedures that we will write in the language but choose to address this material to those who wish to use our procedures and write their own programs.

MATLAB variable names are normally case-sensitive. Thus variable C is different from variable c. A variable name can have up to 19 characters, including letters, numbers and underscores. While it is tempting to use names such as Fund Returns it is safer to choose instead fund returns or to use the convention from the C language of capitalizing only second and subsequent words, as in fund Returns. In any event, adopt a simple set of naming conventions so that you won't write one version of a name in one place and another later. If you do so, you may get lucky (e.g. the system will complain that you have asked for the value of an undefined variable) or you may not (e.g. you will assign the new value to a newly-created variable instead of the old one desired). In programming languages there are always tradeoffs. You don't have to declare variables in advance in MATLAB. This avoids a great deal of effort, but it allows nasty, difficult-to-detect errors to creep into your programs.

Matlab is a program that was originally designed to simplify the implementation of numerical linear algebra routines. It has since grown into something much bigger, and it is used to implement numerical algorithms for a wide range of applications. The basic language used is very similar to standard linear algebra notation, but there are a few extensions that will likely cause you some problems at first.

The goal of the tutorials here is to provide a simple overview and introduction to matlab. The tutorials are broken up into some of the basic topics. The first includes a few examples of how Matlab makes it easy to create and manipulate vectors. The tutorials move from the simple examples and lead to more complicated examples.

Description: `guide` initiates the GUI design environment (GUIDE) tools that allow you to create or edit GUIs interactively. `guide` opens the GUIDE Quick Start dialog where you can choose to open a previously created GUI or create a new one using one of the provided templates. `guide('filename.fig')` opens the FIG-file named `filename.fig` for editing if it is on the MATLAB path. `guide('fullpath')` opens the FIG-file at `fullpath` even if it is not on the MATLAB path. `guide(HandleList)` opens the content of each of the figures in `HandleList` in a separate copy of the GUIDE design environment.

Matlab

A basic introduction on how to define and manipulate vectors in matlab. This is the most basic way that numbers are stored and accessed in matlab.

Matrices: An introduction on how to define and manipulate matrices. We demonstrate how to create matrices and how to access parts of a matrix.

Vector operations: Here we bring together elements of the first two tutorials. The real power of matlab is that the basic operations defined in linear algebra can be carried out with similar notation and a minimal number of programming steps.

Loops: We introduce the basic loop construct used in matlab. We show how to define a for loop and provide an example of a how it can be used to solve a problem.

Plots: A general overview of the basic plotting commands is given. This is a very basic overview given to demonstrate some of the ways data can be plotted.

Executable Files: An introduction is given on how to define files that contain command that matlab can execute as if they had been typed in at the command prompt.

Subroutines: An introduction to subroutines is given. This is a more general way to provide an executable file in which generic arguments are passed back and forth through the command line.

If statements: The basic control structure in matlab is the "if" statement which allows for conditional execution of certain parts of a code. This is useful when you have to check conditions before deciding what actions should be taken.

Data Files: Matlab allows a number of ways to access data files for use in a session. The different ways to save all of the data, a particular matrix, and C style read write statements is examined. Also, the diary command is examined to demonstrate how a text copy of a session can be saved.

5. RESULT

5.1 SAMPLE OUTPUTS

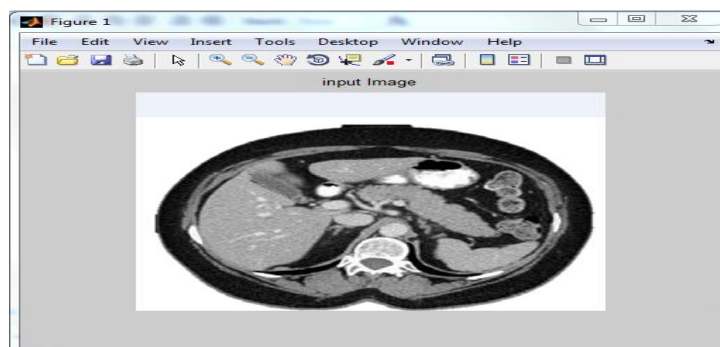


Figure A1.1 Input Image

This figure shows the input image of the CT colon with cancer image (having any type of edges and lines). Using the proposed algorithm to get this output with the help of super pixels. Dense Scale Invariant Feature Transform the algorithm was proposed to solve this issue. GBM can offer a bigger edge. GBM is a boosting method, which builds on weak classifiers. The idea is to add a classifier at a time, so that the next classifier is trained to improve the already trained ensemble.



Figure A1.7 Dense-Sift Output

6. CONCLUSION

From the analyses of the problem in medical data there were problem like image segmentation and classification. To solve the issue the proposed system segments the colon with the considerations of effective several feature points. The proposed system uses WPCA based segmentation and classification process. The proposed system is an automated bottom-up approach for colon segmentation with the consideration of optimal features in abdominal computed tomography (CT) scans. The method generates dynamic cascaded and macro super-pixel segmentation information's by classifying image patches at different resolutions. The proposed system effectively detects the colon from the abdomen CT images with effective feature selection. The best feature selection is performed using WPCA concepts. The classification process is performed after the colon segmentation.

7. FUTURE WORK

However, the frameworks created till today do not apply in case of CT colon with least resolution and quality. Hence, this research work concentrates not only on segmentation of Medical images but also on unifying it under one roof. Thus it gives a combined or a unified Framework of segmentation which can take as input different types of colon and medical images. This system will utilizes histogram and texture features. The method generates dynamic cascaded and macro super-pixel segmentation information's by classifying image patches at different resolutions. In future for Fast organ detection WASIFT (Weighted Adaptive Scalable Invariant Feature Transform) algorithm can be used. This algorithm helps to detect the weighted descriptors from the image for fast image segmentation. Using this system, the organ from the CT images can be segmented quickly. With the consideration of additional features of other organs and classifying those images into different classes with high precision will be

performed in the future work. The future work will use semi-supervised classification techniques to reduce the time of training process. Using the above, the accuracy can be improved and time can be reduced.

8. ACKNOWLEDGMENT

The authors wish to thank the respondents who have spent their precious time and patience for participating in this project.

REFERENCES

- [1] AHARON, M., ELAD, M. and BRUCKSTEIN, A.M. (2006) K-SVD: an algorithm for designing over complete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54, pp. 4311–4322.
- [2] Amal Farag, (2015)“A Bottom-up Approach for Colon Segmentation using Cascaded Superpixels and (Deep) Image Patch Labeling”.
- [3] Amal Farag, Le Lu et al, (2014)“A Bottom-Up Approach for Automatic Colon Segmentation in Abdominal CT Scans.
- [4] Anders Lindbjerg Dahl, (2011)“Learning Dictionaries of Discriminative Image Patches”.
- [5] ASHBURNER, J. and FRISTON, K. J. (1997) Multimodal image co-registration and partitioning - a unified framework. *NeuroImage*, 6(3), pp. 209-217.
- [6] BUADES, A., COLL, B. and MOREL, J.M. (2005) A non-local algorithm for image denoising. *IEEE International Conference on Computer Vision and Pattern Recognition*, 2, pp. 60–65.
- [7] ELSHERIF, M. and ELSAYAD, A. (2001) Wavelet packet denoising for medical image enhancement. *IEEE Transaction Inf. Theory*, 10, pp.180-184.
- [8] FRISTON, K.J. (2000) Voxel-based morphometry the methods. *NeuroImage*, 11, pp. 805–821.
- [9] GERIG, G., KUBLER, O., et al., (1992) Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11, pp. 221–232.
- [10] GULERYUZ, O.G. (2003) Weighted over complete denoising. *Proceedings of Asilomar Conference on Signals and Systems*, 2, pp. 1992-1996.