Colorography: an Adaptive Approach to classify and detect the Breast Cancer using Image Processing

Ganesh Choudhari,   Debabrata Swain, Dipali Thakur, Kiran Somase

Student   Asst. Prof.   Student   Asst. Prof.

Department of Computer Engineering ,Sharadchandra Pawar College of Engineering

University of Pune

ABSTRACT

Nowadays we are finding that mammography technique is best available technique for breast cancer detection. Breast abnormalities are defined over wide range of features and it may happen that radiologist might be easily missed or misinterpreted it.The ability to improve diagnostic information from medical images can be enhanced by designing image processing algorithms that is why we proposed new algorithm to detect lesions in mammogram breast cancer images. In this paper we proposed an algorithm which is implemented on MATLAB. In developing the algorithm, we focused on color pixel intensity. This paper gives a survey of image processing algorithm and comparison among all of them. Lastly we compare all the results of different algorithm (results are taken as standard according to previous work by researchers on them) which are explained in this paper with our algorithm result.

Keywords

breast cancer,X-ray mammography, image processing, segmentation, colorography, tumor.

1. INTRODUCTION

Diagnosis and prognosis of breast cancer in its early stage provides possibility of successful treatment and complete recovery of the patient from the dieses. Breast cancer is one of the major causes of death among women since the last decades and it has become a serious problem for the healthcare systems of different countries. This disease became a commonest cancer among women. If the cancer can be detected early, the possibility of treatment and recovery will increase a lot. Intra-operative diagnosis of the disease has steadily become more important because of the introduction of sentinel lymph node biopsy. A sentinel lymph node is classed as any node that has a direct lymphatic connection to the cancer, and would therefore be the most likely location of cancer spreading from the breast. A number of research works have already done in the area of breast cancer detection and classification because now a days breast cancer is becoming the most common form of cancer disease of today's female population. Thus an adaptive system is required which can classify and detect breast cancer which becomes a challenging and difficult task now a days. Imaging techniques can be used to perform breast biopsies, especially of abnormal areas which can be seen on a conventional mammogram or with ultrasound. Stereotactic-guided biopsy is a technique which can be used to see abnormal area of precise location in three dimensions using conventional mammography. Stereotactic means the use of a computer and scanning devices for the creation three-dimensional images. After that a needle is inserted in to breast to obtain samples and by moving needle in abnormal areas, additional samples can also be obtained. In this technique the problem is that the patent gets a lot of pain. There are serious complications resulting from the Mammotone breast biopsy system.

COLOROGRAPHY is a technique for recording x-ray images in computer code instead of on x-ray film, as with conventional mammography. The images are displayed on a computer monitor and then enhanced (lightened or darkened) before they are printed on film or on monitor. Images can also be manipulated; the radiologist can magnify or zoom in on an area. From the patient's perspective, the procedure for a mammogram with a digital system is the same as for conventional mammography and the patent need not suffer.

COLOROGRAPHY has some pros over conventional mammography. The images can be stored and retrieved electronically, which makes long-distance consultations with other mammography specialists easier. Because the images can be adjusted by the radiologist, subtle differences between tissues may be noted. The improved accuracy of our technique may reduce the number of follow up procedures. Despite these benefits, studies have to be done to shown that COLOROGRAPHY is more effective in finding cancer than conventional mammography. Screening mammography is currently the best available radiological technique for early detection of breast cancer [1]. However it is hard work for doctor to mark or detect all suspicious region of breast cancer from tons of mammograms in a group medical examination. Since the incidence of breast cancer ranks first in Japanese women [2], CAD (Computer Aided Diagnosis)systems for the mammography are widely developed.Breast cancer is broadly categorized into two types; a tumor and a micro-calcification. The CAD system finding a micro-calcification has already achieved a sufficient detection rate, while that finding a tumor has not. This paper proposes a detection system of breast cancer by marking a tumor. The system extracts all suspicious regions of tumors using a Colorography algorithm. As far now we have put forward some of the facts and put light on the present system. Now COLOROGRAPHY will be took into hands for proposal.

2. BACKGROUND

Images are the input to the segmentation and, attributes extracted from those images are the output. Segmentation divides image into its constituent regions or objects. The

level to which segmentation is carried out depends upon the problem being solved. For the segmentation of intensity images like digital mammograms, there are four main approaches, namely, watershed, threshold techniques, boundary-based methods, region- based methods, and hybrid techniques which combine boundary and region criteria.

1) **WatershedTransform-:**
Watershed is a topographic technique [5] by which a pixel is interpreted. The different watershed lines may be calculated. In a segmentation process the length of gradient can be represented as elevation information. Image values are relived in successive flooding technique by construction of watersheds with adjacent catchment basins. A starting point of area is called as marker. In these method expansions of area is done in upward direction of pixel value. Center of cancer effected tissue has higher brightness than a neighbor related pixels, because in mammography the strength of brightness is more as we are going from region toward center. In this technique shape of cancer can be calculated by expanding area in the lower region of pixel value.

2) **Threshold Technique-:**
This technique considered pixels lies within a certain range which belongs to a class, but neglects all spatial information and do not suit with noise of boundaries. In mammogram segmentation of background andbreast tissue can be done by selecting a single gray level value from analysis of grey level histogram. Pixel with grey level less than threshold is taken as background and rests are breast. [3] Here only grey level values are considered, so there is often overlapping done between grey level of objects in breast and background which is a short coming of this policy.

3) **Boundary Based Method-:**
In boundary-based methods used to look for explicit or implicit boundaries between regions corresponding to different tissue types. The two most commonly used boundary based methods are known as *ridge detection* and *edge-detection*:

- Ridge detection follows the peaks (local maxima) in the original image. For example, one way to outline the shape of a crater in a two-dimensional surface plot is to ``walk along'' the crest of the crater's rim. The circular path that would be traced out would be a much more compact representation of the crater than say, the set of all points inside the crater.

- When objects don't have large enough ridges at the region boundaries for ridge detection, the *gradient operator* can be used to enhance the boundaries (i.e., edges) between distinct regions. Edge detection is identical to ridge detection except that peaks are tracked in the gradient space of the image instead of the original image space. The gradient space can be computed by simply applying the gradient operator to the entire image.

4) **Hybrid Technique-:**
These techniques combine boundary and region criteria. This class includes morphological watershed segmentation and variable-order surface fitting. The watershed method is generally applied to the gradient of the image.

3. THE RECOMMEND SCHEME.

Our system is divided into two main processes. They are image segmentation and integration process. First, our system segments image into some regions based on mammography characteristics. Segmentation Process is divided into two sub process. They are morphology and Colorography. Fig.1 shows the proposed system.

Mammography
Image

↓

Morphology

↓

Colorography

↓

Integration

↓

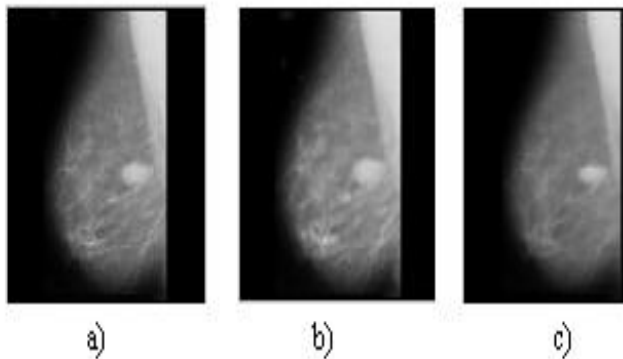Output
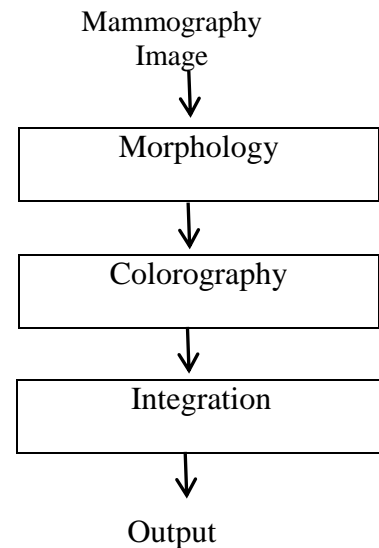
Fig1. System architecture of recommended scheme.

1) Segmentation-:

From a viewpoint of image characteristics, a tumor shown in a mammogram has a feature that higher gradient toward center in brightness. Since the Colorography is a segmentation method to expand areas along the direction of a gradient, we apply the algorithm to a tumor marking. However, a Colorography algorithm is too much sensitive to the change of a gradient, thus the image is separated more than needed due to environmental disturbances such as noise. In order to eliminate the noise influence, we apply a morphology operator [6] as preprocessing.

2) Morphology-:

In the second stage we perform morphology operation. Morphology is an operation of image processing based on shapes. The value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors [7]. The morphological operations are applied on the grayscale mammography images to segment the abnormal regions. Erosion and dilation are the two elementary operations in Mathematical Morphology [8]. An aggregation of these two represents the rest of the operations. The symbols _, Θ and o respectively denote the three fundamental binary morphological operations: dilation, erosion and opening which act on the structuring element [9]. This may result in contrast enhanced mammogram image.

This reduces high frequencies in image (i.e. noise). Figure.2 shows the dilated and eroded image.

Figure.2: a) Normal Image b) Dilated Image c) Eroded Image

3) Colorography-:

In the third step we have used our technique that is called colorography, in which we are taking input from its previous steps output i.e. morphology. We are identifying each and every tissue of muscle by its color code. We have taken or performed one survey upon different medical reports generated by medical science upon breast cancer diagnosis and we came to know that, if any tissue is affected by cancer then it can be identified by its abnormality or by its color which is different from other tissue color which are not been effected by cancer. As shown in figure.3 the abnormal part of breast is surrounded by dark black line which has color dark white and other normal tissue which are not affected by cancer has color grayish.
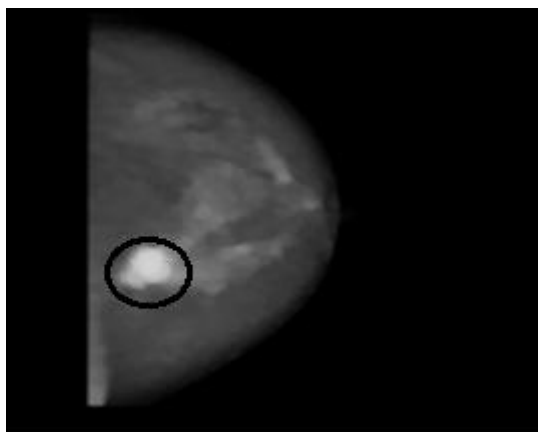
Figure3. Showing normal and abnormal tissue.

❖ **Algorithm for COLOROGRAPHY-:**

1) Start.
2) Accept the image.
3) Store the image in the data structure 2-Dimensional Array, (where each value in an array is a color code representing the color of pixel and that value can be accessed with row index and column index).
4) For each color code in an array,
   i. If color code value is 255 (i.e. white color)

Then increment counter1 (counter1 is a variable where it represents total number of white colored pixels).
   ii. If color code with value between 238 to 254 (i.e. close to white color)
Then increment counter2 (counter2 is a variable where it represents total number of pixels close to white colored).
   iii. If color code with value between 221 to 237 (i.e. just white color)
Then increment counter3 (counter3 is a variable where it represents total number of less white colored pixels).
   iv. Otherwise.
5) Calculate,
   i. The percentage of total number of white colored pixels as
   = (counter1/total number of pixels present in image)*100,
   ii. The percentage of total number of close to white colored pixels as
   = (counter2/total number of pixels present in image)*100,
   iii. The percentage of total number of less white colored pixels as
   = (counter3/total number of pixels present in image)*100.
6) Display the result obtained in step 5.
7) Stop.

So in our technique we have identified,
   i. The muscle tissue by its color and classified muscle tissue into different classes based upon its color and abnormality,
   ii. We have to categorized tissues into 4 classes,
   a. Class A-: in which tissues are purely affected by cancer and the color of tissue is dark white color.
   b. Class B-: in which tissues are partially affected by cancer (i.e. in middle stage) and the color of tissue is close to white color.
c. Class C-: in which tissues are just affected by cancer (i.e. in starting stage) and the color of tissue is just white color.
d. Class D-: in which tissues are not been affected by cancer and the color of tissue is grayish color.

In our simulator we are reading one image file, and then we have used certain function of MATLAB which will divide the image into number of pixels[10]. Then identify each and every pixel based upon its color and color code intensity. After calculating how many number of pixels in each and every class we are using following formula to calculate the percentage of each class,

= (total number of pixels in particular class/total number of pixels in image)*100.

4) Integration and Analysis-:

After applying colorography on an image which is the output of morphology, we get result in four levels they are as classes described in above section. After analyzing these entire steps doctor comes to know actually how much percent a patient has cancer.
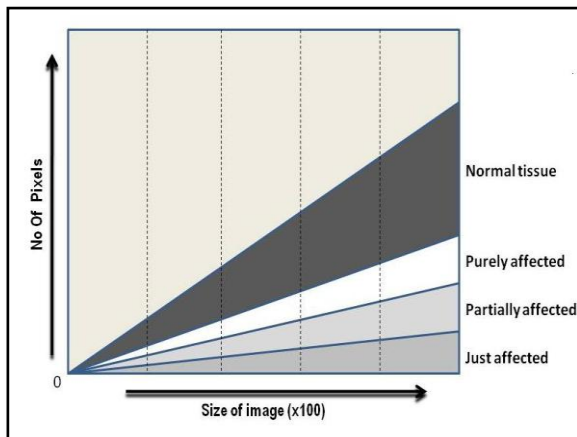
4. COMPARISON
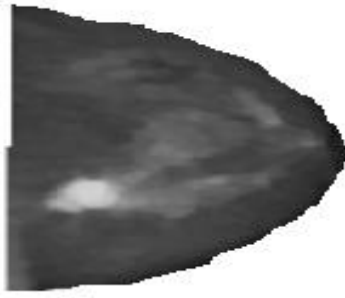
Figure 4. Sample Image

We took one image shown in following figure. 3, on which we performed our algorithm and calculated.

Table. Comparison of Each Class

Class	CounterValue	Percentage
Class A (Purely affected tissues)	8613	36.8502
Class B (Partially affected tissues)	72	0.30805
Class C (Just affected tissues)	06	0.025671
Class D (Normal tissues)	14682	62.816069824
Total	23373	99.999990824 (100)$_{Approximately}$

Figure5. Graph Showing Affected Cells Of Each Class

The value of each and every class which is explained in recommended scheme above table shows comparison between each and every class on the basis of counter value and we calculated the percentage of each class by formula given in recommended scheme.

5. CONCLUSION

The system we present here cannot be used in a hospital setting at the current levels of performance, at Least not as a primary method of diagnosis. However, since the system is fully automatic and works with very low resolution images, it can function as a high-speed preprocessor for a more complex (hence, higher performing) diagnostic system. From the relevant works we can see that noticeably better classification result can be obtained from magnifications ×400 upwards. Our system thus processes 100 times less data than these systems and can be expected to be about 100 times faster, even without taking into account the need for human intervention in other systems.

Our work may be improved in several ways. First, the software on which we are going to apply the colorography algorithm should be platform independent; it is likely that the addition of other characteristics of the images may improve the representation. Other features we will consider and that may be useful for similar results. Our experiments with using perimeters for the features showed performance similar to that of areas. We believe that significant improvements in accuracy can be obtained by considering an ensemble of classifiers each working on a different magnification level. The aggregation of results from these classifiers can be made using simple majority voting or by using more sophisticated ensemble techniques. Our results on classification indicate that this approach is effective. However, these results were obtained using a straightforward approach. It is likely that these results can be improved using more sophisticated optimization techniques. Also, the simplicity of our algorithm is natural than in earlier algorithms which produce probabilities. Therefore, the application of colorography algorithm may improve the performance. Finally, we note that while the dataset we used is not considered small (compared to related work) it would be essential to test the system on larger and complex images before any attempt is made to use the system in a clinical setting.

6. FUTURE WORKS

We implemented the detecting system, which marks all suspicious regions of a tumor from a mammogram. In the near future we will be focusing on various ways to enhance our system and to consider various aspectsof cancer varieties that have been found and those that is to be discovered in the times to come. The system is in its rudimentary state and our efforts will make the system evolve in more comprehensive. In the future, we will need to improve algorithm to decrease average numbers of fault detection in the proposed system.

7. ACKNOWLEDGEMENT

8. REFERENCES

[1] Acha, B., Rangayyan, R.M., Desautels, J.E.L.: Detection of MicrocalcificationsinMammograms. In: Suri, J.S., Rangayyan, R.M. (eds.) Recent Advances in Breast Imaging,Mammography, and Computer-Aided Diagnosis of Breast Cancer. SPIE, Bellingham(2006)

[2] Hidefumi KOBATAKE, Morphology, Corona Ltd., 1996

[3] Sumit Chopra, PankajBhambri, Baljit Singh, Segmentation of the Mammogram Images to find Breast Boundaries , IJCST Vol. 2, Issue 2, June 2011.

[4] IndraKantaMaitra, Sanjay Nag, Samir Kumar Bandyopadhyay, Technique for preprocessing of digital mammogram, computer methods and program and medicine,2011.

[5]Leila Shafarenko and Maria Petrou, " Automatic Watershed Segmentation of Randomly Textured Color Images", IEEE Transactions on Image Processing, Vol.6, No.11, pp.1530-1544, 1997.

[6] L.Vincent, P.Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations", IEEE Trans, PAMI, vol. 13, no 6, pp.583-598, June 1991.

[7] R.B. Dubey, M. Hanmandlu, S.K. Gupta, A comparison of two methods for the segmentation of masses in the digital Mammograms, computerized Medical Imaging and Graphics 34 (2010) 185–191.

[8] R.C. Gonzalez, R.E. Woods, Digital Image Processing, second ed, Prentice-Hall, (Englewood Cliffs, NJ, 2002

[9] Shavi Gupta1, Mohd. Sadiq2, Mona Gupta3 and NaseemRao, Semi Automatic Segmentation of Breast Cancer for Mammograms Based on Watershed Segmentation, Proceedings of the 5th National Conference; INDIACom-2011, Computing For Nation Development, March 10 – 11, 2011.

[10] Dhananjay K. Thekedath, Image Processing through MATLAB Codes, third edition.