Global Journal of Science Frontier Research: A
Physics and Space Science
Volume 22 Issue 1 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Segmentation of Cancerous Mammography using MATLAB

By Samuel Yemoh Tetteh-Abaku, Calvin Kwesi Gafrey & Moses Jojo Eghan

Abstract- Breast cancer is one of the main causes of cancer death in women. Detection is efficiently performed by using digital mammograms. Small clusters of micro calcifications appearing as a collection of white spots on mammograms show an early warning of breast cancer. Early detection performed on X-ray mammography is the key to improving breast cancer diagnosis. To increase radiologists' diagnostic performance, several computer-aided diagnosis (CAD) schemes have been developed to improve the detection of primary identification of this disease. In this research, an attempt is made to develop an adaptive K-means clustering algorithm for breast image segmentation to detect microcalcifications. The method was tested over several images of image databases taken from Mammocare, Ghana for cancer research and diagnosis. The algorithm works faster so that any radiologist can take a clear decision about the appearance of microcalcifications by visual inspection of digital mammograms and detection accuracy has also improved as compared to some existing works.

GJSFR-A Classification: FOR Code: 249999

SEGMENTATIONOFCANCEROUSMAMMOGRAPHYUSINGMATLAB

Segmentation of Cancerous Mammography using MATLAB

Samuel Yemoh Tetteh-Abaku α, Calvin Kwesi Gafrey σ & Moses Jojo Eghan ρ

Abstract- Breast cancer is one of the main causes of cancer death in women. Detection is efficiently performed by using digital mammograms. Small clusters of micro calcifications appearing as a collection of white spots on mammograms show an early warning of breast cancer. Early detection performed on X-ray mammography is the key to improving breast cancer diagnosis. To increase radiologists' diagnostic performance, several computer-aided diagnosis (CAD) schemes have been developed to improve the detection of primary identification of this disease. In this research, an attempt is made to develop an adaptive K-means clustering algorithm for breast image segmentation to detect microcalcifications. The method was tested over several images of image databases taken from Mammocare, Ghana for cancer research and diagnosis. The algorithm works faster so that any radiologist can take a clear decision about the appearance of microcalcifications by visual inspection of digital mammograms and detection accuracy has also improved as compared to some existing works.

I. Introduction

Breast cancer is a type of cancer with the highest incident rates in women. It has been one of the major causes of death among women since the last decades and it has become an emergency for the healthcare systems of countries. It is commonly classified into four stages according to the size of tumors and degree of cancer spread from the breast to other body parts and takes years to develop[5].

Mammography is an imaging study that uses X-rays to image the breast to look for cancer. There are two main types of mammography: film-screen mammography and digital mammography also called full-field digital mammography or FFDM. The technique for performing them is the same. What differs is whether the images take the form of photographic films or digital files recorded directly onto a computer. Mammography also has its limitations. It is less reliable on the dense breast of young women or women who underwent surgical intervention in the breast because glandular and scar tissues are as radiopaque as abnormalities[9]. Furthermore, there is low-dose X-Ray radiation. The estimated sensitivity of radiologists in breast cancer screening is only about 75%. Double reading has been suggested to be an effective approach to improve sensitivity. To improve the accuracy of interpretation, a variety of Computer Assisted Detection (CAD) techniques have been proposed. Interpretation of mammograms mainly involves two major processes: Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADi)[20]. It would be valuable to develop a CAD algorithm using extracted features from the breast profile region; region of interest (ROI). This would reduce the number of biopsies in patients with benign disease and thus avoid patients' physical and mental suffering, with a bonus of reducing healthcare costs.

Initial detection of the cancerous mammogram helps in the early diagnosis of a disease a diseased person which can reduce death possibilities. Methods developed for the detection of the malignant region of the mammograms may not be able to provide results successfully[1,15]. Finding an accurate, robust and efficient breast profile segmentation remains a challenging problem in digital mammography. Hence mammography misses about 17% and up to 50% of breast cancers due to the subtle and unstable appearances of breast cancer in their early stages[8]. To overcome this limitation. It is necessary to develop an approach that can segment malignant regions properly.

A significant method that first detects the cancerous region and then segment the area covered by malignant tissues was proposed. In this paper, the focus was placed on detecting malignant tissues which represent higher intensity values compared to background information and other regions of the breast. However, in the case of some normal dense tissues having similar intensities to the tumor region, it is necessary to detect tumor region excluding those regions successfully. In this research work, an attempt was made to study the effect of L*a*b color space K-means clustering on colour image segmentation. Several general-purpose algorithms have been developed for image segmentation including detection followed by segmentation of mammogram images based on simple image processing techniques using the L*a*b colour space K-means clustering algorithm which provides good results in real-time[10].

II. Materials and Method

a) Sample of images used

Mammogram images were retrieved from the website of Mammocare Ghana. The mammogram images were acquired from Ghanaian patients. The images consist of left and right breast images of fatty, fatty-glandular and dense-glandular breasts, true positive and true negative breast images, false positive

Author σ: e-mail: calvin.gafrey@stu.ucc.edu.gh

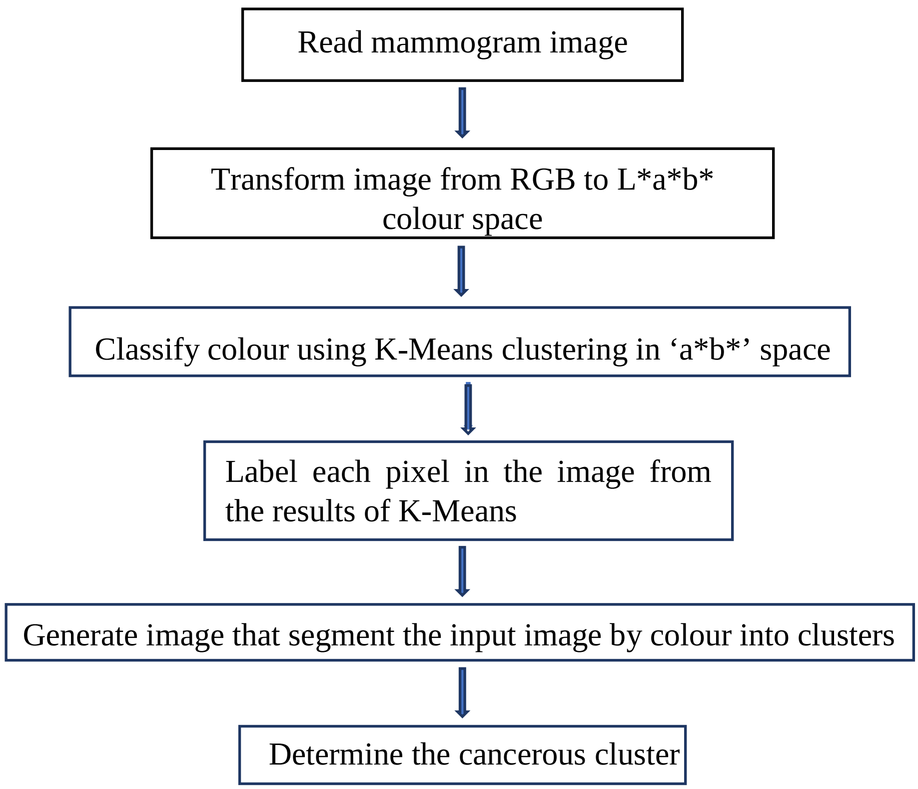
Global Journal of Science Frontier Research (A) Volume XXII Issue I Version I Year 2022

and false negative breast images. The retrieved mammogram images are classified into left breast lesion, right breast lesion, and non-palpable left breast lesion.

The images were retrieved in *Joint Photographic Experts Group*(JPEG)format. The pixels in the images were represented as an 8-bit word. The images were retrieved in Red, Green, Blue (RGB) format each with a pixel size of 500 $\times$ 500. The mammogram images were from diagnoses conducted by radiologists and clinicians using the Breast-i device. In taking the images, a patient undresses and in a darkened room, sitting slightly forward places the Breast-i light source on the inferior surface of (underneath) the breast. The patient views the superior aspect (top surface) of each breast, which should be uniformly bright except for typically a few darker lines corresponding to superficial blood vessels. The mammogram images are classified into three major cases: malignant, benign and normal.

b) Developed Matlab algorithm for image analysis

A flowchart of the method which is implemented in MATrix LABoratory (Matlab) application software (R2013a Matlab, Math Works Inc) is described in figure 3. The basic aim of the proposed approach is to segment colors automatically using the K-means clustering technique and L*a*b* colour space.



The mammogram images were read into MATrix LABoratory (Matlab) application software (R2013a Matlab, MathWorks Inc) from a folder in which they were saved. The images were transformed from RGB to L*a*b* colour space. The L*a*b* colour space was used because it consisted of a luminosity layer 'L*' and two chromaticity layers in 'a*' and 'b*'. Using the L*a*b* colour space is computationally efficient because all of the colour informationis present in the 'a*' and 'b*' layers only[12]. The colors were then classified using K-Means clustering in the 'a*b*' space. To measure the difference between the two colors, the Euclidean distance metric was used.Each Pixel was labelled in the image from the Results of K-Means. For every pixel in the input, K-means computed an index corresponding to a cluster. Every pixel of the image was labelled with its cluster index,also the mean 'a' and 'b' value for each area was extracted. These values served as colour markers in the 'a*b' space. The index image was further processed to generate 3clusters based on colour information[12]. The pixels in the image was separated by colour using pixel labels, which resulted in different images based on the number ofclusters. The results of the nearest neighbor classification were displayed. The labelled matrix contained a colour label for each pixel in the mammogram image. The labelled matrix was then used to separate objects in the original image by colour. The index of each cluster containing the cancerous part of the mammogramwas determined because K-means doesnot return the same cluster index value every time but this was done using the center value of clusters, which contained the mean value of 'a*' and 'b*' for each cluster[12].

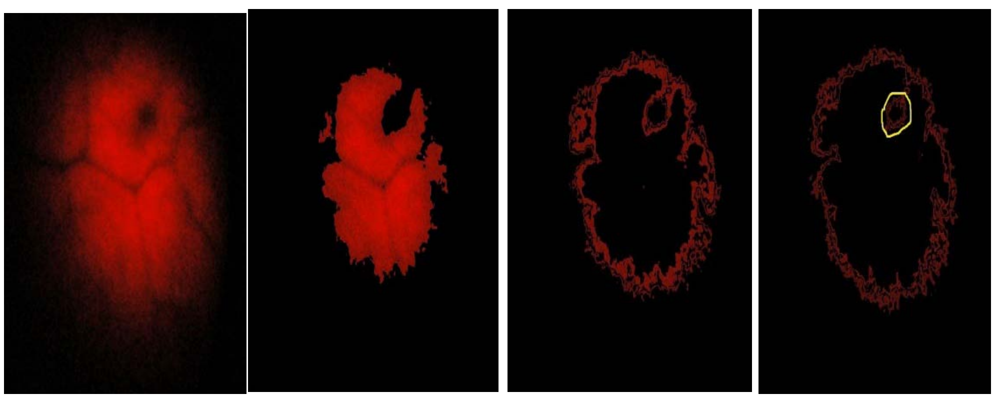
III. Results and Discussions

The proposed methodology has been evaluated on images collected from the database that belongs to Mammocare Ghana, emphasizing the importance of K-Means clustering algorithms in cancerous mammography segmentation. In the proposed methodology, specifically, the effectiveness of the segmentation methods was evaluated on the RGB image, based on the intensity levels of the segmented output.

The result of clustering intensities of the colour bands into different groups using the K-means clustering algorithm were a set of three distinct RGB level regions. These regions, referring to the colors existing in the original image are presented in figure 4, samples 1 to 8. Each region is relatively homogeneous in terms of pixel intensity. These regions were breast intensities (cluster 1), background intensities (cluster 2) and tumor intensity (cluster3) referring to the colors existing in the original image[18]. Therefore, it was assumed that there were three classes of objects to be separated with the K-means clustering algorithm.

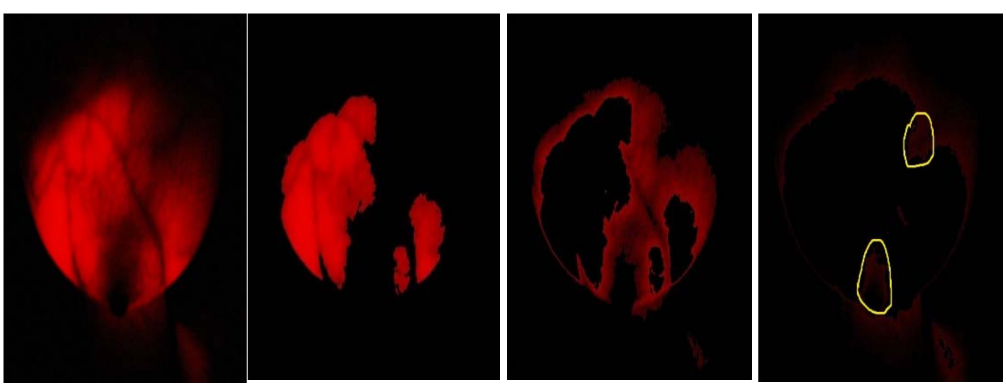
The figures show an original image from the image database and results for clusters using the K-means clustering method with only 4 clusters, and varying the values of classes are shown. Four clusters were used because using three clusters was not sufficient in that case due to the natural variability of sharpness in the input mammogram image. It can be seen that mass and lesion elements in the breast image became clearer by increasing the number of classes keeping the constant value of bins, visual appearance and classification of microcalcification get improved[6].

The following figures show the original images and their processed images using K-means clustering in L*a*b* colour spaces.



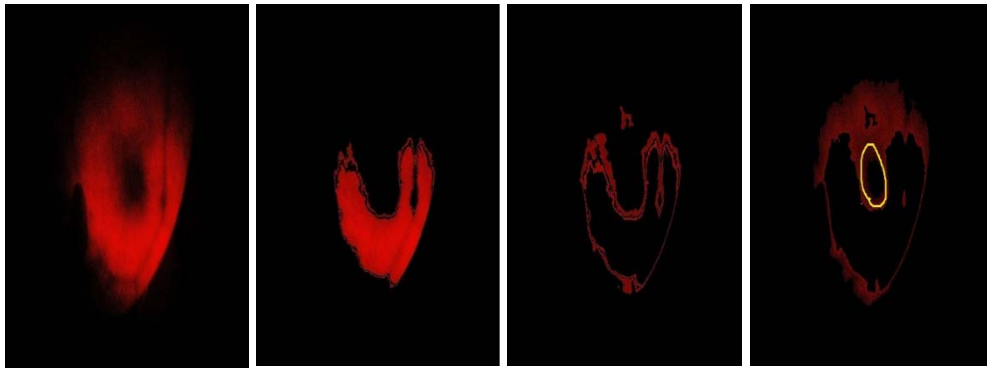
(A) (B) (C) (D)

Figure 4: Sample 1 shows, (A) the original image, (B) cluster 1, (C) cluster 2 and (D) cluster 3 after separating objects by colour using K-means clustering technique in L*a*b* colour space with the tumor region circled in yellow



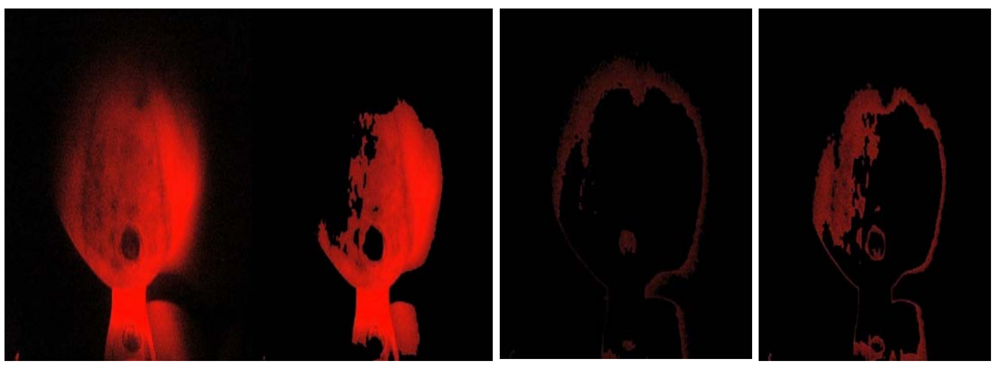
(A) (B) (C) (D)

Figure 5: Sample 2 shows, (A) the original image, (B) cluster 1, (C) cluster 2 and (D) cluster 3 after separating objects by colour using K-means clustering technique in L*a*b* colour space with the tumor region circled in yellow



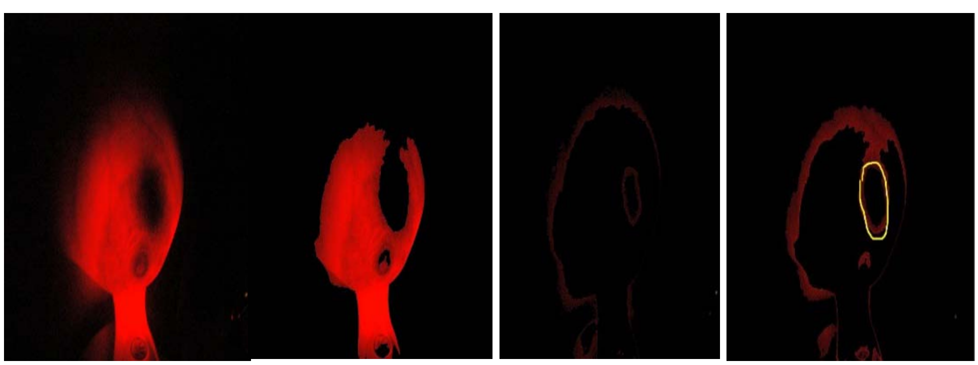
(A) (B) (C) (D)

Figure 6: Sample 3 shows, (A) the original image, (B) cluster 1, (C) cluster 2 and (D) cluster 3 after separating objects by colour using K-means clustering technique in L*a*b* colour space with the tumor region circled in yellow



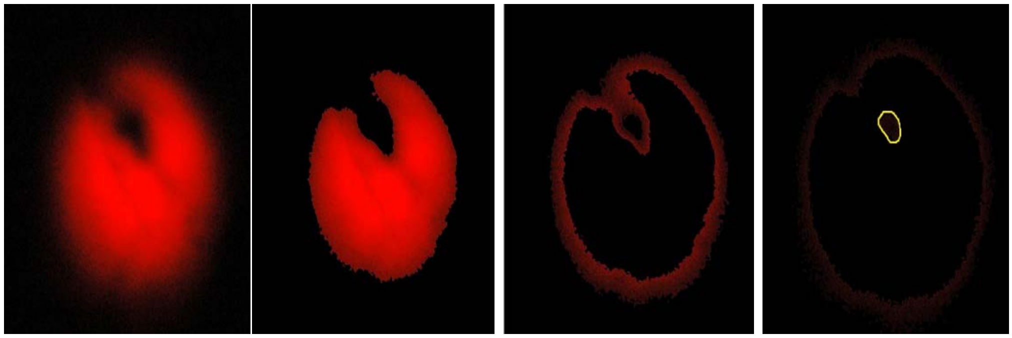
(A) (B) (C) (D)

Figure 7: Sample 4 shows, (A) the original image, (B) cluster 1, (C) cluster 2 and (D) cluster 3 after separating objects by colour using K-means clustering technique in L*a*b* colour space with no tumor found



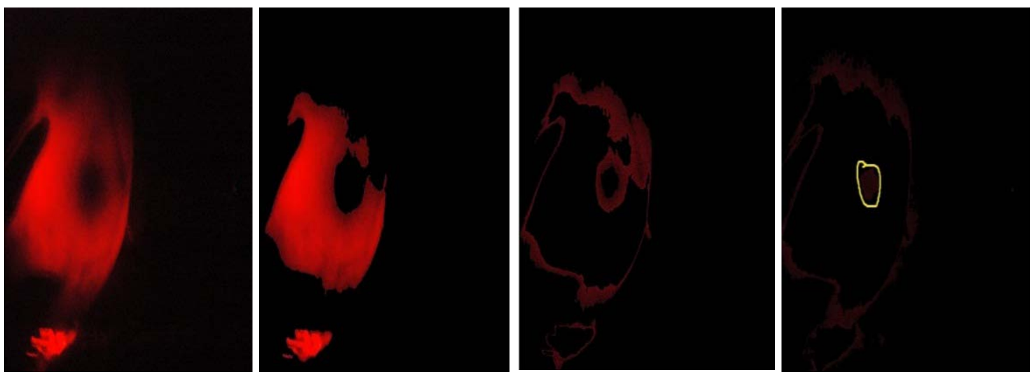
(A) (B) (C) (D)

Figure 8: Sample 5 shows, (A) the original image, (B) cluster 1, (C) cluster 2 and (D) cluster 3 after separating objects by colour using K-means clustering technique in L*a*b* colour space with the tumor region circled in yellow



(A) (B) (C) (D)

Figure 9: Sample 6 shows, (A) the original image, (B) cluster 1, (C) cluster 2 and (D) cluster 3 after separating objects by colour using K-means clustering technique in L*a*b* colour space with the tumor region circled in yellow



(A) (B) (C) (D)

Figure 10: Sample 7 shows, (A) the original image, (B) cluster 1, (C) cluster 2 and (D) cluster 3 after separating objects by colour using K-means clustering technique in L*a*b* colour space with the tumor region circled in yellow

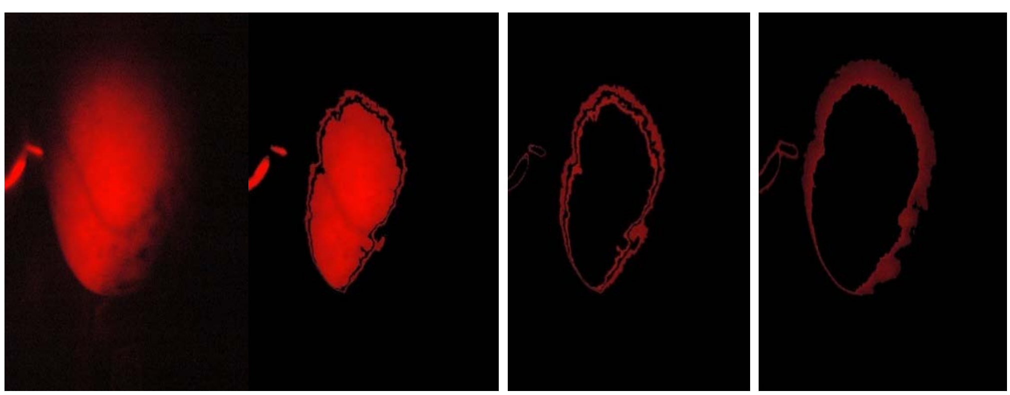


Figure 11: Sample 8 shows, (A) the original image, (B) cluster 1, (C) cluster 2 and (D) cluster 3 after separating objects by colour using K-means clustering technique in L*a*b* colour space with no tumor region

Figures 4 to 11 show the defect segmentation result of the breast with masses and lesions using the K-means clustering technique. After segmenting the input image into four clusters in figures 4 to 11, it was affirmative that the fourth cluster correctly segmented the tumor portion of the image[1]. From the empirical observations, it was observed that using 3 or 4 clusters yielded good segmentation results. Thus, in this experiment, the input images were partitioned into four segments as it also shows the detection result on an image mass while considering a different number of clusters for K-Mean clustering[4]. When the number of clusters is set to 2, one cluster contains the breast part while another one contains mass and background.

If the number of clusters is increased to 3, the defective part is separated with a background. Hence, we further increased the number of clusters to 4. In Figures 4 to 11, samples 1 to 8, the segmentation result was better for 4 clusters than 3 clusters because the area of masses and lesions in the breast is less and the colour of the mass part is quite similar to the colour of the rest of the mammogram. So, if the mass area is larger, fewer clusters will be required while in the mass area is smaller more clusters will be needed. It means the number of clusters required for the mass detection from the mammogram image is inversely proportional to the mass area[3,7,11].

The following shows the results for the fourth clusters of all the images and their respective histograms also obtained using the MATLAB R2013a *imhist* tool. This can be observed that the affected regions are more accurately located i.e. the identification of affected areas with malignant effects gets more prominent.

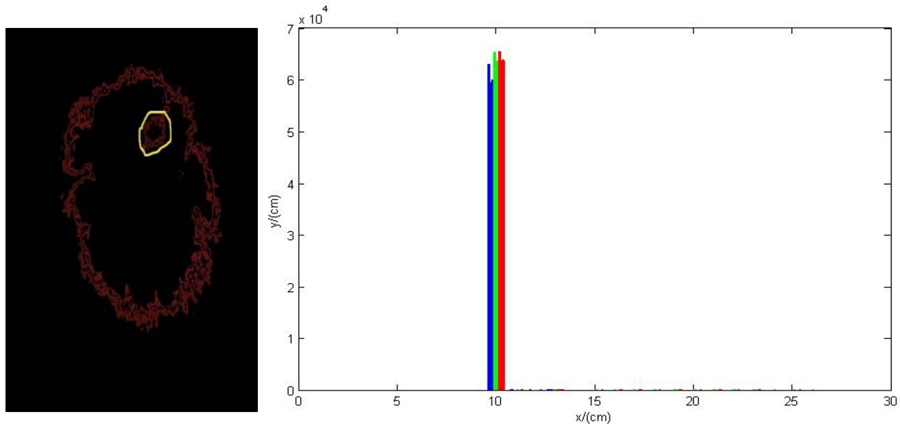


Figure 12: The processed image of sample 1 and its histogram

From figure 12, The tumor was seen when the fourth cluster was applied as compared to the 3rd cluster.

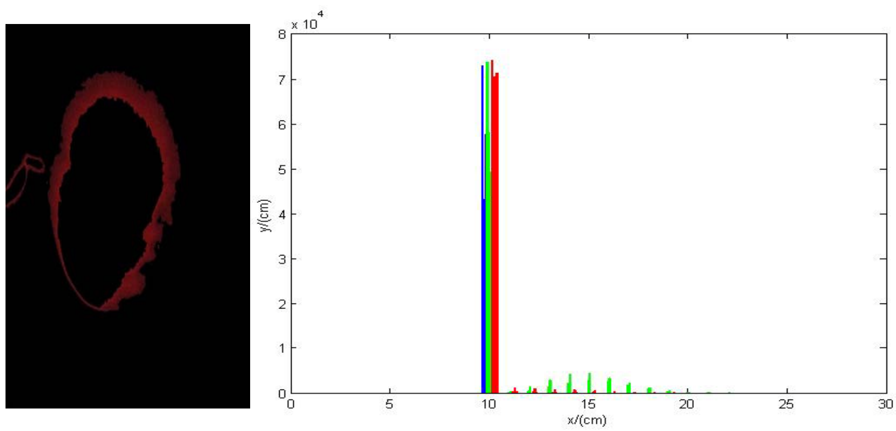


Figure 13: The processed image of sample 2 and its histogram

From figure 13, no tumor was seen in the breast as compared to the original image.

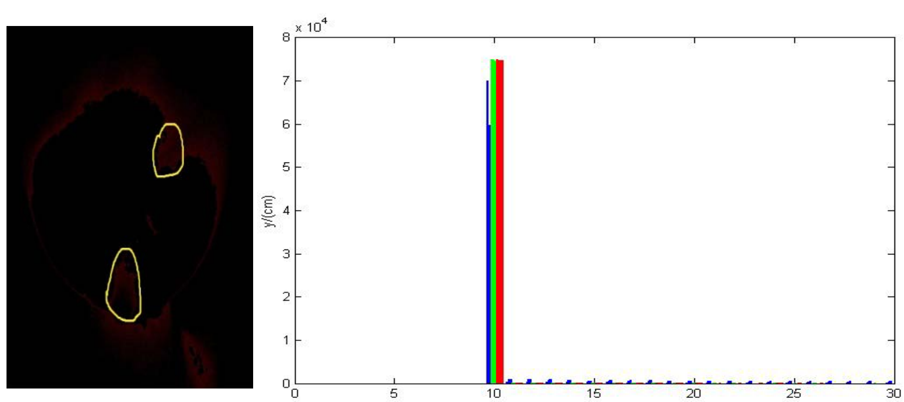


Figure 14: The processed image of sample 3 and its histogram

From figure 14, two masses were identified in the breast as compared to the original images acquired from Mammocare Ghana, and the histogram was indicated.

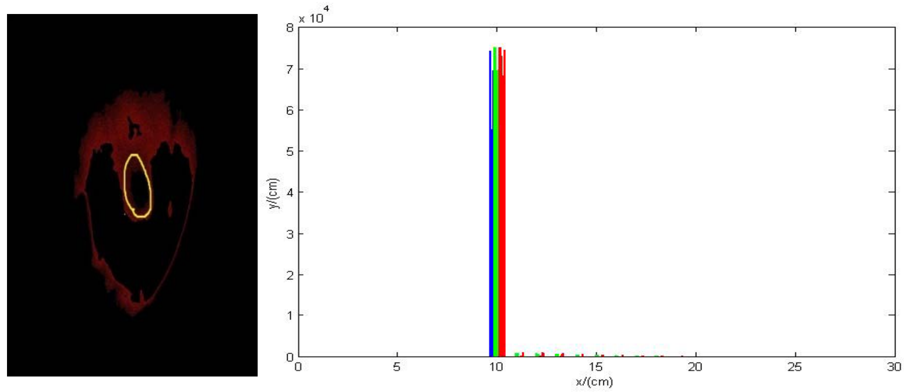


Figure 15: The processed image of sample 4 and its histogram

In figure 15, the tumor was identified in the center of the breast with the red boundary around it and it can be seen in the histogram. Although the tumor was well extracted, the red boundary was a result of the unclear nature of the image acquired.

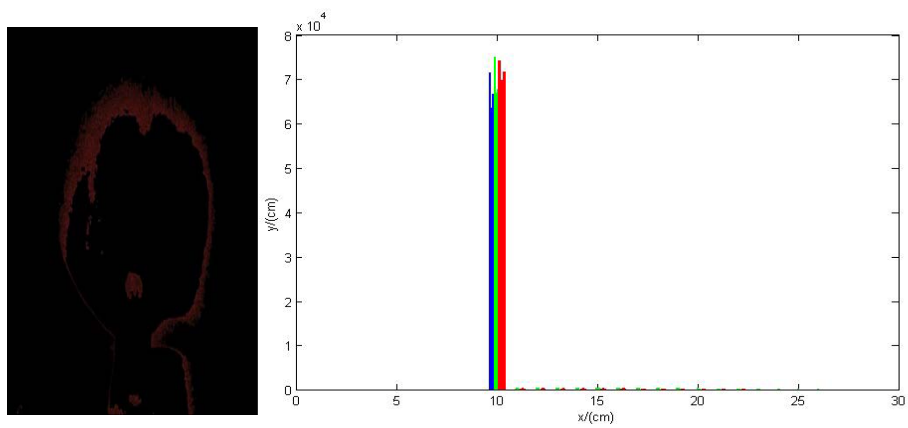


Figure 16: The processed image of sample 5 and its histogram

From figure 16, no tumor was identified as compared to the original image acquired from Mammocare Ghana. The histogram indicates the red boundary seen in the image.

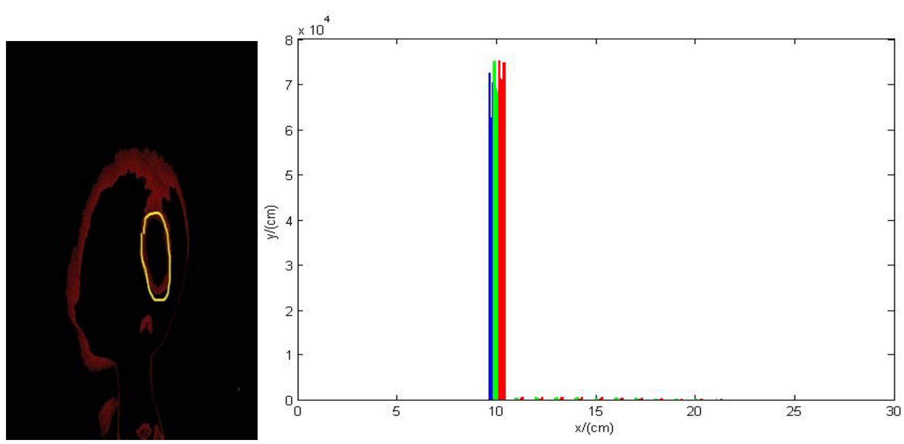


Figure17: The processed image of sample 6 and its histogram

From figure 17, although the area was large, the histogram was not clear enough this is due to the red circle around the tumor. This is because of the blurred nature of the original image and the direction in which the image was taken and hence the mass could be identified within the red circle, further increase of the cluster to be able to well identify the tumor only resulted in the whole image becoming black.

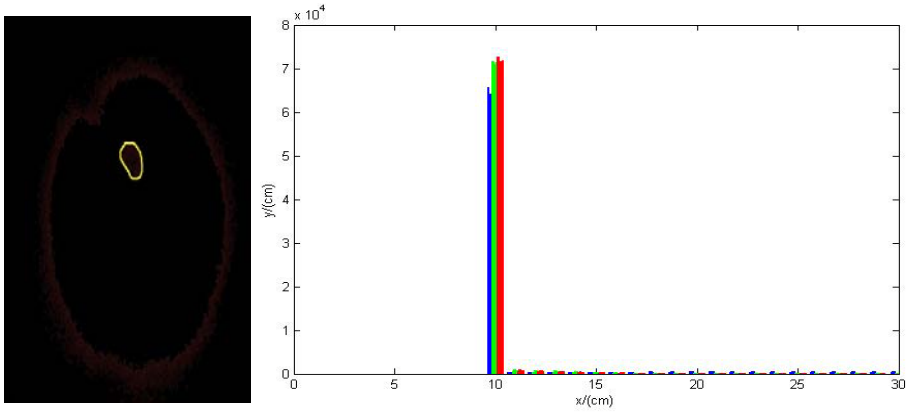


Figure 18: The processed image of sample 7 and its histogram

From figure 18, it can be seen that the mass was seen when the fourth cluster was applied, the histogram also indicated that the tumor has gotten to the mass stage but smaller as compared to the one in figure 19.

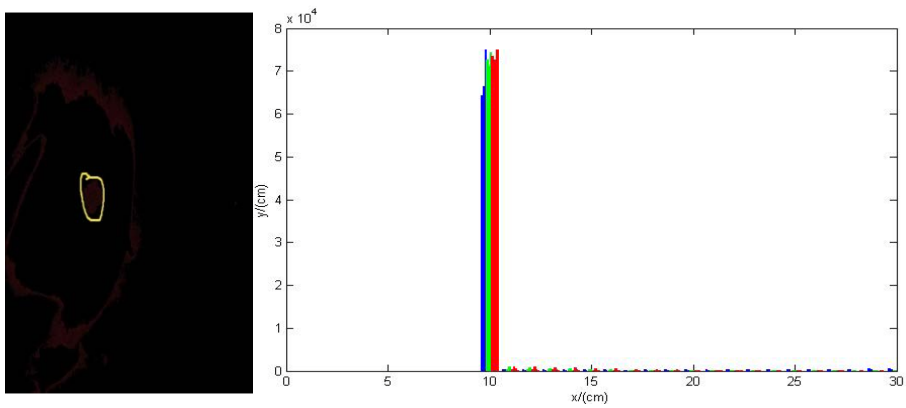


Figure 19: The processed image of sample 8 and its histogram

From figure 19, the tumor was identified when the cluster was increased to 4 the when the histogram for the image indicate that the vibrant colors showed that the tumor was large as compared to the other tumors in the other images.

The experimental results suggest that the introduced method for defect segmentation in this research is robust because it can accurately segment the cancerous part with the breast region, background and the blurred region of interest (ROI) boundary[19].

Finally, the detection accuracy was estimated and compared the performance with previous similar research works emphasizing the detection accuracy. The results obtained are also in support of anticipation with the findings and diagnosis by the radiologist of Mammocare Ghana of cancer research. The accuracy of detection has increased.

IV. Conclusion

In this work, the cancerous mammography segmentation of mammograms using K-means cluttering based on L*a*b* colour space was proposed and evaluated. The proposed approach used the K-means clustering technique for segmenting mammogram image four clusters. Mammograms images were used for the experimental observations and the introduced method was evaluated considering a cancerous mammogram as a case study. Experimental results suggest that the proposed approach is capable of accurately segmenting the tumor(mass) area of mammograms present in images. K-means based tumor segmentation approach is to also segment the cancerous area of the mammogram.

References Références Referencias

1. Abdel-Mottaleb, M., Carman, C, S., Hill, C, R., & Vafai, S. (1996). "Locating the Boundarybetween the Breast Skin Edge and the Background in Digitized Mammograms", in Proc. of the 3rd International Workshop on Digital Mammography (WDM), pp. 467–470.
2. Adam A., Omar, R. (2006.) "Computerized breast cancer diagnosis with Genetic Algorithm and Neural Network", in Proc. of the 3rd International Conference on Artificial Intelligence and Engineering

Technology (ICAIET), 22–24 Nov, Malaysia, pp. 533–538.
3. Basheer N. & Mohammed M., (1998) "Segmentation of Breast Masses in Digital Mammograms Using Adaptive Median Filtering and Texture Analysis", International Journal of Recent Technology and Engineering (IJRTE), Vol. 2, No. 1, pp. 39-43.
4. *Bethapudi*, P., (1998). 'Malignant masses in *mammograms* for detection of breast cancer digital *mammography'* in Mammographic Image Processing*,* Norwell, MA, Vol 15.
5. Bick, U., Giger, M, L., Schmidt, R, A., Nishikawa, R, M., Wolverton, D, E., & Doi, K. (1997). "Automated Segmentation of Digitized Mammograms", Academic Radiology, vol. 2, no. 2, pp. 1–9.
6. Dalmiya., (2012). Learning Techniques for *Mammogram* Classification using wavelet and soft. International Journal of Recent Technology and Engineering (IJRTE), vol. 1, no. 1, pp. 50 – 96.
7. Eklund, G, W., Cardenosa G., & Parsons W. (2001). "Assessing Adequacy of Mammographic Image Quality", Radiology, vol. 190, pp. 297–307,
8. Giger, M, L., Nishikawa, R, M., & Kupinski,M., (1997). "Computerized Detection of Breast Lesions in Digitized Mammograms and Results with a Clinically-Implemented Intelligent Workstation", in Computer Assisted Radiology and Surgery, H.U. Lemke, K. Inamura., M.W. Vannier, eds., Elsevier, Berlin, Germany, pp. 325–330.
9. Hassanien, A.., Bader, A., (2003). "A Comparative study on digital mammography: Enhancement algorithms based on Fuzzy Theory", International Journal of Studies in Informatics and Control, vol. 12, no. 1, pp. 21–31.
10. Kanungo, T., Mount, D, M., Netanyahu, N., Piatko, C., Silverman, R., & Wu, A, Y. (2002). "An efficient k-means clustering algorithm: Analysis and implementation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), pp. 881-892.
11. Lou, S, L., Lin, H, D., Lin, K, P., & Hoogstrate, D. (2000). "Automatic Breast Region Extraction from Digital Mammograms for PACS and Telemammography applications", Computerized Medical Imaging and Graphics, vol. 24, pp. 205–220.
12. Mary P, S., Vennila, I. (2010). "Optimization Fusion Approach for Image Segmentation Using K-Means Algorithm", International Journal of Computer Applications (0975 – 8887), Volume 2 – No.7.
13. Mirzaalian, H., Ahmadzadeh, M, R., & Jafari,M.,"Pre-processing Algorithms on Digital Mammograms", in Proc. of the IAPR Conference on Machine Vision Applications (MVA), Japan, pp. 118–121.
14. Mendez, A, J., Tahoces, P, J., Lado, M, J., Souto, M., Correa, J, L., & Vidal, J, J. (1996). "Automatic Detection of Breast Border and Nipple in Digital Mammograms", Computer Methods and Programs in Biomedicine, vol. 49, pp. 253–262.
15. Ojala, T., Näppi, J., & Nevalainen, O. (2001). "Accurate Segmentation of the Breast Region from Digitized Mammograms", Computerized Medical Imaging and Graphics, vol. 25, no. 1, pp. 47–59.
16. Petrick,N., Chan, H,P., Sahiner, B., & Wei, D., (1996). An Adaptive Density-Weighted Contrast Enhancement Filter for Mammographic Breast Mass Detection, *IEEE Trans. Med. Image.*, Vol.15, pp. 59–67.
17. *Serhat* Özekes., (2005). A new method for automated mass detection in digital *mammographic* images using templates". ICGST International Journal on Graphics, Vision and Image Processing (GVIP), vol. 1, pp. 12.
18. Thangavel, K., Karnan, M., & Sivakumar, R., (2002). "Automatic Detection of Microcalcification in Mammograms - A Review", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), vol. 5, no. 5, pp. 23–53.
19. Wirth, A, M., Stapinski, A., (1995-2006). "Segmentation of the breast region in mammograms using active contours", in Visual Communications and Image Processing.
20. Vyborny, C, J., Yin, F, F., Giger, M, L., Doi, K., Metz, C, E., & Schmidt, R, A. (1991). "Computerized Detection of Masses in Digital Mammograms: Analysis of Bilateral Subtraction Images", Medical Physics, vol. 18, no. 5, pp. 955–963, 1991.