ORB ACADEMIC PUBLISHER

Int. J. Image Video Process. Theor. App., vol. 2014, p. 11-18.
Available online at www.orb-academic.org

International Journal of
Image and Video Processing:
Theory and Application
ISSN: 2336-0992

Detection of Regions of Interest in Mammograms by Using Local Binary Pattern and Dynamic K-Means Algorithm

Abdelali Elmoufidi, Khalid El Fahssi, Said Jai-Andaloussi, Abderrahim Sekkaki

Department of Mathematics and Computer Science, Faculty of sciences, Casablanca Hassan II University, Kingdom of Morocco
Email: abdelali.elmoufidi09@univcasa.ma, elfahssi@etude.univcasa.ma, said.jaiandaloussi@etude.univcasa.ma, abderrahim.sekkaki@etude.univcasa.ma

ABSTRACT

This paper presents a method for the detection of regions of interest (ROI) in mammograms by using a dynamic K-means clustering algorithm. This method is used to partition automatically an image into a set of regions (clusters or classes). Our method consists of three phases: firstly, preprocessing images by using thresholding and filtering methods; secondly, generating range of number of clusters by using Local Binary Pattern (LBP) and Applying k-means with its features to automatically generating the optimal number of clusters (thereafter k is The number of clusters generating); thirdly, partition the mammograms images into k clusters by applying the dynamic k-means clustering algorithm, we end by detecting the regions of interest (ROI) in mammograms images. To demonstrate the results of our proposed method we used the Mini-MIAS (Mammogram Image Analysis Society, UK) database, consisting of 322 mammograms. Our method's performance is evaluated using Free response ROC (FROC) curves. The archived results are 2.84 false positives per image (FPpI) and sensitivity of 85%.

KEYWORDS

Mammography Images — Breast Cancer — K-means clustering — Local Binary Pattern (LBP)— Region Of Interest (ROI).

1. Introduction

Breast cancer is the most common type of cancer among women worldwide. Recent statistics show that breast cancer affects one of every ten women in Europe and one of every eight in the United States [1]. More specifically, breast cancer is the second most common type of cancer and the fifth most common cause of cancer death according to Nishikawa [2]. Women can have the highest chance of survival if physicians are able to detect the cancer at its early stages, because early detection is key in the treatment of breast cancer. For that reason, the mammography remains the best and most accurate tool in detecting breast cancer. One of these tools is the Computer-Aided-Diagnosis (CAD). It has a direct impact on the analysis and treatment of early breast cancer.

Generally the procedure Computer-Aided-Diagnosis (CAD) for the detection mass takes place in three stages:(1) Detection of the regions of interest (ROI), (2) Segmentation of the (ROI), and (3) Classification. The regions of interest (ROI) extraction is a capital step in the mammography segmentation. For this there are several techniques that have been published for the detection of the regions of interest, Such as: Edge-Based Techniques [3], Region-Based Techniques (region growing [2,4-5], Split-and-merge [6], and clustering [1-7,8]).

In this paper, a method of the detection of (ROI) was developed, based on dynamic k-means clustering algorithm. In the literature, many methods have been developed for the detection of regions of interest (ROI) in mammograms. Roula M. Alayli et al [9] are using thresholding algorithm for breast cancer detection, Abdu Gumaei et al [1] proposed a method based on K-means algorithm with a mixture of Gamma distributions for (ROI) detection. Nalini Singh et al [10] used K-means and Fuzzy C-means clustering for mass detection in mammograms. Our method consists of three steps: firstly, we enhance the images by applying image preprocessing techniques ((a) separate the breast profile from the background image, (b) remove digitization noises, (c) enhance the contrasts of breast profile). Secondly, we generate a number k of clusters by using LBP and K-means algorithm, and we split the image mammography on k clusters by using k-means clustering algorithm, we end by detecting the regions of interest (ROI).

The setup of the paper is as follows: Section II-A describes the database that we used for evaluation. Section II-B describes the k-means algorithm.Section II-B describes the Local Binary Pattern (LBP) algorithm. In section III we present our image preprocessing techniques. Section IV-A describes our dynamic k-means clustering algorithm.Detection of regions of interest (ROI) is described in section IV-B. Experiments results and discussion are presented in section V.

2. Methods and Materials

2.1 Database

In this paper we used the mini-MIAS database [11], which contains 322 digitized mammogram images consisted of left and right breast images. The acquired mammogram images are classified into three major cases: malignant, benign and normal. Size of these images is 1024 × 1024 pixels in Portable Greymap (PGM) format. Each pixel in the images is represented as an 8-bit word, where the images are in grayscale format with a pixel intensity of range [0, 255] [2].
Fig.1 shows of different components in the image mammography.

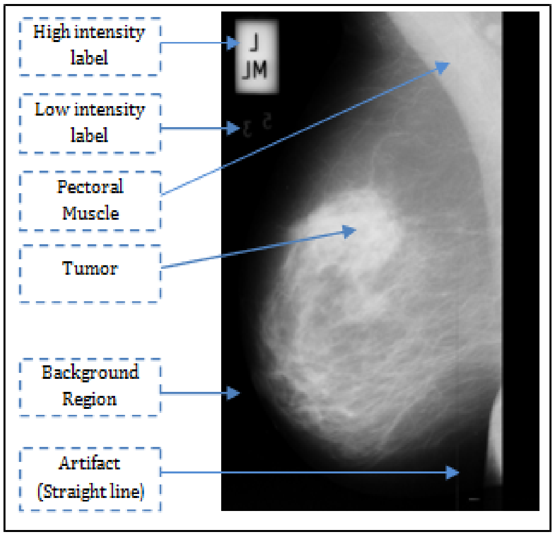

Figure 1. *Example of the elements that constitute a mammogram image*

2.2 K-means Algorithm

The k-means algorithm is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed cunningly, because different locations generate different results. So, the best choice would be to place them as far away from each other as possible. The principle of k-means algorithm is given below:

1. Define K cluster centers, either randomly or based on some heuristics.
2. Assigning each pixel to the nearest cluster is based on the minimum Euclidean distance between the point and the k cluster centers.
3. Re-compute the cluster centers.
4. Repeat step 2 and 3, a loop has been generated is the criterion stops the loop when the center does not move.

For a given set of n observation $\{S_1, S_2, ..., S_k\}$, k-means algorithm segments the observation into k cluster $\{C_1, C_2, ..., C_k\}$, your cluster center is $\{\mu_1, \mu_2, ..., \mu_k\}$, $(k < n)$ So as to minimize the within cluster sum of squares in equation (1).

$$V = arg_s min \sum_{i=1}^{k} \sum_{j=1}^{c_i} || x_j - v_i ||^2 \tag{1}$$

where:

- K : Is the number of cluster centers;
- c_i : Is the number of data points in i^{th} cluster;
- $|| x_j - v_i ||$: Is the Euclidean distance between x_i and v_i;
- v_i : Is the mean of i^{th} in C_i during each iteration; it is as follows:

$$v_i = \frac{\sum_{j=1}^{C_i} x_j{}^i}{n_i} \tag{2}$$

$$Ratio = \frac{\textbf{intra-cluster}}{\textbf{inter-cluster}} \tag{3}$$

The intra-cluster distance: is the sum of squared distance from all points to their cluster centers(see equation 4).

$$\textbf{intra-cluster} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{c_i} || x_j - v_i ||^2 . \tag{4}$$

where: N is the number of pixels in the image, k is the number of clusters, and v_i is the cluster centre of cluster c_i.
The inter-cluster distance: is the distance between cluster centres (see equation 5).

$$\textbf{inter-cluster} = min(\|v_j - v_i\|)^2 \tag{5}$$

where: $i = 1, 2, ..., k-1$ and $j = i+1, ..., k$.

2.3 Local Binary Pattren (LBP)

LBP operator combines the characteristics of statistical and structural texture analysis. The LBP operator is used to perform gray scale invariant two-dimensional texture analysis. The LPB operator labels the pixel of an image by Thresholding the neighborhood (i.e. 3 × 3) of each pixel with the center value and considering the result of this Thresholding as a binary number [14-16]. When all the pixels have been labeled with the corresponding LBP codes, histogram of the labels are computed and used as a texture descriptor. Formally, given a pixel at (x_c, y_c), the resulting LBP can be expressed in decimal form as follows:

$$LBP_{P,R}(x_c,y_c) = \sum_{P=0}^{P=1} S(i_p - i_c)2^P \qquad (6)$$

where i_c and i_p are, respectively, gray-level values of the central pixel and P surrounding pixels in the circle neighborhood with a radius R, and function s(x) is defined as:

$$S(x) = \{^{1,x\succeq 0}_{0,x\prec 0} \qquad (7)$$

From the aforementioned definition, the basic LBP operator is invariant to monotonic gray-scale transformations, which preserve pixel intensity order in the local neighborhoods. The histogram of LBP labels calculated over a region can be exploited as a texture descriptor [15].

3. our proposed approach

This section describes the details of our proposed method for detection of regions of interest (ROI) in the mammogram images. The proposed method consists of two phases, Firstly, we start by preprocessing the mammogram images, in order to:

1. separate the breast profile from the background,
2. remove the digitization noises,
3. enhance the contrast of breast profile.

Secondly, generate dynamic number of clusters by using Local Binary Pattern (LBP) and k-means algorithm. Thirdly, detecting of the regions of interest requires the extraction of the breast profile by using the Thresholding algorithm, then we used our dynamic k-means clustering algorithm to classify pixels of the mammogram images into homogeneous sets, which allows to detect the regions of interest (see Fig.2)

3.1 Mammogram image pre-processing

As in typical film scanned, the digitization of the mammography images can cause some noises at the result image. In the MIAS database several types of noise and imaging artifacts are present [12]. Therefore, computer image processing techniques will be applied to enhance the quality of images:

3.1.1 Breast profile extracted

In this phase, the aim is to extract breast profile region from background, firstly a threshold value is used to transform gray mammogram image to binary mammogram image. The value of this threshold is calculated from the minimum intensity values between the initial two most significant peaks of mammogram image histogram; these peaks represent the background and the breast [1]. Secondly, connected component is used to extract the largest component which is the breast (see Fig.3 (c)).

3.1.2 Digitization Noise Removal and Breast Contrast Enhancemant

In this phase we used a two-dimensional (2D) median filtering in a 3-by-3 neighborhood connection to remove noise. Additionally, the mammogram is usually basically low contrast [1], therefore a step of enhancement of contrast is necessary, for this purpose we used filter means.(see Fig.4).

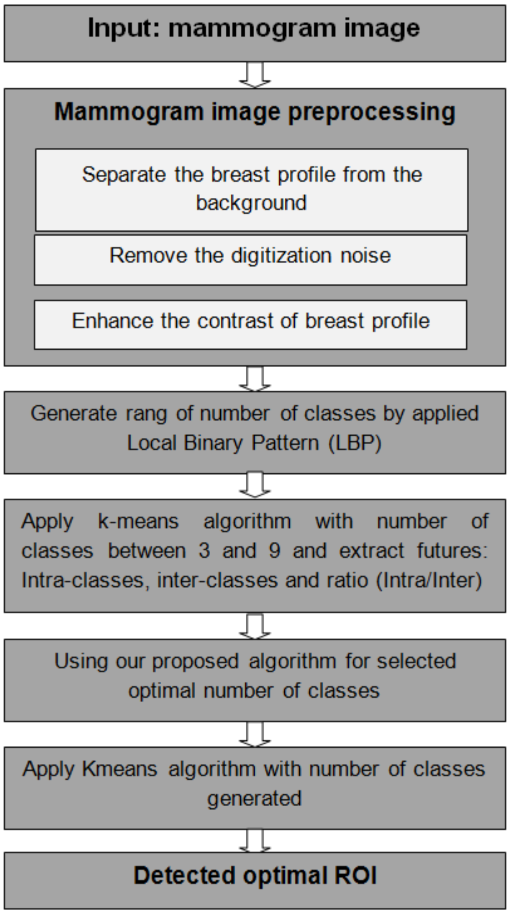
Input: mammogram image
Mammogram image preprocessing
Separate the breast profile from the background
Remove the digitization noise
Enhance the contrast of breast profile
Generate rang of number of classes by applied Local Binary Pattern (LBP)
Apply k-means algorithm with number of classes between 3 and 9 and extract futures: Intra-classes, inter-classes and ratio (Intra/Inter)
Using our proposed algorithm for selected optimal number of classes
Apply Kmeans algorithm with number of classes generated
Detected optimal ROI

Figure 2. *Organization chart of the proposed method.*

4. Detection of Regions of Interest (ROI)

4.1 Choosing the number of clusters in K-Means algorithm

One of the drawbacks of the k-means algorithm is the number of clusters (k), that must be determined by the user as an input parameter. For example, Ali et al are using kmeans algorithms to segment the breast masses with a (k=6) number of clusters, as well as Nalini Singn et al who are using kmeans algorithms to segment the breast masses with a (k=3) number of clusters. Knowing that, in database of mammogram images, the intensity, texture and shape are changing from one image to another. Therefore, taking a fixed number of clusters for all images of database is irrelevant. For this reason we have proposed an algorithm that determines the number of clusters (k) automatically for each image in the mammograms database based only on these characterisic (see algorithm 1).

Many criteria have been proposed to determine the number of clusters (k) which will be used as input parameter for K-means algorithm. Examples of these are: Hartigan criteria

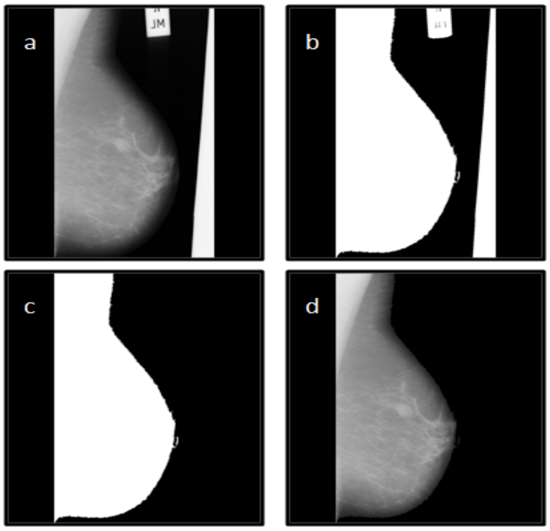

Figure 3. *Separation of breast profile region from background: (a) Original mammogram image, (b) The connected components after thresholding, (c) The largest component extracted and (d) Breast profile separated*

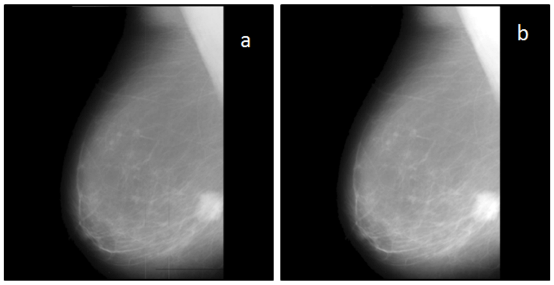

Figure 4. *Noise suppression and contrast enhancement:(a) Original image, (b) Noise removal and breast contrast enhancement.*

algorithm, square error IK-Means, absolute error IK-Means. [7, 11]. In this paper we used a method based on texture of image mammography, starting by Local Binary Pattern (LBP) for determination of initial number of clusters. We finish by characterisic intra-cluster,inter-cluster and the ratio between intra-cluster and inter-cluster to changes or not of value of number of clusters.

4.1.1 Appely Local binary Pattern

In this phase, at first we applied LBP on all images mammograms Mias-database, and represent histograms of the resulting images in a second time, the results obtained show that these histograms are similar for all images, here are two examples :

After analysis of these histograms, we found that all image contains 3 clusters as the minimum value over the background as a fourth cluster and 7 clusters as a maximum over the background as eighth cluster (see Fig.7., Fig.8.).

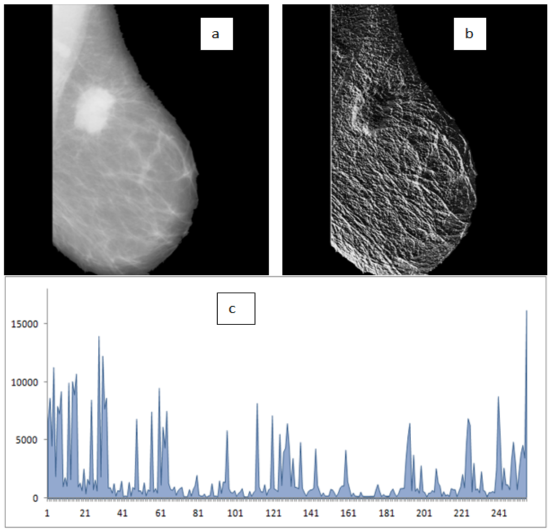

Figure 5. *Example 1: Histogram of image obtained by applied LBP algorithm:(a) Original image, (b) LBP appeled and (c) histogramm of (b).*

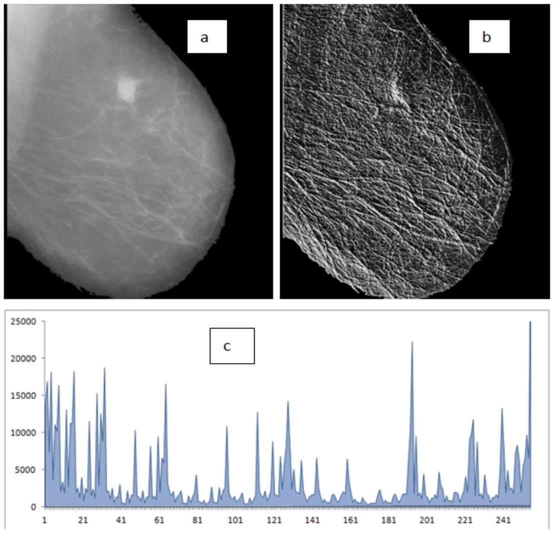

Figure 6. *Example 2: Histogram of image obtained by applied LBP algorithm:(a) Original image, (b) LBP appeled and (c) histogramm of (b).*

4.1.2 our proposed algorithm

In this algorithm, we apply the following steps:

4.2 Detection of regions of interest

Fig.10, Fig.11 and Fig.12 show three examples of the detection of regions of interest (ROI) by using our method. In Fig.10 (c), Fig.11 (d) and Fig.12 (e) the red and blue circles represent the (ROI) detected by expert. The number of clusters for these examples are presented below:

- In Fig. 10, we have 4 clusters.
- In Fig. 11, we have 5 clusters.

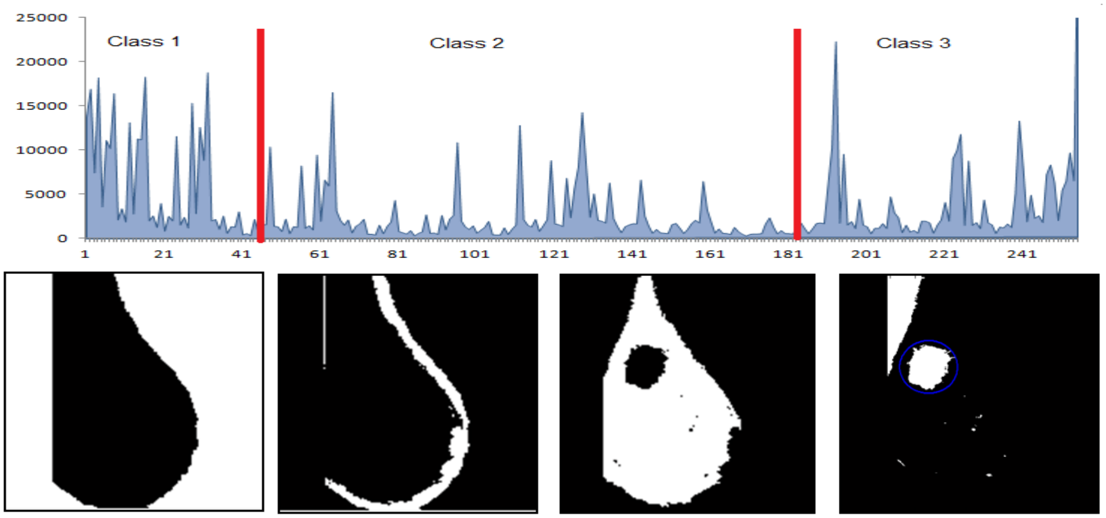

Figure 7. *Minimum number of clusters is 3 plus the background.*

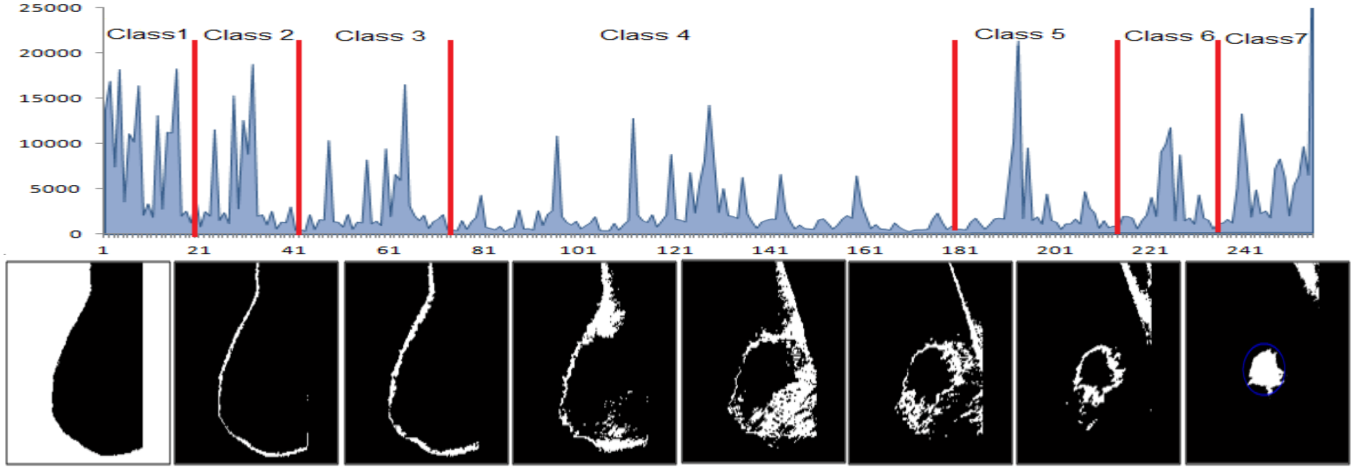

Figure 8. *Maximum number of clusters is 7 plus the background.*

- In Fig. 12, we have 6 clusters.

If the ratio (Intra clusters and inter clusters) contains a local minimum we take the value corresponding k, otherwise we will use a new indicator is the difference between inter-clusters of two successive clusters, if this value is maximum so clusters are identified in the case (number of clusters =K) and the contrary, The value of k is incremented (k = k +1) and the algorithm reapplied.

5. RESULTS AND DISCUSSION

The proposed method is tested by using the Mini-MIAS database [11], as previously mentioned in section II-(A). For evaluating our approach we used some informations offered by the (MIAS) database, such as:

- cluster of abnormality;
- image-coordinates of centre of abnormality;
- radius (in pixels) of a circle enclosing the abnormality.

To evaluate the performance of our algorithm, we calculated the percentage of pixels marked by the radiologist which are located inside the (ROI) detected by our method (the Correct Detection Rate CDR). This value (CDR)is bound between zero (no pixel is located inside the detected (ROI)) and one (all pixels are located inside the detected (ROI)) [5]. The evaluation criterion is the overlapped area ratio, which is the ratio of the overlapped area between the detected region and the criterion region segmented by the radiologists manually. In the case of (MIAS) database, the criterion region is the circle formed by the coordinates of center and radius. See equation 7.

$$CDR = \frac{TP}{TP+FN} * 100\%. \quad FPPI = \frac{Nb_F P}{NbImage} \tag{8}$$

where TP True Positives, FN False Negatives and Nb FP is the number of False Positives and Nb Image is the total number

Table 1. The detailles of the automatically selection of number of classes of three examples mdb134, mdb144 and mdb184.

image	Number of clusters	Intra cluster	Inter cluster	Ratio (Intra-cluster/Inter-cluster)	Diff (Inter cluster(k) -Inter cluster(k+1))
mdb184.pgm	K=3	253,1751	5,67E+03	0,0447	*****
	K=4	121,4879	2,79E+03	0,0435	1,11E+03
	K=5	76,5332	2,68E+03	0,0286	3,52E+02
	K=6	52,9781	2,33E+03	0,0228	2,88E+02
	K=7	38,2921	2,04E+03	0,0188	3,41E+02
	K=8	29,8588	1,70E+03	0,0176	3,51E+02
	K=9	24,2117	1,35E+03	0,018	******
mdb134.pgm	K=3	415,924	5,01E+03	0,083	*****
	K=4	208,5798	4,70E+03	0,0444	******
	K=5	136,7785	4,24E+03	0,0323	******
	K=6	101,1903	3,00E+03	0,0337	******
	K=7	67,1566	2,70E+03	0,0248	******
	K=8	52,312	2,32E+03	0,0225	******
	K=9	40,1115	2,27E+03	0,0176	******
mdb144.pgm	K=3	272,3894	4,36E+03	0,0625	*****
	K=4	165,7001	1,33E+03	0,1244	-2,23E+02
	K=5	108,6551	1,55E+03	0,0699	1,20E+01
	K=6	73,1955	1,54E+03	0,0475	3,29E+02
	K=7	53,4502	1,21E+03	0,0441	8,69E+01
	K=8	53,4502	1,21E+03	0,0441	2,80E+02
	K=9	41,6472	1,13E+03	0,037	*****

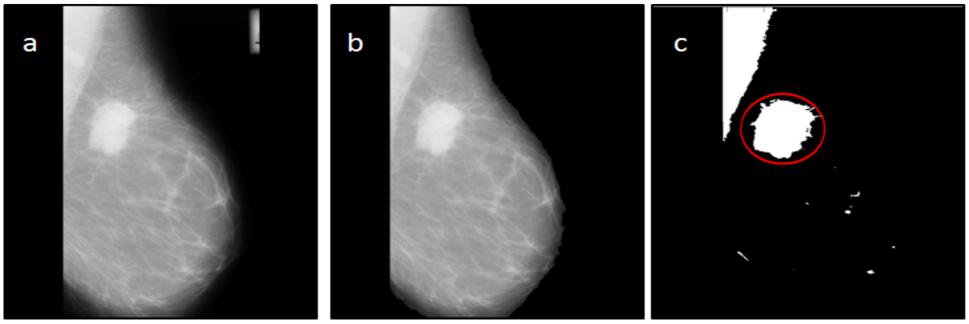

Figure 10. *(a) Original Mammography image, (b) Mammography image after preprocessing step, (c) Optimal(ROI) detected with k=4.*

of image per cluster.

The table below provides the precision percentage of detection of regions of interest that can be found on each cluster of abnormality. As you can see in some cases the precision percentage is higher than 96%. As for the mean precision of all cases, the percentage reaches 85% with 2,84 False Positive Per Image FPpI.

Other statistical methods known as Receiver Operating Characteristics (ROC) and Free-response ROC (FROC) curves are also used to analyse the experimental results. ROC curve is a graphical plot of the sensitivity against specificity for a binary classifier system as its discrimination threshold is varied [15-16].

To evaluate the performance of a CAD system, a Free-response ROC(FROC) curves is used. The free-response ROC (FROC) curve provides the performance of the overall computer aided diagnosis system in detecting the masses, as it reports the mass sensitivity against the FPpI.

6. Conclusion

In this work we presented an approach of the detection of (ROI) based on dynamic k-means clustering algorithm. This approach, consists of three main stages which are: image enhancing, gener-

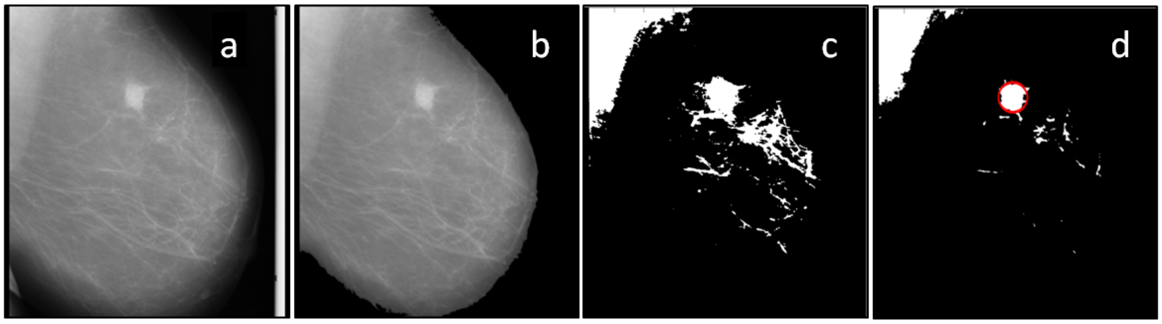

Figure 11. *(a) Original Mammography image, b) Mammography image after preprocessing step,(c) No optimal (ROI) detected with k=4 and (d) Optimal(ROI) detected with k=5.*

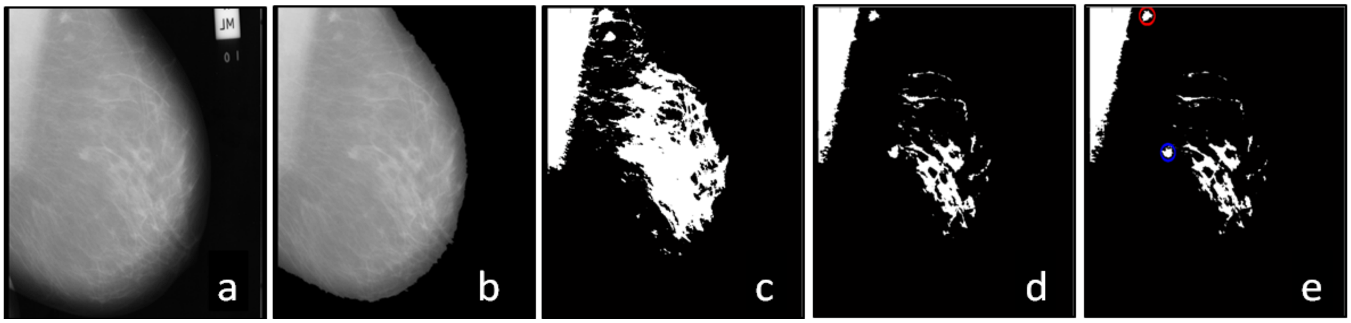

Figure 12. *(a) Original Mammography image, b) Mammography image after preprocessing step, (c) No optimal (ROI)detected with k=4,(d) No optimal(ROI) detected with k=5 and (e)Optimal(ROI) detected with k=6.*

Table 2. The details of obtained results regrouped by average in percentage of each cluster of anomaly.

cluster of abnormality present	Number of images	mean of CDR	FPpI
ARCH	19	75%	2,05
ASYM	15	93%	2,54
CALC	25	66%	3,27
CIRC	23	87%	4,12
MISC	14	96%	2,93
SPIC	19	94%	1,85

ation of a number of clusters and detection of regions of interest. The classic version of K-means algorithm requires the determination of the number of clusters (k) that be used as an input parameter. In this work we added the dynamic aspect to determine the k parameter that was based only on the numeric content of the processed image. The testing results proved that the algorithm is efficient for detection of regions of interest in mammography with a mean of 85% with 2,84 False Positive Per Image FPPI. From this results, we believe that the proposed approach can play an important role in improving the quality of the computer-aided diagnosis.

References

[1] A. Gumaei, A. El-Zaart, M. Hussien, M., and M. Berbar, “Breast Segmentation using K-means Algorithm with A Mixture of Gamma Distributions”, *IEEE 3rd SBNFI*, Lebanon, 28-29 May, 2012, pp. 97-102.

[2] J. Nagi, S. Abdul Kareem, F. Nagi, and S. K. Ahmed, “Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms”, *IECBES 2010* , 30 Nov.- 2 Dec., 2010, Kuala Lumpur, Malaysia, pp. 87-92.

[3] T. F. Chan, and L. A. Vese, “Active contours without edges”, *IEEE Transactions on Image processing*, vol. 10, no. 2, pp. 266-277, 2001.

[4] A. K. Mohanty, S. Sahoo, A. Pradhan, and S. K. Lenka, “Detection of Masses from Mammograms Using Mass shape Pattern”, *International Journal of Computer Technology and Applications*, vol. 2, no. 4, pp. 1131-1139.

[5] R. Jahanbin et al, “Automated Region of Interest Detection of Spiculated Masses on Digita Mammograms”, *IEEE Southwest Symposium on Image Analysis and Interpretation*, Santa Fe, NM, USA, 24-26 Mar., 2008.

[6] M. M. Abdelsamea, “An Automatic Seeded Region Growing for 2D Biomedical Image Segmentation”, *IPCBEE* vol.21, Singapore, 2011, pp.1-5.

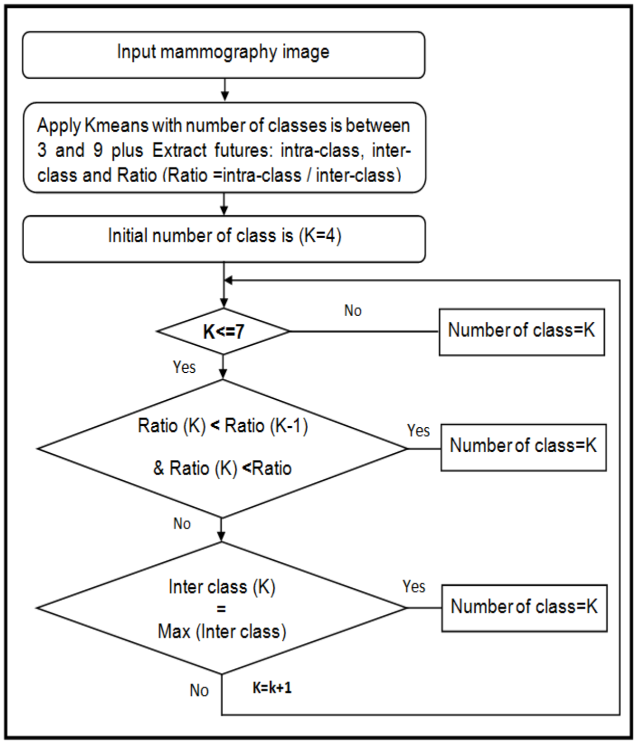
Input mammography image
Apply Kmeans with number of classes is between 3 and 9 plus Extract futures: intra-class, inter-class and Ratio (Ratio =intra-class / inter-class)
Initial number of class is (K=4)
K<=7
No
Number of class=K
Yes
Ratio (K) < Ratio (K-1) & Ratio (K) <Ratio
Yes
Number of class=K
No
Inter class (K) = Max (Inter class)
Yes
Number of class=K
No
K=k+1

Figure 9. *Algorithm 1: Our proposed algorithm for the selection of the number of clusters that is used as a parameter in K-means algorithm.*

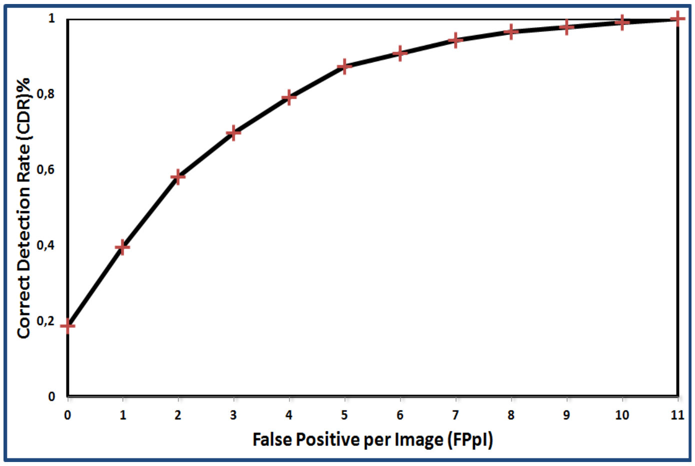

Figure 13. Curve FROC : Correct Detection Rate (CDR) against False positive per Image (FPpI)

[7] R. Siddheswar and R.H. Turi, "Determination of Number of Clusters in k-means Clustering and application in Color Image Segmentation", *ICAPRDT'99*, Calcutta, India, 1999, pp. 137-143

[8] A. T. Bon, "Developing K-Means Clustering on Beltline Moulding Contours", *Journal of Applied Sciences Research*, vol. 5, no.5, pp. 2189-2193, 2009.

[9] M. A. Roula, and A. Y. El-Zaar, "An Iterative Mammographic Image Thresholding Algorithm For Breast Cancer Detection", *ACIT'2013*, Oman, 10-13 Dec., 2013.

[10] N. Singh, A. G. Mohapatra, and G. Kanungo, "Breast Cancer Mass Detection in Mammograms using K-means and Fuzzy C-means Clustering", *International Journal of Computer Applications*, vol. 22, no.2, May 2011.

[11] J. Suckling et al, "The Mammographic Image Analysis Society digital mammogram database", *Exerpta Medica., International Congress Series 1069*, pp. 375-378, 1994.

[12] S. D. Tzikopoulos, M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, and S. Theodoridis, "A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry", *Computer Methods and Programs in Biomedicine*, vol. 102, no. 1, pp. 47-63, 2011.

[13] S. Jai andaloussi, A. Sekkaki, G. Quellec, M. Lamard, G. Cazuguel, and C. Roux, "Mass Segmentation in Mammograms by Using Bidimensional Emperical Mode Decomposition BEMD" , *The 35th IEEE International Conference of the Engineering in Medicine and Biology Society (EMBC13)*, Osaka, Japon, 3-7 July, 2013.

[14] A. M. Sabu, N. Ponraj, "Poongodi. Textural Features Based Breast Cancer Detection : A Survey ", *Journal of Emerging Trends in Computing and Information Sciences*, vol. 3, no. 9, pp. 1329-1334, Sep, 2012.

[15] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey" *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews: Applications and Reviews*, vol. 41, no. 6, pp. 765-781, Nov. 2011.

[16] X. Llado, A. Oliver, J. Mart, and J. Freixenet. "Dealing with false positive reduction in mammographic mass detection". *In Medical Image Understanding and Analysis*, pp. 81-85, 2007.

[17] A. M. Khuzi, R. Besar, W. W. Zaki, and N. N. Ahmad, "Identification of masses in digital mammogram using gray level co-occurrence matrices" , *Biomedical Imaging and Intervention Journal*, vol. 5, no. 3, 2009.

[18] T. Fawcett, ROC Graphs: Notes and Practical Considerations for Researchers. Palo Alto, USA: *HP Laboratories*, 2004.

[19] X. Llado, A. Oliver, J. Mart, and J. Freixenet. "Dealing with false positive reduction in mammographic mass detection". *In Medical Image Understanding and Analysis*, pp. 81-85, 2007.

[20] V. D. Nguyen, D. T. Nguyen, T. D. Nguyen, and V. T. Pham, "An Automated Method to Segment and Classify Masses in Mammograms", *World Academy of Science, Engineering and Technology* vol. 3 no. 4, pp. 776-781, Apr. 2009.