



A New Approach to Image Segmentation Mammogram

Mohammed Rmili^{1, 2, *}, Abdellatif Siwane³, Fatiha Adnani², Fatiha Essodegui³,
Abdelmajid El Moutaouakkil¹

¹Laboratory of Research on Optimization of Emergent Systems, Network and Imaging of computer science Department of ChouaibDoukkali University, EL Jadida, Morocco

²Laboratory of Mathematics Applied on Physics and Industry, Mathematics Department of ChouaibDoukkali University, EL Jadida, Morocco

³Radiology Departement Faculty of Medicine and Pharmacy of Hassan II University and Central Service of Radiology Ibn Rochd University Hospital Casablanca, Morocco

Email address:

gmmili@gmail.com(M. Rmili)

To cite this article:

Mohammed Rmili, Abdellatif Siwane, Fatiha Adnani, Fatiha Essodegui, Abdelmajid El Moutaouakkil. A New Approach to Image Segmentation Mammogram. *American Journal of Nano Research and Applications*. Vol. 3, No. 4, 2015, pp. 78-81.

doi: 10.11648/j.nano.20150304.12

Abstract: Breast cancer continues to be one of the main causes of death among women. Various studies have confirmed that the early detection of sub-clinical cancers may improve the prognosis. X-ray mammography in this case is the best diagnostic technique. It's based on the interaction of a cone beam X-ray with the mole tissue. The projection image obtained can be analyzed qualitatively by the radiologists. But, an automatic treatment and quantitative analysis of this kind of images is suitable. For this reason several studies are conducted to develop tools to help with diagnosis of this disease (CAD: Computer-Assisted Diagnosis). We propose in this paper a new method to segment mammographic images based partly on a pyramidal architecture. The original image is fragmented (quadtree) initially to homogeneous regions. Each region is then associated with a peak of graph. It gathers data within homogeneous groups named regions classes' c, then we use HCA (Hierarchical classification ascendant) and k-means to find the optimal partition for the largest possible value of c at the initial stage. This technique gives good results, and allows calculating morphological parameters of the breast cancer.

Keywords: Mamography, Image Segmentation, K-means, Irregular Pyramid, HCA

1. Introduction

Breast cancer is the most frequent tumor among women. It is the leading cause of female mortality from cancer. In Morocco, between 30 and 45 thousand people are affected each year by the breast cancer according to official information emanating from the 2001 statistics of the National Institute of Oncology Sidi Mohammed Ben Abdellah in Rabat, and seems to be the first female cancer with an incidence of 60 to 90/100 000 women per year. Worldwide, there is 540,000 cases per year and about 300,000 women die. Therefore, it is a real major global public health problem.

Various studies have confirmed that the early detection of sub-clinical cancers can improve prognosis and that mammography; in this case it is the best diagnostic technique. All radiologists recognize the difficulty of the mammogram that increases by the type of tissue examined in the conditions of implementation, the number of available shots,

etc. Radiologists detect around 70% of cancer cases. Due to the increasing cases, and in order to have good diagnostic quality, the automatic assisted diagnosis becomes necessary. In this regard, several studies haddeveloped recent diagnostic tools (CAD: Computer-Assisted Diagnosis) of this disease [6].

2. Materials and Methods

Images obtained by mammography use low-dose X-ray. It consists on producing X-ray beam by a source, and measuring its intensity by a camera after passing through the breast compressed between two plates (figure 1).

This technique is specially designed to visualize breast structures. The breast tissue absorbs X-rays differently according to their texture: dense areas appear in white, fat regions in black and intermediate structures in different gray scales.

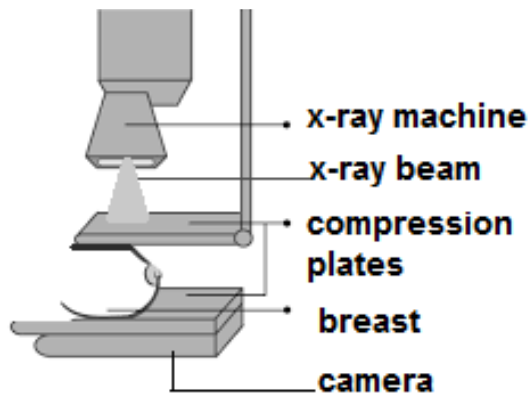


Figure 1. Mammography.

The segmentation stage is very decisive since it will allow the localization of the micro classification. We propose a new segmentation approach that uses the pyramidal organization of regions with two classification algorithms k-means and HCA.

The algorithm of HCA use in its first steps a most expensive routine in time, since they require a large number of distance calculations. For large data sets, the calculation time of this algorithm may become prohibitive. We use to reduce the number of initial regions the K-means, then we apply the HCA on the classes remained.

2.1. Pyramidal Organization

The pyramid [9][10] can be seen as an iterative processing of a level relative to its nearest lower level. Setting the pyramid regions amounts to create a 1st population of regions constitute the 1st level and then iteratively the evolution of this population regions is established of a level N to level N+1. Finally, we specify the stop condition in order to stabilize the population. The pyramidal treatment is divided into three key stages: firstly, the original image is fragmented (quadtree) in homogeneous regions. Secondly, each region R_i is associated with a class C_i in the initial level of the pyramid regions formed after the classification of regions. Finally, each region will make the best possible merger with the regions surrounding it.

2.2. Segmentation by Unsupervised Classification

The segmentation process is to create an image containing the partition of regions according to a specific criterion: color, texture, gray level, etc. A good segmentation helps the radiologist to make the right decision before the intervention.

Segmentation methods based on classification process are seen as a subdivision of the image into different classes so that: the elements of the same class are as similar as possible (minimizing the intra-class variance), and the elements of two separate classes are as different as possible (maximization of inter-class variance). Two types of methods can be distinguished: supervised and unsupervised methods (automatic).

For mammographic image segmentation, supervised classification of the images requires the creation of a learning

base for each class and for a large population of patients; it's a very tedious task for experts [6]. For this reason, our study is focused on unsupervised methods.

2.3. K-means Classification

K-means [4] is one of the learning algorithms without supervision, easy to implement and solves many problems of classification. It classifies objects based on attributes or characteristics in K class number. Subjects within each class are as close as possible to each other, and as far as possible from objects of other classes. Each class of the partition is defined by its objects and its centroid.

The algorithm is composed of the following steps:

Firstly, Place K points into the space represented by the region that is being clustered. These points represent initial group centroids. Secondly, assign each region to the group that has the closest centroid. Thirdly, when all regions are assigned, recalculate the positions of the K centroids. The last two steps are repeated until the centroids become stable.

2.3.1. Hierarchical Cluster Analysis (HCA)

It consists on providing a reduced set of classes obtained by successive combination of similar parts [5](figure 2).

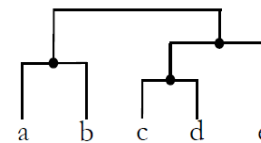


Figure 2. dendrogram resulting from hierarchical cluster analysis (HCA) of 5 Individuals.

In the initial step, the n individuals are obtained by the k-means algorithm. Then the HCA algorithm calculates distances in pairs between individuals, and the two closest ones are merged in a same class. The same process is then applied to the new class and n-2 remaining individuals.

This process is repeated until there is only a single class containing all individuals. It is noted that the novelty here comes from the need to set two distances: the usual distance between two individuals, and a distance between classes. The choice of a distance between individuals, that it has been discussed already, remains to determinate a distance between classes, taking into account that our goal is to find the partition into K classes of observations that minimizes intra-class inertia.

Median object

Each class is characterized by its median.

A x_0 subject is called a median if:

$$\sum_{y \in C_i} d(x_0, y) = \min\{\sum_{y \in C_i} d(x, y)\} \quad (1)$$

d is the distance function.

2.3.2. Topographical Distance

All distances that we use in our method are topographical.

The topographical distance $d_T(p, q)$ between two regions p

and q is the minimum cost path connecting p to q , as follows:

$$dT(p,q)=\min(w(Ppq)) \quad \forall Ppq \quad (2)$$

$W(P_{pq})$ is the topographic path cost pq .

$$W(P_{pq})=\sum_{i=1}^m (kd(p_{i-1},p_i)+|f(p_{i-1})-f(p_i)|),$$

$$P_{pq}=\{p_i\} \quad 1 \leq i \leq m$$

$d(p_{i-1}, p_i)$ is the Euclidean distance dE . K is the weighting factor (usually $k = 1$) [11].

2.3.3. Entropy

Entropy ($E(p)$) is the function that we use to choose the best group. A low value of entropy corresponds to a better clustering [3].

$E(p)$ is defined as:

$$E(p)=-\frac{1}{n} \sum_{j=1}^N \sum_{i=1}^c U'_{ij} \log(U'_{ij})$$

The degrees of membership (U'_{ij}) of pixel x_j in class i is calculated as follows [1] :

$$U'_{ij} = \frac{U_{ij}}{\sum_{i=1}^c U_{ij}} \quad , i=1, \dots, c.$$

3. Our Segmentation Strategy

In the following we present the steps of our method:

Step1: image is divided initially by quadtree into homogeneous regions, and these regions are organized in an adjacency graph.

Step 2: the algorithm maxmin [8] is applied to find the initial centroids ($R_1 \dots R_n$) of groups as separate as possible, with n the number of class given a priori. Then the k-means is applied to find all regions of each class.

Step 3: we apply the HCA algorithm on the result of step 2; it iteratively performs the following steps until obtaining a single class:

1. Determination of the median of each class ($M_1 \dots M_k$).
2. Calculation of the semblance matrix MAT between the median M_i classes.
3. Merging two nearest C_i and C_j classes when $d_T(C_i, C_j) = \min(MAT)$.

Step 4: we cut the tree made by HCA when we find the minimum value of entropy.

Step 5: merging the same class regions.

4. Experimentation Results

Our segmentation method was applied on mammographic images as showed in figure 3, the image is initially divided in homogeneous regions in figure 4, the figure 5 shows the result of k-means and figure 6 illustrate the HCA result.

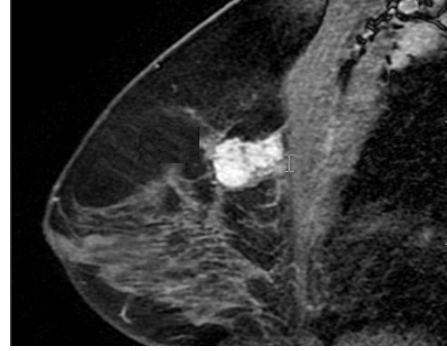


Figure 3. Original image.

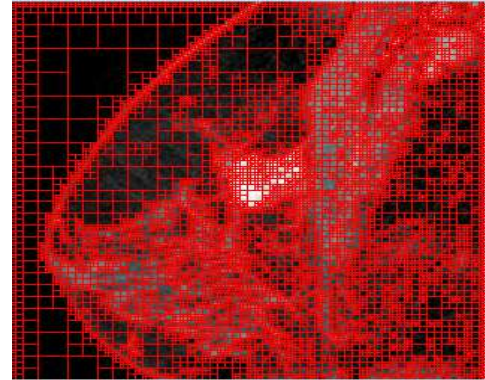


Figure 4. Quadtree result.

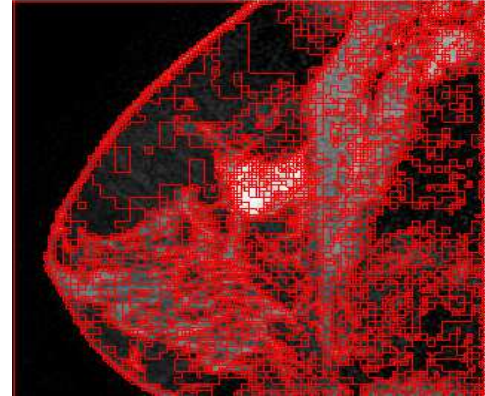


Figure 5. Result of K-means.

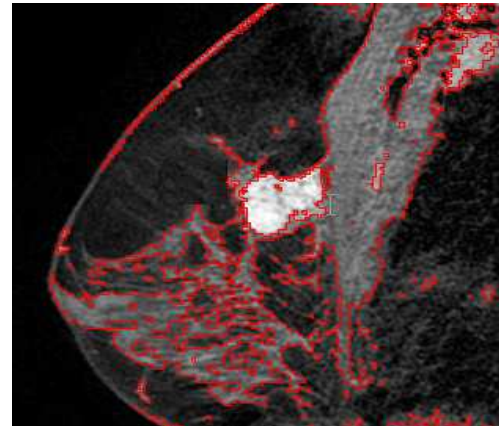


Figure 6. Result of HCA.

We test the approach on offer digital mammography screening images database of South Florida University (<http://marathon.csee.usf.edu/Mammography/Database.html>). This database is provided with each of mammography information about breast density and nature of the lesion

determined by an expert.

Figure7 shows some comparisons of classifications of results between the mass marking the expert represented by the red line and detection of our approach represented by the green line.

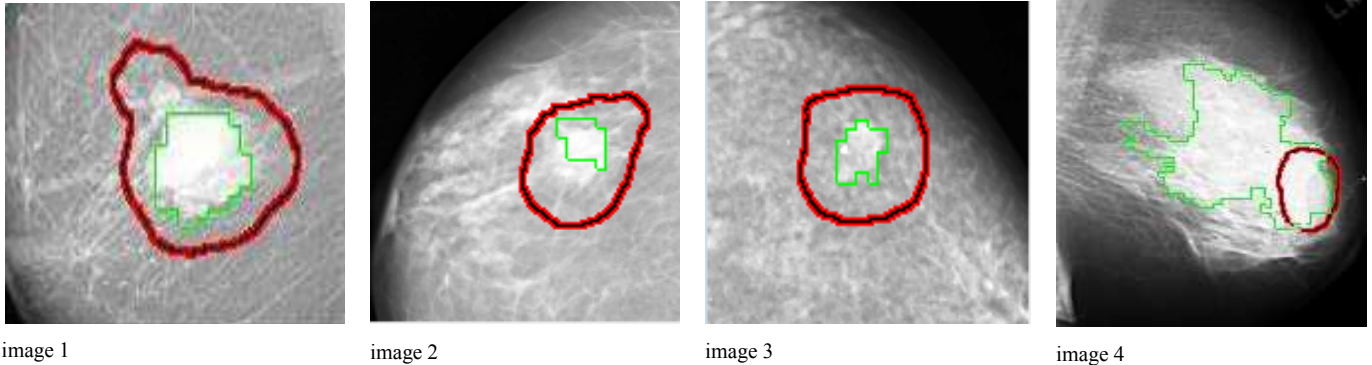


Figure 7. Comparing the shape of the masses detected by the expert (red line) and our approach (green line).

The results of 1, 2 and 3 images show that the lesions are highlighted. Otherwise, the image 4 has false negative detections. The anomalies in the proposed approach reflect that it uses only the intensity property and topographic distance, but specifications are ignored: like shape, size and position. Therefore, this work requires a continuity of adding other geometrical criteria such as the shape or position for a good detection of lesions in a mammogram.

5. Conclusion

In this article, we present a new method for automatic detection of masses in mammograms. This method has two main steps. The first one is the division algorithm, which aims to dividethe image into homogeneous regions. The second one is the use of HCA and K-means algorithm to find the optimal partition after each merging process.

The detection of regions of interest seems to be satisfactory; nevertheless, this study will be continued in order to work on the classification of masses detected in malignant and benign lesions.

References

- [1] Krishna R . "The Possibilistic C-Means Algorithm: Insights and Recommendations". IEEE transactions on Fuzzy Systems 1996.
- [2] Hechmi SHILI*, Lotfi BEN ROMDHANE** et Béchir AYEB*(Une Nouvelle Approche pour la Segmentation des Images Mammographiques. Mars 2012)
- [3] Mizutani and Miyamoto, 2005 Mizutani, K., Miyamoto, S. "Possibilistic Approach to Kernel-Based Fuzzy C-Means Clustering with Entropy Regularization", Lecture Notes on Artificial Intelligence, vol. 3558, pp. 144–155. 2005.
- [4] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297.
- [5] S. C. Johnson (1967): "Hierarchical Clustering Schemes" Psychometrika, 2:241-254.
- [6] Verma, B. "Novel network architecture and learning algorithm for the classification of mass abnormalities in digitized mammograms". In: Artificial Intelligence in Medicine, Vol.42, pp. 67--79 (2008).
- [7] Mizutani and Miyamoto, 2005 Mizutani, K., Miyamoto, S. "Possibilistic Approach to Kernel-Based Fuzzy C-Means Clustering with Entropy Regularization", Lecture Notes on Artificial Intelligence, vol. 3558, pp. 144–155. 2005.
- [8] TouJ.T. et Gonzalez R. « Pattern Recognition Principle », Addison-Wesley, Reading, MA, 1972.
- [9] A. Montanvert, P. Meer, and A. Rosenfeld, "Hierarchical image analysis using irregular tessellations," IEEE Transactions on Pattern Analysis andMachine Intelligence, vol. 13, no. 4, pp. 307–316, April 1991.
- [10] J.M. Jolion and A. Montanvert, "The adaptive pyramid a framework for 2d image analysis," CVGIP Image Understanding, vol. 55, no. 3, pp. 339–348, 1992.
- [11] S. Lefèvre. "A new approach for unsupervised classification in image segmentation". Advances in Knowledge Discovery and Management, Studies in Computational Intelligence. Springer-Verlag, 2009.