

INTERPRETATION DES IMAGES DE MAMMOGRAPHIE PAR L'ALGORITHME SEARCH HARMONY

H.S. Rahli, N. Benamrane

Laboratoire Imagerie et Vision Artificielle, Département d'Informatique,
Université des Sciences et de la technologie Mohamed Boudiaf, Oran

Résumé: L'interprétation des images médicales est un des domaines de recherche les plus encourageants, étant donné qu'il offre des facilités pour le diagnostic et les décisions thérapeutiques d'un grand nombre de maladies tel que le cancer. En particulier, on s'intéresse au cancer du sein qui est une maladie la plus redoutée chez la population féminine. Cet article présente une nouvelle approche hybride pour la détection des lésions mammaires basée sur la méthode k-means et la méthode Search-Harmony.

Mots-clés: Interprétation, détection, K-means, Search Harmony, image de mammographie.

Abstract: The interpretation of medical images is one of the research areas most encouraged, as it offers facilities for diagnosis and therapeutic decisions of many diseases such as cancer. In particular, we are interested in breast cancer which is a disease most feared among the female population. This paper presents a new hybrid approach for the detection of breast lesions based on the k-means method and the method Search harmony.

Keywords: Interpretation, detection, K-means, Search Harmony, image of mammography.

INTRODUCTION

Le cancer du sein est l'un des cancers les plus répandus chez les femmes. C'est une tumeur maligne du sein qui se développe au niveau de la glande mammaire (adénocarcinome). Il n'existe pas encore de moyen d'éviter son apparition. Cependant, un cancer du sein dépisté tôt est plus facile à traiter, engendre moins de séquelles et permet de retrouver une bonne qualité de vie après le traitement.

Les radiologues spécialistes détectent autour de 70% de cas de cancer du sein à cause de la difficulté de l'examen mammographique. A ce titre, plusieurs recherches ont été menées ces dernières années afin de développer des outils d'aide au diagnostic (CAD *Computer Assisted Detection*) de cette maladie qui a pour but l'interprétation des images de mammographies qui donne une description précise des anomalies.

Un système d'interprétation d'images numériques peut être divisé en plusieurs étapes :

- une phase de prétraitement qui suit l'acquisition et la numérisation de l'image. Elle consiste essentiel-

lement à éliminer une quantité importante de « bruits ».

- une phase de segmentation qui permet d'isoler les régions qui peuvent contenir des masses. Certains chercheurs [K+05] ont utilisé un nouveau type d'algorithme de seuillage (minimisation de la somme de l'inertie) pour segmenter les régions de masses où la valeur du seuil convenable est déterminée après plusieurs itérations.

- une phase où certains chercheurs [KLN10] présentent une nouvelle approche pour détecter les microcalcifications. Ils ont utilisé l'algorithme de segmentation ligne de partage des eaux basé sur le gradient morphologique multi-échelle.

- une autre étape où il est nécessaire d'identifier le tissu mammaire de manière à extraire des caractéristiques liées à des lésions de microcalcifications et de masse. C'est l'étape de détection de masse par le CAD (computer aided diagnostic). Les auteurs [PPP09] ont utilisé la technique de transformation en ondelettes pour extraire les régions d'intérêt (ROI). La texture est l'une des caractéristiques les plus importantes utilisées pour identifier un objet dans une image. La dépendance de niveaux de gris spatiale (SGLDM) et la dépendance de la région environnante (SRDM), la méthode de déroulement de niveaux de gris (GLRM) et la différence de niveau de gris (GLMD) sont utilisées pour extraire les caractéristiques de l'image segmentée [T+06].

Pour la détection des tumeurs, on peut modéliser ce problème comme un problème de classification, pour ce faire, on doit choisir parmi plusieurs algorithmes et méthodes. La littérature est très riche à ce sujet, des chercheurs ont hybridé les algorithmes génétiques et le système immunitaire pour tester le diagnostic du cancer du sein [4]. L'avantage majeur des systèmes flous est qu'ils favorisent l'interprétation et permettent de fournir ce qui est appelé mesure de gravité qui signifie dans leur cas, le degré de bénignité ou de malignité. C'est là que les algorithmes génétiques (AG) et les systèmes immunitaires adaptatifs (SIA) permettent la production automatique de systèmes flous. Ils ont hybridé les algorithmes génétiques et le système immunitaire pour tester le diagnostic du cancer du sein.

Une approche utilise l'algorithme Support de Vecteur Machine (SVM) pour la détection des masses mammographiques qui ont démontré leur efficacité dans le diagnostic du cancer du sein [WSH09]. Ils ont introduit ce modèle structuré pour déterminer si chaque région mammographique est normale ou cancéreuse en considérant les structures amassées dans l'ensemble d'apprentissage.

Notre approche ici est basée sur trois outils qui sont la matrice de cooccurrence qui est appliquée sur chaque région afin de calculer ses caractéristiques, une hybridation entre la méthode K-means et l'algorithme Search Harmony.

1 L'ALGORITHME DE K-MEANS

K-means est l'un des algorithmes d'apprentissage sans superviser les plus simples qui permettent de résoudre le problème de classification du fait de sa simplicité de mise en œuvre. K-means est un algorithme permettant de classer ou de regrouper des objets en fonction des attributs/caractéristiques en nombre K de classe, dans laquelle les objets à l'intérieur de chaque classe sont aussi proches que possible les uns des autres et aussi loin que possible des objets des autres classes. Chaque classe de la partition est définie par ses objets et son centroïde.

2 ALGORITHME RECHERCHE D'HARMONIE (HARMONY SEARCH ALGORITHM)

En écoutant et en appréciant une pièce de musique classique, on s'est demandé s'il pleut y avoir un lien entre le fait de jouer de la musique et celui de trouver une solution optimale à un problème complexe tel que la conception d'un réseau d'eau potable ou d'un autre problème d'ingénierie moderne ? Ce lien est représenté par un algorithme nommé « Harmony Search Algorithm (HSA) » développé par Geem et al [GKL01].

3 APPROCHE PROPOSEE

Notre approche d'interprétation est basée sur trois outils qui sont la matrice de cooccurrence qui est appliquée sur chaque région afin de calculer ses caractéristiques, une hybridation de la méthode de K-means et la méthode d'optimisation Search Harmony afin de détecter les tumeurs.

3.1 L'extraction des caractéristiques

Dans le but de décrire les régions de l'image, nous nous intéressons à extraire les caractéristiques de chaque région. L'espace de caractéristiques peut être divi-

sé en trois sous-espaces: les caractéristiques d'intensité, les caractéristiques géométriques et les caractéristiques de texture. Il s'agit d'obtenir la matrice de moyennes d'espace du second ordre, appelée matrice de cooccurrence. Cette matrice contient une masse très importante d'informations difficilement manipulable. C'est pour cela qu'elle n'est pas utilisée directement mais à travers des mesures dites indices de textures. Nous étalons dans ce qui suit quelques indices qui peuvent être tirés de ces matrices comme l'homogénéité, le contraste, l'entropie et l'hétérogénéité.

3.2 La détection par la méthode de K-means-harmony Search

La structure de notre approche de détection des tumeurs s'inspire de l'approche du médecin lors de l'examen radiologique. Notre système de détection est basé sur une méthode hybride K-means qui une méthode de classification très connue et une méthode d'optimisation pour optimiser la détection. Dans ce qui suit, nous détaillerons notre approche étape par étape.

a) Initialisation de la mémoire d'harmonie

Elle doit être initialisée de manière aléatoire, chaque ligne de la mémoire correspond à une classe spécifique des régions dans laquelle la valeur de l'élément i dans chaque ligne est choisie au hasard de la distribution uniforme sur l'ensemble {1,2} et indique le numéro de la classe. On attribue n régions aléatoirement à chaque classe.

On a un ensemble de régions

$$R_i = \{R_1, R_2, \dots, R_n\}$$

la matrice

$$M = [a_{ij}] = \begin{cases} 1 & \text{si la } j^{\text{ème}} \text{ région appartient à la } i^{\text{ème}} \text{ classe} \\ 0 & \text{sinon} \end{cases}$$

b) Calcul de la probabilité d'appartenance des régions à chaque classe

On va calculer pour chaque région $R_i = \{R_1, R_2, \dots, R_n\}$, la probabilité d'appartenance de chaque région dans chaque classe (Tumeur/Non Tumeur).

$$\Sigma_i(i=1)^k \cong [((D_{i,max} - D(R_{ij})) - R_{ij}, C_{jk}))$$

$D_{i,max} = \text{[max]}_i (i) \{D(R_{ij}, C_{jk}), R_j\}$ est la région récemment improvisée.

c) Calcul de la fonction de fitness

La fonction de fitness est la distance minimale des centroïdes des classes.

$$f = \frac{\sum_{i=1}^K \left\{ \frac{\sum_{j=1}^n [D(R_j), C_k]}{N_i} \right\}}{K}$$

K : Nombre de classes pour notre cas est égal à 2.

N_i : Nombre de régions dans la classe i ;

$$N_i = \sum_{j=1}^n a_{ij}$$

d) Improviser une nouvelle harmonie et la mise à jour de la HM

On a pris HMCR (harmony memory considering rate) (HMCR=0.95) et PAR (pitch adjusting rate) (PAR = 0.3).

Algorithme de l'improvisation et la mise à jour des HM :

Tant que (i<=nombre de critère maximal) faire :

Si(rand ∈ (0,1)<=HMCR) alors

Choisir une valeur de la HM pour la valeur de niveau de gris

Choisir une valeur de la HM pour la valeur de la surface

Si (rand ∈ (0,1)<=PAR) alors

Ajuster la valeur de niveau de gris

Ajuster la valeur de la surface

$$val_{new} = val_{old} + rand \in (0,1) * (val_{max} - val_{min})$$

Fsi

Else

Choisir une variable aléatoire ∈ (0,1).

$$val_{new} = val_{min} + rand \in (0,1) * (val_{max} - val_{min})$$

Fsi

Fin tant que

On réitère ce processus par le nombre d'itérations, et on calcule à chaque fois ces paramètres, on obtient cette matrice qui contient la valeur de la fonction de fitness, la matrice de M qui contient la probabilité d'appartenance pour chaque région et les valeurs de C1,C2 qui ont été ajustés.

f1	M1	C1	C2
f2	M2	C1	C2
.	.	.	.
.	.	.	.
.	.	.	.
f30	M30	C1	C2

A la fin de ce processus on prend la valeur minimale de f, et on affiche les régions de la matrice M qui contient les régions appartenant à la classe 1 (Tumeur).

4 RESULTATS

Dans la figure ci-dessous, s'affichent les caractéristiques morphologiques et de texture des régions issues de la phase de segmentation.

Dans la figure 2 s'affichent les résultats de la détection des tumeurs en utilisant une hybridation entre la méthode K-means avec la méthode Search Harmony qui est une nouvelle méthode d'optimisation s'inspirant du processus musical. Notre méthode a été appliquée sur des images mammographiques de taille 256*256 prétraitées.

La Region :	La Surface (Pixels) :	Le Niveau Du Gris Moyen :	La Varia
N°: 1 ->	2329 ->	77 ->	207,51
N°: 2 ->	15967 ->	125 ->	42,97
N°: 3 ->	20910 ->	132 ->	39,35
N°: 4 ->	24 ->	156 ->	79,88
N°: 5 ->	2 ->	82 ->	3444,50
N°: 6 ->	13 ->	94 ->	71,62
N°: 7 ->	2 ->	82 ->	3362,00
N°: 8 ->	4 ->	126 ->	1325,00
N°: 9 ->	4 ->	103 ->	902,00
N°: 10 ->	8 ->	90 ->	152,25
N°: 11 ->	1149 ->	36 ->	71,29
N°: 12 ->	6 ->	86 ->	266,50
N°: 13 ->	27 ->	129 ->	47,56
N°: 14 ->	102 ->	135 ->	11,91
N°: 15 ->	2 ->	50 ->	1300,50
N°: 16 ->	2 ->	68 ->	2312,00
N°: 17 ->	11 ->	123 ->	161,36
N°: 18 ->	121 ->	131 ->	23,93
N°: 19 ->	4 ->	102 ->	901,50

	Homogeneite :	Heterogeneite :	Contraste :	ENTROPIE :
> 0,04 ->	109,03 ->	277,03 ->	57,01 ->	
> 0,45 ->	1487,58 ->	418,81 ->	268,44 ->	
> 0,73 ->	3383,03 ->	538,26 ->	436,59 ->	
> 0,65 ->	3383,04 ->	538,94 ->	436,60 ->	
> 0,72 ->	3383,03 ->	538,52 ->	436,60 ->	
> 0,04 ->	112,89 ->	280,18 ->	58,88 ->	
> 0,73 ->	3383,03 ->	538,46 ->	436,60 ->	
> 0,68 ->	3383,04 ->	538,94 ->	436,60 ->	
> 0,94 ->	3363,74 ->	497,55 ->	423,93 ->	
-> 0,04 ->	112,89 ->	280,18 ->	58,88 ->	
-> 0,06 ->	89,14 ->	163,56 ->	46,16 ->	
-> 0,06 ->	137,71 ->	306,70 ->	71,24 ->	
-> 0,94 ->	3376,96 ->	523,37 ->	432,04 ->	
-> 0,93 ->	3384,14 ->	529,20 ->	435,06 ->	
-> 0,06 ->	135,55 ->	304,96 ->	70,12 ->	
-> 0,93 ->	3319,78 ->	496,28 ->	412,43 ->	
-> 0,95 ->	3358,13 ->	507,68 ->	422,57 ->	
-> 0,92 ->	3392,82 ->	532,39 ->	436,53 ->	
-> 0,94 ->	3351,91 ->	502,12 ->	418,83 ->	

Fig. 1. Caractéristiques morphologiques et de texture

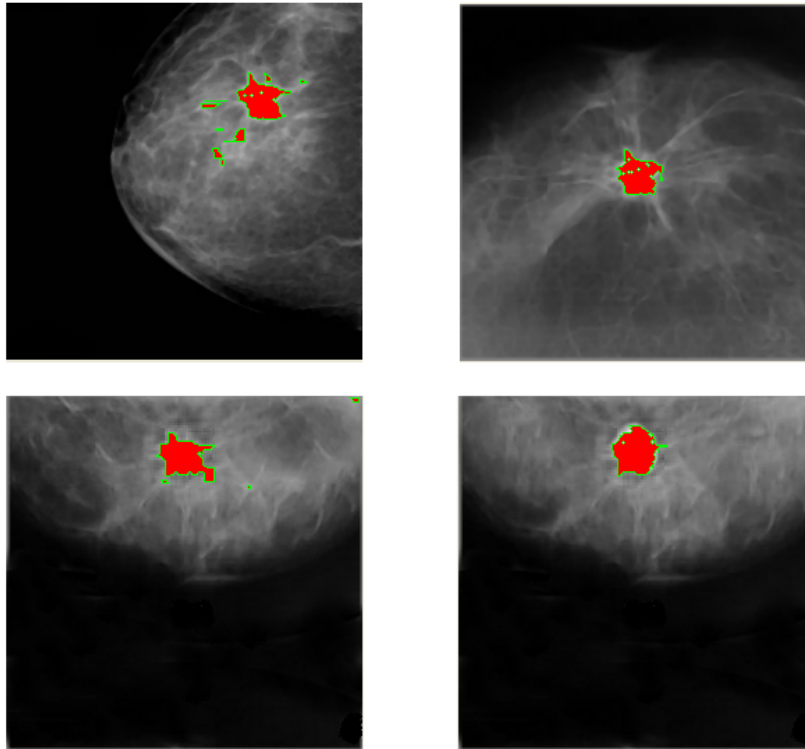


Fig. 2. Module de détection

CONCLUSION

Dans cet article nous avons proposé un système d'interprétation des images de mammographie qui permet de détecter l'anomalie. Ce système combine plusieurs méthodes « K-moyenne, Search Harmony ». L'originalité de notre méthode réside dans l'utilisation des méthodes Search Harmony et K-means qui n'ont jamais été utilisées pour la détection des zones suspectes. Les résultats que nous avons obtenus sont très intéressants et cette approche peut être améliorée en utilisant d'autres caractéristiques des régions pour la phase de segmentation et de détection des lésions mammaires.

REFERENCES BIBLIOGRAPHIQUES

- [BB09] **L. Belkhodja, N. Benamrane** - *Approche d'Extraction de la Région Globale d'Intérêt et Suppression des Artefacts Radiopaques dans une Image Mammographique*, Université de Biskra, 5^{ème} Symposium International Images Multimédias Applications Graphiques et Environnements (IMAGE'09), 04 Novembre 2009, Algérie.
- [C+03] **H. D. Cheng, Xiaopeng Cai, Xiaowei Chen, Liming Hu, Xueling Lou** - *Computer-aided detection and classification of microcalcifications in mammograms: a survey*, Pattern Recognition, Volume 36, issue 12, Pages 2967-2991, December 2003.
- [DJ10] **J. Dheeba, G. Wiselin Jiji** - *Detection of Microcalcification Clusters in Mammograms using Neural Network*, International Journal of Advanced Science and Technology, Vol. 19, June 2010.
- [GKL01] **Z. W. Geem, J. H. Kim, G. V. Logathan** - *A new heuristic optimisation algorithm: Harmony Search*, Simulation 76, pp.60-68, 2001.
- [KLN10] **S. Vijaya Kumar, M. Naveen Lazarus, C. Nagaraju** - *A Novel Method for the Detection of Microcalcifications Based on Multi-scale Morphological Gradient Watershed Segmentation Algorithm*, International Journal of Engineering Science and Technology, Volume 2, Issue 7, Pages 2616-2622, 2010.
- [K+05] **G. Kom, A. Tiedeu, M. Kom, C. Nguemgne, J. Gonsu** - *Détection automatique des opacités dans les mammographies par la méthode de minimisation de la somme de l'inertie*, ITBM-RBM, Vol. 26, Issues 5-6, Octobre-Décembre 2005, Pp. 347-356.

- [N+08] **Emad Nabil, Amr Badr, Ibrahim Farag, Mohamed Ossama** - *A Hybrid Artificial Immune Genetic Algorithm with Fuzzy Rules for Breast Cancer Diagnosis*, INFOS2008, March 27-29, 2008 Caire Egypte.
- [PPP09] **Rabi Narayan Panda, Bijay Ketan Panigrahi, Manas Ranjan Patro** - *Feature Extraction for Classification of Microcalcifications and Mass Lesions in Mammograms*, International Journal of Computer Science and Network Security (IJCSNS), Volume 9, No 5, May 2009.
- [T+06] **K. Thangavel, M. Karnan, R. Sivakumar, A. Kaja Mohideen** - *Ant Colony System for Segmentation and Classification of Microcalcification in Mammograms*, International Journal on Graphics Vision and Image Processing, 2006.
- [VMK09] **Brijesh Verma, Peter McLeod, Alan Klevansk** - *A novel soft cluster neural network for the classification of suspicious areas in digital mammograms*, Pattern Recognition, Volume 42, Issue 9, Pages 1845-1852, September 2009.
- [WSH09] **Defeng Wang, Lin Shi, Pheng Ann Heng** - *Automatic detection of breast cancers in mammograms using structured support vector machines*, Neurocomputing, Volume 72, issue 13-15, Pages 3296-3302, August 2009.