

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique

Université de Batna
Faculté des Sciences
Département d'Informatique

Thèse

En vue de l'obtention du diplôme de
Doctorat en Sciences en Informatique

Détection Multicritères des Anomalies sur des
Mammographies.

Présentée Par
Khamsa DJAROUDIB
Soutenue le 23 octobre 2016

Devant le jury composé de :

<i>Président</i>	<i>Azeddine BILAMI</i>	<i>Prof., Université de Batna</i>
<i>Rapporteur</i>	<i>Abdelmadjid. ZIDANI</i>	<i>Prof., Université de Batna</i>
<i>Co Rapporteur</i>	<i>Abdelmalik TALEB.AHMED</i>	<i>Prof., Université de Valenciennes</i>
<i>Examineurs</i>	<i>Mohamed BENMOHAMED</i> <i>Elkamel MERAH</i> <i>Samir ZIDAT</i>	<i>Prof., Université de Constantine-2</i> <i>Dr., Université de Khenchela</i> <i>Dr., Université de Batna-2</i>

Dédicaces

A ma mère et à ma fille.

Remerciements

Mes premiers remerciements s'adressent à mon encadreur Mr Abdelmadjid ZIDANI, Professeur à l'université de Batna qui a accepté de diriger ce travail scientifique. Je lui exprime ma grande et profonde reconnaissance de m'avoir correctement dirigée aussi bien scientifiquement qu'humainement. Je souhaite lui dire merci pour sa disponibilité, ses conseils judicieux et sa patience.

Je ne cesserai de remercier mon co-encadreur Mr Abdelmalik TALEB AHMED, Professeur à l'université de Valenciennes qui m'a permis de poursuivre mes travaux de recherche dans son Laboratoire. Ces conseils m'ont permis d'avoir confiance et de persévérer dans mes idées tout en me dirigeant de manière efficace. J'aimerais lui dire que c'est grâce au séjour passé dans son laboratoire que j'ai pu réellement avancer dans ce travail de thèse, même si la tâche n'a pas du tout été facile et évidente au départ, et jusqu'à la fin du séjour.

J'exprime ma sincère gratitude à Mr Azeddine BILAMI, Professeur à l'université de Batna d'avoir accepté de présider et juger ce travail de recherche. Je le remercie pour ses conseils et sa disponibilité pour toutes mes questions.

Je remercie aussi Mr Mohamed BENMOHAMED, Professeur à l'université de Constantine et Mr Elkamel MERAH, Docteur à l'université de Khenchla d'avoir accepté l'invitation de mon université pour juger cette présentation de thèse. Je reste à l'écoute de tous leurs conseils.

Enfin, je remercie Mr Samir ZIDAT, Docteur à l'université de Batna, de m'avoir encouragée lors de ma première présentation d'un papier dans une conférence internationale. Des mots simples mais qui m'ont donnée beaucoup d'énergie dans la suite de mon travail de recherche. Merci aussi d'avoir accepté d'être parmi les membres de jury de la soutenance de cette thèse.

Résumé

DETECTION MULTICRITERES DES ANOMALIES SUR DES IMAGES DE MAMMOGRAPHIES

La vision par ordinateur est devenue un outil incontournable pour l'aide au diagnostic des experts médicaux. En particulier, les systèmes automatiques d'aide au diagnostic du cancer du sein (CAD) deviennent de plus en plus utilisés. L'étude menée sur les pathologies mammaires dans des images de mammographie, a montré que le diagnostic des masses mammaires et leur classification suscitent actuellement un grand intérêt. En effet, la complexité, la diversité des formes traitées et la difficulté rencontrée afin de les discerner nécessitent l'usage de descripteurs appropriés. Une revue des différentes méthodes de détection et/ou de segmentation des masses, nous a permis de distinguer tout d'abord entre les différents schémas des systèmes d'aide à la détection (CADE) et des systèmes d'aide au diagnostic (CADx) des anomalies. Pour nous orienter ensuite vers un premier modèle qui introduit le descripteur de texture extrait des matrices de cooccurrence de niveaux de gris (Grey Levels Cooccurrence Matrix, GLCM) dans une approche de détection par contours. Alors que le plus souvent, ces matrices sont utilisées seulement pour l'extraction des caractéristiques de texture dans les phases de segmentation ou de classification des masses. Le deuxième modèle proposé introduit les Modèles de Mixtures des Gaussiennes (GMM), l'Espérance-Maximisation (EM) et la méthode du Maximum de vraisemblance (ML) dans une approche de segmentation par régions des anomalies de cancer, c'est le modèle GMM/ML. L'algorithme consiste à initialiser les paramètres pour l'algorithme itératif EM par des statistiques sur le tissu suspect et des statistiques sur le tissu normal. Ce qui a permis d'ajuster au mieux les paramètres d'initialisation du modèle GMM classique. En plus de ces deux modèles proposés, et dans la continuité de ce travail, une contribution a été apportée aux problèmes d'amélioration des images de mammographie. Il s'agit d'extraire la région mammaire en coupant la région du muscle pectorale et la région bruitée du fond de l'image. L'extraction de ces deux dernières régions est une étape importante dans le processus global des systèmes d'aide au diagnostic CADx ou d'aide à la détection CADE. L'approche GLCM a été utilisée pour la détection des frontières du muscle pectorale et de la région mammaire et leur extraction. Les approches proposées sont à chaque fois testées et validées sur des images issues de mammographies de la base de données BIRADS MIAS. Et pour donner des résultats plus significatifs à cette étude, des tests ont été réalisés sur des images représentant différentes densités de tissu (Gras, Dense et Glandulaire). Enfin, une évaluation et une comparaison de ces approches avec des méthodes récentes de la littérature ont été réalisées.

Mots-clefs: Cancer du sein, Systèmes d'Aide à la Détection (CADE), Systèmes d'Aide au diagnostic (CADx), Détection, Contours, Segmentation, masse, muscle pectoral, région mammaire, Descripteurs de texture, Grey Levels Cooccurrence Matrix (GLCM), Modèle de Mixtures des Gaussiennes (GMM).

Abstract

MULTICRITERIA DETECTION OF THE ABNORMALITIES ON MAMMOGRAPHIC IMAGES

The computer vision became a major tool for the help to the diagnosis of the medical experts. In particular, the automatic Computer Aided Diagnosis (CAD) of the breast cancer, become more and more used. The study on the pathologies mammary, in mammographic images showed that the diagnosis of the mammary masses and their classification arouse at present a big interest. Indeed, the complexity, the diversity of the handled forms and the difficulty met to discern them require the use of appropriate descriptors. A review of the various methods of detection and/or segmentation of the masses, allowed us to distinguish first of all between the Computer Aided Detection systems (CADE) and the Computer Aided Diagnosis systems (CADx) of the abnormalities. Then we propose a first model which introduces the texture descriptor extracted from the Grey Levels Co-occurrence Matrix (GLCM) in a contour detection approach. While most of the time, these matrix are only used for the extraction of texture characteristics in a mass segmentation step or a mass classification step. The second proposed model introduces the Gaussian Mixtures Models (GMM), Expectation-Maximization (EM) method and Maximum likelihood method (ML) in an approach of segmentation of abnormality, by region. It is the GMM/ML model. The algorithm consists in initializing the parameters for the iterative algorithm EM by statistics on the suspicious tissue and statistics on the normal tissue. What allowed the best fitting of the parameters of initialization of the classic model GMM. Besides these two proposed models, and in the continuity of this work, a contribution was brought to the problems of improvement of the images of mammography. It is about of extracting the mammary region by cutting the pectoral region of the muscle and the region made noisy effects by the bottom of the image. The extraction of these last two regions is an important stage in the global process of the CADx systems or the CADE systems. The GLCM approach was used to detect the pectoral borders and the mammary region borders and their extraction. The proposed approaches are every time tested and validated on images from the BIRADS MIAS database. And in order to give more significant results to this study, tests were realized on images representing various densities of tissue (Fatty, dense and glandular). Finally, an evaluation and a comparison of these approaches with recent methods of the literature were realized.

Keywords: Breast Cancer, CADE systems, CADx systems, Detection, Contour, Segmentation, mass, pectoral muscle, mammary region, texture Descriptor, Grey Level Co-occurrence Matrix (GLCM).

Remerciements**Résumé****Abstract****Introduction Générale** 6

Chapitre 1 : Définitions et Etat de l'art	9
1.1. Introduction.....	9
1.2. La mammographie.....	9
1.3. Les principales lésions.....	10
1.4. Les Pathologies de masses.....	10
1.4.1. Les formes de la masse.....	11
1.4.2. Les contours de la masse.....	12
1.4.3. Les formes et les contours résumés.....	13
1.4.4. La taille.....	13
1.5. La classification des pathologies mammaires.....	14
1.5.1. Les différentes classifications.....	14
1.5.2. La classification ACR des masses.....	16
1.6. Les systèmes de Détection ou de Diagnostic Assistés par Ordinateur CAD.....	16
1.6.1 Les avantages des systèmes d'aide CAD.....	16
1.6.2 Les CADe et les CADx.....	17
1.6.3 Les méthodes des systèmes d'aide CAD.....	18
1.7. Les algorithmes de détection et/ou segmentation des masses mammaires.....	19
1.7.1 Techniques de détection basées Pixels pour les CADe.....	19
1.7.2. Techniques de détection basées Régions pour les CADe.....	22
1.8. Conclusion et Choix de méthodes.....	24

Chapitre 2 : Intégration des descripteurs de texture extraits des Matrices de Cooccurrence de Niveaux de Gris (GLCM) dans une approche de détection de contours des anomalies de masses	28
--	----

Résumé	28
2.1. Introduction	28
2.2. Définitions de la texture	28
2.2.1. Les difficultés liées à la définition de la texture	28
2.2.2. Définition de la texture selon les domaines d'application	29
2.2.3. Définition de la texture par le dictionnaire	29
2.2.4. Définition de la texture comme un phénomène à deux dimensions	29
2.3. Descripteurs statistiques de la texture.....	30
2.3.1. Descripteurs Simples.....	30
2.3.2. Descripteurs Complexes : Les Matrices de cooccurrence des niveaux de gris	31
2.3.2.1. Inconvénients des descripteurs statistiques simples	31
2.3.2.2. La matrice de cooccurrence de niveaux de gris.....	32
2.3.2.3. Les descripteurs de texture extraits des matrices de cooccurrence	34
2.3.2.3.1. Moments	34
2.3.2.3.2. Moments de la différence inverse	34
2.3.2.3.3. Contraste	34
2.3.2.3.3. Entropie	34
2.3.2.4. Performances des descripteurs de texture des matrices de cooccurrence	34
2.4. La base de données MIAS	35
2.5. Approche texturale proposée pour la détection des masses	35
2.5.1. Position de la Problématique	35
2.5.2. Hypothèses de travail	36
2.5.3. Schéma de l'approche proposée	37
2.5.4. Les étapes de l'approche proposée	37
2.5.4.1 Amélioration du contraste des images	37
2.5.4.2 Détection des contours des masses par le calcul des GLCM.....	45
2.5.4.2.1. Choix de la distance	45
2.5.4.2.2. Choix de la direction	45
2.5.4.2.3. Choix de la taille du voisinage ou du masque	46
2.5.4.2.4. Calcul du descripteur de texture	47

2.5.4.3 Algorithme général de Calcul des images texturales	48
2.5.4.4 Post-traitements	48
2.6. Résultats de l'approche de détection des contours des masses par le Descripteur GLCM	50
2.6.1 Détection et segmentation des contours des masses dans des tissus de différentes densités	50
2.6.2 Evaluation Quantitative de la détection des contours des masses	52
2.6.2.1 Evaluation Quantitative par le Coefficient du Dice	53
2.6.2.1.1. Présentation du coefficient de Dice.....	53
2.6.2.1.2. Evaluation de la détection et de la segmentation des masses par le Dice Coefficient.....	53
2.6.2.2 Evaluation Quantitative par le F-mesure.....	56
2.6.3 Premières comparaisons avec des travaux similaires	58
2.7. Discussion et Conclusion	59

Chapitre 3 : Les Modèles de Mixture des Gaussiennes (GMM) Pour la Segmentation des Anomalies de Masse sur des Images de Mammographie 61

Résumé	61
3.1. Introduction	61
3.2. Les Modèles GMM de base	62
3.2.1. Principe des Modèles GMM	62
3.2.2. Exemple d'un Modèle GMM	62
3.2.3. Principe de l'algorithme d'Espérance-Maximisation	63
3.2.4. Principe de l'algorithme K-moyennes	65
3.2.4.1. Description	66
3.2.4.2. L'algorithme K-means: Algorithme standard.....	66
3.2.4.3. Convergence	67
3.2.4.4. Avantages et inconvénients	67
3.2.5. Estimation du Maximum de Vraisemblance	67
3.2.5.1. Principe	67

3.2.5.2. Généralisation	68
3.2.5.3. Propriétés	69
3.2.5.4. Loi normale	69
3.2.5.5. Fonction de vraisemblance	70
3.3. Le Modèle GMM/ML proposé pour la segmentation des anomalies de masses	70
3.3.1 Les principales étapes du Modèle GMM/ML	71
3.3.2. Amélioration des images de mammographies	72
3.3.3. Extraction des images « ROIs »	72
3.3.4. Apprentissage du Modèle GMM/ML	73
3.3.4.1. Estimation de la densité de la Masse	73
3.3.4.2. Initialisation du paramètre θ_0	74
3.3.4.3. Paramètre de la densité du tissu	75
3.3.5. GMM/ML: Expectation Maximisation (EM).....	75
3.3.5.1. L'algorithme EM	76
3.3.5.2. Initialisation des paramètres de l'algorithme EM	76
3.3.5.3. Les étapes de l'EM	77
3.4. Résultats de la segmentation des masses par le Modèle GMM/ML	78
3.4.1. Segmentation des masses dans le tissu Dense	78
3.4.2. Segmentation des masses dans le tissu Glandulaire	79
3.4.3. Segmentation des masses dans le tissu Gras	80
3.4.4. Evaluation de la segmentation par le coefficient du Dice	81
3.4.5. Comparaison des résultats GMM/ML	82
3.5. Discussion et Conclusion	82
Chapitre 4 : Intégration des descripteurs de texture extraits des Matrices de Cooccurrence de Niveaux de Gris (GLCM) dans une approche de détection du muscle pectoral et du sein	84
Résumé	84

4.1. Introduction	84
4.2. Etat de l'art	85
4.3. Conclusions sur l'état de l'art	85
4.4. Méthodologie: Détection et segmentation de muscle pectoral et de la région mammaire par les GLCM	86
4.4.1. Introduction	86
4.4.2. Schéma de l'approche proposée	87
4.4.3. Les différentes étapes de l'approche texturale GLCM	87
4.4.3.1. Etape 1: Orientation du sein	87
4.4.3.2. Etape 2: Amélioration du contraste des images	88
4.4.3.3. Etape 3: Détection des contours du muscle pectoral et du sein par les GLCM .	90
4.4.3.3.1. La taille du masque	90
4.4.3.3.2. La direction	91
4.4.3.3.3. La distance	91
4.4.3.3.4. Le Descripteur de contraste	91
4.4.3.4. Etape 4: Post-traitements	92
4.5. Résultats de l'approche de détection des contours du muscle pectoral et de la région mammaire par le Descripteur GLCM de Contraste.....	92
4.5.1. Base de données	92
4.5.2. Résultats de la détection du muscle pectoral et du sein	93
4.5.3. Evaluation quantitative des résultats de la détection du muscle pectoral	96
4.6. Discussion et Conclusion	98
 Conclusions Générales et Perspectives	99
Bibliographie	103
Liste des Figures	110
Liste des Tableaux	113

Introduction Générale

La vision par ordinateur est devenue un outil incontournable pour l'aide au diagnostic des experts médicaux. Elle se base sur des fondements mathématiques qui permettent de résoudre des problèmes de correction géométrique des images, de compression des volumes d'images et de reconnaissance de formes qui est essentiellement la segmentation de l'image. La recherche scientifique réunit dans ce domaine les informaticiens, les physiciens et les mathématiciens pour répondre au mieux aux besoins des experts.

L'objectif de cette thèse est la contribution au développement de méthodes de segmentation d'images. Le problème majeur de la segmentation est la reconnaissance d'objets qui sont le plus souvent confondus avec d'autres objets de l'image. Pour les images de mammographie, le souci est que les anomalies cancéreuses sont dans la majorité des cas, semblables aux tissus normaux. En premier lieu, ce problème a été abordé en introduisant l'information de texture extraite des matrices de cooccurrence de niveaux de gris (GLCM) dans une approche de segmentation par contours. Alors que le plus souvent, ces dernières sont utilisées dans une approche par régions, afin d'extraire des caractéristiques de texture et de les utiliser dans les phases de segmentation ou de classification d'un système CAD (Computer Aided Detection/Diagnosis). En deuxième lieu, les Modèles de Mixtures des Gaussiennes et la méthode de Maximum de vraisemblance (Maximum Likelihood, ML), ont été introduits dans une approche de segmentation des anomalies de cancer, c'est le modèle GMM/ML. L'algorithme Espérance-Maximisation (EM) est utilisé pour ajuster au mieux les paramètres du modèle.

Les approches proposées ont été validées dans des problèmes liés à la détection et à la segmentation des anomalies de masses sur des images de mammographie. Les systèmes concernés sont les systèmes d'aide à la détection (CADe) et/ou les systèmes d'aide aux diagnostics (CADx). Dans ce domaine, il était difficile de classer les méthodes utilisées. L'étude a été abordée selon quatre chapitres :

Dans le premier chapitre, nous avons d'abord éclairci la problématique liée à la confusion entre les systèmes d'aide à la détection ou les systèmes d'aide au diagnostic. Un

état de l'art sur les algorithmes des CAD, des CADx et des CADe, a été brièvement donné. Nous avons évoqué le choix des anomalies de masses par rapport aux anomalies de micro-calcifications, et avons classé les travaux liés à la détection et à la segmentation de ces anomalies. Ils ont été répartis selon les méthodes de segmentation, les méthodes de détection et les méthodes de détection et de segmentation. Cette tâche fût fastidieuse car il n'existait pas au préalable une telle classification de ces travaux antérieurs.

Dans le deuxième chapitre, une approche de détection et de segmentation des anomalies de masses a été proposée. L'approche consiste à utiliser l'information spatiale ou texturale des niveaux de gris de l'image et la robustesse des matrices de cooccurrences de niveaux de gris GLCM dans une approche de segmentation par détection des contours de l'anomalie. Alors que dans la majorité des travaux antérieurs, ces matrices sont utilisées pour extraire des caractéristiques de texture et de les introduire comme entrées supplémentaires à l'étape de segmentation ou à l'étape de classification d'un CAD. Dans ces cas, les approches de segmentation sont des approches par régions. Le travail essaye donc d'expérimenter la méthode GLCM dans une approche à base de contours.

Dans le troisième chapitre, nous contribuons et proposons une méthode semi-automatique de segmentation des masses basée sur les Modèles de Mélange des Gaussiennes (GMM), la méthode du Maximum de Vraisemblance (ML) et l'algorithme espérance-maximisation (Expectation-Maximisation, EM), c'est le modèle GMM/ML. Les modèles GMM classiques sont robustes dans le domaine du traitement d'images. Leur force majeure est l'information donnée par des paramètres initiaux à l'approche itérative. Ainsi, la méthode proposée peut être divisée selon deux étapes. La première est l'étape d'initialisation qui consiste à déterminer des paramètres importants pour le modèle. Ceux sont des statistiques sur les anomalies de masse et des statistiques sur le tissu mammaire normal. Dans la deuxième étape, ces statistiques sont incorporées et apprises par l'algorithme d'Expectation-Maximisation, et l'algorithme du maximum de vraisemblance, pour ajuster au mieux les paramètres du modèle GMM classique. Cette approche a permis de réaliser la segmentation des masses.

Dans le quatrième chapitre, une technique d'amélioration des images de mammographie a été proposée, elle entre aussi dans tout le processus d'un système CAD. En effet, toute méthode de détection et/ou segmentation d'une anomalie cancéreuse nécessite au

préalable une préparation des images pour augmenter et améliorer les performances de la méthode développée. L'étape d'amélioration consiste à extraire la région mammaire de l'image globale, en enlevant la région du muscle pectoral et la région du fond de l'image. Ces deux dernières régions qui peuvent gêner les méthodes de détection et/ou de segmentation des anomalies cancéreuses dans leur objectif. Dans cette perspective et dans la continuité de ce travail, nous avons voulu tester l'approche de détection de contours GLCM afin de détecter les frontières du muscle pectoral, celles de sein et ainsi pour extraire la région mammaire cherchée.

Dans cette étude, les modèles proposés sont à chaque fois testés et validés sur des images issues de mammographies de la base de données BIRADS MIAS. Et dans un souci de donner des résultats plus significatifs, nous avons pris des images représentant différentes densités de tissu (gras, dense et glandulaire). Les résultats des modèles ont été comparés et évalués avec des méthodes similaires et récentes de la littérature. Cette comparaison est réalisée par des méthodes classiques existantes.

Chapitre 1 : Définitions et Etat de l'art

1.1. Introduction

Avant de donner les différentes méthodes numériques d'aide à la détection ou au diagnostic des pathologies mammaires, il est utile de donner quelques définitions sur la mammographie et ces pathologies. Et particulièrement sur la pathologie des masses ainsi que leur classification. Par la suite, l'ambiguïté sur la confusion entre les systèmes d'aide à la détection (CADe) et les systèmes d'aide au diagnostic (CADx), a été soulevée et éclaircie. Pour aboutir enfin à un état de l'art sur les méthodes utilisées dans ces systèmes, et finir avec les motivations de nos choix de méthodes.

1.2. La mammographie

La mammographie est l'imagerie radiographique haute résolution du sein compressée. Ceci implique la transmission de radiation par le tissu et la projection de structures anatomiques sur un écran de film ou un capteur d'image. Deux projections d'image de chaque sein sont obtenues, vue Cranio-Caudal (CC) et vue Médio-Latéral Oblique (MLO). La figure (Fig 1.1) est un exemple d'une image MLO, prise da la base de données MIAS.

La mammographie de haute qualité (haute résolution spatiale) perme d'observer de bonnes structures. Les études ont montré que le taux de mortalité pourrait diminuer de 30 % si toutes les femmes dont l'âge est supérieur ou égal à 50 ans, avaient des mammographies régulières (American Cancer Society, 2002).



Fig.1.1. Image mdb002, MLO (MIAS Database).

1.3. Les principales lésions

Les cancers du sein apparaissent d'habitude avec des structures canalaire (ductal structures) déformées. Il y a trois types majeurs de cancer du sein : les masses circonscrites/ovales, les lésions spiculées ou distorsions architecturales et les micro-calcifications. Les lésions malignes ont généralement une forme plus irrégulière, ou mal définie, que les lésions bénignes.

1.4. Les Pathologies de masses

Nous nous intéressons à la pathologie mammaire de masse, appelée aussi opacité qui est une lésion importante du sein. Sa disposition dans ce dernier varie et peut être vue sur deux incidences différentes. Si une opacité potentielle est vue seulement sur une seule incidence, alors elle est appelée asymétrie jusqu'à ce que son caractère tridimensionnel soit confirmé. Les différentes caractéristiques de ces masses sont à décrire à savoir la forme, le contour et la taille.

Ainsi, les masses sont caractérisées par leur emplacement, leur taille, leur forme et leur contour (Sampat, Markey, & Bovik, 2005(a)), (Naga Mudigonda et al., 2001). Nous avons choisi de contribuer à résoudre les problèmes liés à la détection et/ou la segmentation des anomalies de masses car il s'avère que jusqu'à nos jours, c'est une tâche qui reste plus difficile pour les chercheurs que la détection et/ou la segmentation des micro-calcifications.

En effet, la grande variation dans les paramètres de la taille et la forme, en laquelle la masse peut apparaître sur une image issue de mammographie, fait que sa détection et/ou sa segmentation par des méthodes numériques reste une tâche délicate. La méthode à développer doit prendre en considération tous ces paramètres.

1.4.1. Les formes de la masse

Selon la description du BIRADS (D'Orsi CJ, 2003), les masses mammaires peuvent avoir la forme ronde ovale lobulée ou irrégulière la figure (Fig.1.2) donne des exemples sur ces formes.

- a) Ronde : Il s'agit de masse sphérique, circulaire ou globuleuse.
- b) Ovale : Elle présente une forme elliptique (ou en forme d'un œuf).
- c) Lobulée : La forme de la masse présente une légère ondulation.
- d) Irrégulière : Cette appellation est réservée aux masses dont la forme est aléatoire Elle ne peut être caractérisée par les termes cités ci-dessus.

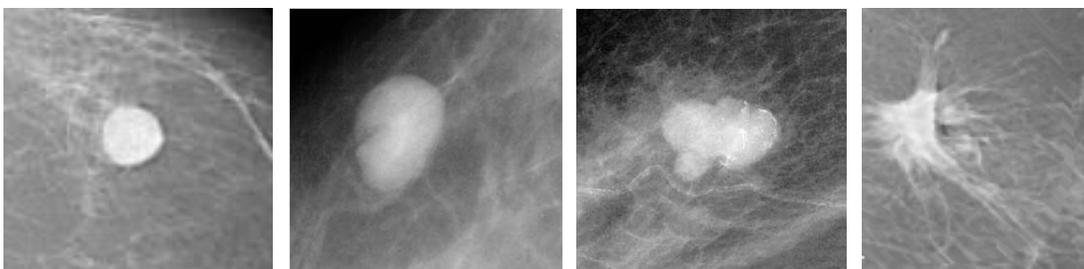


Fig.1.2. Les différentes formes possibles d'une masse (de gauche à droite) : a) Ronde, b) Ovale, c) Lobulée et d) Irrégulière

1.4.2. Les contours de la masse

Le contour des masses mammaires est soit circonscrit, soit micro-lobulé, masqué, indistinct, ou encore spiculé, la figure (Fig.1.3) présente quelques exemples de ces contours. On détaille dans ce qui suit ces différentes notions :

a) Circonscrit : Il s'agit d'une transition brusque entre la lésion et le tissu environnant. Le contour est alors net et bien défini. Pour qu'une masse soit qualifiée de circonscrite, il faut qu'au moins 75% de son contour soit nettement délimité.

b) Micro-lobulé : Dans ce cas, de courtes dentelures du contour créent de petites ondulations.

c) Masqué : Un contour masqué est un contour qui est caché par le tissu normal adjacent.

Ce terme est employé pour caractériser une masse circonscrite dont une partie du contour est cachée.

d) Indistinct : dans ce cas, le contour est mal défini. Ce caractère indistinct (le contraire de circonscrit) peut correspondre à une infiltration.

e) Spiculé : La masse est caractérisée par des lignes radiaires prenant naissance sur le contour de la masse. Ces lignes radiaires sont appelées les spicules.

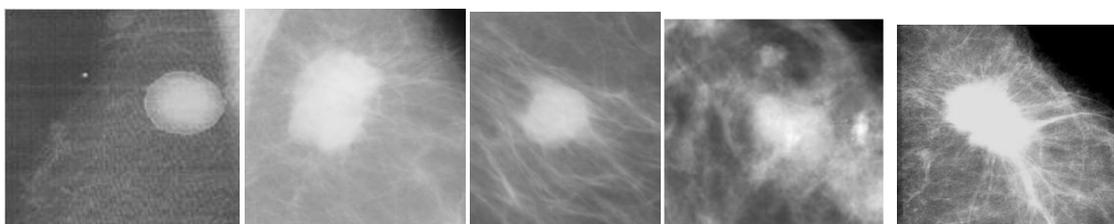


Fig.1.3. Les différents contours possibles d'une masse (de gauche à droite) : a) Circonscrit, b) Micro-lobulé, c) Masqué, d) Indistinct et e) Spiculé.

1.4.3. Les formes et les contours résumés

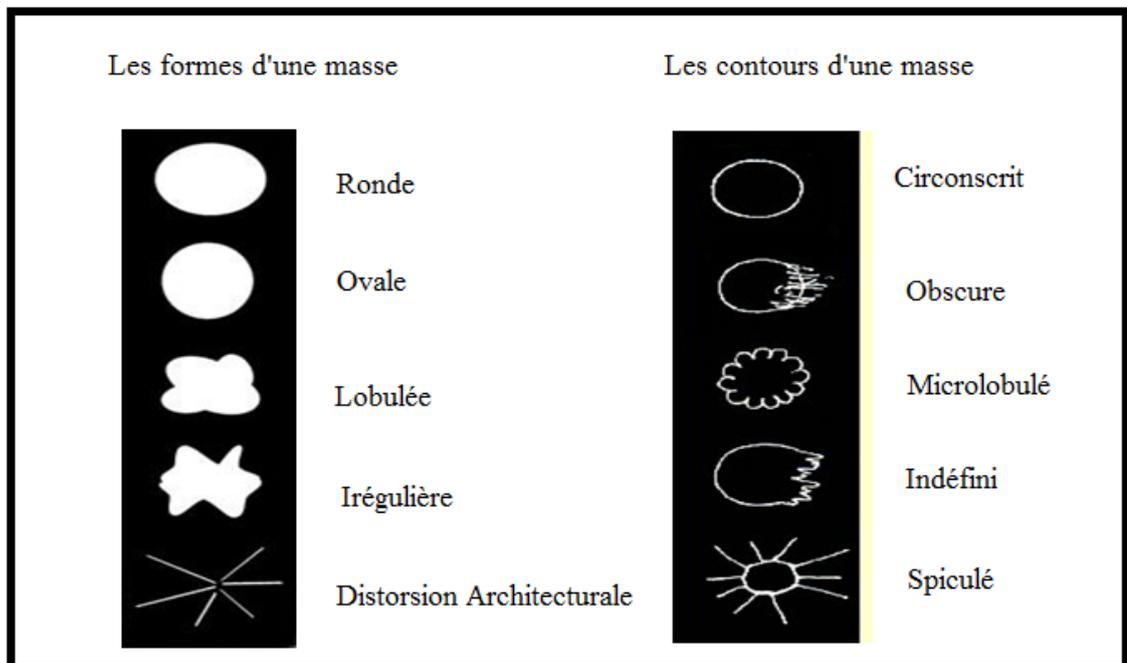


Fig.1.4. Les formes et les contours résumés

1.4.4. La taille :

La taille des masses est donnée par la classification TNM. Cette dernière est un système international qui donne un classement des cancers selon leur extension anatomique. Les trois lettres symbolisent la propagation de la maladie cancéreuse sur le site de la tumeur primitive (**T**), dans les ganglions lymphatiques voisins (**N** pour *node* en anglais) et à distance pour d'éventuelles métastases (**M**). Chaque lettre est affectée d'un coefficient (Classification_TNM, 2015).

Le tableau suivant donne un exemple d'informations sur les tailles des tumeurs (Cabarro.E, 2000):

Classification TNM	Description
Tx	Tumeur ne pouvant être appréciée par défaut de renseignements
T0	Pas d'évidence de tumeur primitive
Tis	Carcinome canalaire in situ ou carcinome lobulaire in situ ou maladie de Paget du mamelon sans tumeur sous-jacente
T1 T1a T1b T1c	Tumeur inférieure à 2 cm dans sa plus grande dimension - T < 0,5 cm - T entre 0,5 et 1 cm - T entre 1 et 2 cm
T2	Tumeur supérieure à 2 cm et inférieure à 5 cm dans sa plus grande dimension
T3	Tumeur de plus de 5 cm dans sa plus grande dimension
T4 T4a T4b T4c T4d	Tumeur de toute taille avec extension : - à la paroi thoracique - œdème ou ulcération de la peau ou nodule de perméation - à la peau et à la paroi thoracique - Cancer inflammatoire

Tab.1.1. Exemple de Classification TNM.

1.5. La classification des pathologies mammaires

1.5.1. Les différentes classifications

Il est important d'adopter un lexique standard et une classification commune afin de mettre à la disposition des radiologues une description claire et précise des lésions mammaires. L'étude morphologique de ces lésions a fait l'objet de plusieurs classifications à savoir la classification de LeGal (LeGal, M et al., 1984), de Lanyi (Lanyi, 1985) et de BIRADS (D'Orsi CJ, 2003). Les classifications les plus connues et les plus pratiquées sont celles de Le Gal et De BIRADS.

Pour palier aux lacunes de la classification de Le Gal, l'American College of Radiology (ACR) a souligné l'importance d'un protocole mammographique standardisé et complet qui prend en compte les différents facteurs de malignité. En novembre 1998, l'ACR a établi le système Américain BIRADS (Breast Imaging Reporting and Data System) qui a été rédigé par un groupe d'experts réunis par l'ANAES2 (Feig, S. A. et al., 1998) (D'Orsi CJ, 2003). Ce système permet de classer les images mammographiques en six catégories en fonction du degré de suspicion de leur caractère pathologique (Tab.1.2 et Tab.1.3). La classification de l'ACR résume les formes des différentes masses et des différentes micro-calcifications, les différents aspects de la distorsion architecturale ainsi que leur degré de malignité. Une fois que le radiologue arrive à reconnaître la catégorie d'une mammographie,

il sait automatiquement les directives et les recommandations associées à cette classe (Tab.1.2). Enfin la pratique a démontré que l'utilisation de ce système permet d'augmenter le taux de reconnaissance des masses malignes et celles bénignes.

Catégorie	Diagnostic	Critères
0	Incomplet	La mammographie ou les ultrasons n'ont pas donné au radiologue assez d'informations pour faire un diagnostic clair; Le suivi de l'image est nécessaire.
1	Négatif	Il n'y a rien pour faire des remarques; examen de routine recommandé.
2	Benin	Une découverte bénigne définie; examen de routine recommandé.
3	Probabilité Benin	Les découvertes à haute probabilité d'être bénignes (> 98 %); suivi d'intervalle court de six mois.
4	Anomalie suspecte	Pas de caractéristique de cancer du sein, mais une probabilité maligne (3 à 94 %); considérer la biopsie.
5	Fortement Suspecte à Maligne	La lésion à haute probabilité d'être maligne (> = 95 %); prendre l'action appropriée.
6	Biopsie Connue Malignité Prouvée.	Lésions connues pour être malignes; s'assurer que le traitement est complété.

Tab.1.2. Catégories et Diagnostiques

1.5.2. La classification ACR des masses

ACR 0	Cliché insuffisant pour vérifier la présence ou non de lésions
ACR 1	Absence de masses suspectes.
ACR 2	-Opacité ronde avec macro-calcifications - Ganglion intra mammaire. - Images de densité graisseuse ou mixte. - Opacités rondes correspondant à un kyste en échographie.
ACR 3	Masse circonscrite ronde ou ovale sans micro-lobulations non liquidienne en échographie
ACR 4	- Masse spiculée sans centre dense. - Masse ronde ou ovale à contour lobulé ou masqué.
ACR 5	- Masse spiculée à centre dense. - Masse à contour flou et irrégulier.

Tab.1.3 Classification BIRADS de l'ACR correspondant à la masse

1.6. Les systèmes de Détection ou de Diagnostic Assistés par Ordinateur CAD

1.6.1 Les avantages des systèmes d'aide CAD

Le but des systèmes de détection assistés par ordinateur est d'aider les radiologues à éviter d'oublier les caractéristiques anormales qui ne sont pas visibles sur l'image de dépistage (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013). Pour ces systèmes, les anomalies de masses restent plus difficiles à détecter que celles des micro-calcifications. La première raison est que les masses présentent une grande variabilité dans leur taille, leur forme et leur contour. Les chercheurs doivent alors envisager toutes ces caractéristiques pour le développement de leurs algorithmes. La deuxième raison est que les anomalies de masses sont le plus souvent confondues avec le tissu mammaire normal. Les CAD restent un outil efficace pour le radiologue (Zheng, B. et al., 2006), (Caroline & Vaijyanthi N, 2012),

(Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013), pour donner un avis supplémentaire et plus fiable pour le diagnostic.

1.6.2 Les CADe et les CADx

La détection et/ou la segmentation est la première étape dans le processus complet des systèmes CAD (Arnau, et al., 2010) (Etler & Horsch, 2009). C'est la fonction clé de ces systèmes.

Cependant, quelques confusions existaient dans la littérature quant à la définition de la fonction exacte des CAD, ils ont donc été classifiés récemment en deux catégories (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013):

- les systèmes CADe, ou d'aide à la détection automatisée des anomalies dans des images de mammographie (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013). Le but des CADe est d'aider les radiologues à détecter et à localiser des anomalies dans des images de dépistage du cancer du sein. Pour les anomalies de masses, il y aura deux étapes: la détection puis la classification en tissu normal ou tissu avec lésion.

- les systèmes CADx, ou d'aide au diagnostic automatisé des anomalies, comme étant bénignes ou malignes (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013). Le but des CADx est d'aider les radiologues à caractériser la lésion, de donner un diagnostic final et de réduire ainsi le nombre de recommandations de biopsies sur des lésions bénignes. Pour les anomalies de masses, il y aura deux étapes: la segmentation de l'anomalie puis sa classification en masse bénigne ou en masse maligne.

Seulement, pour les CADe, nous remarquerons que dans la définition des méthodes de détection, il y a des méthodes qui détectent et qui segmentent les anomalies de masses. Donc on dira qu'il y a détection et/ou segmentation de ces anomalies. Nous avons résumé ces définitions dans la fig.1.5.

1.6.3 Les méthodes des systèmes d'aide CAD

Toutes les méthodes développées dans la littérature pour les systèmes CADE et les systèmes CADx, ont été résumées par les auteurs dans (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013). La figure Fig.1.5 est un résumé schématisé pour éclaircir la différence entre ces deux systèmes.

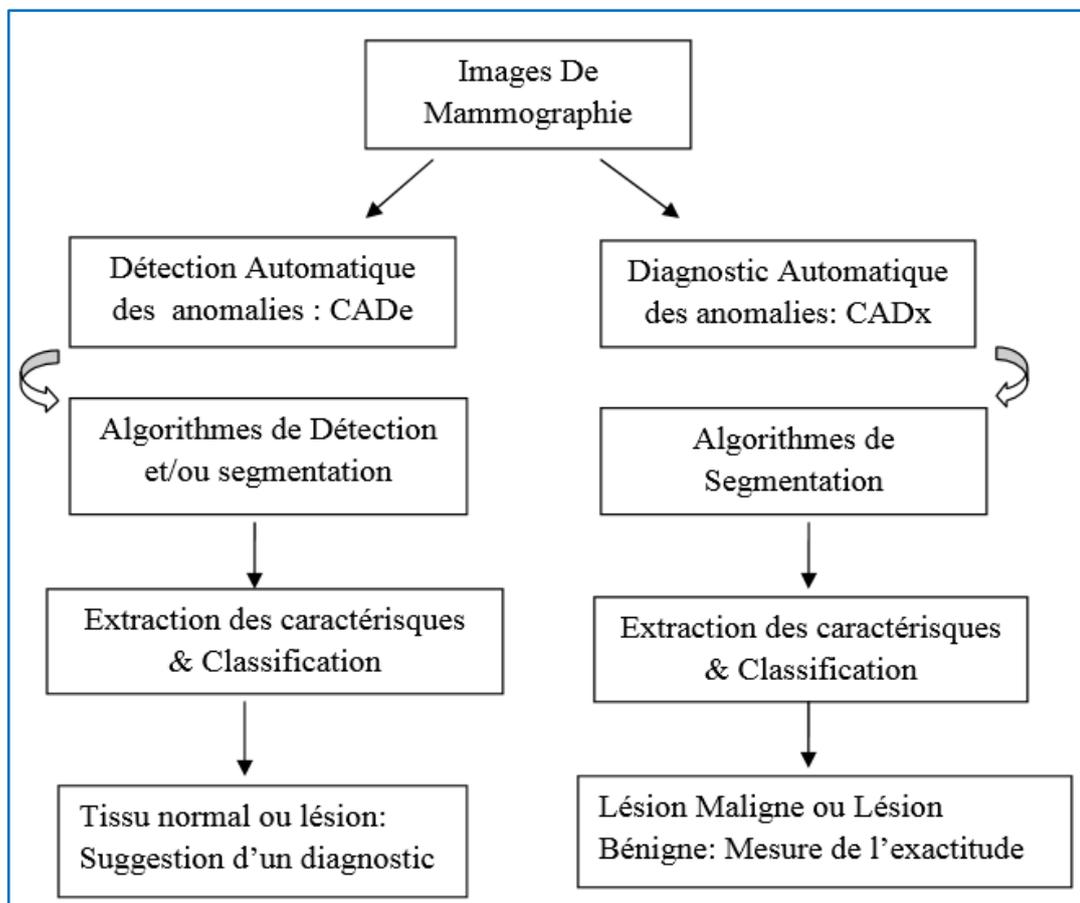


Fig.1.5. Les différents systèmes CAD

Dans notre étude, nous nous sommes intéressés à la détection et/ou la segmentation de l'anomalie de masse d'un CADE.

La section suivante est un état de l'art sur les différents algorithmes de détection et/ou segmentation des masses mammaires.

1.7. Les algorithmes de détection et/ou segmentation des masses mammaires

Comme cité précédemment, les masses sont caractérisées par leur emplacement, leur taille, et leur forme (Sampat, Markey, & Bovik, 2005(a)) (Naga Mudigonda et al., 2001). La grande variation dans la taille et la forme dans laquelle la masse peut apparaître, fait que la segmentation reste une tâche difficile pour les chercheurs. Les masses spiculées ont la particularité d'avoir une haute probabilité de malignité (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013). En plus, dans la plupart des cas les mammographies donnent des images à faible contraste, alors le tissu peut être confondu avec la région de la tumeur du sein (Yao, Chen, & Chow, 2009) comme l'anomalie massive (Vyborny & Giger, 1994). A cause de ces problèmes, beaucoup de méthodes de détection et/ou segmentation des masses ont été développées. Et beaucoup de chercheurs se sont concentrés sur l'identification des régions d'intérêt (ROIs) avec l'anomalie suspecte. Nous pouvons voir une revue de ces méthodes ainsi que l'état d'avancement récent de ces dernières dans (Cheng, et al., 2006), (Arnau, et al., 2010), (Tang, Rangayyan, Xu, El Naqa, & Yang, 2009), (Bandyopadhyay, 2010) (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013).

Les algorithmes les plus communs pour identifier les régions d'intérêt (ROIs) sont des méthodes à base de pixels ou à base de régions. Toutes les approches, dans chaque méthode, ont été détaillées par J. Tang et al. Dans (Tang, Rangayyan, Xu, El Naqa, & Yang, 2009) et A. Jalalian et al. (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013).

1.7.1 Techniques de détection basées Pixels pour les CADe

Dans les approches basées pixels, les caractéristiques sont extraites pour chaque pixel et sont étiquetées comme suspectes ou normales (Sampat, Markey, & Bovik, 2005(b)).

Kegelmeyer et al. (Kegelmeyer, et al., 1994) ont présenté une approche dans laquelle les caractéristiques de texture de Laws (Laws, 1980) et des caractéristiques locales des bords orientés ont été extraites des régions d'intérêt (ROIs) et un classificateur d'arbre de décision binaire a été employé pour classifier des lésions spiculées des tissus normaux. Les auteurs ont

utilisé les vues de multiples d'images mammographiques. Ils ont obtenu une mesure de la sensibilité égale à 100 % avec une spécificité de 82 % dans 85 cas cliniques à quatre vues, contenant 36 cas de cancer et 36 cas normaux.

Liu et al. (Liu, Babbs, & Delp, 2001) ont proposé un schéma multi-résolution pour détecter des lésions spiculées. L'image a été décomposée dans une représentation de multi-résolution et quatre caractéristiques ont été extraites pour chaque pixel à chaque résolution. Les auteurs ont utilisé juste des mammographies contenant des lésions spiculées de la Société d'Analyse d'Image Mammographique (MIAS). Les auteurs ont obtenu un taux des TP (True Positif) TP= 84.2 %, et un taux des FP (False Positif) inférieur à 1.

Sampat et Bovik (Sampat & Bovik, 2003) ont présenté une approche pour détecter des masses spiculées dans des mammographies numérisées. L'approche est présentée en deux étapes. Dans la première étape, un algorithme de filtre a été utilisé pour améliorer les images. Dans la deuxième étape, un filtre radial de spiculation a été utilisé pour détecter l'emplacement spatial de la masse avec les caractéristiques améliorées. Les auteurs ont utilisé juste des mammograms contenant des lésions spiculées de la Base de données Numérique pour la Mammographie de dépistage (DDSM). Les auteurs n'ont pas donné de mesures quantitatives de leur méthode, mais les résultats ont montré que des régions de masse pouvaient être correctement localisées par l'algorithme proposé.

Campanini et al. (Campanini, et al., 2004) a présenté une approche basée sur la méthode SVM avec moins de caractéristiques pour la détection des masse dans des mammographies numériques. Au lieu d'extraire des caractéristiques de ROIS, les auteurs ont utilisé une approche multi-résolution, une représentation d'ondelette plus complète pour codifier l'image avec une redondance d'informations. Deux classificateurs SVM ont été utilisés dans cette approche. Le premier classificateur SVM a été utilisé pour détecter des masses candidates et le second classificateur SVM a été utilisé pour réduire le nombre de Faux Positifs. Les auteurs ont utilisé la base de données DDSM, mais aucun Az n'a été donné, les auteurs ont rapporté que l'algorithme a réalisé un TP de presque 80 % avec un FP de 1.1 par image pour des mammographies contenant des tumeurs malignes.

Kom et al. (Kom, Tiedeu, & Kom, 2007.) a proposé un algorithme de détection des masses en utilisant en premier lieu un algorithme de filtre de transformation linéaire pour améliorer l'image; l'image améliorée a été soustraite de l'image originale pour obtenir une image de différence. Une technique de seuillage adaptatif local a été développée pour détecter la masse dans l'image de différence. Les auteurs ont utilisé une méthode supervisée et aucune remarque pour la base de données utilisée. Les auteurs ont donné un coefficient ($Az = 0.94$ et $Az = 0.938$) correspondant à un taux de sensibilité de 95.91 % et 93.87 %, respectivement, quand l'étape de prétraitement a été ou n'a pas été appliquée.

Eltonsy et al. (Eltonsy, Tourassi, & Elmaghrab, 2007) a proposé un algorithme basé sur des couches concentriques multiples pour détecter des masses dans des mammographies. L'algorithme se présente en trois étapes. D'abord, les régions du sein issues des mammographies de dépistage ont été prétraitées par des techniques de segmentation et de granulation. Alors, les zones suspectes focales ont été détectées utilisant le raisonnement basé sur la connaissance. Les auteurs se sont concentrés à l'élimination des Faux Positifs dans la base de données DDSM.

Ces dernières méthodes ont été résumées par J. Tang et al. (Tang, Rangayyan, Xu, El Naqa, & Yang, 2009). Et plus récemment, toujours pour détecter les masses spiculées, la technique la plus utilisée est le calcul de l'orientation des bords à chaque pixel. Cette approche permettra de vérifier si la forme est spiculée selon toutes les directions (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013). Beaucoup de méthodes différentes sont disponibles pour calculer l'orientation des bords comme l'analyse statistique d'une carte d'orientation de pixel (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013), (Thangavel & Roselin, 2012) et l'extraction de caractéristique, à une représentation multi-résolution utilisant la transformée d'ondelette (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013), (Mencattini, Salmeri, Lojacono, Frigerio, & Caselli, 2008). Le seuillage à valeur multiple est un autre algorithme qui a été développé pour la détection de masse (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013), (Yu & Acton, 2002).

1.7.2. Techniques de détection basées Régions pour les CADe

Dans les approches basées région, les 'ROIs' sont segmentés et ensuite, les caractéristiques sont extraites de chaque région, qui seront par la suite utilisées pour classifier les régions comme suspect ou non suspect (Tang, Rangayyan, Xu, El Naqa, & Yang, 2009).

Chan et *al.* (Chan, et al., 1995) a présenté un algorithme à base de région dans lequel huit caractéristiques de texture ont été calculées des matrices SGLD (Spatial Grey Level Dependence) et une discrimination linéaire point par point a été utilisée pour déterminer l'importance de chaque caractéristique dans la distinction de masses du tissu normal. Les auteurs ont rapporté $Az = 0.84$ pour un ensemble d'images connues et $Az = 0.82$ pour l'ensemble d'images test. Les images connues sont des images dont le diagnostic est connu à l'avance et sont prises à base d'entraînement des algorithmes à tester. (Texture comme caractéristiques).

Dans le travail de Sahiner et *al.* (Sahiner, et al., 1996), quatre caractéristiques de texture sont extraites des matrices GLDS (Grey Level Difference Statistic), et trois caractéristiques de texture des matrices SGLD ont été utilisées pour la détection de la masse. Une circonvolution de réseaux de neurones a été employée comme la méthode de classification pour distinguer entre le tissu masse et le tissu normal. Les auteurs ont rapporté un $Az = 0.87$. (Texture comme caractéristiques).

Mudigonda et al. (Mudigonda, Rangayyan, & Desautels, 2000) a proposé une méthode utilisant des caractéristiques à base de gradient et des caractéristiques à base de texture pour différencier des masses bénignes des tumeurs malignes. Les caractéristiques à base de gradient sont le gradient directionnel et le coefficient de la variation du gradient. Vingt caractéristiques de texture basées sur des matrices GCMs (Grey Cooccurrence Matrix) ont représenté les informations de texture des régions de masses. Après l'union des caractéristiques à base de gradient et à base de texture, une probabilité postérieure calculée en utilisant la distance de Mahalanobis a été employée pour classifier des masses comme bénignes ou malignes. Ici, les auteurs se sont concentré pour classifier des masses comme

bénignes ou malignes, et ont rapporté un $Az = 0.84$. (Malin/bénin et Texture comme caractéristiques (fonctions)).

Un système d'aide à la détection des masses dans des images FFDM a été développé par Wei et *al.* (Wei, et al., 2005). Des images de FFDM brutes ont été améliorées en utilisant des méthodes multi-échelle. Puis une méthode de segmentation en deux étapes, qui a combiné des informations du domaine du gradient et des informations de niveau gris, a été utilisée pour détecter des masses suspectes sur ces images. Dans la troisième étape, des caractéristiques de texture morphologiques et des SGLD ont été extraites pour chaque masse suspecte. Une analyse de discrimination linéaire point par point (LDA) avec une optimisation de la transmission unidirectionnelle ont été employées pour choisir les caractéristiques les plus utiles. Les auteurs ont rapporté un Az entre 0.72 et 0.85. (Texture comme caractéristique).

Bellotti et *al.* (Bellotti, et al., 2006.) a proposé un système complètement automatisé pour la détection des masses. Le système a inclus trois étapes. D'abord, un algorithme de segmentation à base de bord a été mis en œuvre pour choisir les régions suspectes. Puis, huit caractéristiques indépendantes de texture des ROIS ont été extraites des GLCM selon quatre angles $\theta = k\pi/4$ ($k = 0, 1, 2, 3$). Finalement, une classification supervisée par réseaux de neurons, formés avec un apprenant de la descente de gradient, a été employée pour classifier des masses et des tissus normaux. Les auteurs ont rapporté un $Az = 0.783$. (Texture de glcm comme caractéristiques).

Récemment, une méthode de détection des masses automatisée a été présentée par Timp et al. (Timp, Varela, Karssemeijer, & Dacolian, 2007) pour détecter des changements temporels de masses sur des mammographies entre deux tours de dépistage consécutifs. Les auteurs ont utilisé des vues multiples d'images mammographiques et ont rapporté un $Az = 0.74$ sans caractéristiques temporelles et un $Az = 0.77$ avec l'utilisation de caractéristiques temporelles. (Vue multiples pour détecter des changements temporels).

Plus récemment, les chercheurs (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013), (Haralick, 1979), (American Cancer Society, 2002) ont développé des méthodes à base de région pour la détection des masses basées uniquement sur des

caractéristiques de frontières particulières des masses (Afsaneh, Syamsiah B.T., Hajjah Rozi, Abdul Rahman, & Babak, 2013).

1.8. Conclusion et Choix de méthodes

Après cette revue des méthodes les plus importantes utilisées dans la littérature, nous ne pouvons malheureusement pas donner une comparaison objective entre ces dernières car pour chaque méthode les auteurs ont utilisé des critères différents. Nous les résumons comme suit:

1. Des bases de données différentes.
2. Un nombre d'images de test différent.
3. Un contenu d'images différent, des auteurs prennent uniquement des images contenant des masses malignes par exemple.
4. Des vues d'images différentes, des auteurs testent leur algorithme sur une seule vue, alors que d'autres sur des vues multiples.
5. Parfois des objectifs finaux différents, par exemple, des auteurs commencent à détecter des masses, pour aboutir:
 5. a. Soit à une classification du tissu mammaire en région soupçonneuse c'est à dire une masse ou en tissu normal.
 5. b. Soit à une classification de ces masses en bénignes ou malignes.
 5. c. Soit à une réduction des FP
 5. d. Soit à une analyse temporelle de l'évolution des régions suspectes des images prises à des dates différentes, pour finalement détecter des masses.
6. Des méthodes sont supervisées, semi-supervisée ou non supervisée.
7. Les méthodes d'évaluation de l'exactitude et de la performance des algorithmes développés est différentes. Il y a des auteurs qui donnent le coefficient Az, d'autres qui donnent uniquement le nombre des FP et le nombre des TP. D'autres auteurs présentent uniquement des images sans mesure quantitative de l'exactitude de leurs algorithmes.

A.Oliver dans (Arnau, et al., 2010) a essayé de donner un résumé de quelques méthodes, en tenant compte de quelques critères que nous avons cités, mais la prise en considération de tous les critères fait que la comparaison reste une tâche difficile et fastidieuse pour les auteurs. Ainsi, ce domaine de recherche reste ouvert à toute contribution.

Pour notre étude, nous avons récapitulés les méthodes basées pixel, par exemple dans (Kupinski & Giger, 1998), (Li, Kallergi, Clarke, Jain, & Clark, 1995), (Rangayyan, et al., 1997), comme la croissance de régions et ses extensions. Les méthodes basées région (Petrick, Chan, Sahiner, & Wei, 1996), (Kobatake, Murakami, Takeo, & Nawan, 1999), telles que les méthodes basées sur des filtres. Et enfin, les méthodes qui se basent sur des bords simples (Rocha, Tong, & Yan, 2000), tels que les filtres de gradient qui sont largement employés au début de la segmentation des masses. Nous avons remarqué que quoique ces types de méthodes semblent faciles à mettre en œuvre, il est toujours difficile d'acquérir des résultats de segmentation satisfaisants pour les masses à frontières ambiguës. Cela est dû au fait que la caractéristique simple ne peut pas traiter les distributions de densité complexes et la topologie des masses ainsi que celle du tissu normal du sein. Pour détecter des frontières plus précises de masses, quelques chercheurs utilisent des méthodes par contour actifs (Xiao, Xia, & Wang, 2005), (Shi, et al., 2008), (Zou, Zheng, Zhou, & Agyepon, 2008), où l'efficacité de la méthode dépend de l'ajustement de paramètres.

Suite à ces inconvénients, beaucoup de méthodes citées ci-dessus utilisent des informations de texture, parce que les caractéristiques de texture sont des informations plus riches pour le processus de segmentation (Laws, 1980), (Mirmehdi, Xie, & Suri, 2009) et particulièrement dans des images médicales (Ilea & Whelan, 2011).

Ces informations de texture ont été prouvées pour être utiles dans la différenciation entre le tissu normal et le tissu de masses. Les auteurs dans (Pfisterer & Aghdasi, 1999) montrent que la région contenant une tumeur expose une texture typiquement 'basse' comparée à une région où le tissu est normal. Les auteurs dans (Yao, Chen, & Chow, 2009) ont conclu que les caractéristiques de texture démontrent des différences plus importantes entre la tumeur et des tissus normaux que la caractéristique d'intensité des pixels de l'image de mammographie. Dans cette idée, nous avons donné un résumé de quelques méthodes utilisant des caractéristiques de texture dans l'étape de détection et/ou l'étape de

segmentation, soit celles basées pixels ou celle basées régions. La plupart des méthodes incluent des caractéristiques de textures extraites des matrices GLCM dans l'étape de segmentation ou l'étape de classification d'un CAD (Xiaoming, Jun, Dongfeng, & Tang, 2010). Et la plupart de ces méthodes utilisent les matrices GLCM dans des approches basées région, pour extraire des caractéristiques de texture dans une étape précédant l'étape de segmentation ou de la classification de la masse. Nous citons quelques exemples dans (Kanta Maitra, I. et al., 2011), (Meenalosini, Janet, & Kannan, 2012), (Aswini kumar mohanty et al., 2011) (Thangavel & Roselin, 2012). Cependant, il n'y a aucun travail significatif qui a utilisé ces matrices dans une approche de détection par contours.

Ainsi, l'objectif dans cette thèse est la contribution au développement de méthodes de segmentation d'images selon deux grandes parties :

Dans la première partie, nous avons abordé le problème de la détection des anomalies de masses en utilisant la robustesse des matrices de cooccurrence de niveaux de gris (GLCM) et en introduisant l'information de texture extraite de ces matrices dans une approche par contours. En plus, nous expérimentons l'approche proposée sur des images présentant diverses densités de tissus (Gras, Dense et Glandulaire) pour apporter des résultats plus significatifs. À cette fin, quelques images types de la base de données MIAS ont été explorées. L'approche est décrite dans le chapitre 2 de ce travail.

Dans la seconde partie de cette étude, nous contribuons et proposons une méthode semi-automatique pour la segmentation des masses, basée sur le Modèle de Mélange des Gaussiennes (GMM). Le modèle GMM nous donne des informations importantes sur la distribution des probabilités d'intensités des images de mammographie. Par l'initialisation des paramètres de l'algorithme espérance-maximisation (EM) avec des statistiques sur la région suspecte et sur la région normale, et en utilisant l'algorithme du maximum de vraisemblance (ML), les paramètres du modèle GMM classique ont été ajustés au mieux. Le modèle sera nommé GMM/ML.

La troisième partie présente une nouvelle technique de segmentation du muscle pectorale et de la région mammaire. La détection et la délimitation de ces deux régions est un processus nécessaire et une étape essentielle pour les systèmes d'aide au diagnostic CAD, CADE ou CADx. Beaucoup d'auteurs se sont intéressés et ont développé des approches pour

résoudre les problèmes liés soit à la segmentation du muscle pectoral, soit à la segmentation de la région mammaire. Un état de l'art sur les différentes méthodes utilisées est présenté dans cette partie. Nous avons contribué en testant l'utilisation des matrices de cooccurrence de niveaux de gris dans une approche qui n'a pas encore été utilisée dans ces problèmes. L'approche proposée a fourni des résultats satisfaisants, notamment en vue de résoudre certains problèmes non encore résolus, telle la présence d'anomalies de masses ou de micro-calcifications dans les limites pectorales, habituellement ignorées dans les étapes de segmentation mammaires. D'un autre côté, la méthode proposée permet de détecter le muscle pectoral et la région mammaire à la fois. Les tests ont été appliqués sur des images de la base de données « mini MIAS », constituée de mammographies avec des vues « MLO » (Medio Lateral Oblique).

Enfin, une discussion et conclusions de nos travaux sont données, ainsi qu'une proposition des travaux futurs pour améliorer les approches proposées dans chacune des trois parties.

Chapitre 2 : Intégration des descripteurs de texture extraits des Matrices de Cooccurrence de Niveaux de Gris (GLCM) dans une approche de détection de contours des anomalies de masses.

Résumé

Dans ce chapitre, le critère « arrangement spatial des pixels » des images de mammographie est utilisé, et une approche texturale de détection des anomalies de masses est proposée. Elle est basée sur les descripteurs de texture extraits des Matrices de Cooccurrence de Niveaux Gris (GLCM). Actuellement, dans la plupart des travaux réalisés dans le domaine de la détection et/ou la segmentation des masses, les matrices GLCM sont utilisées comme des caractéristiques de texture avec une approche à base de région. Ces caractéristiques entrent dans la phase précédant l'étape de segmentation ou elles sont utilisées comme des données pour l'étape de classification. Le travail discuté dans ce chapitre essaie d'expérimenter la méthode GLCM dans une approche à base de contour. Et vu la complexité des tissus mammaires ainsi que leur importance, l'approche proposée a été testée sur des images à diverses densités de tissu, de la base de données MIAS.

2.1. Introduction

Le Chapitre présente d'abord les définitions principales de la texture dans la littérature, ainsi que toutes les caractéristiques s'y attachant. L'approche proposée pour la détection des masses est ensuite donnée, ainsi qu'une description des données utilisées à cette fin.

2.2. Définitions de la texture

2.2.1. Les difficultés liées à la définition de la texture

Donner une définition précise et universelle de la notion de texture est un problème qui relève presque de l'impossible et ce dernier n'a toujours pas été résolu. Il existe une profusion des approches trouvées actuellement dans la littérature du traitement d'images en général. Pourtant une meilleure compréhension de la notion de texture permettrait d'aborder son analyse de façon plus méthodique et plus concentrée; un formalisme mathématique de cette notion serait un idéal pour les chercheurs, chose qui malheureusement n'existe pas encore.

2.2.2. Définition de la texture selon les domaines d'application

Le terme texture admet des acceptions différentes selon le domaine dans lequel il est utilisé (médecine, géologie, pétrologie, textile, agro-alimentaire,...). Et au sein d'un même domaine, particulièrement en traitement d'images, il n'existe pas de véritable consensus en ce qui concerne la définition d'une texture. Dire qu'une texture est une région d'une image présentant une organisation spatiale homogène des niveaux de luminance est correct mais très peu précis.

I. Claude (Claude, Pouletaut, Huault, & Boulanger, 2001) a bien résumé les raisons du manque crucial d'une définition réelle de la notion de texture. Cela est dû, en partie, à la difficulté de comprendre et modéliser le système de perception visuelle humain; malgré les progrès effectués en physiologie du système nerveux et en sciences cognitives, nous ne connaissons toujours pas, exactement, la nature de l'information que l'être humain capte pour analyser et interpréter une image. Nous ne connaissons donc pas l'information qu'il faudrait fournir à un système de vision artificielle pour qu'il puisse analyser et comprendre une image, comme nous le faisons nous-mêmes.

C'est pourquoi, et malgré la présence importante des textures dans les images, les efforts de nombreux chercheurs, depuis plusieurs décennies, ont conduit principalement à des solutions « ad-hoc » adaptées à une application précise, ou, tout au moins, à une famille d'images (Claude, Pouletaut, Huault, & Boulanger, 2001).

2.2.3. Définition de la texture par le dictionnaire

En termes de définition de la notion de texture, on trouve d'abord celle donnée par le dictionnaire, qui précise simplement qu'une texture est la reproduction spatiale d'un motif de base dans plusieurs directions.

Puis, d'autres plus précises, telle que cette définition: une texture est une structure spatiale constituée par l'organisation de primitives (ou motifs de base) ayant chacune un aspect aléatoire. Une texture est alors une « structure hiérarchique » à deux niveaux.

2.2.4. Définition de la texture comme un phénomène à deux dimensions

Ou bien encore, cette définition qui considère une texture comme un « phénomène » à deux dimensions:

- . La première concerne la description d'éléments de base ou primitives, à partir desquels est formée la texture.
- . La deuxième est relative à la description des relations spatiales entre ces

primitives.

On distingue en fait, selon l'élément de base et les règles de placement, deux types de textures :

. Les macro-textures (ou textures structurées) pour lesquelles il est facile d'extraire visuellement le motif de base et les lois d'assemblage des primitives entre elles. Ces textures peuvent même présenter une certaine périodicité ou cyclo-stationnarité (processus aléatoire plaqué sur un processus périodique). Certains exemples sont représentatifs de ce type de textures, comme la texture d'un mur de brique, de certains tissus ou d'un grillage.

. Les micro-textures (ou textures aléatoires) qui présentent un aspect désorganisé mais dont l'impression visuelle reste globalement homogène. Les différentes régions d'une image aérienne, les bois, les champs, etc., représentent des textures microscopiques.

2.3. Descripteurs statistiques de la texture

2.3.1. Descripteurs Simples

Une approche dans l'étude d'une texture est de calculer la moyenne des niveaux de gris des pixels appartenant à une fenêtre B , ou le calcul de l'écart-type sur cette même fenêtre. Mais cela n'est pas suffisant pour décrire les textures. La figure (Fig.2.1.) montre des exemples de fenêtres avec des textures différentes mais ayant même moyenne et même écart type.

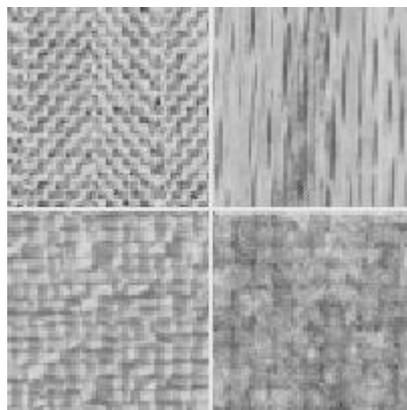


Fig.2.1. Textures ayant même moyenne et même écart type (M. Van Droogenbroeck, 2003).

Une définition plus précise de la notion de moyenne et d'écart-type sur un voisinage (ou fenêtre) a été introduite. Il s'agit de considérer une image f contenant au moins une texture, une fenêtre B complètement incluse dans cette texture. Nous noterons $\#(B)$ le cardinal de B . L'étude de la texture se fait via l'étude de l'image f sur la fenêtre B de manière à fournir une information locale pertinente.

. Moyenne locale

$$\mu_f = \frac{1}{\#(B)} \sum_{(x,y) \in B} f(x,y) \quad (2.1)$$

. Écart-type local

$$\sigma_f = \sqrt{\frac{\sum_{(x,y) \in B} [f(x,y) - \mu_f]^2}{\#(B)}} \quad (2.2)$$

Ces exemples de paramètres peuvent être utilisés comme descripteurs de texture. Cependant et en raison de leur faible pouvoir discriminant, il est nécessaire de combiner plusieurs descripteurs.

2.3.2. Descripteurs Complexes : Les Matrices de cooccurrence des niveaux de gris

2.3.2.1. Inconvénients des descripteurs statistiques simples

Les descripteurs statistiques décrits précédemment ne fournissent pas d'information concernant la nature périodique qu'une texture peut présenter. Ainsi, la texture de fond d'images (petits carrés) de la figure (Fig.2.2.) présente un caractère périodique qu'aucun descripteur introduit précédemment ne pourra mettre en lumière.

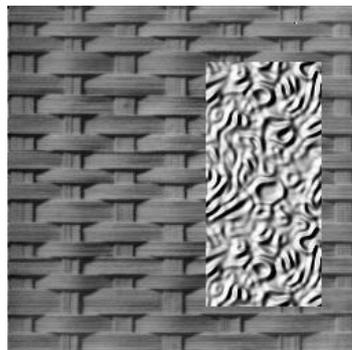


Fig.2.2. Exemple d'image comportant deux textures (M. Van Droogenbroeck, 2003).

La matrice de cooccurrence introduite ci-après va offrir un moyen de caractériser la périodicité et la directivité des textures.

2.3.2.2. La matrice de cooccurrence de niveaux de gris

L'idée consiste à étudier la texture au moyen d'une fenêtre B d'observation de la texture en comptabilisant le nombre de paires de pixels distants de d qui présentent une différence Δz en niveaux de gris. En plus de la distance, on tient compte de la direction définie par toute paire de pixels.

Une matrice de cooccurrence est définie au moyen d'une relation géométrique R entre deux pixels (x_1, y_1) et (x_2, y_2) . Un exemple de relation géométrique simple est :

$$x_2 = x_1 + 1 \quad (2.3)$$

$$y_2 = y_1 \quad (2.4)$$

Pour lequel le pixel (x_2, y_2) est le pixel voisin de droite de (x_1, y_1) .

La matrice de cooccurrence $C_R(i, j)$ est carrée et de dimension $L \times L$, où L est le nombre de niveaux de gris présents dans B . Les indices de la matrice de cooccurrence sont donc les niveaux de gris de la texture étudiée. La matrice $C_R(i, j)$ se construit comme suit (telecom, 2013):

1. Initialisation de la matrice : $\forall i, j \in [0, L[: C_R(i, j) = 0$.
2. Remplissage de la matrice. Si la relation R entre deux pixels (x_1, y_1) et (x_2, y_2) est respectée:

$$C_R(f(x_1, y_1), f(x_2, y_2)) = C_R(f(x_1, y_1), f(x_2, y_2)) + 1 \quad (2.5)$$

Dès lors, $C_R(i, j)$ contient le nombre de fois que l'on a rencontré dans B deux pixels $(x_1, y_1), (x_2, y_2)$ vérifiant la relation géométrique R et tel que $f(x_1, y_1) = i$ et $f(x_2, y_2) = j$. En pratique, on utilise des relations géométriques assez simples comme celle citée comme exemple. Ainsi, on définit une matrice orientée :

$$P_{90^\circ, d}(i, j) = \# \{(x_1, y_1), (x_2, y_2) \in B \mid x_1 = x_2, |y_2 - y_1| = d, f(x_1, y_1) = i \text{ et } f(x_2, y_2) = j\} \quad (2.6)$$

, où 90° correspond à une direction verticale et d détermine la distance entre les deux pixels en question.

De même, on définit :

$$P_{0,d}^{\circ}(i,j) = \#\{(x_1, y_1), (x_2, y_2) \in B \mid y_1 = y_2, |x_2 - x_1| = d, f(x_1, y_1) = i \text{ et } f(x_2, y_2) = j\} \quad (2.7)$$

Il est également possible de définir d'autres matrices correspondant aux directions 45° et 135° .

Cette définition fournit une matrice symétrique, ce qui permet de ne stocker qu'une partie de la matrice.

Exemple. Soit l'image suivante comportant $L = 4$ niveaux de gris ($l = 0, 1, 2, 3$) :

$$f(x,y) = \begin{array}{cccc} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & 2 & 3 \\ 2 & 2 & 3 & 3 \end{array} \quad (2.8)$$

Les matrices $P_{0,1}^{\circ}$ et $P_{90,1}^{\circ}$ sont donc de dimension 4 et valent :

$$P_{0,1}^{\circ} = \begin{bmatrix} 6 & 2 & 1 & 0 \\ 2 & 2 & 0 & 0 \\ 1 & 0 & 4 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix} \quad P_{90,1}^{\circ} = \begin{bmatrix} 6 & 1 & 2 & 0 \\ 1 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix} \quad (2.9)$$

On peut remarquer que les matrices de cooccurrence sont bien symétriques.

L'analyse de textures sur la base de la matrice de cooccurrence est fondée sur l'occurrence répétée d'une configuration typique dans la texture; cette configuration varie rapidement en fonction de la distance pour des textures fines et lentement pour des textures plus grossière.

2.3.2.3. Les descripteurs de texture extraits des matrices de cooccurrence

2.3.2.3.1. Moments

Le moment d'ordre k d'une matrice de cooccurrence $C_R(i, j)$ est défini par :

$$(m_R)_k = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j) \cdot C_R(i, j) \quad (2.10)$$

2.3.2.3.2. Moments de la différence inverse

$$(m_R^{-1})_k = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{C_R(i, j)}{(i-j)^k}, i \neq j \quad (2.11)$$

2.3.2.3.3. Contraste

$$C(k, n) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} |i-j|^k \cdot C_R(i, j)^n \quad (2.12)$$

2.3.2.3.3. Entropie

$$H = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} C_R(i, j) \log_2 C_R(i, j) \quad (2.13)$$

2.3.2.4. Performances des descripteurs de texture extraits des matrices de cooccurrence

Les performances de toute méthode utilisant ces descripteurs de texture sont étroitement liées au choix de la taille de la fenêtre, de la distance et de la direction. Ces choix qui dépendent souvent de l'application et du pouvoir discriminant de ces descripteurs. Des tests sur des images issues de la base de données « Brodatz » sont détaillés au paragraphe § 2.5.4.2.

Il existe plusieurs formes de matrices GLCM, dérivées de la définition donnée ci-dessus, nous avons pensé inutile de les développer dans cette étude car nous avons pris la définition de base des GLCM introduite par Haralick (Haralick, 1979).

2.4. La base de données MIAS

L'approche proposée dans cette étude a été appliquée sur des données de la base MIAS, disponible en ligne librement pour des buts scientifiques (J Suckling et al., 1994) (MIAS, 2012). Les films ont été numérisés et les images correspondantes ont été annotées par des radiologues. L'annotation concerne la densité du tissu mammaire, la nature de l'anomalie ainsi que sa position et sa forme.

Trois classes du tissu mammaire sont distinctes: Gras (Fatty, F), Gras-glandulaire (Glandular, G) et Dense-glandulaire (Dense, D). Toutes les anomalies ont été aussi détectées et décrites : des calcifications, bien définies, spiculées ou des masses mal définies, ou encore les distorsions architecturales ou l'asymétrie. Chaque paire d'images dans la base de données est annotée comme Symétrique ou Asymétrique. La base de données fournit aussi la disposition de chaque anomalie, par son centre de gravité dans un rayon r . Et chaque anomalie est décrite par sa sévérité, maligne ou bénigne. La base de données contient 322 images qui sont réparties en 207 images normales, 38 images contenant des masses et 169 images contenant d'autres anomalies.

2.5. Approche texturale proposée pour la détection des masses

2.5.1. Position de la Problématique

Les problèmes les plus importants se rapportant à la détection des anomalies cancéreuses de masses, ont été résumés :

- Il y a des confusions entre le tissu contenant la masse et le tissu normal.
- Il est très difficile de détecter la masse surtout dans un tissu gras.
- Il est nécessaire de détecter la forme de la masse, mais cette dernière se présente sous des formes très variées d'une personne à une autre. Et il est parfois impossible de donner une forme exacte.

Il est à noter que la forme de la masse peut donner une information pour mesurer le degré de malignité lié à cette dernière (une masse de forme spiculée est hautement maligne) et que la détection de la forme aidera l'expert à donner un meilleur diagnostic.

- En finalité, les méthodes numériques (et donc les systèmes d'aide à la détection CADe et/ou au diagnostic CADx), se basent sur des fondements mathématiques et la

formalisation du problème reste une tâche difficile pour les chercheurs. La définition mathématique de la « texture » en elle-même, reste en cours de recherche.

2.5.2. Hypothèses de travail

D'après tous les travaux antérieurs cités précédemment au chapitre 1, en particulier ceux utilisant l'information de texture pour détecter et/ou segmenter les anomalies de masse, il apparaît que la région contenant une masse présente généralement une texture différente de la région où le tissu est normal. Cela a été démontré de plusieurs façons différentes par les chercheurs. Dans cette étude, l'idée est basée sur le fait que le descripteur de « contraste » peut détecter un éventuel changement spatial au niveau des frontières des deux régions principales, la région contenant une masse et la région contenant le tissu normal. Dans ce cas, le descripteur de texture extrait des GLCM est calculé au niveau de chaque pixel. « Des valeurs grandes de ce descripteur reflètent un changement brusque des valeurs de niveaux de gris. Par ailleurs, des valeurs faibles reflètent une certaine homogénéité des valeurs de niveaux de gris ».

Au niveau des images numériques, cela se traduit par des niveaux de gris qui donnent une couleur blanche pour de fortes valeurs, et des niveaux de gris qui donnent des couleurs qui tendent vers le noir pour de faibles valeurs. Cela a été testé et démontré sur des images de la base de données « Brodatz » et sur des images de mammographie de la base de données MIAS, voir les figures, Fig.2.3 et Fig.2.4.

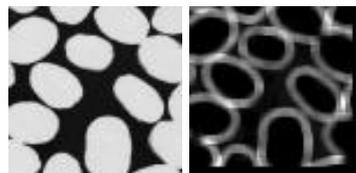


Fig.2.3.Gauche : Image « Brodatz », Droite: Image texturale GLCM

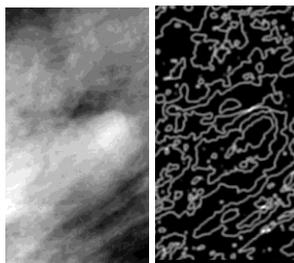


Fig.2.4. Gauche : Image de mammographie, Droite: Image texturale GLCM

Le descripteur de contraste de la figure Fig.2.4 a été calculé dans une région ROI de l'image.

Ces conclusions ont permis de développer une approche de détection des contours des anomalies de masses. Elle est détaillée et testée dans les parties suivantes.

2.5.3. Schéma de l'approche proposée

Les principales étapes de la méthodologie proposée sont récapitulées dans la figure Fig.2.5. :

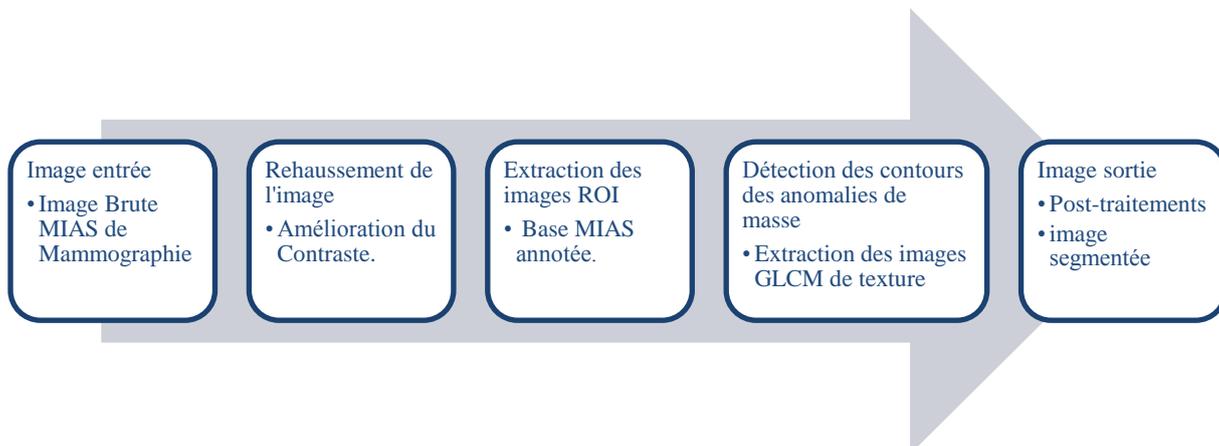


Fig.2.5. Principales étapes de la méthodologie proposée

2.5.4. Les étapes de l'approche proposée

2.5.4.1 Amélioration du contraste des images

La performance des méthodes basées sur les informations de texture dépend fortement de l'étape de prétraitement de l'image à traiter (Pfisterer & Aghdasi, 1999). Cette étape

consiste soit à l'amélioration du contraste des niveaux de gris, soit à l'élimination d'éventuels bruits perturbant l'analyse de l'image. Beaucoup de chercheurs se sont focalisés sur cette première étape de l'ensemble du processus des CAD, afin de développer des méthodes appropriées à cet effet. Il en ressort que la méthode choisie dépend de l'objet à détecter et/ou à segmenter, soit les anomalies de masses, de micro-calcifications ou de distorsions architecturales. Et aussi de la méthode de détection et/ou de segmentation utilisée.

Pour l'approche proposée, cette étape représente notre touche clé. Le but est d'obtenir des meilleurs résultats à l'étape suivante qui est l'étape de détection des contours des masses.

La plupart des images de mammographie présentent un faible contraste d'intensité (Mencattini, Salmeri, Lojacono, Frigerio, & Caselli, 2008) et comme toute image numérique, d'éventuels bruits peuvent désorienter le déroulement de l'étape de détection et/ou de segmentation de l'objet recherché. Par exemple, si l'objet recherché est de petite taille, voire de taille minuscule comme les micro-calcifications, appliquer une méthode de lissage pour enlever du bruit, avant une méthode d'amélioration de contraste de l'image, pourrait supprimer des pixels appartenant à l'objet recherché. Pour cela, comme la taille des anomalies de masses est comparativement grande par rapport à la taille des micro-calcifications, il a fallu d'abord appliquer un lissage de l'image pour éliminer d'éventuels bruits, ensuite une amélioration de son contraste.

Pour le lissage, un filtre de diffusion SRAD (Speckle Reducing Anisotropic Diffusion) (Yu & Acton, 2002), a été utilisé.

Pour l'amélioration du contraste, la méthode d'Egalisation d'Histogramme Adaptative Limitée au contraste, « CLAHE » (Contrast-limited Adaptive Histogram Equalization (CLAHE)), a été utilisée.

Contrairement à la plupart des études similaires, qui extraient les régions d'intérêt ROI, puis y appliquent des améliorations, dans notre approche le lissage et l'amélioration du contraste sont appliqués sur l'image entière de mammographie. Ensuite, nous extrayons l'image ROI suspecte. Ainsi, l'algorithme SRAD peut prendre le bruit pour chaque prise d'image de mammographie, en incluant les régions hors de toute la région mammaire. C'est en général ces régions qui introduisent le bruit le plus gênant pour la reconnaissance des textures des pixels de l'image de mammographie. La démarche proposée donne plus

d'efficacité pour la réduction du bruit dans l'image, et plus de performance pour l'étape suivante de détection des contours des masses.

Pour expliquer cette démarche, c'est-à-dire le choix des étapes d'amélioration du contraste de l'image de mammographie, des tests ont été effectués. Des résultats ont été donnés en appliquant uniquement un lissage, ou uniquement un contraste, ou un contraste puis un lissage, ou un lissage puis un contraste. Ainsi, on pourra faire une distinction et une comparaison entre ces différents résultats.

La figure Fig.2.6 donne exemple d'une image avec les différentes étapes testées pour améliorer le contraste d'une image de mammographie. Les figures Fig.2.6 (b) et Fig.2.6 (c) montrent respectivement le cas où seul un lissage ou un contraste a été appliqué sur l'image.

La figure 2.6 (d) est le cas où un contraste puis un lissage ont été appliqués sur l'image.

La figure 2.6 (e) est le cas où un lissage puis un contraste ont été appliqués sur l'image.

On voit que l'image (e) donne plus de détails et plus de contraste que (b), (c) ou (d). Les régions claires sont encore plus claires, ces régions peuvent correspondre à des régions du tissu contenant des anomalies de masses. Et les régions sombres sont plus sombres, ce qui peut correspondre aux régions du tissu normal, sans anomalies.

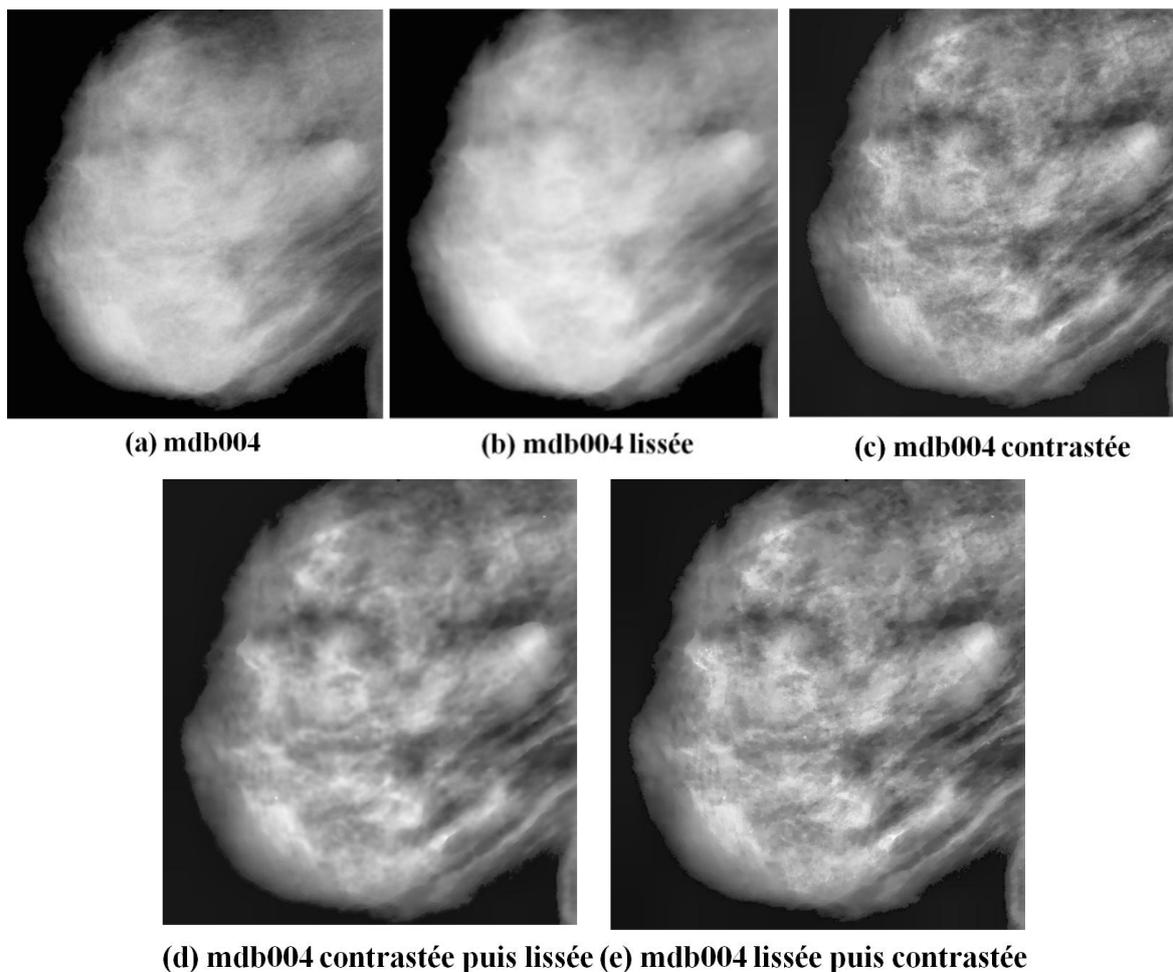


Fig.2.6. Etapes Test d'amélioration du contraste de l'image mdb004.

Les figures suivantes Fig.2.7, Fig.2.8 et Fig.2.9 sont des exemples des images ROI extraits des images prises comme test (respectivement, mdb004, mdb010 et mdb001). Elles montrent aussi l'importance du choix des étapes d'amélioration de contraste de l'image de mammographie d'origine :

La figure 2.7 (a) est une image ROI extraite de l'image mdb004 de mammographie.

La figure 2.7 (b) est une image ROI extraite de l'image mdb004 de mammographie qui a été seulement lissée.

La figure 2.7 (c) est une image ROI extraite de l'image mdb004 de mammographie qui a été seulement contrastée.

La figure 2.7 (d) est une image ROI extraite de l'image mdb004 de mammographie qui a été contrastée puis lissée.

La figure 2.7 (e) est une image ROI extraite de l'image mdb004 de mammographie qui a été lissée puis contrastée.

On voit que dans l'image ROI (e) donne plus de détails et surtout moins de bruit que les images ROI ((b), (c) ou (d)).

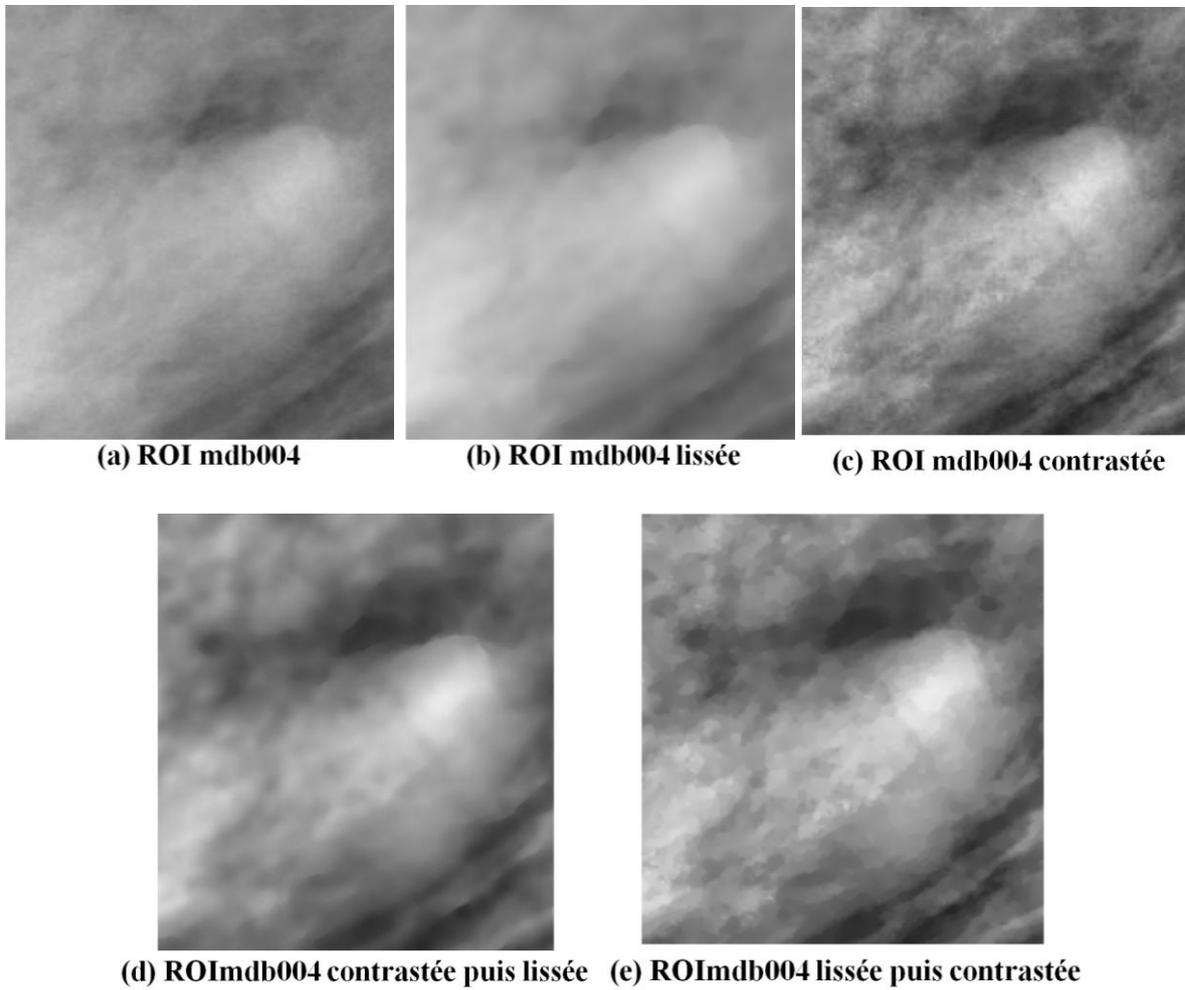


Fig.2.7. Images ROI avec les étapes test d'amélioration du contraste de l'image mdb004.

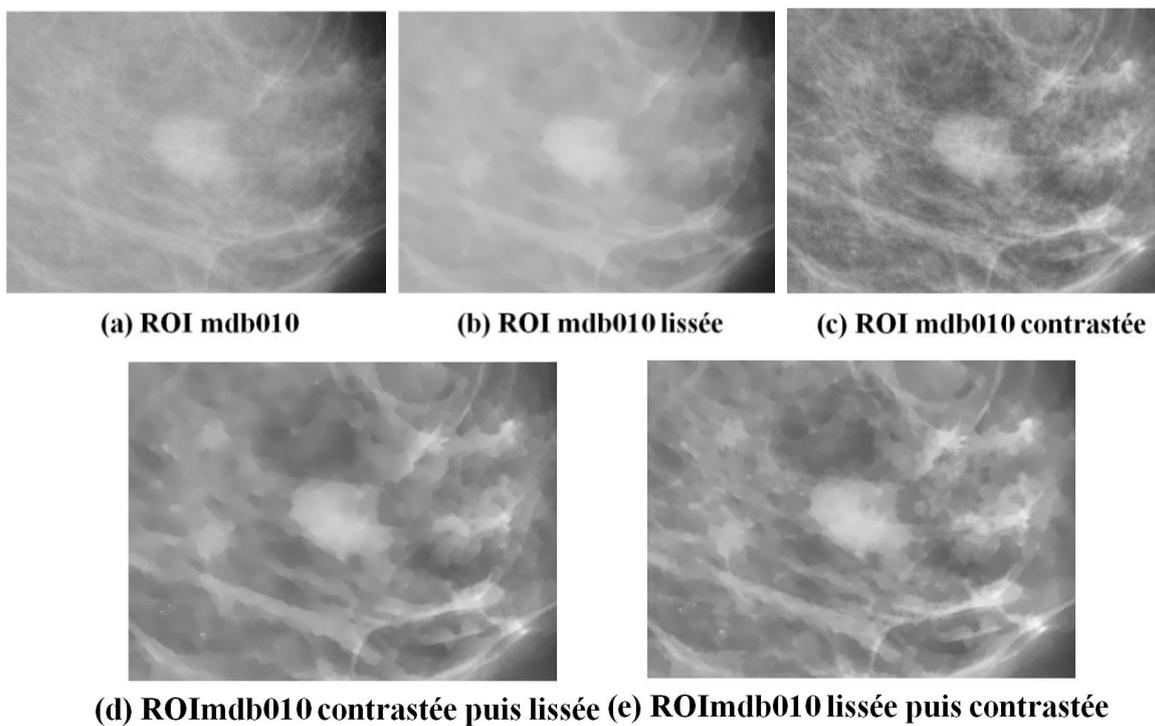


Fig.2.8. Images ROI avec les étapes test d'amélioration du contraste de l'image mdb010.

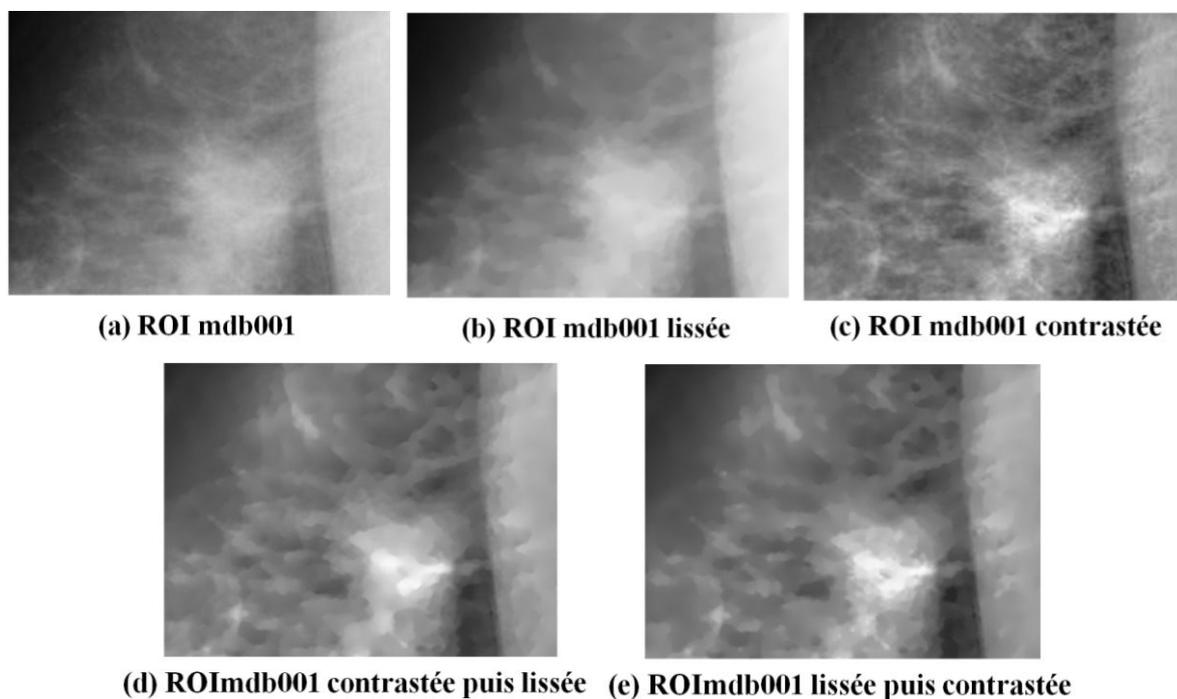


Fig.2.9. Images ROI avec les étapes test d'amélioration du contraste de l'image mdb001.

Concernant l'extraction des images ROI, cela a été réalisé à partir des annotations de la base MIAS. Ces annotations donnent l'emplacement des anomalies, leur rayon, ainsi que leurs contours. En plus de ces annotations, des délimitations des anomalies réalisées par des experts, ont été prises en considération pour vérifier la position de ces dernières. L'image de la figure (Fig.2.10) est un exemple d'une image annotée par l'expert (Fig.2.10 (b)), avec une délimitation du ROI (Fig.2.10 (c)), et un zoom de l'image ROI extraite de l'image améliorée (Fig.2.10 (d)).

Notons que l'image mdb004 présente un exemple de cas où la masse se trouve dans un tissu dense. Ce dernier est représenté par une couleur claire dans l'image de mammographie. C'est l'un des cas les plus difficiles pour la détection des masses. On remarque que l'image améliorée donne des détails importants pour l'étape de détection des contours de la masse.

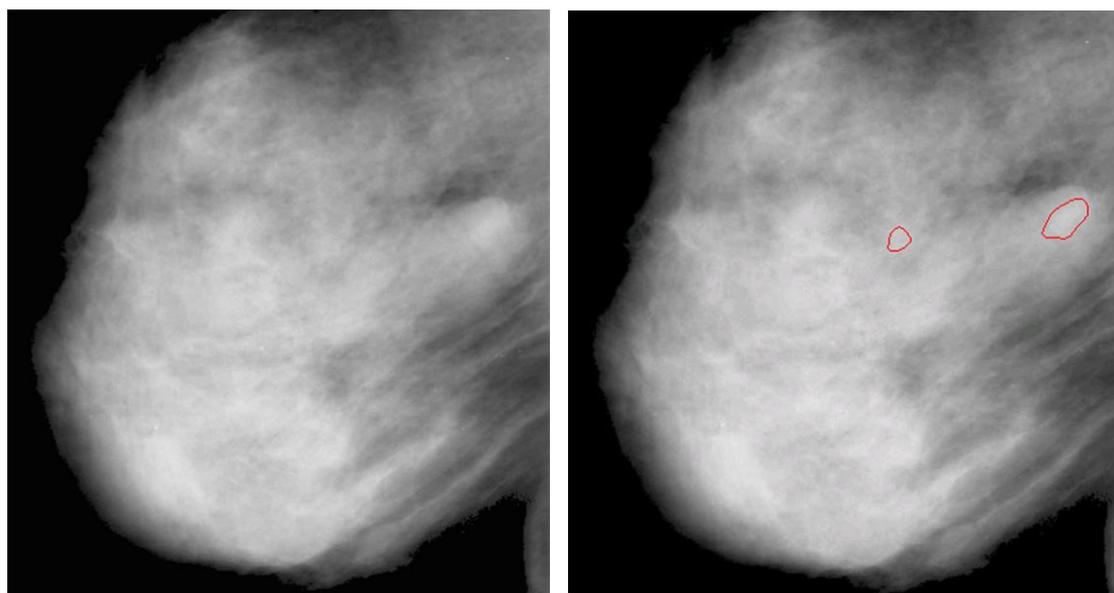
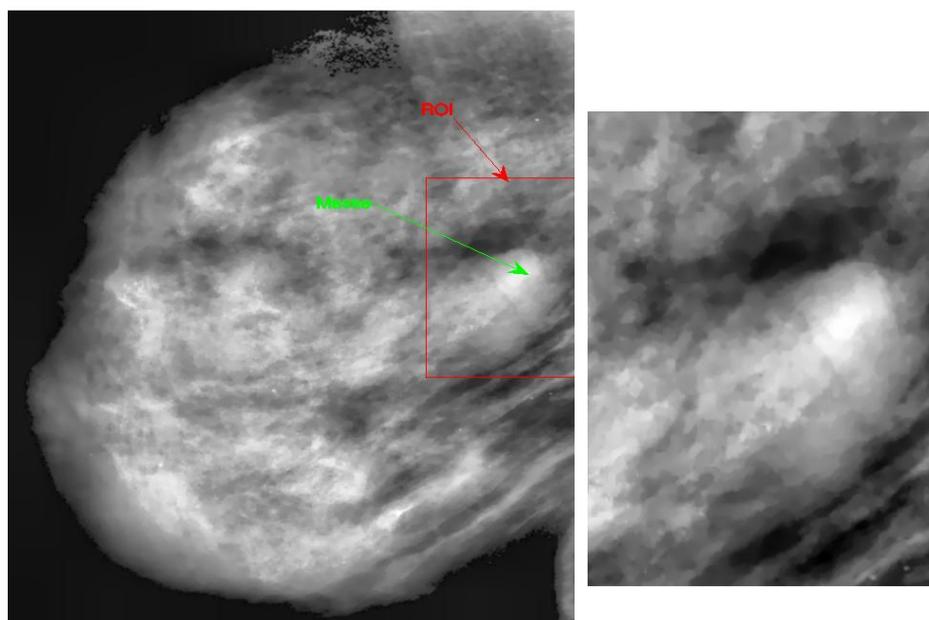
**(a) mdb004****(b) mdb004 annotée****(c) mdb004 améliorée****(d) ROI mdb004**

Fig.2.10. Extraction ROI : (a) MIAS mdb004 input, (b) mdb004 annotée par l'expert, (c) mdb004 contrastée avec ROI (en rouge) et une masse (flèche en vert), (d) image zoom du ROI.

2.5.4.2 Détection des contours des masses par le calcul des GLCM

Dans l'image ROI, les matrices de cooccurrence de niveaux de gris ont été calculées, selon quatre importants paramètres: la distance d , la direction (θ), la taille du voisinage ou le masque glissant tout au long de l'image et le descripteur de texture Haralick (Haralick, 1979).

2.5.4.2.1. Choix de la distance d

La distance d traduit le voisinage du pixel à considérer. Le 1^{er} pixel voisin est traduit par une distance $d=1$, le deuxième pixel voisin est traduit par une distance $d=2$, ...etc. Selon une direction choisie.

Par exemple, si on reprend l'équation (2.6) du paragraphe § 2.3.2.2. , et pour la direction 90° , on peut généraliser le cas du calcul de la distance d et on peut dire que d peut avoir des valeurs variant de 1, jusque la taille de la fenêtre B :

$$\begin{aligned} |y_2 - y_1| &= d \\ d &= 1, d = 2, d = 3, \dots etc \end{aligned} \quad (2.14)$$

Le choix de la distance d dépend de la taille des objets, ou plus précisément de la taille des textures recherchées, grossières ou fines.

Comme nous ne voulons négliger aucune information, nous avons pris $d=1$. En prenant $d=2$, l'information donnée par le premier pixel voisin ($d=1$) sera négligée. Des tests ont été réalisés avec différentes distances et cette hypothèse a été vérifiée.

2.5.4.2.2. Choix de la direction

Calculer les matrices selon une direction unique 0° , ou 45° , ou 90° , ou 135° , ne donnerait pas des traits ou des contours fermés, ils seront plutôt discontinus. Donc les matrices ont été calculées dans les quatre directions, ce qui donne quatre images de texture. Ensuite, et dans le but d'avoir tous les contours détectés, l'image somme a été calculée.

La figure Fig.2.11 confirme notre hypothèse. Elle représente est un exemple de la base de données "Brodatz", connue pour les tests de textures. Les quatre images de texture qui représentent le descripteur de contraste d'Haralick (Haralick, 1979), dans les directions $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$, sont présentées. Ainsi que l'image texture 'somme' de ces quatre images.

L'image somme montre clairement des traits ou contours plus fermés que chacune des quatre images prise seule.

La figure (Fig.2.12), est un ROI extrait de l'image MIAS mdb004. Les remarques faites sur l'image test « Brodatz » de la figure (Fig.2.11), sont là aussi confirmées.

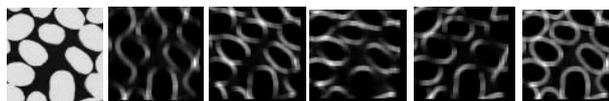


Fig.2.11. De gauche à droite : Brodatz D75 image, Image Contraste 0°, Image Contraste 45°, Image Contraste 90°, Image Contraste 135°, Image Contraste Somme (0°+ 45°+ 90°+135°).

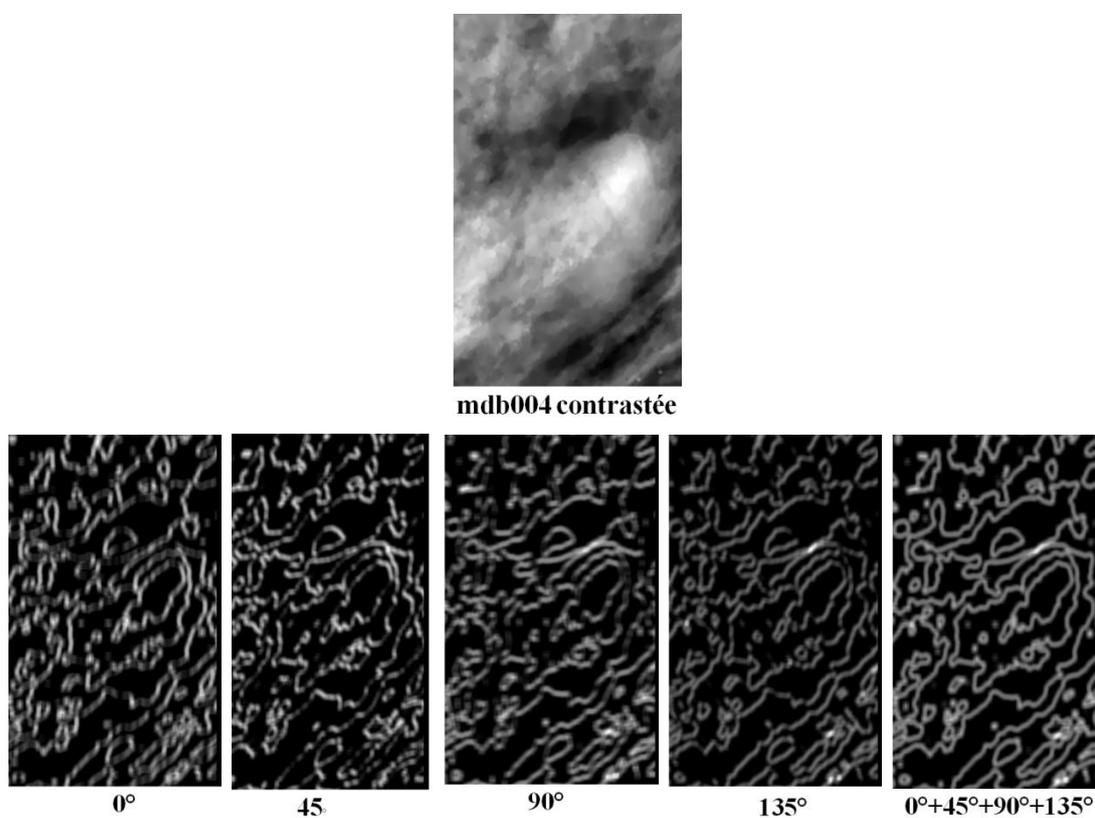


Fig.2.12. Haut: Image mdb004 contrastée, Bas: Descripteur contraste selon les 4 directions puis image somme qui montre des contours fermés.

2.5.4.2.3. Choix de la taille du voisinage ou du masque:

Pour comprendre l'effet du choix de la taille du voisinage sur le calcul du descripteur de contraste, des tests ont été aussi réalisés sur des images « Brodatz » aussi. Les calculs sont

effectués en prenant des tailles de masques différentes (3x3, 7x7, ... etc). Nous pouvons voir dans la figure Fig.2.13 que dans un masque de taille 3x3, les traits et contours sont moins lisses que pour le masque 7x7 et le masque 9x9. Les contours détectés sont d'autant plus flous que la taille du voisinage est grande.

Mais en réalité, le choix de la taille du masque dépend étroitement des textures des objets de l'image, et plus spécialement des textures et tailles des objets recherchés. Pour les images MIAS issues de mammographies, plusieurs tests ont été réalisés, avec plusieurs tailles de masques. Le masque de taille 7x7 a donné des contours plus lisses que le masque 3x3, et des traits plus fins que le masque 9x9.

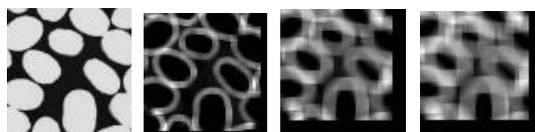


Fig.2.13. De gauche à droite, Brodatz D75 image, Image Contraste dans un masque 3x3, Image Contraste dans un masque 7x7 et Image Contraste dans un masque 9x9.

2.5.4.2.4. Calcul du descripteur de texture:

Dans le but d'avoir plus de caractéristiques de texture des régions suspectes recherchées, la plus part des auteurs calculent tous les descripteurs de texture pouvant être extraits des matrices GLCM. D'autres utilisent des descripteurs les plus connus, soit l'entropie, le contraste, le second moment angulaire, ou le moment des différences inverse. Seulement, les résultats de nos travaux antérieurs (Djaroudib, 1993(a)), (Djaroudib, 1993(b)), ont montré que seul le descripteur de contraste pourrait être suffisant afin de donner l'information recherchée, en l'occurrence, un changement spatial, une variation plus ou moins importante des textures au niveau des frontières, entre la région contenant une anomalie et la région du tissu normal.

En effet, le descripteur de contraste mesure l'hétérogénéité de l'arrangement spatial des niveaux de gris, et ainsi détecte les variations spatiales plus ou moins brusques de ces niveaux de gris (hypothèse de travail). Il résume aussi toute l'information de texture recherchée et utile pour cette étude.

Pour extraire les images texturales, selon les paramètres décrits précédemment (direction, taille du masque et descripteur de texture), le descripteur de contraste a été calculé dans l'image ROI. Cette dernière est balayée par un masque où la matrice de cooccurrence

correspondante est calculée, puis le pixel central du masque est remplacé par le descripteur de contraste extrait de cette matrice. En conclusion, chaque pixel du ROI est remplacé par le descripteur de contraste. Cette procédure permet de produire des images texturales telle que l'exemple vu dans la figure Fig2.12.

Dans ce travail, la formulation du descripteur de contraste est donnée par l'équation (2.15) ci-dessous :

$$contrast = \sum_{i,j} |i - j|^2 p(i, j) \quad (2.15)$$

Où $p(i, j)$: est la valeur de coordonnées (i, j) prise de la matrice de cooccurrence.

2.5.4.3 Algorithme général de Calcul des images texturales

Les étapes principales de notre algorithme sont décrites dans la figure Fig.2.14.

```

Algorithme d'extraction des images texturales
Début
  Entrée d'une image MIAS ROI de mammographie IMG;
  Entrée des paramètres de cooccurrence d,θ,TM,DT; /*distance,direction,taille du masque,Descripteur de texture*/
  Pour toute l'image IMG
    Pour chaque pixel (i,j) ∈ au masque
      Calculer la matrice de cooccurrence correspondante GLCM;
      Calculer DT;
      IMG (i,j) ← DT;
    Fin
  Fin
Fin

```

Fig.2.14. Algorithme d'extraction des images texturales.

2.5.4.4 Post-traitements :

Certaines images texturales ainsi extraites présentent des contours des anomalies parfois légèrement « discontinus », comme le montre la figure (Fig. 2.25). Mais dans la plupart des images, les contours de la masse ont été détectés, tel que l'exemple de la figure

(Fig. 2.24). Seulement, et dans le but de comparer ces contours avec les annotations de l'expert, il a fallu procéder à la fermeture des parties « discontinues » des contours.

Pour des travaux similaires, concernant la segmentation de la masse par détection de contours, les auteurs (*Kavitha & Rejusha, 2015*) utilisent le filtre de Kalman pour affiner le contour initial discontinu, une fois que le contour affiné est obtenu, la masse est détectée en utilisant la distribution de gamma, ensuite le contour est fermé en utilisant le modèle de contour actif basé région pour obtenir le résultat segmenté. Ou encore les auteurs (*Guo, Dong, Wang, Ma, & Gu, 2015*) qui utilisent uniquement les modèles de contours actifs pour fermer les contours discontinus.

Dans notre étude, des fonctions qui prennent en considération les caractéristiques géométriques d'un objet, ont été utilisés. L'objet étant la région suspecte détectée, il s'agit de prendre le centre de gravité de la région intérieure au contours détectés puis de calculé l'aire de la surface de cette région.

Ainsi les coefficients d'évaluation de la détection des contours ont pu être calculés. Ces derniers sont le Dice Coefficient, les coefficients de « Precision », « Recall » et de « F-measure ».

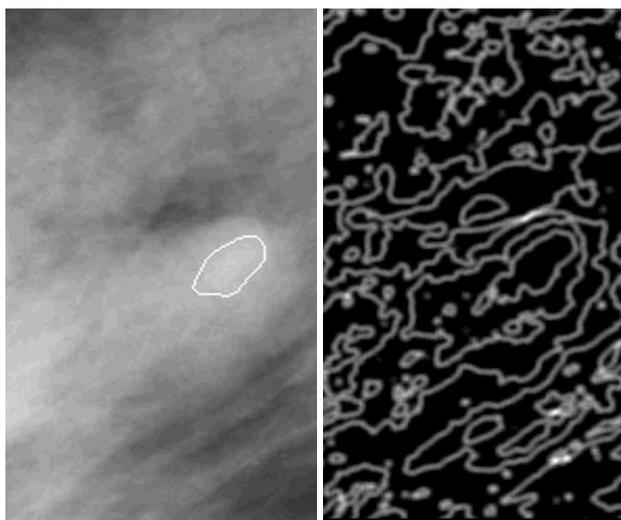


Fig. 2.24. Vue des délimitations de l'expert (à droite), et des contours détectés par l'approche (à gauche).

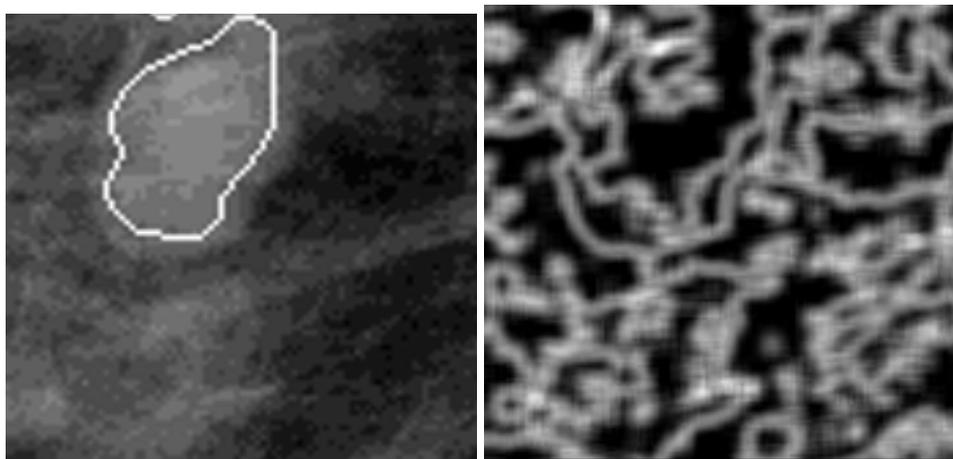


Fig. 2.25. Vue des délimitations de l'expert (à droite), et des contours détectés par l'approche (à gauche).

2.6. Résultats de l'approche de détection des contours des masses par le Descripteur GLCM de Contraste

2.6.1 Détection et segmentation des contours des masses dans des tissus de différentes densités

L'approche proposée a été appliquée sur les images de la base de données MIAS représentant des données « BIRADS ». Seules les 38 images contenant les anomalies de masses sont utilisées. Comme cité précédemment, les tests ont été effectués sur des images de mammographie représentant différentes densités de tissus (Gras, Dense, Glandulaire). Pour

chaque image, le descripteur de contraste est calculé et les images texturales sont extraites. Ces images où les contours de la masse ont été détectés et identifiés. Ensuite, les post-traitements ont été appliqués pour extraire la masse détectée.

Dans la figure (Fig.2.15), des exemples d'images MIAS, mdb004, mdb005 et mdb019 qui représentent respectivement: (Fig.2.15.a), image avec un tissu dense, ce tissu apparaît selon des régions claires en blanc sur l'image de mammographie; (Fig2.15.b), image avec un tissu gras, les régions sont gris, vers le foncé sur les images; et (Fig.2.15.c), image avec un tissu glandulaire, les régions sont de couleurs mixtes, entre le blanc et le gris, gris foncé sur les images de mammographie.

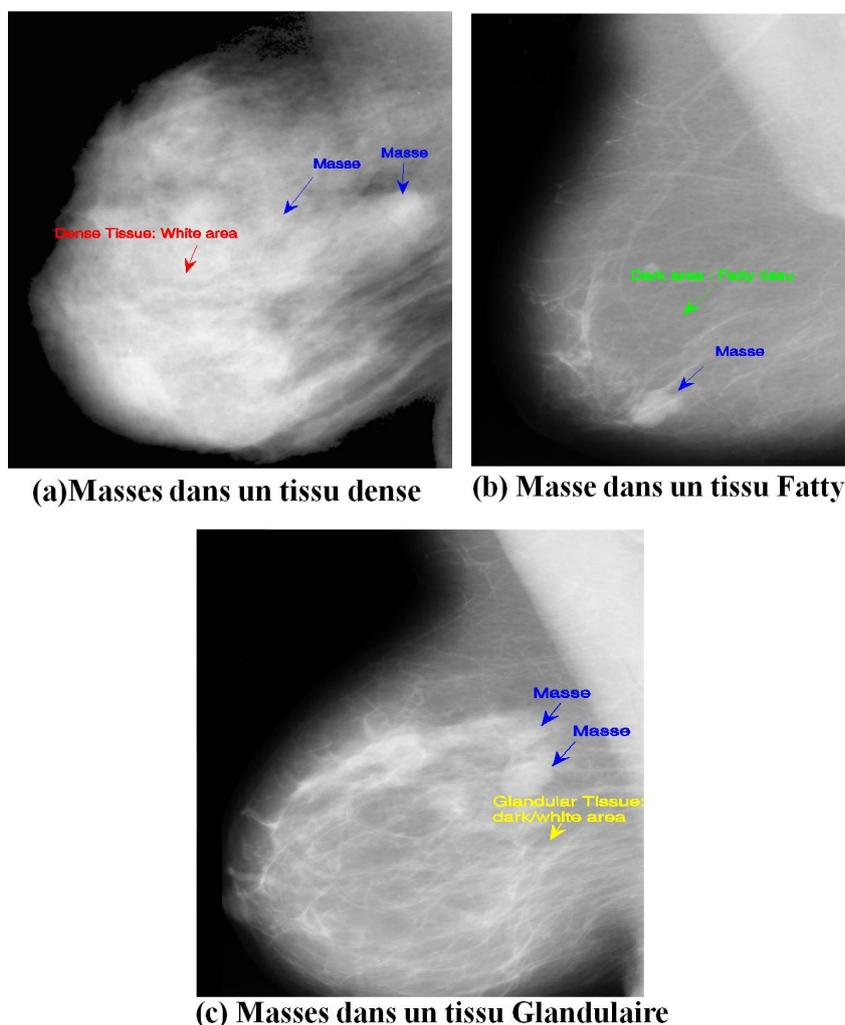


Fig.2.15. (a) mdb004, deux masses (flèche bleu) dans un tissu dense (flèche rouge), (b) mdb005, une masse (flèche bleu) dans un tissu gras (flèche vert), (c) mdb019, deux masses (flèche bleu) dans un tissu glandulaire (flèche jaune).

Quelques résultats finaux de l'approche proposée sont donnés dans les figures Fig.2.16, Fig.2.17 et Fig.2.18. Dans chacune de ces figures, de gauche à droite, l'image ROI, l'image du descripteur de texture, puis l'image de segmentation de la masse.

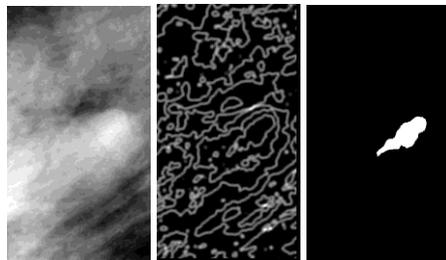


Fig.2.16. Détection et segmentation de la masse dans un Tissu Dense, sur mdb004 Roi image.

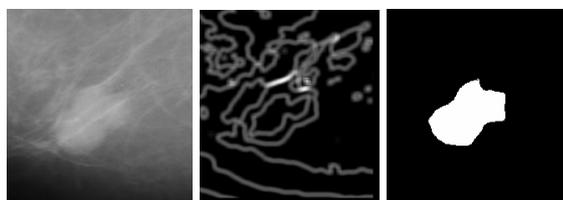


Fig.2.17. Détection et segmentation de la masse dans un Tissu Gras, sur mdb005 ROI image.

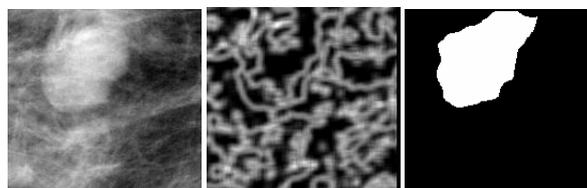


Fig.2.18. Détection et segmentation de la masse dans un Tissu Glandulaire, sur mdb019 Roi image.

2.6.2 Evaluation Quantitative de la détection des contours des masses

Pour l'évaluation de l'approche de détection et de segmentation des masses par contours, l'annotation de l'expert est prise comme base de comparaison. Rappelons que cette annotation produit l'emplacement et les contours tracés par des experts radiologues.

Pour évaluer l'approche nous avons procédé de deux manières :

En premier lieu, le Dice Coefficient est calculé pour comparer les résultats obtenus avec des travaux similaires, des travaux où les auteurs n'ont donné que ce coefficient comme évaluation finale de leurs approches.

En second lieu, les coefficients de « Precision », « Recall » et de « F-mesure », sont calculés. Ces derniers permettent de calculer les courbes ROC et ainsi de donner l'aire sous ces courbes par le coefficient Az (Area under the curve).

2.6.2.1 Evaluation Quantitative par le Coefficient du Dice

2.6.2.1.1. Présentation du coefficient de Dice

Le coefficient de Dice est une mesure de similarité. Ce coefficient est utilisé en statistique pour déterminer la similarité entre deux échantillons.

Le coefficient est toujours compris entre 0 et 1. Il se définit comme le double de l'intersection de deux lots (échantillons de valeurs) divisé par l'union de ces deux lots. La formule ci-dessous résume la méthode de calcul:

$$S = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} \quad (2.16)$$

Dans cette formule, X et Y sont des Classes. Pour calculer le coefficient de Dice de deux régions, il est possible de définir X et Y comme l'ensemble des pixels de chacune des régions respectives.

Dans cette étude, ce coefficient est exprimé en pourcentage.

2.6.2.1.2. Evaluation de la détection et de la segmentation des masses par le Dice coefficient

La masse segmentée est comparée à la segmentation établie par l'expert (annotations des experts sur la base MIAS). Pour ce faire, le coefficient du Dice est calculé.

L'approche proposée a été testée sur les 38 images contenant des anomalies de masses, de la base de données MIAS. Quelques exemples types et représentatifs de nos tests sont

présentés ci-dessous : un cas où la masse est entourée d'un tissu dense dans la figure (Fig.2.19), un cas où la masse est entourée d'un tissu gras dans la figure (Fig.2.20) et un cas où la masse est entourée d'un tissu glandulaire dans la figure (Fig.2.22).

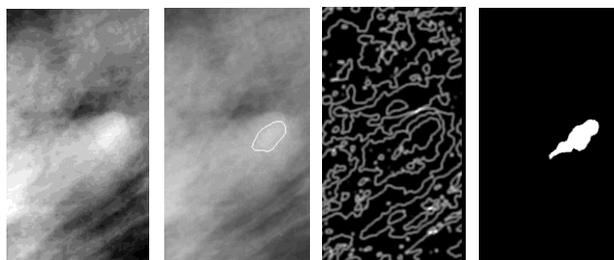


Fig. 2.19. Tissu dense: de gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff = 89.39%.

La figure Fig.2.19 montre le cas du tissu dense qui entoure complètement la masse. Un bon pourcentage de ressemblance avec la segmentation de l'expert radiologue est obtenu, Dice Coefficient = 89.39%. Ce qui aidera certainement l'expert à donner une meilleure interprétation de la forme de l'anomalie ainsi détectée.

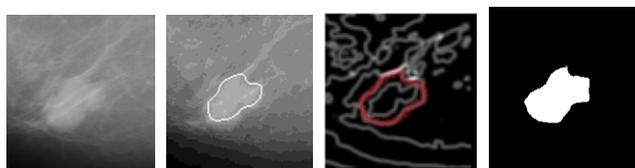


Fig.2.20. Tissu Gras: de gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff. = 91.74%

Les figures Fig.2.20, Fig.2.21, montrent le cas du tissu gras qui entoure la masse. Ici, l'approche donne deux possibilités d'interpréter les contours de la masse. En effet, en comparant avec la segmentation de l'expert, la méthode a détecté d'autres contours, à l'intérieur des contours définis par l'expert. Ces résultats seront utiles à l'expert, ils représentent une information supplémentaire pour décider de la forme et de la taille les plus exactes de la masse.

Si le coefficient du Dice est calculé en suivant les délimitations de l'expert, le coefficient sera évidemment meilleur, même satisfaisant, Dice coefficient = 91.74% (Fig.2.20).

Sinon, il sera moins satisfaisant Dice coefficient = 69.35%. La figure (Fig.2.21) montre le cas où le coefficient est calculé à partir des contours détectés à l'intérieur des annotations de l'expert.

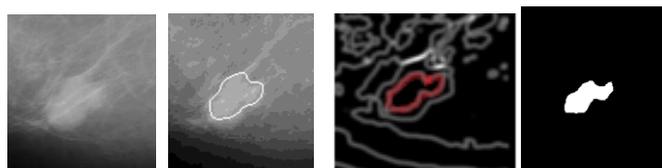


Fig.2.21. De gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff. = 69.35%.

Pour palier à ce problème, des réglages de paramètres ont été effectués, en se basant au départ, sur un préapprentissage à l'aide d'une séquence d'images MIAS.

Les figures Fig.2.22 et Fig.2.23 montrent le cas où le tissu glandulaire entoure la masse. Le même commentaire fait pour les figures Fig.2.20 et Fig.2.21, est aussi fait pour cet exemple. D'autres contours ont été détectés, à l'intérieur des annotations de l'expert.

La figure Fig.2.23 montre le cas où les contours sont pris à l'intérieur des contours délimités par l'expert, Dice coefficient = 86.18%.

La figure Fig.2.22 montre le cas où les contours de la masse sont similaires aux délimitations de l'expert. Le coefficient de ressemblance est meilleur Dice coefficient = 90.67%.

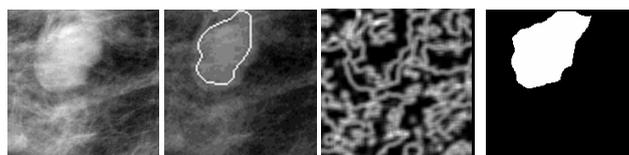


Fig. 2.22. Tissu Glandulaire: de gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff. = 90.67%.

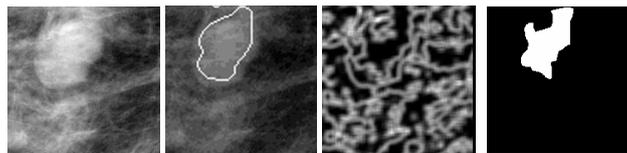


Fig.2.23. De gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff. = 86.18%.

Le tableau ci-dessous récapitule ces résultats de l'approche texturale proposée :

Images	Coefficient du Dice
Masse entourée d'un tissu dense (mdb004)	89.39%.
Masse entourée d'un tissu gras (mdb005)	91.74%
Masse entourée d'un tissu glandulaire (mdb019)	90.67%.

Tab.2.1. Evaluation de la détection des contours des masses à l'aide du Coefficient du Dice.

2.6.2.2 Evaluation Quantitative par le F-measure:

Une autre évaluation de la méthode proposée est donnée, c'est le calcul de l'aire sous les courbes ROC par le coefficient Az.

Cette méthode est plus générale et plus complexe que celle du coefficient du Dice. Pour calculer le coefficient Az, il est nécessaire de calculer plusieurs Taux de la détection et/ou segmentation de la masse. Ces derniers sont les taux des Vrais Positifs (TP, True Positif Rate), les taux des Faux Positifs (FP, False Positif Rate), les taux des Vrais Négatifs (TN, True Négatif Rate) et enfin le taux des Faux Négatifs (FN, False Négatif Rate).

Nous quantifions les *TP*, *FP*, *TN* et *FN* comme:

- *TP: True Positive*, la région des pixels a été segmentée comme masse et qui s'est avérée être la masse.

- *FP: False Positive*, la région des pixels a été segmentée comme masse et qui s'est avérée ne pas être la masse.
- *FN: False Negative*, la région des pixels a été segmentée comme non masse et qui s'est avérée être la masse.
- *TN: True Negative*, la région des pixels a été segmentée a comme non masse et qui s'est avérée ne pas être la masse.
- *TPR: True Positive Rate*, Taux des *TP*.
- *FPR: False Positive Rate*, Taux des *FP*.

$$TPR = \frac{TP}{TP + FN} \quad (2.17)$$

$$FPR = \frac{FP}{FP + TN} \quad (2.18)$$

Par la suite les coefficients de Precision, Recall et de Specificity, sont calculés. Ces derniers ainsi que le coefficient de Sensitivity Eq. (2.20) et le coefficient de Specificity Eq. (2.21), sont nécessaires pour calculer l'aire sous la courbe Az.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.19)$$

$$\text{Recall} = \text{Sensitivity} = TPR \quad (2.20)$$

$$\text{Specificity} = 1 - FPR \quad (2.21)$$

$$F - \text{measure} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)} \quad (2.22)$$

La table (Tab.2.2) montre les valeurs des coefficients de Precision Eq. (2.19), Recall Eq. (2.20) et F-measure Eq. (2.22) obtenues à partir des exemples d'images types citées auparavant et représentant la détection de la masse dans des densités de tissu différentes.

Images	Precision	Recall	F-measure
mdb004, masse dans un tissu dense	0,8978	0,8933	0,8956
mdb005, masse dans un tissu gras	0,8983	0,8781	0,8881
mdb019, masse dans un tissu Glandulaire	0,8997	0,8375	0,8676

Tab.2.2. Evaluation Quantitative de la détection contours des masses à l'aide du F-Mesure.

2.6.3 Premières comparaisons avec des travaux similaires

Comme cité dans le paragraphe conclusion du chapitre 1, il est très difficile de faire une comparaison de l'efficacité des méthodes de détection et/ou de segmentation des anomalies de masses. La raison est que lors du développement de chacune des méthodes, les auteurs ont utilisé des critères différents. Ces derniers ont été résumés au chapitre 1, mais il est important de mentionner que le critère concernant les données utilisées est un paramètre crucial pour l'obtention de résultats significatifs.

En effet les résultats ne seront plus comparables si les auteurs utilisent des bases de données différentes, un nombre d'images de test différent, un contenu d'images différent (que des masses malignes ou que des masses bénignes), ou des vues d'images différentes (une vue ou des vues multiples de l'image).

Néanmoins, l'approche proposée, de détection et de segmentation de la masse par contours, a été comparée avec des travaux similaires d'Arnau Oliver (Arnau, et al., 2010). En premier lieu, ces auteurs ont récemment résumé tous les travaux des méthodes de détection et/ou de segmentation des anomalies de masse.

En second lieu, ils ont testé sept méthodes: 1: méthode basée sur la détection des couches concentriques, 2: méthode basée sur l'approche Laplacian de détection de bords, 3: méthode basée sur le seuillage, 4: méthode basée sur le filtre Iris, 5: méthode basée sur la Différence du Gaussien, 6: méthode basée sur une approche d'appariement de formes, 7: méthode basée sur une approche de classificateur.

Et les résultats de ces méthodes ont été donnés selon trois critères: 1. Influence du tissu mammaire, 2. Influence de la forme de la lésion, 3. Influence de la taille de la lésion. Deux bases de données ont été utilisées à cet effet, « Trueta database » et Mias database ». Pour chacune, un ensemble de données a été utilisé comme des données d'entraînement et un autre ensemble de données comme des données de test. Et pour chaque ensemble, la valeur du coefficient Az a été donnée.

Pour cette étude, seuls les travaux appliqués sur les mêmes données d'images MIAS, ont été considérés pour une comparaison. Il en ressort que les valeurs du coefficient Az qui ont été obtenues varient entre $Az = 0.751$ et $Az = 0.780$ (Arnau, et al., 2010). Pour la méthode proposée dans ce travail, et pour les 38 images tests, le coefficient obtenu est $Az = 0.79$.

2.7. Discussion et Conclusion

L'étape de détection et/ou segmentation des anomalies de cancer dans des images issues de mammographie, est une étape clé qui détermine la performance et le degré de fiabilité des systèmes CAD. Seulement, le faible contraste de ces images et la complexité des types de tissu se trouvant dans la région mammaire (aussi bien visuellement que quantitativement), et pouvant couvrir complètement l'anomalie, font que jusqu'à présent, la tâche la plus difficile pour les chercheurs dans ce domaine est réellement la discrimination entre le tissu mammaire normal et le tissu contenant une anomalie cancéreuse. Plus le tissu est dense, plus la confusion augmente.

Dans cette étude, nous avons contribué en présentant une approche d'identification des masses par la détection de leurs contours. La discrimination entre le tissu normal le tissu contenant l'anomalie est obtenue en utilisant les images texturales du descripteur de contraste extrait des matrices de cooccurrence de niveaux de gris GLCM.

Les résultats du coefficient du Dice, des coefficients de F-Measure, de Precision, de Recall et de l'aire sous la courbe (Az) sont satisfaisants et l'approche proposée peut être améliorée, notamment au niveau de l'étape des post-traitements.

En résumé, nous pouvons confirmer que l'information de texture reste une information importante et nécessaire pour enlever des confusions entre le tissu normal et le tissu

cancéreux. Seulement, jusqu'à présent cette information est difficilement exprimée, il n'existe pas de formalisme mathématique pour définir exactement une texture, ce qui fait que cet axe de recherche reste ouvert à toute contribution scientifique pouvant aider à donner ce formalisme.

Chapitre 3 : Les Modèles de Mixture des Gaussiennes (GMM) Pour la Segmentation des Anomalies de Masse sur des Images de Mammographie.

Résumé

Dans ce chapitre, le critère « intensité des pixels » des images de mammographie est utilisé, et une méthode semi-automatique de segmentation des anomalies de masses est proposée. Elle est basée sur le Modèle de Mélange des Gaussiennes de mélange (GMM - Gaussian Mixture Model). L'utilisation du modèle GMM dans des problèmes de détection et/ou de segmentation des anomalies cancéreuses, n'est pas nouvelle en soi. En effet, le modèle GMM fournit des informations importantes sur les distributions de probabilités d'intensités des images de mammographie. Nous avons voulu contribuer en développant une approche d'utilisation de ce modèle. La méthode proposée dans ce travail peut être subdivisée en deux étapes. Dans la première étape, les statistiques et les formes des masses, ainsi que les statistiques sur le tissu normal, sont "appries" afin de déterminer des paramètres importants pour le modèle GMM. Ces statistiques sont calculées par l'algorithme K-moyen. Dans la seconde étape, ces statistiques sont incorporées pour effectuer une segmentation des masses en utilisant la méthode de maximisation des espérances (EM : Expectation Maximisation) et la méthode du maximum de vraisemblance (ML). L'EM est un algorithme qui calcule et ajuste au mieux les paramètres du modèle GMM. Ainsi les statistiques « apprises » représentent un apprentissage du modèle pour aboutir à une segmentation des tissus mammaires en tissu normal et tissu suspect. L'algorithme proposé a été appliqué sur des images de mammographie de la base MIAS, à différentes densités de tissus. Ensuite, la segmentation des masses obtenue par l'algorithme proposé, est comparée à la segmentation effectuée par un expert en calculant le Dice Coefficient.

3.1. Introduction

Le Chapitre présente d'abord les définitions principales des Modèles de Mixture des Gaussiennes dans la littérature, ainsi que le principe de l'algorithme EM, celui de l'algorithme ML et celui de l'algorithme K-means. L'approche proposée pour la détection et la segmentation des masses est ensuite donnée, ainsi que les données utilisés à cette fin.

3.2. Les Modèles GMM de base

3.2.1. Principe des Modèles GMM

Un modèle de mélange gaussien, GMM (*Gaussian Mixture Model*) est un modèle statistique exprimé selon une densité de mélange. Cette dernière sert usuellement à estimer de manière paramétrique la distribution de variables aléatoires, en les modélisant comme une somme de plusieurs gaussiennes (appelées *noyaux*). Il s'agit alors de déterminer la variance, la moyenne et l'amplitude de chaque gaussienne. Ces paramètres sont optimisés selon un critère de maximum de vraisemblance pour approcher le plus possible la distribution recherchée. Cette procédure se fait le plus souvent itérativement via l'algorithme Espérance-Maximisation (*EM*).

Les Modèles de Mélange Gaussien sont réputés pour reconstruire de manière particulièrement efficace les données manquantes dans un jeu de données expérimentales.

3.2.2. Exemple d'un Modèle GMM :

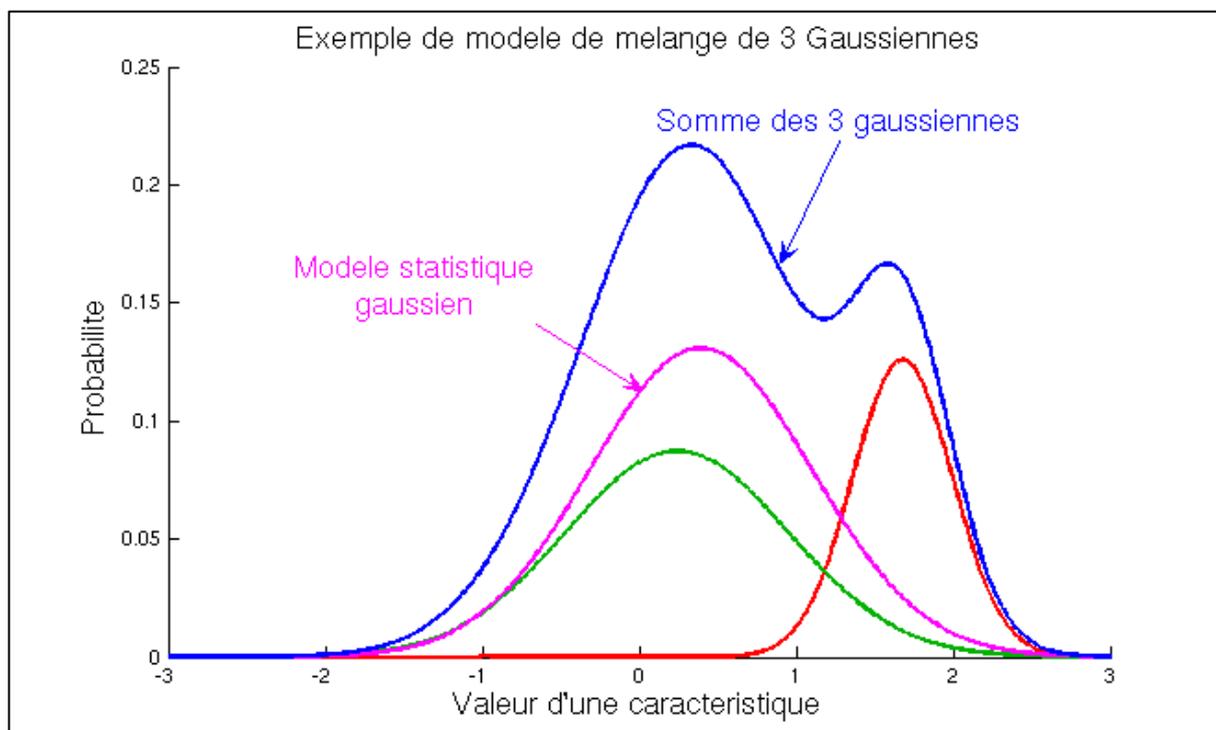


Fig.3.1. Un exemple de GMM à 3 gaussiennes

La figure (Fig.3.1) présente 3 gaussiennes (courbes de couleurs : mauve, vert et rouge), et le mélange de ces dernières sera leur somme (courbe en bleu).

Le principe de GMM est d'adapter l'histogramme de l'image avec un mélange de gaussiennes en utilisant l'algorithme Expectation-Maximisation (EM) (Dempster et al. 1977), ou encore nommé l'algorithme d'Espérance-Maximisation. Le nombre de gaussiennes, qui correspond au nombre de modes de l'histogramme, doit être fixé 'a priori'.

On utilise souvent l'algorithme EM pour la classification de données (Tang, Rangayyan, Xu, El Naqa, & Yang, 2009), (Dempster, Laird, & Rubin, 1977), l'apprentissage automatique (Tang, Rangayyan, Xu, El Naqa, & Yang, 2009), (Dempster, Laird, & Rubin, 1977), ou la vision artificielle (Tang, Rangayyan, Xu, El Naqa, & Yang, 2009). On peut également citer son utilisation en imagerie médicale dans le cadre de la reconstruction tomographique par exemple (Mariano-Goulart, 2013(a)),(Mariano-Goulart, 2013(b)).

L'algorithme d'Espérance-Maximisation comporte :

- une étape d'évaluation de l'espérance (E), où l'on calcule l'espérance de la vraisemblance en tenant compte des dernières variables observées.
- une étape de maximisation (M), où l'on estime le maximum de vraisemblance des paramètres en maximisant la vraisemblance trouvée à l'étape E.

On utilise ensuite les paramètres trouvés à l'étape de maximisation M comme point de départ d'une nouvelle phase d'évaluation de l'espérance, et l'on itère ainsi le processus.

3.2.3. Principe de l'algorithme d'Espérance-Maximisation

En considérant un échantillon $x = (x_1, x_2, \dots, x_n)$ d'individus suivant une loi $f(x_i, \theta)$ paramétrée par θ , on cherche à déterminer le paramètre θ maximisant la log-vraisemblance donnée par:

$$L(x; \theta) = \sum_{i=1}^n \log f(x_i, \theta) \quad (3.1)$$

Cet algorithme est particulièrement utile lorsque la maximisation de L est très complexe mais que, sous réserve de connaître certaines données judicieusement choisies, on peut très simplement déterminer θ .

Dans ce cas, on s'appuie sur des données complétées par un vecteur $z = (z_1, \dots, z_n)$ inconnu. En notant $f(z_i | x_i; \theta)$ la probabilité de z_i sachant x_i et le paramètre θ , on peut définir la log-vraisemblance complétée comme la quantité :

$$L((x, z); \theta) = \sum_{i=1}^n (\log f(z_i | x_i, \theta) + \log f(x_i; \theta)) \quad (3.2)$$

Et donc,

$$L(x; \theta) = L((x, z); \theta) - \sum_{i=1}^n \log f(z_i | x_i, \theta) \quad (3.3)$$

L'algorithme EM est une procédure itérative basée sur l'espérance des données complétées conditionnellement au paramètre courant. En notant $\theta^{(c)}$ ce paramètre, on peut écrire :

$$E[L(x; \theta) | \theta^{(c)}] = E[L((x, z); \theta) | \theta^{(c)}] - E\left[\sum_{i=1}^n \log f(z_i | x_i, \theta) | \theta^{(c)}\right] \quad (3.4)$$

Où l'espérance est prise sur z .

Ou encore :

$$L(x; \theta) = Q(\theta; \theta^{(c)}) - H(\theta; \theta^{(c)}) \quad (3.5)$$

Car $L(x; \theta)$ ne dépend pas de z .

Avec :

$$Q(\theta, \theta^{(c)}) = E[L((x, z); \theta) | \theta^{(c)}] \quad (3.6)$$

Et :

$$H(\theta, \theta^{(c)}) = E\left[\sum_{i=1}^n \log f(z_i | x_i, \theta) | \theta^{(c)}\right] \quad (3.7)$$

On montre que la suite définie par :

$$\theta^{(c+1)} = \arg \max_{\theta} \left(Q(\theta, \theta^{(c)}) \right) \quad (3.8)$$

Fait tendre $L(x; \theta^{(c+1)})$ vers un maximum local.

On peut donc définir l'algorithme EM de la manière suivante:

- Initialisation au hasard de $\theta^{(0)}$
- $c=0$
- Tant que l'algorithme n'a pas convergé, faire:
 - Evaluation de l'espérance (étape E):

$$Q(\theta, \theta^{(c)}) = E \left[L((x, z); \theta) | \theta^{(c)} \right] \quad (3.9)$$

- Maximisation (étape M)

$$\theta^{(c+1)} = \arg \max_{\theta} \left(Q(\theta, \theta^{(c)}) \right) \quad (3.10)$$

- $c=c+1$
- Fin

En pratique, pour s'affranchir du caractère local du maximum atteint, on fait tourner l'algorithme EM un grand nombre de fois à partir de valeurs initiales différentes de manière à avoir de plus grandes chances d'atteindre le maximum global de vraisemblance.

3.2.4. Principe de l'algorithme K-moyennes

L'algorithme des k-moyennes (ou K-means) est un algorithme de partitionnement de données relevant des statistiques et de l'apprentissage automatique (plus précisément de l'apprentissage non supervisé). C'est une méthode dont le but est de diviser des observations en K partitions (clusters) dans lesquelles chaque observation appartient à la partition avec la

moyenne la plus proche. Les nuées dynamiques sont une généralisation de ce principe, pour laquelle chaque partition est représentée par un noyau pouvant être plus complexe qu'une moyenne.

3.2.4.1. Description

Étant donné un ensemble d'observations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, où chaque observation est un vecteur de dimension d , l'algorithme k -means de regroupement vise à partitionner les n observations dans k ensembles $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$ ($k \leq n$) afin de minimiser la distance entre les points à l'intérieur de chaque partition :

$$\arg \min_s \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (3.11)$$

Où μ_i est la moyenne des points dans S_i .

3.2.4.2. L'algorithme K-means: Algorithme standard

- Choisir k points qui représentent la position moyenne des partitions $m_1^{(1)}, \dots, m_k^{(1)}$ initiales (au hasard par exemple)
- Répéter jusqu'à convergence :
 - assigner chaque observation à la partition la plus proche (*i.e* effectuer une partition de Voronoï selon les moyennes).

$$S_i^{(t)} = \left\{ x_j : \|x_j - m_i^{(t)}\| \leq \|x_j - m_{i^*}^{(t)}\| \forall i^* = 1, \dots, k \right\} \quad (3.12)$$

- mettre à jour la moyenne de chaque cluster

$$m_i^{(t+1)} = \frac{1}{S_i^{(t)}} \sum_{x_j \in S_i^{(t)}} x_j \quad (3.13)$$

La convergence est atteinte quand il n'y a plus de changement.

3.2.4.3. Convergence

Il y a un nombre fini de partitions possibles à k classes. De plus, chaque étape de l'algorithme fait strictement diminuer la fonction de coût, positive, et fait découvrir une meilleure partition.

Cela permet d'affirmer que l'algorithme converge toujours en temps fini, vers un minimum local.

Malgré cela, l'algorithme peut être lent à converger, et les auteurs conseillent donc de rajouter un critère d'arrêt lors de son implémentation tel que le nombre d'étapes limité (souvent faible, de l'ordre de 10 pour 2 ou 3 classes sur des ensembles de point 1x2 de taille 200) (wikipedia).

3.2.4.4. Avantages et inconvénients

Un inconvénient de cet algorithme est que les 'clusters' dépendent de l'initialisation et de la distance choisie.

Le fait de devoir choisir à priori le paramètre K peut être perçu comme un inconvénient ou un avantage. Au contraire, dans certains partitionnements de données on préférera s'affranchir d'une telle contrainte.

3.2.5. Estimation du Maximum de Vraisemblance

L'estimation du Maximum de Vraisemblance est une méthode statistique courante utilisée pour inférer les paramètres de la distribution de probabilité d'un échantillon donné. Cette méthode a été développée par le statisticien Ronald Aylmer Fisher en 1922.

3.2.5.1. Principe

Soit une famille paramétrée de distributions de probabilités D_θ dont les éléments sont associés soit à une densité de probabilité connue (distribution continue), soit à une fonction de masse connue (distribution discrète), notée f_θ . On tire un échantillon de n valeurs x_1, x_2, \dots, x_n de la distribution, et l'on calcule la densité de probabilité associée aux données observées :

$$f_\theta(x_1, \dots, x_n | \theta) \quad (3.14)$$

Ceci étant une fonction de θ avec x_1, \dots, x_n fixés, c'est une vraisemblance.

$$L(\theta) = f_{\theta}(x_1, \dots, x_n | \theta) \quad (3.15)$$

Lorsque θ n'est pas observable, la méthode du maximum de vraisemblance utilise les valeurs de θ qui maximisent $L(\theta)$ estimateur de θ : c'est l'*estimateur du maximum de vraisemblance* de θ noté $\hat{\theta}$. Par exemple dans le cas du produit discret, on effectue un tirage de n valeurs, il faut donc trouver le paramètre qui maximise la probabilité d'avoir tiré ce tirage.

Cette méthode se distingue de la recherche d'un estimateur non biaisé de θ , ce qui ne donne pas nécessairement la valeur la plus probable pour θ .

L'estimateur du maximum de vraisemblance peut exister et être unique, ne pas être unique, ou ne pas exister.

3.2.5.2. Généralisation

Pour une variable aléatoire réelle X de loi quelconque définie par une fonction de répartition $F(x)$, on peut considérer des petits voisinages V autour de (x_1, \dots, x_n) dans \mathbb{R}^n , par exemple une boule de rayon ε . On obtient ainsi une fonction de vraisemblance :

$$L(\theta; V) = P \left[(X_{1,\theta}, \dots, X_{n,\theta}) \in V \right] \quad (3.16)$$

dont on cherche un maximum :

$$\theta = \hat{\theta}(V) \quad (3.17)$$

On fait ensuite tendre la taille de V vers 0 dans $\hat{\theta}(V)$ pour obtenir l'estimateur $\hat{\theta}$ de maximum de vraisemblance.

On retombe sur les fonctions de vraisemblance précédentes quand X est à loi discrète ou continue.

3.2.5.3. Propriétés

L'estimateur obtenu par la méthode du maximum de vraisemblance est:

- convergent
- Asymptotiquement efficace, il atteint la borne de Cramer-Rao.
- asymptotiquement distribué selon une loi normale

En revanche, il peut être biaisé en échantillon fini.

3.2.5.4. Loi normale

L'estimateur du maximum de vraisemblance de l'espérance μ et la variance σ^2 d'une loi normale est:

$$\mu_{ML} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.18)$$

$$\sigma^2_{ML} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.19)$$

Démonstration

L'estimateur de la variance est un bon exemple pour montrer que le maximum de vraisemblance peut fournir des estimateurs biaisés : un estimateur sans biais est donné en effet par :

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.20)$$

Néanmoins, asymptotiquement, quand n tend vers l'infini, ce biais, qui est de :

$$\frac{n}{n-1} \quad (3.21)$$

, tend vers 1 et l'estimateur est alors asymptotiquement sans biais.

3.2.5.5. Fonction de vraisemblance

La fonction de vraisemblance, notée $L(x_1, \dots, x_n | \theta_1, \dots, \theta_k)$ est une fonction de probabilités conditionnelles qui décrit les valeurs x_j d'une loi statistique en fonction des paramètres θ_i supposés connus.

Elle s'exprime à partir de la fonction de densité $f(x|\theta)$ par :

$$L(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (3.22)$$

Avec

$$f(x; \theta) = \begin{cases} f_\theta(x) & \text{si } X \text{ est une var. aléatoire continue} \\ P_\theta(X = x) & \text{si } X \text{ est une var. aléatoire discrète} \end{cases} \quad (3.23)$$

$f_\theta(x)$ est la fonction qui représente la densité de X (où θ apparaît) et $P_\theta(X = x)$ représente une probabilité discrète (où θ apparaît).

Cette formule n'est valable que si on suppose que les x_i sont indépendants et identiquement distribués entre eux.

3.3. Le Modèle GMM/ML proposé pour la segmentation des anomalies de masses

Dans le paragraphe § 2.5.1 du chapitre 2, les problèmes liés à la détection des masses ont été soulevés et résumés. L'approche par les modèles GMM proposée est une seconde contribution pour résoudre quelque uns de ces problèmes tels que, les confusions entre le tissu contenant la masse et le tissu normal, et la nécessité de détecter la forme de la masse.

En effet, la force majeure du modèle GMM est l'information donnée "a priori" par les paramètres initiaux à son approche itérative (McLachlan & Basford, 1988), (Dempster, Laird, & Rubin, 1977). Dans ce travail, une contribution a été apportée au Modèle GMM en proposant une étape d'apprentissage du Modèle (Learning the GMM). L'idée est d'introduire des informations sur les tissus mammaires normaux et des informations sur les tissus cancéreux dans les paramètres initiaux des modèles GMM.

L'étape d'apprentissage sera donc définie par l'initialisation des statistiques de la région cancéreuse (anomalie de masse) et l'initialisation des statistiques de la région normale (tissu normal), dans le processus itératif du Modèle GMM. Dans ce sens, la méthode proposée peut être divisée en deux étapes.

En premier, les statistiques et formes de la masse sont estimées pour déterminer d'importants paramètres. Ils sont extraits de la région ROI, et représentent le nombre de regroupements de régions à statistiques semblables, donné par l'algorithme K-means. Les statistiques sur le tissu normal sont représentées par la distribution dans les histogrammes correspondants à ce tissu.

En second, ces statistiques sont « apprises » et incorporées pour améliorer la segmentation de la masse, en utilisant l'algorithme d'Espérance-Maximisation (EM) avec l'algorithme du maximum de vraisemblance ML. Ainsi, l'EM estimera le meilleur ajustement de ces paramètres aux paramètres du modèle GMM.

L'algorithme proposé est appelé modèle GMM/ML. Et le résultat est une image segmentée en régions contenant la masse et en régions ne contenant pas de masse.

Il a été appliqué sur un ensemble de données de la base MIAS décrite au paragraphe § 2.4 du chapitre 2. Les images utilisées sont des cas où la masse est présente dans un tissu dense essentiellement, des cas où le tissu est gras et des cas où il est glandulaire. Les résultats de la segmentation seront comparés aux annotations de l'expert par le biais de la mesure du coefficient du Dice.

Les étapes de l'approche sont détaillées dans le paragraphe suivant.

3.3.1 Les principales étapes du Modèle GMM/ML

Le modèle GMM/ML est inspiré du modèle GMM qui se base sur le principe du Maximum de Vraisemblance (Maximum Likelihood (ML)). Quatre principales étapes définissent l'approche proposée :

1. L'étape d'amélioration du contraste des images de la base de données MIAS de mammographie.
2. L'extraction de la région ROI d'intérêt.
3. L'étape d'apprentissage du Modèle par des statistiques apprises (K-mean).
4. L'étape d'itération de l'algorithme d'Expectation-Maximisation par le Maximum de Vraisemblance, pour obtenir l'image segmentée.

Le schéma de la figure Fig.3.2 résume ces étapes.

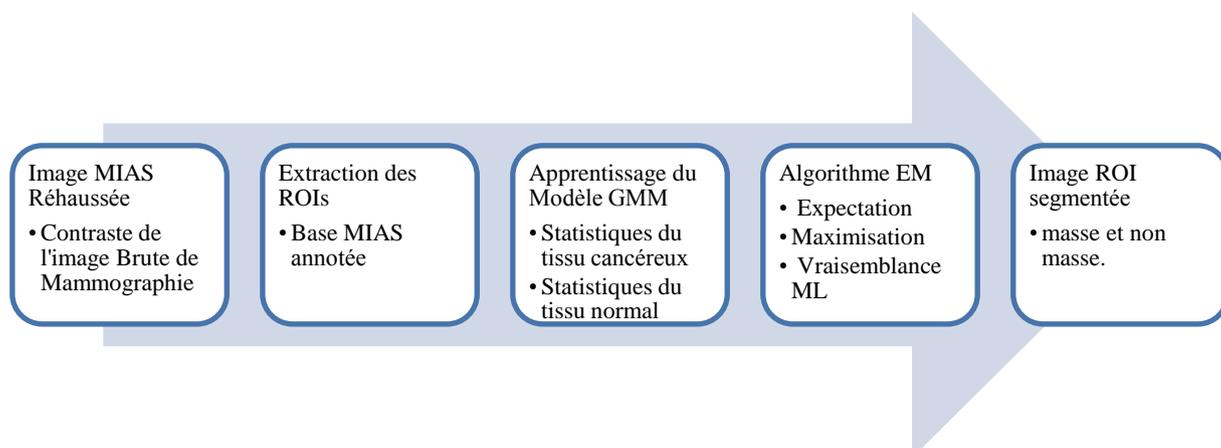


Fig.3.2. Les principales étapes du Modèle GMM/ML

3.3.2. Amélioration des images de mammographies

Toujours dans un but de donner plus de performance à la segmentation de la masse, et pour les mêmes motivations citées dans le paragraphe § 2.5.4.1 du chapitre 2, la même approche a été utilisée pour l'amélioration du contraste des images de la base de données MIAS de mammographie:

1. Tout d'abord un lissage a été appliqué sur chaque image pour la « débruiter » grâce au filtre de diffusion SRAD (Speckle Reducing Anisotropic Diffusion) (Yu & Acton, 2002).
2. En deuxième lieu, grâce à la méthode d'Egalisation d'Histogramme Adaptative Limitée au contraste (CLAHE), une augmentation du contraste de chaque image est réalisée. Plus de détails ont été donnés au paragraphe §2.5.4.1 du chapitre 2.

3.3.3. Extraction des images « ROIs »

Comme cité au paragraphe §2.5.4.1 du chapitre 2, et une fois les images de mammographies améliorées, les régions ROIs ont été extraites grâce aux annotations de la base de données MIAS. La figure fig.3.3 est un exemple d'extraction de région ROI de l'image MIAS mdb004.

Les tests ont montré que la procédure d'extraction des ROIs après avoir appliqué une amélioration des images de mammographie, donne des résultats plus efficaces pour l'étape suivante qui est l'étape de la détection et/ou de la segmentation des anomalies dans un système CAD.

3.3.4. Apprentissage du Modèle GMM/ML

L'apprentissage du Modèle GMM/ML est réalisé par l'apprentissage de l'algorithme EM. Des paramètres importants sont nécessaires pour l'obtention de résultats objectifs. Ceux sont les paramètres d'initialisation de l'algorithme EM.

Trois paramètres initialisent l'algorithme EM :

1^{er} paramètre: c 'est le nombre initial M des composantes de densité des gaussiennes. Ce nombre est estimé par l'algorithme K-moyennes. Il représente le nombre de classes de régions homogènes dans la région intérieure de la masse.

2^{ème} paramètre: Le paramètre θ_0 d'initialisation de l'algorithme, défini par les vecteurs moyenne μ_0 , variance σ_0 et le poids du mélange des gaussiennes π_0 . Ce paramètre représente au départ la classe masse.

3^{ème} paramètre: c 'est le paramètre qui représentera la distribution d'une région du tissu entourant la région contenant la masse. Cette information est donnée 'a priori'. C'est l'histogramme des classes de cette région du tissu. Cette dernière est sélectionnée de manière à prendre des pixels pris aléatoirement dans le voisinage de la région contenant l'anomalie.

Le choix de ces paramètres est détaillé dans les paragraphes ci-dessous.

3.3.4.1. Estimation de la densité de la Masse : 1^{er} paramètre

Le modèle GMM peut donner des informations importantes sur la distribution de probabilité des intensités des images issues de mammographie, et particulièrement sur les variations d'intensités dans la région contenant une anomalie de masse.

En effet, en réalité, la région des masses n'est pas une région complètement homogène. En

d'autres termes, elle peut être représentée par plusieurs sous-régions homogènes (sous-classes). Afin d'estimer leur distribution de densités de plusieurs sous-régions suspectes (ROIs), nous devons introduire le nombre de ces classes de ces sous-régions comme information préalable (a priori information).

Ce nombre M cité au dessus, a été choisi après avoir effectué plusieurs tests, avec plusieurs valeurs de M . Selon la séquence d'images étudiées, un nombre initial max de sous-classes a été pris, soit $M=7$ sous-classes. Il ne sera pas fixe car ce nombre initial M représente le nombre des composantes de densité des gaussiennes. Ce nombre est estimé par l'algorithme K-moyennes. En sortie, nous aurons le nombre M des sous-classes de la région masse.

La figure Fig.3.3 est un exemple du schéma d'extraction d'un ROI, avec un zoom sur la région de l'anomalie dans la figure fig.3.3 (d). Cette dernière montre que la région intérieure à l'anomalie n'est pas homogène et présente bien plusieurs classes de sous régions.

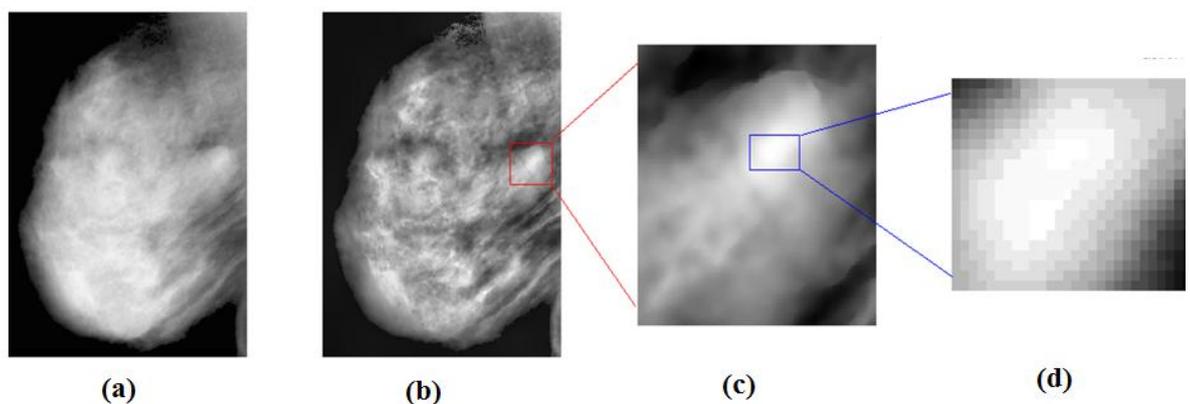


Fig.3.3. Zoom sur la région de l'anomalie:(a) image MIAS mdb004,(b) mdbb04 améliorée,(c) zoom sur la région ROI,(d) Zoom sur l'anomalie.

3.3.4.2. Initialisation du paramètre θ_0 : 2ème Paramètre

Ce paramètre représente au départ la classe masse.

Le paramètre θ_0 d'initialisation de l'algorithme est défini par :

$$\theta_0 = (\pi_0, \mu_0, \sigma_0)$$

Où :

σ_0 : vecteur variance ,

μ_0 : vecteur moyenne ,

π_0 : Le poids de mélange qui satisfait la contrainte suivante:

$$\sum_{k=1}^M \pi_k = 1.$$

3.3.4.3. Paramètre de la densité du tissu : 3^{ème} paramètre

Ce paramètre est donné en choisissant des régions quelconques mais entourant la masse. Ces régions sont considérées comme représentant un tissu normal. Les statistiques de ces dernières sont représentées par leur histogramme.

3.3.5. GMM/ML: Expectation Maximisation (EM)

L'algorithme d'Espérance-Maximisation ou Expectation-Maximisation est l'une des techniques les plus populaires pour extraire les paramètres déterministes d'un mélange avec un nombre donné de composants fournis. Ce nombre est considéré comme une information à priori, pour ce type d'algorithme. Nous avons utilisé un algorithme EM itératif, pour estimer le meilleur ajustement des paramètres du modèle GMM.

D'autre part, comme mentionné précédemment, une des informations principales qui doit être prise en compte est la distribution des intensités des régions de tissu. Ces distributions sont introduites comme données complémentaires à l'algorithme EM.

A ce stade, nous pouvons donner les paramètres du Modèle GMM/M proposé, en partant de la définition du modèle GMM de base par l'équation (3.24) :

$$g(x, \theta) = \sum_{k=1}^M \pi_k f(x, \theta_k) \quad (3.24)$$

Où x est le vecteur des données (valeurs des pixels), $\pi_k, k=1, \dots, M$, sont les poids des mélanges, et $f(x, \theta_k)$, sont les densités des composante Gaussiennes.

Où

$$\theta_k = (\pi_k, \mu_k, \sigma_k), k = 1, \dots, M.$$

Et M représente le nombre de classes dans la région masse, estimé par l'Algorithme du K-means, avec :

μ_k : le vecteur moyennes

σ_k : le vecteur variance

Le poids du mélange satisfaisant la contrainte $\sum_{k=1}^M \pi_k = 1$

Chaque composante de densité est une fonction gaussienne de la forme donnée par l'équation (3.25) :

$$f(x, \theta_k) = f(x, \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k}} \exp\left(-\frac{(x - \mu_k)^2}{2\sigma_k}\right) \quad (3.25)$$

3.3.5.1. L'algorithme EM:

L'algorithme EM a été utilisé avec la fonction du Maximum de Vraisemblance (Maximum Likelihood, ML) donnée par $L(X; \theta)$, les données observées $X = x_1, \dots, x_n$, où n est le nombre de pixels de la région suspecte ROI, comme les équations (3.26) (3.27):

$$L(X; \theta) = L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n \alpha_i f(x_i; \theta) \quad (3.26)$$

$$f(x_i; \theta_k) = f(x_i; \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k}\right) \quad (3.27)$$

Où $i = 1, \dots, n$ et $k = 1, \dots, M$

3.3.5.2. Initialisation des paramètres de l'algorithme EM

Comme cité auparavant, pour le modèle GMM/ML, l'algorithme EM nécessite l'initialisation de trois paramètres:

En premier, le nombre initial M représentant le nombre des composantes des densités Gaussiennes, K-means est l'algorithme le plus souvent utilisé pour estimer ce nombre.

En second, le paramètre initial $\theta_0 = (\pi_0, \mu_0, \sigma_0)$, qui représente en premier la classe de l'anomalie de masse. Ainsi, il est calculé dans une image sélectionné à l'intérieur de la région suspecte, contenant la masse. L'estimation est alors faite sur toute la région suspecte ROI.

Le troisième paramètre représente la distribution "a priori" de la classe du tissu. Elle est donnée par la sélection d'une région quelconque mais autour de la région contenant l'anomalie de masse.

3.3.5.3. Les étapes de l'EM:

L'étape E-:

Calcule la valeur « expectée » du log de la fonction $L(X; \theta)$ du Maximum Likelihood ML, donnée par l'équation (3.26), tout en respectant la condition de distribution de $f(x)$ selon l'estimation en cours du paramètre $\theta^{(k)}$, respectant l'équation (3.24):

$$\Gamma(\theta | \theta^{(k)}) = \log L(X; \theta) = \sum_{i=1}^n \log(\alpha_i f(x_i; \theta)) \quad (3.28)$$

$$\Gamma(\theta | \theta^{(k)}) = \sum_{i=1}^n \log \left(\alpha_i \sum_{k=1}^M \pi_k \cdot f(x_i, \theta_k) \right) \quad (3.29)$$

L'étape M-:

L'étape de Maximisation consiste alors à maximiser les valeurs basses attachées tout en respectant le paramètre $\theta_k = (\pi_k, \mu_k, \sigma_k)$.

En d'autres termes, il s'agit de trouver le paramètre qui maximisera la quantité suivante :

$$\theta^{(k+1)} = \arg \max_{\theta} \Gamma(\theta | \theta^{(k)}) \quad (3.30)$$

$$\pi_{k+1} = \frac{1}{n} \sum_{i=1}^n f(x_i; \theta_k) \quad (3.31)$$

$$\mu_{k+1} = \frac{1}{n\pi_{k+1}} \sum_{i=1}^n x_i f(x_i; \theta_k) \quad (3.32)$$

$$\sigma_{k+1} = \frac{1}{n\pi_{k+1}} \sum_{i=1}^n (x_i - \mu_{k+1})^2 f(x_i; \theta_k) \quad (3.33)$$

3.4. Résultats de la segmentation des masses par le Modèle GMM/ML:

L'algorithme proposé a été appliqué sur une sélection d'images de mammographie dans la base de données MIAS. Seules les 38 images contenant des anomalies de masses ont été considérées. Le choix des images est basé surtout pour les images où l'anomalie est difficile à détecter. Le cas où l'anomalie est entourée d'un tissu dense, le cas où l'anomalie est entourée d'un tissu gras et celui où l'anomalie est entourée d'un tissu glandulaire.

Ensuite, la segmentation des masses calculée par l'algorithme GMM/ML, est comparée à la segmentation effectuée par l'expert, à l'aide d'une évaluation du coefficient du Dice.

Quelques résultats de la segmentation sont présentés dans les figures Fig.3.4, Fig.3.5, et Fig.3.6.

Des résultats représentatifs de l'évaluation du coefficient du Dice sont présentés dans le tableau Tab.3.1.

3.4.1. Segmentation des masses dans le tissu Dense

La figure Fig.3.4 présente une image où l'anomalie est entourée d'un tissu dense :

La figure Fig.3.4 (a) présente la région ROI suspecte.

La figure Fig.3.4 (b) est la segmentation annotée par l'expert.

Les figures Fig.3.4 (c), Fig.3.4 (d), Fig.3.4 (e) représentent la segmentation effectuée par le modèle GMM/ML. Elles montrent le résultat effectué à l'aide de trois régions quelconques (région1, région 2, région3) représentant le paramètre tissu entourant l'anomalie. A chaque région, une indication sur la valeur du coefficient du Dice est donnée. Elle correspondant à la mesure de la segmentation obtenue à l'aide de cette région.

La figure Fig.3.4 (c) représente la segmentation GMM/ML à l'aide d'une région 1.

La figure Fig.3.4 (d) représente la segmentation GMM/ML à l'aide d'une région 2.

La figure Fig.3.4 (e) représente la segmentation GMM/ML à l'aide d'une région 3.

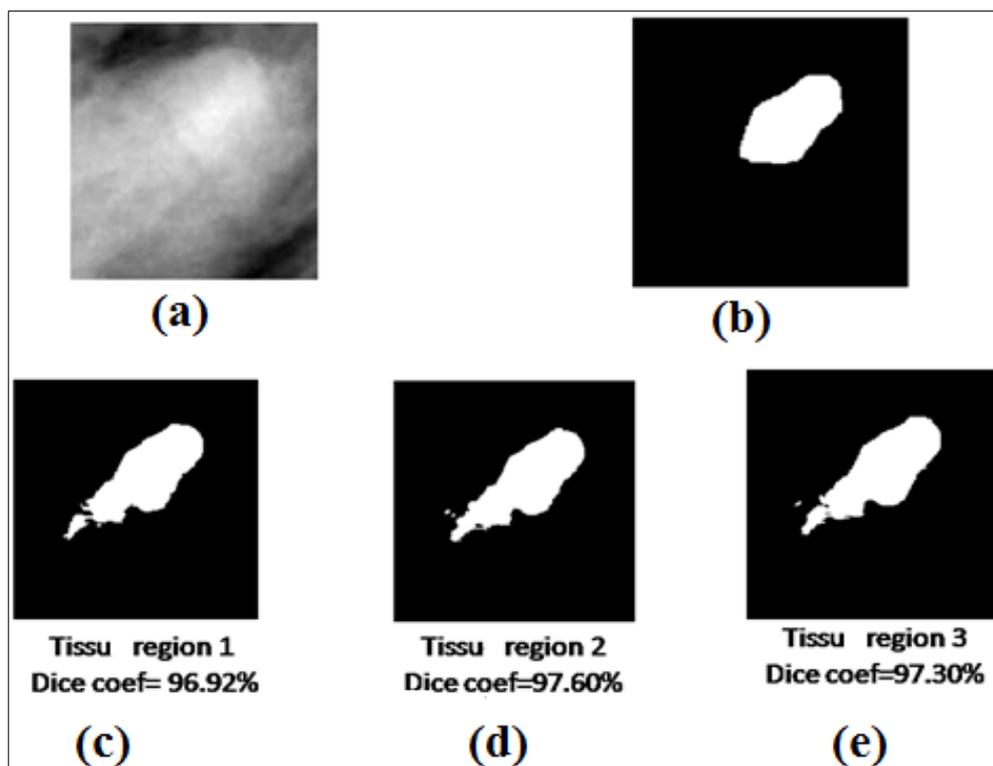


Fig.3.4. Segmentation GMM/ML dans un tissu dense (MIAS mdb004)

3.4.2. Segmentation des masses dans le tissu Glandulaire

La figure Fig.3.5 présente une image où l'anomalie est entourée d'un tissu glandulaire :

La figure Fig.3.5 (a) présente la région ROI suspecte.

La figure Fig.3.5 (b) est la segmentation annotée par l'expert.

Les figures Fig.3.5 (c), Fig.3.5 (d), Fig.3.5 (e) représentent la segmentation effectuée par le modèle GMM/ML. Elles montrent le résultat effectué à l'aide de trois régions quelconques (région1, région 2, région3) représentant le paramètre tissu entourant l'anomalie :

Les figures Fig.3.5 (c), Fig.3.5 (d), Fig.3.5 (e) représentent respectivement la segmentation GMM/ML à l'aide d'une région 1, d'une région 2 et d'une région 3.

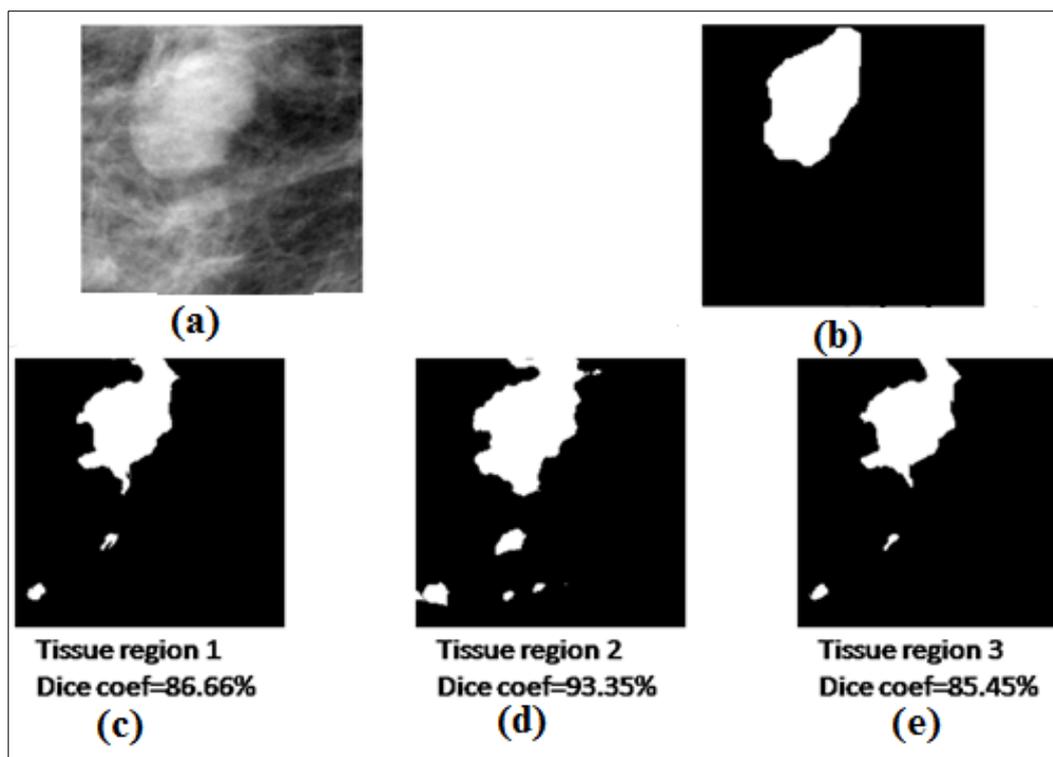


Fig.3.5. Segmentation GMM/ML dans un tissu glandulaire (MIAS mdb019)

3.4.3. Segmentation des masses dans le tissu Gras

La figure Fig.3.6 présente une image où l'anomalie est entourée d'un tissu Gras :

La figure Fig.3.6 (a) présente la région ROI suspecte.

La figure Fig.3.6 (b) est la segmentation annotée par l'expert.

Les figures Fig.3.6 (c), Fig.3.6 (d), Fig.3.6 (e) représentent la segmentation effectuée par le modèle GMM/ML. Elles montrent le résultat effectué à l'aide de trois régions quelconques (région1, région 2, région3) représentant le paramètre tissu entourant l'anomalie :

Les figures Fig.3.6 (c), Fig.3.6 (d), Fig.3.6 (e), représentent respectivement la segmentation GMM/ML à l'aide d'une région 1, d'une région 2 et d'une région 3.

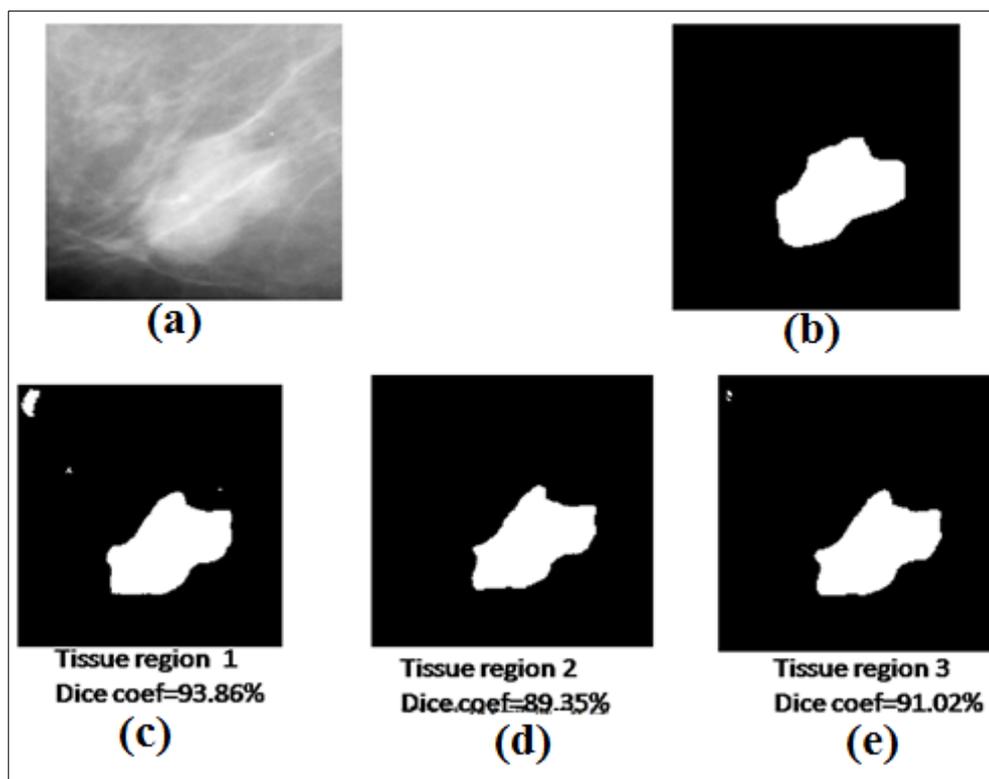


Fig3.6. Segmentation GMM/ML dans un tissu gras (MIAS mdb005)

3.4.4. Evaluation de la segmentation par le coefficient du Dice

Quelques résultats d'évaluation de l'approche de segmentation des masses par la Modèle GMM/ML sont présentés dans un tableau Tab.3.1.

Images	Dice selon une région 1	Dice selon une région 2	Dice selon une région 3
mdb004 anomalie dans un tissu dense	96.92%	97.60%	97.30%
mdb005 anomalie dans un tissu gras	93.86%	89.35%	91.02%,
mdb019 anomalie dans un tissu glandulaire	86.66%	93.35%,	85.45%

Tab.3.1. Evaluation de la segmentation des masses par le Modèle GMM/ML à l'aide du Coefficient du Dice.

3.4.5. Comparaison des résultats GMM/ML :

Dans ce travail, les résultats de la segmentation des anomalies de masses, par le modèle GMM/ML sont comparés avec les résultats de l'approche texturale proposée dans ce travail. Cette dernière qui est une méthode de détection et de segmentation de ces mêmes anomalies, par les matrices de cooccurrence de niveau de gris GLCM. Le paramètre de comparaison est le coefficient du Dice et la table Tab.3.2 récapitule les valeurs obtenues.

Concernant la méthode GMM/ML, les valeurs données représentent la moyenne des valeurs obtenues pour les différentes régions.

Images	Dice pour la méthode GLCM (%)	Dice pour la méthode GMM/ML(%)
Tissu dense (mdb004)	89.39	97.27
Tissu gras (mdb005)	91.74	91.41
Tissu glandulaire (mdb019)	90.67	88.48

Tab.3.2. Tableau de comparaison entre la méthode GMM/ML et la méthode GLCM

Pour les images à tissu dense, la méthode GMM/ML donne de meilleurs résultats

Pour les images à tissu gras et celles à tissu glandulaire la méthode GLCM donne de meilleurs résultats.

3.5. Discussion et Conclusion

L'algorithme du modèle GMM/ML a été appliqué sur différentes images issues de mammographies, incluant différentes densités de tissu. Les résultats de la détection et de la segmentation des anomalies de masses montrent que les statistiques apprises par le modèle induisent des variations du coefficient Dice. Ces variations sont dues essentiellement aux statistiques qui concernent la région du tissu mammaire entourant l'anomalie. Un choix

judicieux de la région qui représente la classe de tissu, fournira une meilleure segmentation de la masse.

Les résultats présentés sont représentatifs des images de la base de données MIAS, celles présentant des densités de tissu différentes. Les valeurs du coefficient du Dice pour un tissu dense, sont généralement meilleures que celles pour un tissu gras ou glandulaire.

Le but initial de la comparaison des valeurs du coefficient du Dice entre la méthode GLCM proposée au chapitre 2, et la méthode GMM/ML présentée dans ce chapitre 3, était de démontrer que l'information donnée par la répartition spatiale des niveaux de gris d'une image, en l'occurrence l'information de texture, et l'information donnée par la valeur de gris des pixels de la même image sont complémentaires. Ainsi, pour obtenir une segmentation judicieuse de l'anomalie de masse, et donc pour différencier le tissu mammaire normal du tissu mammaire suspect, il faudrait combiner entre des méthodes de segmentation qui utilisent le critère « intensité des pixels » des images de mammographie et celles qui utilisent le critère « arrangement spatial des pixels » des images de mammographie.

Chapitre 4 : Intégration des descripteurs de texture extraits des Matrices de Cooccurrence de Niveaux de Gris (GLCM) dans une approche de détection du muscle pectoral et du sein.

Résumé

La segmentation du muscle pectorale et de la région mammaire est un processus nécessaire et une étape importante pour les systèmes d'aide à la détection CADe, ou d'aide au diagnostic CADx. Dans la continuation de ce travail, nous avons voulu tester l'approche GLCM dans des problèmes de détection du muscle pectorale et de la région mammaire. Cette approche n'a pas été utilisée à ces fins. Elle fournit des résultats très intéressants en vue de résoudre certains problèmes pas encore résolus, tels que la présence d'anomalies de masses ou de micro-calcifications dans les limites pectorales. Ces dernières sont habituellement ignorées dans les étapes de détection et/ou de segmentation des régions mammaires suspectes. Ce problème est principalement dû soit à une confusion entre le tissu de la région suspecte et le tissu du muscle pectoral, soit à une confusion entre ligne pectorale et limite pectorale. La technique proposée peut être vue sous deux parties principales. Une méthode de lissage et d'amélioration de contraste est d'abord utilisée en vue d'améliorer l'image de mammographie. Ensuite, des images texturales représentant les descripteurs statistiques extraits des matrices GLCM sont calculées. Elles permettent de détecter des contours du muscle pectoral et de la région mammaire. Et pour affiner les contours détectés, une étape de post-traitements a été développée. Ces techniques ont été appliquées sur des images de la base MIAS, constituée de mammographies avec des vues « MLO » (Médio Latéral Oblique). Certaines images présentent un sein mal orienté, une étape de réorientation du sein est alors proposée. Les résultats sont analysés et comparés avec d'autres travaux actuels et similaires.

4.1. Introduction

Ce chapitre présente d'abord un état de l'art sur les différentes méthodes utilisées pour la détection et/ou la segmentation du muscle pectoral et de la région mammaire qui est le sein. Le but est de faire le point sur les problèmes soulevés par les auteurs quand à la nécessité d'extraire ces deux régions, et les difficultés rencontrées pour l'extraction de la région du

muscle pectoral, ou pour à la délimitation de la région mammaire et du mamelon. Les solutions apportées pour résoudre ces problèmes sont alors résumées. Puis l'approche proposée pour la détection du muscle pectoral et de la région mammaire est présentée et détaillée.

4.2. Etat de l'art

Les systèmes CAD doivent identifier la région mammaire indépendamment du système de numérisation, de l'orientation du sein dans l'image et de la présence de bruit, y compris l'image d'artéfacts. La segmentation du sein consiste à extraire la région mammaire et les frontières pectorales, c'est une étape nécessaire pour les systèmes CAD typiques. Beaucoup d'études se sont concentrées sur ces problèmes. D. Rabat et al. (Raba, D. et al., 2005), ont résumé les travaux sur la segmentation de la région mammaire avec la suppression du muscle pectorale. Plus récemment, une revue des méthodes de segmentation pectorale a été récapitulée par Karthikeyan Ganesana et al. (Karthikeyan, Rajendra Acharyaa, Kuang, Lim, & Thomas abraham, 2013). Des auteurs ont montré que le tissu de la région mammaire et la région du muscle pectoral peuvent avoir les mêmes intensités et une texture semblable dans la mammographie (Peter & Sue, 1992), (Ferrari, F. Frère, Rangayyan, & al., 2004), (Domingues, Cardoso, Amaral, & al., 2010), (Grim, Somol, Haindl, & al., 2009), (Qi, Jiaqing, & Virginie, 2008), (Daniel, Mikael, & Rachid, 1987). Donc cela encombre et désoriente l'étape de détection des anomalies malignes. Peter et Sue (Peter & Sue, 1992) ont utilisé une discrimination à base de texture, entre des types de seins gras et de seins denses. Ils proposent des techniques granulométriques et des masques de texture.

Le problème principal des approches proposées par Ferrari et al. (Ferrari, F. Frère, Rangayyan, & al., 2004), est que le muscle pectoral est approximé par une ligne. Ces méthodes donnent de résultats médiocres quand le contour du muscle pectoral est une courbe. Alors les auteurs se sont focalisés sur la segmentation du muscle pectoral en considérant ce dernier comme une ligne ou comme une courbe.

4.3. Conclusions sur l'état de l'art

Après cet état de l'art sur les travaux relatifs à la segmentation du muscle pectoral, nous pouvons résumer les problèmes soulevés par les auteurs en trois principaux problèmes :

- 1- En premier lieu, la région du muscle pectoral peut avoir la même texture que la région du tissu mammaire.
- 2- En second lieu, la bordure du muscle pectoral doit être considérée comme une courbe et non comme une ligne.
- 3- En troisième lieu, les auteurs utilisent des méthodes pour la détection et/ou la segmentation du muscle pectoral et d'autres méthodes pour la détection et/ou la segmentation du sein.

Dans ce travail, une contribution à résoudre ces problèmes est apportée en testant les paramètres de texture extraits des matrices GLCM, pour la détection des contours du muscle pectoral et du sein. Cette approche n'a pas encore été utilisée à ces fins. Les paramètres GLCM sont généralement utilisés comme des caractéristiques de texture dans des approches de segmentation du muscle pectoral et du sein, basées région. L'approche proposée permet de détecter à la fois les contours du muscle pectoral et du sein.

Quatre étapes décrivent la méthodologie présentée dans ce travail. L'étape de réorientation des mammographies, l'étape d'amélioration des images, l'étape de calcul des images texturales et l'étape des post-traitements. Elles sont détaillées dans les paragraphes qui suivent. Une évaluation des résultats obtenus et leur comparaison avec des travaux similaires sont aussi réalisées.

4.4. Méthodologie: Détection et segmentation de muscle pectoral et de la région mammaire par les GLCM

4.4.1. Introduction

Dans le chapitre 2, une étude a été effectuée concernant l'utilisation des matrices de cooccurrence GLCM dans les méthodes de détection et/ou segmentation des anomalies de masses. Dans la continuation de ce travail, nous avons voulu tester cette approche dans des problèmes de détection et/ou de segmentation du muscle pectoral et de la région mammaire.

Le schéma de la méthodologie utilisée à cet effet ainsi que les détails de ses étapes, sont développés dans les paragraphes qui suivent.

4.4.2. Schéma de l'approche proposée

Les principales étapes de la méthodologie proposée sont récapitulées dans la figure (Fig.4.1) :

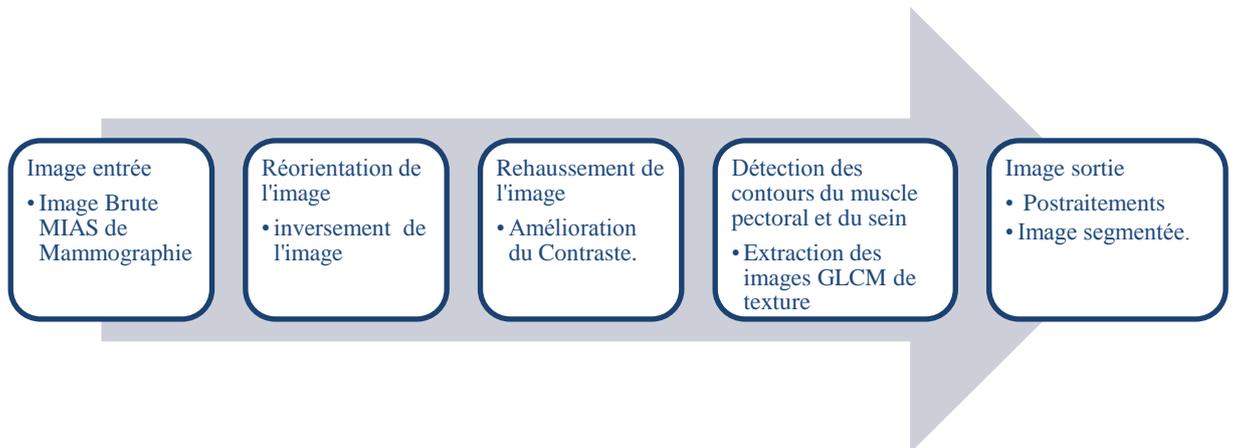


Fig.4.1. Principales étapes de la méthodologie proposée

4.4.3. Les différentes étapes de l'approche texturale GLCM

Quatre étapes principales définissent l'approche:

4.4.3.1. Etape 1: Orientation du sein

Dans la base de données MIAS, certaines images de mammographies présentent un sein 'inversé', on dit que ce dernier est mal orienté, la figure Fig 4.2 en est un exemple. On distingue clairement que dans l'artéfact, le texte est inversé. Ce texte informe sur le sein pris en considération, pour le sein gauche (L: Left), et pour le sein droit (R: Right). Ainsi que la vue de l'image, pour une vue latérale (MLO: La Médio Latérale Oblique) et pour une vue cranio-caudale (CC: Cranio-Caudale). Dans la figure Fig.4.2, l'artéfact montre 'R' et 'ML' mais inversés. pour dire que c'est le sein droit et la vue du sein est Médio Latérale.



Fig.4.2. Exemple d'une image de mammographie MIAS avec sein mal orienté

L'image artéfact est donc le paramètre qui indiquera que dans l'image de mammographie, le sein est bien orienté ou pas. Dans le cas où le texte dans l'artéfact est inversé, l'image de mammographie doit être inversée pour avoir une image de l'orientation réelle du sein.

Avant de procéder à l'application de la méthode proposée, les 322 images de la base de données MIAS ont été lues et réorientées si nécessaire. La figure Fig.4.3 est un exemple d'une image mal orientée avec le résultat de sa réorientation.

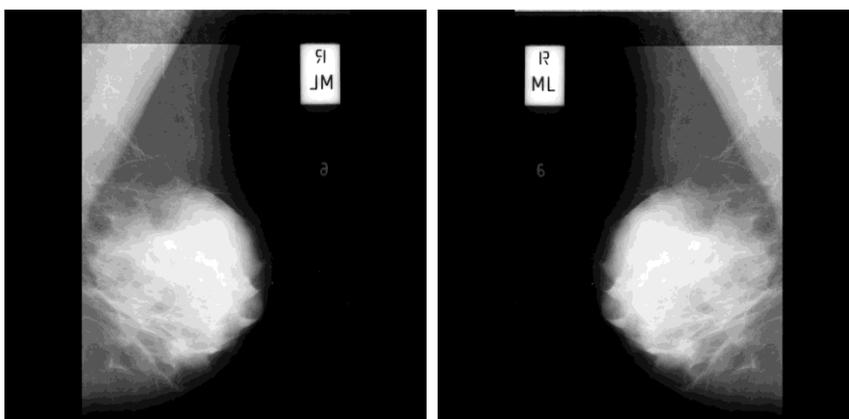


Fig.4.3. A gauche, image avec un sein mal orienté. A droite, image avec un sein bien orienté.

4.4.3.2. Etape 2: Amélioration du contraste des images:

Rappelons que la performance des méthodes se basant sur l'information de texture dépend fortement du prétraitement de l'image prise en entrée (Pfisterer & Aghdasi, 1999). Pour cette approche, la même procédure, présentée au chapitre 2, a été utilisée pour

l'amélioration des images de mammographie. Le but est de détecter des contours précis de la région qui sépare le muscle pectoral du tissu mammaire, et ceux de la région mammaire.

Les images MIAS de mammographie présentent un faible contraste des objets. Pour extraire plus d'information de ces dernières, nous les avons améliorées en appliquant les mêmes techniques et les mêmes méthodes utilisées dans le chapitre 2 au paragraphe §2.5.4.1. Le choix de cette approche d'amélioration des images MIAS de mammographie est motivé par les mêmes raisons citées aussi dans ce paragraphe.

Ainsi, les images ont été améliorées en appliquant tout d'abord un filtre de lissage de l'image, puis une augmentation de son contraste. Le filtre est utilisé pour éliminer d'éventuels pixels de bruit sur l'image, le filtre de diffusion anisotrope SRAD (Speckle Reducing Anisotropic Diffusion)(Yu & Acton, 2002). Et pour augmenter le contraste de l'image ainsi lissée, la méthode d'égalisation d'histogrammes CLAHE (Contrast-limited Adaptive Histogram Equalization).

Des exemples de quelques images améliorées est repris dans les figures Fig.4.4, Fig.4.5. , Fig.4.6. Dans ces figures, les images améliorées présentent plus de contraste et plus de régions. Les régions claires sont encore plus claires et les régions sombres sont encore plus sombres.

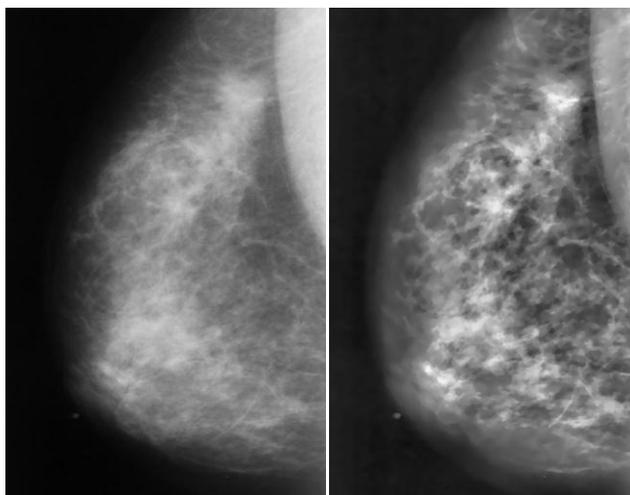


Fig.4.4. A Gauche, image MIAS mdb001. A droite, mdb001 améliorée.

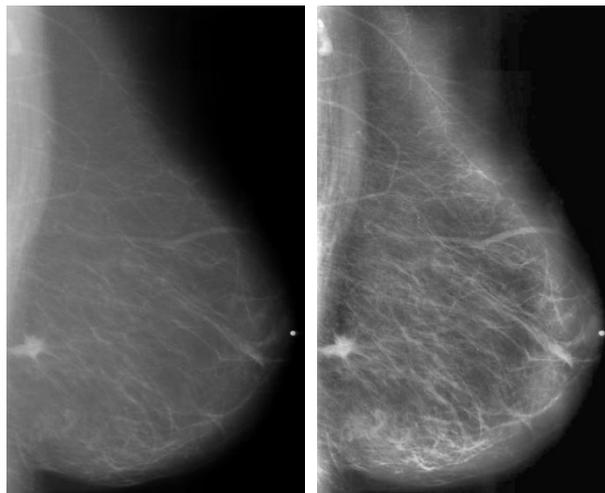


Fig.4.5. A Gauche, image MIAS mdb003. A droite, mdb003 améliorée.

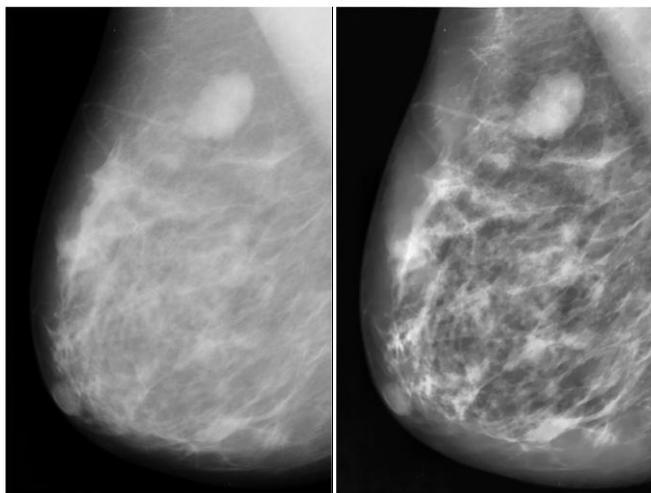


Fig.4.6. A Gauche, image MIAS mdb015. A droite, mdb015 améliorée.

4.4.3.3. Etape 3: Détection des contours du muscle pectoral et du sein par les GLCM

Pour ce faire, et comme pour le chapitre 2, des images de texture représentant le descripteur de contraste ont été obtenues.

Le descripteur de contraste est extrait à partir des matrices de cooccurrence calculées suivant trois importants paramètres: la taille du masque, la direction et la distance. Le choix de ces paramètres a été fait selon les mêmes critères cités aussi dans le chapitre 2 au paragraphe § 2.5.4.2 :

4.4.3.3.1. La taille du masque:

Plus la taille du masque est grande, plus les contours détectés seront flous. Ainsi, la taille minimum 3x3 a été prise pour que tous les détails de l'image soient pris en considération. Des exemples ont été donnés dans chapitre 2 au paragraphe §2.5.4.2.3.

4.4.3.3.2. La direction:

Le calcul des matrices de cooccurrence selon une direction unique $\theta=0^\circ$, ou $\theta=45^\circ$, ou $\theta=90^\circ$, ou $\theta=135^\circ$, donne des contours discontinus. Pour cela, le calcul des matrices a été d'abord effectué pour chacune de des distances, puis la somme des quatre images est calculée. De cette manière, on obtient des contours plus fermés. Des exemples ont été donnés dans le chapitre 2 au paragraphe § 2.5.4.2.2.

4.4.3.3.3. La distance:

Le choix de distance $d=1$ correspondra à prendre le premier pixel voisin. Une distance $d=2$ correspondra au deuxième pixel voisin et donc le premier sera omis, ainsi de suite. Pour les mêmes raisons citées dans le chapitre 2 au § 2.5.4.2.1, le calcul des matrices de cooccurrence a été effectué selon une distance $d=1$.

4.4.3.3.4. Le Descripteur de contraste

Les matrices de cooccurrence GLCM permettent d'extraire descripteurs de texture (Haralick, 1979), les plus connus étant 'le contraste', 'l'entropie', 'l'homogénéité' ou encore 'la corrélation'. Seulement pour cette approche le descripteur de contraste est suffisant pour l'objectif visé, celui de mesurer l'hétérogénéité dans l'image pour détecter les variations spatiales des niveaux de gris de cette image.

Ainsi, l'image est balayée et à partir des matrices de cooccurrence calculées selon les trois paramètres cités au dessus, et pour chaque point de l'image, le descripteur de contraste est extrait de la matrice correspondante. Le point correspondant est remplacé par ce descripteur.

Dans ce travail, l'équation du descripteur de contraste est donnée par la formule (4.1) :

$$contrast = \sum_{i,j} |i - j|^2 p(i, j) \quad (4.1)$$

Et le résultat de ces calculs est une image avec des contours, les figures données au chapitre 2 sur des images ‘Brodatz’, en sont des exemples (Fig.2.11 ou Fig.2.13). Quelques résultats de l’application de cet algorithme sur des images MIAS de mammographie ont été donnés dans les figures Fig. 2.4, 2.16, 2.17 ou 2.18. Elles correspondent à des images ROI. Les résultats pour images complètes de mammographie sont donnés dans les paragraphes qui suivent.

4.4.3.4. Etape 4: Post-traitements

L’étape de détection des frontières du muscle pectoral donne en général des contours parfois discontinus et mal finis. Ce qui est le cas pour certaines images texturales obtenues par le descripteur GLCM. Plusieurs méthodes de post-traitement ont été utilisées par les auteurs pour affiner les frontières du muscle pectoral (Kwok, Chandrasekhar, Attikiouzel, & al, 2004 (a)), (Molinara, Marrocco, & Tortorella, 2013), (Mustra, Bozek, & Grgic, 2009).

Pour notre approche, le souci est de relier les extrémités des contours ainsi détectés par le descripteur textural de contraste, et de les affiner. La méthode itérative “cliff detection” a été utilisée (Kwok, Chandrasekhar, Attikiouzel, & al, 2004 (a)). Elle permet aussi de délimiter plus exactement le contour. Finalement, une région fermée, représentant le muscle pectoral, est produite comme un masque de segmentation (Kwok, Chandrasekhar, Attikiouzel, & al, 2004 (a)).

4.5. Résultats de l’approche de détection des contours du muscle pectoral et de la région mammaire par le Descripteur GLCM de Contraste

4.5.1. Base de données:

Rappelons que l’approche a été testée sur des images de la base de données MIAS. L’algorithme a été appliqué sur les 322 images de cette dernière. Les images sont de type MLO (incidence latérale oblique) et sont numérisées avec une résolution de 50 μm par pixel sur 8 bits. La résolution spatiale de chaque image est de 1024 x 1024 pixels.

Dans les 322 mammographies, 110 images ne contiennent pas les artéfacts radio-opaques, mais souffrent d’un fond bruité (cliché), et 212 images contiennent les artéfacts radio-opaques.

4.5.2. Résultats de la détection du muscle pectoral et du sein:

Les figures Fig.4.7(a) et Fig.4.9(a) représentent deux vues différentes du sein, avec l'image input et l'image améliorée.

Les figures Fig.4.7(b) et Fig.4.9(b) présentent les résultats de la détection des contours, c'est le descripteur de contraste. Et les images avec extraction du muscle pectoral.

Les figures Fig.4.8 et Fig.4.10 présentent respectivement un zoom des figures Fig.4.7(b) et Fig.4.9(b). Le zoom montre les détails de la région pectoral avec la 'ligne' du pectoral et la 'bordure' du pectoral.

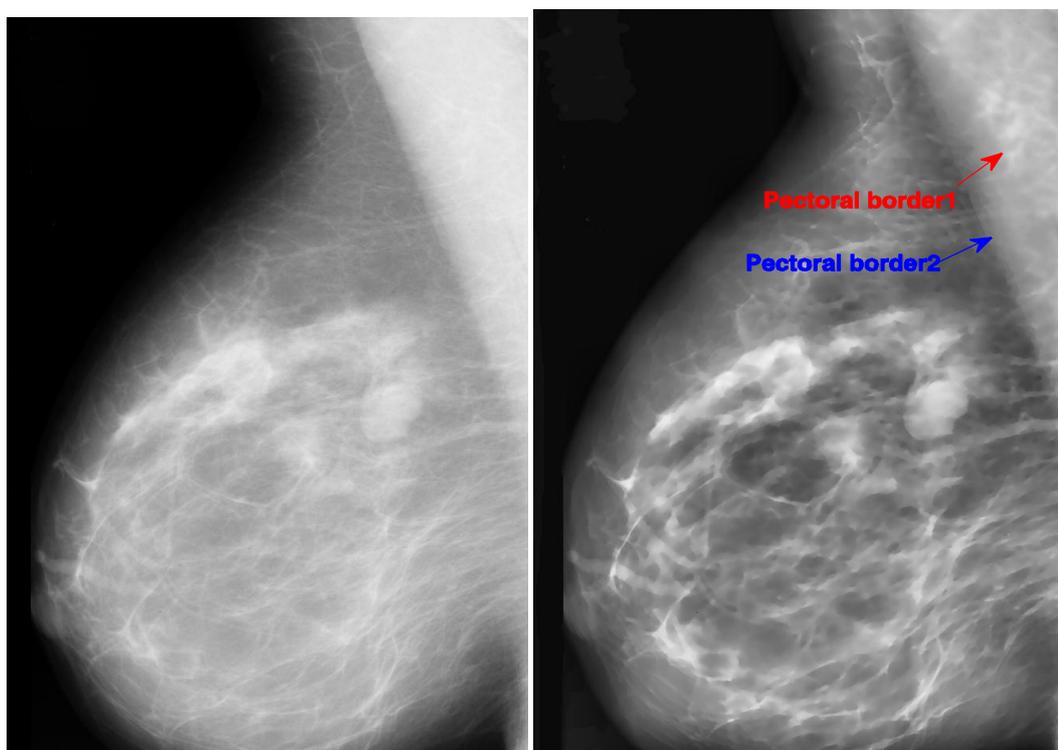


Fig.4.7(a). A gauche, image input mdb019. A droite, mdb019 améliorée.

On remarque déjà que dans l'image améliorée, les frontières du muscle pectoral sont plus visibles. Une frontière qui correspond à la ligne du pectoral (indiquée par une flèche bleue). Et une frontière qui correspond à la bordure du pectoral (indiquée par une flèche rouge).

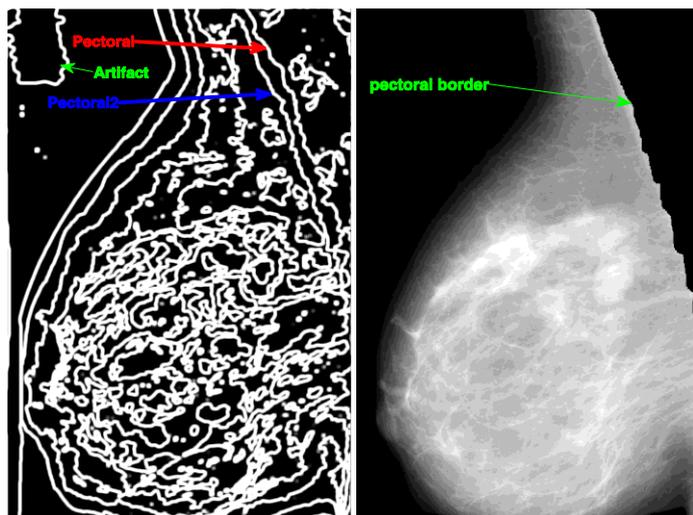


Fig.4.7(b). A gauche, Descripteur de contraste de mdb019. A droite, mdb019 avec extraction du muscle pectoral

Dans la figure Fig.4.7(b), l'image texturale représentée par le descripteur de contraste montre que l'algorithme a détecté les frontières du muscle pectoral (indiquées par des flèches bleues et rouges).

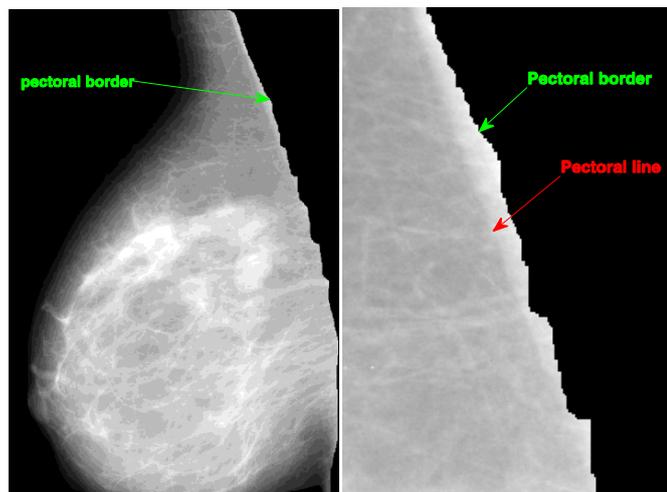


Fig.4.8. A gauche, mdb019 segmentée. A droite, zoom sur les frontières du muscle pectoral.

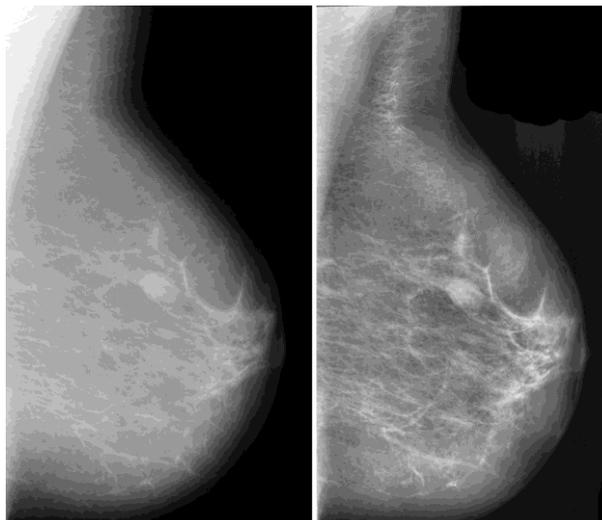


Fig.4.9(a). A gauche, image input mdb012. A droite, mdb012 améliorée.

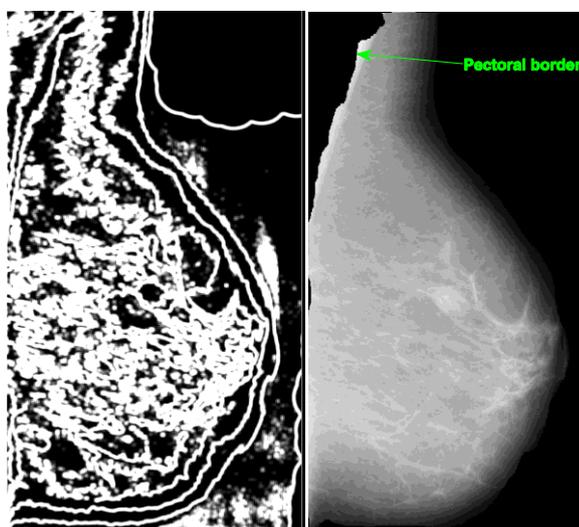


Fig.4.9(b). A gauche, descripteur de contraste de mdb012. A droite, mdb012 avec extraction du muscle pectoral.

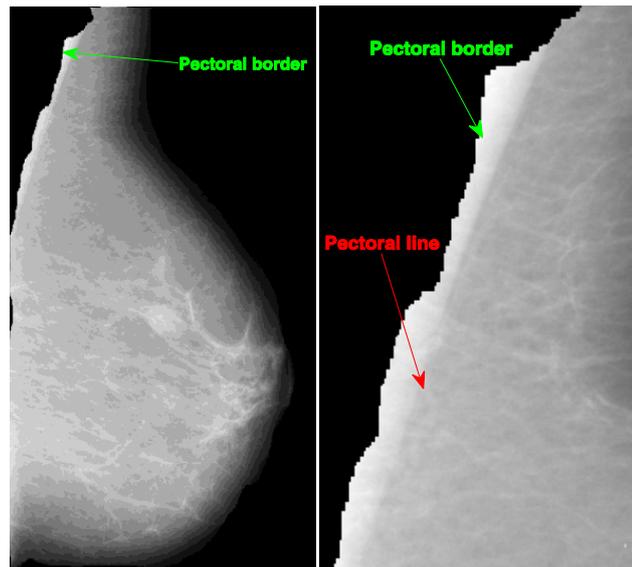


Fig.4.10. A gauche, mdb012 segmentée. A droite, zoom sur les frontières du muscle pectoral.

Dans la figure Fig.4.9(b), l'image texturale montre aussi que l'algorithme a détecté les frontières du muscle pectoral.

Remarque:

La figure Fig4.7(b) montre que l'algorithme a détecté des contours correspondants à l'artéfact. Ce dernier n'est pas visible dans l'image MIAS de mammographie mdb019. Cette dernière fait partie des 110 images qui ne contiennent pas les artéfacts radio-opaques, mais qui souffrent d'un fond bruité (cliché).

4.5.3. Evaluation quantitative des résultats de la détection du muscle pectoral

Finalement seules les 110 images qui ne contiennent pas d'artéfacts ont été testées. Et comme plusieurs d'entre elles souffrent de bruit (telle que l'image mdb019 citée ci-dessus), seules 55 images ont été choisies.

Pour évaluer les résultats de l'approche proposée de détection du muscle pectoral, deux catégories ont été utilisées, Acceptable et Inacceptable. 'Acceptable', représente le pourcentage des résultats où la détection des frontières du muscle pectoral est bonne ou acceptable. 'Inacceptable', représente le pourcentage des résultats où les frontières du muscle pectoral n'ont pas été détectées ou ont été mal détectées. Dans la table 4.1, ces catégories sont représentées par % acc. , et % Unacc.

Ensuite, les valeurs obtenues sont comparées avec les résultats de quelques travaux similaires à nos objectifs. Molinara et al. (Molinara, Marrocco, & Tortorella, 2013), ont résumé récemment ces travaux. Tels que les méthodes utilisées par Kwok et al. (Kwok, Chandrasekhar, Attikiouzel, & al, 2004 (a)), (Kwok, Chandrasekhar, & Attikiouzel, 2001(b)), qui proposent un seuillage itératif combiné avec une recherche à base de gradient. Une approche statistique basée sur le concept 'goodness of fit' est utilisée par Liu et al. (Liu, Wang, & Wang, 2011), afin de développer une mesure de la distribution spatiale locale dans l'image de mammographie. Le travail de Mustra et al. (Mustra, Bozek, & Grgic, 2009), qui proposent un hybride avec une réduction de la profondeur de bits et une décomposition d'ondelettes. Une approche basée sur les histogrammes adaptatifs proposée par Raba et al. (Raba, D. et al., 2005). Et enfin la méthode de Molinara et al. (Molinara, Marrocco, & Tortorella, 2013), qui présente une étape de détection des bords du muscle pectoral et une régression par 'RANSAC' qui fournit le contour final de la zone de muscle.

Les résultats de ces travaux cités, ainsi que ceux de l'approche texturale GLCM proposée dans ce travail, sont résumés dans la Table.4.1. Ils montrent une bonne performance de l'algorithme proposé, en considérant les 55 images test. Les valeurs étant exprimées en termes de détection acceptable et inacceptable du muscle pectoral.

Méthodes	Images	% acc.	% Unacc.
Kwok et al. (2004)	322	83.6	16.4
Liu et al. (2011)	100	81.0	19.0
Mustra et al. (2009)	40	85.0	15.0
Raba et al. (2005)	322	86.0	14.0
Molinara et al. (2013)	55	89.1	10.0
GLCM	55	89.5	10.5

Table.4.1. Evaluation et Comparaison avec d'autres travaux.

4.6. Discussion et Conclusion

Dans ce chapitre, une contribution a été apportée dans les problèmes de détection et/ou de segmentation du muscle pectoral et de la région mammaire. L'approche utilisant les matrices de cooccurrence GLCM, n'a pas été utilisée pour résoudre ces problèmes. Le descripteur de contraste extrait de ces matrices est généralement utilisé comme information supplémentaire dans des approches orientées région. L'approche de détection des contours du muscle pectoral et de la région mammaire par ce descripteur a donné des résultats intéressants.

Concernant la détection du muscle pectoral, et comparativement à d'autres travaux similaires, l'algorithme a détecté des frontières plus précises entre le muscle pectoral et la région mammaire, pour certaines images de la base MIAS. Rappelons que l'objectif de la détection des contours du muscle pectoral, est l'extraction de cette région pectorale car elle peut présenter le même tissu que les anomalies de masse, et de ce fait elle peut être confondue avec cette dernière. D'un autre côté, il est important de ne pas extraire une région qui pourrait justement être suspecte.

Concernant la détection de la région mammaire, l'algorithme a détecté des frontières intéressantes, notamment au niveau de la glande du mamelon. L'évaluation des résultats et la comparaison avec d'autres travaux similaires n'ont pas été abordées dans ce travail. Il serait cependant intéressant de le faire dans des travaux futurs.

Il est important de noter que dans la littérature, les problèmes de détection et/ou de segmentation du muscle pectorale sont abordés indépendamment de ceux de la détection et/ou la segmentation de la région mammaire. L'approche proposée est une méthode de détection du muscle pectoral et de la région mammaire, ce qui peut être considéré comme un gain dans le temps de réponse d'un système d'aide à la détection CADe et ou d'aide au diagnostic CADx.

Conclusion Générale et Perspectives

Dans ce travail de recherche, nous avons contribué au développement de méthodes de détection et de segmentation des anomalies de masses sur des images de mammographie.

L'étape de détection et/ou de segmentation des anomalies de cancer dans des images issues de mammographie, est une étape clé dans tout le processus des systèmes automatiques d'aide à la détection (CADE) ou d'aide au diagnostic (CADx). Elle détermine la performance et le degré de fiabilité de ces systèmes. Seulement, les images de mammographie sont de faible contraste, les types du tissu mammaire sont complexes aussi bien visuellement que quantitativement, et ils peuvent couvrir complètement l'anomalie. Tous ces problèmes font que jusqu'à présent, la tâche la plus difficile pour les chercheurs dans ce domaine est réellement la discrimination entre le tissu mammaire normal et le tissu contenant une anomalie cancéreuse. Plus le tissu est dense, plus la confusion augmente.

Dans cette étude, deux approches ont été proposées, développées, testées sur des images de mammographies et enfin évaluées par rapport à des travaux similaires :

La première est une méthode statistique, elle se base sur le critère « arrangement spatial des pixels » des images, c'est la texture. La robustesse des matrices de cooccurrence de niveaux de gris (Grey Level Cooccurrence Matrix, GLCM) a été utilisée pour développer une approche de segmentation des anomalies de masses par détection de leurs contours. Ces matrices étant plutôt utilisées dans des approches régions, juste pour extraire des caractéristiques de texture. Ces dernières sont introduites comme données, soit pour l'étape de segmentation ou pour une étape postérieure de classification. Pour l'approche proposée, le descripteur de contraste extrait de ces matrices a permis d'obtenir des images texturales où les contours des anomalies ont été détectés.

La seconde est une méthode probabiliste, qui se base sur le critère « intensité des pixels ». Nous avons utilisé la force des Modèles de Mixture des Gaussiennes (Gaussian Mixture Model, GMM), avec l'algorithme d'Espérance-Maximisation (EM) et l'algorithme

du Maximum de Vraisemblance (ML), pour développer une méthode de segmentation des masses par régions. L'initialisation des paramètres à l'algorithme itératif EM par des statistiques sur la région cancéreuse et des statistiques sur la région à tissu normal, a permis d'ajuster au mieux les paramètres GMM. Ces paramètres « appris » pour le modèle GMM/ML font que cette approche a donné plus de performance au modèle GMM de base.

Dans la continuité de ce travail, l'approche texturale GLCM a été testée pour résoudre des problèmes liés à l'amélioration des images de mammographie. Il s'agit de préparer l'image avant les étapes de détection et/ou de segmentation des anomalies. Le principe consiste à extraire la région mammaire en coupant la région du muscle pectoral et le fond bruité des images. Le bruit contenant des artefacts obtenus à la saisie des images. L'approche GLCM a permis de détecter les frontières du muscle et celles du sein. Le descripteur de contraste n'a jamais été utilisé auparavant pour ce type de problèmes. La technique proposée s'est avérée plus précise pour la détection des formes exactes des frontières recherchées. Quelques exemples ont été donnés, ainsi qu'une première évaluation quantitative et comparative avec des travaux similaires.

Le développement des approches proposées repose sur la phase d'amélioration du contraste de l'image de mammographie et la technique proposée à cet effet est considérée comme une étape « clé » pour l'obtention de bons résultats. Il a été montré que l'utilisation de la méthode "SRAD" et de la méthode "CLAHE" de manières différentes, peut donner des résultats moins performants pour les approches proposées. Aussi, d'autres méthodes d'amélioration de contraste de l'image ne donneraient pas les résultats espérés.

Finalement, vu l'importance des types du tissu mammaire, fréquemment citée par la majorité des auteurs, et dans le but d'éclairer la complexité de ces derniers, les approches ont été testées sur des images à différentes densités de tissu. Les cas où la masse est entourée d'un tissu dense, ceux où elle est entourée d'un tissu gras et ceux où elle est entourée d'un tissu glandulaire. Ceux sont les trois types de tissu mammaire les plus présents dans les images numériques de mammographie.

Les résultats obtenus pour les approches proposées ont été quantifiés par le coefficient du Dice, les coefficients de F-Measure, de Precision, de Recall et de l'aire sous la courbe (Az). Ils sont prometteurs pour des travaux futurs. Il a été montré que l'évaluation et la

comparaison de ces derniers avec des travaux similaires n'est pas une tâche évidente, vu que pour chaque méthode les auteurs ont utilisé des critères différents. Arnaud Oliver (2010), a essayé de donner un résumé de quelques méthodes, en tenant compte de quelques critères que nous avons cités, mais la prise en considération de tous les critères fait que la comparaison reste une tâche difficile et fastidieuse pour les auteurs. Ainsi, ce domaine de recherche reste ouvert à toute contribution.

Les approches proposées peuvent être améliorées avec ces travaux futurs :

L'approche de détection des contours des masses par le descripteur de contraste des matrices GLCM peut être améliorée en temps de calcul, car les matrices GLCM sont certainement très robustes pour donner une information complète sur la texture des images numériques, mais elles sont aussi réputées pour leur temps de calcul. Des travaux s'effectuent dans ce sens par beaucoup d'auteurs. Une autre amélioration peut être effectuée pour cette approche, c'est au niveau des post-traitements. Il serait intéressant de développer des méthodes plus précises pour extraire automatiquement la masse telle que identifiée par ses contours.

Concernant l'amélioration des images, seules l'extraction du muscle pectorale a été évaluée. Il serait intéressant de comparer la détection des frontières du sein et donc son extraction du fond de l'image, avec des travaux similaires.

Pour le modèle GMM/ML, la méthode est semi-automatique. Il a été montré que l'initialisation du modèle par les statistiques relatives à la région mammaire entourant l'anomalie, induit des variations du coefficient Dice. Et qu'un choix judicieux de cette région fournira une meilleure segmentation de la masse. Mais comme les valeurs obtenues pour ce coefficient sont prometteuses, il serait intéressant d'automatiser les statistiques apprises par le modèle en faisant un apprentissage automatique. Ce dernier se basera sur les modèles de l'intelligence artificielle.

Une autre amélioration du modèle GMM/ML, consiste à introduire la notion de distance pour différencier au mieux les gaussiennes. Des travaux sont déjà en cours et les premiers résultats sont prometteurs aussi.

D'autre part, le tableau comparatif de la méthode GLCM et de la méthode GMM/ML montre une fois encore que les critères « arrangement spatial » et « intensité des pixels » sont complémentaires. Mais que l'information de texture reste l'information cruciale et nécessaire pour enlever toute confusion entre le tissu normal et le tissu cancéreux. Seulement, jusqu'à présent la notion de texture est difficilement exprimable, il n'existe pas de formalisme mathématique pour définir exactement une texture. Cet axe de recherche reste ouvert à toute contribution scientifique pouvant aider à donner ce formalisme.

Bibliographie

- Afsaneh, J., Syamsiah B.T., M., Hajjah Rozi, M., Abdul Rahman, B. R., & Babak, K. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Elsevier Clinical Imaging* 37, 420–426.
- American Cancer Society. (2002). *Cancer Facts and Figures*. Atlanta, GA.
- Arnau, O., Freixenet, J., Marti, J., Perez, E., Pont, J., R.E. Denton, E., & Zwiggelaar, R. (2010). A review of automatic mass detection and segmentation in mammographic images. *J.Medical Image Analysis*, 14, 87–110.
- Aswini kumar mohanty et al. (2011). Image Mining for Mammogram Classification by Association Rule Using Statistical and GLCM features. *IJCSI International Journal of Computer Science Issues*, 8(3).
- Bandyopadhyay, S. (2010). Survey on Segmentation Methods for Locating Masses in a Mammogram Image. *International Journal of Computer Applications*, 9(11).
- Bellotti, R., Carlo, F. D., Tangaro, S., Gargano, G., Maggipinto, G., Castellano, M., . . . Nunzio, G. D. (2006.). A completely automated CAD system for mass detection in a large mammographic database. *Med. Phys*, 33(8), 3066–3075.
- Cabarrot.E. (2000). Histoire naturelle des cancers du sein. Dans *Encyclopédie Médico-Chirurgicale Gynécologie* (pp. 865-A-10). Paris: Editions scientifiques et médicales Elsevier SAS.
- Campanini, R., Dongiovanni, D., Iampieri, E., Lanconelli, N., Masotti, M., Palermo, G., . . . Roffilli, M. (2004). A novel featureless approach to mass detection in digital mammograms based on support vector machines. *Phys. Med. Biol.*, 49(6), 961–975.
- Caroline, B., & Vaijyanthi N. (2012). Computer aided detection of masses in digital breast tomosynthesis: A review. *IEEE Conference Publications*, 186-191.
- Chan, H. P., D.Wei, Helvie, M. A., Sahiner, B., Adler, D. D., Goodsitt, M., & Petrick, N. (1995). Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space. *Phys. Med. Biol*, 40(5), 857–876.
- Cheng, H., Shi, X., Min, R., Hu, L., Cai, X., & Du, H. (2006). Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition* 39, 646–668.
- Classification_TNM. (2015). https://fr.wikipedia.org/wiki/Classification_TNM (consulté en 2015).

-
- Claude, I., Pouletaut, P., Huault, S., & Boulanger, J. C. (2001). Integrated color and texture tools for colposcopic image segmentation. *Image Processing, 2001. Proceedings. 2001 International Conference on*, 2, 311-314.
- Daniel, C., Mikael, R., & Rachid, D. (1987). A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *72(2)*, 195–215.
- Djaroudib, K. (1993(a)). Quantitative estimation of woody vegetation in the Sudano-Sahelian region. A textural approach applied in a region of Burkina Faso. *EARSeL. Advances in remote sensing*, 2(3), 56-63.
- Djaroudib, K. (1993(b)). *Une étude de la texture des images satellites appliquée en forêts (Cas d'une région soudano sahélienne (Burkina Faso) et des Monts de Béni-Chougrane (Algérie))*. Thèse de magister, Centre National des Techniques Spatiales (CNTS), Algérie.
- Domingues, ., I., Cardoso, J., Amaral, I., & al., e. (2010). Pectoral muscle detection in mammograms based on the shortest path with endpoints learnt by SVMs. *Annual International Conference of the IEEE*, (pp. 3158–3161).
- D'Orsi CJ, M. E. (2003). *Breast Imaging Reporting and Data System, ACR BI-RADS Mammography* (éd. Fourth Edition). Reston, VA: American College of Radiology: Breast imaging atlas.
- Eltonsy, N. H., Tourassi, G. D., & Elmaghrab, A. S. (2007). A concentric morphology model for the detection of masses in mammography. *IEEE Trans. Med. Imag.*, 26(6), 880–889.
- Etlar, M., & Horsch, A. (2009). CADx of mammographic mass and clustered micro-calcification: a review. *Med. Phy.*, 36(6), 2052-2068.
- Feig, S. A. et al. (1998). American college of radiology guidelines for breast cancer screening. *American Journal of Roentgenology*, 171(1), 29–33.
- Ferrari, R., F. Frère, A., Rangayyan, R., & al., e. (2004). Identification of the breast boundary in mammograms using active contour models. *Medical and Biological Engineering and Computing*, 42(2), 201–208.
- Grim, J., Somol, P., Haindl, M., & al., e. (2009). Computer-aided evaluation of screening mammograms based on local texture models. *IEEE Transactions on Image Processing*, 18 (4), 765–773.
- Guo, M., Dong, M., Wang, Z., Ma, Y., & Gu, Y. (2015). A new method for mammographic mass segmentation based on parametric active contour model. *International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, (pp. 27-33).

-
- Haralick, R. M. (1979). Statistical and Structural Approaches to Texture. *Proceedings of the IEEE*, 67, 786-804.
- Ilea, D. E., & Whelan, P. F. (2011). Image segmentation based on the integration of colour–texture descriptors : A review. *Pattern Recognition*, 44, 2479–2501.
- J Suckling et al. (1994). The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. *International Congress Series 1069*, 375-378.
- Kanta Maitra, I. et al. (2011). Identification of Abnormal Masses in Digital Mammography Images. *International Journal of Computer Graphics*, 2(1).
- Karthikeyan, G., Rajendra Acharyaa, U., Kuang, C. C., Lim, C. M., & Thomas abraham, K. (2013). Pectoral muscle segmentation: A review. *Computer methods and performs in biomedicine*, 110(1), 48-57.
- Kavitha, M., & Rejusha, M. (2015). Segmentation of pectoral muscle and detection of masses in mammographic images. *iee sponsored 2nd international conference on electronics and communication system (ICECS 2015)* , (pp. 1201-1204).
- Kegelmeyer, W., Pruneda, J. M., Bourland, P. D., Hillis, A., Riggs, M. W., & Nipper, M. L. (1994). Computer-aided mammographic screening for speculated lesions. *Radiology*, 191(2), 331–337.
- Kobatake, H., Murakami, M., Takeo, H., & Nawan, S. (1999). Computerized detection of malignant tumors on digital mammograms. *IEEE Transactions on Medical Imaging*, 18 (5), 369–378.
- Kom, G., Tiedeu, A., & Kom, M. (2007.). “Automated detection of masses in mammograms by local adaptive thresholding. *Comput. Biol. Med.*, 37(1), 37–48.
- Kupinski, M., & Giger, M. (1998). Automated seeded lesion segmentation on digital mammograms. *IEEE Transactions on Medical Imaging*, 17(4), 510–517.
- Kwok, S., Chandrasekhar, R., & Attikiouzel, Y. (2001(b)). Automatic pectoral muscle segmentation on mammograms by straight line estimation and cliff detection. *Intelligent Information Systems Conference The Seventh*, (pp. 67–72). Australian et New Zealand.
- Kwok, S., Chandrasekhar, R., Attikiouzel, Y., & al, e. (2004 (a)). Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. *IEEE Transactions on Medical Imaging*, 23(9), 1129–1140.
- Lanyi, M. (1985). *Morphologic analysis of micro-calcifications*. *Early Breast Cancer Springer-Verlag*. (J. Z. Baltzer, Éd.) Berlin.

-
- Laws, K. (1980). *Textured image segmentation*. Report 940, University Southern California, Image Processing Institute, Los Angeles.
- LeGal, M et al. (1984). Valeur diagnostique des micro-calcifications groupées découvertes par mammographie. *Bull Cancer a propos de 227 cas.*, 71, 57–64.
- Li, H., Kallergi, M., Clarke, L., Jain, V., & Clark, R. (1995). Markov random field for tumor detection in digital mammography. *IEEE Transactions on Medical Imaging*, 14(3), 565–576.
- Liu, L., Wang, J., & Wang, J. T. (2011). Breast and pectoral muscle contours detection based on goodness of fit measure. *IEEE 5th International Conference on Bioinformatics and Biomedical Engineering, (ICBBE 2011)*.
- Liu, S., Babbs, C. F., & Delp, E. J. (2001). Multiresolution detection of spiculated lesions in digital mammograms. *IEEE Trans. Image Process.*, 10(6), 874–884.
- Mariano-Goulart, D. (2013(a)). *Introduction au traitement numérique des images médicales, In Traitement De L'image*. (O. H. Mariano-Goulart, Éd.) Paris: Elsevier Masson.
- Mariano-Goulart, D. (2013(b)). *Reconstruction tomographique en imagerie médicale, In Traitement De L'image*. (J. C.-G. livier Hélénon, Éd.) Paris: Elsevier Masson.
- Meenalosini, S., Janet, J., & Kannan, E. (2012). A Novel Approach in Malignancy Detection of Computer Aided Diagnosis. *American Journal of Applied Sciences*, 9(7), 1020-1029.
- Mencattini, A., Salmeri, M., Lojacono, R., Frigerio, M., & Caselli, F. (2008). Mammographic Images Enhancement and Denoising for Breast Cancer Detection Using Dyadic Wavelet Processing. *IEEE transactions on instrumentation and measurement.*, 57(7).
- MIAS. (2012). <http://peipa.essex.ac.uk/info/mias.html>(consulté en 2012).
- Mirmehdi, M., Xie, X., & Suri, J. (2009). *Handbook of texture analysis*. CRC Press.
- Molinara, M., Marrocco, C., & Tortorella, F. (2013). Automatic Segmentation of the Pectoral Muscle in Mediolateral Oblique Mammograms. *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, (pp. 506-509). Porto.
- Mudigonda, N. R., Rangayyan, R. M., & Desautels, J. E. (2000). Gradient and texture analysis for the classification of mammographic masses. *IEEE Trans. Med. Imag.*, 19(10), 1032–1043.
- Mustra, M., Bozek, J., & Grgic, M. (2009). Breast border extraction and pectoral muscle detection using wavelet decomposition. *IEEE EUROCON 2009*, (pp. 1426–1433).

-
- Naga Mudigonda et al. (2001). Detection of Breast Masses in Mammograms by Density Slicing and Texture Flow-Field Analysis. *IEEE transactions on medical imaging*, 20(12), 1215-1227.
- Peter, M., & Sue, A. (1992). Classification of breast tissue by texture analysis. *Image and Vision Computing*, 10(5), 277–282.
- Petrick, N., Chan, H., Sahiner, B., & Wei, D. (1996). An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. *IEEE Transactions on Medical Imaging*, 15(1), 59–67.
- Pfisterer, R., & Aghdasi, F. (1999). Comparison of Texture Based Algorithms for the Detection of Masses in Digitized Mammograms. *IEEE*.
- Qi, G., Jiaqing, S., & Virginie, F. (2008). Characterization and classification of tumor lesions using computerized fractal-based texture analysis and support vector machines in digital mammograms. *Int J CARS*, 4(1), 11–25.
- Raba, D. et al. (2005). *Breast Segmentation with Pectoral Muscle Suppression on Digital Mammograms*. Récupéré sur Robotics and Computer Vision Group: <http://vicorob.udg.es>
- Rangayyan, R., Shen, L., Shen, Y., Desautels, J., Bryant, H., Terry, T., . . . Rose, M. (1997). Improvement of sensitivity of breast cancer diagnosis with adaptive neighborhood contrast enhancement of mammograms. *IEEE Transactions on Information Technology in Biomedicine*, 1(3), 161–170.
- Rocha, A., Tong, F., & Yan, Z. (2000). Logic filter for tumor detection on mammograms. *Journal of Compute Science and Technology*, 15(6), 629–632.
- Sahiner, B., Chan, H. P., Petrick, N., Wei, D., Helvie, M. A., Adler, D. D., & Goodsitt, M. M. (1996). Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images. *IEEE Trans. Med. Imag.*, 15(5), 598–610.
- Sampat, M. P., & Bovik, A. C. (2003). Detection of spiculated lesions in mammograms. *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 1, 810–813.
- Sampat, M., Markey, M., & Bovik, A. (2005(a)). *Computer-Aided Detection and Diagnosis in Mammography in Handbook of Image and Video*. (A. Bovik, Éd.) New York.
- Sampat, M., Markey, M., & Bovik, A. (2005(b)). *Computer-Aided Detection and Diagnosis in Mammography in Handbook of Image and Video*. (A. Bovik, Éd.) New York.

-
- Shi, J., Sahiner, B., Chan, H., Ge, J., Hadjiiski, L., Helvie, M., . . . Cui, J. (2008). Characterization of mammographic masses based on level set segmentation with new image features and patient information. *Medical Physics*, 35(1), 280–290.
- Tang, J., Rangayyan, R., Xu, J., El Naqa, I., & Yang, Y. (2009). Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. *IEEE Transactions on Information Technology in Biomedicine*, 13(2), 236-251.
- telecom.ulg.ac.be. (2013). *www.telecom.ulg.ac.be*. Récupéré sur *www.telecom.ulg.ac.be*: http://www.telecom.ulg.ac.be/teaching/notes/totali/elen016/node113_tf.html, (consulté en 2013)
- Thangavel, K., & Roselin, R. (2012). Fuzzy - Rough Feature Selection with Π -Membership Function for Mammogram Classification. *IJCSI International Journal of Computer Science Issues*, 9(3).
- Timp, S., Varela, C., Karssemeijer, N., & Dacolian, B. (2007). Temporal change analysis for characterization of mass lesions in mammography. *IEEE Trans. Med. Imag.*, 26(7), 945–953.
- Vyborny, C., & Giger, M. (1994). Artificial intelligence in mammography. *AJR* 162, 699-708.
- Wei, J., Sahiner, B., Hadjiiski, L., Chan, H., Petrick, N., Helvie, M., . . . Zhou, C. (2005). Computer aided detection of breast masses on full field digital mammograms. *Med. Phys.*, 32(9), 2827–2837.
- Xiao, M., Xia, S., & Wang, S. (2005). Geometric active contour model with color and intensity priors for medical image segmentation. *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology*, 6496–6499.
- Xiaoming, L., Jun, L., Dongfeng, Z., & Tang, J. (2010). A Benign and Malignant Mass Classification Algorithm Based on An Improved Level Set Segmentation and Texture Feature Analysis. *IEEE*.
- Yao, J., Chen, J., & Chow, C. (2009). Breast Tumor Analysis in Dynamic Contrast Enhanced MRI Using Texture Features and Wavelet Transform. *IEEE Journal of selected topics in signal processing*, 3(1).
- Yu, Y., & Acton, S. T. (2002). Speckle Reducing Anisotropic Diffusion. *IEEE Transaction on image processing*, 11(11).
- Zheng, B. et al. (2006). Multiview-based computer-aided detection scheme for breast masses., 33(9), pp. 3135–3143.

Zou, F., Zheng, Y., Zhou, Z., & Agyepon, K. (2008). Gradient vector flow field and mass region extraction in digital mammograms. *Proceedings of the 21st IEEE International Symposium on Computer-Based Medical Systems*, 41–43.

Liste des Figures

Fig.1.1. Image mdb002, MLO (MIAS Database).....	10
Fig.1.2. Les différentes formes possibles d'une masse (de gauche à droite) : a) Ronde, b) Ovale, c) Lobulée et d) Irrégulière	11
Fig.1.3. Les différents contours possibles d'une masse (de gauche à droite) : a) Circonscrit, b) Micro-lobulé, c) Masqué, d) Indistinct et e) Spiculé	12
Fig.1.4. Les formes et les contours résumés	13
Fig.1.5. Les différents systèmes CAD	18
Fig.2.1. Textures ayant même moyenne et même écart type (<i>M.Van Droogenbroeck, 2003</i>).	30
Fig.2.2. Exemple d'image comportant deux textures (<i>M.Van Droogenbroeck, 2003</i>).....	31
Fig.2.3. Gauche : Image « Brodatz », Droite: Image texturale GLCM	36
Fig.2.4. Gauche : Image de mammographie, Droite: Image texturale GLCM.....	36
Fig.2.5. Principales étapes de la méthodologie proposée	37
Fig.2.6. Etapes Test d'amélioration du contraste de l'image mdb004	40
Fig.2.7. Images ROI avec les étapes test d'amélioration du contraste de l'image mdb004	41
Fig.2.8. Images ROI avec les étapes test d'amélioration du contraste de l'image mdb010	42
Fig.2.9. Images ROI avec les étapes test d'amélioration du contraste de l'image mdb001	42
Fig.2.10. Extraction ROI : (a) MIAS mdb004 input, (b) mdb004 annotée, (c) mdb004 contrastée avec ROI (en rouge) et une masse (en vert), (d) image zoom du ROI	44
Fig.2.11. De gauche à droite : Brodatz D75 image, Image Contraste 0°, Image Contraste 45°, Image Contraste 90°, Image Contraste 135°, Image Contraste Somme (0°+ 45°+ 90°+135°)	45
Fig.2.12. Haut: Image mdb004 contrastée, Bas: Descripteur contraste selon les 4 directions puis image somme qui montre des contours fermés	46
Fig.2.13. De gauche à droite, Brodatz D75 image, Image Contraste dans un masque 3x3, Image Contraste dans un masque 7x7 et Image Contraste dans un masque 9x9.	47
Fig.2.14. Algorithme d'extraction des images texturales.....	48
Fig.2.15. (a) mdb004, deux masses (bleu) dans un tissu dense (rouge), (b) mdb005, une masse (bleu) dans un tissu gras (vert), (c) mdb019, deux masses (bleu) dans un tissu glandulaire (jaune)	51

Fig.2.16. Détection et segmentation de la masse dans un Tissu Dense, sur mdb004 Roi image	52
Fig.2.17. Détection et segmentation de la masse dans un Tissu Gras, sur mdb005 ROI image	52
Fig.2.18. Détection et segmentation de la masse dans un Tissu Glandulaire, sur mdb019 Roi image	52
Fig. 2.19. Tissu dense: de gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff = 93.39%	54
Fig.2.20. Tissu Gras: de gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff. = 97.74%	54
Fig.2.21. De gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff. = 69.35%	55
Fig. 2.22. Tissu Glandulaire: de gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff. = 96.67%	55
Fig.2.23. De gauche à droite: ROI image, Contours de la masse (en blanc) par l'expert, Descripteur de texture, Segmentation de la masse, Dice Coeff. = 86.18%	56
Fig. 2.24. Vue des délimitations de l'expert (à droite), et des contours détectés par l'approche (à gauche)	50
Fig. 2.25. Vue des délimitations de l'expert (à droite), et des contours détectés par l'approche (à gauche)	50
Fig.3.1. Un exemple de GMM à 3 gaussiennes	62
Fig.3.2. Les principales étapes du Modèle GMM/ML	72
Fig.3.3. Zoom sur la région de l'anomalie:(a) image MIAS mdb004,(b) mdbb04	74
Fig.3.4. Segmentation GMM/ML dans un tissu dense (MIAS mdb004)	79
Fig.3.5. Segmentation GMM/ML dans un tissu glandulaire (MIAS mdb019)	80
Fig3.6. Segmentation GMM/ML dans un tissu gras (MIAS mdb005).....	81
Fig.4.1. Principales étapes de la méthodologie proposée	87
Fig.4.2. Exemple d'une image de mammographie MIAS avec sein mal orienté	88
Fig.4.3. A gauche, image avec un sein mal orienté. A droite, image avec un sein bien orienté	88
Fig.4.4. A Gauche, image MIAS mdb001. A droite, mdb001 améliorée	89

Fig.4.5. A Gauche, image MIAS mdb003. A droite, mdb003 améliorée	90
Fig.4.6. A Gauche, image MIAS mdb015. A droite, mdb015 améliorée	90
Fig.4.7(a). A gauche, image input mdb019. A droite, mdb019 améliorée	93
Fig.4.7(b). A gauche, Descripteur de contraste de mdb019. A droite, mdb019 avec extraction du muscle pectoral	94
Fig.4.8. A gauche, mdb019 segmentée. A droite, zoom sur les frontières du muscle pectoral	94
Fig.4.9(a). A gauche, image input mdb012. A droite, mdb012 améliorée	95
Fig.4.9(b). A gauche, descripteur de contraste de mdb012. A droite, mdb012 avec extraction du muscle pectoral	95
Fig.4.10. A gauche, mdb012 segmentée. A droite, zoom sur les frontières du muscle pectoral	96

Liste des tableaux

Tab.1.1. Exemple de Classification TNM	14
Tab.1.2. Catégories et Diagnostiques	15
Tab.1.3 Classification BIRADS de l'ACR correspondant à la masse	16
Tab.2.1. Evaluation de la détection des contours des masses à l'aide du Coefficient du Dice	56
Tab.2.2. Evaluation Quantitative de la détection contours des masses à l'aide du F-Mesure	58
Tab.3.1. Evaluation de la segmentation des masses par le Modèle GMM/ML à l'aide du Coefficient du Dice	81
Tab.3.2. Tableau de comparaison entre la méthode GMM/ML et la méthode GLCM	82
Table.4.1. Evaluation et Comparaison avec d'autres travaux	97