

N° d'ordre :

N° de serie :

République Algérienne Démocratique et Populaire

**Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique**



**UNIVERSITÉ ECHAHID HAMMA LAKHDAR
EL OUED**



FACULTÉ DES SCIENCES EXACTES

Mémoire de fin d'étude

MASTER ACADEMIQUE

Domaine: Mathématiques et Informatique

Filière: Informatique

Spécialité: Systèmes Distribués et Intelligence Artificiel

Présenté par:

MERABET CHAHRAZED

THÈME

**Analyse de la texture des images mammaires par une fusion
Des lois de Zipf et de la transformée de Fourier pour la
classification des tumeurs mammaires via
L'analyse en composantes principales**

Devant le jury composé de :

Dr. HAMOUD MERIEIM

Mr.

Mr.

MAA Univ. ElOued Rapporteur

MAA Univ. El Oued Président

MAA Univ. El Oued Examineur

Remerciements

*J'adresse mes premiers remerciements à mon grand DIEU qui m'a donné la volonté,
le courage ainsi que la patience afin de réaliser ma mémoire master.*

*Nous exprimons toute notre gratitude à docteur **M.HAMOUD** maitre de conférences
classe B à université **Echahid Hamma Lakhdar** – el oued pour nous avoir dirigés et
guidés tout le long de ce travail. Ses conseils et remarques étaient très bénéfiques
pour notre travail. Son soutien permanent ainsi que ses disponibilités pour
l'achèvement de ce travail.*

*Nous remercions vivement les membres du jury qui nous ont fait l'honneur
d'accepter de juger notre travail.*

*N'oublions pas, bien sûr, nos professeurs de tout le cycle
Universitaire , de leur présenter nos meilleurs vœux et enfin, nous tiens à remercier
tous ceux qui ont contribué, de près ou de loin, par leurs encouragements, leurs
conseils et leurs soutiens à mener à bien ce travail.*

Dédicaces

*Avant tout, je remercie **ALLAH** le tout puissant de m'avoir donné le
Courage et la patience pour réaliser ce travail malgré toutes les
Difficultés rencontrées.*

*Tous d'abord Je dédie ce modeste travail À ma l'esprit de mon cher père la
miséricorde de Dieu .*

*À mon cher **mère** Je le dédie aussi À ma très cher sœur **HAYAT** qui m'a encouragé
par tous les moyens à avoir la confiance et l'espoir au cours
de la rédaction de ce travail. Adorée et qui m'a aidé, grâce à leur prière et à leur
bénédiction.*

*À ma très cher sœur **LAMIA** et ta petit **AMINE**.*

*À mes très chers frères **SALIM, WALID**.*

*À toute la famille **MERABET**.*

*À mes meilleurs amies **MEDJDA, OUACILA , NOUR ELHOUDA ,HAYAT**.*

À tous mes coulages de promotion master 2 de l'Informatique 2018.

CHAHRAZED MERABET

إن اختيار العلاقات الغير الخطية له أهمية بالغة في تطوير أدوات تحليل الصور القوية وأدوات رؤية الكمبيوتر. وبهذا الصدد ، فإن مشكلتنا البحثية هي تطبيق قوانين القوة: زيب اف وزيب اف معكوس لتحليل صور الثدي. في الواقع ، نتميز قوانين زيب اف بـالتعقيد الهيكلي لبنية الصورة عبر نمذجة التوزيع الإحصائي لوتيرة ظهور الأنماط وفقاً لتوزيع في قانون القوة . وعلاوة على ذلك ، نحن دمجنا الوصفات التي تم الحصول عليها مع تلك التي تم استخراجها بعد تحليل الصور الثديية عن طريق تحويل فورييه المنفصل.

لقد قمنا باقتراح نظام لفهرسة التصوير بالأشعة السينية ونظام البحث الذي يعزز أداء التشخيص بمساعدة الكمبيوتر في مرحلة تقديم التشخيص إلى أطباء الأشعة. مع العلم ، هم أكثر ثقة في حكم التشخيص المعتمد على الحالة الذي يتم إرجاعه من خلال تصنيف قائم على مطابقة القالب ، تم تشخيصه مسبقاً ، على غرار الحالة التي يتم تحليلها بدلاً من النتيجة المجردة الناتجة عن المصنف. أعطى تقييم العمل المقترح نتائج مشجعة.

الكلمات الرئيسية

تحليل الصور ، رؤية الكمبيوتر ، قانون زيب اف و زيب اف معكوس ، التصنيف ، الفهرسة ، بحث المحتوى ، تحويل فوريير المنفصل ، مطابقة القالب .

Résumé

Le choix de la non-linéarité est d'un intérêt crucial dans le développement de puissants outils d'analyse d'image et de vision par ordinateur. Dans ce sens, notre problématique de recherche consiste à appliquer les lois puissance : Zipf et Zipf inverse à l'analyse des images mammaires. En effet, les lois de Zipf caractérisent la complexité structurelle de la texture d'image par la modélisation de la répartition statistique de la fréquence d'apparition des motifs selon une distribution en lois puissance. En outre, nous fusionons les descripteurs obtenus avec ceux générés suite à l'analyse des images mammaires par la transformée de Fourier discrète.

Nous avons proposé un système d'indexation et de recherche des mammographies par le contenu (CBMIIR) qui renforce la performance du diagnostic assisté par ordinateur au niveau de l'étape de la présentation du diagnostic aux radiologues. En effet, ces derniers sont plus confiants d'un jugement de diagnostic basé sur des cas renvoyé par une classification basée template-matching, diagnostiqués préalablement, similaires au cas en cours d'analyse plutôt que le résultat abstraits généré par un classifieur. L'évaluation des travaux proposés a donnée des performances encourageantes.

Mots clés : Analyse d'image, Vision par ordinateur, Loi de Zipf, Loi de Zipf inverse, Classification, Indexation, Recherche par le contenu, transformée de Fourier discrète, template-matching.

Abstract :

The choice of the non-linearity is of crucial interest in the development of powerful tools for image analysis and computer vision. In this sense, our research problematic is to apply the power laws: Zipf and inverse Zipf for mammogram images analysis. Indeed, the laws of Zipf characterize the structural complexity of the image texture by modeling the statistical distribution of patterns frequency of appearance as power law distribution. In addition, we have performed a fusion of the obtained texture features with those generated once applying discrete Fourier transform for mammogram images analysis. Subsequently, we have proposed a content based mammogram image indexing and retrieval system (CBMIIR) that boosts the performance of a computeraided diagnosis (CADx) at the stage of providing the diagnostic to radiologists. Indeed, radiologists feel more confident in their diagnosis decision based upon case-adaptive classification via the template-matching technique, where similar known cases, to the one under analysis, are retrieved and displayed from indexed databases; rather than the abstract result generated by a classifier. The evaluation of the proposed systems has given encouraging performance.

Keywords

Image analysis, computer vision, Zipf's law, inverse Zipf's law, Classification, Indexing, Retrieval by content, discrete Fourier transform, template-matching.

Remerciements.....	ii
Dédicace	iii
ملخص	iv
Résumé.....	v
Abstract.....	vi
Table des matières.....	vii
Liste des figures	xiii
Liste des tableaux.....	xviii

Table des matières

Introduction générale.....	1
Chapitre I : Analyse d'image et vision par ordinateur	
I.1 Introduction.....	4
I.2. Relation entre l'analyse d'image et la vision par ordinateur.....	5
I.3. Analyse d'image.....	5
I.3.1. Analyse de bas niveau d'image	5
I.3.2. Analyse de haut niveau d'image	6
I.4. L'imagerie médicale.....	7
I.5. Analyse de la texture d'une image.....	7
I.5.1. Définition de la texture.....	7
I.5.2. Les méthodes d'analyse de la texture.....	8
I.5.3. Les descripteurs de la texture	9
I.6. La vision par ordinateur	9
I.7. Conclusion	11
Chapitre II : les lois puissance	
II.1 introduction	13
II.2 définition d'une loi puissance.....	14
II.3 les principales lois puissance	15
II.3.1 la loi de Pareto.....	15
II.3.2 la loi de Zipf	15
II.3.3 la loi de Zipf inverse	15
II.4 application des lois puissance en analyse d'images.....	16
II .4.1. Les lois de Zipf et de Zipf inverse en imagerie.....	16

II.4.2 Codage de l'image	17
II.4.3 Construction des courbes de Zipf et de Zipf inverse.....	18
II.5 conclusion.....	20
Chapitre III : Aide au diagnostic médical du cancer du sein assisté par Ordinateur	
III.1.Introduction	22
III.2. Le cancer du sein.....	23
III.3. Le dépistage du cancer du sein	24
III.3.1 La mammographie.....	24
III.4. Détection et classification assistées par ordinateur (CAdE/CAdx) des tumeurs dans la mammographie.....	25
III.4.1. Détection des tumeurs assistée par ordinateur (CAdE)	26
III.4.2. Diagnostic des tumeurs assisté par ordinateur (CAdx).....	27
III.5. Indexation et recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur.....	27
III.6. L'apport d'un système d'indexation et de recherche de mammographies par le contenu (CBMIIR) par rapport à un système d'aide au diagnostic assisté par ordinateur (CAdx) se basant sur la sortie d'un classifieur.....	27
III.7. Approche proposée basée fusion des lois de puissance : Zipf, Zipf inverse et la transformée de Fourier discrète pour la classification des zones d'intérêt des mammographies.....	28
III.7.1. Analyse et caractérisation de la texture des zones d'intérêt par les lois de Zipf et de Zipf inverse.....	28
III.7.2. Analyse et caractérisation de la texture des zones d'intérêt par la transformée de Fourier discrète.....	31
III.7.3.bilan.....	34
III.8. Conclusion.....	35
Chapitre IV. Conception et Implémentation	
IV.1.Introduction	37
IV.2. Problématique	38
IV.3. Principe de la classification basée template-matching et les K plus proches voisins	39
IV.3.1. Algorithme des K-plus proches voisins.....	42
IV.3.2. L'analyse en composantes principales	42
IV.4. Matériel et outils utilisés pour le développement du système proposé	45

IV.4.1.Le matériel utilisé	45
IV.4.2.Les Outils utilisés	45
IV.5. La base des mammographies utilisée	48
IV.6. Fonctionnement du système développé	48
IV.7. Evaluation des performances	49
IV.8. Conclusion	51
Conclusion générale	52
Références	53

Liste de figures

Figure II.1 Représentation d'une loi puissance dans un repère linéaire	14
Figure II.2 Représentation d'une loi puissance dans un repère bi-logarithmique.....	14
Figure II.3 Motif original (a) et sur (b) le motif codé avec la méthode des neuf classes.....	17
Figure II.4 Motif original (a) et sur (b) le motif codé avec la méthode des rangs généraux.....	18
Figure II.5 Courbe de Zipf de l'image Cameraman.....	19
Figure II.6 Courbe de Zipf inverse de l'image Cameraman.....	19
Figure III.1 Appareil de génération des mammographies numériques.....	24
Figure III.2 Mammographie normale.....	25
Figure III.3 Mammographie portant une tumeur bénigne.....	25
Figure III.4 Mammographie portant une tumeur maligne.....	25
Figure III.5 Différence entre le système CADe et le système CADx.....	26
Figure III.6 Courbes de Zipf et de Zipf inverse de trois zones d'intérêts encodées par le codage des rangs généraux.....	29
Figure III.7 Spectre de Fourier (à droite) généré à partir d'une zone d'intérêt (à gauche) portant une tumeur bénigne.....	33
Figure IV.1 L'architecture du système mammo-ZipDFT.....	41
Figure IV.2 Légende relative à la figure IV.1.....	42
Figure IV.3 L'interface de Matlab 2017a.....	46
Figure IV.4 L'interface principale du système Mammo-ZipDFT.....	48
Figure IV.5 Processus de décision médicale issue de l'utilisation de la sortie de Mammo_ZipDFT par la considération de $k=3$	49

Liste des tableaux

Tableau II.1 Classification des niveaux de gris en 9 classes.....	17
Tableau IV.1 Variation de la précision de classification du système «Mammo_ZipDFT» pour diverses valeurs des k plus proche voisins.....	50

Introduction générale

Durant ces dernières décennies, l'augmentation considérable des taux de cancer du sein est devenu un problème de santé publique dans le monde. En effet, il s'agit d'un fameux cancer menaçant la vie de la plupart des femmes et devenant l'une des causes majeures de mortalités chez les femmes.

Dans l'objectif d'assurer le dépistage précoce de cette maladie dangereuse, les radiologues ont été amenés à augmenter la fréquence des mammographies car le diagnostic tardif amène souvent à un traitement lourd et coûteux provoquant un taux de mortalité élevé.

Dans ce contexte, la vision par ordinateur vise à réaliser la compréhension de haut niveau d'imageries médicale. Ceci, en automatisant les tâches du système visuel humain et impliquant le développement d'une base théorique et algorithmique pour atteindre la compréhension visuelle automatique. A cet effet, les systèmes d'aide au diagnostic médical du cancer du sein assisté par ordinateur semblent un bon remède aux radiologues qui saisissent le diagnostic fournie par l'ordinateur comme un deuxième avis pour l'élaboration de leur diagnostic final.

La problématique prise en considération pour la réalisation de ce travail est l'analyse des images mammaires à travers des techniques non linéaires qui sont plus adéquates à la structure complexe d'une image, ceci, en évitant les techniques linéaires qui montrent vite leurs limites. Effectivement, nous nous intéressons aux lois puissance : Zipf et Zipf inverse pour la caractérisation de la texture des images mammaires. Entre autre, nous proposons une fusion des lois de Zipf et de Zipf inverse avec l'approche de la transformée de Fourier discrète pour bénéficier de l'apport complémentaire que peut avoir ces deux techniques de quantification texturale. Notre ultime but est d'essayer d'améliorer les résultats obtenus par l'application des lois de Zipf et de Zipf inverse uniquement.

Dans ce sens, nous développons un système d'indexation et de recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein. En effet, nous nous sommes basés sur le fait qu'un système d'aide au diagnostic médical du cancer du sein assisté par ordinateur et se basant sur la sortie d'un classifieur (CADx) offre le diagnostic à travers des chiffres abstraits générés suite à la sortie d'un classifieur, auxquels les radiologues ne sont pas familiarisés et que les radiologues préfèrent recevoir l'aide à base de cas déjà traités et stockés dans une base de donnée ce qui désigne le principe des systèmes CBMIIR.

Notre travail est divisé en quatre chapitres principaux :

Chapitre I : Analyse d'image et vision par ordinateur

Nous évoquons dans le premier chapitre les notions et concepts de base de l'analyse d'image et la vision par ordinateur, d'autre part, les méthodes d'analyse de la texture des images seront présentées.

Chapitre II : Les lois puissance

Le deuxième chapitre aborde les lois puissance qui sont utilisées dans ce travail et comment les appliquer en analyse d'image.

Chapitre III : Aide au diagnostic médical du cancer du sein assisté par ordinateur

Dans le troisième chapitre nous présentons les techniques assistées par ordinateur d'aide au diagnostic médical du cancer du sein dans le but de donner un deuxième avis aux radiologues durant l'élaboration de leur diagnostic.

Chapitre IV : Conception et implémentation

Dans le quatrième chapitre nous présentons la conception mise au point pour le développement notre système.

Nous terminons notre mémoire par une conclusion générale.

Chapitre I

Analyse d'image et vision par ordinateur

I.1 Introduction

Durant cette dernière décennie, la vision par ordinateur fait une grande révolution dans le domaine d'analyse et d'interprétation d'image et plus précisément le domaine de l'imagerie médicale qui constitue l'outil principal de la vision par ordinateur, en abordant des problèmes dont les humains résolvent à leur insu.

Nous distinguons deux phases dans la vision par ordinateur : problèmes de "bas niveau" et de "haut niveau". La vision de bas niveau englobe des problèmes tels que le codage et la compression de données en vue de réaliser des processus de stockage et de transmission, la numérisation et bien d'autres. En revanche, la vision de haut niveau concerne des objectifs plus spécifiques, citons la reconnaissance d'objets ainsi que l'interprétation des scènes.

Nous présentons dans le premier chapitre les concepts de base de l'analyse d'image et de la vision par ordinateur. De plus, nous évoquons les descripteurs de la texture étant parmi les descripteurs les plus pertinents en vision par ordinateur. Effectivement, notre travail porte sur la caractérisation de la texture par une fusion des lois puissance : Zipf et Zipf inverse et la transformée de Fourier discrète pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur.

I.2. Relation entre l'analyse d'image et la vision par ordinateur

L'ultime but de l'interprétation d'une image par une machine est de trouver une relation entre l'image d'entrée et les modèles du monde réel (Sonka 2008). En effet, ceci est généralement divisé en plusieurs étapes, donc plusieurs niveaux représentant l'image sont utilisés ; citons : le niveau inférieur contenant les données brutes ainsi que d'autres niveaux élevés permettant l'interprétation des données. La vision par ordinateur intervient pour la concevoir ces représentations intermédiaires et les algorithmes servent à établir et maintenir les relations entre les entités au sein et entre les couches.

I.3. Analyse d'image

Une image est une représentation planaire d'une scène ou d'un objet situé dans un espace à trois dimensions (+ la dimension temporelle). L'acquisition, le traitement et la restitution d'une image s'apparentent à une chaîne de traitement (traitement d'un signal) avec toutes ses problématiques notamment les capteurs (Hamoud 2015) .

Une image numérique est constituée d'une grille rectangulaire d'échantillonnage dont les constituants sont des pixels portant des informations sur l'intensité lumineuse des différents lieux au sein de l'image. De ce fait, une image numérique est représentée par une matrice bidimensionnelle, dont les éléments sont des nombres naturels correspondant à des niveaux de quantification dans l'échelle de l'intensité lumineuse. Tout d'abord une image d'entrée est capturée par dispositif de capture d'image en (2D) et numérisée, étant décrite par une fonction d'image $f(x, y)$ dont la valeur est l'intensité lumineuse en fonction de deux paramètres x, y , représentant les coordonnées de l'emplacement dans l'image. Une image n'est qu'une représentation imparfaite d'une scène et élaborer une image correspond à la volonté de proposer une entité observable par l'œil humain.

En pratique, l'analyse d'images est une suite de phases qui doivent être exécutées l'une après l'autre, certaines de ces phases sont souvent étroitement liées et le but consiste en à partir des informations initiales, de chercher à extraire des informations plus pertinentes, et à les interpréter (Hamoud 2015) .

Comme nous l'avons déjà mentionné, les niveaux d'analyse d'une image peuvent se résumer en analyse de bas niveau et analyse de haut niveau.

I.3.1. Analyse de bas niveau d'image

Elle consiste en trois étapes fondamentales et indissociables :

- **Traitement**

Dans ce niveau l'image d'entrée est décrite par la fonction d'image (x,y) , que nous l'appliquons un traitement qui consiste en l'élimination du bruit au sein de l'image, voir plus, l'amélioration de certains descripteurs des objets jugés pertinents pour interpréter l'image. Généralement l'entrée et la sortie de ce niveau sont des images d'intensité (Hamoud 2015).

A cet effet, l'étape de traitement sert à améliorer la qualité de l'image avant les manipulations. Son objectif principal est d'augmenter le contraste entre la région d'intérêt et le reste de l'image pour faciliter les traitements ultérieurs. Ceci peut être réalisé par l'égalisation d'histogramme ou par l'étirement d'histogramme. Toutefois, ces transformations ont l'inconvénient de faire ressortir le bruit dans les images déjà bruitées (Cheikhrouhou 2012).

- **Segmentation**

La phase de segmentation est très importante puisque les traitements ultérieurs (extraction des descripteurs et l'interprétation) sont fortement liés au résultat de la segmentation, dans laquelle l'ordinateur fait en sorte de séparer les objets de l'arrière plan de l'image. En effet, c'est le processus de regroupement des données d'image dans des régions disjointes (classes de pixels) sur la base de propriétés locales telles que la couleur, la profondeur, la texture ; ou sur la base de critères plus globaux (ou même sémantiques), par exemple concernant les dichotomies telles que l'objet/fond ou bénigne/maligne (Hamoud 2015)

- **Extraction et sélection des descripteurs :**

Le but est d'extraire les caractéristiques qui décrivent au mieux et de façon quantitative ou qualitative les objets présents dans l'image. L'extraction et la sélection des descripteurs sont basées sur la couleur, la texture et/ou la forme (région d'intérêt) (Cheikhrouhou 2012).

I.3.2. Analyse de haut niveau d'image

L'Analyse de haut niveau d'image s'appuie sur les données de l'image issues de l'analyse de bas niveau d'image contenant des connaissances relatives au contenu de l'image. La vision de haut niveau tente de commander les étapes d'analyse d'image à l'aide de toutes les connaissances. Généralement l'analyse de haut niveau d'image considère seulement les données qui sont pertinentes et qui se rapportent à des objectifs de haut niveau réduisant considérablement la quantité de données.

Notons que les méthodes d'intelligence artificielle sont largement applicables dans le domaine de l'analyse et l'interprétation de l'image. En effet, la vision par ordinateur de haut niveau

tente d'imiter la cognition humaine et la capacité à prendre des décisions en fonction de l'information contenue dans l'image.

I.4. L'imagerie médicale

De multiples techniques d'imagerie du corps humain sont couramment employées dans le domaine médical (Cheikhrouhou 2012).

Chacune d'elles est sensible à un type de contraste particulier et trouve ses applications pour des organes différents. Le but de l'imagerie médicale est de fournir une représentation visuelle intelligible qui après analyse révèle une information à caractère médical.

I.5. Analyse de la texture d'une image

I.5.1. Définition de la texture

Malheureusement, à notre jour, les chercheurs tentent toujours de s'entendre sur une définition universelle du terme texture, en effet, les définitions mathématiques proposées pour les concepts de la texture peuvent être trop vastes conduisant à ne pas couvrir tous les cas de texture envisageables.

En conséquence, les définitions de texture proposées varient selon les domaines de recherche et les conceptions des auteurs comme suit(Hamoud 2015) :

- Haralick (Haralick, 1979) a décrit la notion de texture comme suit : « Une image texturée est décrite par le nombre et les types de ses primitives tonales ainsi que leurs orientations spatiales. Elle ne peut pas être analysée sans une vue de la primitive tonale de référence. Au fur et à mesure que la résolution augmente, nous observons une texture fine puis une texture grossière ».
- Dans sa thèse, Unser (Unser, 1984) donne une définition perceptuelle de la texture : « Une texture est une région d'une image pour laquelle il est possible de définir une fenêtre de dimensions minimales, telle qu'une observation au travers de celle-ci se traduit par une perception (impression) visuelle identique pour toutes les translations possibles de cette fenêtre à l'intérieur de la région considérée ».
- Drbohlav et Leonardis (Drbohlav et al. 2010) définissent la texture comme des structures de base spatialement adjacentes dont leur placement et orientation sont gouvernés par certaines règles de génération.
- Dans sa thèse (Hamoud 2015), la notion de texture dépend de trois aspects : le premier est qu'un motif de l'image se composant d'un certain nombre de pixels

représentés par leur niveaux de gris, se répètent quasi périodiquement dans une région de taille supérieure à celle du motif, le deuxième est que nous observons un agencement non aléatoire des motifs de l'image et le troisième est que les motifs considérés sont des entités uniformes ayant les mêmes dimensions dans la région texturée.

I.5.2. Les méthodes d'analyse de la texture

Les méthodes d'extraction et d'analyse à base de texture peuvent être divisées en quatre classes (Chen et al 1997) (Tuceryan et al 1998).

- **Les méthodes statistiques**

Analysent la distribution spatiale des valeurs de niveaux de gris par le calcul des indices locaux dans l'image et déduisent par la suite un ensemble de statistiques (Haralick, 1979). Ces modèles sont importants pour de nombreuses textures naturelles ainsi que pour la caractérisation des structures fines.

- **Les méthodes géométriques**

Elles sont utilisées pour décrire les motifs complexes et déduire les propriétés des textures (Tuceryan et al 1998). Ces méthodes permettent de caractériser les propriétés géométriques des textures et découvrir les règles qui règlent leur organisation spatiale. Ces modèles sont efficaces pour décrire les textures où il y a une grande résolution de l'image, citons les méthodes structurelles classiques, les méthodes syntaxiques et les méthodes ensemblistes (Amroun 2013).

- **Les méthodes à base de modèle**

Estiment un modèle paramétrique en fonction de la distribution d'intensité des descripteurs de texture calculés. Les méthodes à base de modèles probabilistes sont largement utilisées telles que les champs aléatoires conditionnels (Lafferty 2001).

Les Champs de Markov et les fractales sont les deux méthodes les plus utilisées dans cette catégorie. La dimension fractale est utilisée pour mesurer la rugosité d'une texture et la répétitivité (spatiale ou à différentes résolutions) d'un motif (Journet et al 2007).

- **Les méthodes fréquentielles**

Elles analysent les fréquences de l'image. Les méthodes fréquentielles les plus utilisées sont les filtres de Gabor, la transformée de Fourier, les transformées en ondelettes et les méthodes de segmentation à base de calcul de moments (Hamoud 2015).

I.5.3. Les descripteurs de la texture

Dans le processus d'analyse d'image, notamment, après le traitement et la segmentation, l'étape suivante est l'extraction des descripteurs discriminants à partir des zones d'intérêt.

Nous nous intéressons aux descripteurs texturaux. En effet, la transformée de Fourier a été parmi les premières techniques utilisées pour l'analyse de la texture où nous calculons généralement le spectre de puissance qui est défini comme le carré du module de la transformée de Fourier et qui est, d'après (Petrou et Garcia-Sevilla, 2006), très informatif sur la périodicité d'une texture (Hamoud 2015).

Citons également la matrice de co-occurrence (Haralick, 1979) des niveaux de gris de l'image qui indique, dans une texture, le nombre de paires de pixels ayant des niveaux de gris (i, j) selon une direction et une distance donnée d. Les descripteurs calculés sur la matrice de co-occurrence permettent de caractériser la régularité, la répétitivité et le contraste des textures. La fonction d'auto corrélation à son tour permet l'obtention d'informations sur les caractéristiques d'une texture donnée, en effet, si la texture est grossière c'est-à-dire ayant des motifs larges alors la fonction baisse lentement par l'augmentation de la distance d'analyse. Autrement, si la texture est fine c'est-à-dire ayant des petits motifs peu espacés alors la fonction décroît rapidement (Uttama et al 2005).

I.6. La vision par ordinateur

La vision par ordinateur est une discipline d'ingénierie cherchant à appliquer ses théories et ses modèles pour la construction de systèmes de vision par ordinateur, ceci, en se basant sur la compréhension des lois fondamentales de la nature. En effet, son ultime but est la reproduction de l'effet de la vision humaine en percevant électroniquement une image.

La vision par ordinateur vise à réaliser la compréhension de haut niveau d'images numériques, de séquences vidéo, de vues de plusieurs caméras ou données multidimensionnelles. Ceci, en

automatisant les tâches du système visuel humain et impliquant le développement d'une base théorique et algorithmique pour atteindre la compréhension visuelle automatique.

La majorité des techniques de vision par ordinateur se basent sur les résultats issus des méthodes de mathématiques, reconnaissance de formes, intelligence artificielle, psycho-physiologie, informatique, électronique, et d'autres disciplines scientifiques. En fait, le but est la compréhension de l'image, la capacité non seulement de récupérer la structure de l'image, mais aussi de savoir ce qu'elle représente (Hamoud 2015).

Lorsqu'un humain analyse une image, il exprime l'information contenu au sein de cette dernière en utilisant des descripteurs faciles à percevoir, consistants en les descripteurs psychophysiques tels que la couleur, le contraste, les contours, la forme, la texture...etc. Nous allons nous intéresser à l'analyse de la texture qui est un domaine important pour la vision par ordinateur. En effet, l'analyse de la texture a priori un grand intérêt en vision humaine et en psychophysique, mais, les caractéristiques de texture peuvent ne pas être perçues de façon évidente par la vision humaine. Les chercheurs en analyse d'image et vision par ordinateur visent à caractériser cette texture à travers des descripteurs numériques pertinents et discriminants permettant la détection des objets et régions saillantes.

A cet effet, nous allons nous baser sur la caractérisation de la texture par une fusion des lois puissance : Zipf et Zipf inverse et la transformée de Fourier discrète afin de bénéficier de leur apport complémentaire dans un processus d'aide au diagnostic médical du cancer du sein assisté par ordinateur.

I.7. Conclusion

Ce chapitre nous a permis d'ouvrir une fenêtre sur le processus d'analyse d'image numérique qui constitue l'outil principal de la vision par ordinateur. En effet, l'analyse d'image et la vision par ordinateur présentent un champ de recherche actif de nos jours.

Dans ce sens, nous nous focaliserons sur cet enjeu durant l'élaboration de ce mémoire. En effet, nous nous intéressons aux modèles de loi puissance de type: Zipf et Zipf inverse ainsi qu'à la transformée de Fourier discrète pour la caractérisation de la texture des zones d'intérêt extraites à partir d'images mammaires pour offrir un aide au diagnostic médical du cancer du sein assisté par ordinateur.

Chapitre II

Les lois puissance

II .1. Introduction

Dans de nombreuses études, des lois puissance ont été appliquées pour modéliser des phénomènes extrêmement divers. Nous nous intéressons aux lois puissance Zipf et Zipf inverse qui établissent une relation entre les fréquences d'apparition des différentes occurrences d'un phénomène et le rang de ces occurrences dans une suite ordonnée.

Dans ce chapitre, il convient d'abord de donner une définition des lois puissance, spécialement les lois de Zipf et de Zipf inverse, qui sera suivie par le processus d'application de ces lois dans l'analyse d'image et la vision par ordinateur.

II .2. Définition d'une loi puissance

Les lois puissance sont omniprésentes dans différents domaines scientifiques (physique, biologie, psychologie, sociologie, économie, linguistique).

Une loi puissance est une fonction qui relie deux quantités x et y et qui prend la forme :

$$Y = ax^{-b} \quad (\text{II.1})$$

Où a et b sont des constantes. Sa représentation dans un repère linéaire est comme suit (figure II.1):

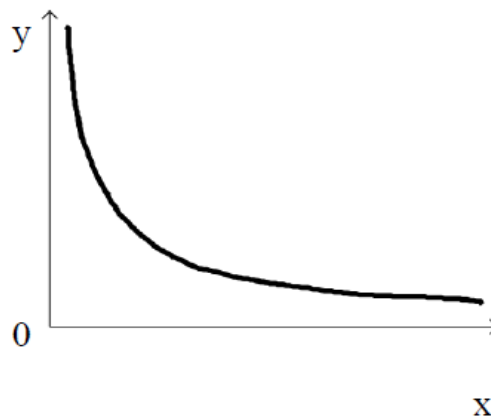


Figure II.1 Représentation d'une loi puissance dans un repère linéaire.

En général, nous représentons les lois puissance dans un repère bi-logarithmique. Dans un tel repère, la représentation graphique d'une distribution en loi puissance est une droite comme indiqué sur la figure II.2.

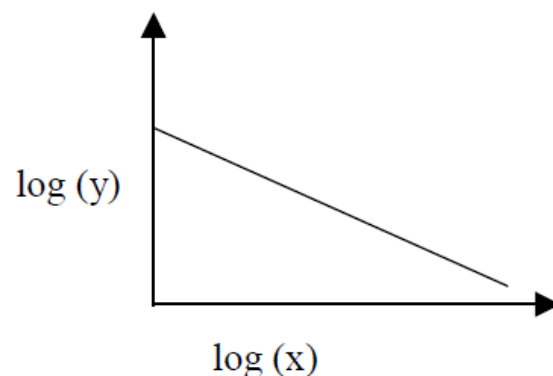


Figure II.2 Représentation d'une loi puissance dans un repère bi-logarithmique.

Nous aurons alors $\ln y = \ln a - b \ln x$. Les modèles de lois puissance les plus connus sont la loi de Pareto et la loi de Zipf (Caron 2004) (Anderson 2004).

II .3. Les principales lois puissance

II .3.1. La loi de pareto

La première mise en évidence d'une loi puissance a été faite en 1897 par l'économiste italien Vilfredo Pareto en étudiant la répartition des revenus personnels des individus dans les principaux pays industrialisés, en effet, Pareto a constaté que cette répartition suivait une loi puissance (Hamoud 2015) (Caron 2004) .

Cette loi est généralement valable pour tous les pays à économie de marché : aussi bien pour les revenus des particuliers que pour les chiffres d'affaire des entreprises.

En 1906 il étudie la répartition des revenus des habitants de divers pays industrialisés et constate que 80% des richesses appartiennent à 20% de la population et ceci, quel que soit le pays observé ; environ 80% des effets est le produit de 20% des causes, 20% des produits fabriqués par une entreprise représentent 80% du volume de production de celle-ci, 20% du réseau routier supporte 80% du trafic, 20% des individus sont responsables de 80% des problèmes, ainsi que d'autres exemples encore (Caron 2004).

II .3.2. La loi de Zipf

La loi de Zipf a été mise en évidence par le linguiste américain George K en 1949. C'est une loi empirique qui s'intéresse à la répartition statistique des fréquences d'apparition des différents mots dans un langage. Selon la loi de Zipf, les fréquences d'apparition des mots classés par ordre décroissant de leur fréquence d'apparition s'organisent suivant une loi puissance. En effet, si nous désignons par F la fréquence d'apparition du mot de rang r dans la suite, la relation est donnée par la formule suivante :

$$F = a r^{-\alpha} \quad (II.2)$$

Dans cette formule, a et α sont deux constantes positives.

II .3.3. La loi de Zipf inverse

Une autre loi puissance a été mise en évidence par Zipf durant ses travaux sur la langue anglaise consistant en la loi de Zipf inverse. Contrairement à la loi de Zipf décrivant la répartition statistique des mots, la loi de Zipf inverse décrit la répartition statistique des fréquences des mots. Selon la loi de Zipf inverse, le nombre de mots *distincts* ayant une

fréquence d'apparition donnée dans un texte suit une loi puissance. En effet, le nombre $I(f)$ de mots apparaissant f fois dans le texte est donné par la formule suivante :

$$I(f) = A f^{-\gamma} \quad (\text{II.3})$$

Où A et γ sont des constantes.

Contrairement à la loi de Zipf qui s'intéresse davantage aux mots les plus fréquents, la loi de Zipf inverse permet donc d'étudier la répartition des mots les plus rares dans le texte.

II .4. Application des lois puissance en analyse d'image

L'application des lois de Zipf en analyse d'images consiste en la modélisation de la fréquence d'apparition des motifs de l'image selon des distributions en loi de puissance; en effet, les lois puissance caractérisent la complexité structurelle des textures dans les images. cependant, cette analyse génère des courbes de Zipf et de Zipf inverse, à partir desquelles nous extrayons une famille de descripteurs de la texture.

II .4.1. Les lois de Zipf et de Zipf inverse en imagerie

- **La loi de Zipf**

Les n -uplets de symboles dans le cas des images seront des motifs constitués de pixels adjacents. Un motif est une matrice carrée de pixels adjacents de l'image centrée sur un pixel. Le motif $M_{2k+1}(i,j)$ de taille $2k+1$ centré sur le pixel $v(i,j)$ est défini ainsi pour $k > 1$:

$$M_{2k+1} = \begin{pmatrix} v(i-k, j-k) & \dots & \dots & \dots & v(i-k, j+k) \\ & & & & \\ & & v(i, j) & & \\ & & & & \\ v(i+k, j-k) & \dots & \dots & \dots & v(i+k, j+k) \end{pmatrix}$$

Les fréquences $N_1, N_2 \dots N_n$ d'apparition des motifs $M_1, M_2 \dots M_n$ de l'image, triées selon l'ordre décroissant en leur attribuant un rang s'expriment par la formule suivante:

$$N_{\sigma(i)} = k \cdot i^{-a} \quad (\text{II.4})$$

Où $N_{\sigma(i)}$ représente le nombre d'apparitions d'un motif de rang i , K et a sont des constantes.

- **La loi de Zipf inverse**

D'après la loi de Zipf inverse [13], le nombre I de motifs différents ayant une fréquence d'apparition f est donné par la formule suivante :

$$I(f) = af^{-b} \quad (\text{II.5})$$

Dans cette formule, a et b sont des constantes positives.

II.4.2 Codage de l'image

Il est nécessaire de définir un codage des motifs de l'image réduisant le nombre des motifs distincts et qui permet l'apparition d'un motif plusieurs fois dans l'image (Caron 2004) (Regniers 2014).

- **Le codage des 9 classes**

L'utilisation de ce codage est basée sur le choix du nombre de classe présentant une ambiguïté pour l'implémentation de ce codage d'image, en effet, il a été constaté que $n=9$ fut le choix optimal en partitionnant l'intervalle $[0, 255]$ des niveaux de gris en un nombre de 9 intervalles, ou de classes de largeur égale, et numérotées en ordre croissant de 0 à 8. Le principe de ce codage consiste à attribuer à chaque pixel du motif, sa valeur de classe $c(x, y)$ en fonction de la valeur $g(x, y)$ de son niveau de gris, en utilisant la formule suivante (Caron 2004) (Hmoud2015) :

$$c(x,y) = \text{int} \left[\frac{n \cdot g(x,y)}{255} \right] \quad (\text{II.6})$$

Le tableau II.1 présente une classification des niveaux de gris en 9 classes:

0-27	28-55	56-83	84-111	112-139	140-167	168-195	196-223	224-255
0	1	2	3	4	5	6	7	8

Tableau II.1 Classification des niveaux de gris en 9 classes (Hamoud 2015).

Nous mentionnons sur la figure suivante un exemple d'un motif avant et après avoir effectué son codage par la méthode des 9 classes (Hamoud 2015).

(a)			(b)		
0	20	250	0	0	8
70	10	213	2	0	7
52	125	120	1	4	4

Figure II.3 Motif original (a) et sur (b) le motif codé avec la méthode des neuf classes.

- **Le codage des rangs généraux**

Le principe du codage des rangs généraux consiste à remplacer les niveaux de gris des pixels par leur rang dans un voisinage.

Le codage d'un motif s'effectue selon les étapes suivantes: nous ordonnons les niveaux de gris des pixels du motif selon un ordre croissant et nous affectons la valeur 0 au niveau de gris le plus bas, tout en incrémente la valeur d'une unité jusqu'à atteindre le niveau de gris le plus élevé. Les pixels ayant la même valeur de niveau de gris recevront le même rang.

La figure II.4 présente un exemple d'un motif avant et après avoir effectué son codage par la méthode des rangs généraux. Nous pouvons remarquer que ce codage permet de mettre en évidence une information plus précise concernant la texture de l'image (Caron 2004) (Hmoud 2015).

(a)			(b)		
0	28	42	0	1	4
29	56	42	2	6	4
41	43	60	3	5	7

Figure II.4 Motif original (a) et sur (b) le motif codé avec la méthode des rangs généraux.

II.4.3 Construction des courbes de Zipf et de Zipf inverse

- **L'algorithme de construction de la courbe de Zipf**

- ✓ Nous rangeons dans un tableau tous les motifs codés rencontrés dans l'image, en associant à chaque motif sa fréquence d'apparition. Notons que le tableau des motifs est initialement vide.
- ✓ Pour chaque nouveau motif rencontré, nous comparons successivement le motif courant à chacun des motifs déjà présents dans le tableau.
- ✓ Si le motif courant est déjà présent dans le tableau, nous incrémentons sa fréquence d'apparition d'une unité. Si le motif n'est pas dans le tableau, nous l'ajoutons en fin de tableau en mettant sa fréquence d'apparition à 1.
- ✓ Lorsque l'image a été entièrement balayée, nous trions le tableau par ordre décroissant de la fréquence d'apparition des motifs. Ensuite, nous traçons la courbe donnant la fréquence d'apparition en fonction du rang dans un repère bi-logarithmique (Caron 2004).

La figure II.5 démontre la courbe de Zipf de l'image Cameraman.

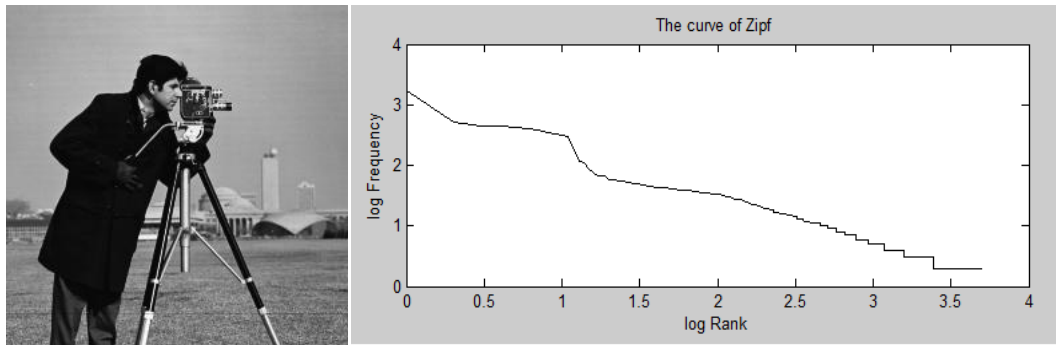


Figure II.5 Courbe de Zipf de l'image Cameraman.

□ L'algorithme de construction de la courbe de Zipf inverse

- ✓ Balayer l'image avec le masque de capture et compter les motifs.
- ✓ Calculer les motifs différents ayant la même fréquence d'apparition de chaque valeur trouvée de la fréquence. Pour cela, nous initialisons la fréquence cherchée à 1 et nous parcourons séquentiellement le tableau des motifs afin de compter les motifs ayant la même fréquence que la fréquence courante.
- ✓ L'algorithme est réitéré en incrémentant à chaque fois la fréquence d'une unité jusqu'à ce que l'on ait atteint la fréquence maximale.
- ✓ Nous traçons la courbe donnant la distribution du nombre de motifs en fonction de leur fréquence d'apparition dans un repère bi-logarithmique (Caron 2004).

La figure II.6 démontre la courbe de Zipf inverse de l'image Cameraman.

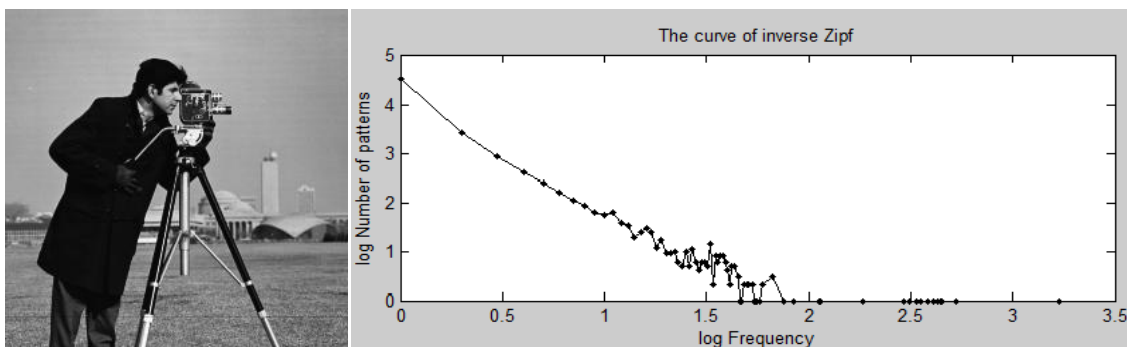


Figure II.6 Courbe de Zipf inverse de l'image Cameraman

II .5. Conclusion

Les lois de Zipf et de Zipf inverse ont été largement employées dans la linguistique, l'économie, la démographie, la biologie, la musique, l'Internet ou l'étude des signaux sonores médicaux.

Nous nous intéressons en particulier à l'application des lois de Zipf et de Zipf inverse au traitement d'image numérique. Cependant, malgré leur adéquation à l'analyse de la structure complexe que présente l'image numérique, les lois de Zipf et de Zipf inverse n'avaient pas fait l'objet de grandes applications dans ce domaine.

Nous contribuons ces lois comme un outil d'aide au diagnostic médical du cancer du sein en établissant une analyse numérique des mammographies pour offrir un second diagnostic aux radiologues. En effet, les prochains chapitres de ce mémoire seront dans ce sens.

Chapitre III

Aide au diagnostic médical du cancer du sein assisté par Ordinateur

III.1. Introduction

Le cancer du sein est le cancer le plus courant chez la femme dans le monde et l'un des principales causes de mortalité féminine. A cet effet, la mammographie est une méthode utilisée par les radiologistes pour le dépistage du cancer du sein en détectant et interprétant les tumeurs dans des images mammaires (Bikesh et al 2011) .

Le taux de mortalité a diminué pendant ces dernières années, due en partie à l'utilisation de la mammographie, en effet, des campagnes de dépistage ont été lancées dans la plupart des pays du monde. De plus, l'un des domaines de recherches les plus vives dans la vision par ordinateur est la contribution des systèmes de diagnostic médical assisté par ordinateur du cancer pour offrir un deuxième avis aux radiologues durant l'élaboration de leur diagnostic.

Dans ce chapitre nous proposons une approche de caractérisation de la texture des images mammaires basée fusion des lois de puissance : Zipf, Zipf inverse et la transformée de Fourier discrète pour la classification des zones d'intérêt extraites à partir des mammographies.

III.2. Le cancer du sein

Un cancer désigne la présence de cellules anormales qui se multiplient de façon incontrôlée. Dans le cas du cancer du sein, les cellules peuvent rester dans le sein ou se répandre dans le corps par les vaisseaux sanguins ou lymphatiques. La plupart du temps, la progression d'un cancer du sein prend plusieurs mois et même quelques années (Passeport sante .2012).

Selon la localisation des cellules à l'origine du cancer, lobule ou canal, on parle de cancer lobulaire ou de cancer canalaire. Le cancer se produit localement, formant une masse cellulaire primitive (néoplasme), parfois appelée (tumeur maligne primaire), qui, avec le temps, envahit les structures avoisinantes saines (infiltration ou invasion). En empruntant le réseau lymphatique les cellules malignes peuvent atteindre des organes très éloignés.

Le nombre de personnes atteintes a progressé légèrement, mais régulièrement, au cours des trois dernières décennies. Par contre, le taux de mortalité a continuellement diminué au cours de la même période, grâce aux progrès réalisés en matière de dépistage, de diagnostic et de traitement.

Durant le diagnostic du cancer du sein, nous distinguons l'étendue de la maladie. En effet, le médecin a besoin de déterminer le stade du cancer afin de prescrire un traitement. Pour la détermination de la stadification du cancer du sein, les médecins se basent sur trois critères : la taille et l'infiltration de la tumeur, l'atteinte ou non des ganglions lymphatiques et la présence ou l'absence de métastases (Merouani 1999). Deux types de tumeur existent :

- **Les tumeurs bénignes** : les tumeurs bénignes ont des contours bien délimités. Elles ont une croissance lente et restent localisées dans le tissu ou l'organe dans lequel elles sont apparues. Elles n'entraînent pas de métastases dans d'autres endroits du corps. Les tumeurs bénignes sont composées de cellules qui ressemblent aux cellules normales du tissu concerné.
- **Les tumeurs malignes**: les tumeurs malignes ont des contours généralement mal délimités. Certaines sont toutefois bien limitées et peuvent alors être confondues avec les tumeurs bénignes ; ceci pouvant retarder le diagnostic du cancer. En effet les cellules cancéreuses changent de forme, taille et contours. Les tumeurs malignes envahissent les tissus voisins et peuvent entraîner des métastases en s'échappant de la tumeur primitive et colonisant une autre région du corps (Cancer du sein 2016).

III.3. Le dépistage du cancer du sein

III.3.1. La mammographie

La mammographie est une technique de radiographie, son objectif consiste à découvrir au plus tôt des anomalies avant qu'elles ne provoquent des symptômes cliniques. La mammographie est non seulement pratiquée dans les campagnes de dépistage du cancer du sein, mais aussi pour le diagnostic et la localisation lors d'interventions chirurgicales. Le point fort d'un tel examen est qu'il permet d'examiner la totalité du tissu mammaire avec une ou deux incidences seulement. L'appareil dédié à la réalisation d'une mammographie est le mammographe comme mentionné sur la figure III.1. Cet appareil se compose d'un tube radiogène générateur de rayons X de faible énergie et d'un système de compression du sein (Cheikhrouhou 2012) .



Figure III.1 Appareil de génération des mammographies numériques
(source : mammographies numériques)

La radiographie est réalisée sur des films argentiques ou sur des systèmes de radiologie digitale de haute qualité. L'analyse de la glande mammaire est réalisée grâce aux différences de l'atténuation des différents types de tissu.

Sur la figure III.2, nous présentons une mammographie de tissu sein ne portant aucune anomalie.



Figure III.2 Mammographie normale

Sur la figure III.3, nous présentons une mammographie portant une tumeur bénigne.

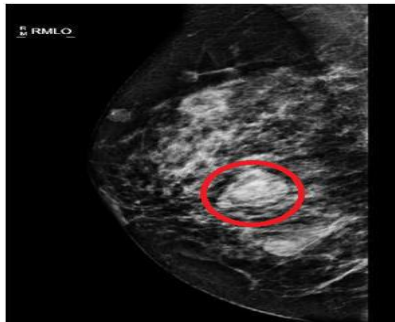


Figure III.3 Mammographie portant une tumeur bénigne

Tandis que sur la figure III.4, nous présentons une mammographie portant une tumeur maligne.



Figure III.4 Mammographie portant une tumeur maligne

III.4. Détection et classification assistées par ordinateur (CAdE/CAdx) des tumeurs dans la mammographie

L'étude des différentes pathologies mammaires et leur classification présente une grande difficulté pour les radiologues, en effet, nous distinguons la complexité de l'étape de détection des lésions et la complexité de l'étape de décision. Pour cela, une grande évolution dans l'analyse automatisée d'image médicale a eu lieu tentant d'améliorer le stade de l'interprétation de l'imagerie médicale ; ceci par l'application de la technologie informatique dans cette tâche

en fournissant une « seconde opinion » au la radiologue. L'utilisation du diagnostic assisté par ordinateur dans de nombreux programmes de dépistage du cancer du sein semble une bonne solution (Hamoud 2015).

Il ya deux grands types de systèmes pour le diagnostic assisté par ordinateur élaborés par plusieurs chercheurs: nous distinguons le (CAdE : Computer Aided Detection) pour la détection des tumeurs assistée par ordinateur et le (CAdx : Computer Aided Diagnosis) pour le diagnostic ou la classification des tumeurs assistés par ordinateur. La figure III.5 est un résumé schématisé pour éclaircir la différence entre ces deux systèmes (Djaroudib 2016).

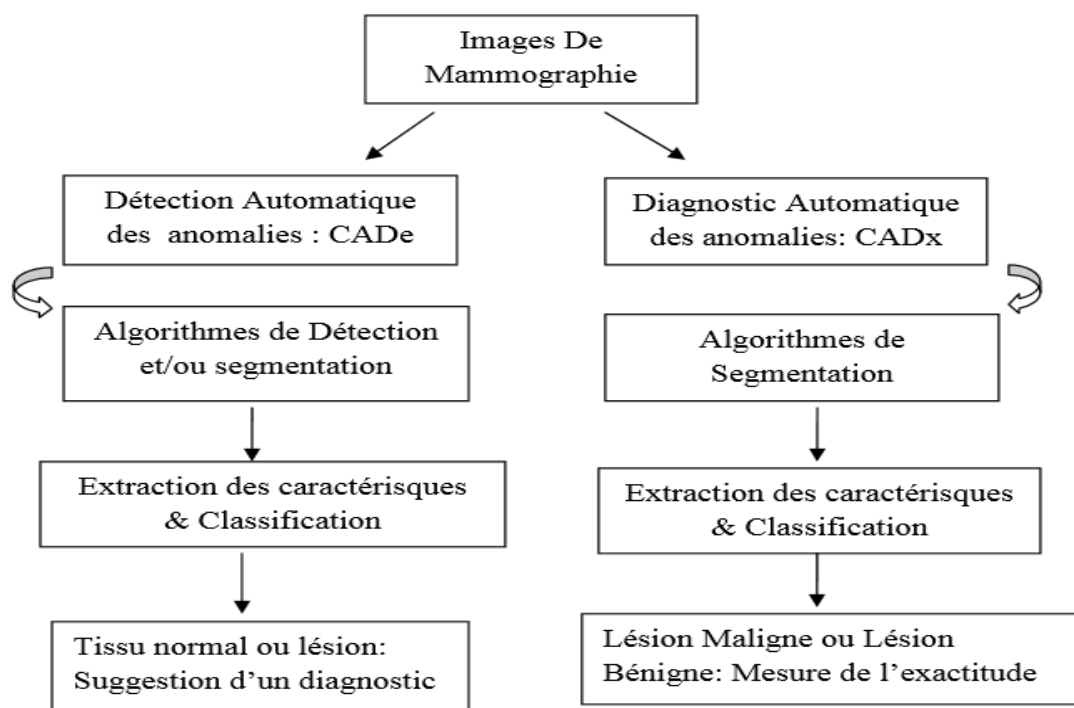


Figure III.5 Différence entre le système CAdE et le système CAdx

III.4.1. Détection des tumeurs assistée par ordinateur (CAdE)

Les systèmes d'aide à la détection automatisée des anomalies dans des images mammaires permettent d'aider les radiologues à détecter et à localiser des régions d'intérêt dans des images de dépistage du cancer du sein pour le diagnostic assisté par ordinateur (CAdx). Ils se basent sur des descripteurs mathématiques pour mesurer les changements de la texture, en effet, les pixels à l'intérieur d'une tumeur présentent des descripteurs différents par rapport aux autres pixels du tissu parenchyme mammaire. Les descripteurs peuvent consister en des valeurs de niveaux de gris, des mesures de texture ou bien des mesures morphologiques (Hamoud 2015) (Djaroudib 2016).

III.4.2. Diagnostic des tumeurs assisté par ordinateur (CADx)

Les systèmes d'aide au diagnostic automatisé des anomalies permettent d'aider les radiologues à caractériser la lésion et donner un diagnostic final en réduisant ainsi le nombre de recommandations de biopsies sur des lésions bénignes. Le principe consiste en donner le résultat de la classification des tumeurs : malignes / bénignes selon deux étapes ; la segmentation de l'anomalie puis sa classification en masse bénigne ou en masse maligne, notons que le résultat est donné à travers des sorties de classifieurs (Djaroudib 2016).

III.5. Indexation et recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur

Suite aux difficultés de la recherche manuelle d'image mammaires basée sur le texte qui effectue la recherche des mammographies par l'annotation textuelle manuelle, donc, posant plusieurs limitations telles que la subjectivité de l'annotation textuelle et sa consommation d'un temps considérable et utilisant un équipement performant pour le diagnostic des patients dans les hôpitaux et les centres de dépistage du cancer du sein. L'indexation et la recherche des mammographies par le contenu (CBMIIR : Content Based Mammogram Image Indexing and Retrieval) semblait fournir une solution automatisée assistée par ordinateur ayant un but ultime d'accéder efficacement aux mammographies archivées souhaitées, pathologiquement similaires, à la mammographie en cours de diagnostic. Ainsi ces systèmes se basent sur le même principe des systèmes d'indexation et de recherche des images par le contenu (CBIR). En effet, la recherche est effectuée en fonction du contenu visuel des images. Notons que durant ces dernières années, les approches de recherche d'images basées sur le contenu (CBIR) a reçu une attention particulière pour l'analyse d'images médicales (Hamoud 2015).

III.6. L'apport d'un système d'indexation et de recherche de mammographies par le contenu (CBMIIR) par rapport à un système d'aide au diagnostic assisté par ordinateur (CADx) se basant sur la sortie d'un classifieur

Vu les développements surprenants dans le domaine du traitement d'image numérique et plus précisément l'image médicale, d'autre part, la grande quantité d'images mammaires, par exemple, produites par jours dans les hôpitaux nécessite le stockage de ces ressources dans

des bases de données, évoluant et devenant des référentiels accessibles. Donc l'indexation représente un remède adéquat pour les futures explorations des mammographies.

Dans ce sens, les systèmes d'indexation et recherche des images mammaires par le contenu (CBMIIR : Content Based Mammogram Image Indexing and Retrieval) ont été développés ayant une structure similaire à celle des systèmes d'aide au diagnostic médical se basant sur la sortie d'un classifieur (CADx) et se différenciant sur la présentation du résultat obtenu.

Dans (Chang et al. 2013), les auteurs présentent les éléments clés d'un système d'indexation et de recherche d'images médicales par le contenu, à savoir : l'acquisition de l'image médicale requête grâce à l'interface du système, le traitement, l'extraction des descripteurs de l'image médicale, la mesure de similarité avec les images médicales diagnostiquées et indexées dans la base, l'indexation du nouveau cas traité et l'affichage des résultats de recherche.

Un système d'aide au diagnostic assisté par ordinateur (CADx) est basé sur un résultat numérique donné par un classifieur dont les radiologues ne sont pas familiarisés (Hamoud 2015). En effet, l'ultime but des systèmes CBMIIR est de renforcer la performance du diagnostic assisté par ordinateur (CADx) où les radiologues reçoivent l'aide à travers des symboles abstraits donnant les classes des mammographies où appartiennent les zones d'intérêt requêtes. Pour palier à l'ambiguïté d'un tel résultat, les radiologues ont affirmé qu'ils sont plus confiants dans leur diagnostic lorsque des cas déjà diagnostiqués et similaires, à celui en cours d'analyse, sont récupérés et affichés ainsi que le résultat de la classe à qui appartient le cas en cours d'analyse. Le raisonnement à base de cas est tiré par le principe que les radiologues répondent mieux aux mammographies similaires au cas en cours d'analyse, que des chiffres abstraits (Cheikhrouhou 2012).

III.7. Approche proposée basée fusion des lois de puissance : Zipf, Zipf inverse et la transformée de Fourier discrète pour la classification des zones d'intérêt des mammographies

III.7.1. Analyse et caractérisation de la texture des zones d'intérêt par les lois de Zipf et de Zipf inverse

Les lois de Zipf et de Zipf inverse font la caractérisation de la texture des zones d'intérêt codées par le codage des rangs généraux, nous allons analyser la représentation graphique des deux lois de Zipf et de Zipf inverse pour une zone d'intérêt du tissu sein, une zone d'intérêt

portant une tumeur bénigne et une autre portant une tumeur maligne comme mentionné sur la figure III.6(Caron 2004).

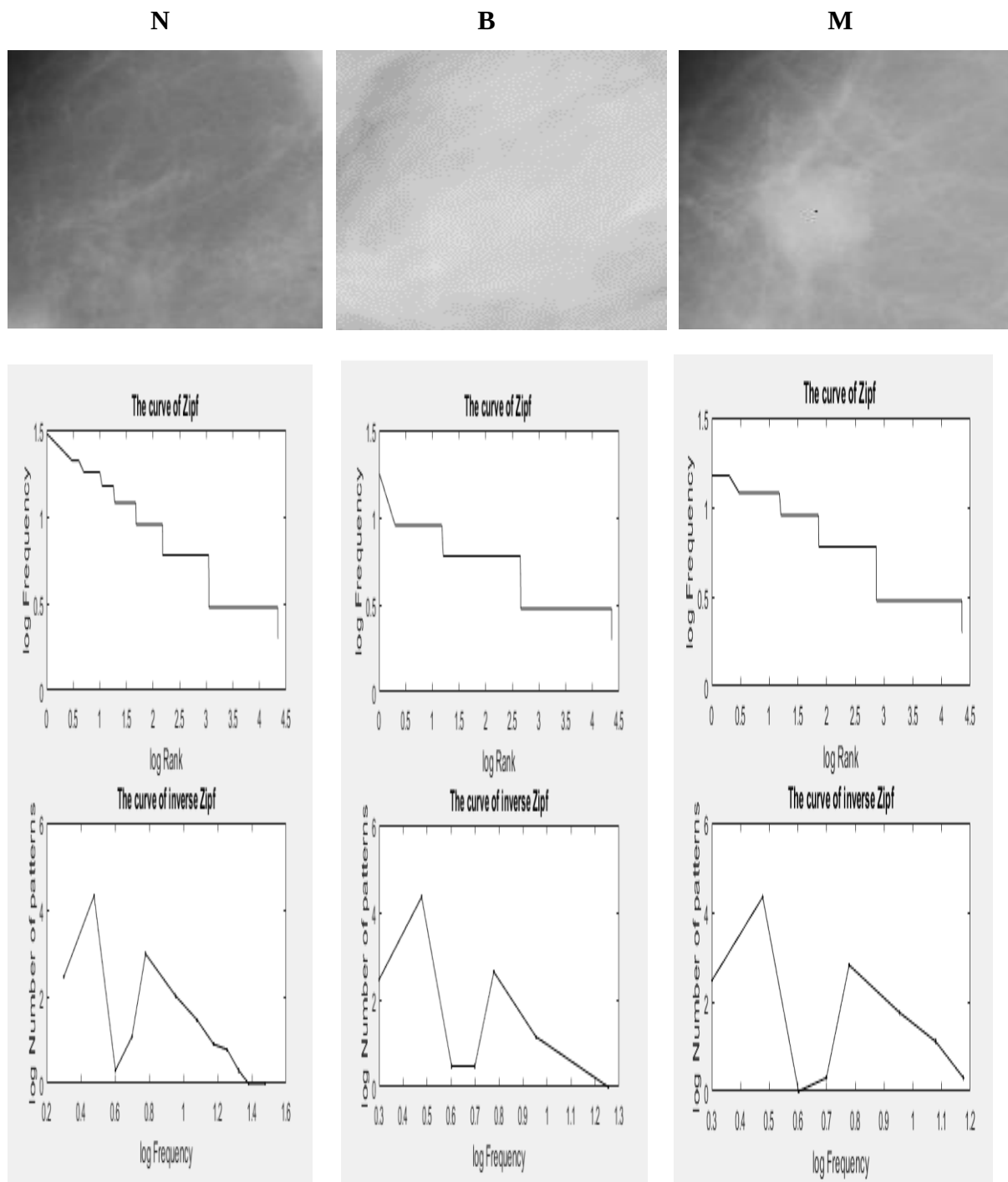


Figure III.6 Courbes de Zipf et de Zipf inverse de trois zones d'intérêts encodées par le codage des rangs généraux : **N** zone d'intérêt sans tumeur, **B** zone d'intérêt portant une tumeur bénigne, et **M** zone d'intérêt portant une tumeur maligne.

Nous pouvons attester que les courbes de Zipf et de Zipf inverse relatives à une zone d'intérêt sans tumeur, une zone d'intérêt portant une tumeur bénigne et une zone d'intérêt portant une

tumeur maligne sont différentes. Effectivement, nous distinguons une ordonnée à l'origine très élevée sur la courbe de Zipf de la zone d'intérêt présentant un tissu normal du sein du fait que la texture est homogène et donc un motif homogène se verra répéter plusieurs fois. D'autre part, la zone d'intérêt portant une tumeur bénigne possède une ordonnée à l'origine plus basse. Si nous analysons à présent la zone d'intérêt portant une tumeur maligne, nous pouvons attester qu'elle possède une ordonnée à l'origine nettement plus basse que celles précédemment citées et ceci sur la courbe de Zipf, en effet, la structure hétérogène de la texture de la tumeur maligne ne génère pas une sur-représentation d'un seul motif homogène mais plusieurs motifs hétérogènes faisant en sorte que l'ordonnée à l'origine soit basse.

Comme indiqué sur la figure, la différenciation des courbes de Zipf inverse est manifestée aussi pour les zones d'intérêts présentant un tissu normal, les zones d'intérêts présentant une tumeur bénigne et les zones d'intérêts présentant une tumeur maligne. Donc, l'analyse des zones d'intérêts avec les lois de Zipf et de Zipf inverse en les encodant par le codage des rangs généraux permet la distinction entre les tumeurs malignes et les tumeurs bénignes ainsi que le tissu normal, cela en utilisant des descripteurs calculés à partir des courbes de Zipf et de Zipf inverse de chaque zone d'intérêt consistants en (Hamoud 2015) :

- **L'aire délimitée par la courbe de Zipf :**

Nous calculons l'aire délimitée par la courbe de Zipf à partir des courbes de Zipf obtenues suite au codage de l'image par le codage des rangs généraux. Soit n le nombre de motifs de la courbe, f_i la fréquence et r_i le rang du motif i , l'aire de la courbe est donnée par la formule III.1 :

$$A = \sum_{i=1}^{n-1} \frac{(f_i + f_{i+1})(r_{i+1} - r_i)}{2} \quad (\text{III.1})$$

- **Les pentes des courbes de Zipf et de Zipf inverse :**

La pente moyenne d'une courbe est le coefficient directeur de la droite des moindres carrés.

Elle est donnée par la formule III.2 :

$$p = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (\text{III.2})$$

- **L'entropie de la courbe de Zipf**

Il est possible de définir une mesure d'entropie à partir des fréquences d'apparition des motifs de l'image. Deux formulations de l'entropie sont possibles, l'entropie relative aux motifs et l'entropie relative aux fréquences d'apparition.

- a) L'entropie relative aux motifs**

L'entropie relative aux motifs est définie par la formule III.3 pour un ensemble de R motifs distincts :

$$H_w = - \sum_{r=1}^R \frac{f(r)}{T} \log_R \frac{f(r)}{T} \quad (\text{III.3})$$

Dans cette formule, $f(r)$ représente la fréquence du motif de rang r , T représente le nombre total de motifs différents ou non, et on utilise un logarithme de base R.

- b) L'entropie relative à la fréquence**

L'entropie relative à la fréquence est définie par la formule:

$$H_f = - \sum_{f=1}^F \frac{I(f)}{R} \log_F \frac{I(f)}{R} \quad (\text{III.4})$$

$I(f)$ représente le nombre de motifs distincts de fréquence f et F le nombre total d'occurrences des motifs dans l'image. On utilise un logarithme à base F afin que l'entropie soit comprise entre 0 et 1.

- **La constante alpha**

Pour tout motif appartenant à l'image, la fréquence d'apparition de ce motif * son rang dans la liste ordonnée décroissante des fréquences des motifs de l'image représente une constante.

- **Les ordonnées à l'origine des courbes de Zipf et de Zipf Inverse**

Nous extrayons l'ordonnée à l'origine à partir des courbes de Zipf et de Zipf inverse.

III.7.2. Analyse et caractérisation de la texture des zones d'intérêt par la transformée de Fourier discrète

Les images peuvent être représentées soit dans le domaine spatial, soit dans le domaine fréquentiel. En effet, la représentation spatiale d'une image numérique correspond au plan des valeurs de chaque pixel représenté par une matrice rectangulaire, Tandis que La représentation fréquentiel analyse les fréquences de l'image. La transformée de Fourier a été parmi les premières techniques utilisées pour l'analyse de la texture dans le domaine fréquentiel.

La transformée de Fourier **F** permet de convertir une image **I** de dimension **N** × **M** depuis son domaine spatial vers son domaine fréquentiel.

La transformée de Fourier de l'image $f(x,y)$, de largeur **N** et de hauteur **M** est donnée par (Barré et al 1999):

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-2\pi i(\frac{ux}{M} + \frac{vy}{N})} \quad (\text{III.5})$$

Où $f(x, y)$ est l'intensité du pixel à la position (x, y) , $0 \leq x \leq M-1$, $0 \leq y \leq N-1$.

(x, y) sont les coordonnées du domaine spatial, (u, v) sont les coordonnées du domaine fréquentiel dans les directions x et y .

Les coefficients $F(u, v)$ dit coefficients de Fourier, fournissent des informations sur les fréquences (u,v) et permettent des manipulations dans le domaine fréquentiel.

Nous calculons la TF en chaque pixel de l'image pour obtenir une image complexe. À une image sinusoïdale 2D correspond un point dans l'espace des fréquences. Pour visualiser cette image, nous représentons sa norme (spectre d'amplitude) et sa phase.

Les valeurs de l'amplitude en général présentent une très grande dynamique (les basses fréquences sont plus importantes que les hautes fréquences), le spectre d'amplitude représente le logarithme du module de la transformée de Fourier.

La fréquence dans une image représente la variation de l'intensité des pixels de l'image.

Les textures présentent une certaine périodicité, cette information se retrouve réorganisée et réduite dans la représentation fréquentielle. Les attributs calculés à partir des spectres filtrés sont ensuite utilisés pour classifier la texture de la zone d'intérêt.

La transformée de Fourier se prête particulièrement bien à la détection et à la caractérisation de parcelles viticoles grâce à leur contenu spatial périodique. Ainsi elle analyse la localisation des pics correspondant à des parcelles viticoles dans l'espace de Fourier (Regniers 2014) (Barré et al 1999) (Chaman et al 2001) .

Nous nous intéressons à la transformée de Fourier discrète durant notre travail, Le spectre de Fourier présente des caractéristiques communes telles que la présence de droites, d'une épaisseur plus ou moins importantes qui passent par l'origine.

- Le spectre informe sur la présence de certaines fréquences (et de leur importance) dans l'image.
- Plus une droite est longue, plus elle porte des fréquences élevées.
- La direction des droites indique les lignes de force de l'image d'origine, qui lui sont perpendiculaires.
- Si une droite est constituée de points ou tries alignés, cela indique une certaine périodicité de l'image dans la direction perpendiculaire. (Barré et al 1999)

Nous devons mentionner que si la texture possède une certaine structure, il en ira de même pour le module de la transformée de Fourier : si la texture est moins ordonnée, voire complètement aléatoire, le spectre du module de la transformée de Fourier de cette texture sera également assez aléatoire et nous ne saurons rien en tirer. Néanmoins, le spectre de la TF présente généralement certains pics. Ces pics nous informent sur l'orientation de la structure ainsi que sur son amplitude. A savoir :

- Le pic dominant dans le spectre donne la direction principale de la texture.
- La localisation des pics permet de déterminer la période spatiale fondamentale de la texture.
- Chaque pic représente une périodicité spatiale dans un sens particulier.

Nous présentons sur la figure III.7, le spectre de Fourier généré à partir d'une zone d'intérêt portant une tumeur bénigne.



Figure III.7 Spectre de Fourier (à droite) généré à partir d'une zone d'intérêt (à gauche) portant une tumeur bénigne.

III.7.3. Bilan

Le but de notre travail consiste en la caractérisation de la texture des régions d'intérêt par une fusion de deux méthodes : les lois puissance Zipf et Zipf inverse ainsi que la transformée de Fourier Discrète pour bénéficier de l'apport complémentaire de ces deux méthodes d'analyse de la texture dans un processus d'aide au diagnostic médical du cancer du sein assisté par ordinateur.

III.8. Conclusion

Une bonne description ou caractérisation de la texture des masses mammaires amènera à une bonne classification ou diagnostic.

Nous avons présenté dans ce chapitre l'approche proposée pour la caractérisation de la texture par l'extraction de descripteurs discriminants pour le processus de classification.

Le chapitre suivant évoquera en détail les étapes de développement de notre système d'aide au diagnostic médical du cancer du sein assisté par ordinateur.

Chapitre IV

Conception et Implémentation

IV.1. Introduction

Les recherches dans le domaine médical à travers la vision par ordinateur ont connu un succès durant la dernière décennie, en raison de l'application du traitement d'image et la vision par ordinateur dans de nombreuses pathologies et plus précisément pour la détection du cancer.

Dans ce chapitre nous allons présenter une vue plus détaillée sur les techniques utilisées pour le développement d'un système classifiant les tumeurs du sein que se basant sur la technique du template-matching ainsi que la caractérisation de la texture des images mammaires par une fusion des lois de puissance : Zipf, Zipf inverse et de la transformée de Fourier discrète.

IV.2. Problématique

Notre problématique de recherche vise à proposer une technique judicieuse d'analyse de la texture d'images mammaires. Généralement, les relations les plus simples à appliquer sont celles linéaires comme les transformées intégrales linéaires. Cependant, ces dernières montrent vite leur limites durant la modélisation d'une image vu sa structure assez complexe.

Dans (Hamoud 2015) (Hamoud et al 2011) , les auteurs ont suggéré les lois de puissance Zipf et Zipf inverse pour la caractérisation de la texture des images mammaires dans une tâche d'aide au diagnostic médical du cancer du sein assisté par ordinateur tout en aboutissant un taux élevés de classification des tumeurs mammaires. La motivation de l'analyse de la texture au sein des mammographies par ces lois provient du fait que les descripteurs texturaux sont de grande importance pour le diagnostic, cependant, difficiles à quantifier par les radiologues.

Le travail proposé dans ce mémoire réside dans l'exploration des perspectives de ces travaux en proposant une fusion des lois de Zipf et de Zipf inverse avec la méthode de la transformée de Fourier discrète pour bénéficier de l'apport complémentaire de ces deux approches de caractérisation de la texture dans le but d'essayer d'améliorer les résultats qui ont été obtenu par l'utilisation des lois de Zipf et de Zipf inverse seulement.

L'imagerie médicale est, certainement, l'un des axes de la médecine qui a le plus progressé. Cette évolution pousse à une demande accrue pour une indexation et recherche des images médicales pour la gestion des données. En effet, les radiologues ont manifesté leur insatisfaction du résultat du diagnostic généré des systèmes d'aide au diagnostic médical du cancer du sein se basant sur un classifieur. A cet effet, nous visons à renforcer les systèmes (CADx) par la nouvelle tendance des systèmes d'indexation et de recherche des mammographies par le contenu (CBMIIR), à qui les radiologues semblent satisfaits vu que le diagnostic sera effectué à base de cas récupérés et affichés à partir de bases de données indexées de cas préalablement diagnostiqués.

Donc nous allons proposer un système d'indexation et de recherche des images mammaires se basant sur la fusion des lois de Zipf et Zipf inverse avec la transformée de Fourier discrète pour la caractérisation texturale des zones d'intérêt extraites des mammographies. Mentionnons que le diagnostic assisté par ordinateur s'effectue par la technique du template-matching.

IV.3. Principe de la classification basée template-matching et les K plus proches voisins

Template Matching pourrait être une méthodologie dans le traitement d'image numérique pour identifier les petits composants d'une figure qui correspondent à une image modèle (Swaroop al 2016). Notre méthodologie pour la classification des zones d'intérêts segmentées au sein des mammographies à des zones d'intérêts portant des tumeurs bénignes, des tumeurs malignes ou des zones d'intérêts normales sans lésions se base sur l'approche template-matching consistant à trouver de petites parties d'une image qui correspondent à une image de modèle. Elle peut être utilisée dans la fabrication dans le cadre du contrôle qualité, l'imagerie médicale, comme un moyen de détecter les bords dans les images (Girod 2013).

Nous utilisons une technique de recherche pour trouver dans la base de données, les zones d'intérêts qui sont pathologiquement similaires à chaque zone d'intérêt requête. Ensuite, nous assignons à la zone d'intérêt requête, sa classe pathologique, selon le résultat donné par l'algorithme des K plus proches voisins (KPPV). En effet, la classe pathologique majoritairement renvoyé par cet algorithme sera la classe pathologique de la zone d'intérêt en cours de diagnostic.

Notre système proposé contient deux phases comme mentionné sur la figure IV.1 où nous exposons l'architecture de notre système d'indexation et de recherche des zones d'intérêt les plus similaires à une zone d'intérêt requête :

- **La phase hors ligne :**

Pour l'analyse des régions d'intérêt par les lois de Zipf et de Zipf inverse dans le but de tracer les courbes de Zipf et Zipf inverse et extraire le vecteur de descripteurs de texture qui contient 8 descripteurs texturaux générés par ces lois puissance pour chaque zone d'intérêt dans la base de données, ensuite, le calcul de la transformée de Fourier discrète à partir de chaque région d'intérêt pour l'extraction des 8 descripteurs texturaux consistant en les coefficients de sortie de la transformée de Fourier discrète. Ensuite, nous diminuons le volume du vecteur descripteurs par la méthode d'analyse en composantes principales (ACP) pour générer 10 descripteurs les plus discriminants.

- **la phase en ligne :**

Dans cette phase, le radiologue intègre une zone d'intérêt requête segmentée à partir d'une mammographie, au système pour identifier sa pathologie. En effet, pour chaque nouvelle zone d'intérêt segmentée à partir d'une mammographie, un vecteur de descripteurs texturaux est calculé suite à son analyse par les lois de Zipf et de Zipf inverse ainsi que l'application de la transformée de fourrier discrète sur cette région d'intérêt. Par la suite nous appliquons l'analyse en composante

principale (ACP) sur le vecteur de descripteurs de la région d'intérêt requête pour générer les nouveaux descripteurs les plus discriminants.

Les cas récupérés par le système suite à la recherche des K plus proches voisins de la zone d'intérêt requête s'utilisent comme prédicteurs utiles de la classe pathologique relative à la requête courante. Les similarités entre le vecteur descripteur de la zone d'intérêt requête et les vecteurs descripteurs de chaque zone d'intérêt indexée dans la base de données sont calculés par la distance Euclidienne intégrée dans l'algorithme des K plus proches voisins.

Le KPPV utilise les étiquettes des classes des voisins similaires pour prédire la classe de la région d'intérêt requête. En effet, il classe les cas voisins selon une mesure de similarité basée sur la distance Euclidienne et affecte la mammographie requête à la classe pathologique dont la plupart des K cas voisins utilisés pour l'apprentissage proviennent.

Notre choix a porté sur l'utilisation de cet algorithme de classification supervisée vu sa simplicité par rapport aux autres algorithmes de classification supervisée tel que les machines à vecteurs de support (SVM) et les réseaux de neurones, tout en aboutissant à de très bons taux de classification.

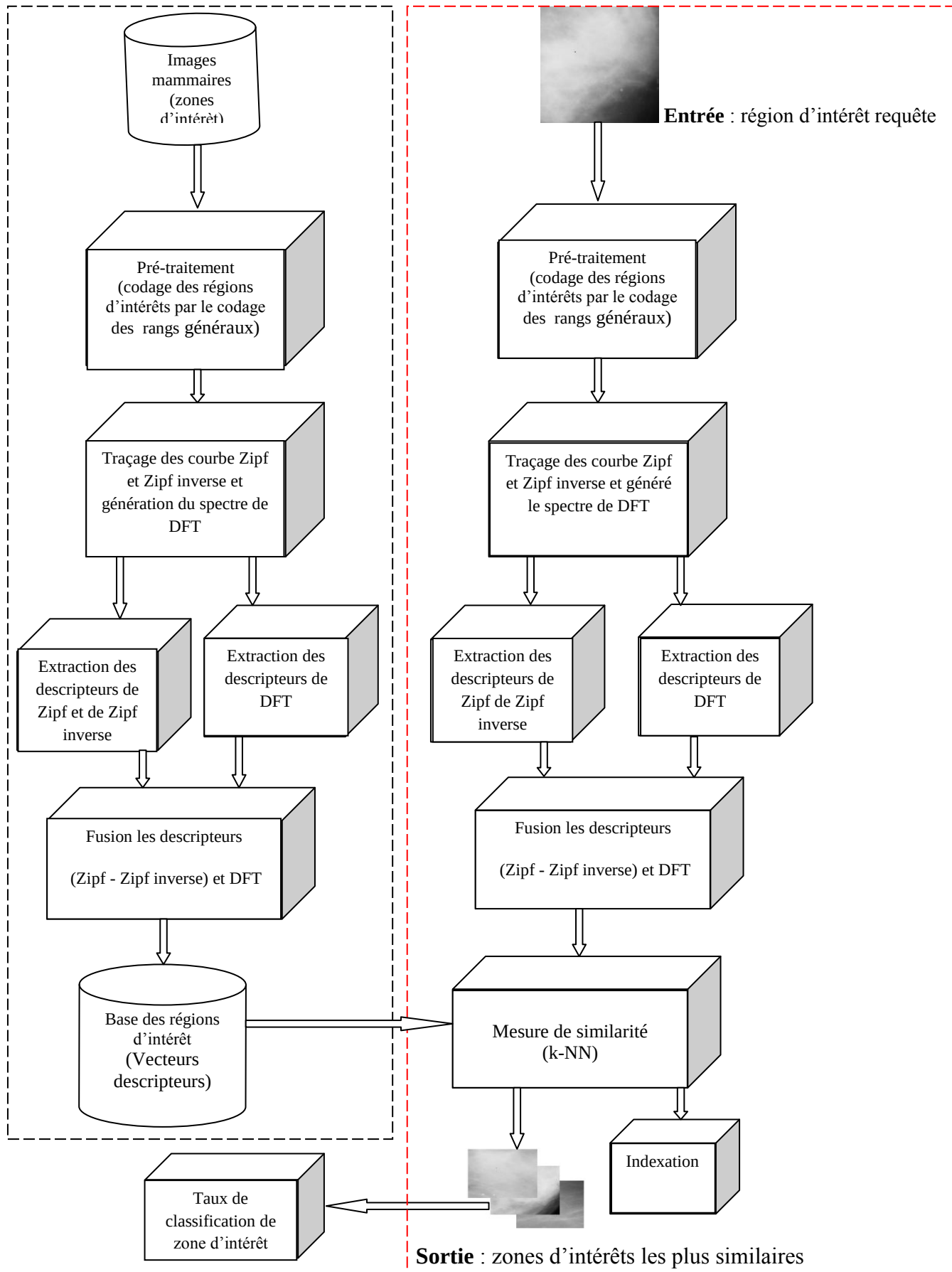


Figure IV.1 L'architecture du système mammo-ZipDFT

Sur la figure IV.2, nous présentons une légende relative à la figure IV.1.

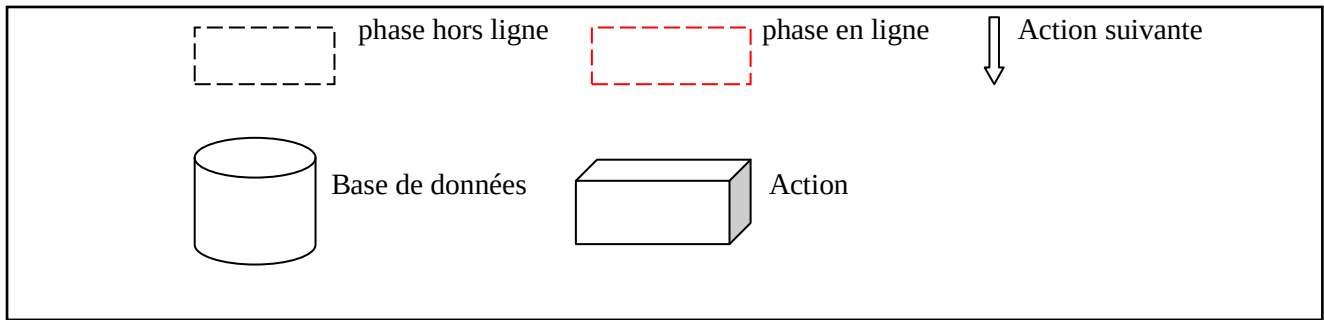


Figure IV.2 Légende relative à la figure IV.1.

IV.3.1. Algorithme des K-plus proches voisins

L'algorithme des k-plus-proches-voisins est l'un des algorithmes les plus simples d'apprentissage automatique supervisé. En supposant qu'une base d'apprentissage correctement étiquetée soit à disposition, son principe est le suivant : étant donnée une base d'apprentissage d'images étiquetées correctement et un entier K (K est un facteur positif entier , généralement petit), le classifieur KPPV détermine la classe d'un nouvel objet en lui attribuant la classe majoritaire des K objets lui ressemblant le plus dans la base d'apprentissage. Afin de trouver les K plus proches d'une donnée à classer, on peut choisir la distance euclidienne. Soient deux données représentées par deux vecteurs x_i et x_j , la distance entre ces deux données est donnée par (Chamroukhi 2013) :

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2} \quad (IV.1)$$

IV.3.2. L'analyse en composantes principales

L'Analyse en Composantes Principales (ACP) est une méthode appliquée dans le cadre de l'apprentissage non supervisé, ce dernier permet de regrouper des objets dans des clusters selon leur similarité dans le but d'explorer les liaisons entre variables et les ressemblances entre individus. Elle crée des nouvelles variables qui sont appelées "composantes principales", ou bien axes principaux et les corrélations entre plusieurs variables seront utilisées pour décrire et résumer l'information contenue dans ces variables. en effet L'Analyse en Composantes Principales (ACP) réduit le nombre de variables et de rendre l'information moins redondante (Miloudi 2014).

Le principe de l'ACP est constitué une séquence des étapes on posées comme suit :

1. Ranger les individus en ligne et les variables quantitatives en colonnes

$$H = \begin{bmatrix} var^1 & var^2 & var^3 \\ x_1^1 & x_1^2 & x_1^3 \\ x_2^1 & x_2^2 & x_2^3 \\ x_3^1 & x_3^2 & x_3^3 \\ x_4^1 & x_4^2 & x_4^3 \end{bmatrix} \quad (IV.2)$$

2. Calculer la moyenne de chaque variable (colonne) :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (IV.3)$$

3. Calculer l'écart type de chaque variable (colonne) :

$$Dev = \left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{1}{2}} \quad (IV.4)$$

4. Centrer les variables :

$$x_c = x - \bar{x} \quad (IV.5)$$

5. Réduire les variables :

$$Red = \frac{x_c}{Dev} \quad (IV.6)$$

6. Calculer les corrélations

$$Cor = Red^t \times Red \quad (IV.7)$$

Il faut transposer la matrice puis la multiplier par elle-même. C'est une matrice carrée dont la diagonale est l'autocorrélation de chaque variable et les autres sont les cross-corrélation.

7. Trouver les valeurs propres et les vecteurs propres correspondants de cette matrice (Matrice de corrélation).

8. Trier les valeurs propres obtenues en ordre décroissant de valeur $\lambda_1 > \lambda_2 > \lambda_3$.

9. Pourcentage d'inertie expliquée

Le pourcentage de variance expliquée par les axes factoriels retenus est obtenu par somme de leurs valeurs propres divisées par la trace. Cette valeur permet d'évaluer en quelque sorte la quantité d'information initiale recueillie par chaque axe.

Le pourcentage d'inertie expliquée par un axe factoriel

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \times 100 \quad (IV.8)$$

10. Détermination des composantes principales

Revient à multiplier la matrice réduite avec la matrice diagonale des vecteurs propres

$$\text{Proj} = \text{Red} * \text{VP} \quad (IV.9)$$

L'opérateur ' * ' est dans ce cas le produit scalaire $x \cdot C_{1,x} + y \cdot C_{1,y}$. Les axes choisis sont les axes principaux trouvés par l'ACP. Nous représentons donc les données sur ces axes. Les coefficients de la projection constituent les caractéristiques des données.

11. Coefficient de corrélation entre les composantes principales et les variables d'origine

Il est souvent utile d'étudier les corrélations entre les variables d'origine x et les nouvelles variables C .

Les corrélations entre les axes factoriels et les variables initiales indiquent la qualité de représentation de la variable sur l'axe. Les coordonnées des variables sont égales aux corrélations avec les axes. Plus une corrélation entre une variable et un axe est forte plus la variable est proche de l'axe. Il ne faut interpréter les axes qu'à partir des variables les mieux représentées, c'est-à-dire celles dont la corrélation est proche de 1 en valeur absolue.

Le coefficient de corrélation entre une composante principale C_j et une variable x_i est égale à :

$$r_{ij} = \frac{\text{covariance } C_j x_i}{\sqrt{\text{variance } C_j} \sqrt{\text{variance } x_i}} \quad \text{tq } r \in [-1, 1] \quad (IV.10)$$

Si r est proche à 1 alors C et x sont très bien corrélés ou liés.

Si r est proche à 0 alors C et x ne sont pas corrélés.

Si r est proche à -1 alors C et x sont très bien anti-corrélés. (Miloudi 2014).

IV.4. Matériel et outils utilisés pour le développement du système proposé :

IV.4.1. Le matériel utilisé

Le système que nous avons proposé a été développé et testé sous un ordinateur plutôt performant qui possède les caractéristiques suivantes :

Nom du modèle : TOSHIBA

Processeur : Intel Core i 3110 M

Fréquence du processeur : 2,40 GHZ

Mémoire vive : 4 Go DDR3

Système d'exploitation : Windows 7 Edition Premium 64-bits

IV.4.2. Les Outils utilisés

- **L'environnement MATLAB**

MATLAB est un langage de développement informatique particulièrement dédié aux applications scientifiques. Matlab est utilisé pour développer des solutions nécessitant une très grande puissance de calcul. Développé par la société [The MathWorks](#), MATLAB permet de manipuler des matrices, d'afficher des courbes et des données, de mettre en œuvre des algorithmes, de créer des interfaces utilisateurs, et peut s'interfacer avec d'autres langages comme le [C](#), [C++](#), [Java](#), et [Fortran](#). L'objet le plus commun dans Matlab est la matrice. C'est donc un tableau d'éléments d'un type donné, par exemple un tableau d'entiers, de réels, de texte, de variable logiques...etc (Loye 2009).

L'interface de Matlab change légèrement selon la version utilisée mais les points centraux resteront identiques comme mentionné sur la figure IV.3.

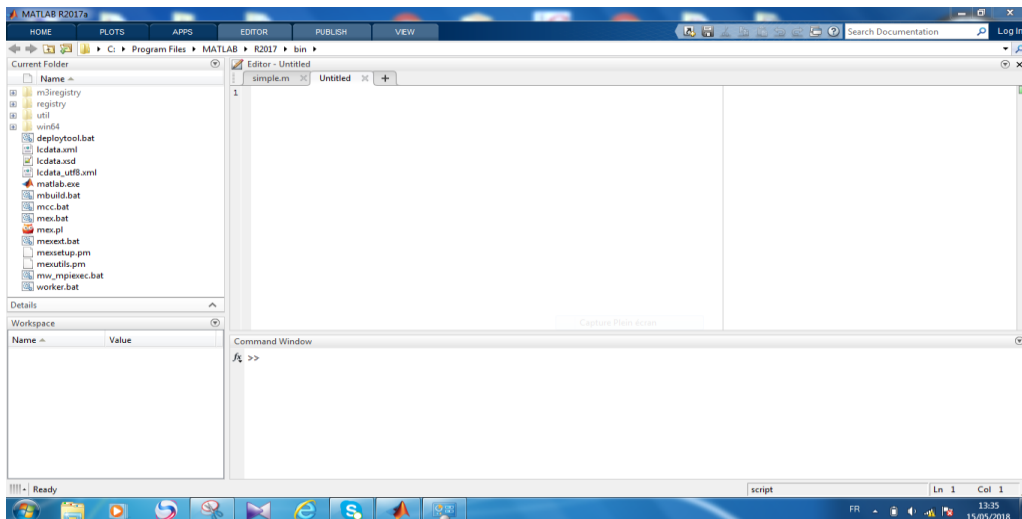


Figure IV.3 L'interface de Matlab R2017a

Le logiciel propose un véritable environnement de travail composé de multiples fenêtres. Nous pouvons distinguer cinq blocs :

Command Window

C'est le terminal dans lequel on doit taper les commandes et sur lequel on verra l'affichage des résultats. Une ligne commence toujours par `>>`

Workspace

Dans cette fenêtre, on obtient la liste des variables connues par Matlab . Il est possible de double-cliquer sur une variable pour l'afficher. Un clic-droit sur les variables offre de nombreuses options.

Command History

Dans cette fenêtre, conserver une trace de toutes les opérations qui ont été réalisées. On peut également remonter dans la liste de commandes en se plaçant dans la Command Window et en pressant les flèches de direction.

Current Folder

Permet de naviguer et de visualiser le contenu du répertoire courant de l'utilisateur. Les programmes de l'utilisateur doivent être situés dans ce répertoire pour être visible et donc exécutable.

Editor

La plupart de votre travail sous Matlab va consister à créer ou modifier des **fichiers .m** qui est le suffixe standard pour les procédures Matlab. Lorsque l'on réalise une tâche sous Matlab, il est très souvent possible de le faire en utilisant uniquement la Command Window.

- **Vecteurs et Matrices**

Vecteurs

Un vecteur sous Matlab est une collection de d'éléments du même type. Un vecteur pourra représenter des valeurs expérimentales ou bien les valeurs discrétisées d'une fonction continue. La méthode la plus simple pour définir un vecteur est de donner sa description explicite à l'aide de la commande `[]`, par exemple : `>> v = [1 2 3 4]` $v = 1 \ 2 \ 3 \ 4$

Matrice

Une matrice va se définir de façon similaire à un vecteur avec la commande `[]`. Une matrice est composée de m lignes et n colonnes. si on souhaite connaître la valeur de m ou n , on utilise la

commande `size(A)`, Ainsi pour définir une variable matricielle : $M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

On écrira : `>> M = [1 2 3 ; 4 5 6 ; 7 8 9];`

- **Types de variables**

Il existe cinq grands types de variables sous Matlab : les entiers, les réels, les complexes, les chaînes de caractères et le type logique. définissons une variable de chaque type :

```
>> a = 1.3; b = 3+i; c = 'bonjour';
```

```
>> d1 = true(1==1); d2 = logical(1);
```

```
>> e = int8(2);
```

a représente un réel, **b** un complexe, **c** une chaîne de caractères, **d1** et **d2** sont deux manières de définir une variable logique (VRAI dans le cas présent) et **e** est un entier codé sur 8 bits (Glorieux 2013).

- **Représentation graphique**

Nous aborderons pour finir la représentation graphique sous Matlab. En effet, pour présenter des données, on est très souvent amené à réaliser des graphes, pour cela Les bibliothèques de Matlab proposent un très grand nombre de fonctions pour la manipulation d'objets graphiques.

Matlab possède un grand choix de commandes pour créer toutes sortes de représentations graphiques : graphiques standards en 2D ou 3D avec axes linéaires, semi-log ou logarithmiques, histogrammes, représentation de coordonnées polaires, de courbes de niveau (contour surface), de mailles régulières (mesh , grid), parmi d'autres (Loye 2009).

IV.5. La base des mammographies utilisée

Nous allons utiliser la base MIAS (Mammograms Image Analysis Society) dans notre travail consistant en 120 mammographies de taille 1024×1024 pixels (Hamoud 2015).

Notons que la méthode de segmentation des zones d'intérêt a été présentée dans (Hamoud 2015).

IV.6. Fonctionnement du système développé

En cliquant sur le bouton RUN de Matlab, nous aurons l'affichage de l'interface principale de notre système. En cliquant sur le bouton « Parcourir », le radiologue pourra sélectionner la zone d'intérêt qu'il désire avoir sa classe pathologique comme mentionné sur la figure IV.4.

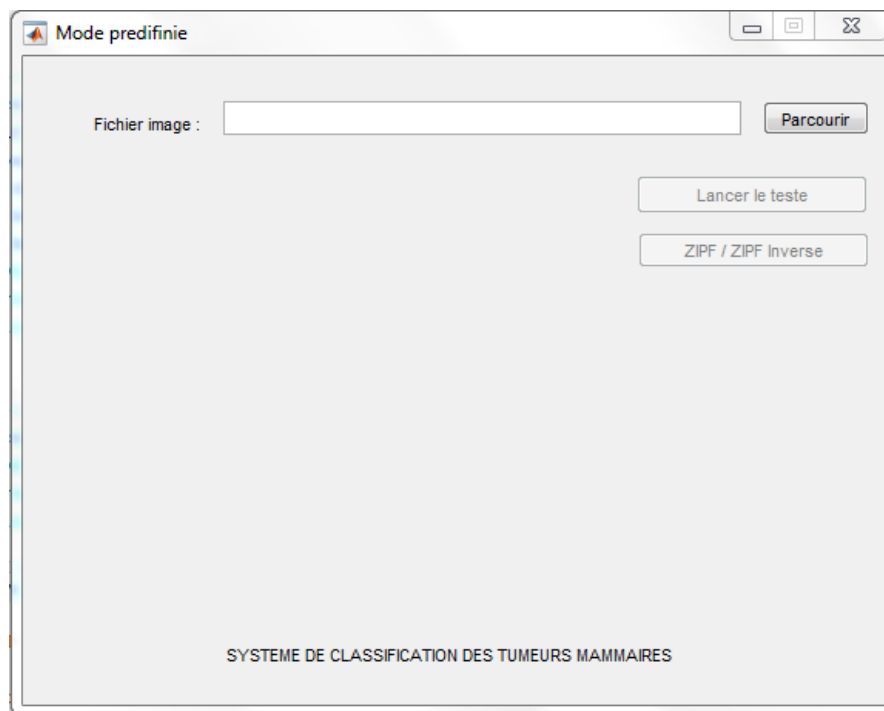


Figure IV.4 L'interface principale du système Mammo-ZipDFT.

Sur la figure IV.5, nous exposons la sortie du système Mammo_ZipDFT. Après avoir introduit la zone d'intérêt requête au système, les zones d'intérêt similaires seront récupérées et affichées afin de donner le résultat de la classification : tumeur maligne, tumeur bénigne ou tissu normal. Ceci en cliquant sur le bouton « lancer le teste ».

Si nous analysons le résultat donné sur la figure, les 3 régions d'intérêt récupérées par le processus de recherche portent un tissu normal, faisant en sorte que le radiologue considère la région d'intérêt requête comme portant un tissu sein et effectivement c'est le cas. De plus, le système Mammo_ZipDFT offre un affichage des voisins appartenant à la même classe pathologique de la région d'intérêt requête.

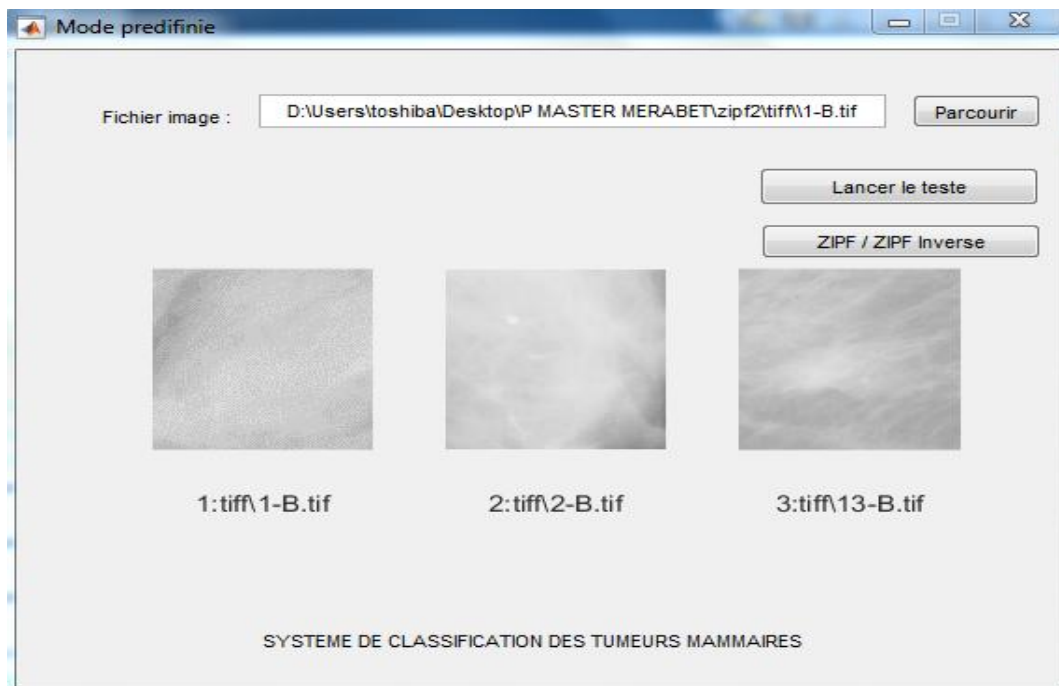


Figure IV.5 Processus de décision médicale issue de l'utilisation de la sortie de Mammo_ZipDFT par la considération de $k=3$.

IV.7. Évaluation des performances

Pour évaluer la performance de notre système, nous utiliserons l'exactitude de classification ou le taux de bonne classification.

Nous avons divisé la base de données MIAS suivant le principe de la validation croisée consistant en la division de l'échantillon de taille n en deux sous échantillons, le premier est celui d'apprentissage (supérieur à 60 % de l'échantillon) et le second est celui de test. À cet effet, nous avons partitionné la base MIAS en deux segments, le premier qui est $> 60\%$ de l'échantillon

consistant en l'ensemble des données d'apprentissage (80 images mammaires) et le deuxième segment (40 images mammaires) pour le test.

Malgré son efficacité dans le traitement des tâches de classification, le KPPV présente l'exigence de la détermination de la valeur optimale des K plus proches voisins.

Nous avons procédé empiriquement en considérant 4 valeurs du K et nous avons obtenus les taux de classification en divisant le nombre de zones d'intérêts dont la pathologie est correctement identifiée par le nombre global de zones d'intérêts.

Le tableau IV.1 démontre les différents taux de classification donnés par notre système pour quatre valeurs différentes du K.

Tableau IV.1 Variation de la précision de classification du système «Mammo_ZipDFT» pour diverses valeurs des k plus proche voisins.

Les valeurs des K plus proches voisins	Taux de classification obtenus
3	75%
5	80%
6	77%

Nous constatons que la précision de la classification varie entre 75% et 80% et que le meilleur taux de classification est obtenu pour K=5.

IV.8. Conclusion

Dans ce chapitre nous avons proposé un système d'indexation et de recherche de mammographies par le contenu basé sur une technique template-matching à travers le classifieur des k plus proches voisins. Notre système a abouti à des taux de classification encourageant variant entre 75% et 80%. Nous pouvons attester que les descripteurs de texture générés par l'analyse des mammographies par la fusion des lois de Zipf et de Zipf inverse et la transformée de Fourier indiquent une pathologie, en effet, lorsque nous avons extrait et exploité ces nouveaux descripteurs, nous avons affirmé qu'ils fournissent assez de puissance de discrimination pour distinguer les tumeurs malignes des tumeurs bénignes ainsi que du tissu parenchymateux normal.

Conclusion générale

Dans ce travail, nous nous sommes intéressés à l'étude de l'imagerie médicale. Plus précisément, nous nous sommes focalisés sur le diagnostic médical du cancer du sein assisté par ordinateur en faisant appel aux techniques de la vision par ordinateur.

Nous avons fusionné les lois puissance : Zipf et Zipf inverse et la transformé de Fourier discrète pour la caractérisation du contenu structurel des zones d'intérêt extraites à partir des mammographies. En effet, nous avons développé un système d'indexation et de recherche des mammographies par le contenu se basant sur la fusion de ces descripteurs texturaux obtenus. Une analyse en composantes principales (ACP) a été appliquée aux vecteurs descripteurs pour diminuer leur dimension, qui est perçue à ce niveau comme une méthode de recherche de nouveaux descripteurs non corrélés et discriminants.

Notre système se base sur l'approche template-matching, qui affecte la région d'intérêt en cours d'analyse à la classe pathologique la majoritairement renvoyée par le processus de recherche, en utilisant le classifieur des K plus proches voisins.

La performance de notre système a été évaluée et estimée à un taux de bonne classification de 80%, ce qui présente des résultats encourageants.

Perspectives

Dans nos futurs travaux, nous explorerons d'autres techniques d'extraction de l'information de la texture à partir du spectre de la transformée de Fourier discrète.

Références

Maria Petrou, M. Garcia Sevilla, P. (2006) image Processing : Dealing with Texture

(mammographies numériques) source site : <http://www.radiologie-mim.fr/fr/133067486713651-mammographie-numerisee->.

Sonka, M. Hlavac, V. Boyle, R. (2008). Image Processing, Analysis, and Machine Vision. International Student Edition. Third Edition. Thomson Learning, part of the Thomson Corporation. USA.

Hamoud, M. (2015). «Indexation et segmentation d'images basées loi de Zipf et Zipf Inverse ». Thèse de doctorat. Université de Annaba.

Cheikhrouhou, I. (2012). « Description et classification des masses mammaires pour le diagnostic du cancer du sein ». Thèse doctorat. Université d'Evry-Val d'Essonne.

Haralick, R.M. (1979). «Statistical and structural approaches to texture». Proceedings of the IEEE, 67. pp786–804.

Ludovic Paulhac. Outils et méthodes d'analyse d'images 3D texturées : application à la segmentation des images échographiques. Interface homme-machine [cs.HC]. Université François Rabelais - Tours, 2009. Français. . <tel-005765

Drbohlav, O. Leonardis, A. (2010). Towards correct and informative evaluation methodology for texture classification under varying viewpoint and illumination. Computer Vision and Image Understanding, 114(4). pp. 439–449.

Chen, C. H. Lee, G. G. (1997). On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis. Graphical Models Image Processing, 59. pp. 349–364.

Tuceryan, M. Jain, A. K. (1998). Texture Analysis. The Handbook of Pattern Recognition and Computer Vision (2nd Edition). World Scientific Publishing Co. pp. 207-248.

Amroun, F. (2013). Extraction de la composante texturée d'une image. Mémoire de magister en automatique. Université Mouloud Mammeri. Tizi Ouzou.

John Lafferty (2001) .Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. Carnegie Mellon University, lafferty@cs.cmu.edu

Journet, N. Ramel, J. Y. Eglin, V. Mullot, R. (2007). Analyse d'images de documents anciens: une approche texture. Traitement du signal, 24. pp. 461-479.

- Uttama, S. Ogier, J.M. Loonis, P. (2005). Top-down segmentation of ancient graphical drop caps: lettrines. In Proceedings of 6th IAPR International Workshop on Graphics Recognition, Hong Kong. Pp. 87–96.
- Caron, Y. (2004). Contribution de la loi de Zipf à l'analyse d'images. Thèse de doctorat. Université de Tour.
- Anderson ,ch.(2004) Les lois de Zipf et de Mandelbrot site :[http ://www.Les lois de Zipf et de andelbrot.html](http://www.Les lois de Zipf et de andelbrot.html)
- Djaroudib ,KH. (2016) «Détection Multicritères des Anomalies sur des Mammographies ». Thèse de Doctorat en Sciences en Informatique Université de Batna
- Merouani, H.F (1999). A Markov Random Field Approach to the Analysis of Texture in Digitized Mammograms. These de doctorat. Robert Gordon University Aberdeen UK.
- Bikesh Kumar Singh , Jitendra Singh Parihar , Pritish Pal (2011) «Wavelet Based information for Retrieval and Classification of Mammographic Images».
- Passeport sante (2012).site [http ://www.passeportsante.net/fr/](http://www.passeportsante.net/fr/).
- Cancer du sein (2016) site [http ://curie.fr/fr/la-fondation/espace-informations/les-dossiers-pedagogiques/cancer du sein/comment-diagnostiquer-un-cancer-du-sein-002331](http://curie.fr/fr/la-fondation/espace-informations/les-dossiers-pedagogiques/cancer_du_sein/comment-diagnostiquer-un-cancer-du-sein-002331)
- Chang, B. Tsai, H. Chou, W. (2013). Using visual features to design a content-based image retrieval method optimized by particle swarm optimization algorithm. Engineering Applications of Artificial Intelligence, 26. pp. 2372–2382.
- Hanifi. M. (2009). Extraction de caractéristiques de texture pour la classification d'images satellites. Thèse de doctorat. Université de Toulouse III.
- Regniers ,O. (2014) Méthodes d'analyse de texture pour la cartographie d'occupations du sol par télédétection très haute résolution». Thèse doctorat. Université de Bordeaux.
- BARRÉ,A .DOLLET,L (1999) analyse de textures par Fourier
site :www.tsi.enst.fr/pages/enseignement/ressources/mti/textures_Fourier/texture.html
- Chaman L. Sabharwal 1 S.R. Subramanya(2001) Indexing Image Databases Using avelet and Discrete Fourier Transform Computer Science department, University of Missouri-Rolla ,pp 434-439
- Girod,B. (2013) « Digital Image Processing » , Stanford University
- CHAMROUKHI,F (2013) « Classification supervisée -Les K-plus proches voisins » Université du Sud Toulon – Var

Miloudi,N. (2014). «Contribution à l'étude de la vulnérabilité des réservoirs en béton par analyse des composantes principale ». Thèse de doctorat. Université Tizi Ouzou.

Glorieux ,q « Outils Mathématiques et utilisation de Matlab » Cours (2013-2014) Licence Université Pierre et Marie Curie - Paris VI

Loye ,A. (2009-2010) «Représentations graphiques » Cours MATLAB UNIL-FGSE – Université laussane.

Hamoud, M. Merouani, H.F. (2011). Detection of a zone of interest in the images based on Zipf laws. Conférence internationale sur Computer Science and its Applications. Saida, Algérie.

MIAS: Database was created by the Mammographic Image Analysis Society (MIAS), United Kingdom. Source: <http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html> mias@svl.smb.man.ac.uk.

P. Swaroop N.Sharma (2016) «An Overview of Various Template Matching Methodologies in Image Processing» International Journal of Computer Applications