Représentation de la variabilité dans le traitement d'images flou

Application à la mammographie numérique

THESE

présentée à par

Andreas RICK

en vue de l'obtention du grade de

DOCTEUR de l'UNIVERSITE PARIS 6

Spécialité :
Informatique

Soutenue le 17 décembre 1999
devant le jury composé de :

Bernadette BOUCHON-MEUNIER	Directrice de Thèse
Isabelle BLOCH	Rapporteur
Michael BRADY	Rapporteur
Herman AKDAG	Examinateur
Patrick GALLINARI	Examinateur
Serge MULLER	Examinateur

Résumé

Représentation de la variabilité dans le traitement d'images flou

Application à la mammographie numérique

Pour de nombreuses tâches de classification dans le cadre du traitement d'images, la variabilité d'une image à l'autre représente une limite de la performance de la classification. Un exemple illustrant ce problème est la détection des lésions en mammographie.

Dans cette thèse, notre but est de réduire l'influence de la variabilité inter-images, en utilisant le contexte dans l'image. A cette fin, nous proposons une méthode de paramétrisation des histogrammes d'attributs calculés sur l'image, qui permet une représentation compacte de ces histogrammes avec peu de paramètres. A partir d'une base d'apprentissage, nous construisons un prototype qui contient un modèle de la variabilité et des liens entre les paramètres.

Le modèle établi de cette manière est ensuite utilisé pour adapter le prototype à chaque nouveau cas.

Cette nouvelle démarche, proposée pour le traitement de la variabilité dans le cadre de la classification floue, est appliquée à une base de données simulée et à une base d'images mammographiques dans le but de détecter des opacités.

Mots clés :

Classification floue	Apprentissage inductif
Variabilité	Mélanges
Algorithme EM	Régression
Prototypes	Modèles de Markov
Mammographie	Aide à la détection

Abstract

Representation of variability in fuzzy image processing

Application on digital mammography

One of the most important factors, which limit the performance of classifiers in many image processing tasks, is the variability. In this thesis, we will propose a new method to reduce the effect of variability in a fuzzy classification system for image processing by using the context in the image. The method consists of the following steps :

- Parameterization of the histograms of all the attributes that are calculated on the images for the classification task which gives a compact description with a small number of parameters.
- Conctruction of a prototype which includes the variability and the interactions between the parameters using the learning database.
- Adaptation of the model to the current case by adapting the parameters of the prototype.

The proposed method is applied to a synthetic data base and to a database of mammography images for the detection of dense lesions.

Key words :

Fuzzy Classification	Inductive Learning
Variability	Mixture Models
EM Algorithm	Regression
Prototypes	Markov Models
Mammography	Computer Aided Detection

à Virginie
à mes parents

Remerciements

La réalisation de cette thèse a été possible grâce à une collaboration entre le laboratoire d'informatique de l'Université Paris VI (Lip6) et le département des applications avancées de GE Medical Systems.

Dans ce cadre, je tiens à remercier en premier lieu *Mme Bernadette Bouchon-Meunier*, ma directrice de thèse, pour l'accueil dans son équipe LOFTI et sa disponibilité tout au long de ce travail.

J'exprime ma profonde gratitude à *M. Serge Muller*, mon tuteur au sein de GE Medical Systems, qui a non seulement initié ces travaux, mais qui par son encadrement, son soutien journalier et ses conseils, a tout simplement rendu possible cette thèse. Malgré sa charge de travail importante, il a toujours été disponible pour une discussion sur des aspect techniques aussi bien que médicaux.

Je tiens à exprimer ma profonde reconnaissance à *Herman Akdag* et *Patrick Gallinari* pour avoir accepté d'examiner ce travail et à *Isabelle Bloch* et *Mike Brady* pour avoir accepté d'en être rapporteurs. C'est un grand honneur pour moi qu'ils aient accepté de participer au jury de cette thèse malgré leurs nombreuses occupations.

Je voudrais dire merci à tous les membres de l'équipe LOFTI, qui au travers de leurs présentations, leurs commentaires lors de mes présentations et des discussions que nous avons eues ensemble, m'ont permis plusieurs fois de trouver les maillons manquants pour la continuation de la chaîne de mon travail. Merci donc à *Christophe·2, Maria, Leila, Nedra, Virginie, Jannick, Nathalie, Marcin* et tous les stagiaires.

Merci à *Sylvie* pour son aide pendant la première phase de ma thèse, les discussions enrichissantes sur les différents aspects du CAD en mammographie et -plus tard- pour son support dans divers problèmes liés à la revue d'images ou à des aspects plus proches du produit FFDM.

Merci aux membres de l'équipe Applications Avancées pour le bon climat de travail, tout le support, les encouragements et les innombrables discussions autour de la machine à café - parfois plus fructueuses que des réunions planifiées une semaine auparavant.

Je tiens à remercier les autres thésards CIFRE de chez GE Medical Systems pour les discussions pendant les semaines CIFRE et surtout *Erwan* et *Armelle* qui ont travaillé aussi sur les sujets de l'équipe Applications Avancées. Avec eux, j'ai pu discuter profondément sur des aspects théoriques de notre travail à de nombreuses occasions, la plus remarquable étant la semaine de *Summer School of Biomedical Engineering* dans l'été 1998.

Merci également à toutes les personnes de GE-Medical Systems, qui m'ont montré parfois avec tact, parfois avec franchise, ce que veut dire *travailler dans l'industrie* et spécialement dans une entreprise américaine.

Sans être directement lié au travail présenté dans ce mémoire, je ne voudrais pas oublier de mentionner *Jean-Luc*. Grâce à son esprit ouvert, il m'a fait participer à son projet black-belt, qui m'a permis de comprendre la méthodologie 6σ avant d'en avoir suivi les formations. Je lui en suis très reconnaissant, car ce projet m'a donné le goût pour ce style de travail - très concret, sur un problème précis d'un produit, sans pourtant sacrifier la rigueur scientifique.

En pensant aux épreuves administratives - qui dans mon cas tordu (un allemand dans une entreprise américaine en collaboration avec une Université française) n'étaient pas toujours évidentes - je ne peux pas oublier de mentionner *Sylvia Bagot* de la direction des ressources humaines de GE et *Ghislaine Mary* du Lip6 et de m'incliner devant un travail si soigneusement fait et qui a permis - entre autres - de remédier à une lettre importante perdue par la Poste au mauvais moment.

Je voudrais remercier les *Dr. Godschalk (Paris)*, *Dr. Levy (Paris)* et les équipes qui ont conçu les bases d'images mammographiques MIAS et DDSM, les clichés de mammographie dont j'ai pu disposer pour cette étude.

Cette thèse a été réalisée dans le cadre d'une convention CIFRE (n^o 96457).

Table des matières

Chapitre 1

Introduction

Depuis l'introduction des sous-ensembles flous (*fuzzy sets*) en 1965 par Zadeh, les méthodes de la logique floue ont été appliquées à presque tous les domaines techniques du traitement de l'information. Le premier succès commercial grand public des applications se reposant sur la logique floue a été dans le domaine des systèmes de contrôle pour des apparails électro-ménagers. Ces systèmes se caractérisent par relativement peu de variables d'entrée et de sortie, mais la logique floue permet d'exprimer et d'évaluer des règles imprécises définies par des experts.

Aujourd'hui, la logique floue est également appliquée à des systèmes beaucoup plus complexes, comme le traitement d'images. Une des idées principales de la logique floue, qui contribue à son succès, est la formalisation partielle de la solution à un problème par l'expert, en utilisant des règles simples et imprécises. Grâce à la logique floue, une telle règle peut être exploitée pour résoudre le problème, même si la solution idéale du problème est impossible à formaliser. Les problèmes trouvés dans le traitement d'images présentent très souvent cette caractéristique. Un expert humain est souvent capable de résoudre un problème de vision avec une vitesse et une précision élevées, mais il est incapable de décrire d'une manière précise les critères qu'il a utilisés pour sa décision. Dans ce contexte, la formalisation de ces critères par des règles imprécises permet souvent d'extraire la partie des connaissances conscientes. Dans d'autres cas, la partie de la connaissance formalisable par des règles, même imprécises, n'est pas suffisante pour résoudre le problème d'une manière satisfaisante.

Une des causes principales de ces échecs est la grande variabilité qui existe entre différentes images.

La variabilité dans le traitement d'images

En traitement d'images, cette variabilité peut être due à un grand nombre de facteurs, comme le changement d'illumination ou le changement des positions des objets. La figure 1.1 représente le schéma d'un système typique de classification dans un problème de traitement d'images. Le but du système est de prendre une décision sur l'image - ici la présence d'une route - dans une situation donnée. Cette situation se caractérise par l'état a vrai, qui décrit la présence ou l'absence d'une route ou sa position, mais aussi par un ensemble d'autres paramètres $(c_1 \cdots c_m)$. Cette scène est observée par une caméra qui, elle-même, est caractérisée par des paramètres $(c_{m+1} \cdots c_n)$. L'image résultante dépend donc de paramètres importants pour l'étape de décision mais aussi d'autres paramètres. Des opérateurs de mesure sont utilisés pour calculer des attributs de l'image et la décision se fait à partir de ces mesures. Les paramètres c de la situation et de la caméra, comme par exemple la position de la caméra ou le changement de l'illumination, se traduisent dans les images par une variabilité qui peut se répercuter également sur les attributs.

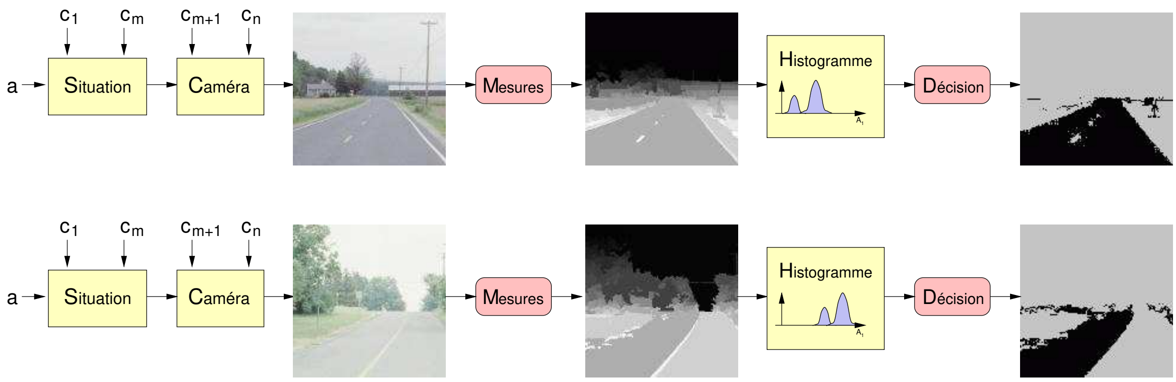


FIG. 1.1 – Problème de la variabilité

Sur la figure 1.1, deux situations similaires sont montrées, mais les variations dans les paramètres c changent l'image. Les histogrammes d'un attribut mesuré sur l'image varient avec les paramètres c de la situation et de la caméra. Un système de décision non-adaptatif ne permet alors pas de retrouver le bon résultat.

Trouver un raisonnement pour prendre une décision basée sur des attributs qui ont une forte variabilité, est généralement difficile.

Dans de nombreux cas, le contexte de l'image contient des informations qui permettent d'adapter le raisonnement au cas concret. Confronté à un problème avec des paramètres très variables, l'expert humain est souvent capable de trouver des règles linguistiques qui s'appliquent à tous les exemples, mais dont les termes utilisés dépendent du contexte.

Le but de nos travaux est de trouver une méthode d'estimation de la variabilité présente dans un problème de traitement d'images, qui permet de prendre en compte cette variabilité pour augmenter la performance du système.

Aide au diagnostic en mammographie

La méthode proposée est appliquée au problème de l'aide au diagnostic en mammographie (figure 1.2). La dépistage systématique du cancer du sein par mammographie nécessite l'analyse d'un très grand nombre de clichés par des experts radiologues. Avec l'arrivée sur le marché de systèmes de mammographie entièrement numériques, les systèmes d'aide à la détection deviennent très importants pour faciliter la tâche du radiologue. Le but de ces systèmes est d'indiquer au médecin des éventuelles zones suspectes (comme celle en haut des images de la figure 1.2), qui pourraient être la traduction d'une lésion cancéreuse sur l'image radiologique.

Dans ce contexte, la grande variabilité des images de mammographie représente une des limitations majeures pour des systèmes de détection et de classification de lésions.

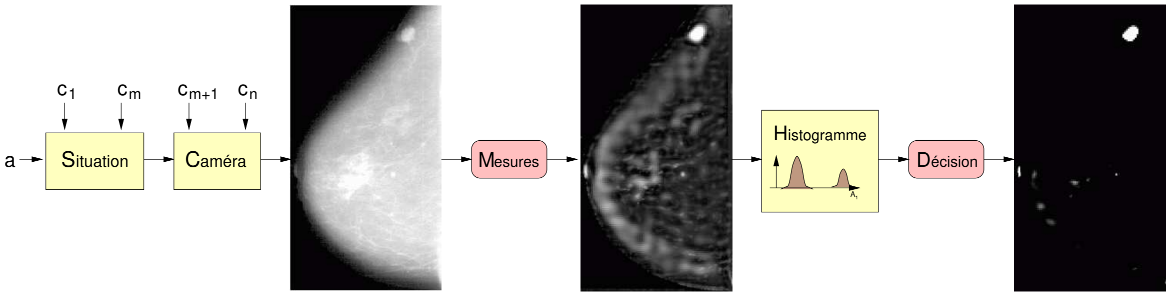


FIG. 1.2 – Problème de la variabilité

Première partie : Présentation du traitement de la variabilité

Dans la première partie de cette thèse, nous présentons une méthode de classification utilisant des fonctions d'appartenance adaptatives, qui permettent d'extraire des règles imprécises pour des données qui contiennent beaucoup de variabilité.

Le chapitre 2 est consacré aux systèmes de classification en général et aux problèmes spécifiques que nous retrouvons dans les problèmes de traitement d'images. Nous regardons en particulier l'incertitude en classification et sa gestion au travers des étapes du traitement. Nous montrons que la variabilité entre images joue un rôle important dans le traitement des incertitudes.

Le chapitre 3 est consacré à la première étape de la construction des fonctions d'appartenance adaptatives : l'extraction d'une description paramétrique des densités d'exemples dans un histogramme.

Dans le chapitre 4, nous utilisons cette description paramétrique pour construire, à partir de la base d'apprentissage, un prototype qui prend en compte la variabilité entre images. Les liens entres les différents paramètres sont également déterminés, ce qui permet de prendre en compte les classes très peu représentées dans les images. Nous proposons un algorithme pour l'adaptation du prototype à l'image courante.

Dans le chapitre 5, différentes méthodes de normalisation sont décrites, qui permettent de convertir un modèle d' histogrammes en fonctions d'appartenance.

Pour résumer le fonctionnement de la méthode de classification, basée sur les fonctions d'appartenance adaptatives, que nous proposons, nous montrons au chapitre 6, toutes les étapes et appliquons la méthode à une base de données synthétisée.

Deuxième partie : Application à l'aide au diagnostic en mammographie

Dans la deuxième partie, nous considérons le diagnostic du cancer du sein. Nous montrons l'application des sous-ensembles flous adaptatifs à la détection de lésions en mammographie.

Le chapitre 7 introduit l'utilisation de la mammographie dans la détection précoce du cancer du sein. Les avantages de la mammographie numérique sont discutés avec la possibilité du traitement des images. Les sources de variabilité, qui peuvent gêner les algorithmes de traitement, sont décrites.

L'aide à la détection des lésions par ordinateur est discutée au chapitre 8 avec une revue des méthodes existantes. Trois méthodes sont présentées avec plus de détails pour la détection des lésions denses, des lésions stellaires et des microcalcifications.

Dans le chapitre 9, nous montrons l'application de la méthode de traitement de la variabilité proposée au problème de la détection des lésions denses.

La réduction de la variabilité due aux paramètres de la caméra (du système d'acquisition) par la mesure de ces paramètres et de la modélisation de leurs effets sur l'image, est illustrée dans le chapitre 10. Nous montrons l'application d'une telle modélisation à la détection des microcalcifications.

Dans la conclusion, nous résumons les principaux apports de nos travaux de thèse, et nous proposons des extentions possibles de ces travaux.

Annexes

Dans l'annexe A, nous décrivons l'évaluation de la performance des systèmes de classification avec la méthodes des courbes COR.

L'annexe B fait le lien de l'algorithme P-EM avec une probabilité a priori.

Première Partie

Classification floue avec adaptation au contexte en traitement d'images

Chapitre 2

La classification et l'apprentissage dans le traitement d'images

Le traitement d'images est utilisé dans des domaines techniques très différents ; l'identification d'un objet (reconnaissance des caractères, reconnaissance des empreintes ou du visage), l'analyse d'un état (fonctionnement normal, erreur) dans la surveillance du processus de fabrication, l'imagerie médicale et les applications multimédia (télévision, jeux, CDROMs)

Un grand nombre d'applications de traitement d'images fait appel à des méthodes de classification. Dans ce chapitre, nous discutons des différents aspects de la classification en nous focalisant sur les méthodes applicables aux problèmes du traitement d'images. Les méthodes d'apprentissage pour la classification sont présentées. Les différentes sources d'imperfections contenues dans les mesures prises dans le traitement d'images sont identifiées. Finalement, l'utilisation du contexte pour l'amélioration des performances est abordée.

2.1 La classification

La classification est l'affectation d'une étiquette à des exemples d'une base de données en utilisant un certain nombre d'attributs connus pour chaque exemple.

Dans le traitement d'images, *exemple* peut signifier soit un pixel d'une image, une zone dans l'image, un objet dans une scène réelle représenté dans l'image ou l'image elle-même. Les *attributs* utilisés sont généralement des mesures prises sur un voisinage de pixels, sur une zone de l'image ou sur l'image entière. Les attributs peuvent provenir de différents traitements sur une seule image comme, par exemple, une décomposition fréquentielle. Les attributs peuvent aussi provenir de différentes images acquises séparément, par exemple des images en couleur, où chaque image est acquise avec un capteur sensible à différentes longueurs d'ondes, ou bien des images médicales acquises par IRM (imagerie de résonance magnétique) avec des pondérations (T1, T2) différentes.

Les *labels* ou *classes* sont les résultats de la classification. Selon l'application, le but est de classer chaque pixel de l'image ou de classer l'image dans son ensemble. Dans le premier cas, on parle souvent de *segmentation* d'images car l'image est séparée en zones appartenant à différentes classes.

2.1.1 Les méthodes de classification

La figure 2.1 donne un résumé des différentes méthodes de classification. Ces catégories ne sont qu'une manière de représenter ces méthodes ; en fait, la figure 2.1 utilise une méthode de classification, les arbres de décision, pour classer toutes les méthodes de classification. Les critères utilisés (ou *attributs*) sont indiqués à gauche. En bas, les feuilles de l'arbre indiquent les catégories ou *classes*.

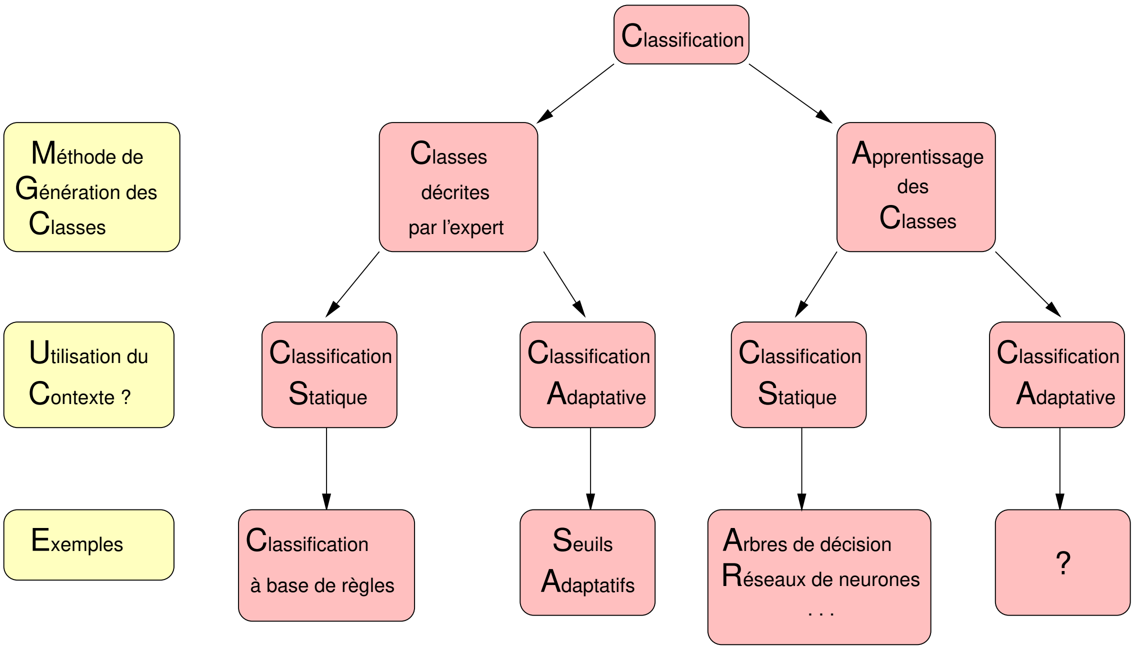


FIG. 2.1 – Les méthodes de classification

Le premier attribut que nous avons choisi ici est la façon dont le système de classification est construit. D'une part, il y a les systèmes qui utilisent directement les connaissances exprimées d'une manière explicite par un expert du domaine, d'autre part il y a les systèmes d'apprentissage qui analysent une base de données pour trouver les caractéristiques des classes.

Le deuxième critère utilisé ici est l'utilisation du contexte. On distingue les systèmes indépendants du contexte où les caractéristiques sont constantes pour tous les exemples, des systèmes dépendants du contexte où les caractéristiques des classes sont adaptées au contexte. Dans des problèmes de traitement d'images, on est confronté fréquemment au besoin de prendre une décision sur une partie d'une image. Dans ce cas, les méthodes de classification dépendantes du contexte permettent d'adapter les décisions à des caractéristiques de l'image en entier.

2.1.2 La classification dans le traitement d'images

La performance des humains dans de nombreuses tâches de traitement d'images en général et de la classification en particulier est supérieure aux performances des systèmes de vision artificielle. Très souvent, cette performance repose sur une connaissance implicite et les experts ne sont pas capables d'exprimer plus qu'une partie de cette connaissance d'une manière explicite. Un exemple pour cet effet se trouve dans [Tay99]. Dans leurs travaux, Taylor et al. ont essayé d'extraire des publications médicales un consensus d'experts pour la classification des microcalcifications en mammographie et de le mettre en correspondance avec des mesures prises sur des mammographies. Ils ont constaté une grande difficulté à trouver une description acceptable pour la plupart des médecins et à mettre en correspondance les mesures prises sur les images avec les descriptions des médecins.

L'extraction directe des connaissances de l'expert pour la classification en traitement d'images est donc un problème difficile. Une autre méthode d'extraction des connaissances est de demander à un expert de faire la classification en question sur une base d'exemples. Cette base de résultats de classification est utilisée ensuite pour un apprentissage des classes. L'avantage de cette approche est que l'expert n'est pas obligé d'exprimer d'une manière explicite sa connaissance, il suffit qu'il les applique sur un nombre suffisant d'exemples. Si le problème le permet, des résultats de classification peuvent aussi être obtenus par d'autres processus de mesure qui sont accessibles pour une base d'apprentissage mais ne le seront pas dans la tâche finale. Pour donner un exemple, nous pouvons imaginer des mesures de qualité supplémentaires d'un système de surveillance d'une chaîne de production pendant la phase d'apprentissage qui ne seront pas possibles dans le fonctionnement normal (par exemple si les tests sont destructifs). Cette manière de construire une description des classes est réalisée dans des systèmes d'apprentissage inductifs.

2.2 Systèmes d'apprentissage inductif

Un système de classification qui utilise l'apprentissage inductif consiste en un module d'apprentissage et en un module de classification.

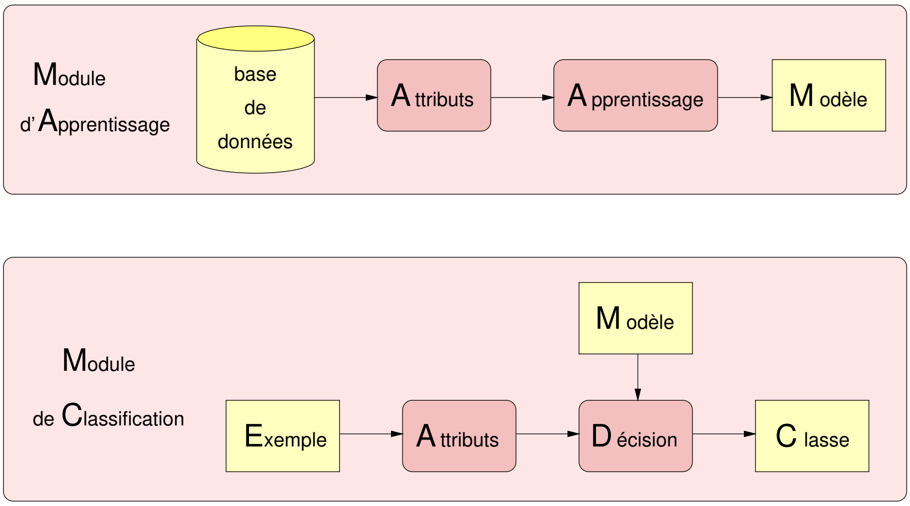


FIG. 2.2 – Apprentissage inductif

Le module d'apprentissage analyse les exemples présents dans la base de données et les généralise pour obtenir un modèle valable sur ces exemples et sur d'autres cas non présents dans la base.

Le module de classification utilise le modèle construit par apprentissage pour classer de nouveaux exemples.

Sans passer en revue tous les systèmes d'apprentissage publiés (une bonne synthèse des méthodes est présentée dans [Tor95] ou [Bri96]) nous allons discuter de quelques points clés :

– le domaine des variables à l'entrée et à la sortie du système
– l'apprentissage supervisé ou non-supervisé
– la répartition du travail entre les modules d'apprentissage et de classification
– le mode de présentation des exemples : global ou incrémental
– la représentation interne des données dans le modèle

2.2.1 Les variables d'entrée et de sortie

Les variables d'entrée d'un système de classification en traitement d'images sont des attributs calculés sur les images. Ces attributs peuvent être calculés sur une position de l'image, une zone ou l'image entière. La variable de sortie est la classe à attribuer au pixel, à la zone ou à l'image.

Pour un système de segmentation par exemple, les variables d'entrée sont typiquement soit directement les valeurs des pixels soit des pixels prétraités (par des filtrages ...). La variable de sortie est la classe de la zone segmentée.

Dans le cas de la classification pour les applications dans le traitement d'images, les relations spatiales entre les classes peuvent être importantes. Dans la suite de ce travail, nous considérons que ces informations spatiales sont extraites entièrement pendant la phase du calcul des attributs. Pour le système de classifications, l'arrangement spatial des exemples n'est donc plus à prendre en compte parce-qu'il est déjà traduit dans les valeurs des attributs. Cette hypothèse forte permet de considérer uniquement les histogrammes des valeurs d'attributs pour tous les exemples de la base.

2.2.2 Apprentissage supervisé ou non-supervisé

On parle d'apprentissage *supervisé* quand, lors de l'apprentissage, le système considère connue la *vérité* sur l'appartenance des exemples aux classes dans la base d'apprentissage. Le système peut alors valider son résultat par rapport à la vérité et adapter son modèle si les résultats ne sont pas corrects.

On parle d'apprentissage *non-supervisé* quand le système doit déterminer lui même les classes. L'apprentissage *non-supervisé* ne peut fonctionner que si les données ont déjà une structure interne.

Quand on fait l'apprentissage sur les bases d'images avec une vérité donnée par l'annotation de l'expert, on se trouve souvent dans un cas intermédiaire, à savoir que les annotations d'expert sont imparfaites, par exemple lorsqu'il les trace sur l'image. Nous discutons de cette situation plus en détail dans chapitre 5. Nous proposons alors une approche basée sur les sous-ensembles flous adaptatifs pour faire une combinaison entre l'apprentissage supervisé et non-supervisé afin d'utiliser à la fois la connaissance apportée par l'expert et la connaissance cachée dans la structure des données de la base d'apprentissage.

2.2.3 La répartition du travail entre les modules d'apprentissage et de classification

Un des principaux choix à effectuer dans les systèmes d'apprentissage inductifs est le moment où telle partie du travail doit être faite.

Dans des systèmes *paresseux*, les exemples de la base sont peu traités au moment de l'apprentissage et tout le travail se fait au moment de la classification. Un bon exemple est celui des systèmes de raisonnement à base de cas. Au moment de l'apprentissage, les exemples sont juste mémorisés. C'est au moment de la classification que la distance avec des exemples existants est évaluée pour arriver à une classification.

Les systèmes à base de règles et les arbres de décision tombent plutôt dans la catégorie de l'apprentissage *diligent*, puisque les informations apportées par les exemples sont transformées dans une structure beaucoup plus adaptée à la prise de décision. L'utilisation de ce type de systèmes est donc en général beaucoup plus rapide. Quand la performance au moment de l'utilisation du système est importante, on aura intérêt à choisir un système *diligent*. Les arbres de décisions en sont un bon exemple, car leur utilisation est très efficace.

2.2.4 Apprentissage global ou incrémental

Le mode de présentation des exemples a un impact important sur le système d'apprentissage. Les systèmes qui demandent une présentation globale de tous leurs exemples peuvent considérer ces exemples comme l'ensemble de l'information disponible et peuvent les transformer d'une manière qui leur permet d'avoir la meilleure performance dans l'utilisation de ces connaissances.

Ils n'ont pas besoin de mémoriser les exemples car ils ne seront plus jamais confrontés avec de nouveaux exemples à apprendre.

Au contraire, les systèmes incrémentaux doivent garder une représentation des exemples déjà vus, ce qui leur permet d'intégrer de nouveaux exemples. Généralement, cela implique de mémoriser d'une manière ou d'une autre la distribution des valeurs des attributs pour les exemples déjà vus. Dans un système de raisonnement à base de cas, tous les exemples sont stockés explicitement. Dans un réseau neuronal, les exemples sont stockés implicitement dans les poids des liens des neurones.

Le mode de présentation des exemples est directement lié à la manière de décrire les classes.

2.2.5 Modèle des distributions ou des frontières

Un choix à faire pour la conception d'un système d'apprentissage est la représentation des classes dans le modèle. Au départ, la base d'apprentissage donne une distribution d'exemples dans l'espace des attributs.

Le but de l'apprentissage est de séparer les classes dans l'espace des attributs par des frontières. Ces frontières peuvent être précises ou floues. Il est intéressant d'utiliser les modèles flous pour représenter la zone de transitions entre les densités d'exemples de deux classes voisines.

La question qui se pose est de savoir si le système de classification a besoin de connaître la répartition des exemples qui ont servi à obtenir ces frontières une fois que l'apprentissage est effectué.

Si le modèle qui est établi pendant l'apprentissage contient uniquement les zones de transitions autour des frontières comme illustré sur la figure 2.3(b), le système de classification ne peut pas donner d'indication sur le nombre d'exemples qui ont servi à apprendre le classement dans une zone de l'univers des attributs donnée. Si le modèle contient également une description des densités d'exemples dans l'univers, comme illustré sur la figure 2.3(a), le nombre d'exemples qui ont servi pendant l'apprentissage est disponible pour la décision.

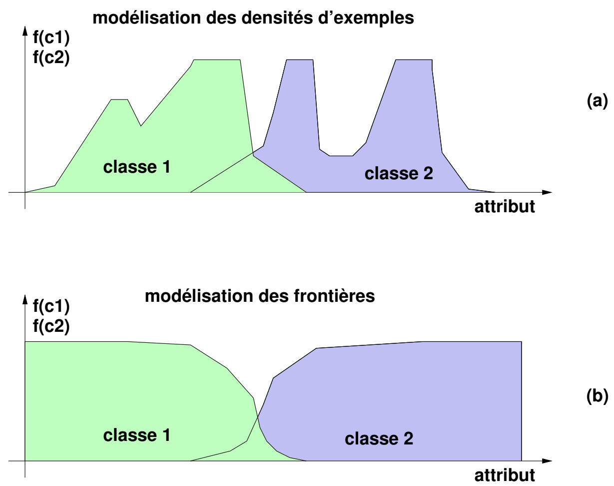


FIG. 2.3 – Modélisation de densités d'exemples ou modélisation de frontières

Nous montrons dans la suite, qu'une modélisation des distributions des exemples sur l'univers des attributs va nous permettre d'adapter les fonctions d'appartenance et donc implicitement les frontières entre les classes

2.3 Les limites des systèmes de classification

Après avoir discuté des systèmes de classification et des systèmes d'apprentissage pour la classification, nous allons regarder les facteurs qui limitent leur performance.

Les limites majeures viennent des données elles mêmes, aussi bien pendant la phase d'apprentissage que pour l'utilisation. Dans de nombreux systèmes de classification, les données ne contiennent pas suffisamment d'information sur la classe pour réussir une classification parfaite. Si deux exemples dans la base ne se distinguent pas dans leurs attributs et n'appartiennent pas à la même classe, il est impossible de les classer correctement uniquement à partir de ces attributs. De plus, les données sont souvent :

– *imprécises*, par exemple quand un expert doit tracer des contours pour délimiter des zones dans l'image,
– *incertaines*, par exemple quand il y a plusieurs décisions possibles,
– *variables* d'une image à l'autre, par exemple du fait des conditions d'éclairage.

2.3.1 Classification par les experts humains

En essayant de construire un système d'apprentissage qui fonctionne dans un contexte imparfait, la question se pose de savoir comment les êtres humains réussissent dans ces tâches.

On pourrait supposer que les experts humains ont encore plus de mal à réussir ces tâches puisqu'ils utilisent pour ces analyses des capteurs (les yeux) qui ajoutent eux-mêmes encore des imprécisions, et qui ne permettent pas d'effectuer des mesures absolues, par exemple sur la taille ou sur des niveaux de gris. Par rapport au lecteur humain, un système automatique de classification peut se baser sur des mesures de la taille et du niveau de gris beaucoup plus précises et même utiliser des critères calculés grâce à une modélisation physique du processus d'acquisition (voir chapitre 10).

Compte-tenu du fait que les résultats des détections humaines sont relativement bonnes, nous pouvons conclure qu'il est possible de réussir dans cette tâche sans mesures absolues et sans une grande précision.

2.4 Méthodes du traitement de l'incertitude, de l'imprécision et de la variabilité

Le modèle de traitement de l'incertitude le plus connu est celui basé sur la théorie des probabilités. Il permet de traiter des incertitudes liées à des fréquences d'apparition d'événements et a été pendant très longtemps le seul moyen de modéliser l'incertitude. Son avantage est sa base mathématique solide et la grande exploration de la méthode qui en a été faite par les statisticiens. Malheureusement, la notion de probabilité n'est pas très bien adaptée à traiter d'autres aspects du raisonnement humain, comme l'imprécision et la confiance.

La notion de degré de confiance a été introduite dans la théorie de l'évidence de Dempster et Shafer. Ils n'imposent plus l'additivité de la confiance comme pour les probabilités et peuvent traiter des données sans attribuer une valeur de probabilité à chaque élément de l'univers.

En ce qui concerne la modélisation de l'imprécision, la notion d'intervalle d'erreur (utilisée par exemple en physique pour décrire l'erreur de mesure) est la plus ancienne. C'est par l'introduction des sous-ensembles flous en 1965 que Zadeh rapproche la notion d'intervalle et le raisonnement humain en considérant plutôt des frontières graduelles que des frontières abruptes. Les innombrables travaux sur la théorie des sous-ensembles flous qui ont été publiés depuis, en ont fait un modèle puissant pour modéliser l'imprécision.

2.4.1 Modélisation de l'imprécision avec la théorie des sous-ensembles flous

Dans le raisonnement humain, les frontières nettes entre deux classes sont rares. Il semble beaucoup plus naturel de raisonner avec des frontières graduelles.

Si, par exemple, nous demandons à un dégustateur de classer des fraises selon leur degré de sucre dans les classes *pas du tout sucrée (p)*, *légèrement sucrée (l)*, *sucrée (s)* et *très sucrée (t)*, et si nous regardons son classement par rapport au degré de sucre mesuré de chaque fraise (figure 2.4), nous constatons que les fraises ne sont pas classées dans l'ordre strictement croissant, mais qu'au contraire il existe des zones de transitions où les fraises sont mélangées.

Nous pouvons utiliser les fonctions d'appartenance pour décrire les classes identifiées par notre expert, par exemple en choisissant les zones des fraises mélangées comme les zones de transition des fonctions d'appartenance trapézoidales (figure 2.4).

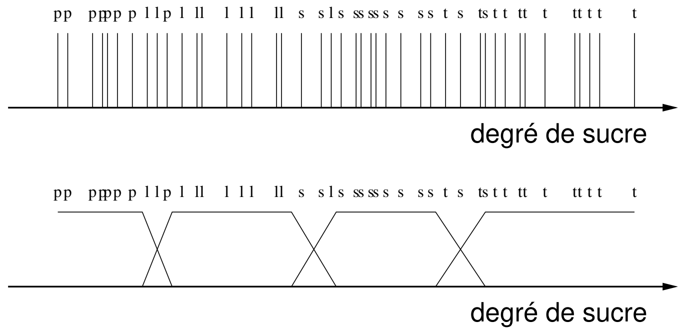


FIG. 2.4 – Construction des fonctions d'appartenance à partir de la classification par un expert

Ces fonctions d'appartenance sont subjectives, c'est-à-dire que le classement n'est pas indépendant de l'expert ; chaque expert nous donne des classes légèrement différentes. Le classement n'est pas non plus indépendant de la situation, parce que si notre expert a mangé des aliments très sucrés juste avant de classer les fraises, nous obtenons des fonctions d'appartenance décalées vers la droite (globalement, les fraises seront jugées moins sucrées).

A première vue, cette subjectivité peut être considérée comme une faiblesse de notre "classificateur", mais souvent cette subjectivité est une adaptation des classes prédéfinies à l'environnement actuel et permet de garder les règles générales valides.

Pour montrer les avantages d'un système adaptatif, nous demandons à l'expert de classer des fraises provenant de quatre fraisiers différents selon leur provenance et nous supposons que la seule différence des quatre types de fraises est leur degré de sucre. Dans un apprentissage supervisé, c'est-à-dire qu'avant le classement des fraises on présente au dégustateur des fraises de chaque fraisier en lui précisant leurs origines, le dégustateur se forme une opinion sur le degré typique des fraises de chaque fraisier. Pendant cette phase d'apprentissage, l'expert adapte les limites de ses fonctions d'appartenance pour refléter les provenances.

L'expert peut adapter ces fonctions même pendant la phase de classification - pendant laquelle il doit trouver la provenance d'autres fraises- si les données elles-mêmes contiennent une structure interne, par exemple un nombre de fraises élevé avec un degré de sucre similaire (figure 2.5).

Dans ce cas, les fonctions d'appartenance peuvent être retrouvées même si la mesure est perturbée (l'expert a mangé quelque chose de très sucré juste avant la dégustation) puisque la structure interne des données - ici l'absence d'exemple d'un degré de sucre intermédiaire - permet de recalibrer la mesure.

Dans le système de gestion de variabilité que nous présentons ici, nous essayons de copier ce fonctionnement qui correspond à un apprentissage à moitié supervisé et à moitié non-supervisé.

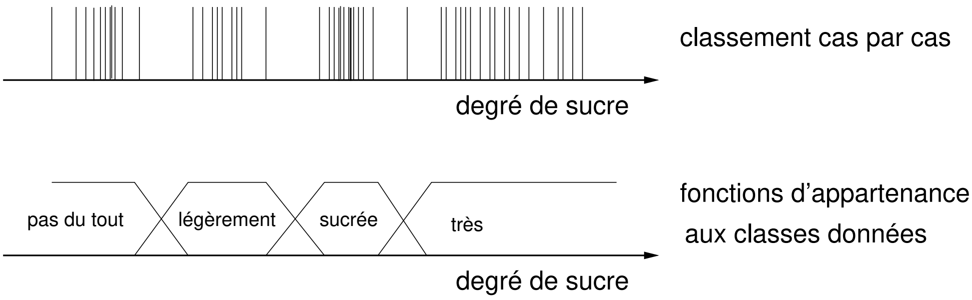


FIG. 2.5 – Adaptation des fonctions d'appartenance aux données

2.5 Apprentissage dans un contexte variable

Pour illustrer l'effet de la variabilité sur la performance d'un système de classification, nous analysons une méthode simple d'apprentissage.

La figure 2.6 montre une façon de construire une partition floue pour un attribut A_1 à partir d'une classification des pixels dans deux classes C_1 et C_2. Les contours qui délimitent les *vraies* classes dans la base d'apprentissage, sont utilisés pour calculer les histogrammes de l'attribut A_1 pour les deux classes. Il y a plusieurs méthodes pour transformer la description des densités d'exemples en des fonctions d'appartenance (voir chapitre 5). Une méthode simple consiste à normaliser ces histogrammes par la somme des histogrammes des deux classes, comme c'est illustré à droite sur la figure 2.6.

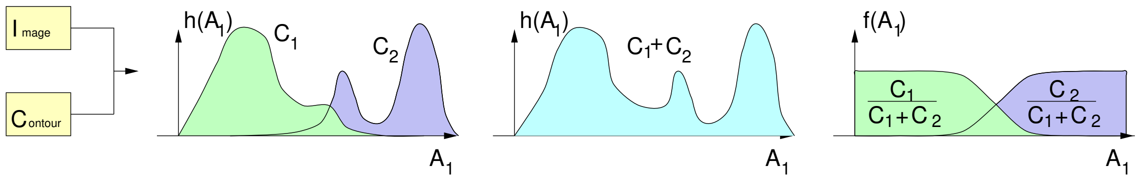


FIG. 2.6 – Construction probabiliste des fonctions d'appartenance

Moins les deux fonctions d'appartenance se recouvrent, plus l'attribut est pertinent pour la distinction des deux classes. Autrement dit, les variations de l'attribut A_1 sont principalement liées à l'appartenance à l'une des classes. Si les conditions de l'acquisition de l'image ne sont pas optimales, les attributs peuvent être soumis à de nombreuses sources de variation dues aux :

– conditions physiques de l'acquisition
– conditions de l'éclairage
– variations de positions relatives
– changement du point de vue
– variations de l'objet à analyser
– ...

Si nous essayons de construire des fonctions d'appartenance pour toute une base de données de la même manière que celle décrite précédemment, en calculant cette fois-ci l'histogramme pour tous les pixels de toutes les images ensembles, la variation entre les images s'ajoute à l'incertitude déjà présente pour les attributs calculés sur une seule image (voir figure 2.7). Le pouvoir de discrimination de l'attribut est dégradé.

Une solution au problème de cette variabilité serait de trouver des attributs qui restent constants à travers toutes les images mais qui sont différents pour les différentes classes. Dans de nombreux systèmes, les attributs sont normalisés d'une manière heuristique avec des mesures prises sur l'image entière ou sur une zone de l'image afin de réduire la sensibilité des attributs aux sources

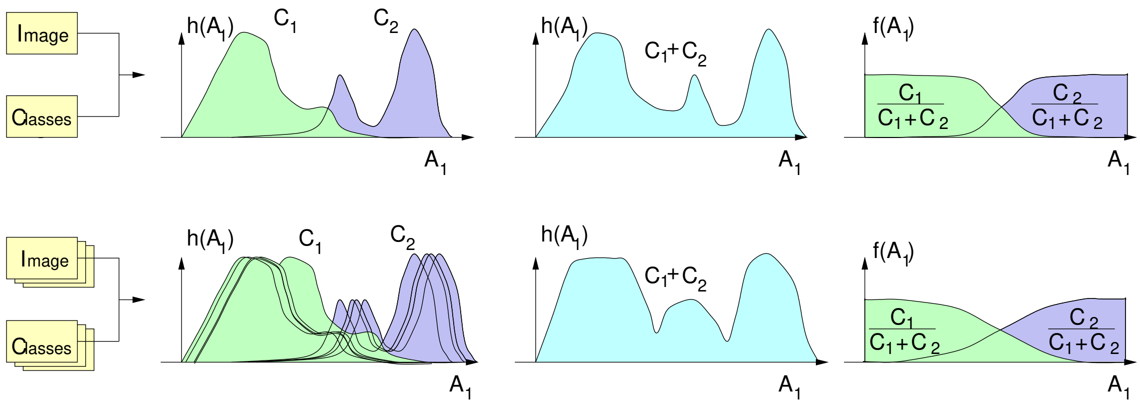


FIG. 2.7 – Classification dans un contexte variable

de variabilité. Malheureusement, cette méthode demande une connaissance explicite des sources et du comportement de la variabilité pour pouvoir l'exprimer dans un calcul de normalisation. Un exemple pour une telle approche est donné dans chapitre 10.

2.5.1 Seuils adaptatifs

Une autre approche pour arriver à une discrimination efficace à travers toute la variabilité des images, est de prendre les décisions faites sur les attributs dépendants de l'image. Dans le cas d'une modélisation des frontières non floues entre les classes, cela revient à des seuils adaptatifs, une méthode qui est couramment utilisée en traitement d'images. Le seuillage adaptatif n'applique pas le même seuil à toutes les images de la base, mais calcule le seuil en fonction de la statistique de l'attribut dans chaque image. Puisque la classification est le but du seuillage, les distributions des classes ne sont pas connues pour calculer le seuil, par contre l'histogramme des deux classes considérées ensemble (au centre de la figure 2.7) est connu. L'adaptation des seuils entre les classes peut se faire pour chaque image ou même pour une partie de l'image. Une revue des méthodes de classification avec seuils adaptatifs se trouve dans le chapitre 10 de [Coc95].

2.6 Sous-ensembles flous adaptatifs

L'inconvénient des seuils - statiques ou adaptatifs - est l'impossibilité de transmettre l'imprécision de ce seuil dans le résultat de classification. Dans des systèmes multi-critères, ceci représente un handicap majeur puisque cette imprécision n'est pas connue pendant la phase de combinaison des différents critères. Les systèmes non flous ne se prêtent donc pas très bien à la combinaison de différents critères parce qu'ils ont tendance à prendre les décisions trop tôt dans la chaîne de traitement.

Dans la suite, nous proposons une méthode généralisant ce seuil adaptatif à des fonctions d'appartenance adaptatives pour pouvoir combiner le traitement de l'incertitude proposé par la logique floue et le traitement de la variabilité.

Les méthodes adaptatives dans la logique floue sont connues depuis longtemps (1979, [Pro79]), mais elles sont principalement appliquées aux problèmes de l'automatisme. Une revue des nombreuses méthodes dans ce domaine se trouve dans le chapitre 5 de [Jag95]. Ces méthodes permettent de réadapter des contrôleurs flous à de nouvelles circonstances soit en modifiant les règles soit en modifiant la position des fonctions d'appartenance. Pour le problème de la variabilité que nous considérons ici, la deuxième approche semble plus intéressante, car elle permet de garder le même raisonnement au travers d'une base d'exemples avec variabilité. Dans les systèmes de contrôle flou adaptatifs, l'adaptation se fait soit sur une nouvelle base d'apprentissage, c'est-à-dire une base avec la vérité connue, soit par retransmission d'une mesure d'erreur acquise en continu entre la valeur de la sortie du système à régler et la consigne. Cela n'est pas applicable au problème d'adaptation des fonctions d'appartenance en classification floue dans le traitement d'images, puisqu'aucune mesure d'erreur n'est disponible quand une nouvelle image arrive.

Le système que nous cherchons à construire pour le traitement d'images doit :
- apprendre d'une manière supervisée les fonctions d'appartenance à partir d'une base d'apprentissage et
- s'adapter aux conditions d'une nouvelle image en utilisant le contexte dans l'image (adaptation non-supervisée)

Pour réussir cette tâche, nous proposons de calculer les fonctions d'appartenance pour les classes à partir de la fonction d'appartenance de la somme de toutes les classes et d'un modèle développé pendant la phase d'apprentissage. Evidemment, cela doit se faire avant la normalisation des fonctions d'appartenance car l'information que nous recherchons à extraire est cachée dans des densités d'exemples sur l'espace des attributs plus ou moins élevées. La méthode que nous proposons consiste en une phase d'apprentissage et une phase d'utilisation (figure 2.8).

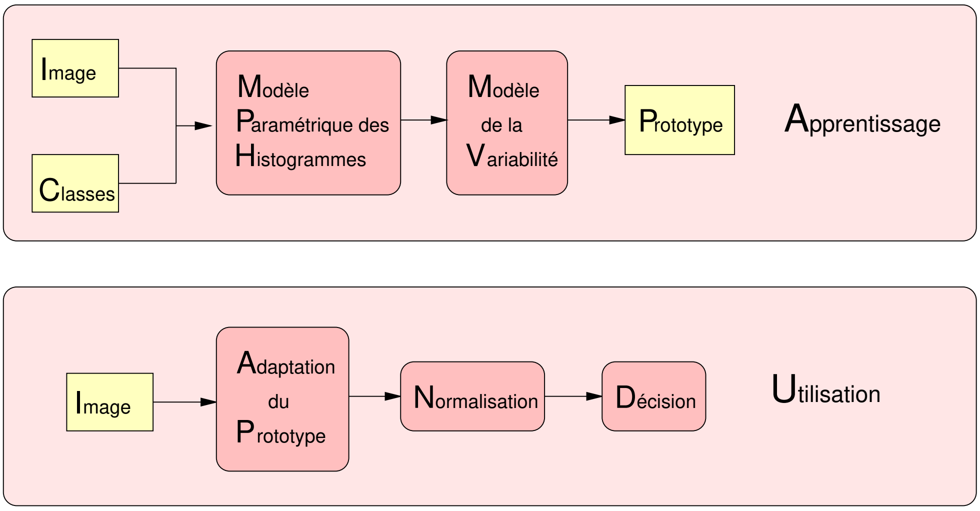


FIG. 2.8 – Sous-ensembles flous adaptatifs

2.6.1 La phase d'apprentissage

La phase d'apprentissage se déroule selon les étapes suivantes :
- modélisation des densités dans les histogrammes pour chaque attribut, chaque classe et chaque image de la base d'apprentissage
- création d'un prototype qui modélise les liens entre les descriptions des différentes classes.

Dans le chapitre 3, nous allons regarder comment il est possible de décrire les densités des exemples avec un modèle paramétrique reposant sur des sommes de gaussiennes. La méthode est généralisée à d'autres formes de composantes.

Dans le chapitre 4, nous abordons le problème d'estimation de la variabilité dans la base d'apprentissage. D'abord, les composantes de la description paramétrique sont regroupées pour retrouver le même ordre des composantes dans chaque image ; ensuite, un prototype est construit, qui représente le comportement commun à toutes les images de la base. Les variations des exemples sont estimées et les liens entre les composantes de la description sont estimés par une régression.

2.6.2 La phase d'utilisation

La phase d'utilisation se déroule selon les étapes suivantes :

– l'adaptation du prototype à l'image actuelle

– une normalisation de la description des densités pour obtenir des fonctions d'appartenance.

L'adaptation du prototype à l'image courante est décrite dans la section 4.4.6. L'algorithme utilisé est le même que celui utilisé pendant la phase d'apprentissage.

Dans le chapitre 5, les méthodes de normalisation qui permettent de passer d'une description de l'histogramme à des fonctions d'appartenance sont décrites. Une interprétation de cette normalisation dans le cadre de l'extraction de l'information à partir des annotations (par des contours) d'un expert est donnée.

Finalement, le chapitre 6 conclut la première partie en montrant l'application concrète de la méthode proposée à une base de données artificielle.

Chapitre 3

Paramétrage des histogrammes

3.1 Présentation du problème

Comme nous l'avons évoqué précédemment, notre but est de faire une modélisation et une gestion rigoureuses de l'incertitude à travers toutes les étapes de notre algorithme de classification. Nous nous basons sur la théorie des sous-ensembles flous pour décrire les valeurs avec leurs incertitudes.

La première étape de notre algorithme est d'estimer l'incertitude qui est apportée dans notre système à partir de la base d'apprentissage. Pour cela, il faut construire les fonctions d'appartenance pour les attributs provenant des exemples de notre base de données.

Les sources d'incertitude inhérentes à un système de classification sont :

- l'erreur de mesure
- la précision de la mesure
- la variabilité des classes à détecter
- le nombre limité d'exemples

Tout d'abord, les attributs mesurés sur les images incluent une erreur de mesure et une imprécision liées à la résolution de la mesure. Par exemple, les valeurs de niveau de gris dans l'image sont codées avec 8, 12 ou 16 bits et le bruit du capteur peut introduire une incertitude de l'ordre de plusieurs bits.

Les phénomènes à modéliser eux-mêmes ont une variabilité interne qui ne peut pas être complètement modélisée à partir d'une base d'apprentissage limitée.

Pour construire des fonctions d'appartenance à partir d'une base d'exemples, il faut franchir la frontière entre une représentation purement fréquentielle, donnée par les exemples de la base, et une représentation par fonctions d'appartenance qui sont définies sur tout l'univers de l'espace des attributs. Cette étape doit donc réaliser une agrégation entre les valeurs des différents exemples pour modéliser une connaissance sur le processus dont les exemples sont tirés.

Le deuxième but du paramétrage des histogrammes est la simplification de ceux-ci jusqu'à ce que les points communs entre les différentes images de l'application en question deviennent mesurables. Si l'histogramme d'un attribut, qui est utilisé pour la classification, est divisé en plusieurs milliers de cases sur l'abscisse, cette représentation n'est pas très adaptée pour trouver des régularités. Le but de la paramétrisation est donc de réduire le nombre de paramètres nécessaires pour décrire la distribution des exemples dans l'espace des attributs, sans pourtant négliger les variations qui peuvent exister entre les histogrammes des différentes images.

La méthode de paramétrisation que nous proposons d'utiliser est d'approcher les histogrammes par une somme de gaussiennes. Nous allons détailler notre méthode dans ce qui suit.

3.2 Modélisation des histogrammes par mélange de gaussiennes

Une des méthodes couramment utilisée en statistique pour la description des distributions est l'approchement par un mélange. Dans ce cas, une distribution inconnue est estimée par une somme de distributions paramétriques c_k :

$$c(x, \Phi) = \sum_{k=1}^{K} p_k \cdot c_k(x, \Theta_k) \tag{3.1}$$

avec :

$$\Phi = (\vec{p}, \vec{\Theta}) \tag{3.2}$$

Le but de l'apprentissage est alors de déterminer à la fois les paramètres Θ_k de chaque distribution et leurs poids p_k respectifs.

3.2.1 Mélanges de gaussiennes

Le mélange le plus utilisé est le mélange de distributions gaussiennes $c_k = g_k(x, \mu_k, \sigma_k)$.

$$c(x|\Phi) = \sum_{k=1}^{K} p_k \cdot g_k(x, \mu_k, \sigma_k) \tag{3.3}$$

avec :

$$g_k(x, \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2} \tag{3.4}$$

et

μ_k moyenne de la gaussienne k

σ_k écart-type de la gaussienne k

p_k poids (ou amplitude) de la gaussienne k

Un exemple d'approximation d'une distribution d'exemples par une somme de trois gaussiennes est donné en figure 3.1.

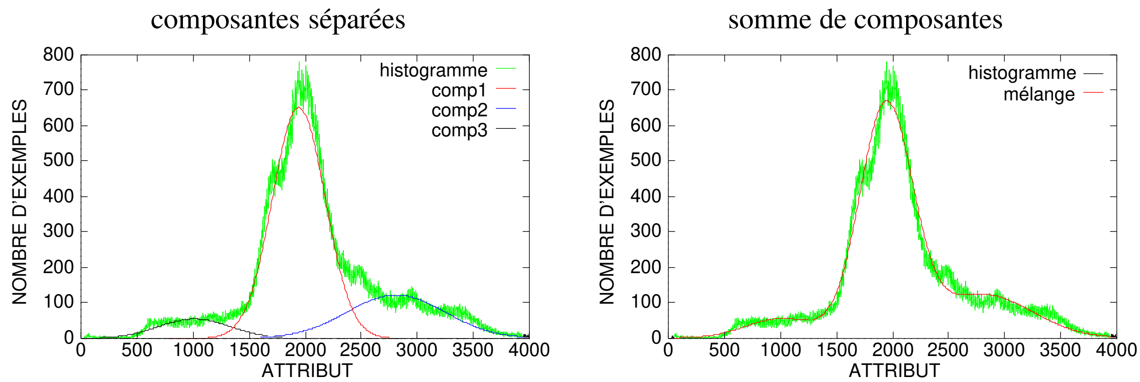


FIG. 3.1 – Approximation d'une distribution par une somme de gaussiennes

3.2.2 Identification des mélanges de gaussiennes

Les mélanges de gaussiennes ont la capacité de permettre l'estimation de n'importe quelle distribution avec une erreur ε choisie. Selon la distribution de départ et l'erreur acceptée, le nombre de gaussiennes nécessaires est plus ou moins important. Il est donc possible d'utiliser les mélanges de gaussiennes pour représenter n'importe quelle distribution de densité d'exemples avec une précision que l'on peut contrôler par le nombre de gaussiennes utilisées. Une grande précision nécessitera l'utilisation d'un grand nombre de composantes, tandis que l'on pourra réduire le nombre de composantes dans le mélange pour modéliser seulement le comportement principal de la fonction.

3.2.3 L'idée de l'algorithme EM

Si le nombre des composantes dans un mélange est connu, l'algorithme *EM* (Expectation Maximation) proposé par Dempster [Dem77] permet d'approcher une fonction à partir d'une initialisation.

Le problème de l'estimation des paramètres des composantes dans les mélanges se base sur le principe de maximisation de la vraisemblance (maximum likelihood) : c'est-à-dire que les paramètres des composantes sont choisis de manière à maximiser la probabilité des exemples d'appartenir au mélange.

Connaissant les paramètres du mélange Φ, la probabilité de l'exemple i de la base d'apprentissage est donnée par :

$$P(x_i|\Phi) = c(x_i|\Phi) = \sum_{k=1}^{K} p_k \cdot \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2} \tag{3.5}$$

Si les probabilités sont considérées indépendantes pour les N exemples, la probabilité globale de la base est donnée par le produit des probabilités de tous les exemples :

$$P(\vec{x}|\Phi) = \prod_{i=1}^{N} \left(\sum_{k=1}^{K} p_k \cdot \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2} \right) \tag{3.6}$$

Malheureusement, une solution directe par maximisation n'est pas envisageable car on est en face de deux niveaux d'incertitude :

- l'incertitude sur l'appartenance d'un exemple donné à chaque composante dans le mélange.
- l'incertitude sur les paramètres des composantes.

L'algorithme EM fonctionne en décomposant le problème en deux étapes nommées étape **E** et étape **M**. Dans l'étape **E**, il calcule une valeur λ_{ki} qui exprime pour chaque exemple i de la base, avec quel degré il correspond à chaque composante k du mélange. Dans l'étape **M**, les λ_{ki} sont considérés constants et les paramètres de chaque composante peuvent être calculés séparément en pondérant chaque élément i de la base par son degré d'appartenance λ_{ki} à la composante k. Cela est visualisé sur la figure 3.2 : le mélange de l'itération précédente (3.2 a) est normalisé pour fournir les deux fonctions de pondération (3.2 b) ; celles-ci seront multipliées par l'histogramme (3.2 c) pour obtenir les histogrammes pour chaque composante (3.2 d). Ensuite ces histogrammes pondérés sont utilisés pour calculer une nouvelle estimation pour les composantes (3.2 a).
Pour l'estimation des paramètres d'une composante, il reste donc à maximiser :

$$P_k(\vec{x}|\Phi,\vec{\lambda_k}) = \prod_{i=1}^{N} \left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{\lambda_{ki}}{2\sum_{i=1}^{N}\lambda_{ki}}\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2} \right) \tag{3.7}$$

Maximiser $P_k(\vec{x}|\Phi,\vec{\lambda_k})$ est équivalent à maximiser son logarithme $\ln(P_k(\vec{x}|\Phi,\vec{\lambda_k}))$:

$$\begin{aligned} \ln(P_k(\vec{x}|\Phi,\vec{\lambda_k})) &= \sum_{i=1}^{N} \ln(\frac{1}{\sqrt{2\pi}\sigma_k}) - \sum_{i=1}^{N} \frac{\lambda_{ki}}{2\sum_{i=1}^{N}\lambda_{ki}} \left(\frac{x_i-\mu_k}{\sigma_k}\right)^2 && (3.8) \\ &= \sum_{i=1}^{N} \ln(\frac{1}{\sqrt{2\pi}\sigma_k}) - \frac{\sum_{i=1}^{N}\lambda_{ki}(x_i-\mu_k)^2}{2\sigma_k^2\sum_{i=1}^{N}\lambda_{ki}} && (3.9) \end{aligned}$$

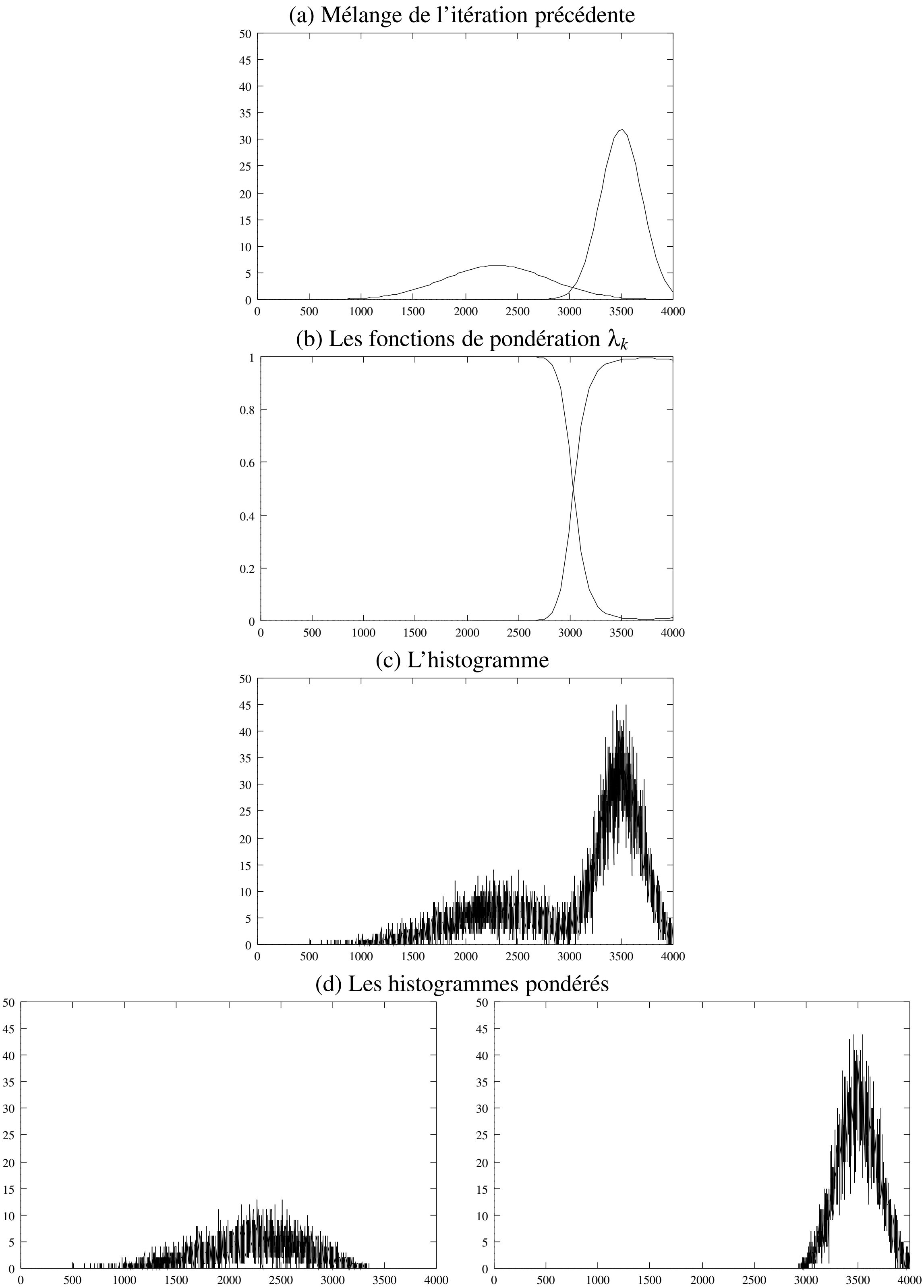


FIG. 3.2 – Décomposition de la maximisation par étapes : le mélange de l'itération précédente (a) est normalisé pour fournir les deux fonctions de pondération (b) qui sont multipliées par l'histogramme (c) pour obtenir les histogrammes pour chaque composante (d) qui sont utilisés pour calculer une nouvelle estimation pour les composantes (a)

Le maximum relativement à μ_k est en une position de dérivée partielle en μ_k nulle :

$$\frac{\partial}{\partial \mu_k} \ln(P_k(\vec{x}|\Phi, \vec{\lambda_k})) = 0 \tag{3.10}$$

$$\Rightarrow \quad -\sum_{i=1}^{N} \frac{\lambda_{ki} \cdot (x_i - \mu_k)}{\sum_{i=1}^{N} \lambda_{ki} \cdot \sigma_k^2} = 0 \tag{3.11}$$

$$\Rightarrow \quad \sum_{i=1}^{N} \lambda_{ki} \cdot x_i = \sum_{i=1}^{N} \lambda_{ki} \cdot \mu_k \tag{3.12}$$

$$\Rightarrow \quad \mu_k = \frac{\sum_{i=1}^{N} \lambda_{ki} \cdot x_i}{\sum_{i=1}^{N} \lambda_{ki}} \tag{3.13}$$

De la même manière, le maximum relativement à σ_k est en une position de dérivée partielle en σ_k nulle :

$$\frac{\partial}{\partial \sigma_k} \ln(P_k(\vec{x}|\Phi, \vec{\lambda_k})) = 0 \tag{3.14}$$

$$\Rightarrow \quad -\frac{1}{\sigma_k} + \sum_{i=1}^{N} \frac{\lambda_{ki} \cdot (x_i - \mu_k)^2}{\sum_{i=1}^{N} \lambda_{ki} \cdot \sigma_k^3} = 0 \tag{3.15}$$

$$\Rightarrow \quad \sigma_k^2 = \frac{\sum_{i=1}^{N} \lambda_{ki} \cdot (x_i - \mu_k)^2}{\sum_{i=1}^{N} \lambda_{ki}} \tag{3.16}$$

$$\Rightarrow \quad \sigma_k = \sqrt{\frac{\sum_{i=1}^{N} \lambda_{ki} \cdot (x_i - \mu_k)^2}{\sum_{i=1}^{N} \lambda_{ki}}} \tag{3.17}$$

$$\tag{3.18}$$

La valeur p_k est simplement calculée à partir des λ_{ki} :

$$p_k = \sum_{i=1}^{N} \lambda_{ki} \tag{3.19}$$

Les formules retrouvées ici sont celles d'une estimation classique du poids p, de la valeur centrale μ et de l'écart-type σ d'une gaussienne à partir d'une série d'exemples avec la seule différence que les exemples sont pondérés selon leur valeur d'appartenance λ_{ki}.

3.2.4 L'implémention de l'algorithme EM

L'implémentation de l'algorithme EM s'effectue selon les deux étapes décrites ci-dessous :

étape E

On effectue la mise à jour des λ_{ki} lors de l'itération (m). Lors de l'initialisation de l'algorithme, on fixe les p_k, μ_k, σ_k pour calculer les valeurs initiales pour des λ_{ki} ;

$$\lambda_{ki}^{(m)} = \frac{p_k^{(m-1)} \cdot g(x_i, \mu_k^{(m-1)}, \sigma_k^{(m-1)})}{\sum_{j=1}^{K} p_k^{(m-1)} \cdot g(x_i, \mu_j^{(m-1)}, \sigma_j^{(m-1)})} \tag{3.20}$$

Ensuite les mélanges eux-mêmes sont mis à jour à partir des exemples pondérés par les λ_{ki} :

étape M

(mise à jour des p_k,μ_k,σ_k)

$$p_k^{(m)} = \sum_{i=0}^{N} \lambda_{ki}^{(m)} \tag{3.21}$$

$$\mu_k^{(m)} = \frac{\sum_{i=0}^{N} \lambda_{ki}^{(m)} x_i}{p_k^{(m)}} \tag{3.22}$$

$$\sigma_k^{(m)} = \sqrt{\frac{\sum_{i=0}^{N} \lambda_{ki}^{(m)} (x_i - \mu_k^{(m)})^2}{p_k^{(m)}}} \tag{3.23}$$

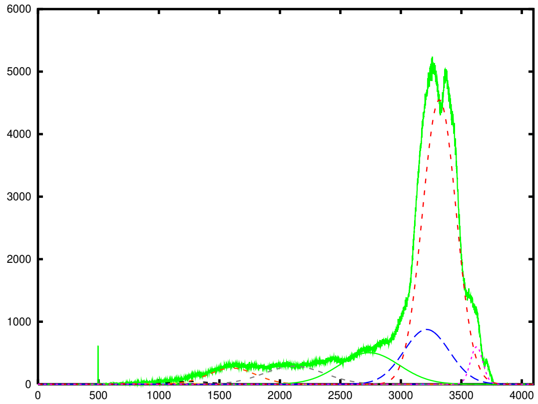
avec :

x_i exemples de la base d'apprentissage

λ_{ki} appartenance de l'exemple i à la gaussienne k

3.2.5 Les résultats de l'algorithme EM

L'algorithme EM donne de bons résultats si l'initialisation de départ est déjà assez proche de la solution optimale. Sinon, il trouve un minimum local proche de l'état initial $(p^{(0)},\mu^{(0)},\sigma^{(0)})$. Le nombre de gaussiennes choisies a une influence majeure sur la répartition des gaussiennes, comme le montre la figure 3.3 qui donne le résultat de l'approximation d'un histogramme par 8 et par 12 composantes. On peut observer que le mode principal de l'histogramme est presque entièrement décrit par une seule composante dans le cas d'initialisation avec 8 composantes et que le même mode est décrit par deux composantes lorsque le nombre de composantes augmente.



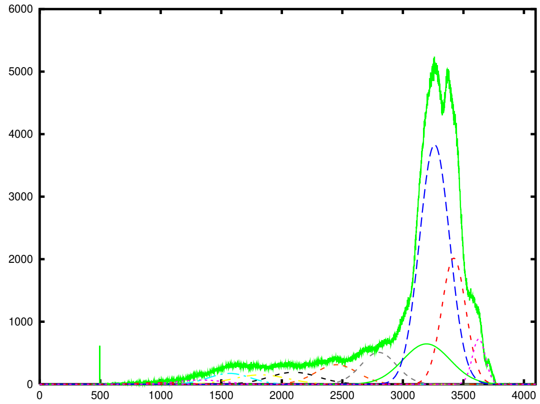


FIG. 3.3 – Approximation d'une distribution par une somme de gaussiennes utilisant 8 ou 12 composants. La façon d'approcher le mode principal change significativement avec le nombre de composantes choisi.

Nous allons utiliser les mélanges de gaussiennes dans deux buts : comme représentation d'une distribution d'exemples dans l'espace des attributs et pour l'identification des principaux modes de ces distributions. En ce qui concerne la représentation, il est acceptable d'avoir des composantes différentes pour des distributions similaires ; par contre concernant l'identification, il est souhaitable de trouver des composantes stables qui varient peu avec les variations des distributions. C'est pourquoi nous avons cherché des algorithmes d'estimation des paramètres qui donnent des résultats plus stables.

3.2.6 Identification hiérarchique des gaussiennes

Dans [Zhu96], un algorithme hiérarchique pour l'identification des gaussiennes est proposé. A chaque itération, il essaie de trouver une gaussienne qui, ajoutée à un bruit de niveau t_s constant, approche la fonction. L'idée de cette approche est que les composantes de l'histogramme qui ne sont pas encore identifiées soient représentées par un modèle de bruit constant à travers de l'univers.

On peut alors utiliser une version simplifiée de l'algorithme EM avec une seule gaussienne :

étape E

$$\lambda_i^{(m)} = \frac{g(x_i, \mu^{(m-1)}, \sigma^{(m-1)})}{g(x_i, \mu^{(m-1)}, \sigma^{(m-1)}) + t_s} \tag{3.24}$$

étape M

$$p^{(m)} = \sum_{i=0}^{N} \lambda_i^{(m)} \tag{3.25}$$

$$\mu^{(m)} = \frac{\sum_{i=0}^{N} \lambda_i^{(m)} x_i}{p^{(m)}} \tag{3.26}$$

$$\sigma^{(m)} = \sqrt{\frac{\sum_{i=0}^{N} \lambda_i^{(m)} (x_i - \mu^{(m)})^2}{p^{(m)}}} \tag{3.27}$$

Puisque le niveau de bruit n'est pas vraiment connu, car il s'agit d'un modèle pour toute la partie de la fonction qui ne va pas être approchée par la prochaine gaussienne, plusieurs valeurs t_s sont utilisées. Le choix entre les résultats obtenus avec les différents t_s est effectué par un test d'adaptation du type Kolmogorov-Smirnov. La figure 3.4 montre les résultats obtenus pour deux niveaux de bruit différents. Pour le niveau $t_s = 0$ (figure 3.4 à gauche), l'histogramme entier est approché par une gaussienne.

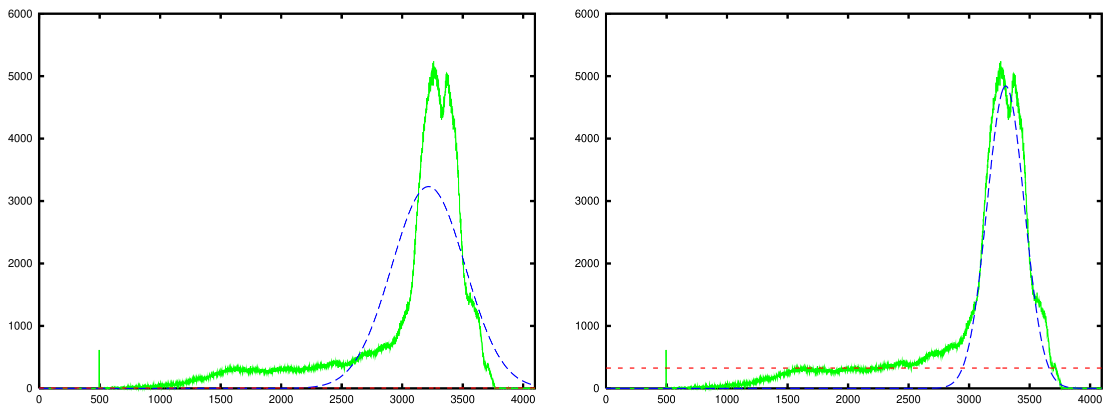


FIG. 3.4 – Résultat pour deux valeurs t_s différents.
à gauche : $t_s = 0$ - l'histogramme entier est modélisé par la gaussienne
à droite : $t_s \neq 0$ - seulement une partie de l'histogramme est modélisée par la gaussienne

Puisque l'histogramme ne ressemble pas dans son ensemble à une gaussienne, cette approximation ne décrit aucune partie de l'histogramme d'une manière satisfaisante. Pour un niveau t_s plus élevé (figure 3.4 à droite) seul la partie de l'histogramme qui est significativement au-dessus du niveau t_s (indiqué par le trait horizontal) est prise en compte pour l'estimation des paramètres de la gaussienne. Logiquement, la gaussienne ne décrit pas l'ensemble de l'histogramme, mais pour le mode le plus important, le résultat est très intéressant.

Une fois que la meilleure gaussienne a été identifiée, elle est soustraite de la fonction initiale et la recherche de la gaussienne suivante est effectuée de la même manière.

Au cours des itérations, l'histogramme est de mieux en mieux approché par la somme des gaussiennes et les gaussiennes trouvées sont de plus en plus petites. De même, le niveau du bruit t_s qui permet de trouver une gaussienne intéressante diminue avec le nombre d'itérations.

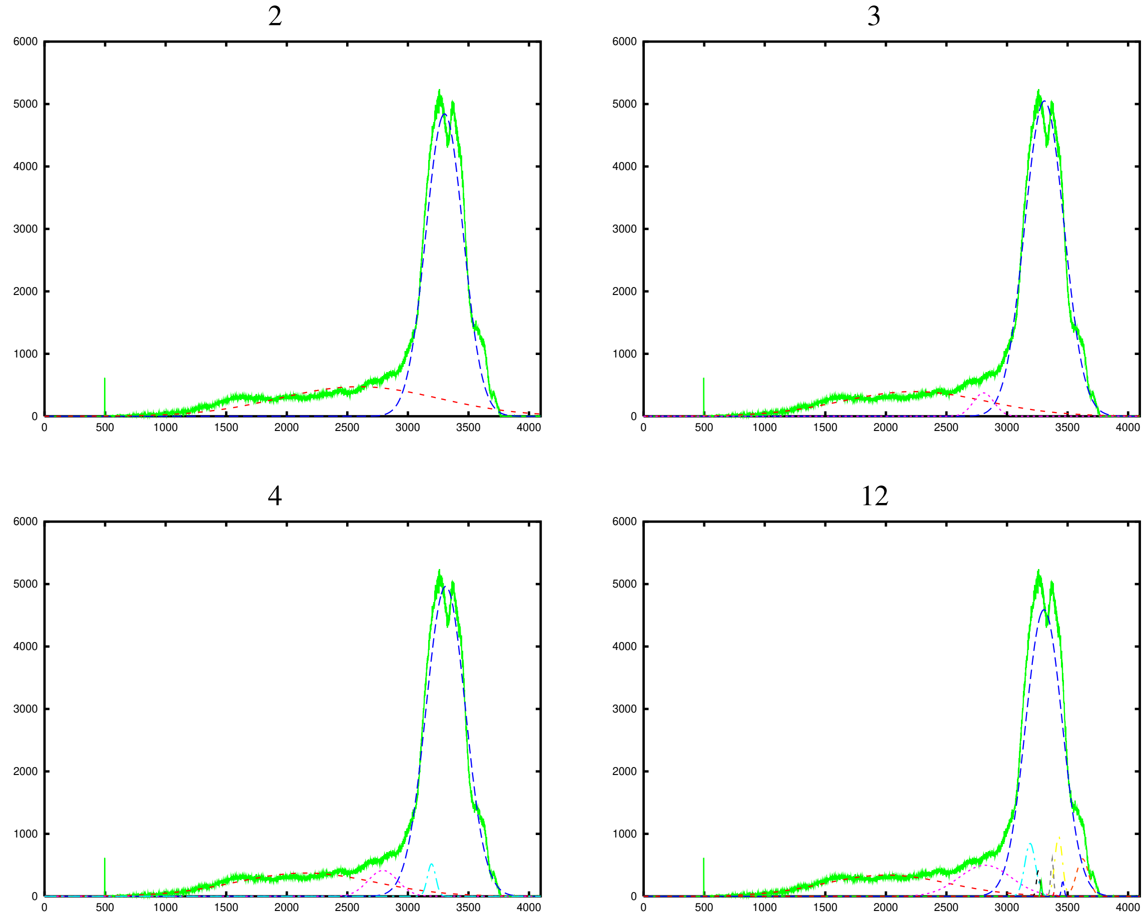


FIG. 3.5 – Résultats des mélanges avec 2,3,4 ou 12 composantes

La figure 3.5 montre les résultats pour une approximation avec deux, trois, quatre et douze gaussiennes. Les principales gaussiennes sont reconnues indépendamment du nombre de composantes. La décomposition hiérarchique permet donc de décomposer une distribution d'une façon presque indépendante du nombre de gaussiennes.

3.3 Une nouvelle approche pour l'identification hiérarchique

Même si l'identification hiérarchique donne des résultats satisfaisants pour la plupart des histogrammes, elle a une faiblesse quand les composantes à estimer sont de tailles très différentes. Cela vient du fait qu'à chaque fois qu'une composante a été estimée, sa distribution d'exemples est soustraite de la distribution d'exemples présents dans l'histogramme. Un premier effet non souhaitable est que l'on peut obtenir des valeurs négatives dans l'histogramme, dues aux fluctuations locales à l'intérieur d'un mode.

Dans des cas extrêmes - si aucune précaution n'est prise - cet algorithme peut arriver à expliquer ces valeurs négatives par des composantes avec des poids négatifs. Le deuxième effet défavorable est le comportement du bruit dans les zones où une composante avec un poids important a été soustraite lors des itérations précédentes, puisque les variations locales d'une composante importante sont plus importantes que les variations dans des zones avec peu d'exemples.

3.3.1 La méthode

Pour remédier à ces problèmes, nous introduisons une nouvelle méthode d'estimation hiérarchique des paramètres à partir de la méthode précédente. Cette méthode prend en compte les gaussiennes déjà identifiées au cours de la décomposition hiérarchique dans le modèle du bruit.

Si le nombre d'exemples n à un endroit sur l'histogramme suit un loi de Poisson, les fluctuations autour de cette valeur nominale de n exemples sont de l'ordre de $\sqrt{n}$. Il est donc raisonnable d'accepter une erreur d'approximation (en absolu) plus grande dans les zones avec une grande densité d'exemples. Au lieu de soustraire les gaussiennes identifiées, il est possible de les intégrer dans le bruit et de cette manière de prendre en compte la variabilité introduite par une composante importante. L'étape E est donc remplacée par :

étape E

$$\lambda_{ki}^{(m)} = \frac{p_k^{(m-1)} \cdot g(x_i, \mu_k^{(m-1)}, \sigma_k^{(m-1)})}{\sum_{j=0}^{K} p_j^{(m-1)} \cdot g(x_i, \mu_j^{(m-1)}, \sigma_j^{(m-1)}) + t_s} \tag{3.28}$$

étape M

(non modifiée par rapport à l'estimation hiérarchique décrite précédemment)

$$p_k^{(m)} = \sum_{i=0}^{N} \lambda_{ki}^{(m)} \tag{3.29}$$

$$\mu_k^{(m)} = \frac{\sum_{i=0}^{N} \lambda_{ki}^{(m)} x_i}{p_k^{(m)}} \tag{3.30}$$

$$\sigma_k^{(m)} = \sqrt{\frac{\sum_{i=0}^{N} \lambda_{ki}^{(m)} (x_i - \mu_k^{(m)})^2}{p_k^{(m)}}} \tag{3.31}$$

Pour chaque niveau de la hiérarchie, cette formulation ressemble fortement à celle de l'algorithme EM standard : les gaussiennes déjà identifiées sont prises en compte comme dans l'algorithme EM standard mais une seule gaussienne est mise à jour dans l'étape M et les gaussiennes non encore identifiées sont approchées par un bruit homogène représenté par t_s.

Ce nouvel algorithme est aussi robuste que l'algorithme hiérachique proposé dans [Zhu96] décrit précédemment, mais la convergence est encore plus rapide puisque le problème des valeurs négatives dans l'histogramme est évité.

Dans la suite, nous avons dénommé cet algorithme *H-EM*.

3.3.2 Généralisation à d'autres formes de composantes

Le modèle paramétrique qui a été choisi jusqu'à présent est la somme de composantes gaussiennes. Ce modèle est très couramment utilisé pour deux raisons :

- il permet de décrire parfaitement les processus où les valeurs à mesurer sont au départ des valeurs non floues auxquelles un bruit gaussien est ajouté. Puisque la plupart des sources de bruit sont gaussiennes (bruit électronique des capteurs, bruit quantique) ce modèle est applicable à beaucoup de problèmes.
- les bases mathématiques pour estimer les mélanges des gaussiennes sont connues depuis longtemps.

Néanmoins, les gaussiennes ne sont pas les seules fonctions paramétriques qui peuvent être utilisées pour modéliser un histogramme et nous allons illustrer dans cette section comment l'algorithme H-EM peut les utiliser pour réaliser l'approximation.

Le raisonnement appliqué aux composantes gaussiennes peut être généralisé sur d'autres composantes $c_k(x,\mu,\sigma)$ en transformant la gaussienne par une exponentielle d'une fonction f monotone et décroissante de la manière suivante :

$$c_k(x,\mu_k,\sigma_k) = \frac{1}{\gamma}\exp\left(-f\left(\left(\frac{x-\mu_k}{\sigma_k}\right)^2\right)\right) \tag{3.32}$$

avec :

μ_k - la valeur centrale de la composante k

σ_k - la largeur de la composante k

p_k - le poids de la composante k

$\gamma = \int_{-\infty}^{\infty}\exp\left(-f\left(\left(\frac{x-\mu_k}{\sigma_k}\right)^2\right)\right)dx$ - facteur de normalisation

La distribution de probabilité de chaque exemple de la base d'apprentissage est donnée par le mélange :

$$P(x_i|\Phi) = c(x_i|\Phi) = \sum_{k=1}^{K} p_k \cdot \frac{1}{\gamma}\exp\left(-f\left(\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right)\right) \tag{3.33}$$

et la distribution de probabilité globale de la base est donc le produit des distributions de probabilité de tous les exemples

$$P(\vec{x}|\Phi) = \prod_{i=1}^{N}\left[\sum_{k=1}^{K} p_k \cdot \frac{1}{\gamma}\exp\left(-f\left(\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right)\right)\right] \tag{3.34}$$

Nous utilisons la même approche que dans la section 3.2.3 pour décomposer le problème en deux étapes. Il nous reste donc pour l'étape **M** à maximiser :

$$P_k(\vec{x}|\Phi,\vec{\lambda}_k) = \prod_{i=1}^{N}\left[\frac{1}{\gamma}\exp\left(-\frac{\lambda_{ki}}{\sum_{i=1}^{N}\lambda_{ki}} f\left(\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right)\right)\right] \tag{3.35}$$

et nous prenons le logarithme pour simplifier :

$$\ln(P_k(\vec{x}|\Phi,\vec{\lambda}_k)) = \sum_{i=1}^{N}\ln(\frac{1}{\gamma}) - \sum_{i=1}^{N}\frac{\lambda_{ki}}{\sum_{i=1}^{N}\lambda_{ki}} f\left(\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right) \tag{3.36}$$

Pour trouver le μ_k pour lequel $\ln(P_k(\vec{x}|\Phi,\vec{\lambda}_k))$ est maximal, la dérivée partielle relativement à μ_k est calculée et annulée :

$$\frac{\partial}{\partial\mu_k}\ln(P_k(\vec{x}|\Phi,\vec{\lambda}_k)) = 0 \tag{3.37}$$

$$\Rightarrow \quad -\sum_{i=1}^{N} \frac{\lambda_{ki} \cdot f'\left(\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right)(x_i-\mu_k)}{\sum_{i=1}^{N} \lambda_{ki} \cdot \sigma_k^2} = 0 \tag{3.38}$$

$$\Rightarrow \quad \sum_{i=1}^{N} \lambda_{ki} f'\left(\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right) \cdot x_i = \sum_{i=1}^{N} \lambda_{ki} f'\left(\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right) \cdot \mu_k \tag{3.39}$$

Avec une définition de β_{ki} pour simplifier :

$$\beta_{ki} = f'\left(\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right) \tag{3.40}$$

nous obtenons :

$$\mu_k = \frac{\sum_{i=1}^{N} \lambda_{ki}\beta_{ki} \cdot x_i}{\sum_{i=1}^{N} \lambda_{ki}\beta_{ki}} \tag{3.41}$$

La même démarche est appliquée pour σ_k :

$$\frac{\partial}{\partial \sigma_k} \ln(P_k(\vec{x}|\Phi, \vec{\lambda}_k)) = 0 \tag{3.42}$$

$$\Rightarrow \quad \frac{1}{\sigma_k} - \sum_{i=1}^{N} \frac{\lambda_{ki} \cdot f'\left(\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right)(x_i-\mu_k)^2}{\sum_{i=1}^{N} \lambda_{ki} \cdot \sigma_k^3} = 0 \tag{3.43}$$

$$\Rightarrow \quad \sigma_k^2 = \frac{\sum_{i=1}^{N} \lambda_{ki}\beta_{ki} \cdot (x_i-\mu_k)^2}{\sum_{i=1}^{N} \lambda_{ki}\beta_{ki}} \tag{3.44}$$

$$\Rightarrow \quad \sigma_k = \sqrt{\frac{\sum_{i=1}^{N} \lambda_{ki}\beta_{ki} \cdot (x_i-\mu_k)^2}{\sum_{i=1}^{N} \lambda_{ki}\beta_{ki}}} \tag{3.45}$$

Cette introduction de la fonction f dans la composante correspond à un changement d'échelle sur l'abscisse. Dans le cas d'une gaussienne, la pondération des exemples est constante sur tout l'univers, mais après le changement d'échelle cette pondération β_{ki} change comme la dérivée de la fonction de changement d'échelle. Dans le cadre de l'algorithme de la famille EM, cette pondération s'ajoute à la pondération par les λ_{ki} provenant du chevauchement de différentes composantes.

3.3.3 Impulsion cosinus

Pour illustrer le fonctionnement avec une autre forme de composantes, nous allons regarder une impulsion de cosinus. Une impulsion de cosinus (figure 3.6) est une période de cosinus centrée autour de zéro. Dans la proximité de zéro elle est très similaire à une gaussienne, le support n'étant pas l'infini mais l'intervalle $]-1,1[$.
Elle est définie par :

$$\mathrm{cosi}(x) = \begin{cases} \frac{1}{2}(1-\cos(\pi x)) & \text{si } |x| < 1 \\ 0 & \text{sinon} \end{cases} \tag{3.46}$$

Pour transformer la gaussienne en impulsion cosinus il suffit de choisir f telle que :

$$\exp\left(-f\left(\left(\frac{x-\mu}{\sigma}\right)^2\right)\right) = \begin{cases} \frac{1}{2}(1+\cos(\pi\frac{x-\mu}{\sigma})) & \text{si } |x-\mu| < \sigma \\ 0 & \text{sinon} \end{cases} \tag{3.47}$$

Où μ est la position centrale de l'impulsion et σ est sa largeur à mi-hauteur. En posant $z = \left(\frac{x-\mu}{\sigma}\right)^2$ nous obtenons pour f :

$$f(z) = \begin{cases} -\ln\left(\frac{1}{2}(1+\cos(\pi\sqrt{z}))\right) & \text{si } |z| < 1 \\ \infty & \text{sinon} \end{cases} \tag{3.48}$$

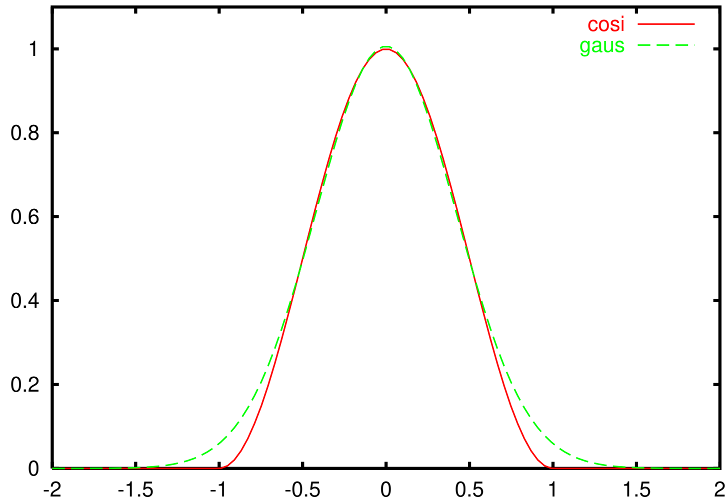


FIG. 3.6 – Impulsion de cosinus comparée à une gaussienne

La fonction de pondération en fonction de la distance normalisée par rapport au centre de la composante est :

$$\beta_k(x) = f'\left(\left(\frac{x}{\sigma_k}\right)^2\right) = \begin{cases} \frac{2\pi\sin(\pi\frac{x}{\sigma_k})}{\sigma_k\left(1+\cos(\pi\frac{x}{\sigma_k})\right)} & \text{si } |x| < \sigma_k \\ 0 & \text{sinon} \end{cases} \tag{3.49}$$

Contrairement aux gaussiennes, les impulsions de cosinus sont définies non nulles uniquement dans l'intervalle $]-1,1[$. Si ces valeurs sont interprétées comme une probabilité, les valeurs nulles posent un problème, parce que cela veut dire que théoriquement on ne doit trouver aucun exemple dans cette zone. Le facteur de pondération pour les exemples à l'extérieur du support d'une composante est donc infini. Dans l'algorithme H-EM cette problématique est cachée par la considération d'une composante avec une densité constante sur tout l'univers. Les valeurs de pondération pour chaque exemple de la base sont égales à :

$$\tilde{\lambda}_{ki} = \lambda_{ki}\beta_{ki} = \begin{cases} \frac{\frac{1}{2}(1+\cos(\pi\frac{x_i-\mu_k}{\sigma_k}))}{\frac{1}{2}(1+\cos(\pi\frac{x_i-\mu_k}{\sigma_k}))+t_s} \cdot \frac{2\pi\sin(\pi\frac{x_i-\mu_k}{\sigma_k})}{\sigma_k\left(1+\cos(\pi\frac{x_i-\mu_k}{\sigma_k})\right)} & \text{si } |x_i-\mu_k| < \sigma_k \\ 0 & \text{sinon} \end{cases} \tag{3.50}$$

$$= \begin{cases} \frac{2\pi\sin(\pi\frac{x_i-\mu_k}{\sigma_k})}{\sigma_k\left(1+\cos(\pi\frac{x_i-\mu_k}{\sigma_k})+2t_s\right)} & \text{si } |x_i-\mu_k| < \sigma_k \\ 0 & \text{sinon} \end{cases} \tag{3.51}$$

La figure 3.7 montre la pondération totale $\tilde{\lambda}_{ki}$ utilisée pour l'identification de la première composante avec l'algorithme H-EM avec différents niveaux t_s de la composante constante. La figure 3.8 montre, en référence, la pondération utilisée par l'algorithme H-EM utilisant les composantes gaussiennes.

Le fait que la fonction de pondération ait un support limité et des valeurs importantes proches des limites du support réduit la zone de convergence de l'algorithme. Les exemples qui sont en dehors du support d'une composante ne sont pas pris en compte par cette méthode d'itération, ceux au bord du support ont un poids extrême. Il est donc nécessaire de bien initialiser les composantes. Une possibilité est de commencer les itérations de l'algorithme H-EM avec les fonctions de pondération d'une gaussienne de support infini qui, pour cette raison, est plus robuste vis-à-vis de l'initialisation. Une fois que les positions des composantes ont été trouvées approximative-

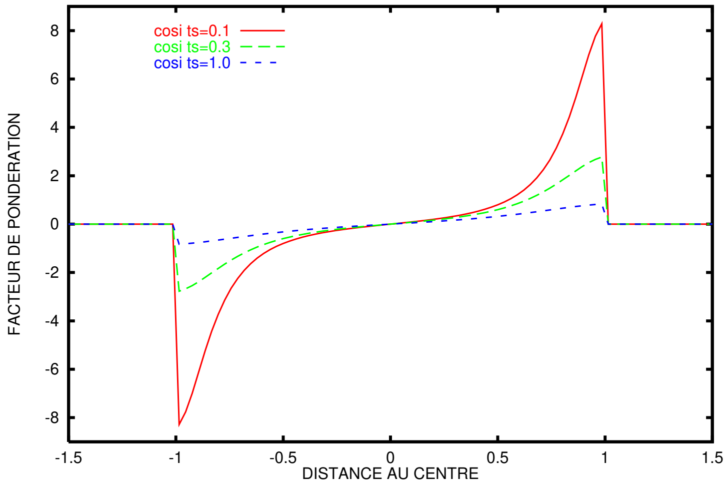


FIG. 3.7 – Fonction de pondération pour les impulsions de cosinus

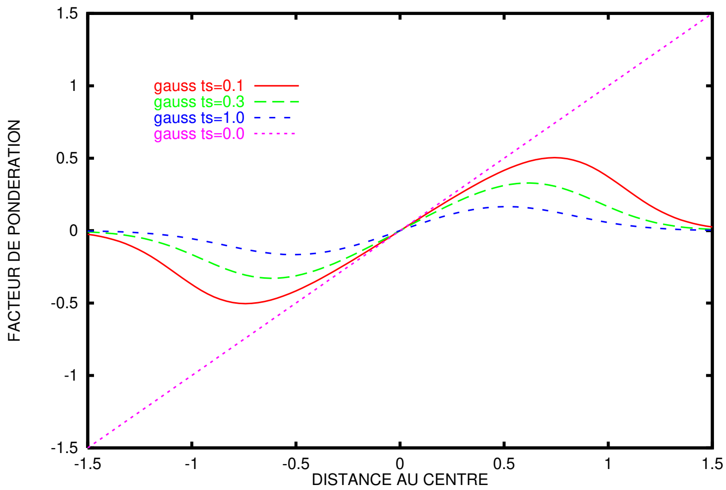


FIG. 3.8 – Fonction de pondération pour les composantes gaussiennes

ment, la pondération peut être changée pour affiner l'approximation avec les formes spécifiques des composantes choisies pour cette application.

3.3.4 Impulsion triangulaire

Les valeurs floues sont souvent représentées avec des formes triangulaires pour exprimer que la valeur d'appartenance à une catégorie diminue linéairement avec la distance. Nous pouvons décrire un triangle symétrique par :

$$\mathrm{tria}(x) = \begin{cases} 1 - |x| & \text{si } |x| < 1 \\ 0 & \text{sinon} \end{cases} \tag{3.52}$$

Pour transformer la gaussienne en impulsion triangulaire, f doit satisfaire :

$$\exp\left(-f\left(\left(\frac{x-\mu}{\sigma}\right)^2\right)\right) = \begin{cases} 1 - \left|\frac{x-\mu}{\sigma}\right| & \text{si } |x-\mu| < \sigma \\ 0 & \text{sinon} \end{cases} \tag{3.53}$$

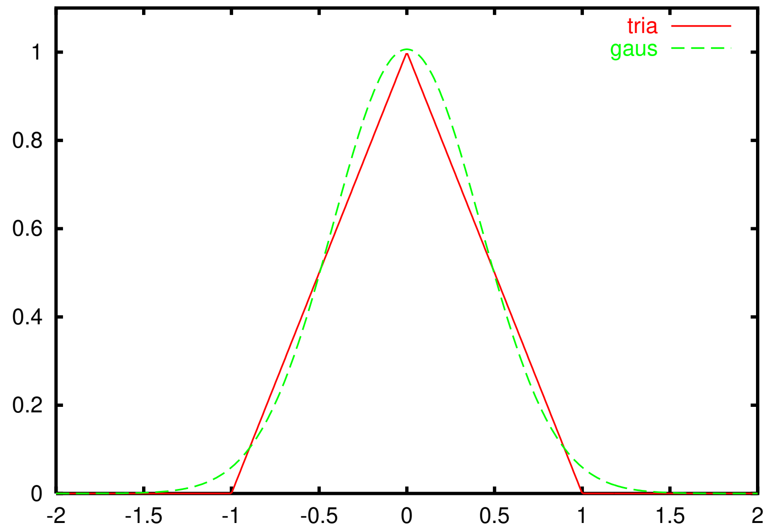


FIG. 3.9 – Impulsion triangulaire comparée à une gaussienne

En posant $z = \left(\frac{x-\mu}{\sigma}\right)^2$ nous obtenons pour f :

$$f(z) = \begin{cases} -\ln\left(1-\sqrt{|z|}\right) & \text{si } |z| < 1 \\ \infty & \text{sinon} \end{cases} \tag{3.54}$$

La fonction de pondération en fonction de la distance par rapport au centre de la composante est :

$$\beta_k(x) = f'\left(\left(\frac{x}{\sigma_k}\right)^2\right) = \begin{cases} \frac{\frac{x}{\sigma_k}}{(\sigma_k - |x|)} & \text{si } |x| < \sigma_k \\ 0 & \text{sinon} \end{cases} \tag{3.55}$$

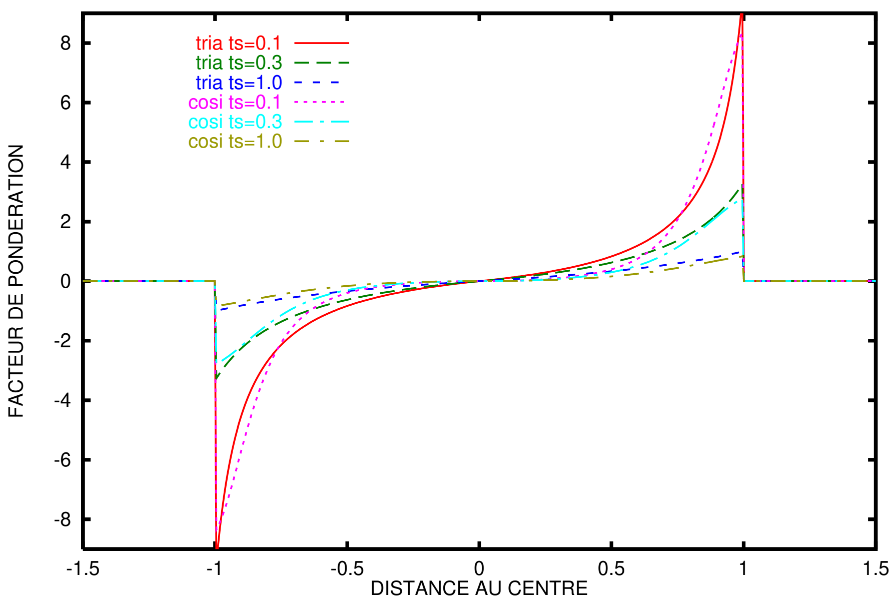


FIG. 3.10 – Fonction de pondération pour les impulsions triangulaires et les impulsions de cosinus

Comme le montre la figure 3.10, les fonctions de pondération des impulsions de cosinus et des impulsions triangulaires sont très similaires.

Cette transformation de l'axe par la fonction f peut être appliquée à n'importe quelle fonction symétrique et décroissante avec la distance par rapport à l'origine. La méthode présentée ici semble intéressante pour des applications où l'on connaît a priori les formes typiques des distributions d'exemples. En choisissant une forme adaptée pour les composantes, le nombre de composantes nécessaires pour décrire l'histogramme avec une erreur maximale donnée est réduit.

Chapitre 4

Modélisation de la variabilité

Dans le chapitre précédent, nous avons décrit une méthode pour trouver une description paramétrique des densités d'exemples dans l'histogramme d'un attribut.

Cette méthode de paramétrisation peut être utilisée pour réaliser une description compacte des distributions d'exemples dans chaque classe, qui peut elle-même être transformée en fonctions d'appartenance aux classes avec les méthodes décrites dans le chapitre 5.

Dans ce qui suit, nous montrons d'abord comment une classification statique peut être réalisée sans ou avec la paramétrisation de l'histogramme. Ensuite, nous présentons une nouvelle approche pour modéliser la variabilité dans la base d'apprentissage grâce à un prototype qui se base sur le modèle paramétrique des histogrammes. Nous proposons également un modèle des liens entre les composantes et un algorithme pour identifier ces liens. Nous terminons le chapitre avec la description d'une version modifiée de l'algorithme EM qui permet d'adapter à l'image actuelle le prototype identifié lors de la phase d'apprentissage.

4.1 Classification statique

Dans les systèmes traditionnels de classification floue, les fonctions d'appartenance sont fixées une fois pour toute pendant la phase d'apprentissage. Avant de proposer une méthode pour créer des fonctions d'appartenance adaptatives, nous décrivons le cas non adaptatif ou statique.

4.1.1 Classification statique non paramétrique

Un système d'apprentissage pour une classification statique est illustré sur la figure 4.1.

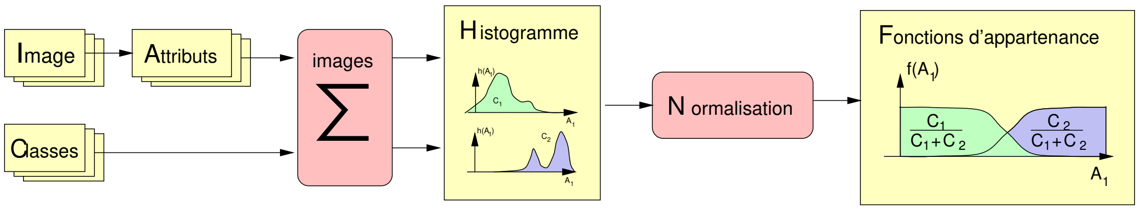


FIG. 4.1 – Classification statique

Les attributs intéressants pour la classification sont calculés pour toutes les images. Les occurrences des différentes valeurs sont calculées pour les exemples de chaque classe dans un histogramme. Les histogrammes sont combinés au travers des images pour chaque classe. Les histogrammes sont transformés en fonctions d'appartenance par une des méthodes de normalisation présentées dans le chapitre 5.

4.1.2 Classification statique paramétrique

L'utilisation de la description paramétrique des histogrammes dans un système de classification statique est illustrée sur la figure 4.2.

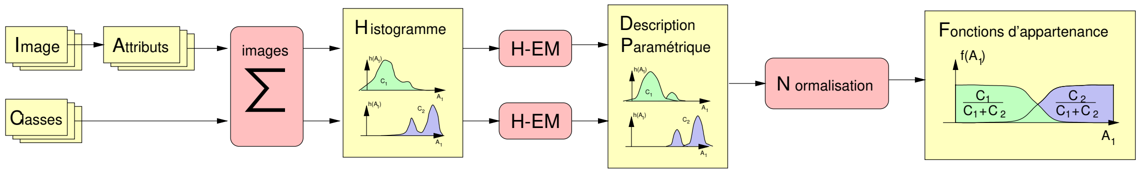


FIG. 4.2 – Classification statique avec description paramétrique

Pour simplifier la description des densités d'exemples dans les histogrammes, une description paramétrique peut être extraite en utilisant l'algorithme H-EM. Finalement, cette description paramétrique qui représente des occurrences dans l'histogramme est utilisée pour construire les fonctions d'appartenance grâce à une normalisation.

Ces fonctions d'appartenance ne sont pas dépendantes du contexte. Si les valeurs des attributs sont soumises à des variations, la performance des attributs pour la classification est réduite. Or, dans la plupart des problèmes de traitement d'images, le contexte des autres éléments dans l'image est très utile pour aider à la tâche de classification. Il convient donc d'adapter la classification au contexte pour en améliorer la performance.

4.2 Une nouvelle approche pour la classification adaptative

Il existe de nombreuses méthodes de classification non floues dans le traitement d'images qui s'adaptent au contexte dans l'image. L'exemple le plus connu est le seuillage adaptatif. Les différentes méthodes proposées dans la littérature (une revue des méthodes se trouve dans [Coc95] chapitre 10) se basent sur l'analyse de l'histogramme - soit de l'image dans sa totalité soit d'une de ses parties - pour calculer des seuils qui séparent les classes. Un inconvénient de ces méthodes non floues réside en la prise de décision prématurée avec des données imparfaites. Un autre inconvénient de la plupart des méthodes provient de la nécessité de définir a priori la façon de calculer les seuils en fonction de l'histogramme.

4.2.1 Classification floue adaptative

Dans notre approche, nous combinons l'idée d'adaptation au contexte avec l'utilisation de modèles flous pour arriver à une classification floue adaptative. Nous allons utiliser la base d'apprentissage pour estimer à la fois la position typique des classes et leur variation autour de cette valeur typique, une information nécessaire pour identifier le degré souhaitable d'adaptation au contexte.

La figure 4.3 est une illustration de cette démarche. Un prototype des densités d'exemples dans les classes est créé à partir de la base d'apprentissage. Il est adapté à l'histogramme de l'image actuelle pour essayer d'extraire la variation qui est indépendante de la classe.

Pour arriver à cette adaptation il y a deux problèmes à résoudre :
- création d'un prototype qui modélise la variabilité à partir de la base d'apprentissage
- adaptation du prototype à l'exemple actuel

Nous allons regarder ces problèmes dans ce qui suit.

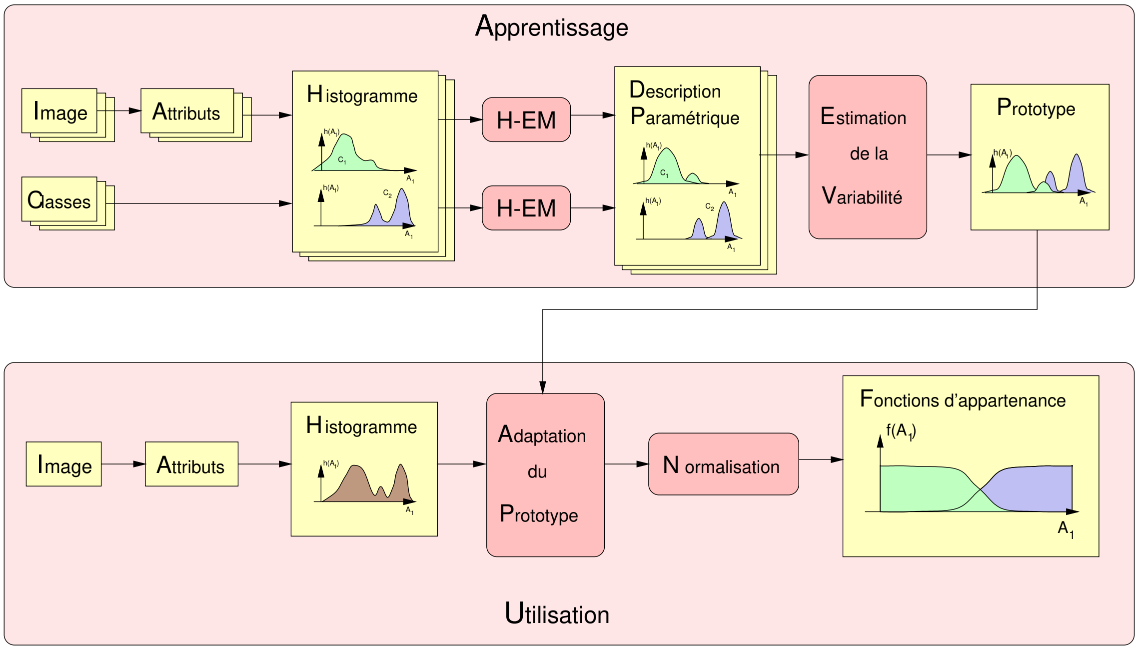


FIG. 4.3 – Classification adaptative

4.2.2 Modélisation de la variabilité

Pour identifier la variabilité existante dans le problème posé, nous utilisons la base d'apprentissage. Premièrement, une description compacte des densités des attributs est faite avec les méthodes décrites dans le chapitre précédent. Si cette description est pertinente, la variation des paramètres du modèle est représentative pour la variabilité dans la base.

Dans la pratique, les paramètres trouvés par l'algorithme H-EM ne sont pas toujours complètement représentatifs pour la variation dans la base d'apprentissage pour deux raisons :
- l'ordre dans lequel les composantes sont identifiées peut changer s'il y a deux composantes d'importance similaire.
- si le nombre de composantes choisies est trop grand, certaines composantes peuvent être utilisées pour décrire des fluctuations locales présentes uniquement dans cet exemple et ne sont pas représentatives pour la base.

Cette constatation nous amène à résoudre un deuxième problème de classification, celle des composantes pertinentes au travers de la base d'apprentissage. Ce problème de classification est un problème de classification non supervisée ou bien un problème de regroupement. Pour cette classification, nous possédons une bonne initialisation qui est donnée par l'ordre des composantes trouvées par l'algorithme H-EM. Nous allons regarder les méthodes de regroupement pour choisir une méthode adaptée à ce problème.

4.3 Algorithmes de regroupement

Les algorithmes de regroupement (*clustering* en langue anglaise) sont utilisés pour l'apprentissage non supervisé. Ils permettent de trouver des structures internes dans une base de données en regardant la distribution des exemples dans l'espace des attributs. L'hypothèse de base utilisée par ces algorithmes est que les exemples qui appartiennent à une même classe sont proches dans l'espace des attributs. Les algorithmes de regroupement se distinguent dans la façon de calculer la distance entre les exemples et une classe qui est souvent décrite par un prototype.

Les algorithmes de regroupement les plus connus sont :

les k-plus proches voisins

les C-moyennes classiques

les C-moyennes floues

les C-moyennes possibilistes

les réseaux de neurones

Il est intéressant de noter que tous ces algorithmes itératifs (sauf les réseaux de neurones) ont une structure similaire. Ils utilisent un prototype pour les classes qu'ils adaptent aux densités d'exemples par des itérations comportant deux étapes : l'estimation des regroupements pour chaque classe et la mise à jour du prototype utilisant l'information des regroupements. La figure (4.4) montre les généralités de ces algorithmes. Dans la suite nous allons formaliser les étapes pour ensuite montrer les particularités de chaque algorithme.

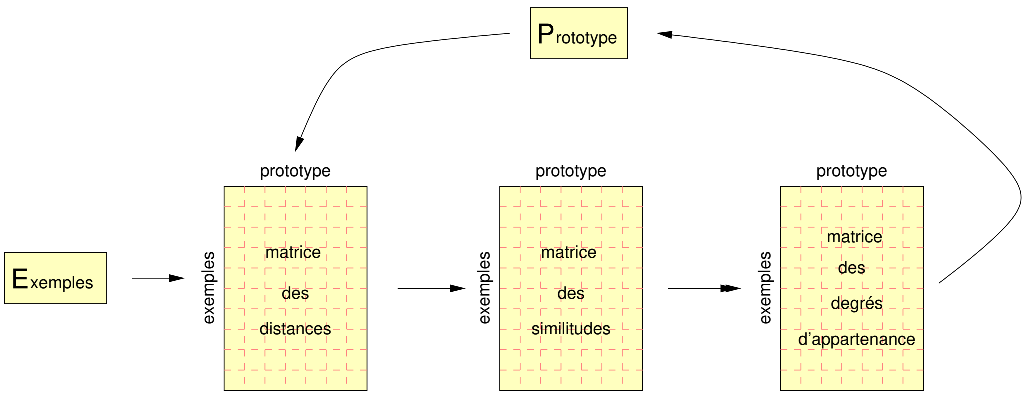


FIG. 4.4 – Algorithmes de regroupement itératifs

Soit $X = \{x_j / j = 1 \dots N\}$ l'ensemble des exemples, représentés par des vecteurs d'attributs $x_j = [x_{j,1}, x_{j,2}, \dots, x_{j,a}]$ et $P = \{p_i / i = 1 \dots C\}$ un ensemble de prototypes - caractérisant chacune des classes. Soit $d(x_j, p_i)$ une mesure de distance entre le prototype p_i et l'exemple x_j et $f_A(x_j, p_i)$ un degré d'appartenance de l'exemple x_j à la classe représentée par le prototype p_i.

Utilisant ces définitions nous pouvons écrire l'algorithme général pour la classification par itération :

---

1 - initialiser les p_i (choix du nombre des classes, et de leurs positions initiales).

2 - regrouper les exemples x_i en calculant les degrés d'appartenance $f_A(x_j, p_i)$ de chaque exemple à chaque classe .

3 - recalculer les prototypes en utilisant les degrés d'appartenance.

4 - tester la convergence et boucler vers le regroupement si le critère d'arrêt n'est pas satisfait.

Dans la phase du regroupement, le degré d'appartenance est souvent calculé par l'intermédiaire d'une mesure de dissemblance $d(x_j, p_i)$ et une mesure de similarité $s(x_j, p_i)$ comme c'est illustré sur la figure 4.4. Dans la suite, nous examinons uniquement les mesures les plus connues. Une bonne description des différentes mesures de similarité et de dissemblance se trouve dans [Rif96].

4.3.1 C-moyennes classiques (CM)

Pour obtenir l'algorithme des C-moyennes classiques il suffit de choisir :

$$d(x_j, p_i) = (x_j - p_i)^2 \tag{4.1}$$

et

$$f_{\text{A}}(x_j, p_i) = \begin{cases} 1 & \text{si } d(x_j, p_i) = \min_k(d(x_j, p_k)) \\ 0 & \text{sinon} \end{cases} \tag{4.2}$$

Cette équation peut aussi être interprétée en définissant la similarité $s(x_j, p_i)$ comme suit, ce qui facilite la comparaison avec les C-moyennes floues décrites dans le paragraphe suivant :

$$s(x_j, p_i) = \frac{1}{d(x_j, p_i)} = \frac{1}{(x_j - p_i)^2} \tag{4.3}$$

et

$$f_{\text{A}}(x_j, p_i) = \begin{cases} 1 & \text{si } s(x_j, p_i) = \max_k(s(x_j, p_k)) \\ 0 & \text{sinon} \end{cases} \tag{4.4}$$

La fonction d'appartenance est donc un singleton à la position de similarité maximale dans chaque ligne de la matrice de similarité. La mise à jour du prototype est réalisée par une simple moyenne pondérée des exemples de la base par leur fonction d'appartenance f_{A} :

$$p_i = \frac{\sum_{j=1}^{N} f_{\text{A}}(x_j, p_i) \cdot x_j}{\sum_{j=1}^{N} f_{\text{A}}(x_j, p_i)} \tag{4.5}$$

4.3.2 C-moyennes flous (FCM)

Le seul changement par rapport aux C-moyennes classiques est que la définition de f_{A} ne s'exprime plus à l'aide de l'opérateur maximum mais par une normalisation :

$$f_{\text{A}}(x_j, p_i) = \frac{s(x_j, p_i)}{\sum_{k=1}^{C} s(x_j, p_k)} \tag{4.6}$$

Ceci signifie que chaque exemple a un degré d'appartenance inversement proportionnel à la distance entre l'exemple et le prototype de la classe, le total de ces degrés étant égal à l'unité. Krishnapuram et Keller ont constaté [Kri93-2] que cette normalisation pose des problèmes pour l'interprétation des résultats obtenus avec $f_{\text{A}}(x_j, p_i)$ comme fonction d'appartenance. La normalisation introduit une dépendance entre les classes et étend le support des fonctions d'appartenance sur l'univers entier. Des exemples très éloignés des centres de tous les prototypes ont des valeurs d'appartenance à chaque classe, qui sont proches de $1/C$ et ainsi, ils influencent les prototypes de toutes les classes avec la même importance que des exemples plus proches. Krishnapuram et Keller ont proposé une autre méthode : les C-moyennes possibilistes.

4.3.3 C-moyennes possibilistes (PCM)

Les valeurs d'appartenance $f_{\text{A}}(x_j, p_i)$ ne sont plus calculées en normalisant dans la direction des prototypes mais indépendamment les unes des autres :

$$f_{\text{A}}(x_j, p_i) = \frac{1}{1 + \left(\frac{d^2(x_j, p_k)}{\eta_i}\right)^{\frac{1}{m-1}}} \tag{4.7}$$

Les distances $d(x_j, p_k)$ du dénominateur sont normalisées par une distance typique η_i de chaque classe. Cette valeur doit être déterminée au départ, mais peut être mise à jour en se basant par exemple sur la distance moyenne intra-classe [Kri93-2] :

$$\eta_i = K \frac{\sum_{j=1}^{N} f_A(x_j, p_i) d(x_j, p_i)}{\sum_{j=1}^{N} f_A(x_j, p_i)} \tag{4.8}$$

La valeur m dans l'exposant, qui modifie la fonction de distance, est généralement fixée d'une manière heuristique (souvent $m = 2$). Comme le remarquent Krishnapuram et Keller [Kri96], les C-moyennes possibilistes sont équivalentes à un ensemble de M-estimateurs robustes indépendants. C'est une des raisons pour laquelle les prototypes obtenus peuvent se superposer et même être égaux.

4.3.4 k-plus proches voisins (k-NN)

L'algorithme k-NN ne calcule pas de prototypes, mais utilise directement les distances à tous les exemples dans la base d'apprentissage pour classer chaque nouvel exemple. Cette méthode est souvent utilisée dans des systèmes de classification à base de cas puisqu'elle ne nécessite pas de transformation des données de la base d'apprentissage pour la classification de nouveaux exemples. Pour une position p donnée dans l'espace des attributs, la classe trouvée est celle de la majorité des k voisins les plus proches à cette position. L'algorithme utilise donc une distance, typiquement la distance euclidienne :

$$d(x_j, p) = (x_j - p)^2 \tag{4.9}$$

Puisque l'algorithme k-NN utilise uniquement l'ordre des exemples, n'importe quelle transformation monotone ne change pas le résultat et il est aussi possible de raisonner sur une mesure de similitude qui est une fonction monotone décroissante de la distance, par exemple :

$$s(x_j, p) = \frac{1}{d(x_j, p)} \tag{4.10}$$

Ensuite, les exemples sont classés selon leur similitude. Soit $B = \{b_i / i = 1 \ldots N\}$tel que b_i est la $i$ème plus grande valeur parmi $s(x_1, p) \cdots s(x_N, p)$. Les k plus grands membres de b_i sont sélectionnés, soit $B_k = \{b_i / i = 1 \ldots k\}$ et la classe la plus représentée dans B_k est attribuée au nouvel exemple à la position p.

Puisque le nombre k est fixé, la densité des exemples dans l'espace des attributs a une influence sur la taille de la zone d'influence des exemples. Pour le même k, la zone d'influence d'un exemple de la base est plus petite lorsque l'on considère un exemple dans une zone dense plutôt que dans une zone avec peu d'exemples.

4.3.5 L'algorithme EM

L'algorithme EM ressemble aux C-moyennes floues dans le sens où il calcule $f_A(x_j, p_i)$ par une normalisation. Le prototype est caractérisé par trois paramètres (w_i, μ_i, σ_i). La distance utilisée est une exponentielle :

$$d(x_j, p_i) = \frac{\sqrt{2\pi\sigma_i}}{w_i} \cdot e^{\frac{1}{2}\left(\frac{x_j - \mu_i}{\sigma_i}\right)^2} \tag{4.11}$$

et la similitude est une gaussienne :

$$s(x_j, p_i) = \frac{1}{d(x_j, p_i)} = \frac{w_i}{\sqrt{2\pi\sigma_i}} \cdot e^{-\frac{1}{2}\left(\frac{x_j - \mu_i}{\sigma_i}\right)^2} \tag{4.12}$$

ensuite le calcul de $f_{\mathrm{A}}(x_j, p_i)$ est strictement équivalent à celui des C-moyennes floues :

$$f_{\mathrm{A}}(x_j, p_i) = \frac{s(x_j, p_i)}{\sum_{k=1}^{C} s(x_j, p_k)} \tag{4.13}$$

La mise à jour des paramètres des prototypes est effectuée sur les trois paramètres $p_i = (w_i, \mu_i, \sigma_i)$ de la manière suivante :

$$w_i^{(m)} = \sum_{j=1}^{N} f_{\mathrm{A}}(x_j, p_i)^{(m)} \tag{4.14}$$

$$\mu_i^{(m)} = \frac{\sum_{j=1}^{N} f_{\mathrm{A}}(x_j, p_i)^{(m)} x_i}{w_i^{(m)}} \tag{4.15}$$

$$\sigma_i^{(m)} = \sqrt{\frac{\sum_{j=1}^{N} f_{\mathrm{A}}^{(m)} (x_i - \mu_k^{(m)})^2}{w_i^{(m)}}} \tag{4.16}$$

4.3.6 L'algorithme EM hiérarchique

L'algorithme EM hiérarchique ne suit pas strictement les étapes de l'algorithme général pour la classification par itérations tel que décrit au début du paragraphe 4.3, puisqu'il estime les prototypes un par un. Il est quand même intéressant de rappeler les relations utiles de chaque étape pour faire le lien avec les autres algorithmes de regroupement. La distance choisie est la même que pour l'algorithme EM classique :

$$d(x_j, p_i) = \frac{\sqrt{2\pi\sigma_i}}{w_i} \cdot e^{\frac{1}{2}\left(\frac{x_j - \mu_i}{\sigma_i}\right)^2} \tag{4.17}$$

mais f_{A} est définie comme :

$$f_{\mathrm{A}}(x_j, p_i) = \frac{\frac{1}{d(x_j, p_i)}}{t_s + \frac{1}{d(x_j, p_i)}} = \frac{1}{1 + t_s \cdot d(x_j, p_i)} \tag{4.18}$$

où t_s est un niveau de bruit représentant les autres prototypes non encore identifiés. On peut constater une équivalence entre les équations 4.18 et 4.7 pour $\eta_i = 1/t_s$ et $m = 2$. Contrairement aux C-moyennes possibilistes, l'algorithme EM hiérarchique ne fixe pas le paramètre t_s a priori, mais cherche la meilleure correspondance entre les prototypes et l'histogramme des exemples avec un test d'adaptation statistique. Cette méthode fonctionne généralement bien quand le nombre d'exemples dans la base est suffisant pour que les tests d'adaptation soient discriminants. Pour la recherche des prototypes suivants, l'algorithme EM hiérarchique soustrait une distribution qui représente les prototypes déjà identifiés, à la distribution des exemples dans la base.

La mise à jour de η dans l'équation 4.8 proposée par Krishnapuram et Keller ressemble à la mise à jour de σ dans l'équation 4.16 de l'H-EM.

Ensuite le calcul de $f_{\mathrm{A}}(x_j, p_i)$ est strictement équivalent à celui des C-moyennes floues :

$$f_{\mathrm{A}}(x_j, p_i) = \frac{s(x_j, p_i)}{\sum_{k=1}^{C} s(x_j, p_k)} \tag{4.19}$$

4.3.7 Discussion

Les algorithmes décrits ci-dessus utilisent tous la même structure générale mais les mesures de distances, de similitude et de normalisation sont différentes et leur comportements se distinguent significativement. Les critères de distinction sont :
- somme des valeurs d'appartenance égale à un
- adaptation de la taille (est-ce que l'algorithme adapte seulement les positions ou aussi la largeur de chaque prototype ?)
- variable la plus stable. (quelle est la variable supposée la plus stable : le nombre d'exemples dans chaque classe ou la taille de chaque classe dans l'espace des attributs ?)
- coïncidence (est-ce que les classes peuvent coïncider ?)
- influence d'exemples extrêmes (quelle est l'influence des exemples qui sont très loin des autres exemples ?)

Le tableau 4.1 résume les caractéristiques des algorithmes de regroupement.

	$hs = 1$	flou	adapte taille	coïncidence de classes	influence d'exemples extrêmes
CM	oui	non	non	jamais	sur une classe
FCM	oui	oui	non	parfois	sur toutes les classes
PCM	non	oui	possible	souvent	faible
k-NN	oui	non	oui	jamais	faible
EM	non	oui	oui	parfois	faible

TAB. 4.1 – Comparaison des méthodes de regroupement

L'ensemble des algorithmes basés sur la minimisation d'une énergie par itération est sensible à l'initialisation, mais la précision requise est différente pour les différentes approches. Plusieurs auteurs proposent par exemple d'utiliser les résultats du FCM pour initialiser le PCM puisque ce dernier demande une initialisation plus précise que le FCM.

Le choix d'un des algorithmes de regroupement s'appuie sur quelques caractéristiques du problème que nous nous posons :
- Nous possédons une bonne initialisation pour un certain nombre de classes, celles qui sont suffisamment importantes pour que l'algorithme H-EM les trouve dans le bon ordre indépendamment des variations locales dans une image.
- Nous avons une bonne idée du nombre d'éléments dans chaque classe. Celui-ci doit être à peu près égal au nombre d'images.

Pour notre problème de regroupement des composantes identifiées par les histogrammes des attributs, nous avons utilisé l'algorithme FCM parce qu'il représente un bon compromis dans le sens où il permet de répartir des exemples sur plusieurs classes et en même temps d'éviter la coïncidence de plusieurs classes.

Le fait que l'algorithme FCM donne des valeurs d'appartenance similaires à toutes les classes pour les exemples loin de tous les centres des classes ne pose pas de difficulté dans notre application, puisque les valeurs d'appartenances pour chaque exemple ne sont pas aussi importantes que les centres des classes qui servent comme prototype dans la suite.

Comme attributs pour chaque composante, nous avons, en plus de p_i, μ_i, σ_i, le pourcentage d'exemples qui se trouvent au dessus de la valeur modale dans tout l'histogramme et le pourcentage du poids de la composante par rapport à l'histogramme total.

Après l'algorithme FCM, les composantes qui représentent la même partie de l'histogramme sont dans la même classe mais les paramètres des composantes ne sont pas constants à l'intérieur de la classe.

4.4 Création d'un prototype

Les composantes identifiées dans toutes les classes et tous les exemples de la base d'apprentissage sont utilisées pour construire des prototypes pour la description des modes dans les histogrammes. C'est ce prototype qui sert à définir un ordre unique des composantes et les décrire avec leurs valeurs centrales et leur variabilité.

4.4.1 Prototype sans interactions

Si les composantes sont considérées comme indépendantes entre elles, on peut décrire le prototype des composantes par une valeur centrale et une variance. Un cas simple consiste à utiliser encore une fois une gaussienne pour modéliser la distribution des valeurs des paramètres. Cette démarche est représentée sur la figure 4.5 : les pixels de toutes les images sont séparés selon leur classe. Ensuite, le mélange des gaussiennes est construit pour modéliser la distribution globale.

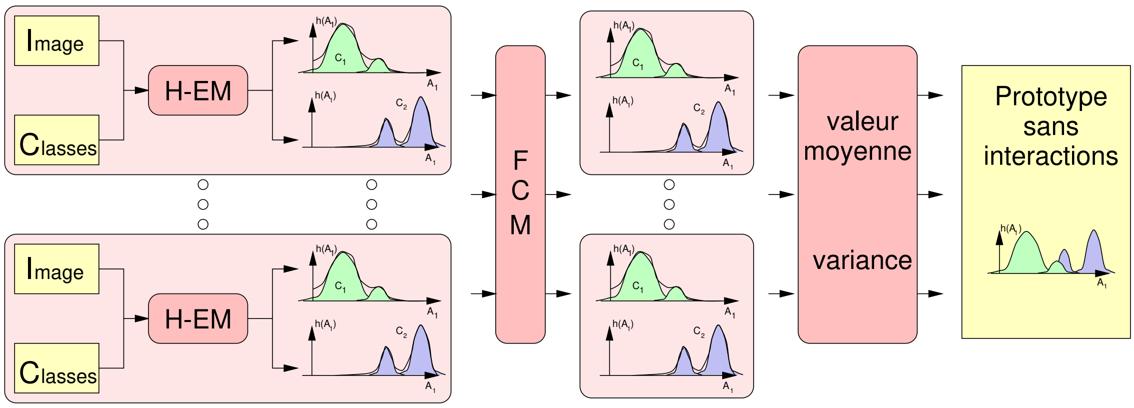


FIG. 4.5 – Création d'un prototype sans interaction

Lors de l'utilisation, l'adaptation de ce prototype à l'histogramme de l'image courante peu se faire avec une version modifiée de l'algorithme EM.

4.4.2 Algorithme EM avec prototype (P-EM)

Une manière simple d'utiliser un prototype pour la paramétrisation de l'histogramme d'un attribut est de le prendre comme initialisation de l'algorithme EM. Cette démarche ne permet pas de prendre en compte la grandeur de la variabilité, c'est-à-dire qu'aucune limite n'est imposée à l'algorithme EM pour la distance d'adaptation des composantes relatives au prototype.

Une autre possibilité est d'intégrer un poids dans l'étape M, ce qui pénalise la distance entre le prototype et les composantes pendant l'itération (algorithme P-EM). Cette pénalisation peut être interprétée comme un facteur de vitesse v_k d'adaptation des composantes vers l'histogramme d'une part et vers le centre du prototype d'autre part.

étape E

(mise à jour de λ_{ki})

$$\lambda_{ki}^{(m)} = \frac{p_k^{(m-1)} \cdot g(x_i, \mu_k^{(m-1)}, \sigma_k^{(m-1)})}{\sum_{j=1}^{K} p_j^{(m-1)} \cdot g(x_i, \mu_j^{(m-1)}, \sigma_j^{(m-1)})} \tag{4.20}$$

Ensuite les mélanges eux-mêmes sont mis à jour à partir des exemples pondérés par λ avec une vitesse qui dépend du prototype.

étape M

On détermine la vitesse $0 < v_k < 1$ de la mise à jour de la composante g_k en fonction de la similarité $S(g_k, \hat{g}_k)$ entre cette composante et la composante correspondante $\hat{g}_k$ du prototype. Plus la similarité entre la composante actuelle et celle du prototype est grande plus la vitesse de la mise à jour est élevée :

$$v_k^{(m)} = S(g_k, g_{k_{proto}}) \tag{4.21}$$

Mise à jour de p_k, μ_k, σ_k en prenant en compte la vitesse v_k.

$$p_k^{(m)} = v_k^{(m)} \cdot \sum_{i=0}^{N} \lambda_{ki}^{(m)} + (1 - v_k^{(m)}) \cdot p_k^{(m-1)} \tag{4.22}$$

$$\mu_k^{(m)} = v_k^{(m)} \cdot \frac{\sum_{i=0}^{N} \lambda_{ki}^{(m)} x_i}{p_k^{(m)}} + (1 - v_k^{(m)}) \cdot \mu_k^{(m-1)} \tag{4.23}$$

$$\sigma_k^{(m)} = v_k^{(m)} \cdot \sqrt{\frac{\sum_{i=0}^{N} \lambda_{ki}^{(m)} (x_i - \mu_k^{(m)})^2}{p_k^{(m)}}} + (1 - v_k^{(m)}) \cdot \sigma_k^{(m-1)} \tag{4.24}$$

avec :

μ_k moyenne de la gaussienne k

σ_k écart-type de la gaussienne k

p_k poids de la gaussienne k

λ_{ki} compatibilité de l'exemple i avec la gaussienne k

Dans le cas d'un prototype dont les variations des paramètres sont elles-mêmes des gaussiennes, il est possible de montrer que l'algorithme P-EM décrit ici est équivalent à maximiser la probabilité globale pour tous les pixels connaissant la probabilité a priori décrite par ces gaussiennes (voir annexe B).

Interprétation comme minimisation d'énergie

Une autre interprétation de la méthode d'adaptation du prototype par l'algorithme P-EM est celle d'une minimisation d'une fonction d'énergie $E(\Phi, \Phi_0)$ qui se décompose en deux parties :

E_{int} l'énergie interne, qui décrit l'éloignement de la position actuelle Φ de la position typique du prototype Φ_0.

E_{ext} l'énergie externe, qui décrit l'éloignement de la forme de la distribution actuelle à l'histogramme de l'exemple.

Le minimisation qui est réalisée par l'algorithme P-EM est celle de la somme des deux énergies : $E(\Phi, \Phi_0) = E_{\text{int}}(\Phi, \Phi_0) + E_{\text{ext}}(\Phi)$

Les limites d'un prototype sans interactions

Le prototype sans interactions donne les positions et les variations typiques des composantes trouvées dans l'histogramme. Dans notre algorithme P-EM, ces composantes sont adaptées à l'histogramme actuel ; par contre, cette représentation ne permet pas de prendre en compte, dans l'adaptation d'une des composantes, les adaptations déjà effectuées dans d'autre composantes. Cela peut poser des problèmes si l'histogramme est très décalé par rapport à sa valeur typique.

L'exemple de la figure 4.6 montre l'adaptation d'un prototype sur un histogramme décalé. A la place d'une adaptation de toutes les composantes dans leur ensemble, quelques composantes sont modifiées en taille et les autres sont utilisées pour expliquer 'le reste' de l'histogramme. Pour le premier histogramme, le mode principal est modélisé avec les composantes 1, 2, 3 et 4, tandis que pour le second histogramme de l'exemple, ce sont les composantes 1, 4, 6 et 7 qui sont mobilisées.

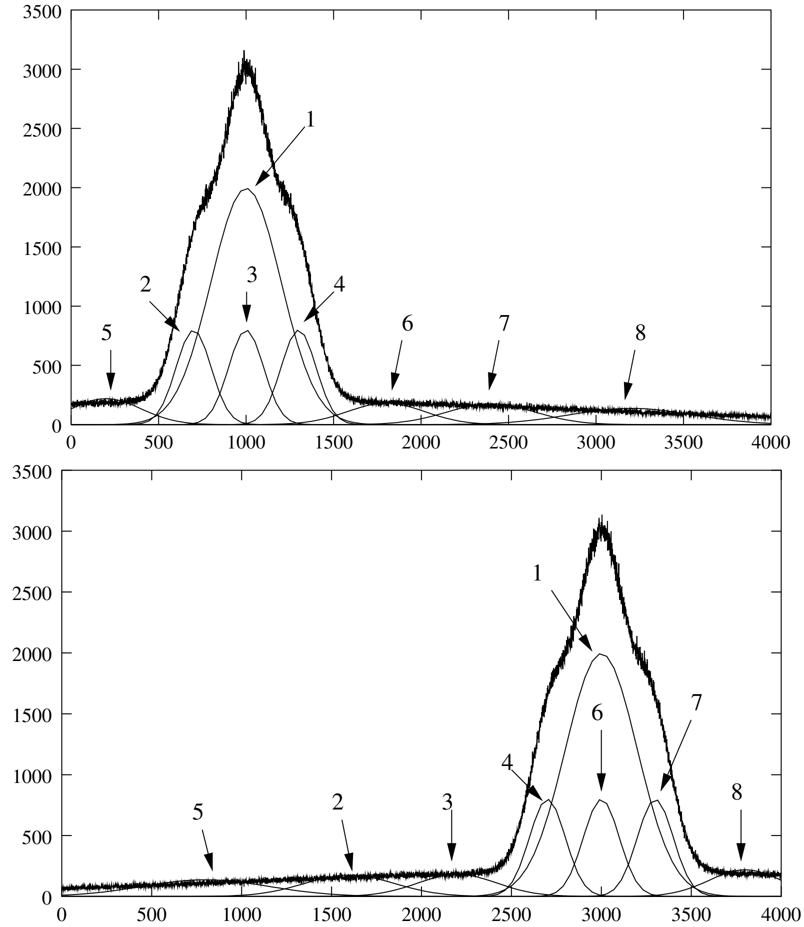


FIG. 4.6 – Limites d'un prototype sans interactions : les composantes sont échangées

Pour resoudre ce problème, il est intéressant d'inclure un modèle d'interactions entre les composantes du prototype.

4.4.3 Prototype avec interactions

L'idée des prototypes avec interactions est que les composantes qui décrivent l'histogramme peuvent s'adapter à l'histogramme mais qu'elles gardent toujours une certaine position et une certaine forme relative les unes par rapport aux autres. La nécessité d'inclure cette interaction devient plus évidente quand la forme typique des densités dans l'histogramme n'est pas une gaussienne. Dans ce cas, la forme peut être estimée par des composantes gaussiennes avec toute la précision souhaitée si le nombre des composantes pour l'approximation n'est pas limité. Par exemple, si l'histogramme respecte toujours une forme triangulaire symétrique dont la position et la largeur varient, toutes les composantes qui sont utilisées pour l'estimation subissent les mêmes translations et le même changement de largeur.

Pour inclure ces dépendances entre les paramètres des composantes dans le prototype, il est nécessaire de trouver un modèle. La dépendance peut être décrite par l'intermédiaire de la fonction d'énergie interne $E_{\text{int}}(\Phi,\Phi_0)$ du prototype. L'énergie interne doit être minimale pour les combinaisons de paramètres qui sont typiques, et de plus en plus élevée lorsque les paramètres s'éloignent des combinaisons courantes.

4.4.4 Modélisation des interactions

Parmi toutes les interactions qui peuvent exister entre les composantes, il faut faire un choix pour définir le modèle qui sera utilisé dans l'algorithme de classification adaptative. Pour un modèle donné, l'estimation des paramètres peut se faire par régression sur la base d'apprentissage. Si le modèle choisi est trop simple, ces paramètres ne permettent pas d'exprimer les liens plus complexes qui peuvent exister entre les composantes. Si le modèle est trop complexe, une grande base de données est nécessaire pour estimer les paramètres avec une bonne précision lors de l'apprentissage.

Modèle linéaire complet

Une relation linéaire est un modèle simple permettant de décrire la relation entre les composantes de deux histogrammes. Le modèle linéaire complet prend en compte toutes les interactions linéaires possibles, même entre les différents types de paramètres p,μ et σ. Dans notre cas, nous allons appliquer le modèle linéaire aux variations des paramètres autour des valeurs typiques :

$$\Delta\Phi = \Phi - \Phi_0 \tag{4.25}$$

Chaque variation est donc décrite par une combinaison linéaire de toutes les autres composantes. Sur la figure 4.7, ceci est illustré pour le lien entre deux paramètres seulement, bien que, dans le modèle linéaire complet, la dépendance entre chaque paire de paramètres soit déterminée. La

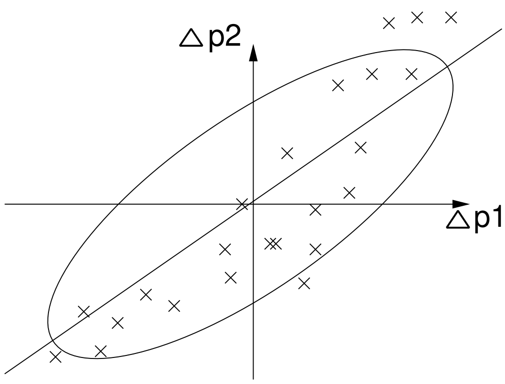


FIG. 4.7 – Régression pour déterminer le lien linéaire entre deux paramètres

figure 4.7 montre également que ce modèle se limite aux liens linéaires bien que des liens plus complexes puissent exister.

Si nous regroupons toutes les interactions dans une matrice $\mathbf{A}$, une dépendance linéaire parfaite est décrite par :

$$\mathbf{A}\cdot\Delta\Phi = 0 \tag{4.26}$$

En pratique, les composantes ne sont jamais parfaitement alignées sur une droite et il est donc nécessaire de permettre un écart par rapport à ce modèle idéal. Nous pouvons donc choisir comme énergie interne E_{int}, l'erreur quadratique par rapport à la droite définie par (4.26). Nous avons alors :

$$E_{\text{int}}(\Phi,\Phi_0) = (\mathbf{A}\cdot\Delta\Phi)^2 \tag{4.27}$$

Pour prendre en compte le fait que le modèle linéaire est plus pertinent pour certains paramètres et moins pour d'autres, les erreurs peuvent être pondérées par la variance R_i mesurée pendant l'apprentissage :

$$E_{\text{int}}(\Phi,\Phi_0) = (\mathbf{A}\cdot\Delta\Phi)^T \cdot \mathrm{R}_i \cdot \mathbf{A}\cdot\Delta\Phi \tag{4.28}$$

Modèle linéaire décomposé

Dans la plupart des applications, ce modèle est trop général. Il suffit souvent de considérer uniquement des interactions entre les paramètres du même type. Un modèle linéaire décomposé ne prend pas en compte les interactions entre les poids, les moyennes et les écart-types. Nous obtenons donc trois équations linéaires séparées :

$$\mathbf{A}_p \cdot (\vec{p}-\vec{p}_0) = 0 \tag{4.29}$$
$$\mathbf{A}_\mu \cdot (\vec{\mu}-\vec{\mu}_0) = 0 \tag{4.30}$$
$$\mathbf{A}_\sigma \cdot (\vec{\sigma}-\vec{\sigma}_0) = 0 \tag{4.31}$$

Avec ce modèle, l'énergie interne se décompose également en trois termes :

$$E_{\text{int}}(\Phi,\Phi_0) = |\mathbf{A}_p\cdot(\vec{p}-\vec{p}_0)|^2 + \left|\mathbf{A}_\mu\cdot(\vec{\mu}-\vec{\mu}_0)\right|^2 + |\mathbf{A}_\sigma\cdot(\vec{\sigma}-\vec{\sigma}_0)|^2 \tag{4.32}$$

Avec un modèle plus réduit, une base de données plus restreinte est suffisante pour avoir une bonne estimation de leur valeurs. En contrepartie, certains liens entre les largeurs et les positions des composantes ne peuvent pas être décrits.

Les paramètres $a_{i,j}$ doivent être estimés à partir des composantes trouvées pour chaque exemple de la base d'apprentissage avec une méthode de régression.

4.4.5 Apprentissage des interactions

La modélisation des interactions peut avantageusement utiliser comme première étape une modélisation avec un prototype sans interaction (figure 4.8). Cette démarche profite du réajustement de l'algorithme P-EM pour extraire les variations autour d'un prototype déjà établi dans la première phase.

Pour trouver la matrice des interactions **A**, on minimise l'erreur quadratique du modèle linéaire sur toutes les images j de la base d'apprentissage :

$$\sum_{j=1}^{J} \left|\mathbf{A}\cdot\Delta\Phi_j\right|^2 \to min \tag{4.33}$$

Si, pour les paramètres $(\Delta\Phi_j)$ de chaque image j, les résidus de l'approximation statique sont assemblés dans une matrice $\mathbf{M}_\Phi$, le critère à minimiser pour trouver la matrice des interactions **A** peut également être reformulé :

$$\sum_{k=1}^{K} |\mathbf{M}_\Phi \mathrm{A}_k|^2 \to min \tag{4.34}$$

où A_k est la k-ème colonne de la matrice A. Ce vecteur décrit les interactions entre le k-ème élément de Φ et tous les autres éléments.

La régression pour ce système linéaire peut être réalisée avec une des approches bien connues, comme la méthode des moindres carrés avec comme intermédiaire, la matrice caractéristique. Le minimisation revient alors à résoudre un système d'équations linéaires. Dans l'application décrite au chapitre suivant, nous avons choisi une réalisation basée sur une décomposition en valeurs singulières(DVS) qui facilite l'analyse de l'erreur lors de la régression [Pre92].

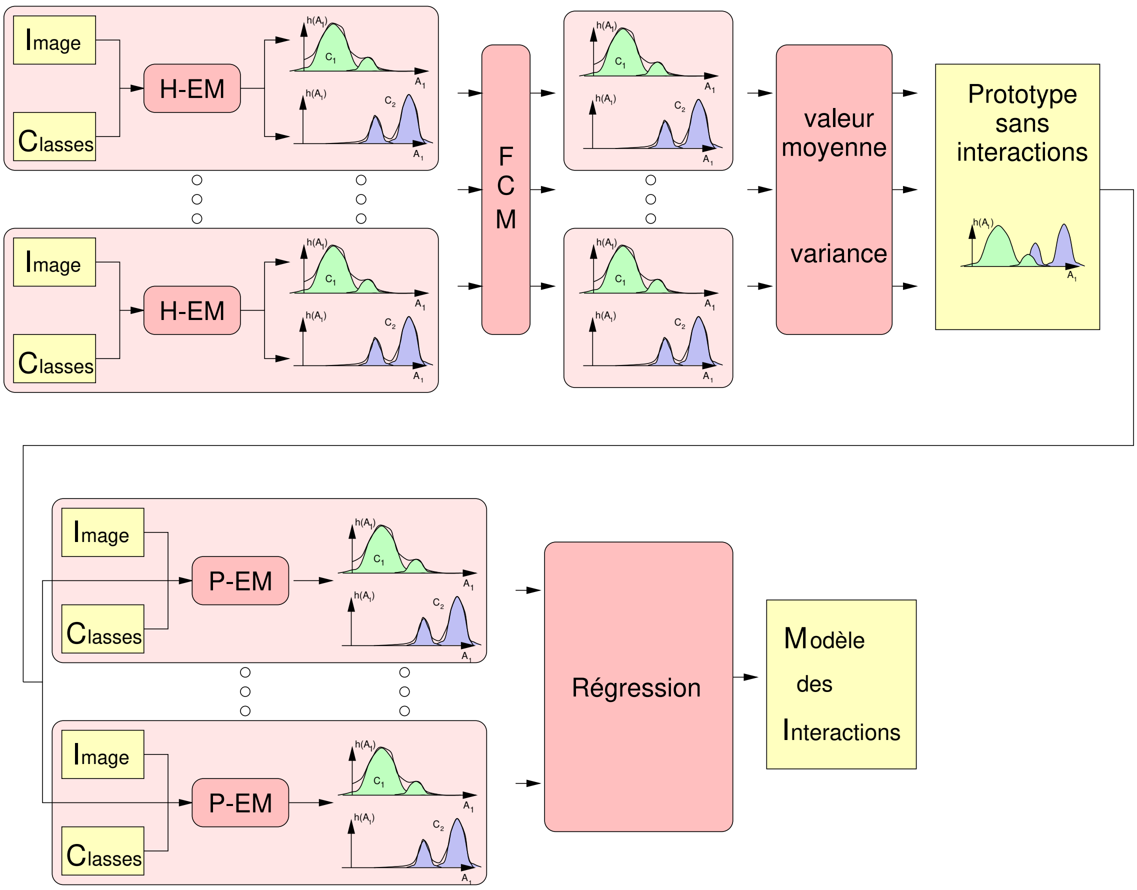


FIG. 4.8 – Création d'un prototype avec interaction

4.4.6 Algorithme P-EM avec interactions

Pour utiliser la connaissance des interactions entre les composantes de la description paramétrique, dans l'approximation d'un nouvel histogramme, l'algorithme P-EM doit être modifié. Puisque les composantes sont liées entre elles par le prototype avec interactions, les paramètres Φ ne peuvent plus être calculés indépendamment. Nous proposons d'ajouter une étape **P** à l'algorithme qui prend en compte le prototype et lie entre eux les paramètres des composantes.
Après l'apprentissage, les paramètres suivants ont été estimés :

- Φ_0 position typique du prototype
- R_e variance du prototype statique autour de Φ_0
- $\mathbf{A}$ matrice des interactions
- R_i variance du prototype avec interactions autour de la droite $\mathbf{A}\Phi = 0$

Pour simplifier les calculs nous allons noter **B** la matrice qui lie les valeurs des composantes par l'énergie interne. Cette matrice donne la modification à appliquer au vecteur des paramètres $\Delta\Phi$ pour réduire l'énergie interne. **B** est calculée à partir des paramètres estimés pendant la phase d'apprentissage :

$$\mathbf{B} = \left[\mathrm{R}_e^{-1} + \mathbf{A}^T \mathrm{R}_i^{-1} \mathbf{A}\right]^{-1} \tag{4.35}$$

Dans le cas du modèle décomposé nous obtenons :

$$\mathbf{B}_p = \left[\mathrm{R}_{e_p}^{-1} + \mathbf{A}_p^T \mathrm{R}_{i_p}^{-1} \mathbf{A}_p\right]^{-1} \tag{4.36}$$

$$\mathbf{B}_\mu = \left[\mathrm{R}_{e_\mu}^{-1} + \mathbf{A}_\mu^T \mathrm{R}_{i_\mu}^{-1} \mathbf{A}_\mu \right]^{-1} \tag{4.37}$$

$$\mathbf{B}_\sigma = \left[\mathrm{R}_{e_\sigma}^{-1} + \mathbf{A}_\sigma^T \mathrm{R}_{i_\sigma}^{-1} \mathbf{A}_\sigma \right]^{-1} \tag{4.38}$$

Avec cette notation, les étapes de l'algorithme sont alors :

- étape **E** : calculer les λ_{ki}
- étape **M** : calculer Φ en minimisant l'énergie externe
- étape **P** : calculer Φ en minimisant l'énergie interne :
  - soustraire le centre du prototype : $\left(\Phi^{(m-1)} - \Phi_0\right)$
  - appliquer l'énergie interne : $\mathbf{B}\left(\Phi^{(m-1)} - \Phi_0\right)$
  - ajouter le centre du prototype : $\Phi^{(m)} = \mathbf{B}\left(\Phi^{(m-1)} - \Phi_0\right) + \Phi_0$

Décrites de cette manière, les étapes **E** et **M** sont équivalentes à celles de l'algorithme EM standard, c'est-à-dire qu'elles minimisent l'énergie externe entre la position des composantes et l'histogramme. L'étape **P** réduit l'énergie interne qui décrit les positions des composantes les unes par rapports aux autres.

Chapitre 5

Construction des fonctions d'appartenance à partir d'un histogramme

Dans les chapitres précédents, nous avons vu comment la distribution des densités d'exemples peut être estimée et comment la variabilité peut être modélisée. Dans ce chapitre, nous allons regarder les différentes possibilités pour passer de la description des densités d'exemples à des fonctions d'appartenance.

Pour choisir la bonne méthode de transformation des densités d'exemples en des fonctions d'appartenance, il y a deux parties à considérer : la façon dont les densités sont obtenues et le raisonnement que l'on veut appliquer dans la suite.

5.1 La base d'apprentissage

Les bases d'apprentissage, pour la détection et la classification d'objets dans le traitement d'images, sont souvent constituées de la base d'images elle-même, enrichie avec les annotations d'experts. A titre d'illustration, une image de mammographie qui contient une lésion est montrée sur la figure 5.1. L'expert peut, par exemple, avoir entouré des parties de l'image pour montrer la présence et la localisation des objets à détecter, comme sur la figure 5.1 à droite.

Dans le cadre d'un raisonnement probabiliste, le contour est considéré comme le vrai contour de l'objet dans l'image. Autrement dit, tous les pixels à l'intérieur de la zone d'intérêt ont une valeur d'appartenance à la classe *pixel de l'objet à détecter* égale à l'unité, et tous les autres ont une valeur égale à zéro. Cela est illustré sur la figure 5.2.

La base d'apprentissage est utilisée pour estimer la fréquence d'apparition des classes en fonction des valeurs d'attributs. Dans le raisonnement bayésien, la probabilité de se tromper, quand on prend une décision basée sur les attributs, est déterminée. Un coût peut être associé à chaque bonne ou mauvaise classification, ce qui permet de définir un risque total (coût multiplié par la fréquence d'apparition). L'objectif est alors de minimiser ce risque total.

L'utilisation des sous-ensembles flous permet d'élargir les possibilités du raisonnement. Le contour donné par l'expert n'est plus interprété comme le bord d'une zone dont la valeur d'appartenance de tout pixel est égale à l'unité. Il délimite une zone dans laquelle la plupart des pixels appartient à l'objet à détecter. En conséquence, les valeurs d'appartenance ne sont plus binaires, mais comprises entre zéro et un.

Pour une base d'apprentissage composée d'images, il se pose le problème de déterminer les fonctions d'appartenance pour chacun des points de l'image. Cette information peut rarement être obtenue en interrogeant l'expert, puisque la quantité d'informations à fournir pour cela est énorme.

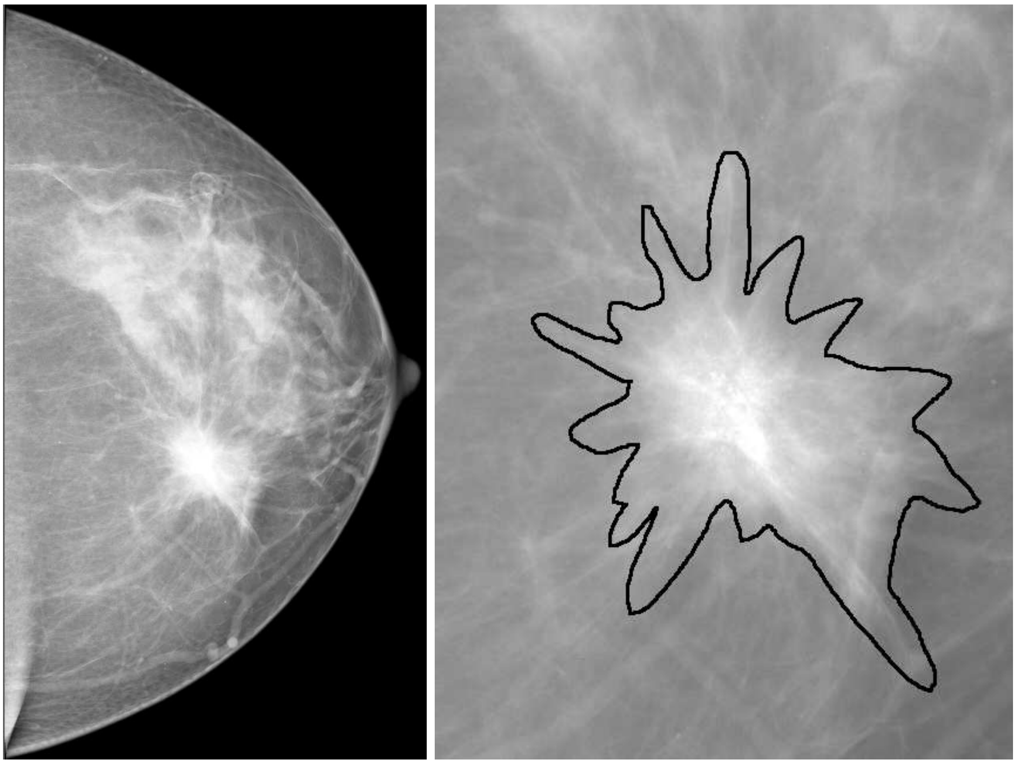


FIG. 5.1 – Image de mammographie et annotations

FIG. 5.2 – Fonction d'appartenance non floue pour le raisonnement probabiliste

Les fonctions d'appartenance doivent donc être reconstruites à partir de l'information que l'expert peut fournir, par exemple, à partir de contours des zones d'intérêt.

Cette reconstruction des fonctions d'appartenance n'est pas unique et la méthode utilisée doit dépendre des hypothèses que l'on peut apporter au problème. Dans la suite, nous allons montrer plusieurs types de raisonnement et en déduire la façon de calculer les fonctions d'appartenance.

5.2 Fonctions d'appartenance géométriques

Si l'on suppose connue la largeur de la zone de transition entre les classes sur l'image, il est possible d'estimer les fonctions d'appartenance de manière purement géométrique, sans prendre en compte les valeurs d'attributs. Un exemple pour ce type de raisonnement est de considérer le contour, comme étant l'alpha-coupe de la fonction d'appartenance. On calcule alors une fonction de distance vers l'extérieur et vers l'intérieur du contour par des dilatations et érosions successives. La fonction d'appartenance peut donc être une fonction de la distance (avec signe) au contour.

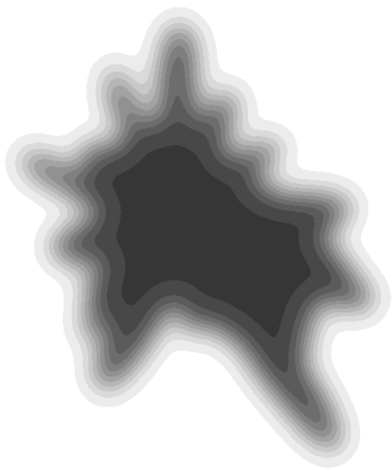


FIG. 5.3 – Fonction d'appartenance géométrique

Cette approche est applicable si l'expert peut tracer des contours avec une certaine précision qui est a priori connue. Pour le contour montré précédemment, la fonction d'appartenance construite de cette manière est illustrée sur la figure 5.3.

5.3 Fonctions d'appartenance probabilistes

Le but final de la construction des fonctions d'appartenance pendant l'apprentissage est la classification de nouveaux exemples en utilisant ces fonctions. Autrement dit, les valeurs d'appartenance à chaque classe pour chaque pixel de la base d'apprentissage ne sont pas indispensables si l'on connaît une méthode pour calculer directement les valeurs d'appartenance en fonction des attributs.

Les informations connues de la base d'apprentissage, autres que les relations géométriques, sont les valeurs des attributs et la position à l'intérieur ou à l'extérieur du contour dessiné par l'expert. Dans le cas le plus général (mais sans prendre en compte la géométrie), nous avons donc l'histogramme des points à l'intérieur et l'histogramme des points à l'extérieur du contour pour construire la fonction d'appartenance sur l'espace de tous les attributs.

Un raisonnement possible pour construire une fonction d'appartenance à partir d'un contour donné, est de définir une fonction entre le rapport des nombres de points à l'intérieur du contour $h_1(a)$ et le nombre totale de pixels $h_1(a)+h_2(a)$ avec les mêmes valeurs d'attributs a.

$$f_{c1}\left(h_1(a), h_2(a), a\right) = F_1\left(\frac{h_1(a)}{h_1(a)+h_2(a)}\right) \tag{5.1}$$

$$f_{c2}\left(h_1(a), h_2(a), a\right) = F_1\left(\frac{h_2(a)}{h_1(a)+h_2(a)}\right) \tag{5.2}$$

Si la fonction F_1 est l'identité, nous retrouvons des fonctions d'appartenance probabilistes. Ce choix fait que les fonctions d'appartenance sont calculées exactement comme des probabilités dans le cas non flou ; ce qui ne veut pas forcément dire que, dans la suite, les opérateurs d'agrégation doivent suivre le raisonnement bayésien (opérateur produit pour la conjonction). Un exemple pour un raisonnement qui nécessite une fonction F_1 différente de l'identité est donné dans la section 5.5.

5.4 Fonctions d'appartenance de surface

Dans ce paragraphe, nous présentons un raisonnement qui prend en compte la façon dont les contours sont obtenus en interrogeant l'expert.

Quand l'expert trace le contour d'un objet compliqué sur une image, comme pour l'exemple présenté dans la deuxième partie de ce travail - les lésions en mammographie - il ne va généralement pas pouvoir le tracer avec une précision extrême. Même pour une structure qui pourrait être caractérisée par un fonction d'appartenance non floue comme la structure de la figure 5.4 à gauche, l'expert a tendance à simplifier ou bien à généraliser en donnant un contour comme celui de la figure 5.4 à droite.

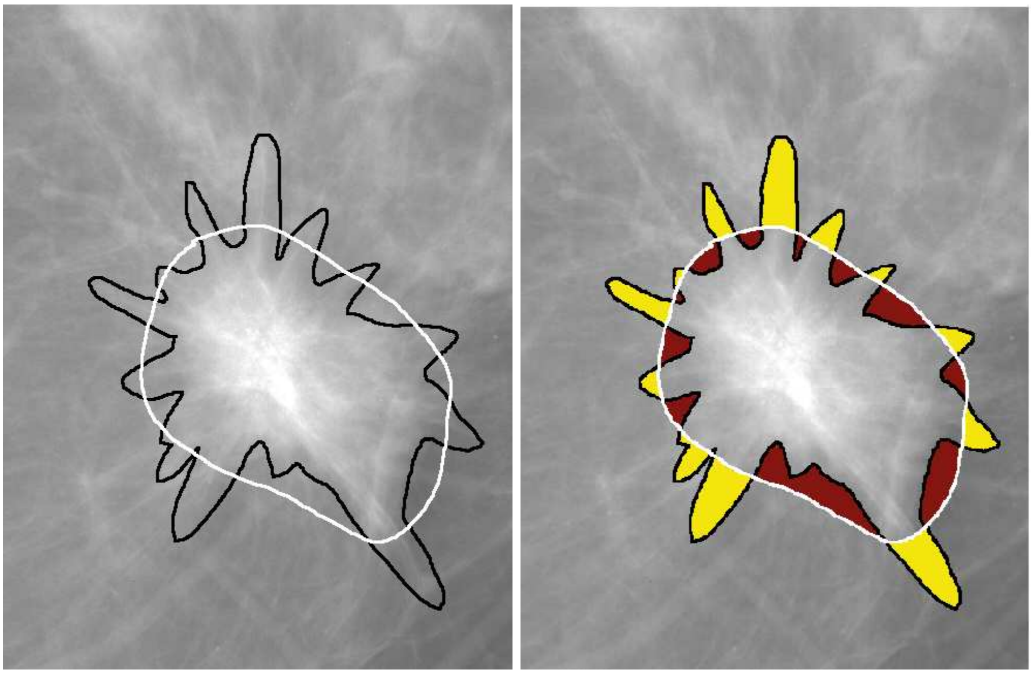


FIG. 5.4 – Contour tracé par l'expert

A proprement parler, le contour ne satisfait ni les conditions demandées pour faire une classification bayésienne (puisque les points à l'intérieur du contour ne sont pas tous entièrement membres de la même classe), ni les conditions pour construire des fonctions d'appartenance purement géométriques (puisque le contour ne représente pas une valeur d'appartenance donnée).

Pour établir une nouvelle méthode de construction des fonctions d'appartenance à partir de ces contours approximatifs, nous allons supposer que l'expert a essayé de tracer un contour qui représente une alpha-coupe (seuil α_1) de la vraie fonction d'appartenance, et qu'il a simplifié ce contour. Nous supposons que la simplification est faite de telle manière que l'aire de la région qui a une valeur d'appartenance au-dessus du seuil α_1, mais qui est restée à l'extérieur du contour à cause de la simplification, est égale à l'aire de la région qui a une valeur d'appartenance en dessous du seuil α_1 mais qui est restée à l'intérieur du contour. Cela est indiqué dans la figure 5.4 à droite par le fait que le contour simplifié coupe en deux l'aire de la *mauvaise classification* .

Pour formaliser cette hypothèse d'un manière générale, les valeurs d'appartenance sont fonction de quatre paramètres. Ces paramètres correspondent aux aires des points classés correctement ou incorrectement pour les deux classes, si l'on applique une classification binaire avec un seuil s, sur les histogrammes h_1 et h_2 des deux classes.

$$a_{1,1}(s) = \int_{-\infty}^{s} h_1(b)db \qquad \text{éléments de la classe 1 classés correctement} \tag{5.3}$$

$$a_{1,2}(s) = \int_{s}^{+\infty} h_1(b)db \qquad \text{éléments de la classe 1 classés dans la classe 2} \tag{5.4}$$

$$a_{2,1}(s) = \int_{-\infty}^{s} h_2(b)db \qquad \text{éléments de la classe 2 classés dans la classe 1} \tag{5.5}$$

$$a_{2,2}(s) = \int_{s}^{+\infty} h_2(b)db \qquad \text{éléments de la classe 2 classés correctement} \tag{5.6}$$

$$f_{c1}(s) = F_3\left(a_{1,1}(s), a_{1,2}(s), a_{2,1}(s), a_{2,2}(s)\right) \tag{5.7}$$

Dans le cas plus spécifique cité au début de cette section où l'on considère uniquement les aires des fausses classifications $a_{1,2}(s)$ et $a_{2,1}(s)$, on pourrait par exemple choisir :

$$f_{c1}(s) = \frac{a_{1,2}(s)}{a_{1,2}(s) + a_{2,1}(s)} \tag{5.8}$$

$$f_{c2}(s) = \frac{a_{2,1}(s)}{a_{1,2}(s) + a_{2,1}(s)} \tag{5.9}$$

Ce faisant, on introduit une dépendance linéaire entre les rapports des aires et la valeur d'appartenance. Les fonctions d'appartenance construites de cette manière sont monotones. Cette méthode est bien adaptée pour construire des fonctions d'appartenance décrivant des zones de transition entre deux classes. Dans le cas d'un histogramme multimodal, où il existe plusieurs zones de transitions, il est utile de séparer les histogrammes entre les maxima, d'appliquer la construction des fonctions d'appartenance sur chaque zone de transition séparément, et ensuite d'agréger les fonctions résultantes avec des conjonctions s'il y a plusieurs zones de transition sur le même maximum, ou des disjonctions s'il y a plusieurs maxima pour la même classe.

5.5 Fonction d'appartenance basée sur la détectabilité

Une autre situation couramment rencontrée, lorsqu'on utilise les annotations de l'expert sur l'image, est celle où la partie à détecter est entourée d'une manière très large. Spécialement pour le marquage des lésions en mammographie, les experts ont tendance à marquer la zone à prélever lors d'une opération, plutôt que le signe à détecter. Cette zone plus large contient le signe à détecter, mais aussi une grande partie de tissu normal. Pour un système de détection, il s'agit donc plus d'une zone d'intérêt que d'une délimitation du signe à marquer. Le but de la détection n'est pas de trouver la zone indiquée par l'expert, mais de trouver un marqueur pour le signe anormal à l'intérieur de cette zone.

Afin de traduire cette description du problème en une méthode de normalisation pour la création des fonctions d'appartenance, nous allons utiliser la notion de détectabilité. Une zone est détectable si il existe au moins un attribut par lequel le nombre de zones dans l'image avec des valeurs d'attribut similaires est faible. Cette détectabilité peut être définie comme la probabilité de trouver un élément de la classe à détecter, pour une valeur donnée des attributs. Nous nous retrouvons donc dans le cas détaillé dans la section 5.3, mais avec une fonction F_1 différente. Si nous supposons que dans la zone délimitée par l'expert se trouve au moins un point qui est complètement suspect, la fonction F_1 doit fournir des fonctions d'appartenance normalisées. D'un autre côté, nous ne recherchons pas tous les points dans la zone délimitée. Cela peut être traduit de différentes manières, en fonction de ce que l'on considère le plus constant :

– la détectabilité minimale
– le pourcentage de surface dans la zone marquée qui est détectable

Dans le premier cas, la fonction d'appartenance est nulle pour toutes les valeurs d'attributs pour lesquelles la détectabilité est en dessous d'un seuil. Dans le deuxième cas, les valeurs d'attributs sont ordonnées selon leur détectabilité et seuls les premiers *n* pourcents sont non nuls.

5.6 Un exemple

Pour illustrer le comportement des différentes méthodes de normalisation permettant l'obtention des fonctions d'appartenance, nous les appliquons au problème de la détection des lésions en mammographie qui est présenté dans la deuxième partie. Nous considérons ici uniquement un attribut très simple : le niveau de gris de l'image originale.

La figure 5.5 montre une image de mammographie, avec une lésion et une annotation ajoutée par un médecin pour montrer la position de la lésion. Puisque la lésion est stellaire, le médecin a entouré la lésion très large pour inclure toutes les pointes rayonnantes autour du noyau de la lésion. Les histogrammes de l'attribut *niveau de gris* pour les régions à l'intérieur et extérieur au contour, sont tracés sur la figure 5.6. Le noyau de la lésion est plus clair (ici représenté par un niveau faible) que le reste de l'image, mais puisque l'ensemble de la région entourée ne contient pas seulement le noyau, l'histogramme de la région marquée se recouvre avec l'histogramme des zones normales.

La figure 5.7 montre le résultat des trois méthodes de normalisation présentées sous la forme de fonctions d'appartenance et sous forme d'images des valeurs d'appartenance à la classe *objet à détecter*.

Les hypothèses requises pour les deux premières méthodes ne sont pas bien respectées dans la façon dont le médecin a tracé le contour.

Pour les fonctions d'appartenance probabilistes (à gauche sur la figure 5.7), la surface à l'intérieur du contour qui n'appartient pas au noyau de la lésion augmente la plage des niveaux de gris qui sont dans le support de la fonction d'appartenance. Le fait qu'il n'y ait pas de niveau de gris à l'intérieur de la lésion qui n'existe pas aussi à l'extérieur, se traduit, dans la fonction d'appartenance probabiliste, par le fait qu'elle n'est pas normalisée.

Les fonctions d'appartenance de surface sont un peu plus sélectives et définissent légèrement mieux la frontière des deux classes. Mais l'hypothèse selon laquelle le contour tracé correspond au contour de la 0,5-coupe simplifiée, n'est pas justifiée. Le fait que le contour tracé par le médecin contient le noyau de la lésion et toute la zone du tissu environnant dans lequel se trouvent les spicules (fibres qui rayonnent du centre de la lésion), se traduit dans une extension de la fonction

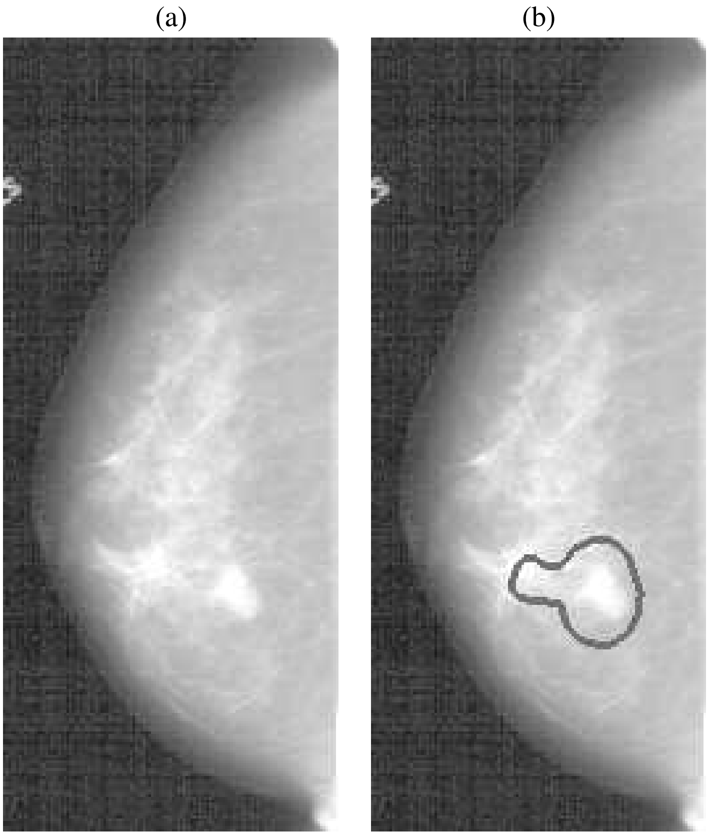


FIG. 5.5 – Image de la base DDSM cancer10 (a) avec l'annotation (b)

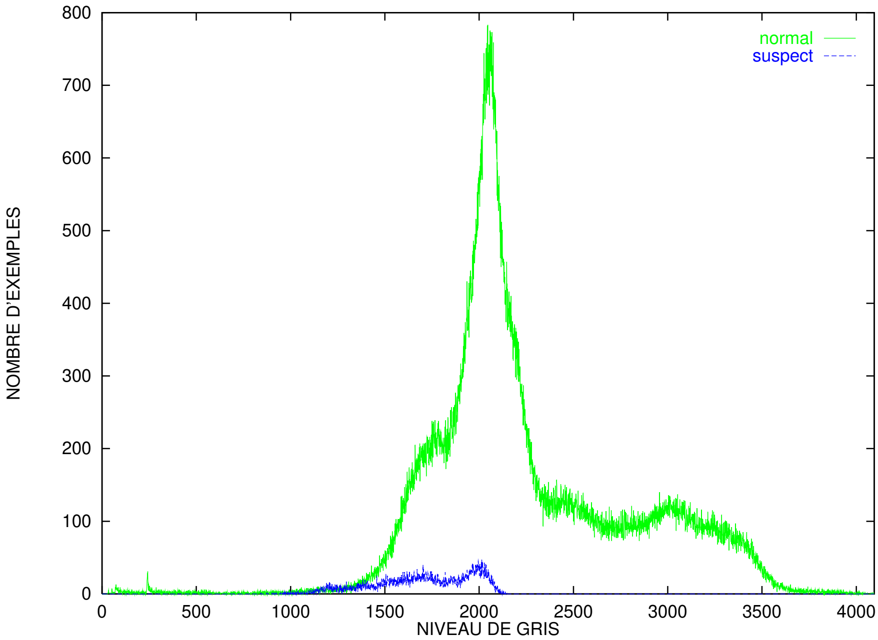


FIG. 5.6 – Histogramme de l'attribut des pixels à l'intérieur et à l'extérieur du contour

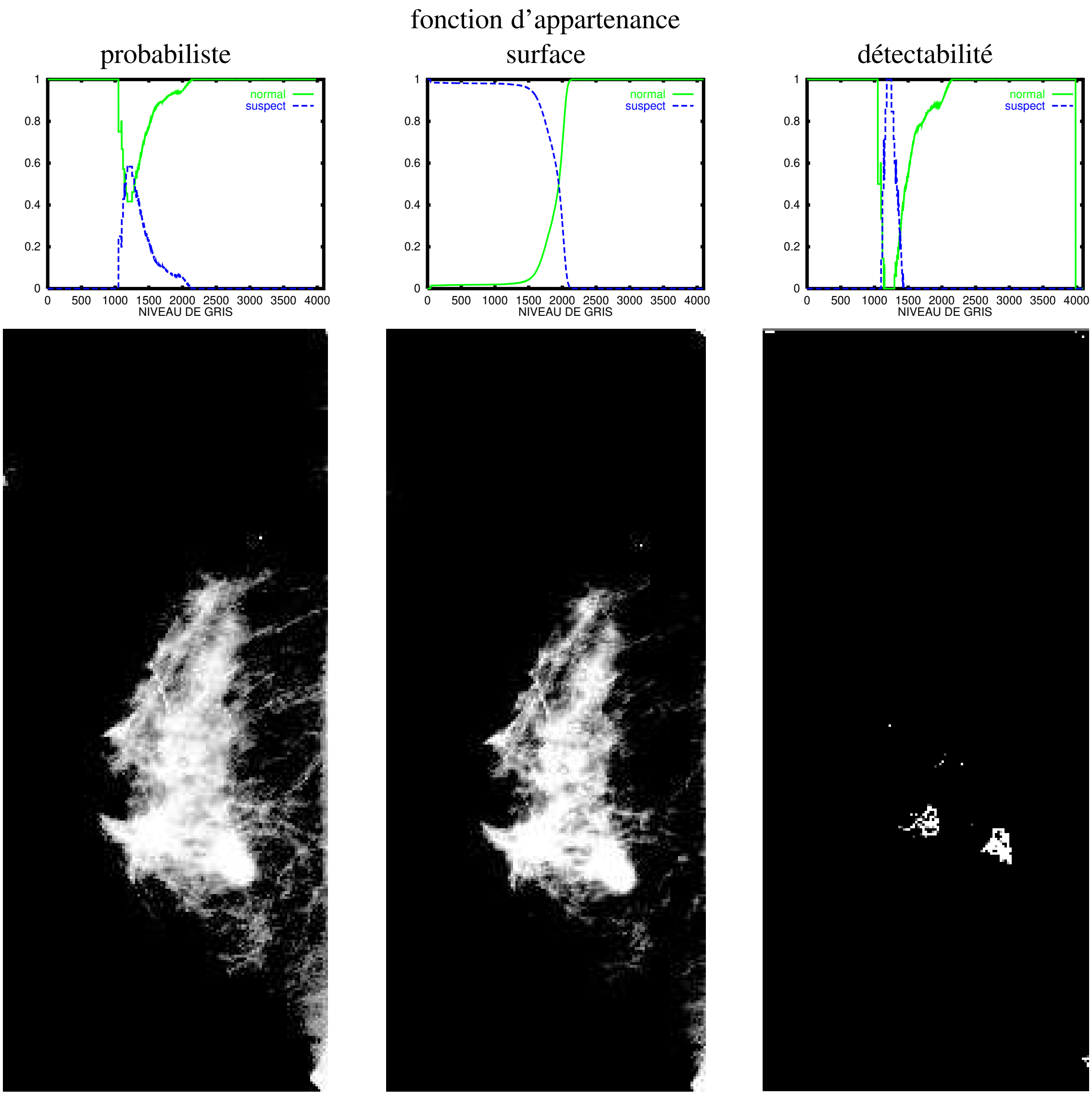


FIG. 5.7 – Comparaison des méthodes de normalisation

d'appartenance vers des valeurs de gris plus élevées (plus sombres). Appliqué à l'image d'origine, les valeurs d'appartenance d'une grande partie de la zone glandulaire sont relativement élevées (voir la partie centrale de la figure 5.7). Pour obtenir des fonctions d'appartenance plus spécifiques au noyau de la lésion, le contour tracé par le médecin devrait plutôt être interprété comme le support de la fonction d'appartenance.

Dans un but de détection, la troisième méthode est très intéressante, parce qu'elle combine l'information du médecin sur la position de la zone suspecte avec les caractéristiques uniques de cette zone par rapport au reste de l'image. La partie de l'univers des niveaux de gris retenue par cette méthode correspond au noyau qui est la partie la plus facilement distinguable dans cette image.

Chapitre 6

Un nouvel algorithme de classification floue adaptative

Dans la première partie de cette thèse, nous avons décrit les différents éléments d'une nouvelle méthode de classification floue adaptative. Cette méthode s'adresse aux problèmes de classification :

- qui nécessitent un traitement de l'incertitude
- dont les attributs sont soumis à une variabilité qui provient de variables non contrôlées
- dont la position des classes dans l'espace des attributs doit être apprise à partir d'une base d'exemples
- pour lesquels l'ensemble des données d'un nouvel exemple contient des informations sur le contexte, qui peut donc aider à réduire les effets de la variabilité.

Une grand nombre d'applications en traitement d'images satisfait à tous ces critères. Dans ce chapitre, nous allons d'abord récapituler la chaîne complète de classification que nous proposons pour ces problèmes (figure 6.1). Ensuite, nous présentons les résultats d'une première validation

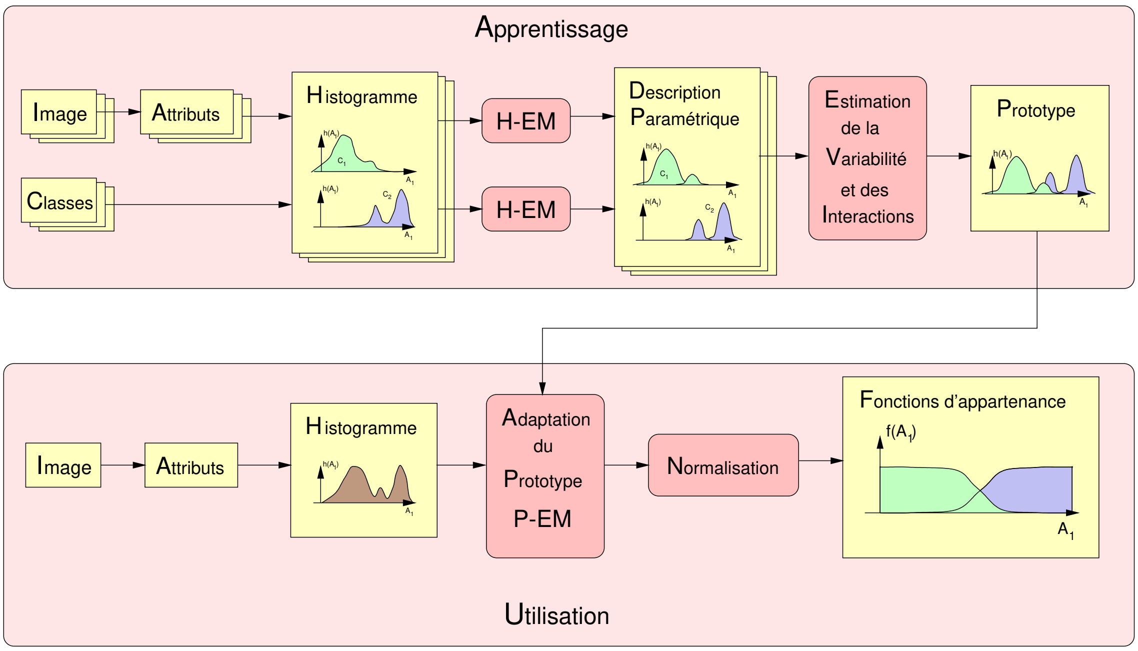


FIG. 6.1 – Classification floue adaptative

de la méthode appliquée à une base de données synthétique. Finalement, nous montrons des liens intéressants avec les modèles de Markov cachés.

6.1 Les étapes de la classification floue adaptative

Nous considérons dans ce chapitre le problème de classification de tous les pixels des images en deux classes. Pour la base d'apprentissage, nous supposons que la vraie classe est connue pour chaque pixel. Dans la pratique, cela est souvent réalisé en demandant à un expert de marquer les zones à détecter ou d'indiquer les classes à attribuer sur différentes parties de l'image.

6.1.1 Calcul des attributs

La première étape est commune à tous les systèmes de classification en traitement d'images : le calcul des attributs. Le calcul des attributs est effectué sur la base d'apprentissage et également pour tous les exemples de test ou de l'utilisation.

Pour la phase d'apprentissage, les histogrammes de chaque attribut sont calculés pour chaque image et chaque classe.

6.1.2 Modélisation de l'histogramme

La deuxième étape est la modélisation des densités d'exemples dans l'histogramme. Pour cela nous utilisons l'algorithme H-EM qui décompose l'histogramme successivement en différentes composantes. Cela est fait pour chaque image, chaque attribut et chaque classe.

6.1.3 Modélisation des liens internes par régression

L'estimation des liens internes se fait en deux étapes.

– un regroupement de composantes
– une régression des paramètres des composantes.

Le regroupement des composantes sert à associer entre elles les composantes identifiées indépendamment sur chaque image. Si les images ont des caractéristiques communes, cela se traduit en une forte concentration des composantes dans l'espace des paramètres des composantes. Nous avons utilisé l'algorithme des C-moyennes floues pour ce regroupement. Les paramètres des composantes sont p, μ, σ, respectivement le poids, la moyenne et l'écart-type pour des composantes gaussiennes.

Si les histogrammes sont suffisamment similaires d'une image à l'autre et les composantes suffisamment différentes, l'étape de regroupement n'est pas nécessaire, car l'algorithme H-EM va fournir les composantes toujours dans la même ordre. Dans le cas contraire, spécialement s'il y a deux composantes importantes bien séparées mais avec un poids similaire, l'algorithme H-EM peut renverser l'ordre dans lequel il identifie les composantes. C'est pourquoi l'étape de regroupement peut devenir nécessaire.

La deuxième partie est la régression afin de trouver un modèle qui décrive les liens entre les paramètres. Nous avons utilisé un modèle linéaire dans les exemples que nous avons testés, mais d'autres hypothèses sont possibles. Dans ce cas, une régression classique peut être appliquée. Nous avons choisi une méthode reposant sur la DVS (décomposition en valeurs singulières) qui est connue pour sa robustesse numérique ([Pre92]). Les méthodes de régression floues peuvent également être utilisées. Cela peut permettre de limiter la zone des valeurs que les paramètres peuvent prendre.

6.1.4 Apprendre les classes

La modélisation des densités des histogrammes par des composantes et des liens entre ces composantes, est effectuée à partir de la base d'apprentissage. Les composantes sont créées pour chaque classe indépendante, bien que le modèle des liens inclue les interactions entre composantes des différentes classes. Pour une classification à un seul attribut, il n'y a donc plus rien à apprendre. Pour les problèmes à plusieurs attributs, cette méthode peut être appliquée attribut par attribut. Par exemple, lors de la construction d'un arbre de décision flou, toutes les fonctions d'appartenance adaptatives sont construites et celle qui maximise un critère de séparation (voir section 6.1.6 et dans [Mar98]) est choisie pour le nœud suivant de l'arbre. Ensuite, les fonctions d'appartenance adaptatives sont calculées pour les éléments dans chaque branche.

6.1.5 L'utilisation du modèle

Après la phase d'apprentissage, le modèle peut être utilisé pour la classification des nouveaux exemples. Les mêmes attributs que pour l'apprentissage sont calculés. Les histogrammes sont alors calculés pour chaque attribut sur l'ensemble de l'image. Grâce à l'algorithme P-EM, les prototypes des classes sont adaptés aux histogrammes de l'image actuelle. Ensuite, le prototype est normalisé, ce qui donne les fonctions d'appartenance adaptées. Ces fonctions d'appartenance peuvent être utilisées dans des systèmes de classification à base de règles ou d'arbres de décisions flous.

6.1.6 Arbres de décisions flous

Dans ce rapport, nous ne décrivons pas l'ensemble de la théorie des arbres de décision, ni celle des arbres de décision flous, mais nous donnons quelques indications pour illustrer comment la gestion de la variabilité proposée peut être appliquée à un classificateur. Une description approfondie des arbres de décisions flous se trouve dans [Mar98].

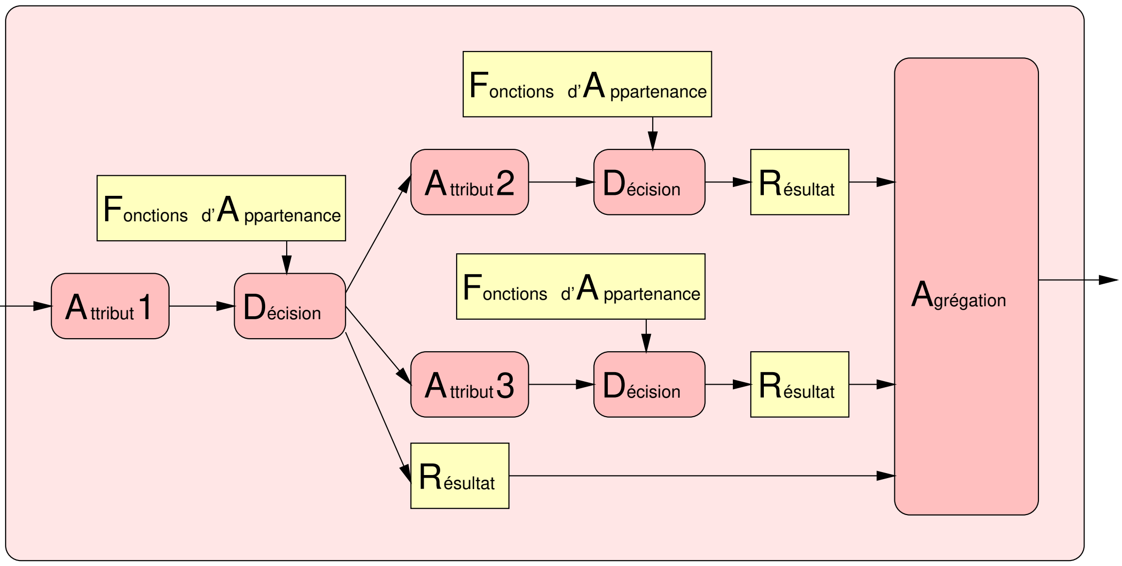


FIG. 6.2 – Arbre de décision flou

La figure 6.2 montre un exemple d'un arbre de décision à trois nœuds. Dans chaque nœud de l'arbre, une décision est prise en fonction d'un attribut.
Dans le cas des arbres de décision classiques avec un attribut numérique, la décision est du type :

Si la valeur pour l'attribut > seuil, **alors** suivre branche 1 **sinon** suivre branche 2

En suivant les branches et en appliquant les décisions à chaque nœud, l'arbre est traversé jusqu'à une feuille qui représente une classe.

Dans le cas des arbres de décision flous, les décisions ne sont plus prises sur un simple seuil mais en calculant une mesure de similarité entre la valeur de l'attribut et un sous-ensemble flou.

Les décisions sont du type :

Si la valeur pour l'attribut **est compatible avec** la modalité 1 **alors** suivre branche 1
Si la valeur pour l'attribut **est compatible avec** la modalité 2 **alors** suivre branche 2

Il est possible que la condition soit satisfaite - au moins à un certain degré - pour plusieurs branches de l'arbre. La propagation de l'exemple à travers l'arbre se fait alors dans plusieurs branches en parallèle et donne des degrés de confiance pour plusieurs feuilles, et donc pour plusieurs classes. Une étape d'agrégation peut ensuite "défuzzifier" le résultat, en déterminant la classe la plus probable, ou garder une représentation floue pour le traitement suivant.

L'arbre de décision flou permet de prendre en compte l'imprécision et l'ambiguïté présentes dans les données. Il n'est pas possible de classer l'ensemble des exemples correctement à partir d'un seul attribut mesuré, puisqu'il y a des zones de recouvrement entre les classes. Dans ces zones de recouvrement, l'arbre de décision flou ne prend pas une décision trop tôt, mais continue à considérer les deux cas possibles.

L'utilisation du modèle de variabilité dans un arbre de décision est représentée dans la figure 6.3. Dans chaque nœud de l'arbre les fonctions d'appartenance sont adaptées à l'image actuelle avant de prendre une décision partielle sur les classes.

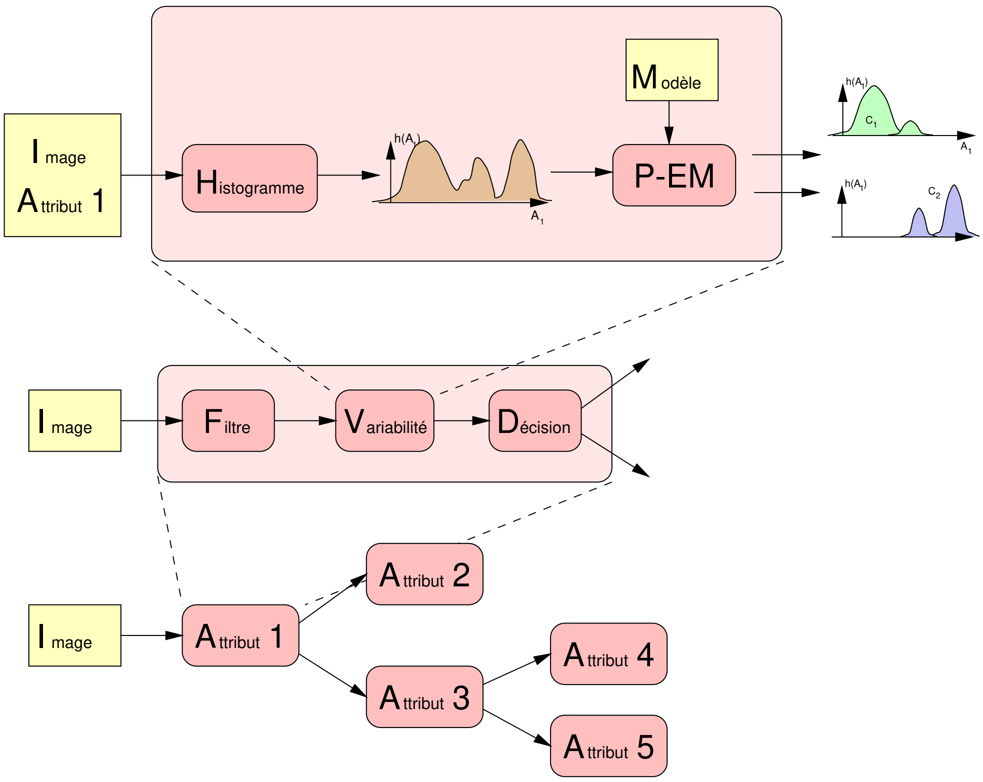


FIG. 6.3 – Utilisation du modèle dans un arbre de décision

6.2 Application à une base synthétisée

Nous avons appliqué la nouvelle méthode proposée à une base de données synthétisée. L'avantage d'une base de donnée synthétisée est la connaissance exacte de sa composition, ce qui permet de vérifier les résultats. Les données que nous avons synthétisées ne représentent pas des images, mais les histogrammes des valeurs d'attributs.

6.2.1 La base de données

La base de données est construite en quatre étapes :

- Définition du modèle caché, notamment : $\mathbf{A}$ la matrice de l'interaction, Φ_0 les centres des prototypes, $\vec{r}$ les variances des prototypes
- choisir un vecteur ε au hasard avec une variance r
- inverser $\mathbf{A} \cdot \Phi = \varepsilon$
- échantillonner le mélange déterminé par Φ ;

En utilisant ce processus, nous retrouvons un ensemble d'histogrammes qui contiennent, pour chacun d'entre eux, l'ensemble des composantes définies par le prototype. Les composantes sont liées statistiquement par la matrice d'interaction $\mathbf{A}$. Elles ne se déplacent pas indépendamment les unes des autres entre les différents histogrammes. Chaque histogramme représente l'histogramme qu'on aurait pu mesurer sur un attribut calculé sur une image.

Une base d'apprentissage de 100 exemples et une base de test de 100 exemples pour la validation ont été construites de cette manière. La figure 6.4 montre les composantes pour 3 exemples de la base. Dans la figure 6.5, les histogrammes échantillonnés sont visualisés.

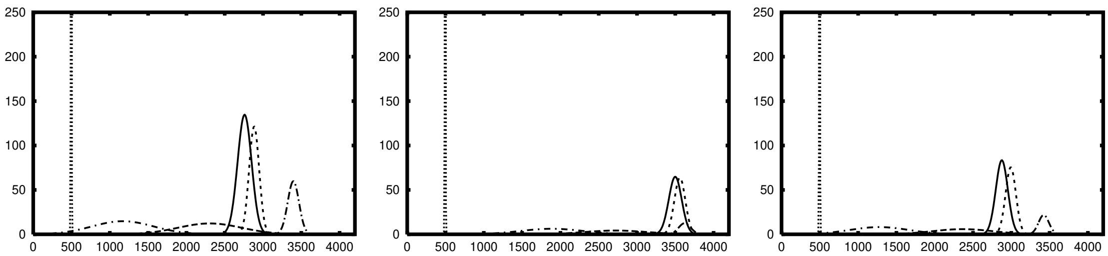


FIG. 6.4 – Les composantes de trois exemples de la base d'exemple

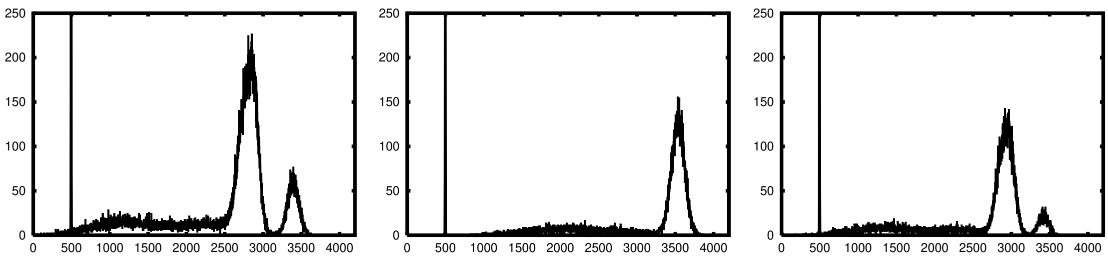


FIG. 6.5 – Les histogrammes correspondants après échantillonnage

6.2.2 Création d'un prototype

Dans la phase d'apprentissage, un prototype est estimé à partir de la base d'apprentissage. Pour cela, les composantes présentes dans tous les histogrammes sont estimées avec l'algorithme H-EM. Ensuite, les composantes estimées sont regroupées avec l'algorithme des C-moyennes floues.

Puis, les interactions sont estimées avec une régression linéaire multidimensionelle dans l'espace de tous les paramètres. Le résultat est une matrice **A**. Le centre du prototype qui correspond à une énergie interne nulle est montré sur la figure 6.6.

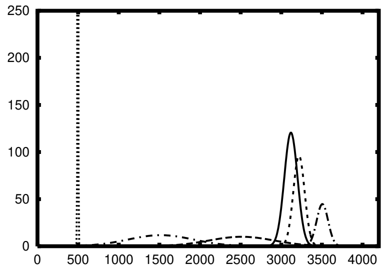


FIG. 6.6 – Le centre du prototype utilisé pour l'initialisation de l'algorithme P-EM

6.2.3 Evaluation sur la base de test

Une fois que le prototype avec les interactions internes a été estimé, l'algorithme P-EM décrit dans le chapitre 4 est utilisé pour adapter le prototype à chaque exemple de la base de test. Les résultats d'adaptation aux trois exemples montrés ci-dessus sont visualisés sur la figure 6.7.

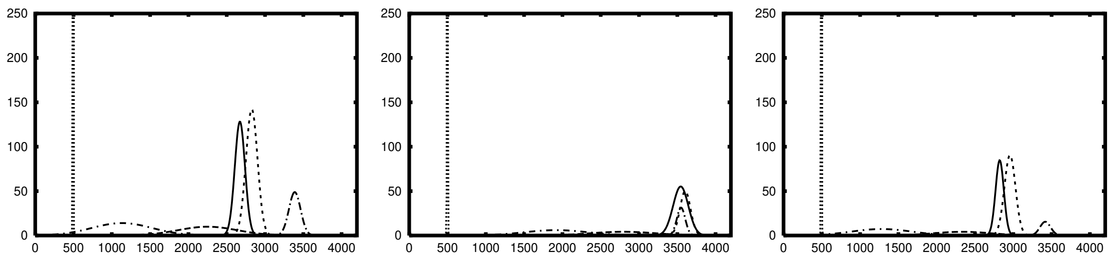


FIG. 6.7 – Résultats après l'adaptation du prototype avec l'algorithme P-EM

On peut observer que les paramètres des composantes retrouvées sont très proches des paramètres des composantes initiales qui ont servi à la génération des histogrammes pour ces trois exemples, malgré le fait que la variabilité soit si importante que, pour l'exemple au milieu de la figure 6.7, les composantes sont très proches les unes des autres.

Finalement, nous avons calculé les courbes COR (annexe A) pour 100 cas de la base de test en utilisant cinq méthodes de classification :

idéale :	Méthode de classification fictive qui utilise la connaissance de la vraie classe pour chaque exemple dans l'histogramme pour la décision. La valeur obtenue donne la limite supérieure de la performance de la classification qui peut être obtenue en utilisant uniquement cet attribut. Cette estimation est légèrement optimiste, car toutes les fluctuations statistiques dans les histogrammes sont exploitées, un effet qui ne peut pas être reproduit dans un système réel.
statique :	Estimation de l'histogramme global à travers tous les exemples de la base. La valeur obtenue donne la limite supérieure des performances obtenues avec une classification statique à partir de la base d'apprentissage. Comme dans le cas **idéal**, la valeur obtenue est légèrement optimiste.
mélange idéal :	Méthode de classification fictive qui utilise la connaissance de la vraie classe pour chaque exemple dans l'histogramme pour la décision. Les histogrammes sont d'abord modélisés avec un mélange de gaussiennes, avant d'être utilisés pour la classification. La différence par rapport à la valeur de performance de la méthode de classification **idéale** donne une idée des pertes qui sont introduites par cette approximation et de la surévaluation introduite par cette première méthode.
mélange statique :	Similaire au cas **statique**, mais avec les histogrammes modélisés par un mélange de gaussiennes avant d'être utilisés pour la classification. Cette valeur donne un résultat réaliste pour un classificateur statique.
mélange adaptatif :	La méthode d'adaptation des fonctions d'appartenance proposée.

Les résultats obtenus montrent (tableau 6.1) que la méthode présentée peut améliorer les résultats de la classification sur cette base par rapport à une classification statique.

idéale	statique	mélange idéal	mélange statique	mélange adaptatif
0,990	0,971	0,989	0,855	0,977

TAB. 6.1 – Surface sous la courbe COR

Les résultats obtenus pour la méthode des fonctions d'appartenance adaptatives sont même meilleurs que pour le cas statique idéal, qui est déjà une estimation très optimiste de ce qui est réalisable avec une méthode de classification statique. L'estimation plus réaliste de la performance d'une méthode de classification statique, qui est donnée dans la colonne 4 du tableau 6.1, montre que le gain potentiel de la méthode proposée est significatif. La différence entre les valeurs **idéales** et les valeurs pour le **mélange idéal** est très faible dans notre cas, puisque les histogrammes ont été générés à partir des mélanges, et nous savons a priori que le modèle de gaussiennes est bien adapté.

6.3 Liens avec les modèles de Markov cachés

La méthode de modélisation de variabilité que nous proposons est liée aux modèles de Markov cachés puisque dans les deux cas il y a une partie observable (les densités dans les histogrammes) et une partie *cachée* (par exemple le temps d'exposition) qui a une influence sur la partie observable. Dans cette section, nous décrivons les modèles de Markov cachés pour établir le lien avec nos travaux.

Les modèles de Markov cachés sont utilisés dans de nombreuses applications pour décrire et analyser des signaux stochastiques. L'idée de base des modèles de Markov cachés est qu'il existe, dans le système à analyser, une chaîne de Markov X_n qui décrit l'évolution du système, mais dont l'état actuel x_n n'est pas accessible. L'état actuel est uniquement accessible au travers des observations Y_n qui sont conditionellement dépendantes de l'état actuel.

Si les variables Y_n sont discrètes, on parle *des modèles de Markov cachés discrets* (ou *discrete hidden Markov models*, DHMM en anglais). Si les variables Y_n sont continues, on parle *des modèles de Markov cachés continus* (ou *continuous hidden Markov models*, CHMM en anglais).

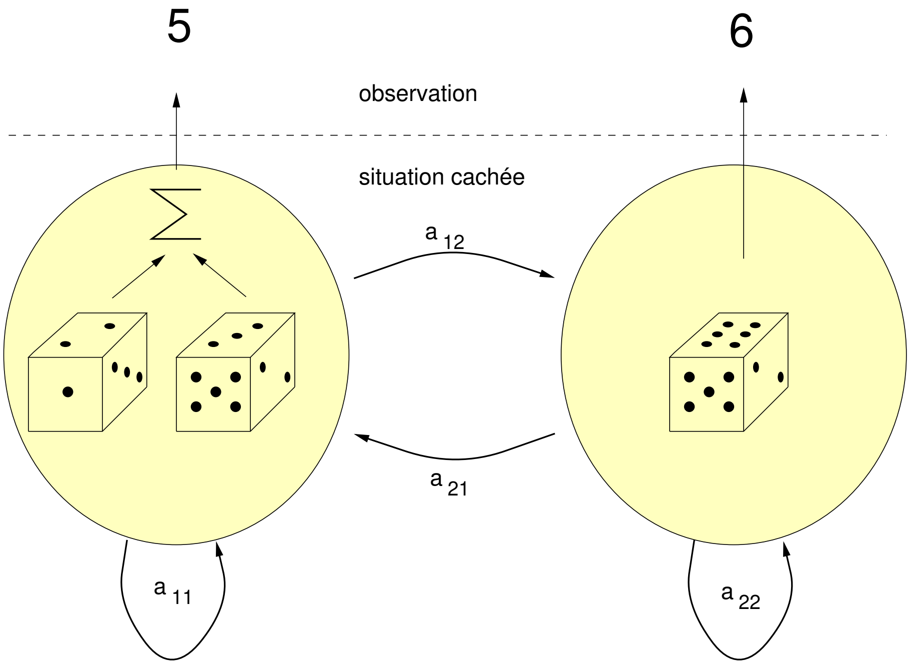


FIG. 6.8 – Modèle de Markov caché discret

La figure 6.8 illustre un exemple de modèle Markov caché discret. Le système caché a deux états : dans le premier il contient un dé, dans le deuxième il contient deux dés. Le nombre de dés n'est pas accessible à l'utilisateur, qui ne peut observer que la somme totale des valeurs de tous les dés Y_n. Les probabilités de transition entre les états (un dé/deux dés) sont représentées par la matrice A :

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \tag{6.1}$$

La figure 6.9 illustre un exemple de modèle Markov caché continu. Le système caché est une personne qui parle. Sa voix est transmise par un canal bruité. Le but du récepteur est de retrouver la parole envoyée. Le signal reçu a une distribution continue qui change en fonction de l'état du système. Si la personne prononce un “a”, la fréquence de la voix n'est pas la même que lorsqu'elle prononce un “u”. Autour de cette fréquence nominale, le récepteur perçoit une distribution qui vient du bruit de la transmission. La matrice de transition **A** contient les probabilités de passage entre les différentes syllabes.

6.3.1 Des mélanges interprétés comme modèles de Markov cachés

Un mélange de fonctions paramétriques peut être considéré comme un modèle de Markov caché où l'état interne caché représente le numéro de la composante de laquelle l'exemple actuel est tiré. Cette information n'est pas accessible dans le cas des fonctions d'appartenance adaptatives. La matrice de transition est composée de la fréquence d'apparition p_i de chaque composante.

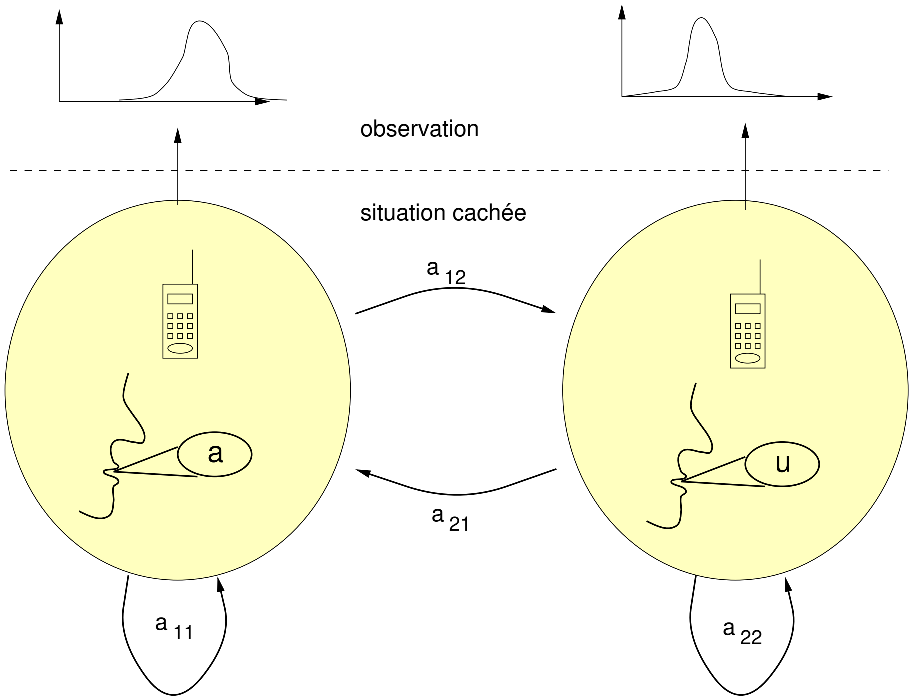


FIG. 6.9 – Reconnaissance vocale : modèle de Markov caché continue

$$\mathbf{A} = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \\ \vdots & \vdots & & \vdots \\ p_1 & p_2 & \cdots & p_n \end{bmatrix} \tag{6.2}$$

L'algorithme le plus utilisé pour l'identification des paramètres des modèles de Markov cachés, l'algorithme *Baum-Welch*, est une implémentation spécifique de l'algorithme EM que nous avons utilisé pour l'estimation des paramètres des mélanges.

Les propriétés de l'algorithme EM et des modèles de Markov ont été analysées par plusieurs auteurs (une synthèse se trouve dans [Cou96-1]). Grâce à l'équivalence entre les modèles de Markov et les mélanges de composantes, toutes les propriétés des modèles de Markov s'appliquent aussi au problème de l'estimation des composantes avec l'algorithme EM.

La modélisation des liens entre les composantes, que nous avons proposée en utilisant une régression, peut être interprétée comme l'estimation d'une variable cachée supplémentaire qui caractérise l'image dans son ensemble. Cette variable en elle-même modifie la position et la forme des densités de probabilité de chaque composante.

6.4 Conclusion

Dans cette première partie, nous avons discuté des méthodes de construction des fonctions d'appartenance pour la classification. Nous nous sommes spécialement intéressés aux problèmes de classification appliqués au traitement d'images pour lesquels la variabilité entre images est importante. Nous avons montré que la variabilité peut dégrader sérieusement le pouvoir de discrimination des attributs quand elle est prise en compte uniquement comme incertitude pendant la phase de construction des fonctions d'appartenance.

Nous avons proposé une nouvelle approche de l'identification des fonctions d'appartenance qui permet d'adapter ces fonctions à chaque image. L'approche proposée permet de modéliser la variabilité entre les images et ce faisant, de réduire l'incertitude des sous-ensembles flous. Cela permet de trouver des résultats plus pertinents et d'utiliser au maximum l'information sur les classes contenues dans les attributs servant à la classification.

La modélisation des fonctions d'appartenance adaptatives passe par une modélisation des densités dans les histogrammes des attributs mesurés sur l'image. Les paramètres de ce modèle sont identifiés à partir d'une base d'apprentissage par l'intermédiaire d'une régression. Une fois que le modèle est identifié, la fonction paramétrique peut être adaptée au cas actuel. Finalement, la fonction paramétrique est convertie en fonction d'appartenance dans une étape de normalisation. Nous avons discuté de plusieurs méthodes de normalisation, dont la méthode probabiliste qui est la plus connue, et proposons deux nouvelles méthodes, l'une qui se base sur l'intégrale de la fonction paramétrique, l'autre sur la détectabilité. Les deux méthodes se prêtent bien à des problèmes d'apprentissage à partir de contours fournis par des experts sur les images de la base d'apprentissage.

Finalement, nous appliquons la méthode proposée à une base de donnée simulée, ce qui permet de comprendre son fonctionnement et illustre le gain potentiel qu'elle peut apporter.

Dans la deuxième partie, nous discutons de l'application des méthodes de réduction de la variabilité à des images de mammographie.

Deuxième Partie

Application à un système de détection des signes radiologiques en mammographie numérique

Chapitre 7

Mammographie et cancer du sein

Dans ce chapitre, nous allons présenter l'approche médicale de l'imagerie du sein, pour situer le contexte dans lequel l'application que nous cherchons à développer va être utilisée. La présentation de l'anatomie du sein et des changements bénins et malins qui peuvent se produire, n'est pas exhaustive et est donnée uniquement dans le but de faciliter la compréhension de la problématique pour en déduire des méthodes de détection et de classification. La littérature de l'aspect médical du cancer du sein est très vaste ([Hey97] [Gra93] [Lam81] [Lan86][Tra95] [Tab85] [Tre89] ...), et l'ensemble des publications ne représente que la partie linguistiquement exprimable de la connaissance actuelle sur ce sujet. L'expertise des médecins, acquise pendant des années d'expérience, est difficilement transmissible par ces derniers, mais elle est pourtant souvent aussi riche, voire plus riche que les publications.

Concevoir un système de détection et de classification automatique des lésions en mammographie, ne doit pas être compris comme une tentative de remplacer le médecin. La question n'est pas de se substituer au médecin, mais de lui proposer des outils performants qui l'aident dans son travail d'analyse des clichés de mammographie.

7.1 L'anatomie du sein

Une représentation simplifiée de l'anatomie du sein est donnée en figure 7.1. Les structures discernables de l'extérieur vers l'intérieur du sein sont :

- la peau qui englobe le sein
- le ligament de Cooper avec les crêtes de Duret
- la matrice conjonctive qui contient :
  - · les galactophores
  - · les lobules
  - · les artères
  - · les veines
  - · les nerfs
  - · les ganglions
- le pectoral et les côtes

Pour expliquer le fonctionnement du sein, il faut regarder l'un des dix à vingt canaux galactophoriques qui partent du mamelon. La structure simplifiée d'un canal galactophorique est montrée sur la figure 7.2. En partant d'un orifice galactophorique au niveau du mamelon, le canal galactophorique principal se dilate d'abord pour former une cavité appelée sinus lactifère. Ensuite, il rétrécit et se sépare en une arborescence de moyens et de petits canaux galactophoriques qui se terminent dans des lobules.

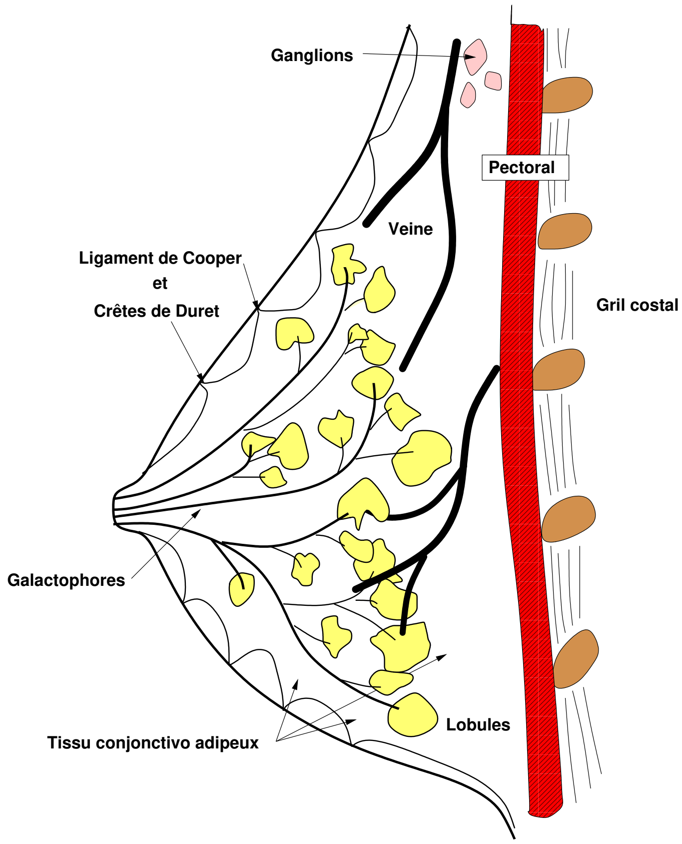


FIG. 7.1 – Anatomie du sein

Un agrandissement d'une coupe au travers d'un canal galactophorique est représenté schématiquement dans la figure 7.3. L'intérieur du canal galactophorique est tapissé par les cellules épithéliales qui sont responsables de la fonction sécrétoire. A l'extérieur de ce cylindre, il y a une membrane de cellules myo-épithéliales et du tissu conjonctif. Cet ensemble est lui-même entourée d'une couche constitué des cellules adipeuses. Ce canal galactophorique est situé dans le tissu conjonctif de soutien.

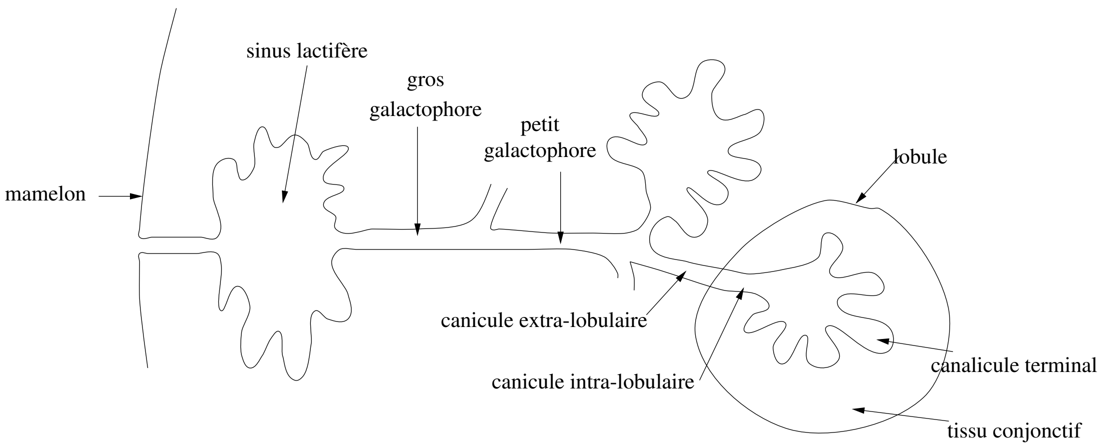
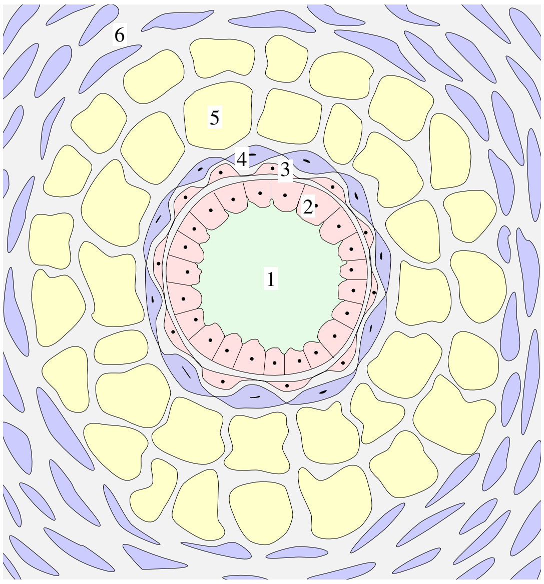


FIG. 7.2 – Schéma d'un galactophore



1 - lumière du galactophore
2 - cellules épithéliales
3 - cellules myo-épithéliales
4 - tissu conjonctif
5 - cellules adipeuses
6 - tissu conjonctif de soutien

FIG. 7.3 – Agrandissement d'un galactophore

7.2 La mammographie

La mammographie réalise une projection du volume du sein sur le plan de l'image. Elle permet d'analyser la glande mammaire grâce aux différences de l'atténuation des différents types de tissu. L'avantage de la visualisation de l'ensemble des tissus mammaires sur une seule image est directement lié à l'un de ses plus grand défauts, la superposition des différents tissus traversés par le même rayon et projetés sur un seul endroit du détecteur. Cette superposition est une source d'incertitude, puisqu'il n'est plus possible de distinguer sur l'image, des sur-densités réelles qui correspondent à une région radio-opaque dans le tissu en trois dimensions, et des superpositions de plusieurs tissus avec des densités relativement faibles.

7.2.1 Corrélation entre l'anatomie et les images de la mammographie

Avant de pouvoir utiliser les connaissances de l'anatomie et de l'histologie du sein dans l'interprétation des images radiographiques, il est nécessaire d'établir le lien entre la structure des tissus et l'atténuation apportée aux rayons X.

Puisque l'interaction des rayons X avec la matière est principalement déterminée par les couches intérieures des électrons dans la structure atomique, elle est presque indépendante des liaisons chimiques qui jouent sur les couches externes. L'atténuation pour des composants chimiques et des structures biologiques peut donc être calculée à partir des atténuations de chaque élément et de la densité de ceux-ci dans le composant ou la structure. Pour les matières prédominantes dans le sein, nous obtenons le tableau de correspondance suivant :

composant	atténuation radiologique
graisse	radio-transparent
eau	légèrement radio-opaque
tissu conjonctif, fibres	radio-opaque
calcium	très radio-opaque

Un autre façon de faire le lien entre la réalité physique et son aspect radiologique est l'approche expérimentale qu'utilise Lamarque [Lam81] en comparant l'analyse histologique et les images obtenues par micro-radiographie pour des exemples de tissus. Il donne le tableau de correspondance suivant :

structures histologiques	traduction radiologique
épithélium galactophorique	peu perceptible
tissu conjonctif	opacité dense homogène
fibrose	opacité linéaire dense
tissu graisseux	image claire

7.2.2 Les incidences en mammographie

Selon la partie du sein à laquelle s'intéresse l'examen, différentes incidences sont utilisées. Les incidences les plus courantes sont l'incidence oblique externe (ou medio-lateral-oblique) et l'incidence de face (ou cranio caudal).

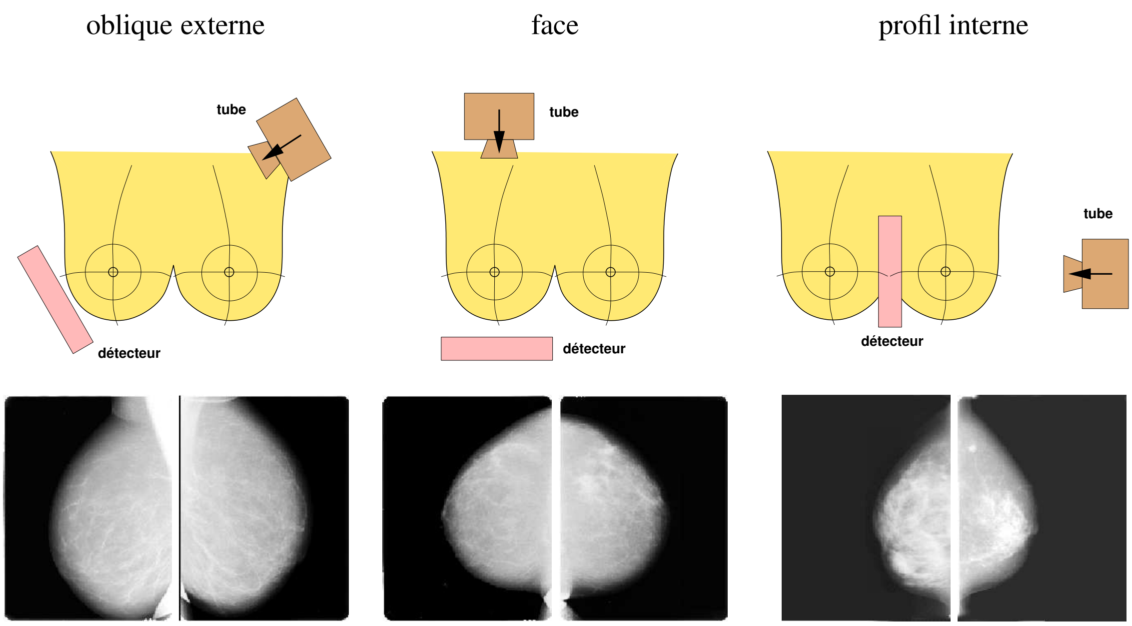


FIG. 7.4 – Incidences en mammographie

7.2.3 Aspect normal

L'aspect du sein normal est très variable d'une femme à l'autre. Le facteur le plus remarquable est la grande variabilité de la densité radiologique de l'aire mammaire. Dans ses travaux, Wolfe [Wol76] propose une classification des types de seins en 4 classes (voir tableau 7.1)

N1	très peu de tissu dense
P1	les éléments denses occupent moins du quart de l'aire mammaire
P2	les structures denses occupent plus du quart de l'aire mammaire
DY	l'aire mammaire est globalement dense

TAB. 7.1 – Classification des tissus par Wolfe

Wolfe essaie de lier la densité du tissu au risque de développer un cancer, mais d'autres études mettent en doute ce lien [Tre89] ; néanmoins cette classification des tissus est utilisée dans d'autres travaux [Suc95][Mil92] sans faire de lien au risque.

7.3 Pathologies mammaires

Dans cette partie, nous allons montrer quelques exemples de pathologies mammaires les plus fréquentes. Nous allons en même temps expliquer les modifications de structures des tissus qui sont liées à ces pathologies, et comment ces modifications se traduisent sur l'image de la mammographie.

7.3.1 Kystes

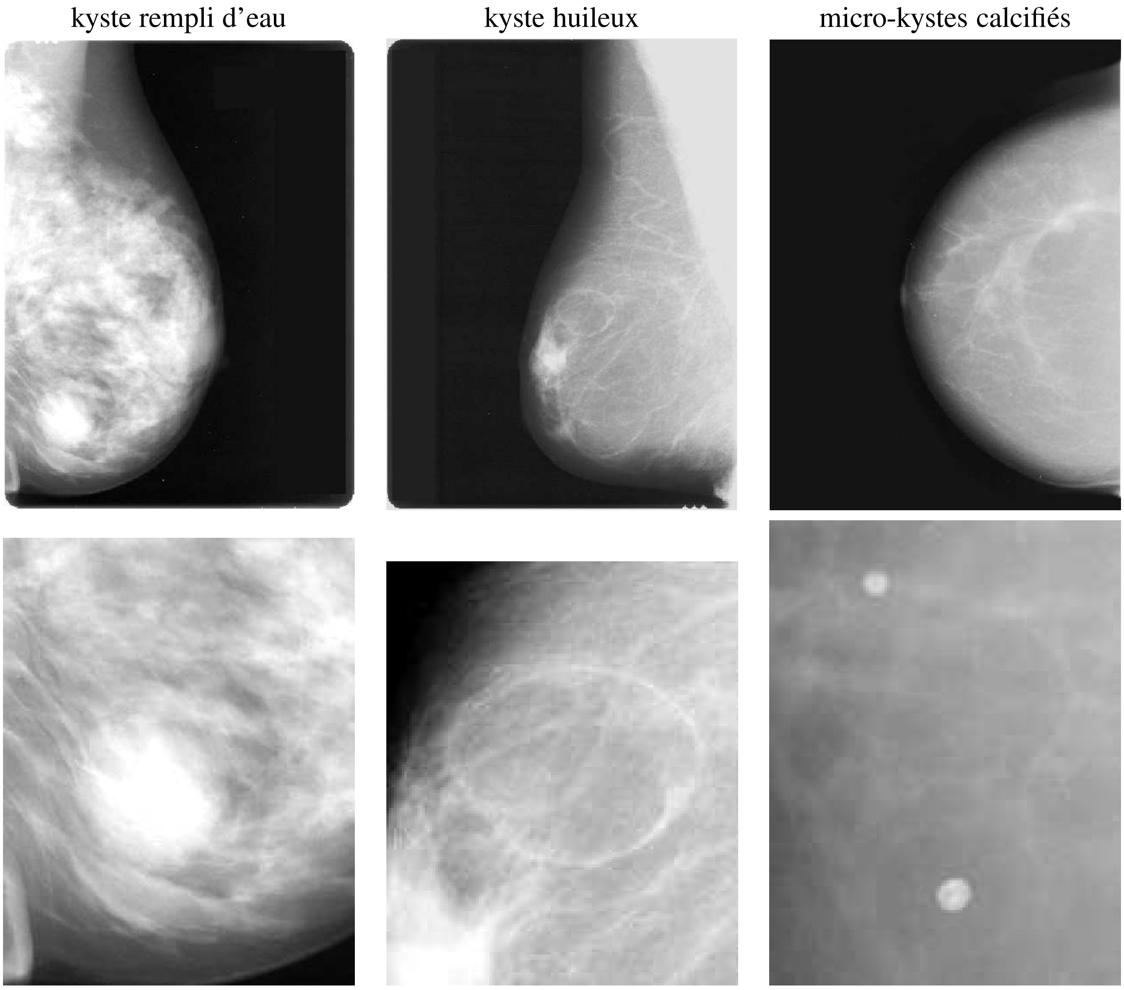


FIG. 7.5 – Différents exemples pour des kystes

Si l'un des canaux galactophoriques est obstrué et qu'un segment du canal ou un lobule n'a plus de sortie, cette cavité peut se remplir avec un liquide. Ce remplissage peut exercer une pression qui gonfle la membrane du canal galactophorique ; la cavité devient alors sphérique ou ovale. Selon le type de liquide qui s'est accumulé dans le kyste, l'atténuation de cette lésion est plus élevée que celle du tissu environnant (kyste rempli d'eau) ou moins élevée que le tissu environnant (kyste huileux)

Les kystes se traduisent presque toujours, sur les clichés de mammographie, par des opacités circonscrites avec des bord nets.

Parfois les cellules du tissu conjonctif qui entourent l'épithélium sont le siège d'un dépôt d'oxalate ou de phosphate de calcium. Dans ce cas on parle de kystes calcifiés. Puisque l'atténuation du calcium est beaucoup plus élevée que celle des autre tissus, les kystes calcifiés sont généralement faciles à détecter.

7.3.2 Fibro-adénome

Le fibro-adénome est une lésion à croissance lente qui est constituée de cellules épithéliales et de cellules conjonctives. La densité radiologique et la forme ovale ou ronde sont comparables à celles d'un kyste rempli d'eau. La distinction d'avec celui-ci n'est pas toujours possible à partir d'un cliché de mammographie.

fibro-adénome fibro-adénome calcifié

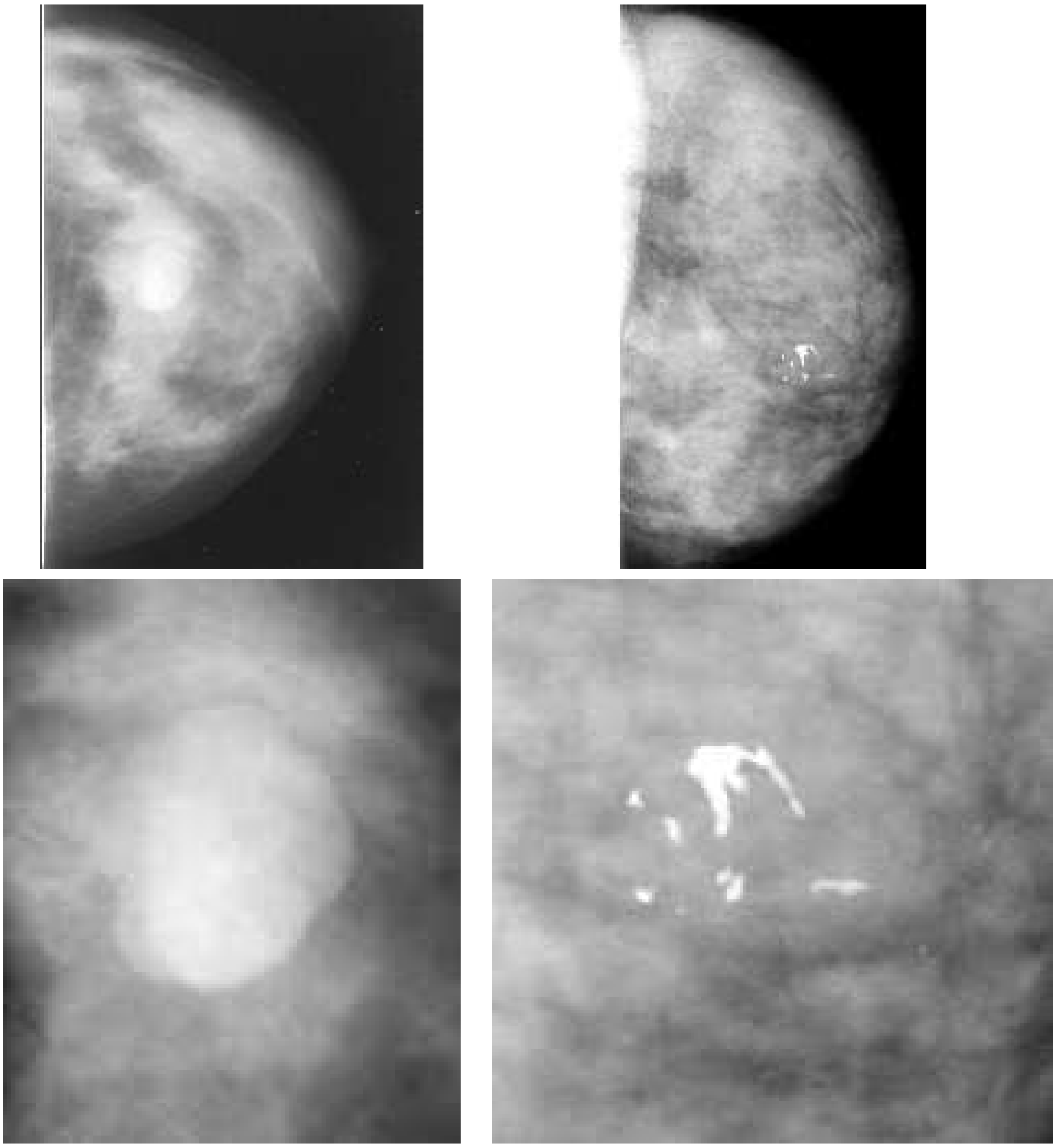


FIG. 7.6 – Deux exemples de fibro-adénomes

Parfois un dépôt de calcium est observé sur ce type de lésion. Dans ce cas, la distinction est possible sur un cliché de mammographie, puisque la calcification d'un fibro-adénome prend une allure caractéristique en “pop-corn”.

7.3.3 Comédo-carcinome

Cette lésion consiste en des cellules cancéreuses qui se propagent à l'intérieur des canaux galactophoriques. Radiologiquement, aucune lésion centrale ne peut être distinguée, puisque le cancer à l'intérieur des canaux ne provoque pas de réaction fibreuse. Le comédo-carcinome n'est visible sur une mammographie que grâce aux calcifications intra-canalaires réparties dans une partie de l'arbre galactophorique. Les calcifications peuvent évoluer d'une forme initiale semblable à des grains de sel jusqu'à un remplissage complet des canaux galactophoriques.

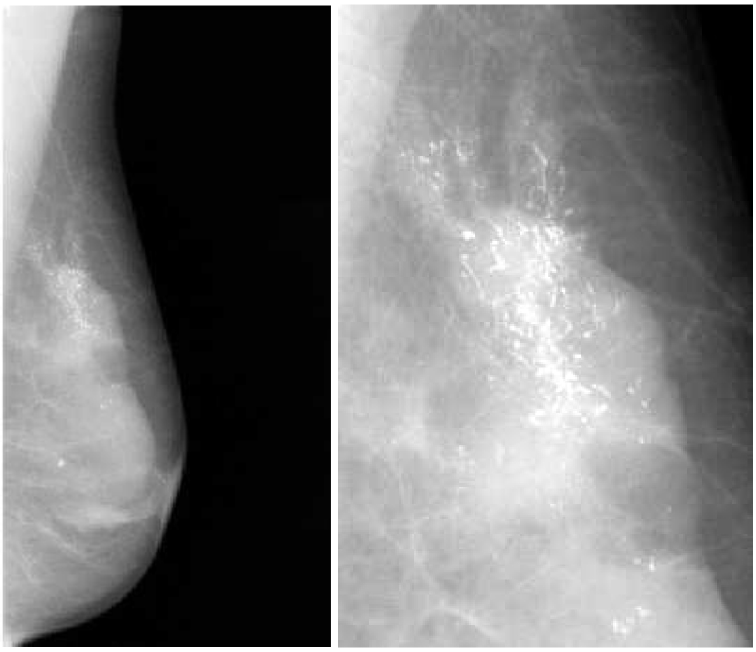


FIG. 7.7 – Exemple d'un comédo-carcinome

7.3.4 Cancer infiltrant

Les cancers infiltrants regroupent environ 70% des cancers du sein. Les cellules cancéreuses se multiplient et forment des amas ou des tubes. Elles sont visibles sur une mammographie grâce à la réaction fibreuse et, dans certains cas grâce aux calcifications. Parmi les cancers infiltrants, trois classes peuvent être distinguées :

- l'épithélioma infiltrant évolué
- le squirrhe
- le cancer colloïde

Pour des raisons didactiques, les exemples montrés dans figure 7.8 sont des cas facilement reconnaissables. Dans la pratique quotidienne du radiologue, la détection et l'identification de ces cas ne représentent aucune difficulté. Les lésions des cancers infiltrants au début de leur développement sont peu visibles et donc difficiles à distinguer du tissu normal.

épithélioma infiltrant évolué squirrhe

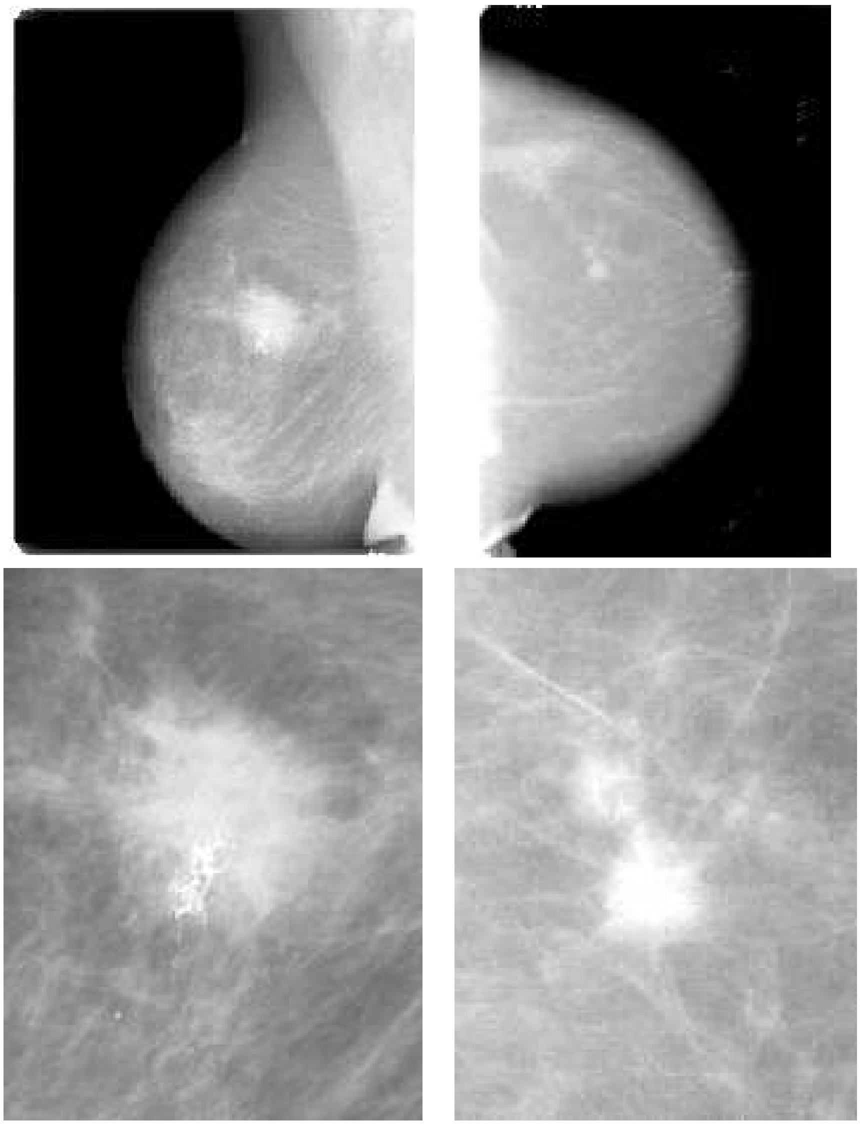


FIG. 7.8 – Exemples de cancers infiltrants

7.3.5 Autres pathologies

Il existe un grand nombre d'autres types de pathologies bénignes ou malignes que nous n'allons pas décrire ici, mais qui doivent être prises en compte lors de la conception d'un système d'aide au diagnostic (pour une description des pathologies plus rares, on pourra consulter [Tab85][Lam81][Tre89] ou les nombreux articles dans la littérature spécialisée).

7.4 Signes radiologiques

Dans la section précédente, nous avons essayé de caractériser les différents types de lésions et d'expliquer leurs représentations sur les clichés radiographiques. Dans [Tre89], les auteurs orientent leurs analyses vers l'aspect radiologique et décrivent les interprétations possibles. Ils

distinguent trois grands groupes de signes radiologiques en relation avec le cancer :
- les opacités arrondies
- les opacités stellaires
- les calcifications

7.4.1 Opacité arrondie

Il y a un grand nombre de structures qui peuvent se traduire sur les images de mammographie par des opacités arrondies :

Opacité arrondie solitaire	
Structures normales	**Pathologies rares**
formation cutanée (verrue, naevus)	Papillome (intra-kystique ou intra-canalaire)
Mamelon non tangentiel	Tumeur phyllode
Ganglion lymphatique normal	Harmatome
Veine vue en coupe	Métastase d'un carcinome
Lobule glandulaire circonscrit	Localisation d'un lymphome
	Adénome vrai
	Abcès
	Hématome
	Galactocèle
Pathologies fréquentes	**Pathologies exceptionnelles**
Kyste	Sarcome
Fibro-adénome	Forme nodulaire de tuberculose
Cancer de forme nodulaire	Hémangiome
	Ganglion lymphatique métastatique
	Neurofibrome

Opacités arrondies multiples
Kystes
Fibro-adénomes
Cancer nodulaire multifocal
Métastases d'un carcinome
Localisation d'un lymphome
Galactocèles

7.4.2 Opacités stellaires

Les opacités stellaires sont généralement associées avec une lésion cancéreuse, bien qu'il existe d'autres structures bénignes qui peuvent donner lieu à un signe stellaire sur une mammographie. La liste suivante donne les lésions le plus courantes avec le traitement qu'on leur associe typiquement :

– **lésion construite**
  · pas de formation palpable
  · centre dense absent
  · vue sur une seule incidence
  · disparaît sur cliché localisé ou sur incidence variable

- **lésion post-opératoire**
  · antécédents chirurgicaux
  · cicatrices
  · évolution clinique
  · évolution mammographique
  → surveillance
- **carcinome**
  · image stellaire évidente
  · clinique évident (mais formes infra-cliniques)
  → cytoponction
  → vérification histologique impérative
- **mastopathie bénigne**
  · rare
  · structure stellaire non ou difficilement palpable
  · centre dense petit ou inexistant
  · modifications cutanées rares
  → cytoponction
  → vérification histologique
- **lésion inflammatoire**
  · clinique évocatrice
  · évolution sous traitement
  → cytoponction
  → surveillance ou drainage

7.4.3 Microcalcifications

Les microcalcifications sont formées par un dépôt de matériaux fortement radio-opaques composés de $Ca_3(PO_4)_2$, $CaCO_3$, ou $Mg_3(PO_4)_2$ dans les tissus. Les calcifications rencontrées sur les clichés de mammographie peuvent être d'origines variées. Il existe :
- des calcifications intra-canalaires liées à un comédo-carcinome
- des calcifications comme réaction à un cancer infiltrant
- des kystes calcifiés
- des fibro-adénomes calcifiés
- des calcifications vasculaires
- des calcifications comme conséquence d'intervention chirurgicale
- des calcifications liées à un traumatisme
- des calcifications cutanées

Certains types de calcifications ont une forme ou un groupement bien caractéristiques, d'autres se ressemblent. Les calcifications intra-canalaires peuvent se présenter d'une manière très caractéristique quand elles remplissent comme une moulage l'espace dans les canaux galactophoriques. La détection de ces microcalcifications est importante car elles représentent le premier signe détectable sur une mammographie pour certains types de cancer (comédo-carcinome). Elles ne restent cependant qu'un signe indirect des pathologies sous-jacentes. On notera aussi que beaucoup de calcifications sont observées dans des seins qui ne présentent aucune pathologie.

7.5 Le cancer du sein

Le mot *cancer* est un terme général utilisé pour toutes sortes de tumeurs malignes qui s'étendent rapidement. Bien que ses causes exactes ne soient pas encore connues à ce jour, on suppose que le cancer se développe quand une cellule, dont l'information génétique a été modifiée, se multiplie rapidement, sans que les mécanismes de défense de l'organisme arrivent à la mettre sous contrôle.

Les causes de modifications de l'information génétique sont dues, par exemple, à des substances cancérigènes ou à des irradiations. De plus, il existe des facteurs de risque qui augmentent la probabilité de développer un cancer.

Grâce aux avancées médicales dans tous les domaines, l'espérance de vie augmente constamment et, avec d'autres facteurs liés au changement de mode de vie et de l'environnement, cela augmente la probabilité de développer un cancer au cours de la vie. En même temps, la probabilité d'apparition d'un cancer varie beaucoup selon la localisation et le type du cancer.

Dans la plupart des pays, un grand effort est fait aujourd'hui pour obtenir des statistiques fiables, afin d'en savoir plus sur les causes et les facteurs de risques du cancer.

De telles statistiques sont présentées dans les figure 7.9 et 7.10. Elle montrent la mortalité et le nombre de nouveaux cas, pour les différents types de cancer chez les femmes en France. Ces statistiques montrent qu'en France, le cancer du sein est de loin le cancer le plus courant chez la femme et qu'il est aussi la première cause de mortalité par cancer. On remarque également une forte croissance des incidences du cancer du sein dans la période 1975-1995 sur laquelle porte la statistique montrée sur le figure 7.9.

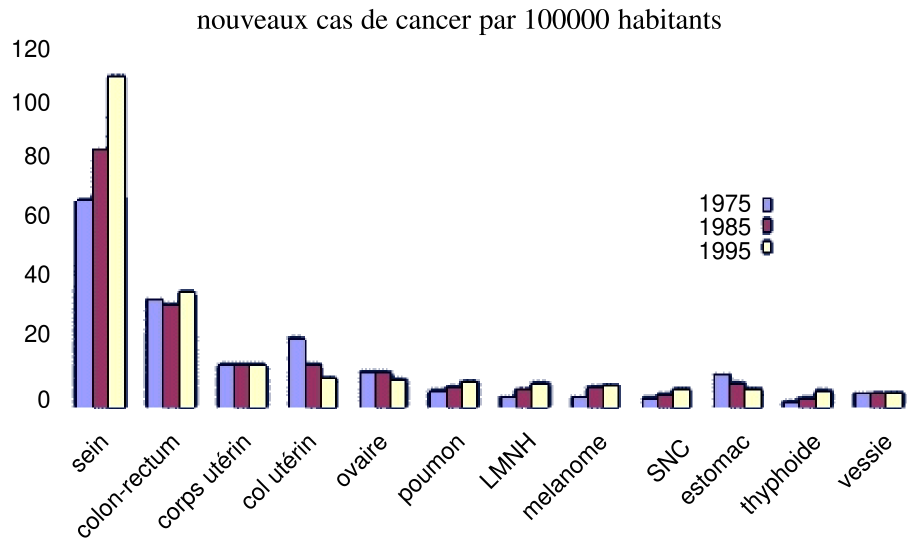


FIG. 7.9 – Evolution de la répartition de nouveaux cas pour les principaux sièges ou types de cancer chez la femme en France (source : Institut de Veille Sanitaire [Che95])

En 1990, environ 26000 cas de cancer du sein ont été enregistrés, ce qui représente environ 30% des incidences de cancer chez les femmes. Plus de 10000 femmes sont décédées à cause du cancer du sein, ce qui représente 18% de la mortalité par cancer. A l'échelle de chaque personne, cela représente un risque de 10% qu'une femme développe un cancer du sein au cours de sa vie.

La chance de survie après dix ans, tous cancers du sein confondus, est de l'ordre de 50%. Par contre, si le cancer est détecté très tôt dans son développement, c'est-à-dire quand il n'est pas encore envahissant, le taux de guérison après traitement devient proche de 100%. Cela explique l'importance de sa détection précoce.

Les méthodes de détection sont l'autoexamination par la femme elle-même et des examens réguliers par le médecin. Les outils principaux du médecin, autre l'examen clinique, sont les outils de l'imagerie du sein :

– la mammographie (imagerie par rayons X)
– l'échographie (imagerie par ultrasons)
– l'imagerie par résonance magnétique (IRM)

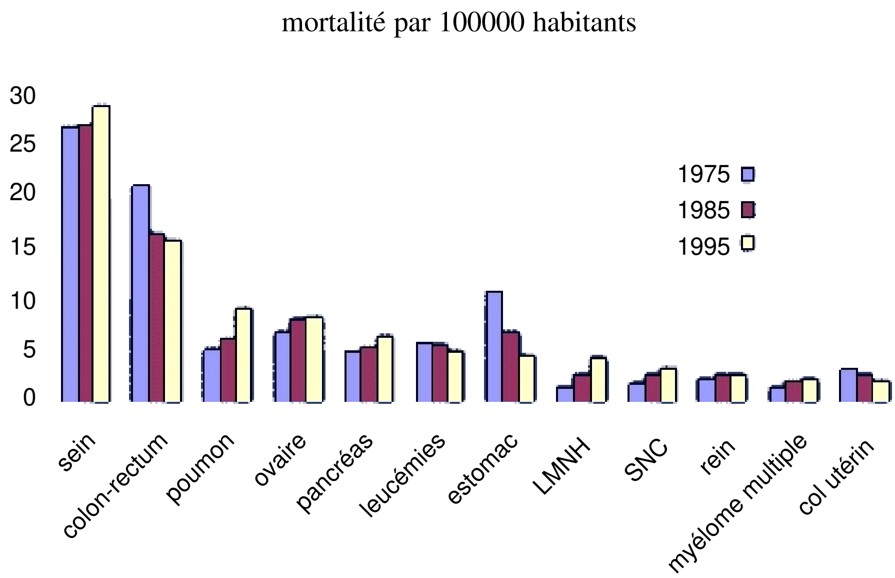


FIG. 7.10 – Evolution de la répartition de décès pour les principaux sièges ou types de cancer chez la femme en France (source : Institut de Veille Sanitaire [Che95])

Parmi ces modalités, la mammographie est la mieux adaptée au dépistage, parce qu'elle permet d'examiner l'ensemble du tissu mammaire avec une ou deux incidences seulement. A un instant donné, l'échographie ne permet la visualisation que d'une coupe du tissu. Il est ainsi difficile d'assurer que le sein ait été examiné dans sa totalité à l'issue d'un examen. Il est possible de manquer une petite lésion, d'autant plus s'il s'agit d'amas de microcalcifications, l'un des signes révélateurs pour certains types de cancers pour lesquels l'échographie ne donne pas d'images aisées à interpréter.

Contrairement à l'examen par IRM qui est très coûteux, les coûts et la durée de l'examen mammographique permettent une utilisation générale dans le cadre d'une campagne de dépistage.

Dans de nombreux pays, des campagnes de dépistage du cancer du sein sont mises en œuvre par les responsables de la santé publique en proposant des mammographies systématiques et gratuites aux femmes dont l'âge est compris entre 40-50 ans et 70 ans, selon les pays.

7.6 Les examens mammographiques

La mammographie est un examen de radiographie qui consiste à obtenir une projection du sein par des rayons X, sur un détecteur plan. Outre son utilisation en dépistage du cancer du sein, elle est aussi pratiquée pour le diagnostic ou pour la localisation lors d'interventions chirurgicales (ponctions).

7.6.1 Dépistage

Aujourd'hui, la mammographie est l'outil principal pour le dépistage du cancer du sein. Dans les campagnes de dépistage, deux ou quatre images sont acquises par patiente, à raison de une ou deux par sein. Le fait de choisir deux incidences par sein, permet de réduire le nombre de lésions non détectées, car certaines lésions ne sont pas toujours visibles sur une seule vue. Les images obtenues sont généralement lues par deux radiologues et, en cas de doute, des examens supplémentaires peuvent être demandés.

7.6.2 Diagnostic

La mammographie est aussi utilisée en diagnostic, lorsqu'il y a déjà une suspition à la suite d'une campagne de dépistage, ou parce que la patiente elle-même a détecté une anomalie. Dans ces conditions, le but de la mammographie est souvent d'analyser plus précisément une lésion détectée cliniquement ou de chercher un signe radiologique dans une zone suspecte. L'objectif de la mammographie devient alors la classification d'une lésion plutôt que sa détection. Une alternative à la mammographie, l'échographie, est souvent utilisée en diagnostic, parce qu'elle permet de distinguer certains types de lésions qui ne peuvent pas toujours être distinguées sur un cliché de mammographie.

7.6.3 Interventions

Une fois qu'une lésion a été détectée, par exemple lors d'une campagne de dépistage, il est souvent nécessaire de vérifier par d'autre méthodes si elle est bénigne ou maligne. Si la distinction n'est pas possible à partir des seules images, cette vérification se fait en prélevant un petit échantillon de tissu par voie percutanée dont l'histologie sera étudiée. Pendant le prélèvement de l'échantillon par aiguille, la localisation de la lésion à échantillonner est effectuée sous échographie ou à l'aide de la méthode de stéréotaxie ; cette dernière consiste à prendre deux clichés de mammographie sous des angles différents. Si la lésion est visible sur les deux vues, il est possible de déterminer sa position en trois dimensions en se basant sur la géométrie connue de l'appareil de stéréotaxie.

7.6.4 Mammographie digitale

Récemment, les avancées technologiques dans le domaine des couches minces ont permis de mettre en œuvre des capteurs d'image de grande taille, sensibles aux rayons X. Ces capteurs ont rendu possible la réalisation des premiers systèmes de radiologie digitale [Mul99]. Pour cela les rayons X n'exposent plus un film, mais une matrice de photodiodes gravées sur un panneau de la taille des films. A ce jour, il existe des panneaux de $40cm \times 40cm$ avec plusieurs millions de pixels. Le panneau utilisé pour la mammographie à GE, duquel proviennent quelques unes des images utilisées dans cette thèse, a une dimension de $19cm \times 23cm$ avec une résolution de $100\mu m$.

Les avantages de cette nouvelle technologie pour les patientes et les médecins sont énormes, car les images sont disponibles juste après la pose, sans attendre le développement qui était nécessaire pour le film. En dépistage et en diagnostic, l'image numérique est disponible sur une station de revue qui permet d'aider le médecin à prononcer son diagnostic, et cela dans les plus brefs délais. En cas de doute, des examens supplémentaires peuvent être envisagés sans que la patiente ait besoin d'être reconvoquée. D'autres avantages de l'utilisation d'images numériques sont la facilité d'archivage, de transmission par réseau (télé-médecine) et de duplication. En stéréotaxie, l'accélération de la procédure grâce à la technologie numérique permet de réduire l'inconfort, car le sein de la patiente doit rester sous compression pendant toute l'intervention pour éviter tout mouvement.

Dans le contexte de la mammographie numérique, les méthodes d'aide à la détection et à la classification deviennent de plus en plus importantes. Une fois la chaîne d'acquisition entièrement numérique, l'introduction d'un système d'aide au diagnostic ne nécessite aucun acte supplémentaire pour le médecin (tel l'introduction d'un film dans un numériseur) et permet donc pour la première fois un vrai gain d'efficacité.

7.6.5 Variabilité dans les images de mammographie

Toutes les méthodes de traitement d'images en mammographie sont confrontées à une grande variabilité dans ces types d'images. Les causes principales de cette variabilité sont :

– les conditions physiques de l'acquisition (kV, mAs, pistes, géométrie)
– les variations anatomiques normales à l'intérieur du sein
– le changement d'incidence et de compression
– les variations entre les seins de la même patiente
– les variations normales entre patientes (âge, taille)
– les variations liées à une maladie (processus cancéreux)

Pour des systèmes de détection de lésions, seule la dernière variation est recherchée, mais elle est masquée par toutes les autres sources de variations. Grâce aux systèmes numériques d'acquisition, un certain nombre de sources de variabilité liée au mode d'acquisition peut être identifié et traité. Nous avons présenté une telle approche pour la détection des microcalcifications dans le chapitre 10 et dans [Ric99-1].

Malgré ces avancées, une des sources principales des variations - la variabilité normale des structures des tissus entre patientes - reste à traiter. Nous utilisons la modélisation de la variabilité par des fonctions d'appartenance adaptatives, présentée dans la première partie de ce document, pour traiter cette variabilité.

Chapitre 8

Aide au diagnostic en mammographie

Comme nous l'avons indiqué dans l'introduction, l'exploration du tissu mammaire par mammographie se fait dans différents contextes. Dans le cas du dépistage, la lecture des clichés se fait généralement par deux radiologues qui n'ont pas d'autres informations sur la patiente que les clichés. Cette situation est la plus difficile au niveau de la détection des lésions, puisqu'aucune aide n'est apportée pour compléter l'information contenue dans l'image.

Les radiologues spécialisés qui lisent les clichés acquis lors d'une campagne de dépistage doivent lire un grand nombre de clichés à la suite. Les films sont chargés dans des négatoscopes à défilement pour être présentés examen par examen au radiologue. La durée de lecture par examen est de l'ordre d'une minute, ce qui montre la grande rapidité avec laquelle les radiologues analysent les images.

Après lecture, un rapport est rédigé dans un format standardisé, comme celui proposé par le *American College of Radiology* et connu sous la dénomination "BI-RAD"[BIRAD95]. La densité radiologique et les signes radiologiques y sont décrits et une proposition pour des analyses supplémentaires est éventuellement donnée.

Dans le contexte d'une campagne de dépistage, la classification complète par type de lésion n'est pas réalisable dans tous les cas, puisque les images seules ne fournissent pas suffisamment d'information sur les lésions. Souvent, seuls les résultats de l'examen clinique et la fusion de ces résultats avec ceux de la lecture des images permettent d'énoncer un diagnostic fiable. Cela ne pose pas de problèmes parce que l'identification d'une lésion peut être une étape supplémentaire.

Par contre, la lecture pour le dépistage doit trouver les signes avec la plus grande sensibilité possible, puisqu'une lésion ratée lors du dépistage ne sera pas retrouvée plus tard si aucune suspicion n'a été exprimée.

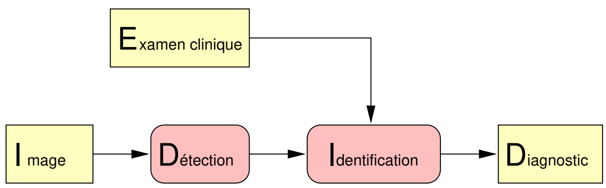


FIG. 8.1 – Chaîne de traitement pour la dépistage

8.1 Lecture des images numériques

Avec l'arrivée des détecteurs numériques sur le marché, le déroulement d'une campagne de dépistage va être significativement changé. En effet, les images ne seront plus lues sur un film accroché au négatoscope mais sur un écran. Cela est à première vue un handicap pour le radiologue, car son expérience était acquise à partir des images sur film. D'un autre côté, la possibilité de pouvoir changer les réglages de l'affichage va lui permettre d'augmenter sa capacité d'analyse sur les images. Par exemple, les filtres adaptés au rehaussement des microcalcifications peuvent lui permettre de détecter plus de calcifications. Ces nouvelles possibilités ne doivent cependant pas prolonger le temps de lecture, puisque le débit de lecture est critique pour le dépistage. Cette condition pose des contraintes sévères sur l'implémentation des fonctionnalités de la revue d'image.

C'est dans ce contexte que l'aide au diagnostic par ordinateur devient intéressante. La première étape est un système d'aide à la détection

8.2 Aide à la détection

Plutôt que de faire lire tous les clichés par deux médecins, il est envisageable qu'un même radiologue fasse une première lecture sans assistance et qu'ensuite le système de l'aide à la détection lui montre les zones classées suspectes pour que le médecin vérifie qu'il n'a rien raté dans ces zones.

L'avantage de cette approche est la combinaison des méthodes automatiques et humaines. Le système automatique peut faire une recherche plus exhaustive sur l'ensemble du cliché pendant que le radiologue dirige son attention sur les zones lui semblant les plus remarquables.

Krupinski et Nishikawa [Kru97] ont comparé les performances d'un système de détection automatique avec celles de six radiologues pour la détection des microcalcifications. Ils ont enregistré les positions des yeux des radiologues pendant la lecture des films et ont analysé le temps pendant lequel les radiologues se focalisent sur des lésions détectées ou ratées. Ils constatent qu'un radiologue expérimenté ne balaye pas l'image entière mais se concentre sur les zones identifiées lors de l'évaluation initiale globale comme contenant probablement des lésions. Leur évaluation sur 80 films a montré que 10% des amas de microcalcification ont été détectés par le système automatique mais par aucun des six radiologues et que 11% des amas ont été ratés par le système automatique mais détectés par au moins un des radiologues. Seulement 5% des amas ont été ratés par les radiologues et le système automatique.

8.3 Aide à l'identification

Une fois qu'une lésion a été détectée, il est nécessaire d'énoncer un diagnostic et si besoin est, de prévoir un traitement. La difficulté de l'identification du type de lésion et de sa bénignité ou malignité résulte du fait que des pathologies très différentes peuvent se traduire sur l'image radiologique de la même manière. Par exemple, les kystes et les fibro-adénomes ne sont pas toujours différenciables sur l'image. Ces deux lésions sont toutes les deux bénignes, mais, comme le tableau des lésions arrondies dans la section 7.4.1 le montre, il y a aussi le risque de confondre des lésions bénignes et malignes.

Une partie importante de la différenciation est liée à l'examen clinique. Par exemple, une différence entre la taille d'une lésion sur l'image radiologique et la taille déterminée par palpation est un signe de malignité.

Dans ce contexte, un système automatique d'aide à la classification peut être utile. Nous retrouvons là une complémentarité entre les points forts du médecin et les points forts des systèmes automatiques.

Le médecin peut plus facilement comparer les informations obtenues par examen clinique et celles par la mammographie. Par contre le médecin est subjectif quant au classement des lésions. Un système d'aide à la classification basé sur les attributs de la lésion mesurés sur l'image donnera toujours les mêmes résultats pour le même cliché et peut donc servir comme référence.

Une autre avantage potentiel est la taille de la base d'apprentissage. Avec l'introduction de la mammographie numérique, il y aura bientôt d'énormes bases de données disponibles. Un radiologue ne pourra jamais les passer toutes en revue pour élargir encore son expérience, ce qui sera par contre possible pour un système automatique.

8.4 Méthodes existantes pour l'aide au diagnostic

En raison de l'importance du cancer du sein, un grand nombre de groupes de recherche travaille sur le problème de l'aide au diagnostic. Les premiers travaux datent de 1967 [Win67] mais la plupart des publications date de la dernière décennie.

Les travaux se distinguent par les types des signes qui sont détectés (opacités, calcifications), le but recherché (détection, classification, présentation d'image) et les méthodes appliquées.

Dans la suite, nous allons regarder uniquement le problème de la détection et décrire à titre illustratif trois méthodes d'extractions d'attributs, une pour chaque type de lésion (opacité ronde, opacité stellaire, microcalcification).

8.4.1 Détection

D'une manière générale, la détection se décompose en une phase de calcul d'attributs et en une phase de décision ou de sélection (figure 8.2).

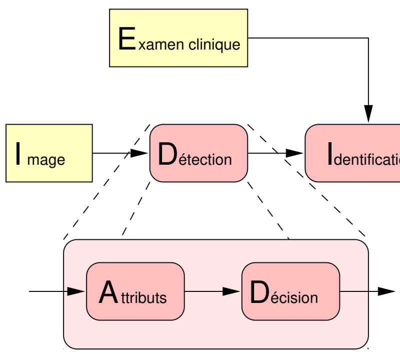


FIG. 8.2 – La détection

Comme nous l'avons indiqué dans chapitre 7, l'ensemble de la chaîne de traitement est confronté à des données qui contiennent beaucoup de variabilité. Il est donc très important que chaque étape soit construite de manière à prendre en compte cette variabilité.

Pour l'étape de calcul des attributs, cela signifie que le choix des attributs doit favoriser les attributs avec une sensibilité minimale aux variations non importantes et une sensibilité maximale pour la présence ou absence de signes radiologiques du cancer.

Pour l'étape de décision, la variabilité peut être prise en compte avec un raisonnement qui s'adapte aux conditions présentes dans l'image actuelle.

8.4.2 Les attributs pour la détection

Les attributs pour la détection proposés dans la littérature sont très variés et on retrouve des adaptations de quasiment toutes les méthodes du traitement d'images :

- les ondelettes [Dha96][Laine94][Qia95][Yu99][Yu98]
- la croissance de région [Ban97][Tao98]
- les moments [Dha96]
- les fractales [Lev95]
- les gradients [Ema95]
- les statistiques de voisinage[Pet94][Kim99]
- les filtres [Kar96][Gup95]
- la morphologie mathématique [Bet97][Bot96-4] [Vac95][Gri91][Mil92]
- les contours actifs [Ban97]
- le template matching [Lai89]
- les réseaux neuronaux [Bou89][Li97][Suc95][Cha95-1][Sah96-1][Bak95][Lo98]

Les attributs sont généralement sélectionnés à la main, mais il existe des travaux qui essaient d'optimiser le choix des attributs par des algorithmes génétiques [Sah96-2].

Pour montrer une application de la méthode des fonctions d'appartenance adaptatives, nous avons choisi le problème de détection des opacités en mammographie. Dans les paragraphes suivants, nous allons décrire trois types d'attributs : un spécialisé dans la détection des opacités denses, un spécialisé dans la détection des lésions stellaires et un spécialisé dans les microcalcifications.

8.4.3 Détection des opacités denses

Les opacités qui sont plus radio-opaques que les tissus environnants, peuvent être détectées avec des filtres adaptés. Dans [Lai89], des filtres circulaires qui ont une zone d'indifférence entre le centre de la masse et l'extérieur sont utilisés. Ils utilisent plusieurs filtres de tailles différentes pour les opacités de 3 à 15 pixels de diamètre. Les coefficients du filtre pour des opacités de 5 pixels de diamètre sont :

$$\begin{array}{ccccccccc}
 & & & -1 & -1 & -1 & & & \\
 & & -1 & -1 & 0 & -1 & -1 & & \\
 & -1 & -1 & 0 & 1 & 0 & -1 & -1 & \\
-1 & -1 & 0 & 1 & 1 & 1 & 0 & -1 & -1 \\
-1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & -1 \\
-1 & -1 & 0 & 1 & 1 & 1 & 0 & -1 & -1 \\
 & -1 & -1 & 0 & 1 & 0 & -1 & -1 & \\
 & & -1 & -1 & 0 & -1 & -1 & & \\
 & & & -1 & -1 & -1 & & &
\end{array} \tag{8.1}$$

Ces filtres peuvent être décrits par trois rayons :

r_n le rayon du noyau de l'opacité. A l'intérieur d'une sphère de rayon r_n, les coefficients sont égaux à 1.

r_e le rayon de l'extension maximale de l'opacité. Entre r_n et r_e, les coefficients sont égaux à 0, c'est-à-dire que les valeurs d'image n'ont pas d'influence sur le résultat. Cela permet de détecter des opacités qui ne sont pas parfaitement circulaires.

r_f le rayon maximal du filtre. Entre r_e et r_f, les coefficients sont égaux à -1. Cette zone correspond au fond de l'image.

Nous avons implémenté une méthode très similaire, mais appliquée sur des images en multi-résolutions. L'avantage des images en multi-résolutions est la possibilité d'utiliser une gamme de tailles pour les opacités plus larges, tout en réduisant la complexité en termes de nombres d'opérations par pixel. Afin d'obtenir cela, une pyramide d'images est créée avec trois niveaux et un facteur de sous-échantillonnage de deux entre chaque niveau. Sur chaque image, nous appliquons un filtre circulaire tel qu'il est décrit ci-dessus. Les résultats appliqués à une image mammographique sont montrés sur la figure 8.3. On remarque la lésion en haut de l'image dont la taille

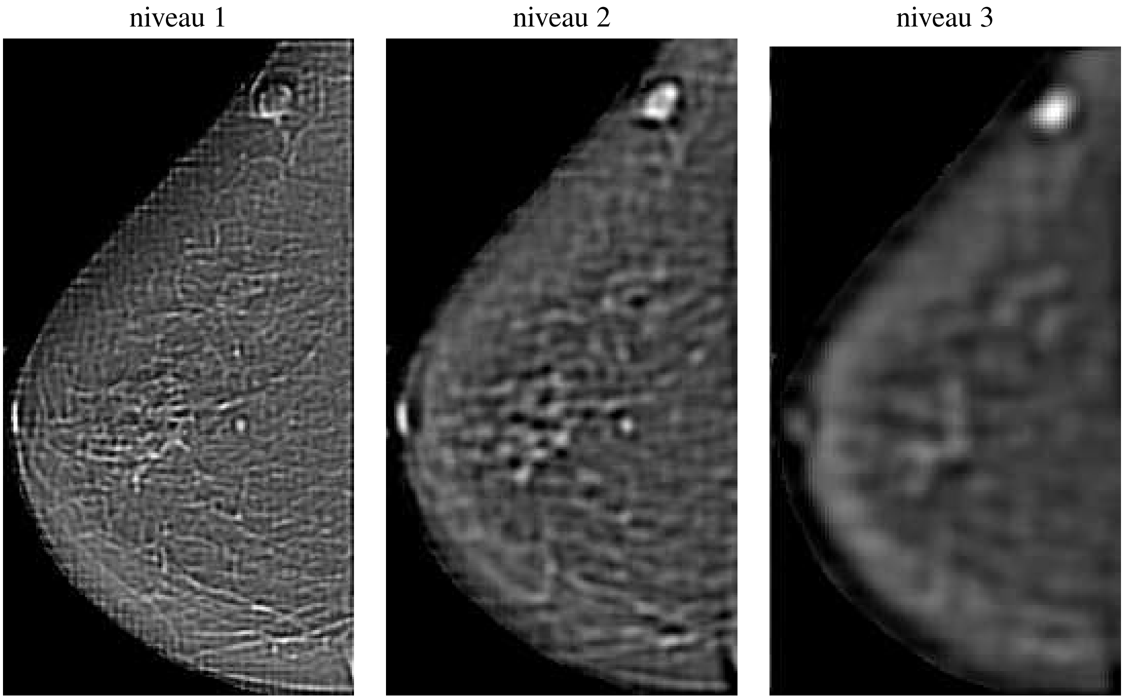


FIG. 8.3 – Détection des lésions denses niveau par niveau

correspond au filtre circulaire appliqué à l'image sous-échantillonée trois fois, et elle est donc très bien rehaussée par ce filtre. Ensuite, les résultats obtenus aux différentes échelles sont fusionnés avec l'opérateur *maximum*. Le résultat est montré sur la figure 8.4.

8.4.4 Détection des opacités stellaires

Dans [Kar96], Karssmeijer et Brake ont présenté une méthode de détection et de classification, qui repose sur la statistique des directions des fibres détectées dans l'image. Pour chaque point de l'image, la direction et le contraste d'une fibre sont estimés avec un ensemble de trois filtres. Les filtres utilisés sont la deuxième dérivée d'un filtre gaussien dont les réponses impulsionnelles sont illustrées sur la figure 8.5.

A partir des résultats de ces trois filtres, la direction θ_{min_max} du maximum de la dérivée est calculée :

$$\theta_{min\_max} = \frac{1}{2}\left[arctan\left(\sqrt{3}\frac{W_{-60} - W_{+60}}{W_{+60} + W_{-60} - 2W_0}\right) \pm k\pi\right] \tag{8.2}$$

Ensuite, la valeur de la dérivée est calculée dans cette direction :

$$\begin{aligned} W(\theta) &= \frac{1}{3}(1 + 2\cos(2\theta))W_0 \\ &+ \frac{1}{3}(1 - \cos(2\theta) + \sqrt{3}\sin(2\theta))W_{+60} \\ &+ \frac{1}{3}(1 - \cos(2\theta) - \sqrt{3}\sin(2\theta))W_{-60} \end{aligned}$$

image originale 1 lésions denses

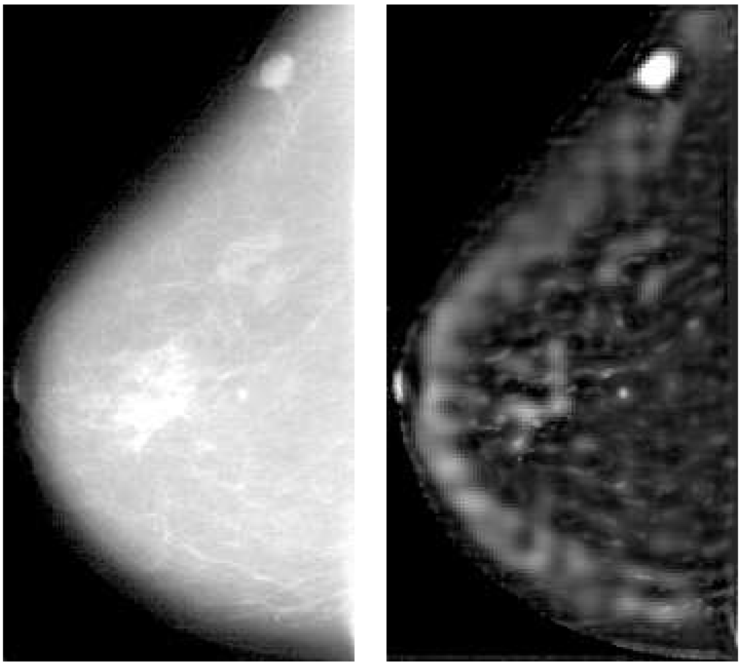


FIG. 8.4 – Détection des lésions denses

W_{-60} W_0 W_{+60}

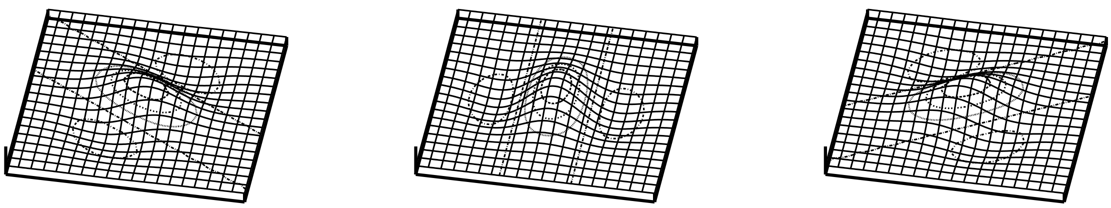


FIG. 8.5 – Filtres directionnels pour la détection des fibres

Puisque ces filtres font intervenir la largeur des fibres détectées, ils sont appliqués sur des images en plusieurs résolutions. Ensuite, les images sont fusionnées avec l'opérateur maximum. Ce procédé est schématisé sur la figure 8.6.

Après l'estimation des directions de fibres pour chaque pixel, une analyse statistique de ces directions est effectuée pour estimer le pourcentage de pixels qui, dans une zone circulaire entre le rayon R_{min} et R_{max}, pointent sur le centre de la zone. Pour prendre en compte l'imprécision dans la mesure de la direction et la taille du noyau des lésions stellaires, le centre est lui-même approché par une zone circulaire plus petite (de rayon R_{centre}). La figure 8.7 illustre le fonctionnement de cet opérateur : à gauche, une lésion stellaire est schématisée, sur laquelle se trouvent superposés les trois cercles de rayons R_{centre}, R_{min} et R_{max}. Le plus petit est inclus dans le noyau tandis que les fibres qui rayonnent du centre se trouvent principalement entre R_{min} et R_{max}. A droite de la figure 8.7, l'analyse statistique est illustrée : pour chaque pixel (ici seulement quelques pixels sont indiqués par des points noirs) la direction des fibres est calculée et pour toute la zone entre R_{min} et R_{max}, on compte le nombre de directions qui coupent le petit cercle de rayon R_{centre}.

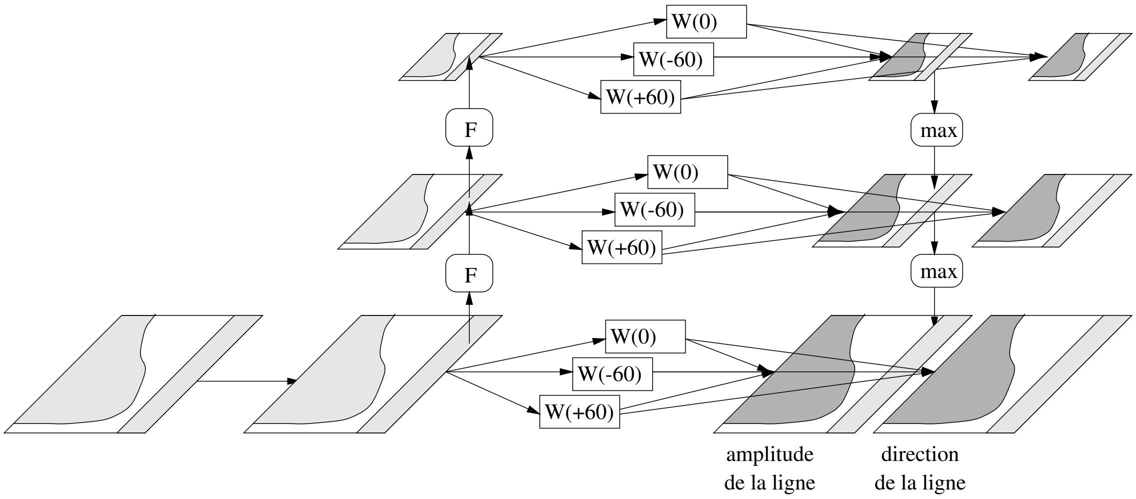


FIG. 8.6 – Calcul de la direction et du contraste des fibres en multi-résolution

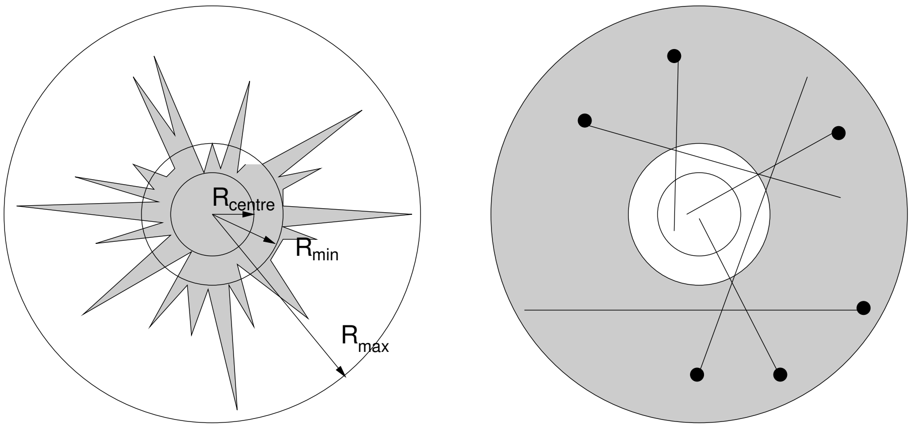


FIG. 8.7 – Statistique des directions des fibres

La figure 8.8 montre l'application de ce filtre à une image mammographique. Dans cet exemple, il y a deux lésions stellaires qui sont indiquées avec des flèches sur l'image originale. Les lésions sont bien visibles sur l'image originale (à gauche), car elles se trouvent dans une zone peu fibreuse. L'amplitude $W(\theta_{max})$ des fibres est donnée au centre de la figure 8.8 et finalement le résultat des statistiques des directions est montré à droite. Les deux lésions sont visibles dans l'image de cet attribut mais il y a d'autres zones qui sont rehaussées également, par exemple à des endroits correspondant à des croisements de fibres. Visuellement, il n'est pas beaucoup plus facile de détecter des lésions sur l'image de l'attribut que sur l'image originale, puisque les lésions dans cet exemple possèdent un noyau dense relativement facilement repérable. L'intérêt de cet attribut est fondé dans sa complémentarité par rapport à l'attribut précédent. La combinaison des deux attributs, c'est-à-dire les zones stellaires avec un noyau dense, correspond à une description typique donnée par des médecins pour des zones très suspectes.

8.5 Détection des microcalcifications

Les deux principales caractéristiques des microcalcifications sont leur petite taille et leur atténuation élevée. Dans [Gri91] et [Bot96-4], une méthode de détection des microcalcifications

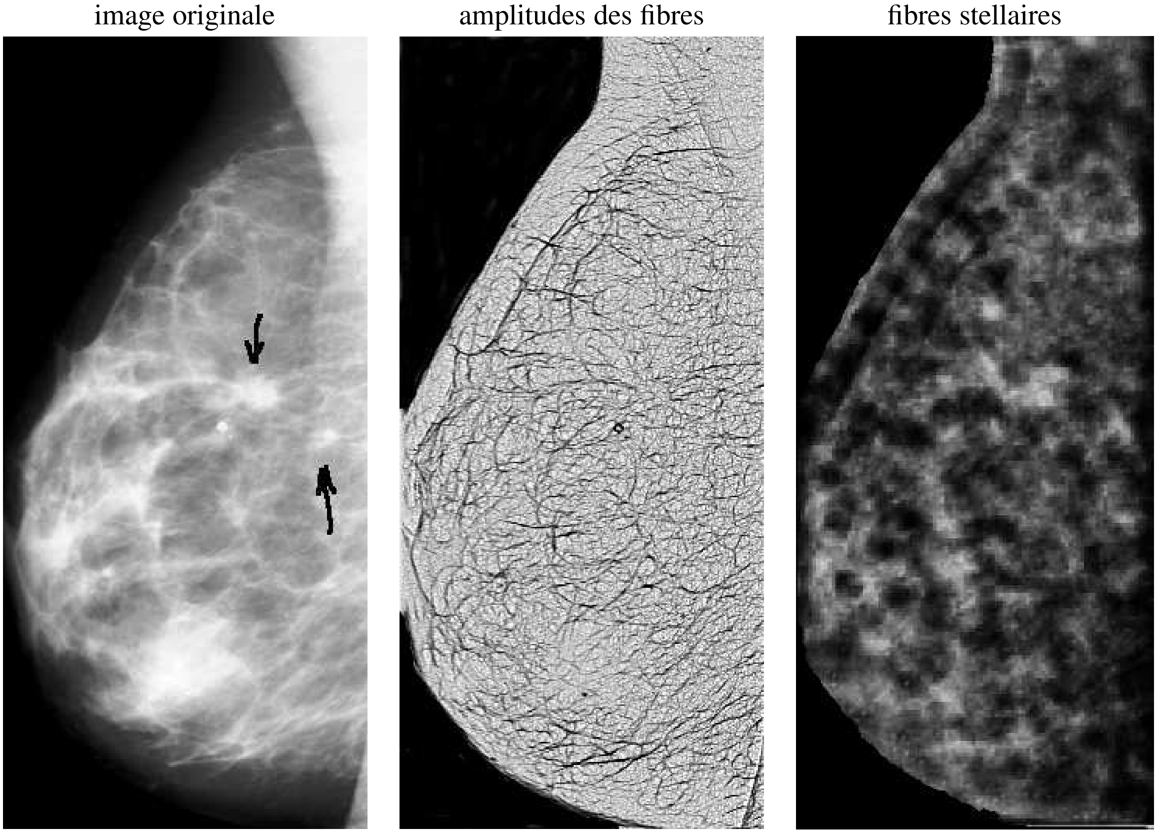


FIG. 8.8 – Statistique des directions des fibres

basée sur l'opérateur *chapeau haut-de-forme* est présentée, qui combine ces deux caractéristiques. L'opérateur *chapeau haut-de-forme seuillé* est composé de trois étapes :

1 - l'ouverture de l'image avec un élément structurant
2 - la soustraction de l'image resultant de l'ouverture à l'image originale
3 - l'application d'un seuil sur le résultat.

L'ouverture élimine toutes les structures plus petites que l'élément structurant et ne garde que les structures plus grandes. La soustraction de cette image à l'image originale contient donc uniquement les petites structures qui ont été effacées par l'ouverture. Finalement, le seuillage ne prend en compte que les structures présentant un certain contraste. Puisque la taille des microcalcifications n'est pas connue d'avance, plusieurs opérateurs *chapeau haut-de-forme* avec des éléments structurants de tailles différentes sont utilisés en parallèle et les résultats sont combinés (vois figure 8.9).

Le niveau de seuillage pour le *chapeau haut-de-forme seuillé* est très sensible au niveau de bruit dans l'image. Dans [Bot96-4], le seuillage se fait sur la somme des images du *chapeau haut-de-forme*, ce qui revient à donner plus d'importance aux petites structures. Puisque ces travaux se basent sur des images de mammographie obtenues à partir de films numérisés, les conditions d'acquisition ne sont pas connues et le seuil de détection est calculé à partir des constantes déduites de chaque image.

Dans le chapitre 10, cette méthode de détection est améliorée pour prendre en compte les paramètres d'acquisition qui sont disponibles sur des systèmes de mammographie numériques.

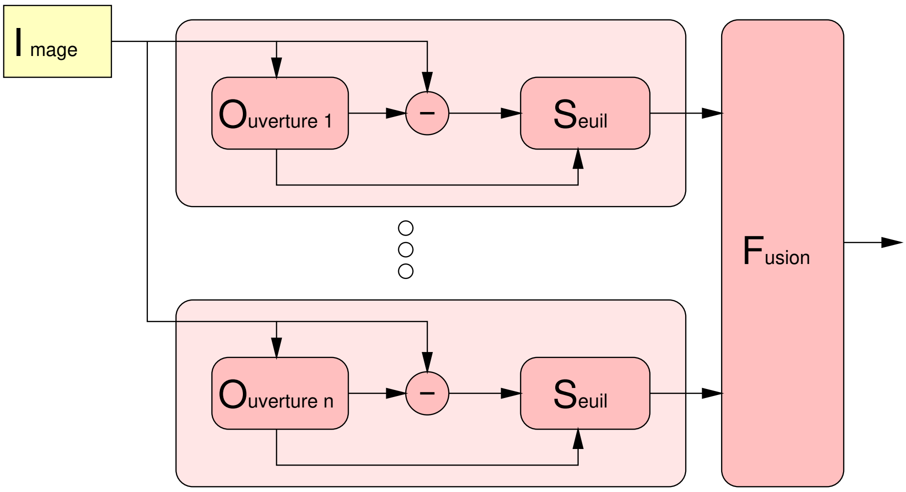


FIG. 8.9 – Détection de microcalcifications avec des opérateurs *chapeau haut-de-forme*

8.6 Conclusion

Les trois attributs détaillés dans ce chapitre ne représentent que des exemples pour des filtres spécialisés à la détection des trois grandes classes de signes que l'on trouve en mammographie. En pratique, on utilise une combinaison des plusieurs attributs puisqu'on ne connaît pas par avance quel type de signe se trouve sur un cliché de mammographie.

Chapitre 9

Utilisation des fonctions d'appartenance adaptatives pour la détection d'opacités

Dans ce chapitre, nous allons appliquer la méthode des sous-ensembles flous adaptatifs au problème de la détection des lésions denses en mammographie.

9.1 Base de données

Au moment de la réalisation des travaux présentés ici (en 1999), les premiers appareils de mammographie numériques offrant un champ de vue comparable au film commencent à être commercialisés. Par conséquent, les bases de données des images de mammographie numériques sont encore rares. Nous avons donc préféré utiliser une base de données construite à partir de films numérisés pour valider notre approche. Pour cela, nous avons travaillé avec une partie de la base *Digital Database for Screening Mammograph (DDSM)* présentée dans [Bow98].

Nous avons utilisé le volume *cancer10* qui contient 236 images provenant de 59 patientes. Les examens ont été réalisés avec un protocole de dépistage, avec 4 clichés par patiente, c'est-à-dire une vue de face et une vue oblique externe par sein. (Voir chapitre 7 pour une explication des vues). Ensuite les films ont été numérisés avec 12 bit par pixel (4096 niveaux de gris) et une résolution spatiale de 43.3 microns. Pour les mesures effectuées dans la suite, nous avons sous-échantillonné l'image par un facteur de deux dans les deux directions en moyennant sur 2 fois 2 pixels. Nous considérons que la résolution obtenue de 86,6 microns par pixel est suffisante pour la tâche de détection des opacités.

La base d'images DDSM est également constituée des annotations données par un expert radiologiste. La description contient une certain nombre d'attributs pour chaque lésion et une annotation par contour pour chaque lésion sur les deux vues. Les attributs fournis pour chaque image sont :

- le nombre des anomalies
- le type de chaque anomalie (calcification / lésion dense)
- évaluation selon le code BI-RADS
- subtilité
- résultat de la biopsie (bénin/ malin)
- les contours des lésions

La figure 9.1 montre le cas 1626 de cette base, avec les annotations du médecin sur les deux clichés du sein gauche.

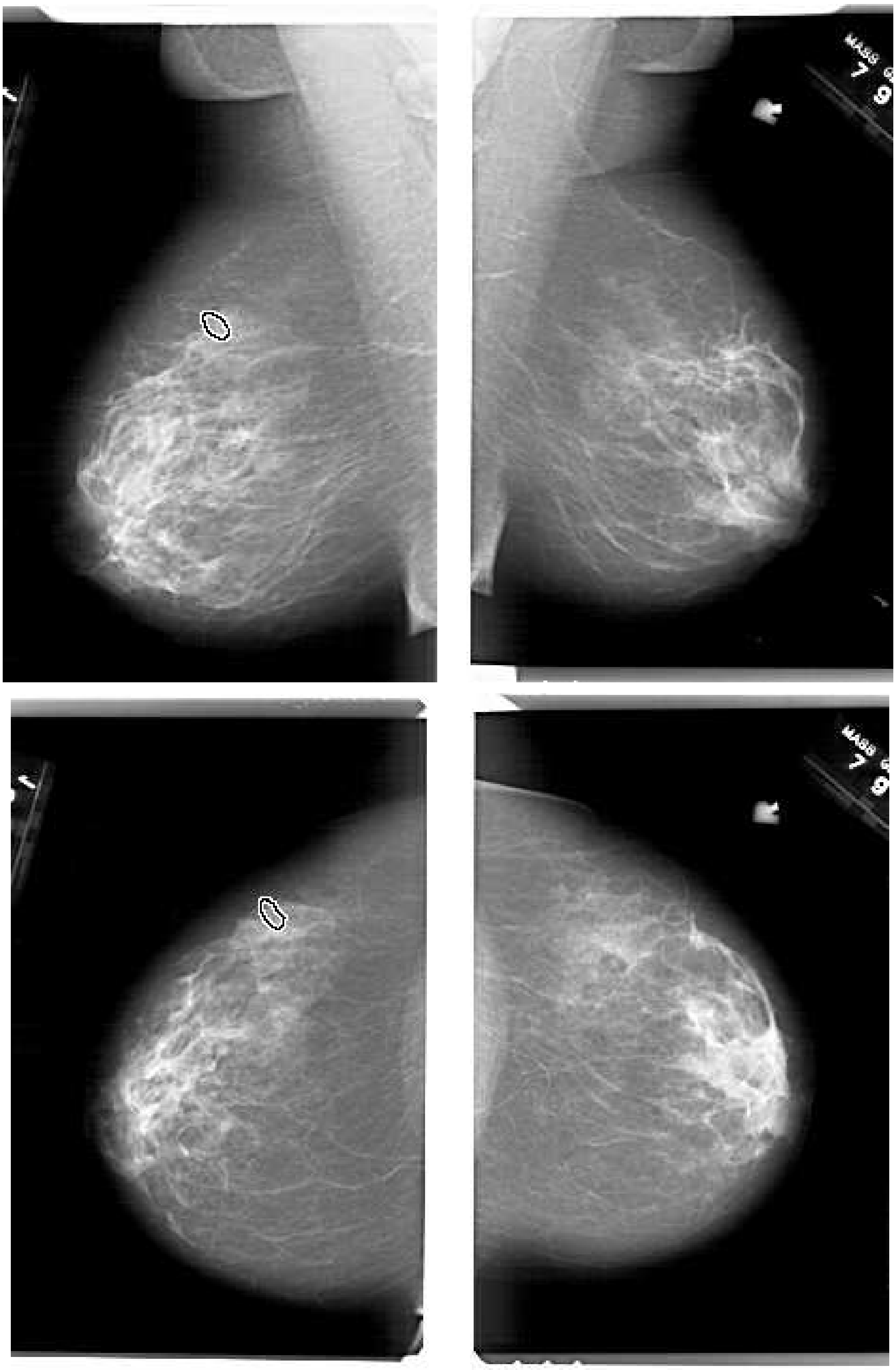


FIG. 9.1 – Cas 1626 de la base DDSM avec annotations

9.2 Apprentissage du modèle de variabilité

La première phase pour l'utilisation des fonctions d'appartenance adaptatives est l'apprentissage du modèle de la variabilité à partir de la base d'apprentissage. Les étapes pour cet apprentissage sont :
- calcul des attributs
- calcul des histogrammes de chaque attribut
- estimation des densités dans l'histogramme, par un mélange de composantes
- regroupement des composantes
- régression pour trouver des interactions entre les composantes

9.2.1 Calcul des attributs

Pour illustrer la méthode des fonctions d'appartenance adaptatives, nous avons choisi l'attribut décrit dans la section 8.4.3 qui rehausse les opacités denses. Cet attribut est calculé sur chacune des 116 images avec au moins une lésion. La figure 9.2 montre les deux images de cet attribut calculé sur les deux images du sein gauche. Les positions de la lésion sont marquées avec des flèches. Les zones qui contiennent la lésion ne se distinguent pas beaucoup du reste de l'image

projection oblique externe — projection de face

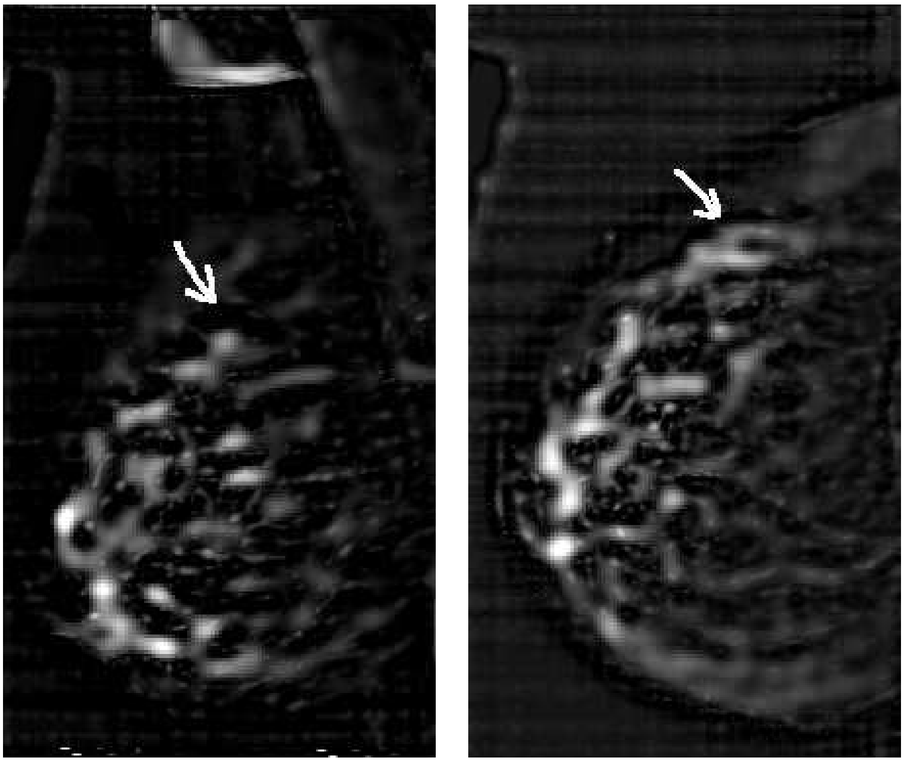


FIG. 9.2 – Image de l'attribut pour opacités denses

pour ce cas plutôt difficile. On ne distingue pas clairement un noyau dense sur l'image originale, et même sur l'image de l'attribut pour les opacités denses les zones sont difficiles à détecter. Il est clair que le résultat basé sur ce seul attribut n'est pas suffisant pour une détection avec une grande précision, mais nous utilisons cet exemple difficile pour illustrer le fonctionnement de la méthode d'adaptation proposée. Dans un contexte d'un système de détection complet avec plusieurs attributs et une bonne gestion des imprécisions, même un attribut peu discriminant peut être utile, si l'information qu'il apporte est complémentaire à celle des autres attributs.

9.2.2 Calcul des histogrammes

En utilisant le contour fourni par l'expert médical, les histogrammes à l'intérieur et à l'extérieur du contour tracé sont calculés. Les histogrammes pour une image de la base sont donnés en figure 9.3.

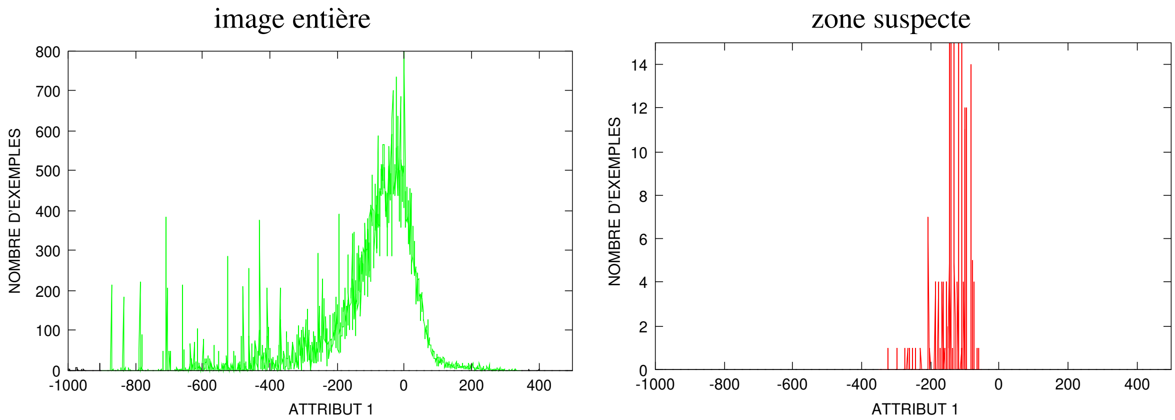


FIG. 9.3 – Histogramme pour une image de la base

Pour la suite il est important de noter la différence d'échelle entre les deux histogrammes, mise en évidence sur la figure 9.4, qui montre les deux histogrammes ensemble. On remarque également que les valeurs de l'attribut pour les pixels de la zone suspecte sont comprises dans les valeurs de la zone normale. Le but du système de classification est alors de trouver la zone suspecte à l'intérieur de l'histogramme des pixels normaux.

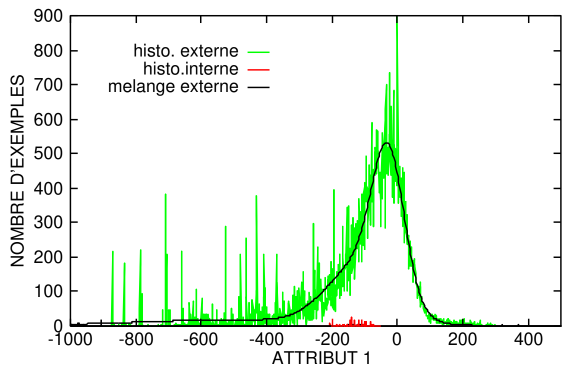


FIG. 9.4 – Comparaison d'échelle

9.2.3 Estimation d'un modèle paramétrique de l'histogramme

Nous utilisons l'algorithme H-EM pour trouver une description paramétrique de tous les histogrammes par un ensemble de composantes gaussiennes. Dans notre cas, nous avons choisi 4 composantes pour la description des histogrammes de l'image entière et 3 composantes pour la description des histogrammes des zones suspectes indiquées par les médecins. Le résultat trouvé pour la même image que précédemment, est donné sur la figure 9.5.

9.2.4 Regroupement des composantes

Pour assurer que l'ordre des composantes est le même pour toutes les images, l'algorithme FCM est appliqué sur les paramètres des composantes. Nous avons choisi de réduire le nombre de

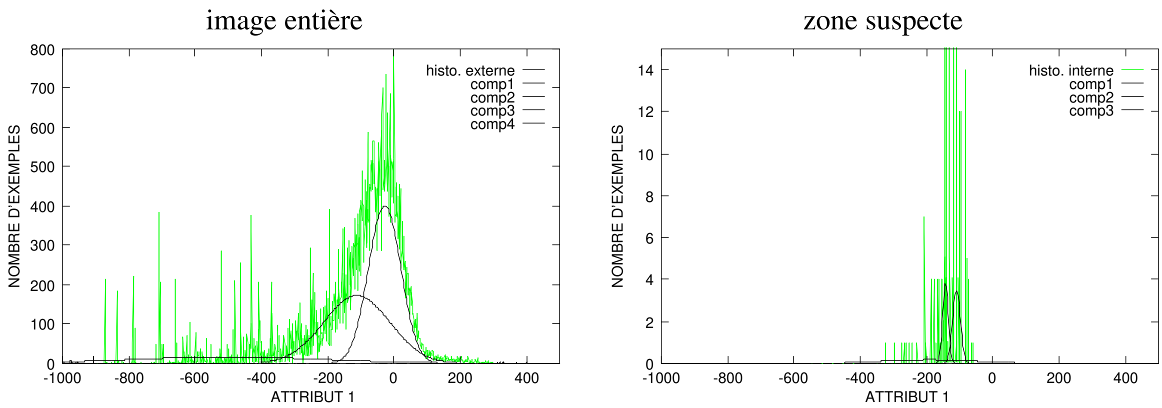


FIG. 9.5 – Approximation initiale avec l'algorithme H-EM

composantes retenues dans ce prototype à trois composantes pour l'image entière et deux composantes pour les zones suspectes. Le résultat de cette étape est un prototype (figure 9.6) qui donne les valeurs typiques pour les composantes et une matrice des valeurs d'appartenance de chaque composante trouvée initialement, associées à chaque composante du prototype.

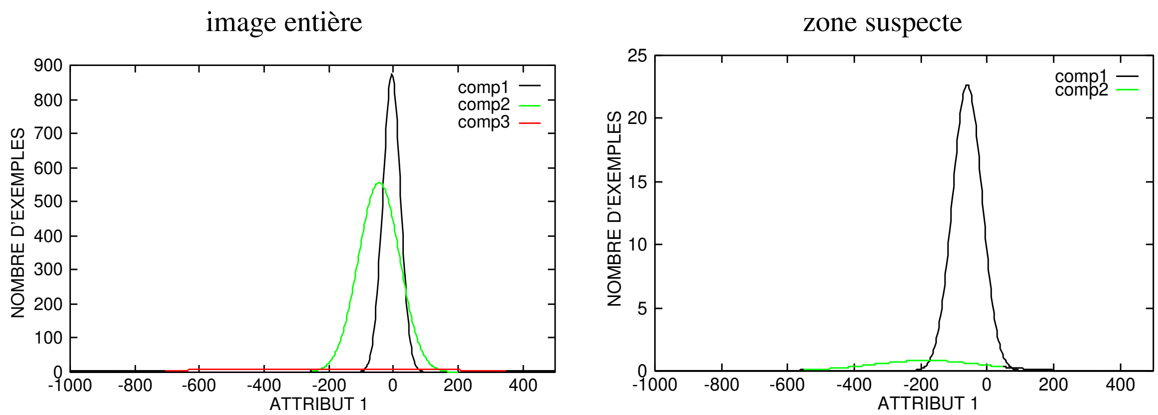


FIG. 9.6 – Prototype issu du regroupement

9.2.5 Estimation des interactions avec une régression

En regardant le rapport entre les histogrammes mis en évidence sur la figure 9.4, il est évident que nous ne pouvons pas espérer trouver la position et la forme de la fonction d'appartenance de la classe de pixels *suspect* indépendamment du modèle pour l'image en entier, car ce signal est complètement caché par la majorité des pixels normaux. Pour cette raison, il est impératif de modéliser les liens entre les paramètres des composantes. Nous pouvons retrouver ceux-ci assez facilement, à partir de l'histogramme de l'image entière et des paramètres des composantes cachées de la classe *suspect*.

Pour modéliser ces interactions, nous avons choisi le modèle linéaire décomposé décrit dans la section 4.4.4.

Les matrices **B** (voir équations 4.36 à 4.38) que nous avons obtenues sont :

$$\mathbf{B}_p = \begin{bmatrix} 0.701 & -0.461 & -0.017 & 0.058 & 0.012 \\ -0.461 & 0.322 & -0.025 & 0.077 & 0.016 \\ -0.017 & -0.025 & 1.018 & 0.004 & 0.002 \\ 0.058 & 0.077 & 0.004 & 0.997 & -0.007 \\ 0.012 & 0.016 & 0.002 & -0.007 & 1.015 \end{bmatrix} \tag{9.1}$$

$$\mathbf{B}_\mu = \begin{bmatrix} 0.973 & 0.007 & -0.177 & 0.034 & 0.006 \\ 0.007 & 1.017 & 0.036 & -0.006 & -0.019 \\ -0.177 & 0.036 & 0.300 & 0.138 & 0.002 \\ 0.034 & -0.006 & 0.138 & 0.992 & 0.028 \\ 0.006 & -0.019 & 0.002 & 0.028 & 0.003 \end{bmatrix} \tag{9.2}$$

$$\mathbf{B}_\sigma = \begin{bmatrix} 1.002 & -0.007 & 0.016 & 0.007 & -0.132 \\ -0.007 & 1.017 & 0.006 & 0.003 & -0.054 \\ 0.016 & 0.006 & 1.005 & -0.006 & 0.120 \\ 0.007 & 0.003 & -0.006 & 1.017 & 0.054 \\ -0.132 & -0.054 & 0.120 & 0.054 & 0.048 \end{bmatrix} \tag{9.3}$$

Les deux premières lignes et colonnes correspondent aux paramètres des composantes qui décrivent la zone suspecte et les trois autres aux paramètres de l'image entière. Si la matrice **B** était égale à la matrice identité, cela signifierait qu'il n'y a pas de liens entre les composantes. Pour la matrice $\mathbf{B}_p$ qui décrit les poids des composantes, nous observons une forte interaction entre les deux composantes de la zone suspecte (avec un signe négatif). Cela traduit le fait que le poids des deux composantes peut varier, mais la somme des deux est relativement stable. Dans la matrice $\mathbf{B}_\mu$, on trouve par exemple une légère interaction entre la première composante de la zone normale (colonne 1) et la première composante de la zone suspecte (colonne 4) avec un signe positif, c'est à dire que les positions centrales de ces composantes ont tendance à bouger dans la même direction. Les interactions dans la matrices $\mathbf{B}_\sigma$ ne sont pas aussi évidentes.

9.3 Test du modèle de variabilité

Pour évaluer la performance de classification, nous avons coupé la base de données DDSM en deux, à savoir la base d'apprentissage et la base de test. L'apprentissage du modèle de variabilité se fait sur une partie, et le test sur l'autre. Pour augmenter la valeur statistique du test, les deux bases sont échangées et les résultats sont calculés comme la moyenne des deux essais.

Le modèle estimé sur la base d'apprentissage est ensuite appliqué sur les images de la base de test.

Pour estimer les fonctions d'appartenance pour un nouveau cas il faut :
- calculer des attributs
- calculer les histogrammes
- estimer les densités dans l'histogramme avec l'algorithme P-EM, en utilisant le modèle qui a été trouvé pendant la phase d'apprentissage
- normaliser pour obtenir des fonctions d'appartenance.

Les deux première étapes sont strictement équivalentes à celles de la phase d'apprentissage.

9.3.1 Adaptation du prototype à l'image actuelle

Pendant l'adaptation du prototype à l'histogramme actuel avec l'algorithme P-EM, les composantes importantes sont réajustées grâce à l'énergie externe qui vient de l'histogramme. Les

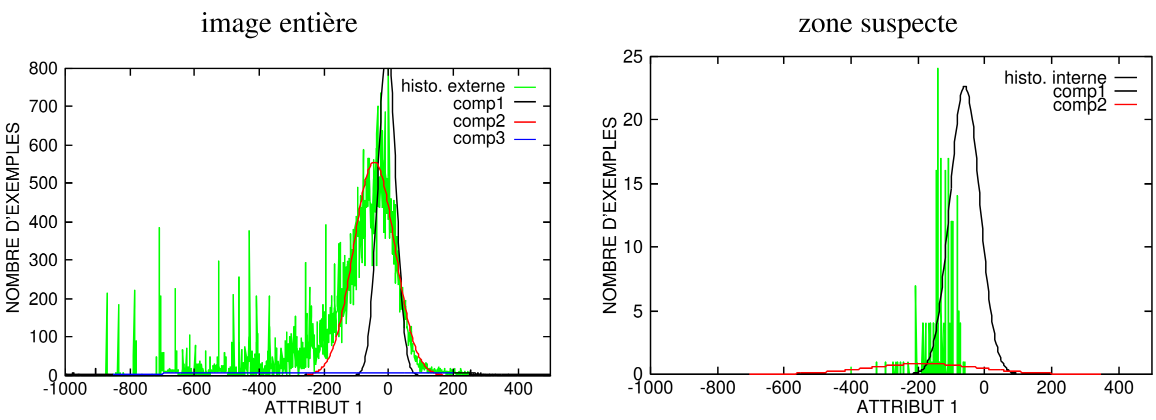


FIG. 9.7 – Histogrammes et prototype avant adaptation

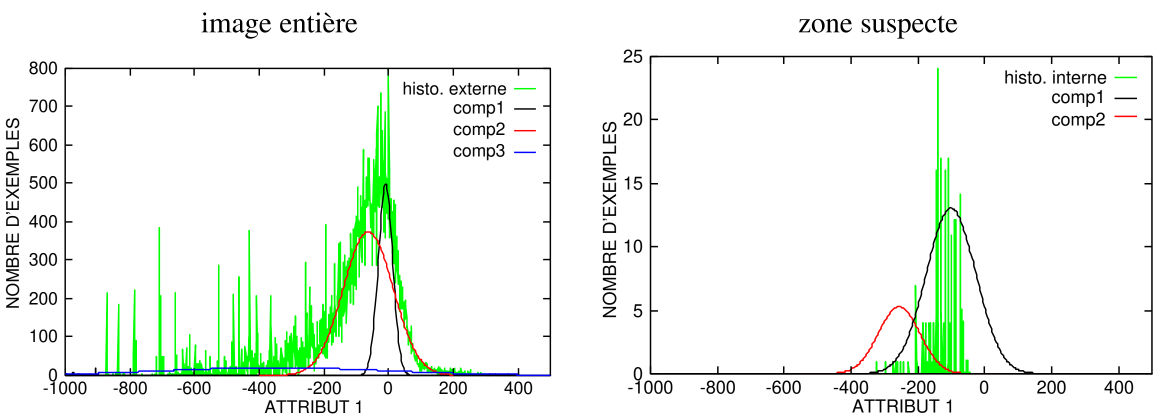


FIG. 9.8 – Histogrammes et prototype après adaptation

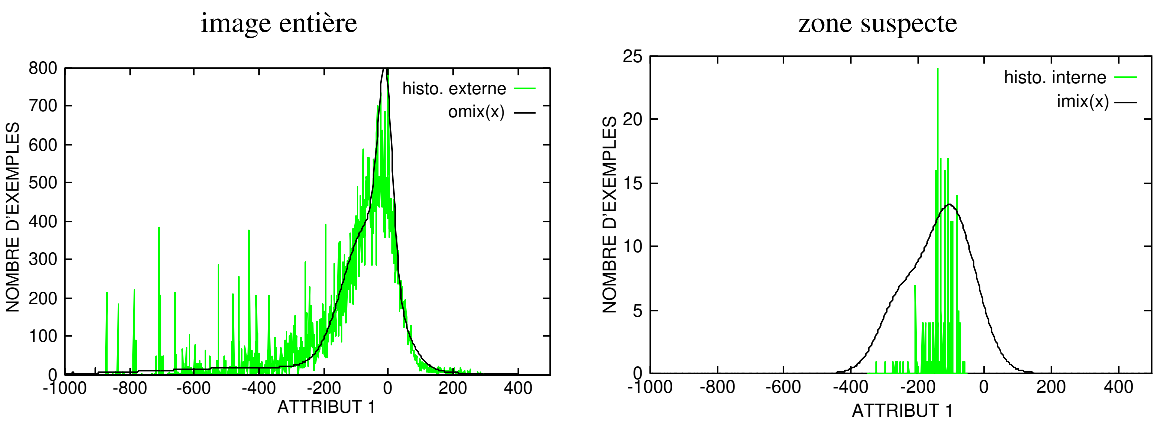


FIG. 9.9 – Mélange après adaptation

positions des composantes qui ont un faible poids, sont réajustées par rapport à ces composantes par le biais de l'énergie interne qui a été établie lors de l'apprentissage.

Pour illustrer cet ajustement, la figure 9.7 montre l'histogramme avec le prototype avant adaptation, et la figure 9.8 celui après adaptation pour le même exemple utilisé dans ce chapitre. La figure 9.9 montre les mélanges (somme des composantes) du prototype adapté.

On peut remarquer que, à cause de l'énergie interne, l'adaptation du prototype avec l'algorithme P-EM n'est pas aussi précise que celle faite pendant l'apprentissage (figure 9.5) - même pour la zone normale. Par contre, cette énergie interne permet d'adapter les composantes pour la zone suspecte, malgré le fait que son importance en nombre de pixels soit de plusieurs ordres de grandeurs inférieure à celle de la zone normale.

La précision des composantes trouvées pour la partie cachée (la zone suspecte) est meilleure que celle du prototype non adapté. Elle est aussi meilleure que celle d'une description par un histogramme établi sur la somme de tous les cas (figure 9.10) cependant elle reste largement inférieure à celle estimée pendant l'apprentissage, puisque sa position est presque entièrement définie par le modèle de variabilité.

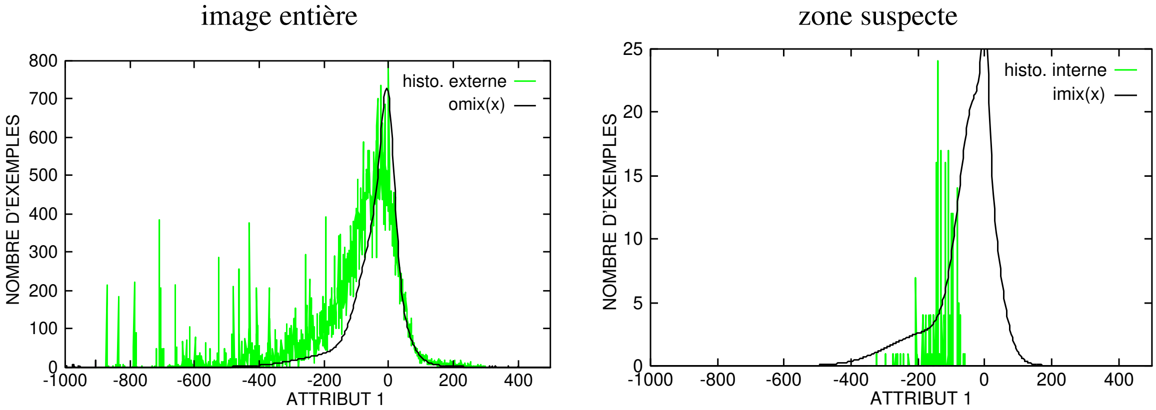


FIG. 9.10 – Histogrammes issus de la somme de tous les cas

9.3.2 Performance de classification

Pour donner une idée de la performance de la méthode des fonctions d'appartenance adaptatives, nous avons utilisé le critère de la surface sous la courbe COR (annexe A). Le tableau 9.1 montre les surfaces sous la courbe COR pour les deux moitiés de la base (moitié d'apprentissage, moitié test), la moyenne des deux et le résultat si la base est utilisée en entier pour l'apprentissage et le test.

	idéale	statique	mélange idéal	mélange statique	mélange adaptatif
1ère moitié	0.783	0.681	0.752	0.642	0.682
2ème moitié	0.809	0.681	0.761	0.642	0.685
moyenne	0.802	0.681	0.757	0.642	0.684
1 seule base	0.802	0.691	0.757	0.643	0.695

TAB. 9.1 – Surface sous la courbe COR

La signification des différentes colonnes est la même que celle décrite dans la section 6.2.3. Les valeurs *idéale* et *statique* sont des estimations optimistes de la performance maximale pour des images séparées (*idéale*) et pour toutes les images ensembles (*statique*). Ces résultats sont fortement surévalués pour cette base, puisque la statistique dans l'histogramme est très faible pour la classe *suspect*. Les valeurs de performance réalistes sont données dans la colonne *mélange idéal* qui considère que la statistique de l'image actuelle est connue (ce qui n'est pas le cas en réalité) et la colonne *mélange statique* qui donne le résultat pour une classification statique. Finalement, la colonne *mélange adaptatif* donne les résultats pour la méthode adaptative proposée.

9.3.3 Discussion

Les résultats présentés dans la section précédente montrent que l'adaptation du prototype à l'image actuelle donne des résultats supérieurs à des méthodes statiques qui ne considèrent pas de variabilité entre les images. Le gain de performance qui peut être obtenu par une telle modélisation dépend d'un certain nombre de facteurs, comme le degrée de variabilité présent dans la base d'images et l'applicabilité d'un modèle simple (comme le modèle linéaire utilisé ici) aux liens entre les composantes.

L'intérêt des résultats présentés ci-dessus ne réside pas dans les valeurs absolues de la performance pour un système de détection de lésions denses. On trouve de nombreuses publications qui présentent des résultats meilleurs pour des systèmes complets de détection des lésions. Pour arriver à des performances comparables, le nombre d'attributs utilisé ici n'est pas suffisant. L'intérêt de l'application de notre méthode sur le problème de la détection des opacités est illustré par le gain de performance de classification pour un seul attribut. Pour généraliser ce gain, la méthode doit être appliquée à d'autres attributs.

La méthode présentée ici est intéressante pour des applications où les données ont une variabilité importante qui n'est pas connue au départ. Le système proposé est capable d'estimer à la fois les positions des classes, leur variabilité et les interactions entre les différentes classes.

Chapitre 10

Réduction de la variabilité pour la détection des microcalcifications

Dans le chapitre précédent, nous avons présenté l'application des fonctions d'appartenance adaptatives pour réduire la variabilité introduite par des paramètres non-mesurables dans le cadre de la détection d'opacités en mammographie. Dans la suite, nous proposons une méthode de réduction de la variabilité pour les paramètres mesurables du système d'acquisition, qui utilise une simulation afin de prédire l'effet de ces paramètres sur l'image. Nous appliquons cette méthode à l'autre groupe de signes importants pour la détection du cancer du sein, les microcalcifications.

10.1 Détection des microcalcifications

Une microcalcification est un dépôt de substances (composé de $Ca_3(PO_4)_2$, $CaCO_3$ et $Mg_3(PO_4)_2$) qui sont très radio-opaques et se traduisent, dans les clichés de mammographie, par de petits points clairs. Sur une mammographie, les caractéristiques qui distinguent les microcalcifications des autres éléments sont leur petite taille et leur fort contraste. Les microcalcifications n'ont pas de taille minimale, ce qui fait que les plus petites d'entre elles peuvent facilement être confondues avec le bruit présent dans les images de mammographie.

Le problème de la détection des microcalcifications peut être divisé en deux sous-problèmes séparés :

– trouver les signes radiologiques qui leur correspondent dans une image bruitée
– distinguer ces signes des autres signes normaux dans les images de mammographie, comme les croisements de fibres.

Dans chacun des deux problèmes, on est confronté à la variabilité d'une image à l'autre. Pour la détection des microcalcifications par rapport au bruit, la variabilité vient principalement du niveau de bruit présent dans l'image. Pour la distinction des microcalcifications des autres signes normaux, la variabilité vient des variations normales des différents tissus.

Avec l'introduction des systèmes numériques en mammographie, les paramètres physiques du système d'acquisition sont connus d'une manière beaucoup plus précise que pour les systèmes à base de film. Pendant l'acquisition, tous les paramètres de réglage de l'appareil de mammographie sont enregistrés avec l'image elle-même, et une calibration du système est faite d'une manière régulière. Cette connaissance détaillée du système d'acquisition peut être utilisée pour réduire la première source de variabilité mentionnée au début du paragraphe, celle liée au bruit.

Dans ce qui suit, nous décrivons un modèle de la chaîne d'image d'un système de mammographie numérique qui permet de prédire le niveau de bruit à un endroit donné sur l'image, en prenant en compte tous les paramètres d'acquisition disponibles. Nous montrons également comment il est possible d'utiliser ce modèle pour réduire l'influence de la variabilité du bruit sur le résultat de la détection. La figure 10.1 montre un tel modèle.

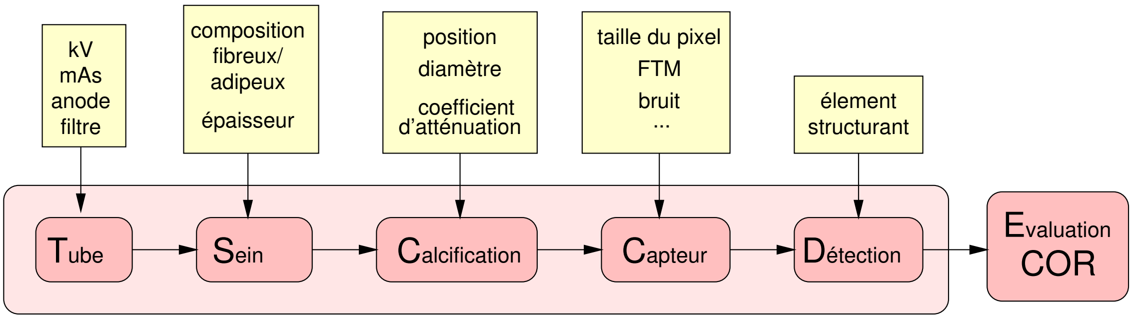


FIG. 10.1 – Modèle complet d'un système de mammographie numérique, du tube au système de détection des calcifications

10.2 Modèle de la chaîne d'acquisition

La chaîne d'acquisition est composée d'un tube, de l'objet à analyser et du capteur de rayons-X. L'image acquise est ensuite utilisée pour détecter des éventuelles zones suspectes. Dans la suite, nous allons décrire les différentes composantes du modèle.

10.2.1 Modèle du tube

Les tubes à rayons X génèrent les rayons X par projection d'électrons, accélérés dans un champ électrique, sur une anode (composée de Mo, Rh ou W). Le spectre d'émission ($n_0(u,\alpha,E)$), qui décrit le nombre de photons X émis en fonction de leur énergie (E en keV) par quantité de charges envoyées au tube, est fonction du matériau de l'anode, de la tension u d'accélération pour les électrons et l'angle de sortie du faisceau par rapport à la normale à la surface de l'anode (α). Pour notre application, ces spectres, qui ont été mesurés, sont stockés pour toutes les combinaisons des paramètres *tension d'accélération* et *angle de sortie*. La fenêtre de sortie du tube et le filtrage extérieur, qui est utilisé pour sélectionner la partie du spectre qui donne le meilleur contraste, sont pris en compte au travers de leur transmission μ_{filtre} et leur epaisseur l_{filtre} pour obtenir le spectre $n_1(u,\alpha,E)$ à l'entrée de l'objet à analyser :

$$n_1(u,\alpha,E) = n_0(u,\alpha,E) \cdot e^{-\mu_{\text{filtre}}(E)\cdot l_{\text{filtre}}} \tag{10.1}$$

10.2.2 Modèle de l'objet

Si l'on connaît la disposition tridimensionnelle des différents tissus de l'objet à analyser, on peut calculer pour chaque épaisseur de tissu l_{tissu}, de la même manière que pour le filtre, sa transmission et le spectre $n_i(u,\alpha,E)$ derrière cette épaisseur :

$$n_i(u,\alpha,E) = n_{i-1}(u,\alpha,E) \cdot e^{-\mu_i(E)\cdot l_i} \tag{10.2}$$

$$n(u,\alpha,E) = n_0(u,\alpha,E) \cdot e^{-\sum_i \mu_i(E)\cdot l_i} \tag{10.3}$$

Nous allons détailler notre modèle dans la section 10.3 et montrer quelques simplifications utiles.

10.2.3 Modèle du détecteur

En utilisant les modèles décrits ci-dessus et en connaissant la géométrie de l'appareil de mammographie, il est possible de calculer le nombre de photons ($n(x,y,E)$) de chaque énergie (E) qui arrivent à la surface d'un pixel (x,y) du capteur de rayons X. Pour trouver le niveau de gris $I_s(x,y)$ retourné par le capteur, nous utilisons la fonction d'efficacité $\eta(E)$ du capteur en fonction de l'énergie des photons et nous intégrons sur tout le spectre :

$$I_s(x,y) = \int_0^{E_m} n(x,y,E) \cdot \eta(E) dE \tag{10.4}$$

10.2.4 Modèle du bruit

L'analyse du bruit du capteur a montré que la source principale du bruit est le bruit quantique généré par le nombre limité de photons X arrivant sur un pixel, suivi par le bruit électronique (σ_e) du circuit de lecture et le bruit de quantification du convertisseur analogique numérique (σ_a). La variance du bruit quantique (σ_q) est donnée par la racine carré du nombre de photons :

$$\sigma_q(x,y) = \sqrt{\int n(x,y,E) dE} \tag{10.5}$$

Les autres sources de bruit sont considérées constantes. Puisque les sources de bruit peuvent être considérées indépendantes, le bruit total est égal à :

$$\sigma_q(x,y) = \sqrt{\sigma_q^2(x,y) + \sigma_e^2 + \sigma_a^2} \tag{10.6}$$

10.3 Modèle d'une microcalcification

Le modèle présenté jusqu'ici est très général, et permet même de prendre en compte les effets de durcissement du spectre, mais il a deux inconvénients :

– la composition de l'objet à visualiser doit être connue en trois dimensions.
– les calculs sont complexes

Pour arriver à des résultats analytiques, et pour simplifier les calculs de manière à ce qu'ils soient réalisables en temps réel dans une application comme l'aide à la détection, nous proposons quelque simplifications.

Pour simplifier le modèle de l'interaction entre les rayons X et la matière, on peut considérer uniquement un rayonnement monochromatique. L'intensité du rayonnement après une couche d'épaisseur l est alors donnée par :

$$n(l) = n(0) \cdot e^{-\mu \cdot l} \tag{10.7}$$

avec :

μ coefficient d'atténuation linéaire

$n(0)$ intensité du faisceau incident

$n(l)$ intensité après traversée du tissu d'épaisseur l

Si le coefficient d'atténuation n'est pas constant à travers le tissu, nous obtenons une intégrale :

$$n(l) = n(0) \cdot e^{-\int_{z=0}^{l} \mu(z) \cdot dz} \tag{10.8}$$

En première approximation, nous allons supposer que la microcalcification a une forme sphérique et qu'elle se trouve dans un tissu homogène. L'intensité reçue en un point du détecteur

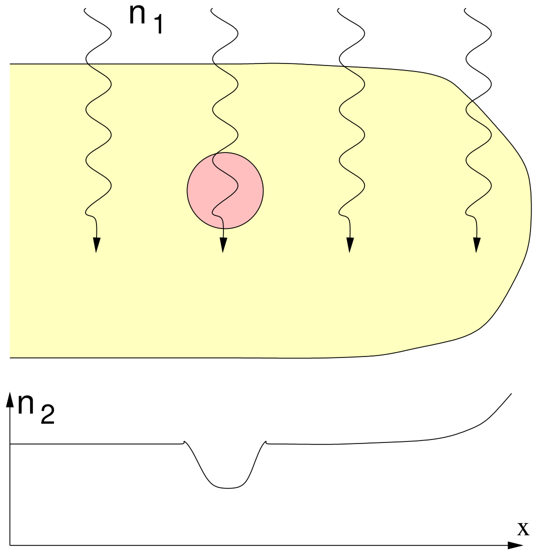


FIG. 10.2 – Projection d'une microcalcification

peut alors être calculée par :

$$n(x,y) = n(0) \cdot e^{-\mu_{tissu} \cdot l_{tissu}(x,y) - \mu_{calc} \cdot l_{calc}(x,y)} \tag{10.9}$$

En supposant une épaisseur totale $l_{tot} = l_{tissu} + l_{calc}$ constante dans un environnement autour de la microcalcification et une différence des coefficients d'atténuation de $\Delta\mu = \mu_{calc} - \mu_{tissu}$, nous pouvons écrire :

$$n(x,y) = n(0) \cdot e^{-\mu_{tissu} \cdot l_{tot} - \Delta\mu \cdot l_{calc}(x,y)} \tag{10.10}$$

L'intensité n_{env} dans l'environnement est alors :

$$n_{env} = n(0) \cdot e^{-\mu_{tissu} \cdot l_{tot}} \tag{10.11}$$

et l'intensité sous la calcification :

$$n(x,y) = n_{env} \cdot e^{-\Delta\mu \cdot l_{calc}(x,y)} \tag{10.12}$$

L'hypothèse d'une microcalcification sphérique et le choix d'un système de coordonnées centré autour du centre de cette sphère de rayon r nous donnent :

$$n(x,y) = n_{env} \cdot e^{-\Delta\mu \cdot \sqrt{(r^2 - x^2 - y^2)}} \tag{10.13}$$

Si la microcalcification est suffisamment petite pour que l'exposant soit beaucoup plus petit que 1, nous pouvons simplifier cette expression avec :

$$e(-ax) \approx 1 - ax \tag{10.14}$$

pour obtenir :

$$n(x,y) \approx n_{env} \cdot (1 - \Delta\mu \cdot \sqrt{(r^2 - x^2 - y^2)}) \tag{10.15}$$

Puisque la réponse du détecteur utilisé est linéaire et que le facteur de conversion est connu, nous avons, avec l'équation (10.15), une approximation de l'image de la microcalcification.

L'extraction des points centraux de ces sphères peut se faire avec l'opérateur *chapeau haut-de-forme* comme décrit dans la section 8.5. Pour trouver les calcifications avec un diamètre d_m (voir figure 10.3), il suffit de faire une ouverture avec un cercle d'un diamètre légèrement supérieur

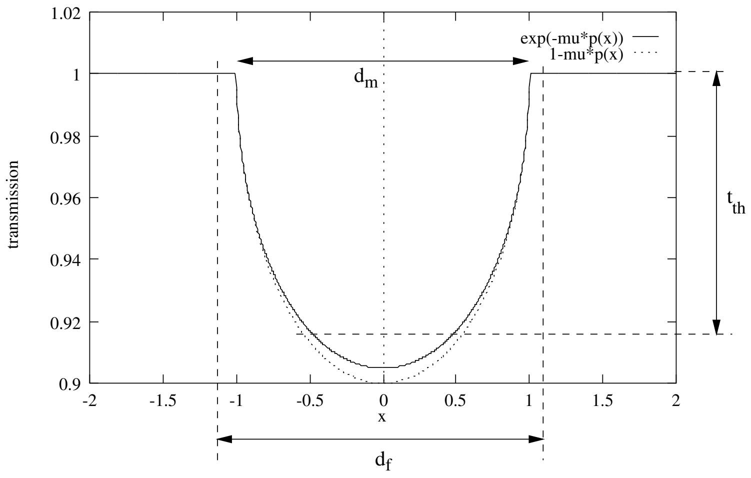


FIG. 10.3 – Détection des microcalcification avec un *chapeau haut-de-forme*

$d_f = d_m + \Delta d$. Cette fermeture nous fournit l'image du tissu, en effaçant les microcalcifications plus petites que d_f. La soustraction avec l'image d'origine nous donne alors toutes les structures plus petites que d_f, et donc les microcalcifications recherchées.

Pour les séparer du bruit et des autres structures fines, nous appliquons un seuil t_{th}, que nous pouvons calculer à partir de l'équation 10.15. Pour bien choisir ce seuil, nous allons regarder deux détails que nous avons négligés jusqu'ici : la résolution spatiale et la position de la microcalcification par rapport aux bords des pixels du capteur.

10.3.1 Influence de la discrétisation de l'espace

Les microcalcifications n'ont pas de taille minimale et il est intéressant de les détecter le plus tôt possible, c'est-à-dire quand elles sont encore très petites. Cependant, la taille des pixels du détecteur n'est pas infiniment petite. La plupart des images utilisées dans ce travail a été acquise à l'aide d'un détecteur dont la taille des pixels est de $100\mu m \times 100\mu m$.

En partant de l'équation simplifiée (10.15), qui représente la transmission d'un objet sphérique et en superposant cette sphère à l'espace du détecteur comme dans la figure 10.4, il est possible de calculer les valeurs de l'intensité obtenues par intégration sur l'aire du pixel.

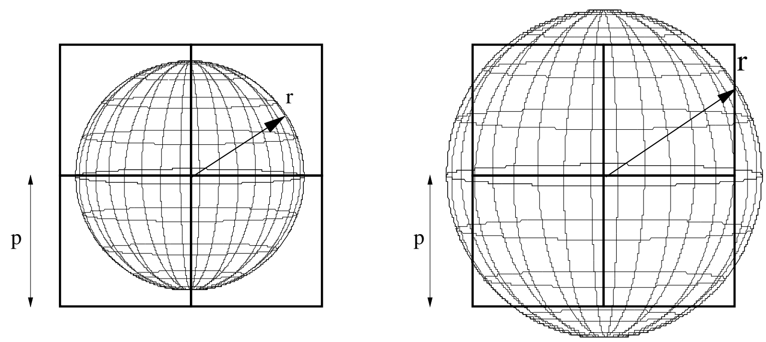


FIG. 10.4 – Rapport de taille entre les microcalcifications et les pixels

Dans une première phase, nous considérons seulement le cas le plus défavorable qui est donné quand l'atténuation maximale de la sphère est partagée sur quatre pixels, c'est-à-dire que le centre de la sphère est aligné avec le bord des pixels. L'intensité sur un pixel $p_{(x,y)}$ est alors donnée par :

$$n(p_{(x,y)}) = n_{env} \cdot t(p_{(x,y)}) \approx n_{env} \int_0^p \int_0^p (1 - \Delta\mu \cdot \sqrt{(r^2 - x^2 - y^2)})dydx \tag{10.16}$$

ce qui donne :

$$t(p_{(x,y)}) \approx \begin{cases} 1 - \Delta\mu\left(\frac{\pi}{3}r^3\right) & \text{pour} \quad r < p \\ 1 - \Delta\mu\left(\frac{\pi}{3}r^3 - \frac{\pi(r-p)}{6}(3r^2 + (r-p)^2)\right) & \text{pour} \quad d < r < \sqrt{2}p \\ 1 - \Delta\mu(r) & \text{pour} \quad r >> p \end{cases} \tag{10.17}$$

Pour des microcalcifications plus petites que le pixel, l'atténuation augmente avec r^3, comme le volume de la sphère. Pour les plus grandes microcalcifications, l'atténuation augmente avec r, comme l'épaisseur maximale de la sphère. Si l'atténuation devient importante, l'approximation (10.14) n'est plus très précise (la différence est illustrée sur la figure 10.5).

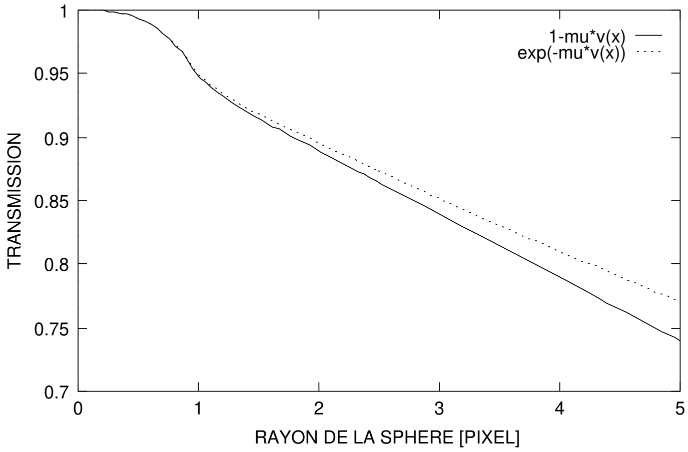


FIG. 10.5 – Transmission en fonction de la taille de la sphère

10.3.2 Influence de la position par rapport aux pixels

Sur la figure 10.6, la distribution du contraste maximal en niveau de gris est tracée en fonction de la taille de la sphère de calcium et de la position par rapport aux frontières des pixels. Dans cet histogramme, chaque paquet représente une sphère d'un rayon compris entre 0,3 pixel et 1,7 pixels. La largeur des paquets montre la gamme de variation du contraste maximal en fonction de la position de la calcification par rapport aux pixels.

On peut en déduire, par exemple, que si toutes les microcalcifications d'un rayon de 1,1 pixels doivent être détectées (dans les mêmes conditions d'acquisition que celles utilisées pour la construction de la figure 10.6), le seuil à appliquer doit être de l'ordre de 7 niveaux de gris, puisque la limite inférieure du paquet de rayon de 1,1 pixels est 7.

10.4 Estimation de la performance du système de détection

Pour évaluer la performance d'un système de détection de microcalcifications, il est intéressant de connaître le nombre de faux positifs que le système produit pour un taux de vrais positifs donné. Pour estimer ces taux, on peut utiliser des simulations reposant sur la méthode Monte-Carlo. On construit tout d'abord un ensemble d'images à partir du modèle du sein et de la microcalcification

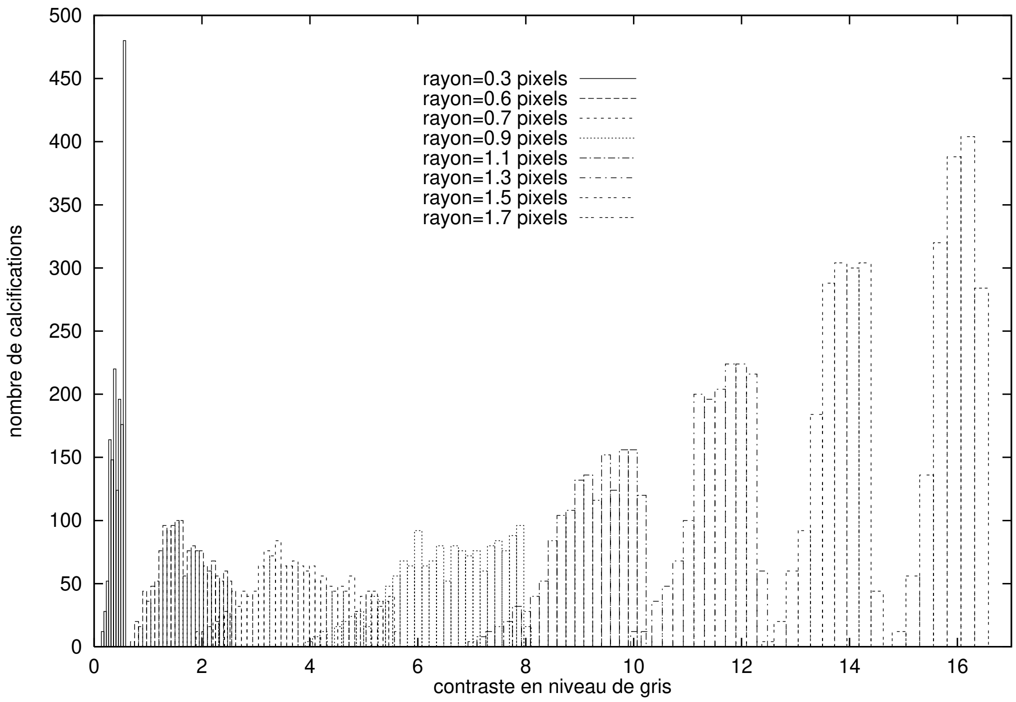


FIG. 10.6 – Contraste en fonction de la taille de la sphère et de la position de sa projection sur le détecteur

à détecter, en prenant en compte le système d'acquisition et son bruit, et ensuite on applique le système de détection sur ces images.

Si l'on considère la détection des microcalcifications, représentées par des sphères radio-opaques, par l'opérateur *chapeau haut-de-forme*, les taux de vrais et de faux positifs peuvent être calculés d'une manière beaucoup plus simple :
La probabilité que le seuil t appliqué sur un pixel de l'image de *chapeau haut-de-forme* soit en dessous de la valeur réelle du pixel, bien qu'à cet endroit il n'y ait que du bruit dans l'image, est donnée par :

$$\tilde{p}_{fpr}(x,y) = \int_{t+bg(x,y)}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_n(x,y)} e^{-\frac{1}{2}\left(\frac{g-bg(x,y)}{\sigma_n(x,y)}\right)^2} dg \qquad (10.18)$$

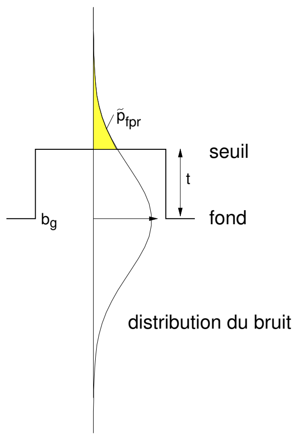


FIG. 10.7 – Calcul du taux de faux positifs par pixel

Cette probabilité p_{fpr}, qui est calculée par l'intégrale du bruit comme illustré sur la figure 10.7, correspond au taux de faux positifs par pixel. Pour les N pixels de l'élément structurant, la probabilité est donnée par :

$$p_{fpr}(x,y) = 1 - (1 - \tilde{p}_{fpr}(x,y))^N \tag{10.19}$$

Et le nombre total de faux positifs dans l'image est alors :

$$n_{fpr} = \sum_y \sum_x p_{fpr}(x,y) \tag{10.20}$$

Le calcul du nombre de faux négatifs utilise le même principe, mais, cette fois-ci, la probabilité de ne détecter aucun pixel de l'élément structurant est calculée, si cet élément est placé à l'endroit d'une microcalcification sphérique. Sur la figure 10.8, qui montre une illustration de ce calcul, on voit le niveau de fond auquel est ajoutée l'atténuation d'une sphère opaque.

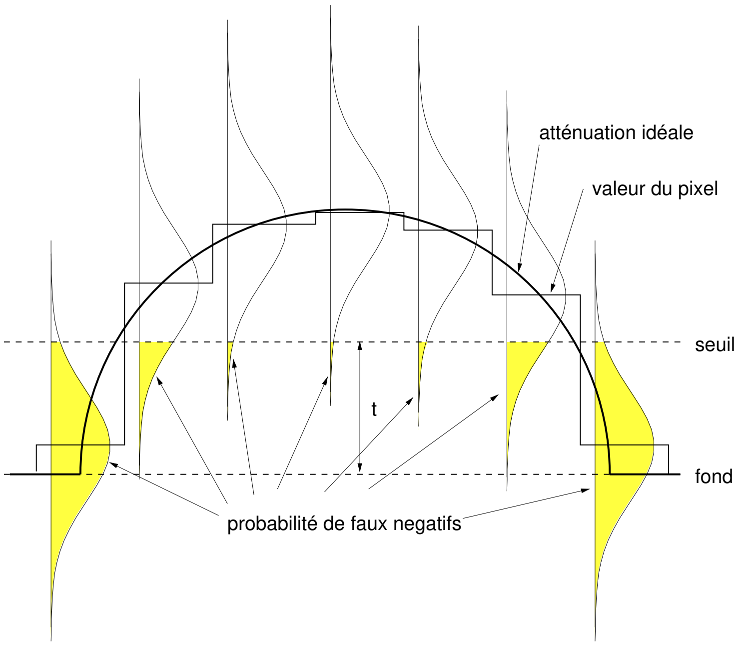


FIG. 10.8 – Calcul du taux de faux négatifs par pixel

Ensuite le résultat est discrétisé, et pour chaque pixel, le niveau de bruit est ajouté. La probabilité p_{fnr}, pour qu'un pixel ne soit pas détecté, est alors :

$$\tilde{p}_{fnr}(x,y) = \int_{-\infty}^{t+bg(x,y)} \frac{1}{\sqrt{2\pi}\sigma_n(x,y)} e^{-\frac{1}{2}\left(\frac{g-s(x,y)}{\sigma_n(x,y)}\right)^2} dg \tag{10.21}$$

Si le bruit des pixels voisins est considéré comme non-correlé, la probabilité de ne détecter aucun des pixels de l'élément structurant est alors :

$$p_{fnr}(C) = \prod_{\forall(x,y)\in C} \tilde{p}_{fpr}(x,y) \tag{10.22}$$

A partir des taux de faux positifs et de faux négatifs, en fonction du seuil appliqué à l'opérateur *chapeau haut-de-forme*, il est possible de calculer la courbe COR (annexe A), qui donne un bon résumé de la performance de détection du système.

10.5 Applications du modèle de simulation

Le modèle présenté ici contient un ensemble de modèles des différentes parties d'un système complet de détection des microcalcifications. Bien que certaines parties soient particulières au système spécifique présenté ici, il est possible d'extraire quelques informations générales qui sont valables au-delà de cette configuration spécifique. Le système de détection des microcalcifications réalisé ici par des opérateurs *chapeau haut-de-forme* peut être remplacé par une autre méthode de détection, comme par exemple un moyennage local avec seuillage, sans que cela change fondamentalement l'influence des autres paramètres du système. Il est donc intéressant d'utiliser le modèle pour explorer les limites de la détectabilité des microcalcifications dans le mode de fonctionnement qui est aujourd'hui appliqué au dépistage du cancer du sein.

Une question importante, pour obtenir un bon compromis entre le nombre de cancers trouvés et le risque pour les patientes dû aux effets ionisants des rayons X, est la dose utilisée pour un cliché de mammographie.

10.5.1 Influence de la dose et du choix de l'anode

Nous avons utilisé le modèle pour prédire la dose qui est nécessaire pour détecter les microcalcifications avec une certitude acceptable, en fonction de la taille des microcalcifications et de la combinaison de l'anode et du filtre du tube.

Les courbes de la figure 10.9 montrent la dose (en mAs) qui est nécessaire pour détecter des microcalcifications dans un sein dense, avec une surface sous la courbe COR égale à 0,999.

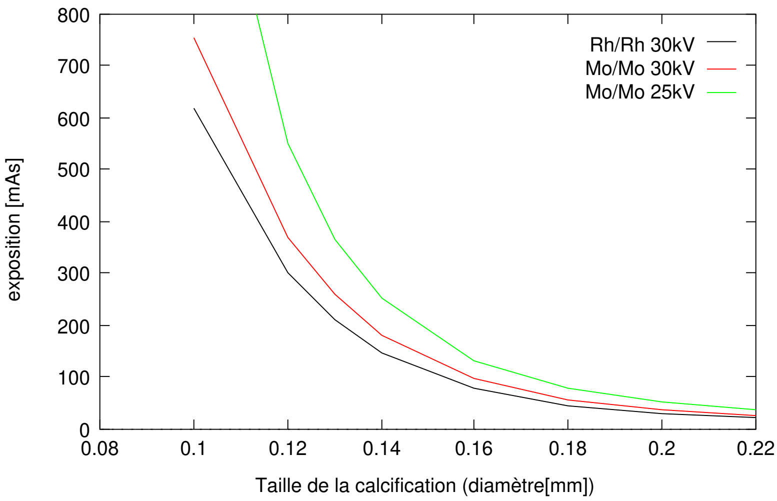


FIG. 10.9 – Dose nécessaire pour la détection des microcalcifications de différent taille dans un sein dense

On peut observer que la dose nécessaire pour la détection des microcalcifications augmente d'une manière considérable si le diamètre des microcalcifications descend au dessous de 0,15 mm.

10.5.2 Influence de la taille du pixel

Le modèle est utilisé également pour analyser l'effet de la taille du pixel sur la performance de détection, en gardant tous les autres paramètres, comme anode et dose, constants.

La figure 10.10 montre la surface A_z sous la courbe COR pour trois tailles différentes de pixel en fonction du diamètre de la microcalcification. Dans la figure 10.10, à gauche, les conditions choisies correspondent à un sein dense avec des paramètres d'acquisition typiques pour ce cas

(Rh/Rh, 30kV, 100mAs). Puisque la courbe COR est calculée pour la classification pixel par pixel, la surface A_z qui correspond à une performance acceptable est de l'ordre de $0,999$, c'est-à-dire que la partie intéressante se trouve en haut de la figure 10.10. On peut observer que la surface sous la courbe COR augmente généralement avec une diminution de taille de pixel ; par contre, pour des valeurs A_z élevées, la performance n'augmente plus quand la taille du pixel est en-deçà de $100\mu m$. Sous ces conditions, la limite de détectabilité est déterminée par le bruit quantique et est alors directement liée à la dose. Si la dose est augmentée - ou si on regarde dans les zones adipeuses avec les conditions de l'acquisition typique pour un sein dense - ce qui a été simulé pour obtenir la figure 10.10 à droite, la plus petite taille de microcalcifications détectable diminue et l'influence de la taille du pixel est visible au-dessous de $100\mu m$. Ce cas, qui n'est pas représentatif de l'utilisation clinique actuelle de la mammographie, montre qu'il faut augmenter la dose pour pouvoir profiter d'une taille de pixel inférieure à $100\mu m$.

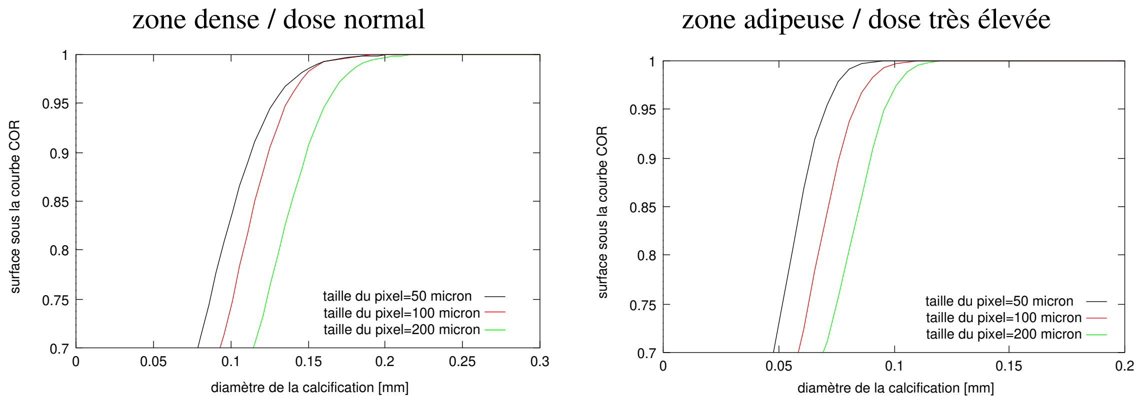


FIG. 10.10 – Surface A_z sous la courbe COR en fonction de la taille des microcalcifications et de la taille du pixel du capteur, pour une zone dense (à gauche) et une zone adipeuse (à droite)

Il est intéressant de mentionner dans ce contexte, une étude menée par Heang-Ping Chan et.al. [Pin99] sur la performance de lecteurs humains pour la caractérisation des lésions (bénin ou malin). Leur résultats, bien qu'ils n'aient pas encore atteint une certitude statistique de 95%, indiquent que la taille optimale du pixel se trouve entre $70\mu m$ et $105\mu m$, et ils suggèrent qu'une taille de pixel encore plus petite ($35\mu m$) donnerait des résultats plus mauvais.

10.6 Choix des seuils adaptatifs basé sur le modèle

Les applications du système de simulation présenté dans ce chapitre ne se limitent pas à l'exploration des limites de détection et à l'optimisation des paramètres du système d'acquisition ; il permet également d'optimiser les paramètres de la méthode de détection des microcalcifications.

La méthode de détection présentée dans la section 8.5 nécessite une stratégie pour fixer les seuils des opérateurs *chapeau haut-de-forme* pour chaque taille de calcification. Ces seuils sont très dépendants du bruit dans l'image, qui lui-même dépend des paramètres d'acquisition. En utilisant le système de simulation, il est possible de calculer la performance théorique du système de détection dans les conditions précises de l'acquisition. Autrement dit, cette simulation peut prédire, par exemple, le nombre de faux positifs qui seront générés à cause du bruit par un certain choix de seuil. Cette connaissance du taux théorique de faux positifs indique une nouvelle façon de réaliser le seuillage, illustré sur la figure 10.11. Pour chaque pixel de l'image, le nombre de faux positifs est calculé. Les valeurs réelles de l'opérateur *chapeau haut-de-forme* et du niveau de fond dans l'environnement (donné par l'ouverture) sont introduits dans le système de simulation pour trouver quel taux de faux positifs théorique serait obtenu si on détectait ce pixel comme une microcalcification. Un faible nombre de faux positifs indique qu'un pixel est suffisamment contrasté pour qu'il soit détectable dans une image acquise dans les conditions actuelles. Ce taux de faux positifs est une bonne mesure pour la combinaison des résultats de détection, à la fois à l'intérieur d'une image, pour une taille d'élément structurant donnée, et pour les résultats des élément structurants de taille différente. La détection, comme indiquée sur la figure 10.11, se fait donc sur l'image des minima des taux de faux positifs pour les différents éléments structurants. Cela revient à choisir en chaque endroit sur l'image, l'élément structurant le mieux adapté à la détection de la microcalcification qui se trouve potentiellement à cet endroit. Le seuillage global sur l'image ainsi obtenue peut se faire avec un seuil fixe, puisque l'ensemble de la variabilité liée aux conditions d'acquisition a été pris en compte en utilisant le modèle de simulation.

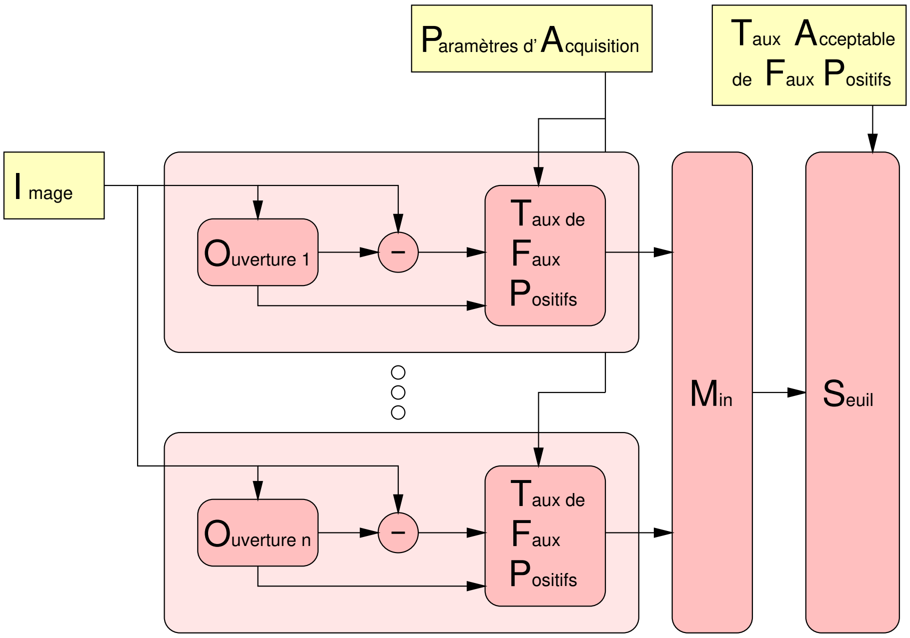


FIG. 10.11 – Détection des microcalcifications en utilisant le taux théorique de faux positifs par pixel

10.7 Evaluation sur des images cliniques

Pour valider la chaîne de simulation, nous comparons le résultat théorique à des résultats obtenus avec le même système de détection sur des images réelles acquises avec le système *Senographe2̃000D* de mammographie plein champ de GE (pour plus de détail sur le système voir [Mul99]).

10.7.1 Comparaison du taux de faux positifs théorique au taux de positifs mesuré sur des images cliniques

Le nombre théorique de faux positifs a été calculé pour des conditions d'acquisition qui correspondent à celles d'une vraie acquisition d'un sein asymptomatique avec le système Senographe 2000D. Le résultat pour le taux de faux positifs, en fonction du seuil appliqué à l'opérateur *chapeau haut-de-forme* et en fonction du niveau de fond, est donné sur la figure 10.12 à gauche. Le nombre de positifs réellement mesurés avec le même système de détection est donné sur la figure 10.12 à droite. Puisque le sein est asymptomatique, il n'y a donc pas de microcalcifications et le nombre de positifs peut être considéré comme un nombre de faux positifs (en supposant qu'il n'y ait pas de microcalcifications ratées par les experts). Dans l'échelle utilisée, les niveaux de fond qui correspondent aux pixels à l'intérieur du sein, sont compris entre 0 et 5000. On peut constater que le taux de faux positifs calculé correspond assez bien aux valeurs mesurées sur une image réelle pour les niveaux de gris à l'intérieur du sein (entre 0 et 5000). Autrement dit, pour cette région, le nombre de faux positifs générés par la détection est très proche du nombre induit par le bruit quantique. La détection des microcalcifications à l'intérieur du sein, et spécialement dans les zones denses, est donc principalement le problème de la séparation du bruit quantique du signal utile.

Les différences entre la simulation et la réalité sont plus importantes pour les zones au bord du sein, où la dose qui arrive sur le détecteur est importante. Dans ces conditions, le bruit quantique n'est pas la raison majeure du nombre de faux positifs, et le problème de la détection des microcalcifications est plutôt le problème de la distinction entre les microcalcifications et les autres structures, comme les croisements de fibres.

simulation :
taux de faux positifs

mesure :
taux de positifs (vrais et faux)

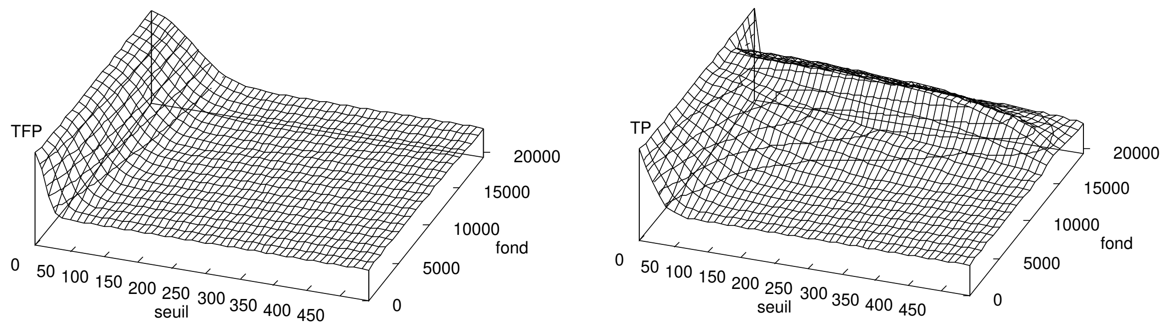


FIG. 10.12 – Comparaison entre le taux de faux positifs prédit par la simulation et le nombre de pixels détectés sur une image réelle

10.8 Conclusions

Le système de simulation de la chaîne entière d'acquisition d'image de mammographie, du tube jusqu'à première étape d'un système d'aide au diagnostic, permet à la fois de comprendre les limites physiques de la détection et d'optimiser les paramètres de cette chaîne. L'inclusion de ce modèle de simulation dans le système de détection des microcalcifications, permet de prendre en compte tous les paramètres du système d'acquisition lors de la détection, par l'intermédiaire du taux de faux positifs théorique.

La comparaison des résultats théoriques de performance aux valeurs obtenues sur des images réelles montre que le système prédit correctement le nombre de faux positifs dans les zones denses du sein, ce qui correspond aux zones les plus critiques sur les images de mammographie. Dans ces conditions, le paramètre le plus important pour la détection est le niveau de bruit quantique, et ce niveau est prédictible à partir des paramètres d'acquisition. Pour la détection des microcalcifications, la variabilité dans les images peut donc être réduite d'une manière significative, si les paramètres d'acquisition sont connus. Avec la généralisation des systèmes numériques de mammographie, l'utilisation des paramètres d'acquisition dans le système d'aide au diagnostic devrait donc devenir inévitable.

Chapitre 11

Conclusion et Perspectives

La performance des experts humains, face à la plupart des tâches de traitement d'images, a été longtemps très supérieure à celle des systèmes de vision artificiels. Cette supériorité commence à diminuer, d'une part grâce à l'accroissement de la puissance de calcul disponible, et d'autre part grâce aux algorithmes de traitement de plus en plus sophistiqués.

Un des points forts de l'observateur humain est sa possibilité d'apprendre une tâche avec très peu de cas d'apprentissage et d'étendre cette connaissance à de nouveaux cas. Cette extension des connaissances fonctionne même si la base d'apprentissage d'origine contient beaucoup de variabilité.

11.1 Nos travaux

Dans nos travaux, nous avons essayé de concevoir un système d'apprentissage pour des tâches de classification en traitement d'images, qui prend en compte la variabilité entre images.

Dans un système de classification, il y a deux endroits dans lesquels la variabilité peut être prise en compte :

- au moment du calcul des attributs
- pendant la décision.

La réduction de la variabilité au moment du calcul des attributs demande, en général, une très bonne connaissance explicite du domaine. Nous avons montré un tel exemple pour la problématique de la détection des microcalcifications face au bruit dans des images de mammographie. Le modèle de simulation que nous proposons, permet de prendre en compte l'ensemble des paramètres d'acquisition disponibles pour ce type d'examen.

Pour prendre en compte la variabilité lors de la prise de décision, nous proposons un système qui permet d'apprendre à la fois la position des classes dans l'espace des attributs disponibles pour la tâche de classification, et son adaptation à l'image courante.

L'approche que nous avons choisie repose sur la modélisation des densités des histogrammes des attributs par une somme de fonctions paramétriques. Nous avons utilisé des sommes de composantes gaussiennes pour lesquelles il existe des méthodes d'estimation (algorithme EM), mais nous avons montré comment cet algorithme peut être appliqué à d'autres formes de composantes.

Les liens entre les paramètres des composantes et leurs variations sont estimés à partir de la base d'apprentissage avec une démarche en deux étapes : le regroupement des composantes dans des groupes, pour établir un ordre unique de ces composantes pour toutes les images, et une régression linéaire des interactions entre paramètres.

L'utilisation du système de classification sur de nouveaux exemples passe par l'introduction de cette connaissance dans l'algorithme d'estimation des paramètres, que nous avons baptisé P-EM (prototype-EM). Finalement, les densités sont converties en fonctions d'appartenance par une étape de normalisation. Nous avons décrit la méthode de normalisation la plus connue basée sur les probabilités, et nous avons proposé deux nouvelles méthodes qui se basent sur la façon dont la vérité sur les lésions est recueillie auprès des experts.

11.2 Les résultats

Nous avons appliqué cette méthode à une base de données synthétique et à des images de mammographie. Dans les deux cas, nous avons pu constater une amélioration des performances de classification par rapport aux fonctions d'appartenance non-adaptatives et même par rapport aux limites théoriques d'un classificateur statique en général.

Les résultats montrent que la variabilité entre images peut être estimée, au moins partiellement, par les fonctions d'appartenance adaptatives proposées et que cette estimation permet une augmentation de la performance de classification.

Nous avons également montré que l'utilisation d'un système de simulation de l'acquisition d'une image de mammographie, permet de prendre en compte les paramètres d'acquisition lors de la détection de microcalcifications. La comparaison de la performance du système de détection aux valeurs théoriques déterminées par simulation, a montré que la précision du modèle de simulation est suffisante pour prédire la performance dans les zones denses du sein.

11.3 Perspectives

Après avoir constaté l'utilité des fonctions d'appartenance adaptatives dans le traitement d'images, plusieurs directions de poursuite des travaux sont intéressantes. Les questions qui se posent sont :

– Quelles sont les fonctions paramétriques les plus efficaces pour modéliser les histogrammes ? Il est probable que cette question ne pourra pas être résolue indépendamment du problème posé.
– Quels sont les modèles d'interaction qui permettent d'exprimer le mieux la variabilité et qui peuvent être déterminés à partir d'une petite base d'apprentissage ?
– Quelle est la méthode de régression la mieux adaptée pour estimer ces interactions ?
– Comment est-il possible de trouver le nombre de composantes optimales pour décrire d'une manière efficace les densités dans les histogrammes ?

11.3.1 Construction du système d'apprentissage complet

Les travaux présentés ici montrent uniquement la faisabilité de la modélisation de la variabilité sur un seul attribut et très peu d'exemples. Dans l'extension de ce travail, il sera intéressant d'appliquer la méthode à d'autres attributs pour construire un système plus puissant de détection des lésions en mammographie. Il sera intéressant d'évaluer l'augmentation de la performance pour un système complet utilisant plusieurs attributs, par exemple dans un système à base d'arbre de décision. Avec l'adaptation des fonctions d'appartenance au cas actuel, on peut espérer trouver des règles plus générales, c'est-à-dire qui s'appliquent à une partie plus grande de la base. En théorie, cela devrait permettre de réduire le nombre des règles ou le nombre de nœuds nécessaires dans le système.

11.3.2 Évaluation sur les images du détecteur numérique

Les premières évaluations ont été réalisées sur la base d'images acquises sur film et numérisées ensuite, étant donné que des images du détecteur numérique n'étaient pas encore disponibles. Dès maintenant, une base d'une taille intéressante (>4000 images) est à notre disposition, mais leur préparation pour l'utilisation comme base d'apprentissage ou base de test demande encore un travail important (établissement de la *vérité clinique* par radiologues et par des examens d'histologie ou du suivi à 5 ans.)

11.3.3 Application à d'autres domaines

Finalement, il est aussi intéressant d'appliquer cette méthode dans des domaines de traitement d'image complètement différents pour évaluer à quel point la méthode est généralisable.

Théoriquement, la méthode peut s'appliquer également à des applications autres que le traitement d'images comme des problèmes de classement avec un nombre importantes d'exemples, affichant des propriétés similaires. Un champ d'application possible concerne toutes les méthodes de caractérisations des matériaux par spectrométrie (par exemple diffraction par rayons-X, ou spectrographie de masse).

Annexes

Annexe A

Les courbes COR

Pour les problèmes de classification en médecine (mais aussi dans d'autres domaines) l'a priori de chaque décision n'est pas absolu mais basé sur le jugement de l'expert. Plus concrètement dans notre application, il s'agit de faire le compromis entre le risque de ne pas détecter un cancer et le désagrément de faire passer des examens supplémentaires sur un grand nombre de patientes.

Les courbes COR (caractéristiques opérationnelles du récepteur) [Ber94] sont souvent utilisées pour évaluer la performance de plusieurs compromis. Dans la courbe COR, le taux de vrais positifs t_{tvp} (cancers trouvés par rapport aux cancers présents) est tracé en fonction du taux de faux positifs t_{fp} (cancers indiqués qui ne le sont pas par rapport à tous les signes négatifs) :

$$t_{vp} = \frac{n_{vp}}{n_{vp} + n_{fn}} \tag{A.1}$$

$$t_{fp} = \frac{n_{fp}}{n_{fp} + n_{tn}} \tag{A.2}$$

La courbe COR pour un classificateur idéal est un rectangle, celle pour un classificateur aléatoire une droite sur la diagonale. L'intégrale de la surface sous la courbe COR est souvent utilisée comme mesure globale de la performance d'un algorithme de classification.

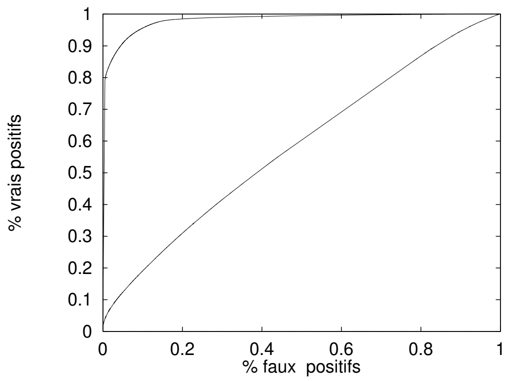


FIG. A.1 – Exemple d'une courbe COR pour un bon et un mauvais classificateur

A.1 Evaluation de la performance maximale d'un attribut

Pour estimer la puissance de discrimination maximale d'un attribut, nous utilisons les deux histogrammes calculés à partir de la connaissance de la vérité, c'est-à-dire du bord de la région à détecter. Pour chaque valeur de l'attribut dans les histogrammes, le nombre de pixels à l'intérieur et à l'extérieur de la région est connu. Si le système de classification décide d'attribuer la classe *à l'intérieur de la région* aux pixels ayant une certaine valeur pour l'attribut, il est donc possible de calculer le taux de vrais positifs et de faux positifs pour cette partie des pixels. Dans le cas opposé, c'est-à-dire quand le système de classification décide d'attribuer la classe *à l'extérieur de la région* aux pixels ayant cette valeur pour l'attribut, il est possible de calculer le taux de vrais négatifs et de faux négatifs.

Pour calculer la surface maximale sous la courbe COR, il faut prendre les décisions dans le bon ordre puisque la surface en dépend.

A.2 L'ordre de la prise de décision

Pour illustrer cet effet, nous considérons un cas extrême dans lequel les attributs permettent la classification parfaite sur la moitié des exemples et ne donnent aucune indication sur les classes pour la deuxième moitié. Si les décisions du classement sont d'abord prises sur les exemples faciles à classer, la courbe COR obtenue ressemble à celle de gauche dans la figure A.2. Dans

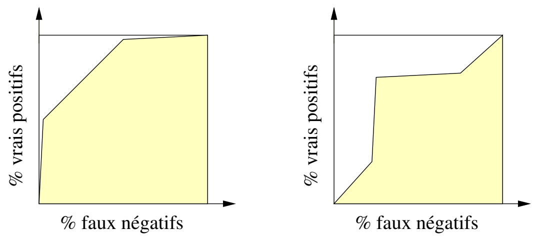


FIG. A.2 – Importance de l'ordre des décisions

le cas inverse, c'est-à-dire quand les décisions sont prises d'abord sur les exemples difficiles à classer, le résultat est plutôt celui indiqué à droite dans la figure A.2. Avec le critère de la surface sous la courbe COR, le deuxième ordre de décision est défavorable.

A.2.1 Le pouvoir de discrimination pour la décision suivante

En poursuivant ce raisonnement un peu plus loin, il est possible de définir une mesure de discrimination d'un attribut pour une classification de l'exemple suivant. Considérons une situation au milieu du processus de classification dans lequel la décision est prise sur une partie des exemples. Dans les exemples déjà classés, il y a un nombre de vrais positifs n_{vp}, de vrais négatifs n_{vn}, de faux positifs n_{fp} et de faux négatifs n_{fn}, ce qui limite la zone dans laquelle peut évoluer la courbe COR avec les décisions suivantes.

La figure A.3 illustre cette situation : la courbe COR est déjà déterminée dans ses limites. Si les exemples classés sont enlevés de la base, la courbe COR est vide, mais elle couvrait seulement le grand rectangle marqué en pointillé dans la figure A.3. Pour la décision suivante, il existe aussi un taux de vrais positifs, de vrais négatifs, de faux positifs et de faux négatifs, ce qui permet de calculer les angles a et b qui déterminent la direction dans laquelle la courbe COR va progresser en prenant cette décision. Ceci est indiqué avec les vecteurs de la figure A.3. L'angle du vecteur par rapport à la diagonale est une mesure pour la qualité de la classification et la longueur du vecteur indique la quantité de décision qui peut être prise avec cette qualité. Si la décision est basée seulement sur un seul attribut, les angles a et b sont une mesure de la pertinence de cet attribut pour la décision suivante à prendre.

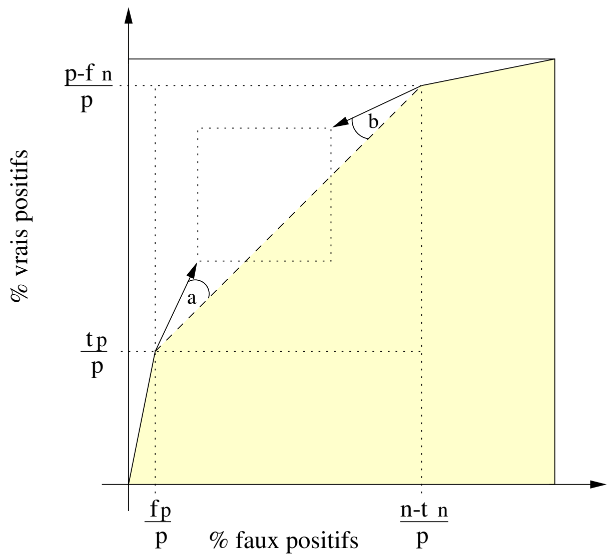


FIG. A.3 – Visualisation de la pertinence dans la courbe COR

A.2.2 Estimation du pouvoir de discrimination maximale

Pour obtenir la surface maximale sous la courbe COR l'algorithme suivant est utilisé :

1- Calculer les histogrammes à l'intérieur et à l'extérieur de la région

2- Classer les valeurs de l'histogramme dans l'ordre croissant du rapport entre les pixels à l'intérieur et à l'extérieur de la région

3- Ajouter toutes les valeurs de l'histogramme dans la classe *à l'extérieur*

4- Calculer les vrais positifs et les vrais négatifs et ajouter le point dans la courbe COR

5- Enlever la valeur avec le rapport maximal entre les pixels à l'intérieur et à l'extérieur de la région de la classe *à l'extérieur* et l'ajouter à la classe *à l'intérieur*

6- S'il y encore des valeurs dans la classe *à l'extérieur* recommencer à partir de 4

Annexe B

Liens de l'algorithme P-EM avec la probabilité à priori

Si nous considérons connue la densité de probabilité a priori de rencontrer les valeurs μ_k, p_k, σ_k : $P(p_k), P(\mu_k), P(\sigma_k)$ et que nous appliquons les méthodes de maximum de vraisemblance en utilisant ces connaissances, nous devons maximiser la probabilité globale, donnée par ce produit des probabilités :

$$P(\vec{x},\Phi) = P(\vec{x}|\Phi)\cdot P(p_k)\cdot P(\mu_k)\cdot P(\sigma_k) \tag{B.1}$$

En suivant un raisonnement similaire à celui du chapitre 3.2.3 nous obtenons :

$$\begin{aligned}\ln(P_k(\vec{x},\Phi,\vec{\lambda}_k)) &= \sum_{i=1}^{N}\ln(\frac{\lambda_{ki}}{\sqrt{2\pi}\sigma_k}) - \sum_{i=1}^{N}\frac{\lambda_{ki}}{2\sum_{i=1}^{N}\lambda_{ki}}\left(\frac{(x_i-\mu_k)}{\sigma_k}\right)^2 \\ &+ \ln P(p_k) + \ln P(\mu_k) + \ln P(\sigma_k) \end{aligned} \tag{B.2}$$

Si nous choisissons des gaussiennes pour les $P(p_k), P(\mu_k), P(\sigma_k)$:

$$P(p_k) = \frac{1}{\sqrt{2\pi}\sigma_{p,k}}e^{-\frac{1}{2}\left(\frac{p_k-\mu_{p,k}}{\sigma_{p,k}}\right)^2} \tag{B.3}$$

$$P(\mu_k) = \frac{1}{\sqrt{2\pi}\sigma_{\mu,k}}e^{-\frac{1}{2}\left(\frac{\mu_k-\mu_{\mu,k}}{\sigma_{\mu,k}}\right)^2} \tag{B.4}$$

$$P(\sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_{\sigma,k}}e^{-\frac{1}{2}\left(\frac{\sigma_k-\mu_{\sigma,k}}{\sigma_{\sigma,k}}\right)^2} \tag{B.5}$$

et appliquons la méthode du maximum de vraisemblance :

$$\ln(P_k(\vec{x},\Phi,\vec{\lambda}_k)) = \sum_{i=1}^{N}\ln(\frac{1}{\sqrt{2\pi}\sigma_k}) - \sum_{i=1}^{N}\frac{\lambda_{ki}}{2\sum_{i=1}^{N}\lambda_{ki}}\left(\frac{(x_i-\mu_k)}{\sigma_k}\right)^2 \tag{B.6}$$

$$+ \ln(\frac{1}{\sqrt{2\pi}\sigma_{p,k}}) - \frac{1}{2}\left(\frac{(p_k-\mu_{p,k})}{\sigma_{p,k}}\right)^2 \tag{B.7}$$

$$+ \ln(\frac{1}{\sqrt{2\pi}\sigma_{\mu,k}}) - \frac{1}{2}\left(\frac{(\mu_k-\mu_{\mu,k})}{\sigma_{\mu,k}}\right)^2 \tag{B.8}$$

$$+ \ln(\frac{1}{\sqrt{2\pi}\sigma_{\sigma,k}}) - \frac{1}{2}\left(\frac{(\sigma_k-\mu_{p,k})}{\sigma_{\sigma,k}}\right)^2 \tag{B.9}$$

nous retrouvons les formules d'iteration de l'algorithme P-EM :
(dérivée de μ)

$$\frac{\partial}{\partial \mu_k} \ln(P_k(\vec{x}, \Phi, \vec{\lambda}_k)) = 0 \tag{B.10}$$

$$\Rightarrow \quad -\sum_{i=1}^{N} \frac{\lambda_{ki} \cdot (x_i - \mu_k)}{\sum_{i=1}^{N} \lambda_{ki} \sigma_k^2} + \frac{\mu_k - \mu_{\mu,k}}{\sigma_{\mu,k}^2} = 0 \tag{B.11}$$

$$\Rightarrow \quad \mu_k \left(\sum_{i=1}^{N} \frac{\lambda_{ki} \cdot x_i}{\sigma_k^2} + \frac{1}{\sigma_{\mu,k}^2} \right) = \sum_{i=1}^{N} \frac{\lambda_{ki} \cdot \mu_k}{\sigma_k^2} + \frac{\mu_{\mu,k}}{\sigma_{\mu,k}^2} \tag{B.12}$$

$$\Rightarrow \quad \mu_k = \frac{\sum_{i=1}^{N} \lambda_{ki} \cdot x_i + \mu_{\mu,k} \frac{\sigma_k^2}{\sigma_{\mu,k}^2}}{\sum_{i=1}^{N} \lambda_{ki} + \frac{\sigma_k^2}{\sigma_{\mu,k}^2}} \tag{B.13}$$

$$\Rightarrow \quad \mu_k = v_k \cdot \frac{\sum_{i=1}^{N} \lambda_{ki} x_i}{\sum_{i=1}^{N} \lambda_{ki}} + (1 - v_k) \cdot \mu_{\mu,k} \tag{B.14}$$

avec :

$$v_k = \frac{\sum_{i=1}^{N} \lambda_{ki}}{\sum_{i=1}^{N} \lambda_{ki} + \frac{\sigma_k^2}{\sigma_{\mu,k}^2}} \tag{B.15}$$

Bibliographie

Intelligence artificielle

[Bri96] Garry Briscoe, Terry Caelli, *A Compendium of Machine Learning, Volume 1 : Symbolic Machine Learning*, Ablex Publishing Corporation 1996 (ISBN 1-56750-178-8)

[Cla89] Peter Clarke, Tim Niblett, *The CN2 induction Algorithm*, Machine Learning Journal, Vol. 3, No. 4, 1989, pp. 261 - 283
http ://www.cs.utexas.edu.users/pclark/papers/cn2.ps

[Tor95] I.S. Torsum, *Foundations of Intelligent Knowldge-Based Systems*, Academic Press 1995 (ISBN 0-12-696060-7)

Logique floue

[Bel96] Gleb Beliakov, *Fuzzy Sets and Membership Functions Based on Probabilities*, Intelligent Systems, Vol 91, 1996, pp. 95-111

[Blo99] Isabelle Bloch, *On fuzzy distances and their use in image processing under imprecision*, Pattern Recognition, Vol 32, 1999, pp. 1873-1895

[Bou93] B. Bouchon-Meunier, *Fuzzy similitude and approximate reasoning*, Advances in Fuzzy Theory and Technology, Editor P. P. Wang, Bookwrights Press, pp. 161–166, 1993

[Bou95] Bernadette Bouchon-Meunier, *La logique floue et ses applications*, Addison-Wesley 1995

[Dub80] D. Dubois, H. Prade, *Fuzzy Sets and Systems : Theory and Applications*, Academic Press, 1980

[Gat89] I. Gath, A. B. Geva, *Unsupervised Optimal Fuzzy Clustering*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, July. 1989, pp. 773-779

[Hal98] Lawrence O. Hall, Anand Namasivayam, *Using Adaptive Fuzzy Rules for Image Segmentation*, WCCI98, FUZZ-IEEE 98, International Conference on Fuzzy Systems May 1998, Vol. 2, pp. 1560-1565

[Jag95] René Jager, *Fuzzy Logic in Control*, PhD-thesis, Univeristy Delft, 1995

[Jua98] Chia-Feng Juang, Chin-Teng Lin, *An On-Line Self-Constructing Neural Fuzzy Inference Network and its Applications*, IEEE Transactions on Fuzzy Systems, Vol 6, No. 1, Feb. 1998, pp. 12-32

[Kar98] Nilesh N. Karnik, Jerry M. Mendel, *Introduction to Type-2 Fuzzy Logic Systems*, WCCI98, FUZZ-IEEE 98 International Conference on Fuzzy Systems May 1998, Vol. 2, pp. 915-921

[Koe98] Ludovic Koehl, Xianyi Zeng, Christian Vasseur *A Method for Optimizing Fuzzy Membership Functions*, Studies in Informatics and Control Vol. 7. No. 2, June 1992

[Kri93-1] F.C.H. Rhee, R. Krishnapuram, *Fuzzy Rule Generation Methods for High-level Computer Vision*, Fuzzy Sets and Systems, Vol 60, 1993, pp. 245-258

[Kri93-2] Raghu Krishnapuram, *A Possibilistic Approach to Clustering*, IEEE Transactions on Fuzzy Systems, Vol 1, No. 2, May 1993, pp. 98-110

[Kri95] R. Krishnapuram, H. Frigui, *Fuzzy and Possibilistic Shell Clustering Algorithms and Their Application to Boundary Detection and Surface Approximation : Part 1*, IEEE Transactions on Fuzzy Systems, Vol 3, No. 1, Feb. 1995, pp. 29-43

[Kri96] R. Krishnapuram, J. M. Keller, *The Possibilistic C-Means Algorithm : Insights and Recommendations*, IEEE Transactions on Fuzzy Systems, Vol 4, No. 3, Aug. 1996, pp. 385-393

[Kos92] Bart Kosko, *Neural Networks and Fuzzy Systems : A Dynamical System approach to Machine Intelligence*, Prentice-Hall 1992.

[Lee95] E. Stanley Lee, Qing Zhu, *Fuzzy and Evidence Reasoning*, Physica-Verlag 1995

[Pro79] T. Procyk, E. Mamdani, *A linguistic self-organising process controller*, Automatica, Vol. 15, pp. 15-30, 1979

[Rif96] Maria Rifqi, *Mesures de comparaison, typicalité et classification d'objets flous : théorie et pratique*, Thèse de doctorat de l'informatique, Université Paris 6, 1996

[Wan92] Li-Xin Wang, Jerry M. Mendel, *Fuzzy Basis Functions, Universal Approximation, and Othogonal Least Squares Learning*, IEEE Transactions on Neural Networks, Vol 3, No. 5, September. 1992, pp. 807-814

[Wu99] Tsu-Ping Wu, Shyi-Ming Chen, *A new method for Constructing Membership Functions and Fuzzy Rules from Training Exemples*, IEEE Transactions on Systems, Man, and Cybernetics B, Vol 29, No. 1, Feb. 1999, pp. 25-40

Régression

[Blo96] Michael Blount, Chiahung Lin, *Solving Fuzzy Regression Equations Using a Fuzzy Neural Network*, dans : International Conference on Fuzzy Systems, Fuzz96IEEE, 8.-11. Sep 1996, New Orleans

[Cel92] Aivars Celmins, *Nonlinear Least Square Regression in Fuzzy Vector Spaces*, dans : J. Kacprzyk, M. Fedrizzi, "Fuzzy Regression Analysis", Physica-Verlag Heidelberg, 1992

[DeV95] Richard De Veaux, *A Guided tour of Modern Regression Methods*, Turorial at 1995 Fall Technical Conference, St Luis.

[Fri91] Jerome H. Friedman, *Multivariate Adaptive Regression Splines*, The Annals of Statistic, Vol. 19, No. 1, 1991, pp 1 - 67

[Pre92] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery, *Numerical Recipes in C*, Cambridge Press, 1992

[Sak92] Masatoshi Sakawa, Hitoshi Yano, *Fuzzy Linear Regression and its Applications*, dans : J. Kacprzyk, M. Fedrizzi, "Fuzzy Regression Analysis", Physica-Verlag Heidelberg, 1992

[Sav92] Dragan Savic, Witold Pedrycz, *Fuzzy Linear Regression Models : Construction and Evaluation*, dans : J. Kacprzyk, M. Fedrizzi, "Fuzzy Regression Analysis", Physica-Verlag Heidelberg, 1992

[Tan82] Hideo Tanaka, S. Uejima, K. Asai, *Linear Regression Analysis with Fuzzy Model*, IEEE Transactions on System, Man and Cybernetics, Vol 12, 1982, pp. 903 - 907

[Tan92] Hideo Tanaka, Hisao Ishibuchi, *Possibilistic Regression Analysis Based on Linear Programming*, dans : J. Kacprzyk, M. Fedrizzi, “Fuzzy Regression Analysis”, Physica-Verlag Heidelberg, 1992

[Yag88] R. Yager, *On Ordered Weighted Averaging Operators on Multicriteria Decision-Making*, IEEE Transactions on System, Man and Cybernetics, Vol 18, 1988, pp. 183 - 190

Fouille de données

[Buh95] J. Buhmann, T. Hofmann, *Pairwise data clustering by deterministic Annealing*, Tech. Rep. IAI-TR-95-7, Institut für Informatik III, Universität Bonn

[Hol94] M. Holsheimer, A. Siebes, *Data Mining, The Search for Knowledge in Databases*, Tech. Rep. CS-R9406, CWI

Arbres de décision

[Mar98] Christophe Marsala, *Apprentissage inductif en présence de données imprécises : construction et utilisation d’arbres de décision flous*, Thèse d’informatique, Université Paris 6, 1998

[Mur94] Sreerama K. Murthy, Simon Kasif, Steven Salzberg, *A System for Induction of Oblique Decision Trees*, Journal of Artificial Intelligence Research, August 1994,1-13

Mélanges

[Aka95] Shotaro Akaho, *Mixture model for image understanding and the EM algorithm*, TR-95-13 Electronical Laboratory, Tsukuba, Ibaraki, Japan

[Amb96] Christophe Ambroise, Gerard Govaert, *Spatial Clustering and the EM-Algorithm*, Technical Report (DRAFT) Feb. 1996 Universite de Technologie de Compiegne

[Ayl98] Stephen R. Aylward, Bradley M. Hemminger, Etta D. Pisano *Mixture Modeling for Digital Mammogram Display and Analysis*, in : Nico Karssemeijer, Martin Thijssen, Jan Hendriks, Leon van Erning (Eds.) Digital Mammography, Nijmegen 1998, Kluwer Academic Publishers, 1998, pp. 305 - 312

[Cai97] Hélène Caillol, Wojciech Pieczynski, *Estimation of Fuzzy Gaussian Mixture and Unsupervised Statistical Image Segmentation*, IEEE Transactions on Image Processing, Vol. 6, No. 3, March 1997, pp. 425 - 440

[Cou96-1] Christophe Couvreur, *Hidden Markov Models and Their Mixtures*, Raport de Diplôme d’études approfondies en mathématiques, Université catholique de Louvain, 1996

[Cou96-2] Christopher Couvreur, *The EM Algorithm : A Guided Tour*, IEEE European Workshop on Computationaly Intensive Methods in Control and Signal Processing, Pragues, August 28-30, 1996

[Dem77] M. Dempster, N.M. Laird, D.B.Jain, *Maximum likelihood for incomplete data via the EM algorithm*, J. Royal Statistics Society, Series B. Vol. 39, 1977, pp 1 - 38

[Hel95] David P. Helmbold, Robert E. Schapire, Yoram Singer, Manfred K. Warmuth, *A comparison of new and old algorithms for a mixture estimation problem*, ACM, 1995, pp. 69 - 78

[Zhu96] Xinhua Zhuang, Yan Huang, K. Palaniappan, Yunxin Zhao, *Gaussian Mixture Density Modeling, Decomposition, and Applications*, IEEE Transactions on Image Processing Vol. 5, No.9, Sep 1996, pp. 1293-1302

Imagerie du sein

[Ack72] L. V. Ackerman, E. E. Gose, *Breast lesion classification by computer and xeroradiograph*, Cancer, Vol. 30, 1972, pp. 1025 - 1035

[Bak95] Jay A. Baker, Phyllis J. Kornguth, Joseph Y. Lo, Margaret E. Willford, Carey E. Floyd, *Breast Cancer : Prediction with Artificial Neural Network Based on BI-RADS Standardized Lexicon*, Radiology, Vol. 196, September 1995, pp. 817 - 822

[Ban97] Isaac N. Bankman, Tanya Nizialek, Inpakala Simon, Olga B. Gatewood, Irvin N. Weinberg, William R. Brody, *Segmentation Algorithms for Detecting Microcalcifications in Mammograms*, IEEE Transactions on Information Technology in Biomedicine, Vol. 1, No. 2, pp. 141 - 149

[BIRAD95] BI-RAD, *Breast Imaging Reporting and Data System*, American College of Radiology, sep. 1995

[Bot96-1] S. Bothorel , B. Bouchon Meunier , S. Muller, *Fuzzy logic based approach for mammographic images*, EUFIT'96, September 1996, Aachen, Germany.

[Bot96-2] S. Bothorel , B. Bouchon Meunier , S. Muller, *Extraction d'attributs flous pour l'analyse d'images mammographiques*, LFA'96, Dec. 1996, Nancy

[Bot96-3] S. Bothorel, B. Bouchon, S. Muller, *A Fuzzy Logic-based Approach for Semiological Analysis of Microcalcification in Mammographic Images*, International Journal of Intelligent Systems, Vol. 12, 1997, pp. 819-843

[Bot96-4] Sylvie Bothorel, *Analyse d'image par arbre de décision floue - Application à la classification sémiologique des amas de microcalcifications en mammographie numérique*, Thèse de l'université Paris 6, 1996

[Bou89] C. Bourrely, S. Muller, *Detection of microcalcifications in mammographic images*, NATO ASI, Series in System & Computer Sciences, Springer Verlag, Mar. 1989

[Bow98] M. Heath, K.W. Bowyer, D. Kopans et al, *Current status of the Digital Database for Screening Mammography*, in : Nico Karssemeijer, Martin Thijssen, Jan Hendriks, Leon van Erning (Eds.), Digital Mammography, Nijmegen 1998, Kluwer Academic Publishers, 1998, pp. 457-460

[Brz90] D. Brzakovic, X. M. Luo, P. Brazakovic, *An Approach to Automated Detection of Tumors in Mammograms*, IEEE Transactions on Medical Imaging, Vol. 9, No. 3, September 1990, pp. 233 - 241

[Cha95-1] Heang-Ping Chan, Shih-Chung B. Lo, Berkman Sahiner, Kwok Leung Lam, Mark A. Helvie, *Computer-aided detection of mammographic microcalcifications : Pattern recognition with an artificial neural network*, Medical Physics, Vol. 22, No.10, October 1995, pp. 1555 - 1567

[Cha95-2] Heang-Ping Chan, Datong Wei, Mark A. Helvie, Berkman Sahiner, Dorit D. Adler, Mitchel M. Goodsitt, Nicolas Petrick, *Computer-aided classification of mammographic masses and normal tissue : linear discriminant analysis in the texture feature space*, Phys. Med. Biol., Vol. 40, 1995, pp. 857 - 876

[Che95] L. Chérié-Challine, F. Ménégoz, K. Malek, P.Grosclaude, E. Jougla, M. Parkin, R. Black, *Incidence et mortalité des cancers en France : Situation en 1995 et évolution entre 1975 et 1995*, Institut de Veille Sanitaire,

[Cup96] Tommy E. Cupples, G.W. Eklund, Gilda Cardenos, *Mammographic Hallo Sign Revisited*, Radiology, Vol. 199, No. 1, April 1996, pp. 105 - 107

[Dha96] Atam P. Dhawan, Yateen Chitre, Christine Kaiser-Bonasso, Myron Moskowitz, *Analysis of Mammographic Microcalcifications Using Gray-Level Image Structure Features*, IEEE Transactions on Medical Imaging, Vol. 15, No. 3, June 1996, pp. 246 - 259

[Din97] J.M. Dinten, G. Montemont, M. Darboux, *Opacities detection and characterization in mammograms using a bilateral comparison and local characteristics*, SPIE, Vol. 3035, 1997, pp. 56 - 66

[Doi98] K. Doi, H. MacMahon, M.L. Giger, K.R. Hoffmann *Computer-Aided Diagnosis in Medical Imaging*, Proceedings of the First Intrenational Workshop on Computer-Aided Diagnosid, Chicago, USA, Sep. 1998, Elsevier, 1998.

[Ema95] Takehiro Ema, Kunio Doi, Robert M. Nichikawa, Yulei Jiang, John Papaioannou, *Image feature analysis and computer-aided diagnosis in mammography : Reduction of false-positive clustered microcalcifications using local-gradient analysis*, Medical Physics, Vol. 22, No.2, February 1995, pp. 161 - 169

[Gav97] Marios A. Gavrielides, Maria Kallergi, Laurence P. Clarke, *Automatic Shape Analysis and Classification of Mammographic Calcifications*, SPIE, Vol. 3034, 1997 , pp. 869 - 876

[Gra93] A. Grad, M. Rosselli Del Turco, *Breast Cancer Screening in Europe*, Springer Verlag 1993

[Gup95] R. Gupta, P. E. Undrill, *The use of texture analysis to delineate suspicious masses in mammography*, Phys. Med. Biol., Vol. 40, 1995, pp. 835 - 855

[Gur97] M. Nafi Gurcan, Yasemin Yardimci, A. Enis Cetin, Rashid Ansari, *Automated Detection and Enhancement of Microcalcifications in Mammograms Using Nonlinear Subband Decomposition*, Proceedings of IEEE ICASSP97, International Conferece on Acoustics, Speech, and Signal Processing, April 20-24, 1997, München, Germany.

[Hey97] Sylvia H.Heywang-Köbrunner, Ingrid Schreer, David Dershaw, *Diagnostic Breast Imaging*, Georg Thieme, Germany, 1997

[Hig96] Ralph Highnam, Michael Brady, Brasil Shepstone, *A representation for mammographic image processing*, Medical Image Analysis, Vol. 1, No.1, 1996, pp. 1 - 18

[Hig99] Ralph Highnam, Michael Brady, *Mammographic Image Analysis*, Kluwer Academic, 1999

[Huo95] Zhimin Huo, Maryelle L. Giger, Carl J. Vyborny, Ulrich Bick, Ping Lu, Dulcy E. Wolverton, Robert A. Schmidt, *Analysis of spiculation in the computerized classification of mammographic masses*, Medical Physics, Vol. 22, No.10, October 1995, pp. 1569 - 1579

[Kar96] Nico Karssmeijer, Guido M. te Brake, *Detection of Stellate Distortions in Mammograms*, IEEE Transactions on Medical Imaging, Vol. 15, No.5, October 1996, pp. 611 - 619

[Keg94] Philip Kegelmeyer, Joe M. Pruneda, Philip D. Bourland, Argye Hillis, Mark W. Riggs, Michael L. Nipper, *Computer-Aided Mammographic Screening for Spiculated Lesions*, Radiology, Vol. 191, December 1994, pp. 331 - 337

[Kil93] Judy Kilday, Francesco Palmieri, Martin D. Fox, *Classifying Mammographic Lesions Using Computerized Image Analysis*, IEEE Transactions on Medical Imaging, Vol. 12, No.4, December 1993, pp. 664 - 669

[Kob96] Hidefumi Kobatake, Yukiyasu Yoshinaga, *Detection of Spicules on Mammogram Based on Skeleton Analysis*, IEEE Transactions on Medical Imaging, Vol. 15, No.3, June 1996, pp. 235 - 245

[Kob97] Hidefumi Kobatake, Kenichi Okuno, Masayuki Murakami, Masamitsu Ishida, Hideya Takeo, Shigeru Nawano, *CAD system for full-digital mammography and its evaluation*, SPIE, Vol. 3034, 1997 , pp. 745 - 752

[Kop96] Daniel B. Kopans, Cynthia A. Swann, George White, Kathreen A. McCarthy, Deborah A. Hall, Stephen J. Belmonte, William Gallagher, *Asymmetric Breast Tissue*, Radiology, Vol. 171, No. 3, June 1989, pp. 639 - 643

[Kru97] Elizabeth A. Krupinski, Robert M. Nishikawa, *Comparison of eye position versus computer identified microcalcification clusters on mammograms*, Medical Physics, Vol. 24, No. 1, January 1997, pp. 17 - 23

[Lai89] Shuk-Mei Lai, Xiaobo Li, Walter F. Bischof, *On Techniques for Detecting Circumscribed Masses in Mammograms*, IEEE Transactions on Medical Imaging, Vol. 8, No.4, December 1989, pp. 377 - 386

[Laine94] Andrew F. Laine, Sergio Schuler, Jian Fan, Walter Huda, *Mammographic Feature Enhancement by Multiscale Analysis*, IEEE Transactions on Medical Imaging, Vol. 13, No.4, December 1994, pp. 725 - 740

[Lam81] J.L.Lamarque, *Le sein, radiodiagnostic clinique*, Medicine et Sciences Internationales, 1981

[Lan86] Marton Lanyi, *Diagnosis and Differential Diagnosis of Breast Calcifications*, Springer Verlag 1986

[Lev95] François Lefebvre, Habib Benali, René Gilles, Edmond Kahn, Robert Di Paola, *A fractal approach to the segmentation of microcalcifications in digital mammograms*, Medical Physics, Vol. 22, No.4, April 1995, pp.381 - 390

[Li97] Huai Li, Shih-Chung Lo, Yue Wang, Matthew T. Freedman, Seong K. Mun, *Mammographic Mass Detection by Stochastic Modeling and a Multi-Modular Neural Network*, SPIE, Vol. 3034, 1997 , pp. 480 - 490

[Lo98] Shin-Chung B. Lo, Huia Li, Akira Hasegawa, Yue J. Wang, Matthew T. Freedman, Seong K. Mun, *Detection of Mammographic Masses Using Sector Features With A Multiple Circular Path Neural Network*, SPIE, Vol. 3338, 1998, pp. 1205 - 1214

[Mau99] M. Poissonniere, R. Highnam, M. Brady, B. Shepstone, R. English, *Integration of Low-Level Processing to Faciliate Microcalcification Detection*, in : Nico Karssemeijer, Martin Thijssen, Jan Hendriks, Leon van Erning (Eds.), Digital Mammography, Nijmegen 1998, Kluwer Academic Publishers, 1998, pp. 185-188

[Mul99] S. Muller, *Full-Field Digital Mammography designed as a complete system.*, European Journal of Radiology, Invited Paper, Vol. 31 , pp 25-34, 1999.

[Nag98] Rufus H. Nagel, Robert M. Nishikawa, John Papaioanou, Kunio Doi, *Analysis of methods for reducing false positives in automated detection of clustered microcalcifications in mammograms*, Medical Physics, Vol. 25, No. 8, August 1998, pp. 1502 - 1506.

[Pet96-1] Nicolas Petrick, Heang-Ping Chan, Berkman Sahiner, Datong Wei, *An Adaptive Density-Weighted Contrast Enhancement Filter for Mammographic Breast Mass*

Detection, IEEE Transactions on Medical Imaging, Vol. 15, No.1, February 1996, pp. 59 - 67

[Pet96-2] Nicolas Petrick, Heang-Ping Chan, Berkman Sahiner, Datong Wei, Mark A. Helvie, Dorit D. Adler, *Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification*, Medical Physics, Vol. 23, No.10, October 1996, pp. 1685 - 1696

[Pin99] Heang-Ping Chan et.al. *Digital Mammography : Observer Performance Study of the Effect of Pixel Size on Radiologists Characterization of Malignant and Benign Microcalcifications*, SPIE Medical Imaging 99, International Conference on Acoustics, Speech, and Signal Processing, Feb. 20-26 1999, San-Diego, CA. USA.

[Poh96] Scott Pohlman, Kimerly A. Powell, Nancy A. Obuchowski, William A. Chilcote, Sharon Grundfest-Broniatowski, *Quantitative classification of breast tumors in digitizes mammogramms*, Medical Physics, Vol. 23, No 8, August 1996, pp. 1337-1345

[Qia95] Wei Qian, Maria Kallergi, Laurence P. Clarke, Huai-Dong Li, Priya Venugopal, Dansheng Song, Robert A. Clark, *Tree structured wavelet transform segmentation of microcalcifications in digital mammography*, Medical Physics, Vol. 22, No.8, August 1995, pp. 1247 - 1254

[Ric99-1] Andreas Rick , Serge Muller, Sylvie Bothorel, Michel Grimaud, *A Quantitative Model of Microcalcification Detection in Digital Mammography*, Medical Image Computing and Cumputer-Assisted Intervention 1999, (MICCAI99), Cambridge, England 19-22 Sep. 1999.

[Ric99-2] A. Rick, B. Bouchon Meunier, S. Muller, *Fonctions d'appartenance adaptatives pour le traitement d'images*, LFA'99, Octobre 1999, Valenciennes

[Ric99-3] A. Rick , S. Bothorel , B. Bouchon Meunier, S. Muller, M. Rifqi, *Fuzzy Techniques in Mammographic Image Processing*, in : *Fuzzy Techniques in Image Processing*, series : "Studies in Fuzziness and Soft Computing" Ed. Mike Nachtegael, Etienne Kerre, Physica-Verlag (Springer), (to be published).

[Sah96-1] Berkman Sahiner, Heang-Ping Chan, Nicolas Petrick, Datong Wei, Mark A. Helvie, Dorit D. Adler, Mitchel M. Goodsitt, *Classification of Mass and Normal Breast Tissue : A Convolutional Neural Network Classifier with Spatial Domain and Texture Images*, IEEE Transactions on Medical Imaging, Vol. 15, No.5, October 1996, pp. 598 - 610

[Sah96-2] Berkman Sahiner, Heang-Ping Chan, Nicolas Petrick, Datong Wei, Mark A. Helvie, Dorit D. Adler, Mitchel M. Goodsitt, *Image feature selection by a genetic algorithm : Application to classification of mass and normal breast tissue*, Medical Physics, Vol. 23, No.10, October 1996, pp. 1671 - 1683

[Suc95] J. Suckling, D. R. Dance, E. Moskovic, D. J. Lewis, S.G. Blacker, *Segmentation of mammograms using multiple linked self-organizing neural networks*, Medical Physics, Vol. 22, No.2, February 1995, pp. 145 - 152

[Tab85] L. Taber, P. B. Dean, *Teaching atlas of mammography*, Georg Thieme, Germany, 1985

[Tao98] Eric Tao, Chester Ornes, Jack Sklansky, *Automatic detection of microcalcifications in digital mammography*, SPIE, Vol. 3338, 1998, pp. 1450 - 1458

[Tay99] Paul Taylor, John Fox, Andrew Todd-Pokropek, *Evaluation of a Decision Aid for the Classification of Microcalcifications*, in : Nico Karssemeijer, Martin Thijssen, Jan Hendriks, Leon van Erning (Eds.) Digital Mammography, Nijmegen 1998, Kluwer Academic Publishers, 1998, pp. 237 - 244

[Tra95] A.Travade, A.Isnard, H. Gimbergues, *Imagerie de la pathologie mammaire*, Masson, 1995

[Tre89] A. Le Treut, M. H. Dilhuydy, *Mammographie, guide d'interprétation*, arnette, 1989

[Wei95] Datong Wei, Heang-Ping Chan, Mark A. Helvie, Berkman Sahiner, Nicolas Petrick, Dorit A. Adler, Mitchell M. Goodsitt, *Classification of mass and normal breast tissue on digitial mammograms : Multiresolutional texture analysis*, Medical Physics, Vol. 22, No.9, September 1995, pp. 1501 - 1513

[Win67] F. Winsberg, M. Elkin, J. Macy, V. Bordaz, W. Weymouth, *Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis*, Radiology, Vol. 89, 1967, pp. 211 - 215

[Wol76] J. N. Wolfe, *Breast patterns as an index of risk for developping breast cancer*, Am. J. Roentgenology, 1976 : 126 : 1130-1139

[Yam99] M. Yam, R. Highnam, M. Brady, R. English, *Detecting calcifications in mammogramms using the h_{int} representation*, Proceedings of CARS'99, Computer Assisted Radiology and Surgery, Paris 1999, pp. 373 - 377

[Yin91] Fang-Fang Yin, Maryellen L. Giger, Kunio Doi, Charles E. Metz, Carl J. Vyborny, Robert A. Schmidt, *Computerized detection of masses in digital mammograms : Analysis of bilateral substraction images*, Med. Phys. , Vol. 18, No.5, Sep/Oct 1991, pp. 955 - 963

[Yu98] Songyang Yu, Ling Guan, Stephen Brown, *Automatic Detection of Clustered Microcalcifications in Digital Mammography Based on Wavelet Features and Neural Network Classification*, SPIE, Vol. 3338, 1998, pp. 1540 - 1546

[Yu99] Songyang Yu, Ling Guan, Stephen Brown, *Automatic detection of clustered microcalcifications in digitized mammogram film*, Journal of Electronic Imaging, Vol. 8, No. 1, Jan. 1999, pp. 76-82

[Zwigg97] Reyer Zwiggelar, Tim C. Par, Caroline R. M. Boggis, Susan Astley, Christopher Taylor, *Statistical modelling of lines structures in mammogramms*, SPIE, Vol. 3034, 1997 , pp. 510 - 521

Traitement d'images

[Abm94] Wolfgang Abmayr, *Einführung in die digitale Bildverarbeitung*, Teubner, 1994

[Cho93] Zang-Hee Cho, Joie P. Jones, Manbir Singh, *Foundations of Medical Imaging*, Wiley Interscience, 1993

[Coc95] J.P.Cocquerez, S.Philipp, *Analyse d'images : filtrage et segmentation*, Masson 1995

[Dud73] Richard O. Duda, Peter E. Hart, *Pattern Classification and Scene Analysis*, Wiley 1973

[Fol90] James D. Foley, Andries van Dam, Steven K. Feiner, John F. Huges, *Computer Graphics, Principles and Practice*, Addison Wesley, 1990

[Jae91] Bernd Jähne, *Digitale Bildverarbeitung*, Springer Verlag, 1991

[Koe92] Jan J Koendrink, Andrea J. van Doorn, *Generic Neighbourhood Operators*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No.6, June 1992, pp. 597 - 605

[Rip96] B.D.Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996

Statistique

[Ber94] Kevin S. Berbaum, Donald D. Dorfman, E.A. Franken, *Mesuring Observer Performance by ROC Analysis, Indications and Complications*, Investigative Radiology, Vol.24, No.3 March 1989, pp.229- 245

[Bra92] Siegmund Brandt, *Datenanalyse*, BI Wissenschaftsverlag, 1992

[Zha95] Zhengyou Zhang, *Parameter Estimation Techniques : A Tutorial with Application to Conic Fitting*, INRIA rapport de recherche n 2676, Oct. 1995

Morphologie mathématique

[Bet97] D. Betal, N. Roberts, G.H. Whitehouse, *Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology*, British Journal of Radiology, Vol. 70, pp. 903 - 917

[Gri91] Michel Grimaud, *La géodesie numérique en morphologie mathématique. Application à la détection automatique de microcalcifications en mammographie numérique.*, Thèse de doctorat à l'Ecole Nationale Supérieure des Mines de Paris, 1991

[Har87] Robert M. Haralick, Stanley R. Sternberg, Xinhua Zhuang, *Image Analysis Using Mathematical Morphology*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-9, No.4, July 1987, pp. 532 - 549

[Mar87] Petros Maragos, Ronald Schafer, *Morphological Filters - Part I : Their Set-Theoretic Analysis and Relation to Linear Shift-Invariant Filters*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-35, No.8, August 1987, pp. 1153 - 1184

[Naj96] Laurent Najman, Michel Schmitt, *Geodesic Saliency of Watershed Contours and Hierarchical Segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No.12, December 1996, pp. 1163 - 1173

[Sal94] Philippe Salembier, Montse Pardàs, *Hierarchical Morphological Segmentation for Image Sequence Coding*, IEEE Transactions on Image Processing, Vol. 3, No.5, September 1994, pp. 639 - 651

[Vac95] Corinne Vachier, *Extraction de caractéristiques, segmentation d'image et morphologie mathématique*, Thèse de de doctorat à l'Ecole Nationale Supérieure des Mines de Paris, 1995

Textures

[Kim99] Jong Kook Kim, Hyun Wook Park, *Statistical Textural Features for Detection of Microcalcifications in Digitized Mammograms*, IEEE Transactions on Medical Imaging, Vol. 18, No.3, March 1999, pp. 231 - 238

[Mil92] Peter Miller, Sue Astley, *Classification of breast tissue by texture analysis*, Image vision and computing, Vol. 10, No. 5, June 1992

[Pet94] Arthur Petrosian, Heang-Ping Chan, Mark A. Helvie, Mitchell M. Goodsitt, Dorit A. Adler, *Computer-aided diagnosis in mammography : classification of mass and normal tissue by texture analysis*, Phys. Med. Biol. Vol.39, 1994, pp. 2273 - 2287

[Smi97] John R. Smith, *Integrated Spatial and Feature Image Systems : Retrieval, Analysis and Compression* , PhD Thesis, 1997 (http ://disney.ctr.columbia.edu/jrsthesis/)