



HAL
open science

Computer-aided detection for digital breast tomosynthesis

Gero Peters

► **To cite this version:**

Gero Peters. Computer-aided detection for digital breast tomosynthesis. domain_other. Télécom ParisTech, 2007. English. NNT: . pastel-00002938

HAL Id: pastel-00002938

<https://pastel.hal.science/pastel-00002938>

Submitted on 28 Sep 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

présentée par

Gero PETERS

pour l'obtention du
GRADE DE DOCTEUR

Specialité : Signal et Images

**Computer-aided Detection for Digital Breast
Tomosynthesis**

**Détection Automatique des Signes Radiologiques
pour la Mammographie Numérique Tridimensionnelle**

Soutenue le : 21 juin 2007

devant un jury composé de:

Mme Christine GRAFFIGNE

M. Abdelkader ZIGHED

Mme Isabelle BLOCH

M. Serge MULLER

Sir Michaël BRADY

Mme Isabelle MAGNIN

Rapporteur

Rapporteur

Directeur de thèse

Co-Directeur de thèse

Examineur

Examineur

To my parents
To Fatima

Acknowledgements

This work was funded by GE Healthcare (Buc, France) in collaboration with the Ecole Nationale Supérieure des Télécommunications (Paris, France), with the support of the ANRT (Association Nationale de la Recherche Technique).

I wish to thank Mrs. Isabelle Bloch for being my research advisor and for welcoming me to the TSI research group of the ENST (Paris, France). She helped me during my work with her passion for research, her great availability, her constant encouragement and her seemingly unlimited energy. Her profound knowledge in medical imaging and fuzzy set theory led to many fruitful discussions and helped me to constantly question my approaches. I would also like to thank my co-advisor, Mr. Serge Muller, for the trust he put in me by accepting me in his team at GE Healthcare. His many years of experience in mammography and his guidance were of great help in countless discussions. His scientific curiosity was an inspiration to my work. I am not only grateful for an outstanding working relationship, but also to have found a good friend.

I wish to express my sincere gratitude to Mrs. Christine Graffigne and Mr. Abdelkader Zighed who accepted the challenging and time-consuming task to assess my PhD work. I would also like to express my appreciation to Mrs. Isabelle Magnin and Sir Michaël Brady for accepting to become members of my jury.

I would like to thank Mr. Jean-Michel Milles and Mr. Christian Moreau who allowed me to do my PhD in the mammography department of GE Healthcare. I thank my colleagues of the AMA-IQ team, and more generally all my colleagues at GE Healthcare for their friendship and their support. Thanks to Razvan Iordache for asking the scientific questions that hurt and which make you evolve, for all the coffee breaks, and for being so much more than just a colleague. Thanks to Xavier Bouchevreau for answering a mind-boggling number of computer-related questions, and for reanimating my workstation at several instances. Thanks to Fanny Patoureaux for her friendliness, her unlimited food resources and her willingness to share her PhD experience with a junior PhD. Thanks to Sylvain Bernard for the discussion about CAD. Thanks to Benedicte Grosjean and Sylvie Puong for sharing the PhD experience and supporting each other in our respective PhD depressions. Thanks to Giovanni Palma for helping me out when I really needed him, and for taking all my wise-guy-advice with a smile. Thanks to Vincent Bismuth for his relentless efforts to communicate in German, and for our many discussions about life, the universe and everything. Thanks to Cyril and Jean whose office I invaded more than once looking for (and finding) answers. Many thanks also to Celine, Myriel, Guillaume, Veronique, Elisabeth and all the others I did not mention for the great time we spent together.

I would like to thank Mike Grabb and his team at GE global research center for accepting me in their midst for three weeks; an inspiration to a young PhD. Special thanks go to Frederick Wheeler and Amitha Perera for many discussions on CAD as well as to Jeffrey W. Eberhard for answering my numerous questions on the DBT prototype.

Furthermore, I need to thank all the members of the GE basketball team; what would I have done over the years without having a lunchbreak of basketball to look forward to. Many thanks to Julien, Joonwon, Marc, Richard, Fred, Judikael, and all the others.

Thanks to the PhD students of the TSI department, those that accompanied me to eat

outside, and those who never grew tired of trying to convince me to give the cafeteria just one more shot. Antonio for sharing the suffering of countless weekends away from our partners, Julien, François and Jeff for many hours of gaming, Jamal and Jeremy for intense basketball play at Glacière, Sylvie, Nathalie and Vincent for receiving me in their office, listening to my stories, and exchanging milestones of motion-picture history, and all the others for many seminars, discussions and celebrations we shared over the years.

Thanks to my friends and family who listened to my stories without ever understanding what fuzzy logic is or why anyone would care to spend so much time with it. To Julian for most excellent phone calls. To Momo, Chris B. and Daniel from the Kicker group for inspiring discussions. To the Turbine 97 group for the valuable advice and for broadening my cultural horizon. To Jörg, Simone, Céline, Carlo, Chris L., Flo and to all the others for their support and their friendship.

And finally very special thanks to the people that matter most, and without whom I could have never done it. To my brother and sisters, for always being there for me. To my parents without whom I would not be. Who have always encouraged me, and who never doubted that I will succeed. And of course, to Fatima, whose love and support over the last years I cannot describe in words. For all the weekends and evenings she spent without me, for all the work she took off my shoulders, for the encouragements and for always believing in me. I could not have done it without you.

Contents

Acknowledgements	v
Table of Contents	vii
List of Figures	x
List of Tables	xvi
Abstract	xix
Résumé	xxi
Synthèse	xxiii
1 Introduction	1
2 Context and Motivation	5
2.1 The Breast	5
2.1.1 Anatomy of the Breast	5
2.1.2 Breast Cancer	7
2.2 Breast Imaging	8
2.2.1 Mammography	8
2.2.2 Digital Mammography	11
2.2.3 Digital Breast Tomosynthesis	12
2.2.4 Other Breast Imaging techniques	18
2.3 Computer-Aided Detection	18
2.3.1 CAD for 2D Mammography	19
2.3.2 CAD for DBT	24
3 Extracting Image Information	29
3.1 Introduction	29
3.2 Markers	30
3.2.1 Candidate Calcification Detection	30
3.2.2 Candidate Mass Detection	32
3.3 Segmentation	36

3.3.1	Active Contour Models	38
3.3.2	Experiments and Results	44
3.3.3	Fuzzy Contour Extraction	51
3.3.4	A New Class of Fuzzy Contours: Fuzzy Active Contours	55
3.4	Feature Extraction	63
3.4.1	Extraction of 2D Attributes	63
3.4.2	Computing 2D Fuzzy Attributes	64
3.5	Conclusion	68
4	Information Fusion	69
4.1	Image Reconstruction	69
4.1.1	Back-projection Methods	70
4.1.2	Transform Algorithms	71
4.1.3	Algebraic Reconstruction Techniques	72
4.1.4	Statistical Reconstruction Techniques	73
4.2	Aggregation of Fuzzy Data - A New Pixel-Based Approach	74
4.2.1	Partial Defuzzification	75
4.2.2	Aggregating 2D Fuzzy Particle Maps	78
4.3	Aggregation of Fuzzy Data - A New Particle-Based Approach	84
4.3.1	Linking 2D Fuzzy Particles	84
4.3.2	A Novel Approach for Aggregating Fuzzy Attributes	87
4.4	Conclusion	91
5	Decision Making	93
5.1	Introduction	93
5.2	Decision Trees	95
5.2.1	Introduction	95
5.2.2	Tree Construction	95
5.2.3	Classification of Test Data	98
5.3	Fuzzy Decision Trees	99
5.3.1	Crisp Input Data	99
5.3.2	Fuzzy Input Data	105
5.4	An Improved Fuzzy Decision Tree	111
5.4.1	Synthetic Data	111
5.4.2	Node Tests	112
5.5	Classifying Mass Attributes from DBT Projection Data	121
5.6	Performance Evaluation and Comparison	126
5.7	Conclusion	126
6	CAD for Clinical DBT Data	129

6.1	The Database	129
6.2	The Algorithm	132
6.3	Algorithm Performance	134
6.3.1	Feature Selection	135
6.3.2	Results Using the Leave-One-Out Method	136
6.3.3	Obtaining a Final Decision	137
6.3.4	The Contribution of Fuzzy Logic	139
6.3.5	Conclusion	141
7	Conclusion	143
7.1	Main Contributions	143
7.2	Suggestions for Further Work	145
A	Criteria to Compare Volumes and Surfaces	151
A.1	Volume comparison	151
A.2	Surface comparison	152
A.3	Illustrative examples	153
Appendices		
List of Patents and Publications		157
Bibliography		158

List of Figures

1	Principe de la mammographie numérique tridimensionnelle	xxiii
2	Défis présentés par les propriétés des images de projection	xxiv
3	Pathologies du sein	xxv
4	Stratégies de CAD pour DBT	xxv
5	Test d'hypothèse	xxvii
6	Ensemble de contours candidats	xxviii
7	Calculer le sous-ensemble flou <i>fort gradient</i>	xxix
8	Extraction des ensembles de niveau candidat	xxx
9	Conversion de la description de contour vers une description en chaque pixel	xxxii
10	Carte de particules floues 2D por détection d'opacités	xxxii
11	Stratégie d'agrégation d'information	xxxiii
12	Schéma de l'agrégation au niveau des particules	xxxiv
13	Arbre de décision flou complet	xxxvii
14	Un échantillon traité par l'algorithme complet proposé	xxxviii
1.1	CAD Strategies for 3D Data	3
2.1	Simplified Anatomy of the Breast	6
2.2	Schematic of a Branching Duct	6
2.3	X-Ray Attenuation	9
2.4	Principle of a Mammography Acquisition	9
2.5	Overlaying Structures in Mammography	9
2.6	Mammographic Views	10
2.7	Different Kinds of Breast Densities	11
2.8	Morphology of Microcalcifications	11
2.9	Different Types of Breast Lesions	12
2.10	Principle of Digital Breast Tomosynthesis	13
2.11	Image Dynamic of 2D Mammography	13
2.12	Image Dynamic of DBT Projections	14
2.13	Eliminate Overlay and Masking	14
2.14	Limited Angular Range of DBT	15
2.15	DBT Projection Images	15

2.16	Point Spread Functions for Different Geometries	16
2.17	DBT Reconstructed Slices	16
2.18	Maximum Z-Resolution	17
2.19	Tomographic Projection Images	17
2.20	Basic Modules of a CAD Scheme	19
2.21	CAD Strategies for DBT	25
2.22	Algorithm Scheme of the Proposed Approach	28
3.1	Mexican Hat Wavelet	31
3.2	Dense Kernel Detector	33
3.3	Clinical Example of the Dense Kernel Detector	34
3.4	Sliding Window Process	35
3.5	Gabor Wavelet	36
3.6	Challenges of DBT Projection Data	37
3.7	Limitations of Multi-Thresholding	37
3.8	Problems in Obtaining Closed Contours	38
3.9	The Function ϕ	39
3.10	Advantages of the Level-Sets	40
3.11	Drawbacks of Different Active Contour Models	41
3.12	Examples of Fast-Marching Level-Sets	41
3.13	Examples for the Chan-Vese Model	42
3.14	Comparison of Active Contour Models	43
3.15	Example of Applied Preprocessing	44
3.16	Manual Segmentation by an Expert	45
3.17	Different Types of Data	45
3.18	Impact of Parameters	47
3.19	Results on a Clinical ROI	49
3.20	Results on a Clinical ROI	49
3.21	Results on a Clinical ROI	50
3.22	Results on a Simulated ROI	50
3.23	Imprecision and Uncertainty	52
3.24	Set of Candidate Contours	53
3.25	Fuzzy Contour Extraction	53
3.26	Computing the Fuzzy Subset <i>High Gradient</i>	54
3.27	Computing the Fuzzy Subset <i>Contour</i>	55
3.28	Extracting the Candidate Level-Sets	56
3.29	Membership Function to the Class <i>Contour</i>	57
3.30	Example of Fuzzy Active Contour Hypothesis	58
3.31	Example of Fuzzy Active Contour Hypothesis	58

3.32 Hypothesis Testing	60
3.33 Fuzzy Active Contour Result for a Clinical ROI	61
3.34 Plot of the Level-Set Function ϕ	61
3.35 Fuzzy Active Contour Result for a Clinical ROI	62
3.36 Fuzzy Active Contour Result for a Clinical ROI	62
3.37 Computation of Attributes	63
3.38 Example of Feature Extraction	65
3.39 Applying the Extension Principle to Derive the Fuzzy Attribute	66
3.40 Example of Fuzzy Attribute Extraction	67
4.1 The Principle of Shift-and-Add Reconstruction	70
4.2 Limitations of the DBT geometry	71
4.3 Order Statistics Based Back-Projection	72
4.4 The Principle of Iterative Reconstruction Techniques	73
4.5 Fusion of Fuzzy Image Data	74
4.6 Dimensionality of Fuzzy Images	74
4.7 Conversion of Contour- to Pixel-Description	75
4.8 Problem of Overlapping	77
4.9 2D Fuzzy Particle Map for Microcalcification Detection	77
4.10 2D Fuzzy Particle Map for Mass Detection	78
4.11 2D Fuzzy Particle Maps for Microcalcification Detection	78
4.12 2D Fuzzy Particle Maps for Mass Detection	79
4.13 Information Aggregation Strategy	80
4.14 3D Fuzzy Particle Maps for Microcalcification Detection	81
4.15 3D Fuzzy Particle Maps for Mass Detection	81
4.16 Fuzzy Particle Maps for Microcalcification Detection on a clinical ROI	82
4.17 Fuzzy Particle Maps for Microcalcification Detection on a clinical ROI	82
4.18 Fuzzy Particle Maps for Mass Detection on a clinical ROI	83
4.19 Fuzzy Particle Maps for Mass Detection on a clinical ROI	83
4.20 Comparison of Different Aggregation Operators	84
4.21 Strategy for Aggregation of Fuzzy Data	85
4.22 Different Overlap Scenarios	86
4.23 Resemblance Between two Fuzzy Subsets	87
4.24 Algorithm Scheme for Aggregation on Particle Level	87
4.25 Aggregation of Fuzzy Attributes	89
4.26 Aggregation of Fuzzy Attributes	90
5.1 Basic Terminology of Decision Trees	95
5.2 Example of Attributes in a Decision Tree	96
5.3 Symbolic and Numerical Values at a Node Test	96

5.4	Automatic Node Test Computation	97
5.5	Example of a Decision Tree	98
5.6	A Crisp Decision Tree Processing a Test Sample	98
5.7	Comparison of a Crisp and a Fuzzy Decision Tree	100
5.8	From Histograms to Node Tests	101
5.9	Processing Applied to Fuzzy Histograms	101
5.10	Choosing the Most Discriminating Attribute	102
5.11	Construction of a Fuzzy Decision Tree	102
5.12	Computation of the Cumulated Degrees of Membership	103
5.13	Computing the Class Densities in the Leaves	104
5.14	Obtaining a Final Decision	105
5.15	Fully Constructed Fuzzy Decision Tree	106
5.16	A Fuzzy Decision Tree Processing a Crisp Sample	106
5.17	Satisfiability of a Fuzzy Subset to a Class	107
5.18	Computation of Fuzzy Histograms	108
5.19	Computation of a Weighted Fuzzy Histogram	109
5.20	Contrast Between Two Classes	109
5.21	A Fuzzy Decision Tree Processing a Fuzzy Sample	110
5.22	Synthetic Fuzzy Training Data Sets	112
5.23	The Impact of Noise on the Training Data	113
5.24	Non-Monotone Fuzzy Attributes	114
5.25	Problems of Multiple Attribute Use	115
5.26	Processing of Dependent Attributes	116
5.27	Fully Constructed Fuzzy Decision Tree	118
5.28	Example of the Classification of Synthetic Fuzzy Data	119
5.29	Classification Results on Synthetic Data	120
5.30	Example of Fuzzy Tree Construction	123
5.31	Example of Fuzzy Tree Construction	124
5.32	Example of Fuzzy Tree Construction	125
6.1	Spiculated Masses from the Database	130
6.2	Circumscribed Masses from the Database	130
6.3	One Example processed by the Entire Algorithm Chain	131
6.4	Flow-Chart of the Proposed CAD Algorithm	133
6.5	Examples of Fuzzy Decision Trees on Clinical Data	137
6.6	Flow-Chart of the Reference CAD Algorithm	140
6.7	Examples of Binary Decision Trees on Clinical Data	140
7.1	Fuzzy Microcalcification Cluster Detection	147
7.2	The Importance of Shape Features for Microcalcification Clusters	147

7.3 Orientation of the Cluster Object 148

A.1 Illustrative Examples of Volume and Surface Comparison 153

List of Tables

3.1	Active Contour Segmentation Results on Clinical Data	48
3.2	Active Contour Segmentation Results on Simulated Data	48
6.1	Discriminating Power of the Cumulated Fuzzy Attributes	135
6.2	Classification Results for a Single Node using Leave-One-Out	136
6.3	Classification Results for a Full Fuzzy Decision Tree	138
6.4	Classification Results for a Full Fuzzy Decision Tree	139
6.5	Classification Results for a Full Binary Decision Tree	141
A.1	Details of the Examples in Figure A.1	153
A.2	Results of the Comparison of the Examples in Figure A.1	155

Abstract

Digital Breast Tomosynthesis (DBT) is a new three-dimensional limited-angle tomography breast imaging technique that will substantially overcome the superimposition problem for lesion detection in digital mammography. This work focuses on developing an automated detection scheme for the very recent images obtained with DBT.

Since DBT is an emerging modality, clinical data sets are still rare, and few works have been published in the domain of computer-aided detection for these data sets. The three dimensional nature of the data offers a number of advantages. At the same time, issues like a reduced patient dose per scan, the limited-angle acquisition geometry and the resulting reconstruction artifacts, as well as the problem of an adapted reconstruction technique need to be addressed.

We develop an algorithm for reconstruction-independent computer-aided detection on DBT data. Working directly on the tomographic projection images, we make use of fuzzy set theory to preserve the ambiguity present in the images until the information from the entire set of projected views can be jointly considered. A framework is proposed that allows for detection of different kinds of radiological findings using the same general framework.

We present a method for fuzzy contour extraction for mass lesions. The originality of our method lies in the use of a set of dedicated hybrid active contour models. Each model is constructed using a priori knowledge about a class of objects of interest. For each radiological finding, we apply the different active contour models resulting in a set of fuzzy contours for each object of interest. A feature vector is extracted for each fuzzy contour. Using the extension principle we establish a set of fuzzy attributes based on this feature vector.

We propose an aggregation strategy for combining the information corresponding to a given three-dimensional object over the entire set of projected views. To this end we introduce a partial defuzzification operation that provides a pixel-based representation of the fuzzy information extracted for each object. This operation, combined with a back-projection/re-projection step, provides the links between the particles in the projection images, that correspond to the same three-dimensional particle. Cumulated fuzzy attributes and a confidence degree for each particle are computed. A fuzzy decision tree working on fuzzy data is then applied to obtain a final decision.

The presented approach has been validated on a number of simulated images as well as on clinical images. We show that the algorithm is capable of distinguishing between different radiological signs. The availability of a larger clinical database will allow a better quantitative evaluation of the algorithm performance in a near future.

Keywords: Computer-aided Detection, Digital Breast Tomosynthesis, Digital Mammography, Fuzzy Set Theory, Active Contour Models, Reconstruction, Fuzzy Decision Tree

Résumé

La mammographie numérique tridimensionnelle est une nouvelle modalité d'angulation limitée d'imagerie du sein qui est censée pouvoir surmonter le problème de superposition pour la détection des lésions en mammographie numérique. Dans cette thèse, notre but est de développer une méthode de détection automatique de lésions à partir de données récentes de mammographie tridimensionnelle.

Il n'y a aujourd'hui que très peu de données cliniques disponibles et peu de travaux scientifiques ont été publiés dans le domaine. Les données tridimensionnelles comportent de nombreux avantages. En même temps, des problèmes dus à la réduction de dose par image, la géométrie d'angulation limitée, et les artefacts de reconstruction qui en résultent doivent être gérés.

Nous développons un algorithme de détection automatique des signes radiologiques pour la mammographie numérique tridimensionnelle qui est indépendant de la reconstruction utilisée pour la visualisation du volume. En traitant directement les images de projection, nous appliquons la théorie des sous-ensembles flous afin de modéliser l'ambiguïté présente dans les images, ce qui nous permet en même temps de retarder la prise de décision jusqu'à une étape du traitement où nous pouvons prendre en compte la totalité des informations extraites à travers l'ensemble des projections. Nous proposons un système permettant la détection de différents types de lésions en utilisant une même méthodologie de haut niveau.

Nous présentons une méthode d'extraction de contours flous pour des opacités. L'originalité de l'approche présentée réside dans l'application d'un modèle dédié de contours actifs hybrides. Chaque modèle est construit à partir de connaissances a priori que l'on a sur une classe d'objets donnée. Nous appliquons les différents modèles pour chaque signe radiologique, fournissant un ensemble de contours flous pour chaque objet traité. Un vecteur d'attributs est extrait que l'on utilise pour obtenir, par le principe d'extension, un vecteur d'attributs flous.

Nous proposons un opérateur d'agrégation afin de combiner l'information correspondant à un objet tridimensionnel à travers l'intégralité des images de projection. A cette fin, nous introduisons un opérateur de défuzzification partielle, qui nous permet d'obtenir une représentation de l'information floue extraite pour chaque objet en chaque pixel. En combinant cet opérateur avec une étape de rétroprojection/reprojection, nous obtenons le lien entre les particules dans les différentes images de projection qui correspondent à une même particule tridimensionnelle. Nous montrons comment on peut calculer des attributs cumulés et un degré de confiance associé à chaque particule. Afin d'obtenir une décision à partir des informations extraites, nous utilisons un arbre de décision flou, capable de prendre des valeurs floues en entrée.

Nous avons validé notre outil de détection automatique en traitant de nombreuses images simulées, aussi que des images cliniques. Nous montrons que la méthode est capable de distinguer différents signes radiologiques. Une meilleure évaluation quantitative de la performance de l'algorithme dépendra de la disponibilité d'une base de données cliniques plus importante dans un avenir proche.

Mots clés: détection automatique de signes radiologiques, mammographie numérique tridimensionnelle, théorie des sous-ensembles flous, modèles de contours actifs, mammographie numé-

rique, reconstruction, arbres de décision flous

Synthèse

Le cancer du sein continue à être l'une des causes principales de mortalité de cancer parmi les femmes (Imaginis 07). Comme les causes de cette maladie restent inconnues, un dépistage précoce est le seul moyen pour réduire la mortalité parmi la population féminine. La mammographie joue un rôle clé pour la détection des cancers à un stage de développement préliminaire. C'est d'autant plus important que la probabilité de développement de métastases augmente avec la taille de la tumeur (Duffy 03).

La mammographie numérique tridimensionnelle (DBT) est une nouvelle modalité d'imagerie du sein en angle limité qui a le potentiel de surmonter le problème de superposition pour la détection des lésions (Dobbins III 03; Wu 03b; Suryanarayanan 00; Niklason 97). Plusieurs images de projection sont acquises pour différentes positions de la source de rayonnement (Niklason 97; Wu 03b; Dobbins III 03; Ren 05) comme le montre la figure 1. Une reconstruction des données tomographiques conduit à une représentation tridimensionnelle des données.

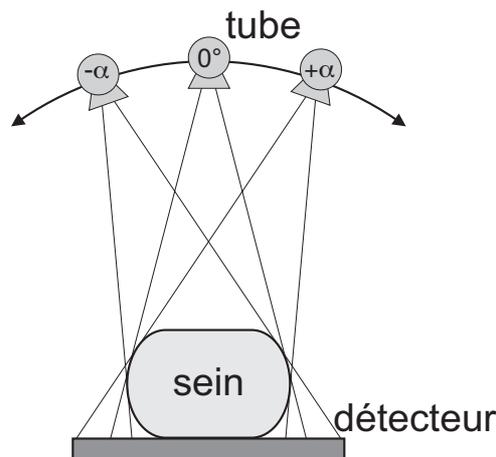


Figure 1: Principe d'une acquisition en mammographie numérique tridimensionnelle : plusieurs images de projection sont acquises pour différentes positions de la source de rayonnement.

En général 11 à 45 images sont acquises de manière équirépartie sur une ouverture angulaire de 15 à 45 degrés. La dose totale délivrée à la patiente est comparable à celle des deux vues (CC, MLO) de la mammographie standard, mais elle est répartie sur l'ensemble des images de projection. La dose par image acquise est ainsi sensiblement réduite par rapport à la mammographie 2D, pour maintenir une dose totale comparable. Cela représente un impact majeur pour le traitement des images de projection, car la caractéristique de ces images change fortement, et des algorithmes développés pour la mammographie 2D ne peuvent pas être appliqués sans modification (voir figure 2).

L'avantage principal de cette nouvelle modalité d'imagerie du sein est sa capacité à éliminer le recouvrement des tissus ainsi que l'effet masquant. L'ouverture angulaire du chemin décrit par la source de rayons-X est limitée. Pour des positions de tube loin de la position centrale, la pelote de compression peut entrer dans le champ visuel du détecteur et causer des artefacts dans

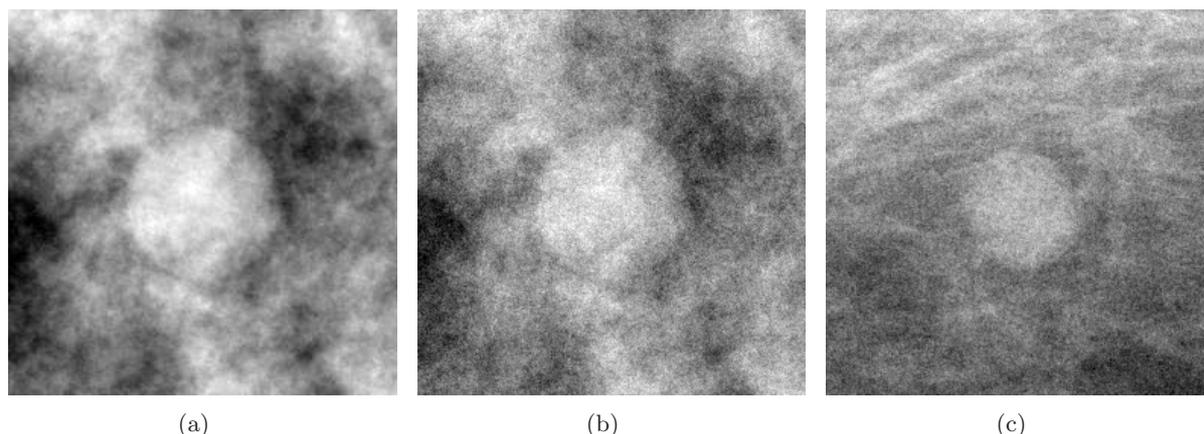


Figure 2: Défis présentés par les propriétés des images de projection : (a) une opacité simulée à dose standard, (b) les mêmes opacité, composition de sein et technique d’acquisition que en (a) mais image simulée à dose d’une projection tomographique, (c) ROI d’une image de projection clinique. La différence en niveau de bruit et de contraste entre l’objet et le fond est bien visible.

les images. En outre, il n’est pas possible de projeter le volume entier du sein sur un détecteur statique pour des positions de source arbitraires. Ces contraintes ont conduit à adapter une gamme angulaire limitée pour des acquisitions de DBT à $\pm 15 - 45^\circ$, comme mentionné ci-dessus.

L’augmentation du volume de données, en passant de 4 vues en mammographie 2D à plus de 100 coupes reconstruites en DBT, demande de nouveaux outils afin de guider les radiologues dans leur processus de décision.

Stratégies de CAD Dans environ deux-tiers des cas où des lésions ont été manquées pendant le dépistage, le radiologue n’a pas détecté un cancer qui était visible rétrospectivement (Martin 79; Buchanan 83). Un tel manque pourrait être dû à la nature des résultats visuels, aux images de basse qualité, à la fatigue, ou à l’inadvertance des radiologues. Une double lecture (par deux radiologues) sert à augmenter la sensibilité (Bird 90; Murphy Jr 90; Thurfjell 94). Ainsi, un but du CAD est d’augmenter l’efficacité du dépistage du cancer du sein en employant un système informatique en tant que deuxième lecteur pour aider le radiologue en indiquant les endroits suspects dans les images (Vyborny 94). Examiner des mammogrammes est une tâche répétitive qui implique dans la plupart du temps des images normales, ce qui en fait une tâche bien adaptée à l’analyse par ordinateur. Des exemples de lésions du sein sont donnés dans la figure 3.

Dans le CAD 2D nous pouvons généralement identifier trois modules classiques de traitement. D’abord une *détection* de candidats est effectuée. Une fois que des candidats ont été identifiés, une information supplémentaire doit être recueillie. Cela est fait pendant l’étape d’*analyse* qui peut inclure la segmentation aussi que l’extraction d’attributs. Enfin cette information est employée pendant l’étape de *décision* pour classifier les candidats. Pendant cette étape, la sélection d’attributs et la classification peuvent être effectuées.

Quand on compare le CAD pour DBT au CAD conçu pour la mammographie standard, la particularité est la nature tridimensionnelle des données. Nous avons donc choisi de classifier les différentes stratégies de CAD selon leur manière de prendre en compte l’aspect 3D des données.

Etant donné les trois modules de base décrits ci-dessus, et leur ordre qui ne devrait pas être changé, pour établir un arrangement complet du CAD 3D, un composant supplémentaire doit être inclus pour effectuer la transformation entre la représentation 2D et la représentation 3D des données. Cela est réalisé par une rétroprojection des données tomographiques dans l’espace

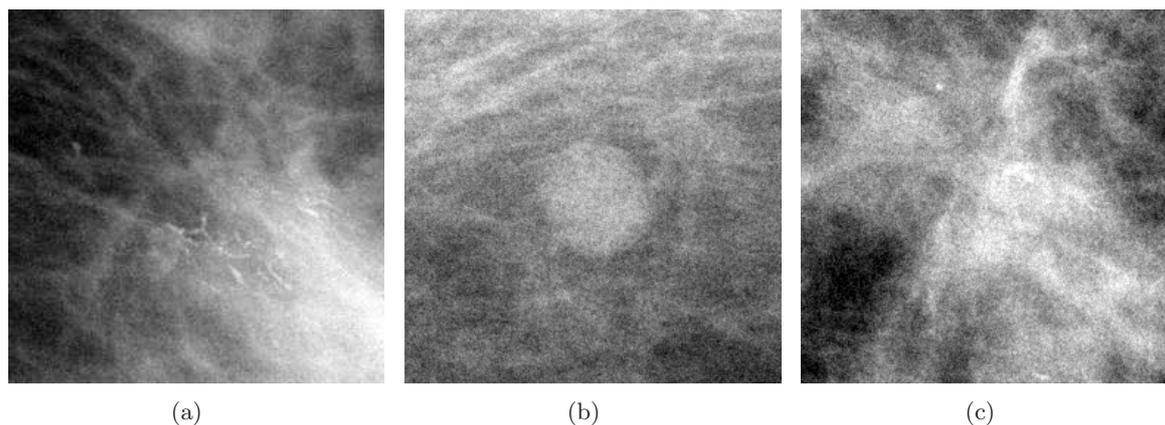


Figure 3: Différents types de lésion présents dans le sein : (a) amas de microcalcifications, (b) opacités circonscrites, (c) opacités stellaires.

3D. Avec ces quatre modules, trois stratégies pour le CAD sur des données de DBT ont été identifiées. Celles-ci sont illustrées dans la figure 4.

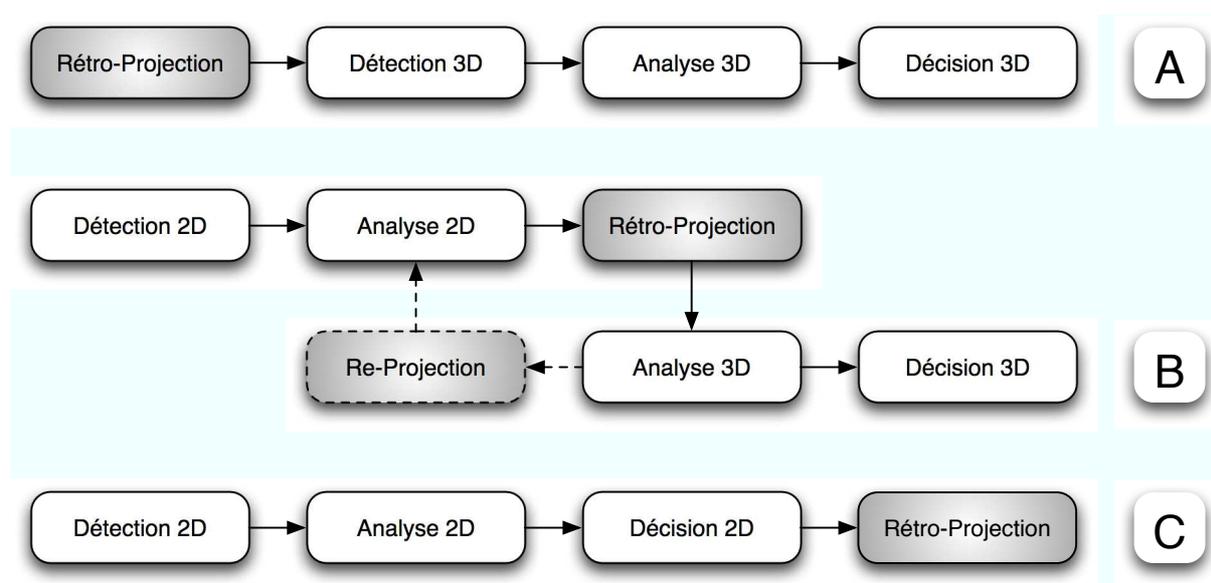


Figure 4: Stratégies de CAD : (A) traitement effectué directement sur des données 3D, (B) effectuer la détection et une partie de l'analyse en 2D, rétro-projeter les résultats intermédiaires et effectuer des analyses supplémentaires ainsi que la décision en 3D (possibilité d'extension vers une approche itérative en rajoutant une étape de re-projection), (C) effectuer le traitement complet en 2D et rétro-projeter uniquement les résultats obtenus.

La première stratégie décrite dans la figure 4 (A) travaille directement sur la représentation 3D des données. L'avantage de cette méthode est de tirer profit du contraste amélioré dans les coupes reconstruites par rapport aux projections tomographiques. Un inconvénient majeur est dû à l'anisotropie du volume reconstruit provoqué par l'ouverture angulaire limitée. En outre, les algorithmes de reconstruction pour cette nouvelle modalité ne sont pas encore entièrement optimisés.

Dans la figure 4 (B) une partie du traitement est effectuée sur les images de projection. Puis le résultat de ce traitement est rétroprojeté dans l'espace 3D. Cette méthode combine plusieurs avantages. Travailler directement sur les images de projection présente un effort informatique

réduit comparé à un traitement des coupes reconstruites qui sont généralement beaucoup plus nombreuses. En outre, travaillant en 2D, l'approche est indépendante d'une technique spécifique de reconstruction. En même temps, la combinaison des résultats de différentes projections tomographiques permet de tenir compte de la nature 3D des données.

Dans la figure 4 (C) le traitement complet est effectué dans l'espace 2D, et seulement les résultats sont rétroprojetés en 3D. Une limitation forte de cette approche est cependant qu'elle considère chaque image de projection séparément, et ignore ainsi complètement la nature 3D des données.

Nous avons donc choisi de développer un algorithme selon la stratégie (B).

Principe de l'approche développée L'un des inconvénients des techniques classiques de segmentation est qu'elles exigent une prise de décision dans chaque image de projection. Nous voulons retarder cette décision au moyen de l'extraction de contours flous afin de recueillir de l'information supplémentaire avant qu'une décision soit prise. Cependant, ce choix des paramètres guidant les algorithmes de segmentation constitue une décision elle aussi. Cela vaut particulièrement pour une approche de contour actif comme celle qui sera discutée plus tard dans cette section.

Nous avons développé une nouvelle approche pour obtenir des contours actifs flous pour différents types d'objet d'intérêt. Dans le cas de la détection assistée par ordinateur, nous ne savons pas quel genre d'objet nous recherchons. Cependant, nous connaissons généralement l'ensemble des objets attendus possibles. Nous procédons alors en construisant un modèle de contour actif pour chacun de ces objets. Chacun des modèles sera appliqué à tous les objets détectés.

La suite de la procédure vise à valider l'hypothèse prise pendant la segmentation. Une illustration de ce processus pour deux exemples de classes (opacité circonscrite et opacité stellaire) est donnée dans la figure 5.

Dans les paragraphes suivants nous discutons les différentes étapes de cette approche.

Contours flous La théorie des sous-ensembles flous (Zadeh 65) peut être employée comme outil pour gérer l'imperfection dans les images. Nous pouvons d'abord traiter l'ambiguïté au niveau de la segmentation, puis pendant la classification des particules segmentées en utilisant des attributs mesurés à partir des contours. L'ambiguïté des contours se propage aux attributs mesurés à partir des particules segmentées et donc à la classification qui s'appuie sur ces attributs. Notre approche présente l'avantage de tenir compte de l'imperfection dans l'image des contours de particules, présentant des contours flous, et de transférer cette imperfection jusqu'à l'étape de classification par des attributs flous.

Notre approche de segmentation floue se compose de trois étapes :

- fournir un marqueur pour chaque particule à segmenter ;
- déterminer un ensemble de contours candidats pour chaque particule ;
- construction de contours flous pour chaque particule par l'évaluation des contours candidats en s'appuyant sur leurs degrés d'appartenance à la classe *contour*.

La manipulation de l'imprécision et de l'incertitude des particules au niveau du contour, et non pas au niveau du pixel, permet la réduction de la dimensionalité de la segmentation floue, puisque nous passons d'une description bidimensionnelle (région) à une description unidimensionnelle (contour).

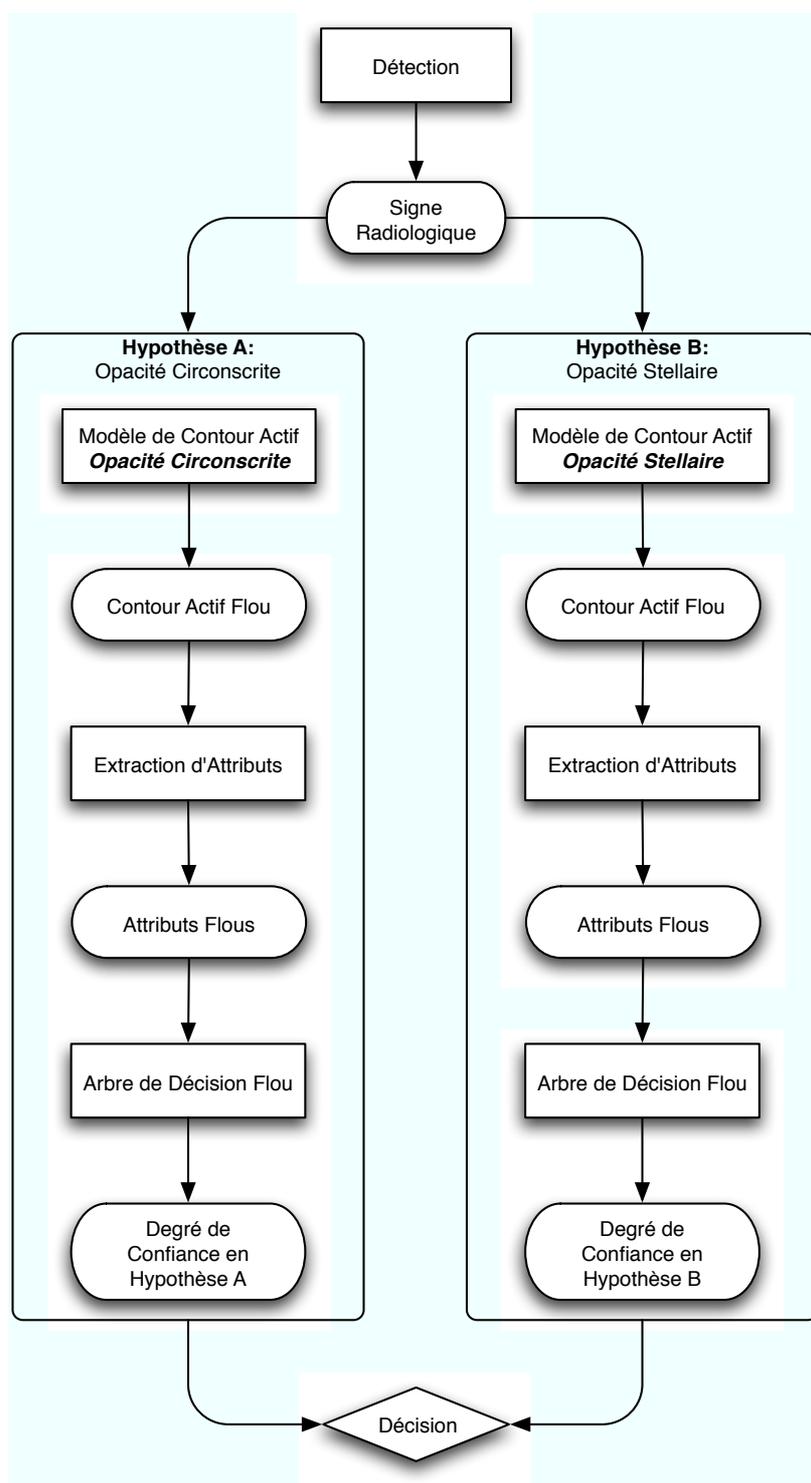


Figure 5: Test d'hypothèse pour un signe radiologique. Deux modèles de contour actif sont appliqués. Chaque modèle prend en compte des connaissances a priori concernant un objet d'intérêt donné (opacité circonscrite/stellaire). Dans la suite, chaque contour flou est traité indépendamment de l'autre. Extraction d'attributs, agrégation et classification fournissent une base pour pouvoir valider l'hypothèse initiale.

L'ensemble des maxima régionaux de l'image d'intensité I est noté $Max(I)$. Nous considérons comme marqueurs les maxima régionaux $M_I \in Max(I)$ de l'image d'intensité dans le support des particules candidates détectées par un filtre d'ondelettes.

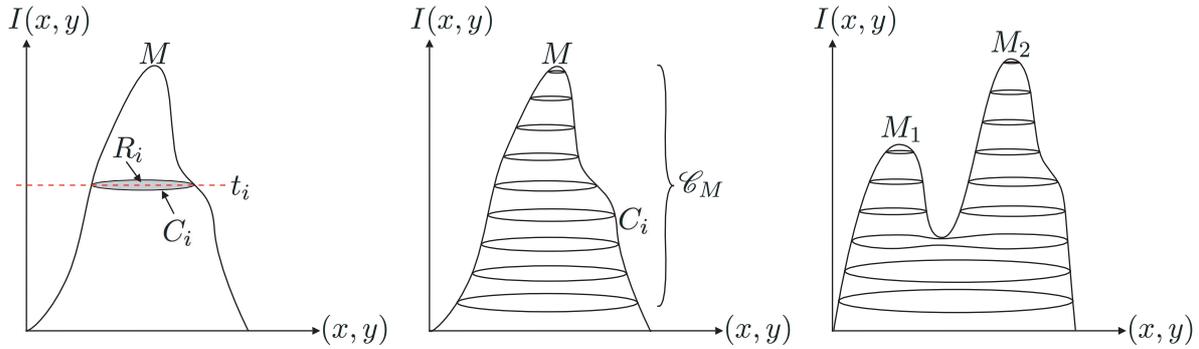
Pour chaque particule marquée par un marqueur M_I on obtient un ensemble de contours candidats par croissance de région. Le seuil de I au niveau t_i est défini comme :

$$I_{t_i} = \{(x, y) \mid I(x, y) \geq t_i\} \quad (1)$$

On détermine un ensemble de contours \mathcal{C}_{M_I} en appliquant un opérateur de multi-seuillage :

$$\mathcal{C}_{M_I} = \left\{ C_i = \partial R_{M_I}(I_{t_i}), t_1 = \max_{M_I}(I) > t_2 > \dots > t_{n_I} \right\} \quad (2)$$

où $\partial R_{M_I}(I_{t_i})$ est la frontière de la région R_{M_I} pour l'ensemble $R_{M_I}(I_{t_i})_{i=1, \dots, n_I}$ qui inclut le marqueur M_I de l'image I , et $t_1 = \max_{M_I}(I)$ est le niveau de gris maximal de l'image I dans le marqueur M_I . Ce processus est illustré dans la figure 6(a).



(a) Profil en niveaux de gris d'une particule candidate et seuil appliqué au niveau t_i .

(b) Contours extraits pour différents niveaux t_i .

(c) Contours appartenant à plusieurs particules marquées par différents marqueurs.

Figure 6: Ensemble de contours candidats

Extraction des contours flous Chaque maximum de l'image est traité indépendamment. Par conséquent, certains contours C_i peuvent appartenir à plusieurs ensembles de contours pour des particules identifiées par différents marqueurs (figure 6(c)). Cette propriété permet de gérer la superposition dans des images radiologiques.

La fonction d'appartenance à la classe *contour* $\mu_{\mathcal{C}|\mathcal{A}_1}$ est définie à partir d'un sous-ensemble flou \mathcal{A}_1 qui représente des connaissances a priori sur les contours (critères de gradient par exemple) (figure 7) .

Contours actifs flous Une des difficultés de la segmentation des images de tomosynthèse provient du fond texturé et du faible rapport contraste à bruit (Cheng 06).

Un certain nombre d'attributs d'image peuvent être pris en considération en concevant le terme d'énergie externe pour un modèle de contour actif. Les attributs les plus populaires incluent le gradient d'image (l'information de bord), l'intensité de pixel (l'information d'homogénéité de région), la texture et la couleur.

Afin d'obtenir une formulation du modèle de contour actif hybride, nous avons choisi de nous appuyer sur une approche qui maximise l'homogénéité de région et nous ajoutons une

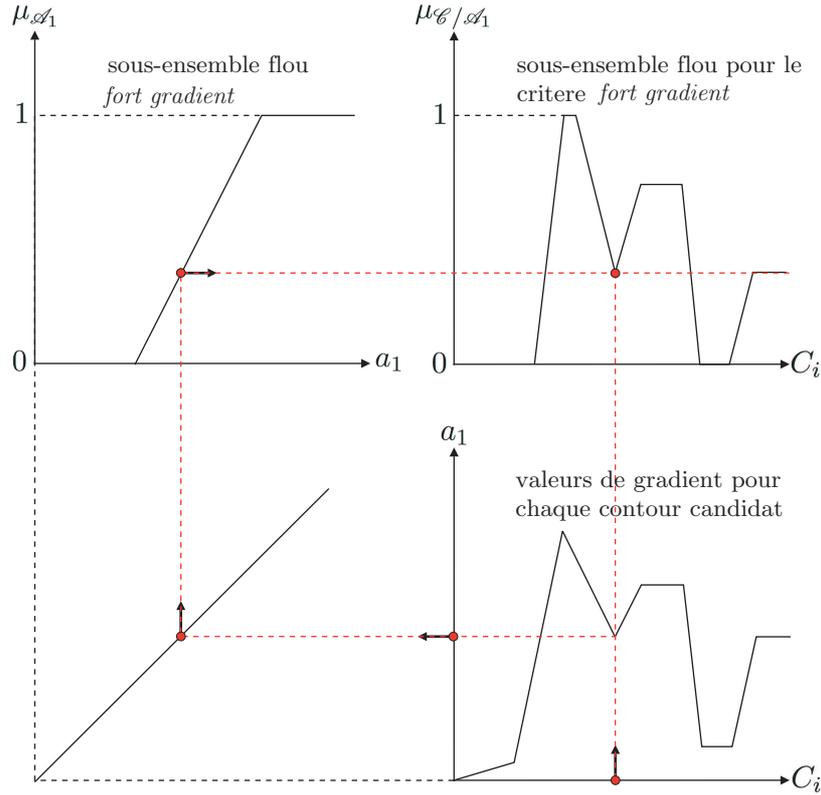


Figure 7: Des connaissances *a priori* sur l'image permettent la construction du sous-ensemble flou *fort gradient* (en haut à gauche). Pour une particule donnée la valeur du gradient est mesurée pour chaque contour candidat (en bas à droite). Ainsi le sous-ensemble flou *contour* relatif au critère *fort gradient* est calculé pour la particule (en haut à droite).

force externe prenant en compte l'information de gradient. La fonctionnelle d'énergie devient:

$$\begin{aligned}
 E(\phi) &= \mu \int_{\Omega} \delta(\phi(x, y)) |\nabla \phi(x, y)| dx dy \\
 &+ \nu \int_{\Omega} H(\phi(x, y)) dx dy \\
 &+ \alpha E_{region} + \beta E_{bord} + \gamma E_{pression}.
 \end{aligned} \tag{3}$$

où μ , ν , α , β et γ sont des paramètres de l'algorithme. Le terme d'homogénéité de région est donné par (Chan 01b) :

$$\begin{aligned}
 E_{region} &= \lambda_1 \int_{\Omega} |I_{org}(x, y) - \bar{c}_1|^2 H(\phi(x, y)) dx dy \\
 &+ \lambda_2 \int_{\Omega} |I_{org}(x, y) - \bar{c}_2|^2 (1 - H(\phi(x, y))) dx dy.
 \end{aligned} \tag{4}$$

Le terme de gradient est défini comme :

$$E_{bord} = - \int_{\Omega} \delta(\phi(x, y)) g(|I_{org}(x, y)|) dx dy, \tag{5}$$

avec $g(|I_{org}(x, y)|)$ une fonction d'arrêt conçue pour ralentir le contour dans le voisinage des forts gradients :

$$g(|I_{org}(x, y)|) = |\nabla(G_{\sigma}(x, y) * I_{org}(x, y))| \tag{6}$$

avec $G_\sigma * I_{org}$, une version lissée de I_{org} , la convolution de l'image I_{org} avec la gaussienne G_σ . La fonction $g(|I_{org}(x, y)|)$ vaut zéro dans des régions homogènes et de valeur importante sur les bords.

Enfin, le terme de pression est donné par :

$$E_{pression} = \int_{\Omega} \delta(\phi(x, y)) dx dy, \quad (7)$$

ce qui représente un terme de pression constante commandé par le paramètre constant γ . Ce terme est nécessaire pour conduire le contour vers les bords où le terme d'arrêt compensera le terme d'inflation et l'évolution du contour s'arrêtera.

Le choix d'un ensemble bien adapté de paramètres μ , ν , α , β et γ est crucial pour la bonne application de ce modèle aux données.

Nous employons ce modèle hybride de contour actif en tant que point de départ pour notre approche de segmentation floue. Nous ajustons les paramètres de l'algorithme selon des connaissances a priori sur un objet d'intérêt donné. Pour la démonstration nous choisissons ici une opacité circonscrite comme objet d'intérêt. Une opacité circonscrite est généralement compacte, avec un bord net. Les paramètres correspondant devraient donc refléter une forte régularisation (μ), et un poids fort sur le terme de gradient (β).

Une fois que le modèle converge et l'évolution s'arrête, nous avons trouvé le premier contour candidat. C'est le niveau zéro, ce qui est équivalent au contour extrait pour le cas classique. Ensuite nous tirons profit des propriétés de l'implémentation des contours actifs géométriques par ensembles de niveau en choisissant d'extraire des niveaux $\phi \neq 0$ de cette fonction (voir figure 8).

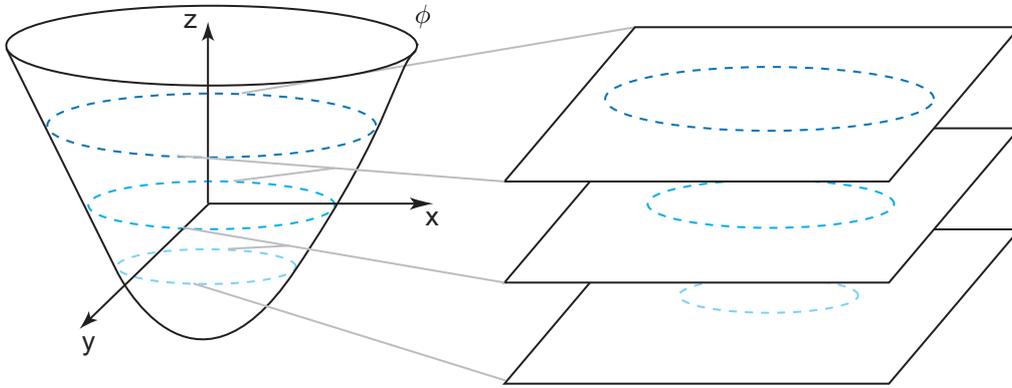


Figure 8: Extraction des ensembles de niveau: Nous pouvons extraire différents contours candidats correspondant à des valeurs de ϕ différentes.

Pour des raisons de simplicité nous avons choisi une approche plus directe. Nous calculons des niveaux ϕ_{l_i} également espacés dans une bande étroite autour du niveau zéro. Chaque contour est alors associé à un degré d'appartenance qui est dérivé de l'énergie du contour candidat.

Défuzzification partielle et agrégation A ce point nous disposons d'un ensemble d'images où chaque pixel peut être associé avec un grand nombre de valeurs. Nous voulons agréger ces images afin d'obtenir une représentation 3D de l'information extraite. Dans les projections chaque contour flou se compose d'un ensemble de contours candidats ainsi que de leur degré d'appartenance à la classe *contour* associée. L'agrégation directe de cette information n'est pas une tâche facile. Si nous envisageons une représentation équivalente dans l'espace 3D (par exemple un ensemble des surfaces candidats) nous allons rencontrer des difficultés topologiques

exigeant une stratégie d'agrégation très complexe. Afin de surmonter ces problèmes topologiques nous avons choisi de présenter une méthode pour simplifier la représentation des particules floues. Cette étape peut être vue comme défuzzification partielle.

Afin de transformer l'information d'image disponible en images appropriées à une méthode de rétro-projection, nous proposons de convertir la description actuelle de contour en une description par pixel (voir figure 9).

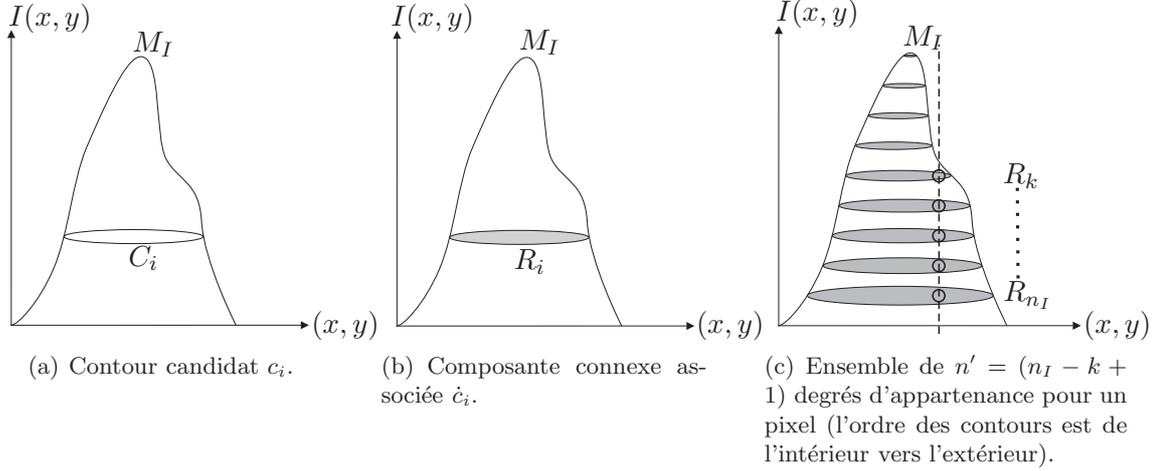


Figure 9: Conversion de la description de contour vers une description en chaque pixel.

La transformation que nous introduisons ici est une défuzzification de certains aspects des contours flous. Ainsi nous réduisons le degré de flou des particules candidates. L'ensemble des points appartenant à la particule est donc défini par :

$$\mathcal{R}_{M_I} = \left\{ R_i = R_{M_I}(I_{t_i}), t_1 = \max_{M_I}(I) > t_2 > \dots > t_{n_I} \right\} \quad (8)$$

avec $R_{M_I}(I_{t_i})$ la région pour l'ensemble $R_{M_I}(I_{t_i})_{i=1, \dots, n_I}$ qui inclut le marqueur M_I de l'image I , et I_{t_i} le seuil de I au niveau t_i .

Afin d'obtenir une valeur unique pour chaque position de pixel, nous définissons un opérateur qui combine l'ensemble des degrés d'appartenance $\mathcal{M}_{M_I}(x, y)$ en un degré d'appartenance unique $\mu_{\mathcal{C}}$ par pixel. Puisque nous n'avons aucune connaissance antérieure sur le *meilleur* contour, nous choisissons de procéder d'une façon prudente, qui nous amène à choisir un opérateur disjonctif. Ce type d'opérateur, c.-à-d. une t-conorme, réalise une union de l'ensemble d'information (Dubois 80).

En utilisant une t-conorme notée \perp , nous calculons un degré d'appartenance unique pour chaque pixel d'une particule candidate donnée par :

$$\forall (x, y) \in \mathcal{I}, \quad I^f(x, y) = \perp_m I_m^f(x, y) = \perp_m \perp_i \mu_{R_{m,i}}(x, y) \quad (9)$$

où $R_{m,i}(x, y)$ est la i ème composante connexe de la particule candidate marquée par le même marqueur qui contient le pixel (x, y) . $I^f(x, y)$ est la carte de particules floues résultante. Nous avons ainsi obtenu une représentation bien appropriée à une étape d'agrégation, combinant des informations acquises dans différentes vues de tomosynthèse en utilisant une méthode de rétro-projection.

Des exemples de défuzzification partielle appliquée aux données cliniques sont donnés pour une opacité circonscrite dans la figure 10.

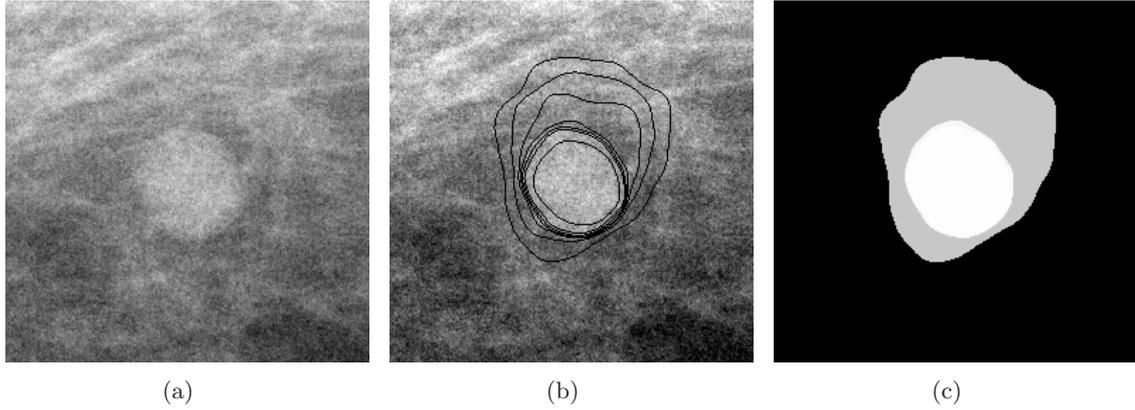


Figure 10: (a) Région d'intérêt dans une projection de tomosynthèse, (b) contours flous extraits, (c) carte de particules floues correspondante pour la détection d'opacités.

Sur la figure 10(c) nous remarquons que seulement deux degrés d'appartenance différents (non-nuls) sont présents dans la carte de particules floues. Puisque pour chaque pixel nous prenons le max de tous les degrés d'appartenance des contours contenant ce pixel, les contours n'auront pas tous un impact sur la carte de particules floues.

Après avoir effectué des détections floues individuelles dans chacune des N_p images de projection de l'acquisition DBT, nous avons produit une carte de particules floues correspondant à chaque image de projection. La correspondance spatiale entre l'ensemble de cartes de particules floues et le volume de particules floues résultant est établie en utilisant des connaissances a priori au sujet de la géométrie d'acquisition. Notre but est de trouver pour chaque voxel l'information correspondante dans toutes les N_p cartes de particules floues qui ont été créées. Chaque voxel (x_v, y_v, z_v) est projeté sur le plan du détecteur pour chacune de N_p positions du tube, ayant pour résultat un pixel (x_k, y_k) dans la k ème image de projection :

$$\forall (x_v, y_v, z_v) \in V^f, \quad \Lambda_k : (x_v, y_v, z_v) \mapsto (x_k, y_k), \quad k = \{1, \dots, N_p\} \quad (10)$$

où V^f est le volume de particules floues et Λ_k est la transformation géométrique qui établit le lien entre le voxel (x_v, y_v, z_v) et le pixel (x_k, y_k) défini par :

$$(x_k, y_k) = \Lambda_k(x_v, y_v, z_v) = (s_k(z_v) \cdot x_v + \xi_k(z_v), s_k(z_v) \cdot y_v + \eta_k(z_v)) \quad (11)$$

avec $\xi_k(z_v)$ et $\eta_k(z_v)$ les facteurs de décalage dans les directions x_v et y_v et $s_k(z_v)$ le facteur d'agrandissement, déduit de la géométrie d'acquisition (connue).

L'agrégation des informations recueillies dans les cartes de particules floues pour un voxel donné est exprimée comme:

$$V^f(x_v, y_v, z_v) = \Psi_{k=1}^{N_p} \left[I_k^f(x_k, y_k) \right] = \Psi_{k=1}^{N_p} \left[I_k^f(\Lambda_k(x_v, y_v, z_v)) \right] \quad (12)$$

où $V^f(x_v, y_v, z_v)$ est l'intensité du voxel à la position (x_v, y_v, z_v) du volume de particules floues, $I_k^f(x_k, y_k)$ est l'intensité du pixel à la position (x_k, y_k) de la k ème carte de particules floues, correspondant à la projection de la position (x_v, y_v, z_v) , et Ψ est l'opérateur d'agrégation. La figure 11 illustre cette opération d'agrégation.

Pour des raisons de simplicité, nous avons choisi la moyenne arithmétique comme opérateur d'agrégation pour combiner l'ensemble de cartes de particules floues. En utilisant cet opérateur

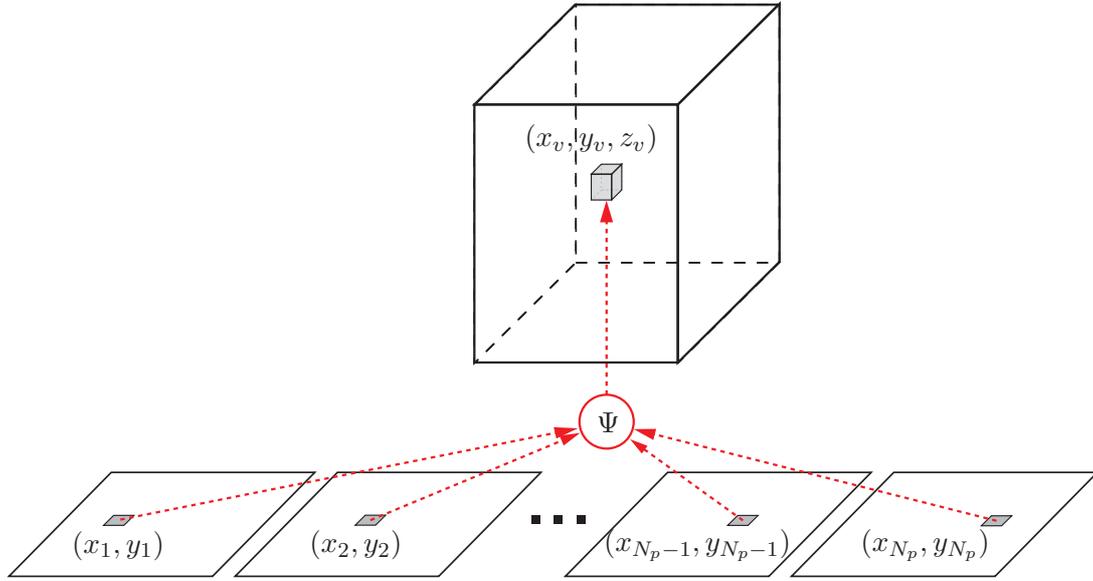


Figure 11: Stratégie d'agrégation d'information : pour un voxel (x_v, y_v, z_v) l'information de l'ensemble des pixels (x_k, y_k) correspondants est agrégée en utilisant l'opérateur Ψ . La position du pixel (x_k, y_k) qui correspond à la projection d'un voxel (x_v, y_v, z_v) donné dans la k ème carte de particules floues est calculée en prenant en compte la géométrie d'acquisition.

d'agrégation, l'équation 4.12 peut être réécrite comme:

$$\begin{aligned} V^f(x_v, y_v, z_v) &= \frac{1}{N_p} \sum_{k=1}^{N_p} \left[I_k^f(x_k, y_k) \right] \\ &= \frac{1}{N_p} \sum_{k=1}^{N_p} I_k^f(s_k(z_v) \cdot x_v + \xi_k(z_v), s_k(z_v) \cdot y_v + \eta_k(z_v)) \end{aligned} \quad (13)$$

avec $\xi_k(z_v)$ et $\eta_k(z_v)$ les facteurs de décalage et $s_k(z_v)$ le facteur d'agrandissement. Ces facteurs sont connus puisqu'ils sont directement liés à la géométrie d'acquisition.

Nous avons ainsi établi un cadre où la géométrie du système d'acquisition est prise en considération et utilisé pour rétro-projeter les cartes de particules floues dans l'espace 3D. Nous sommes cependant très flexibles concernant l'opérateur d'agrégation appliqué.

Établir un ensemble de particules 2D pour chaque particule 3D (quoique cette particule 3D ne soit jamais réellement calculée ou visualisée) est équivalent à trouver pour chaque particule 2D d'une projection donnée les particules correspondantes dans l'ensemble des autres projections. Nous pouvons alors passer toute l'information extraite pour une particule donnée en tant qu'un seul ensemble de données à la prochaine étape de traitement. En conséquence, la décision finale peut s'appuyer plus facilement sur l'ensemble des informations disponibles.

Le cadre géométrique général dérivé de la géométrie d'acquisition a été décrit ci-dessus. Cependant, notre but final n'est pas d'établir un lien entre les positions des pixels dans les vues projetées. En fait, nous voulons établir un lien entre différents candidats détectés dans les images de projection qui correspondent à la même particule tridimensionnelle.

Nous tirons profit de la technique de défuzzification partielle et de la stratégie d'agrégation correspondante décrite ci-dessus. Une fois que nous avons calculé le volume de particules floues, nous allons projeter les particules 3D (une par une) sur les cartes de particules floues 2D. Si la projection d'un objet tridimensionnel se recouvre avec une seule particule floue 2D ou avec aucune particule, la conclusion est simple. Dans le cas de chevauchements multiples des essais

supplémentaires doivent être réalisés afin de décider quelle particule nous allons associer à l'objet tridimensionnel.

L'étape d'agrégation est complète quand nous avons établi une description complète de chaque objet 3D qui tient compte de l'information extraite sur l'ensemble de projections.

Attributs flous cumulés Une particule floue 2D est caractérisée par un ensemble d'attributs flous $\{\mathcal{A}_{j,\mathcal{C}}\}_{j=1,\dots,N_a}$ extraits à partir du contour flou de la particule. N_a étant le nombre d'attributs flous mesurés à partir du contour flou \mathcal{C} . Nous recherchons un opérateur qui combine les fonctions d'appartenance d'un attribut flou donné $\mathcal{A}_{j,\mathcal{C}}$ à travers l'ensemble de projections. La fonction d'appartenance résultante doit représenter l'*attribut flou cumulé* $\mathcal{A}_{j,\mathcal{C}}^c$ pour la particule 3D.

Puisque nous développons une application de détection de cancer du sein, où les cancers manqués représentent un risque considérable pour la santé de la patiente, nous préférons procéder d'une façon prudente. Nous avons donc choisi une approche possibiliste, où chaque information est considérée comme observation possiblement correcte. Cela nous amène à une combinaison des différents attributs flous par l'intermédiaire d'une t-conorme. Nous avons choisi l'opérateur maximum (la plus petite t-conorme) pour combiner les fonctions d'appartenance d'un attribut donné à travers l'ensemble de projections.

L'*attribut flou cumulé* d'un objet 3D P_i^V est alors défini par :

$$\forall a_j \in \mathcal{A}_j, \quad \mu_{\mathcal{A}_{j,\mathcal{C}}^c}(a_j) = \frac{1}{k} \mu_{\mathcal{A}_{j,k,\mathcal{C}}}(a_j), \quad k = \{1, \dots, N_p\} \quad (14)$$

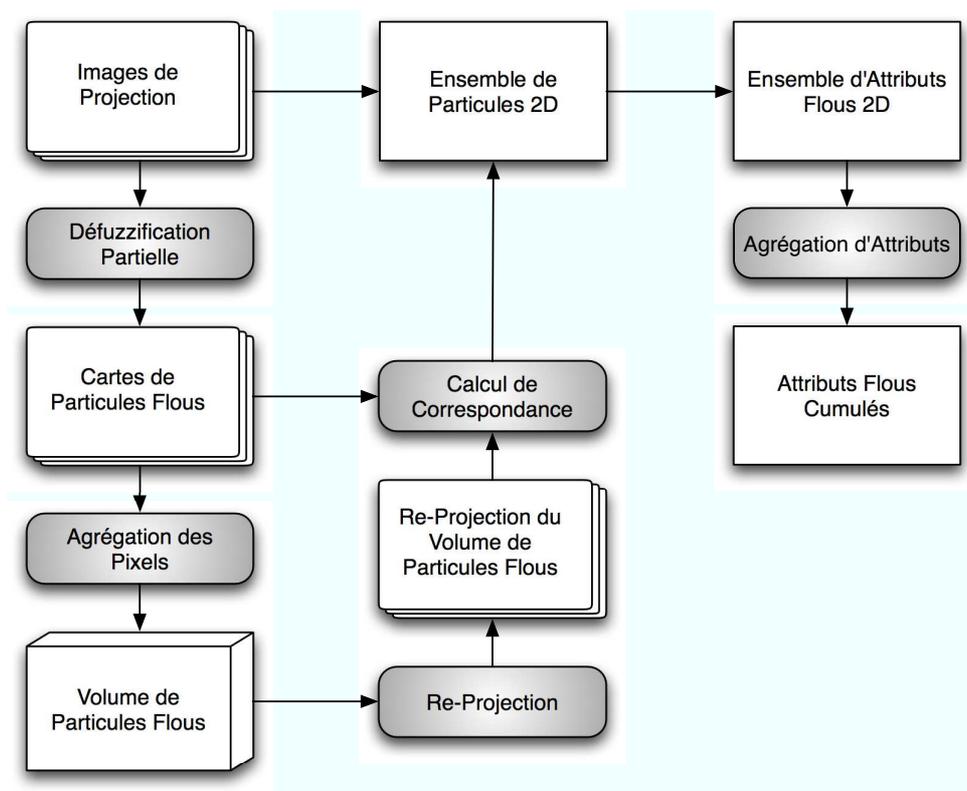


Figure 12: Schéma de l'agrégation au niveau des particules.

Arbre de décision flou Aujourd'hui, de nombreuses techniques de classification essayent de tirer profit des avantages du traitement flou. L'idée fondamentale d'introduire la théorie

des sous-ensembles flous dans le concept des arbres (classiques) de décision est de compenser une des imperfections centrales de ces arbres : la sévérité avec laquelle les valeurs examinées sont seuillées à chaque nœud de l'arbre. Des exemples ayant des valeurs d'attribut proches l'un de l'autre, mais de différents côtés du seuil de nœud, prendront des chemins complètement différents dans l'arbre de décision classique. Afin de surmonter cette limite, nous employons une des propriétés bien connues de la théorie des sous-ensembles flous : l'appartenance graduelle à une classe ou à un ensemble donné.

L'arbre de décision flou accepte donc les exemples qui suivent avec un certain degré différents chemins à travers l'arbre. Dans les arbres de décision classiques, chaque exemple finit son chemin en une feuille unique de l'arbre, alors que dans un arbre de décision flou chaque exemple descendra tous les chemins possibles de l'arbre pour arriver dans l'ensemble des feuilles de l'arbre. Le degré avec lequel un exemple donné arrive dans une feuille donnée reflète son chemin à travers l'arbre, et les tests au nœud le long du chemin.

Dans un arbre de décision flou le test d'attribut $attribut > seuil$ est remplacé par le test $attribut\ flou = valeur$, où la valeur est représentée par une fonction d'appartenance.

La mesure de discrimination que nous nous appliquons s'appuie sur le contraste entre deux sous-ensembles flous (Bothorel 96):

$$C(\mathcal{A}_1, \mathcal{A}_2) = \frac{\int \min(\mu_{\mathcal{A}_1}, \mu_{\mathcal{A}_2})}{\int \max(\mu_{\mathcal{A}_1}, \mu_{\mathcal{A}_2})} \quad (15)$$

où \mathcal{A}_k est le sous-ensemble flou qui correspond à la classe c_k (avec la fonction d'appartenance $\mu_{\mathcal{A}_k}$) et $C(\mathcal{A}_1, \mathcal{A}_2)$ est une mesure de contraste entre les deux sous-ensembles flous \mathcal{A}_1 et \mathcal{A}_2 . Les intégrales sont calculées pour les domaines de définition des sous-ensembles flous. L'attribut qui minimise $C(\mathcal{A}_1, \mathcal{A}_2)$ est considéré le *meilleur*.

Une fois que nous avons identifié le *meilleur* attribut, nous créons le nœud de racine ayant comme test au nœud la fonction de référence de cet attribut. Puis, la base d'apprentissage entière est examinée par ce nœud. Pour chaque échantillon nous calculons le degré d'appartenance à chaque classe. Chaque échantillon de la base d'apprentissage passera ainsi dans chaque sous-arbre avec le degré d'appartenance associé à la classe.

Pour un arbre ayant plus de deux niveaux hiérarchiques (au moins un nœud interne sans compter le nœud de racine), une autre particularité doit être prise en considération. À chaque nœud, un degré d'appartenance à chaque classe sera calculé pour les échantillons. Pour un échantillon passant par plusieurs nœuds consécutifs, nous devons combiner ces degrés d'appartenance pour refléter le chemin que l'échantillon a parcouru à travers l'arbre. Cette combinaison des degrés d'appartenance est appelé *degré cumulé*. Le degré cumulé réalise une conjonction des tests au nœud pour un échantillon et un chemin à travers l'arbre donné. Puisque nous travaillons sur des sous-ensembles flous, la conjonction des sous-ensembles flous qui servent de tests au nœud est réalisée par une t-norme.

L'algorithme pour la construction de l'arbre est répété jusqu'à ce qu'un des critères d'arrêt définis soit satisfait. Ceux-ci peuvent inclure :

- le nombre d'échantillons par feuille est inférieur à une valeur donnée,
- la pureté de chaque feuille est supérieure à une valeur donnée,
- tous les attributs ont été employés (au moins) une fois.

Si tous les nœuds internes ont été traités, nous trouvons chaque échantillon dans chaque feuille avec un degré cumulé donné. Nous pouvons alors calculer la densité de classe pour la classe L_c

dans la feuille i par :

$$n_i^{L_c} = \sum_{j=1}^{\aleph_{L_c}} \mu_{i,j}^c, \quad (16)$$

où \aleph_{L_c} est la cardinalité de la base d'apprentissage pour la classe L_c et $\mu_{i,j}^c$ est le degré cumulé de l'échantillon j dans la feuille i .

Chaque chemin à travers l'arbre peut être interprété comme jeu de règles décrivant un objet qui appartient à l'une des classes. Une classe peut alors contenir des définitions très différentes d'objets. Ici, nous voulons calculer le degré d'appartenance à la classe pour un échantillon donné. Puisqu'il est suffisant que l'échantillon satisfasse une des définitions de l'objet pour être considéré comme un membre de la classe, nous effectuons une union floue des différents degrés d'appartenance à la classe. Cela nous amène à appliquer une t-conorme à chaque ensemble de degrés cumulés.

En construisant un arbre de décision flou pour traiter des données floues nous devons tenir compte de la nature floue des échantillons. Ainsi, des modifications concernant le calcul des degrés d'appartenance et la construction des fonctions de séparation pour les tests aux nœuds doivent être appliquées.

Dans le cas des données floues les valeurs d'attribut de chaque échantillon sont elles-mêmes des sous-ensembles flous. Nous devons donc calculer le degré d'appartenance d'un sous-ensemble flou à un autre sous-ensemble flou. C'est un problème assez commun dans la théorie des sous-ensembles flous, qui est généralement exprimé comme degré de *satisfiabilité* S d'un sous-ensemble flou concernant un autre :

$$S(\mathcal{A}, \mathcal{B}) = \frac{\int \mu_{\mathcal{A} \cap \mathcal{B}}(x) dx}{\int \mu_{\mathcal{B}}(x)} \quad (17)$$

où \mathcal{A} et \mathcal{B} sont des sous-ensembles flous et les intégrales sont calculées sur le domaine de définition des sous-ensembles flous.

Afin d'établir une fonction séparatrice, un histogramme flou est calculé de façon analogue à un histogramme classique. Pendant le procédé de construction d'histogramme nous devons simplement ajouter une fonction pour chaque échantillon au lieu d'une seule valeur. Avant que nous ajoutions la valeur d'attribut d'un échantillon donné à l'histogramme, nous normalisons l'attribut en divisant par l'aire sous la fonction d'appartenance.

En outre, les aires des histogrammes d'attributs flous sont également normalisés à 1. C'est équivalent à dire que nous imposons que les histogrammes contiennent un nombre égal d'attributs afin de pouvoir les comparer.

À chaque nœud de l'arbre, l'histogramme flou est calculé pour chaque attribut. La contribution de chaque échantillon est pondérée par le degré cumulé de l'échantillon en ce nœud.

Nous avons maintenant tous les composants nécessaires pour construire un arbre de décision flou capable de traiter des données floues. Il est important de noter qu'on a fait quelques hypothèses a priori sur les attributs qui justifient le traitement appliqué. Ces hypothèses incluent des propriétés importantes des attributs qui caractérisent les objets d'intérêt :

- monotonie,
- régularité,
- monomodale.

Ces conditions sont vraies pour la majorité des attributs utilisés dans la détection assistée par ordinateur dans des images de mammographie. Des exemples incluent *le gradient*, *la compacité*, *l'élongation*, *l'homogénéité*, et de nombreux attributs s'appuyant sur l'histogramme et le niveau de gris. Un exemple pour un attribut qui ne satisfait pas cette condition est *l'aire* d'une opacité.

Une autre limitation importante de l'algorithme est que les attributs ne peuvent pas être employés plus d'une fois dans le même chemin à travers l'arbre (du nœud de racine à une feuille donnée).

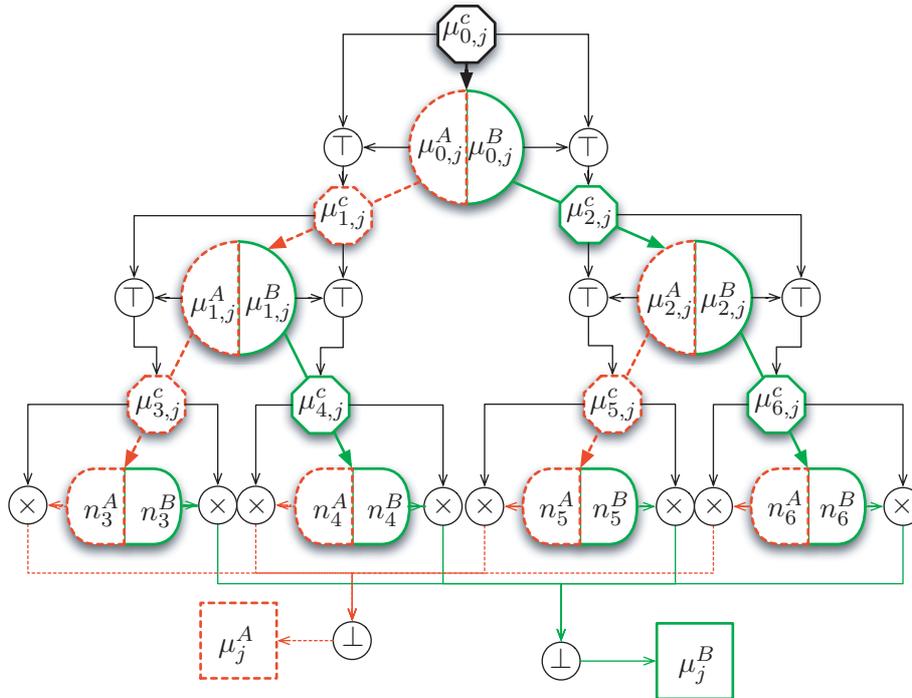


Figure 13: Arbre de décision flou complet : à l'intérieur des nœuds se trouvent les degrés d'appartenance $\mu_{i,j}^A$ et $\mu_{i,j}^B$ à chaque classe calculés par le test au nœud correspondant. Les octogones contiennent les degrés cumulés $\mu_{i,j}^c$ pour l'échantillon donné. Chaque feuille contient les densités de classe n_i^A et n_i^B pour les deux classes. Dans notre implémentation, les degrés cumulés sont calculés en utilisant l'opérateur min. En raison de la base de données limitée, nous pondérons les densités de classe dans chaque feuille. La décision finale est obtenue en appliquant l'opérateur max.

Résultats Un algorithme de CAD est censé servir comme outil pour aider des radiologues dans leur tâche de détection et de diagnostic. Afin d'évaluer un tel outil nous devons créer une situation représentative pour son usage prévu et mesurer la performance. Nous envisageons de calculer certains indices de performance. Ceux-ci sont généralement exprimés en termes de faux positifs (FP) et faux négatifs (FN). Les FN sont souvent remplacés par la sensibilité (Se) de l'algorithme. Dans notre cas, quoique nous disposions d'un grand nombre d'images cliniques, la vérité clinique est absente entièrement. Nous ne pouvons donc pas évaluer l'exécution de l'algorithme proposé selon des techniques classiques d'évaluation à l'heure actuelle.

Nous avons réalisé un examen visuel de cas choisis de DBT avec un groupe d'experts en mammographie afin d'identifier les opacités. Nous évaluons la performance de l'algorithme et sa capacité à distinguer les différents genres d'opacités (les opacités circonscrites et les opacités stellaires) qui ont été identifiées pendant la revue d'expert.

Puisque nous avons seulement un nombre très limité de cas cliniques à notre disposition, nous

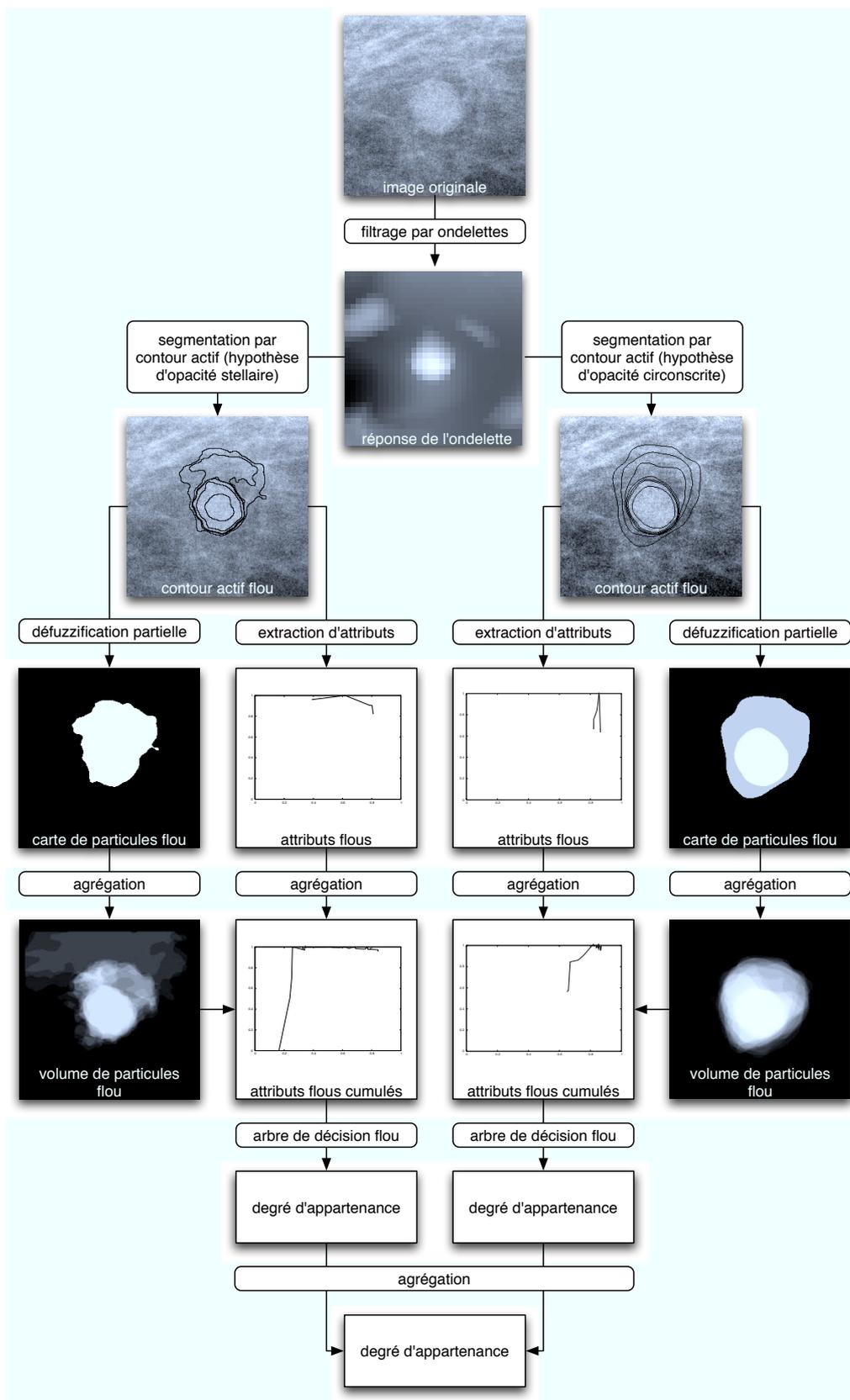


Figure 14: Un échantillon traité par l'algorithme complet proposé. Les résultats sur un cas clinique contenant une opacité circonscrite. De haut en bas : image originale brute, réponse du filtre d'ondelettes (sous-échantillonnée), contours actifs flous, carte de particules flous 2D, coupe reconstruite du volume de particules flous. Un attribut flou extrait pour chaque modèle et l'attribut flou cumulé correspondant est montré à côté des cartes/volumes de particules flous.

avons dû exécuter l'apprentissage et les tests avec la même base de données. Nous avons appliqué la méthode *leave-one-out*, où un échantillon est examiné avec un arbre qui était construit avec la base de données entière excepté l'échantillon à examiner. Ce processus est répété pour chacun des échantillons dans la base de données.

Après avoir calculé la puissance de discrimination pour chaque attribut, nous pouvons construire un arbre entier pour chaque modèle (circonscrit et stellaire). L'arbre de décision flou pour le modèle des opacités stellaires comporte deux attributs (*compacité* et *ratio de Feret*), alors que l'arbre pour le modèle des opacités circonscrites était construit à partir de trois attributs (*compacité*, *ratio de Feret* et *homogénéité*). Puisque nous appliquons la méthode *leave-one-out* nous devons construire un nouvel arbre pour chaque échantillon d'essai (construit avec les échantillons restants de la base de données).

La classification de la base de données par l'arbre de décision flou en utilisant la méthode *leave-one-out* résulte dans deux arbres (modèle circonscrit et modèle stellaire) pour chaque exemple de la base de données. Ainsi nous obtenons quatre degrés d'appartenance associés. Afin de calculer l'erreur de classification nous devons combiner ces degrés pour obtenir une décision concernant l'exemple. Nous avons besoin alors de défuzzifier les résultats de classification de l'arbre de décision flou. Pour chaque arbre nous considérons que le max de deux valeurs d'appartenance correspond à l'étiquette résultante de classe de l'exemple examiné.

Nous obtenons une erreur de classification de 17.6 % (3 de 17 cas faussement classifiés) pour le modèle d'opacité circonscrite et une erreur de classification de 17.6 % (3 de 17 cas faussement classifiés) pour le modèle d'opacité stellaire. Puisque nous disposons de multiples arbres de décision, ces valeurs seules ne suffisent pas pour évaluer la performance de classification. En fait nous devons étudier la combinaison des résultats de classification. Nous obtenons 11.8 % (2 sur 17) de cas contradictoires et 11.8 % (2 sur 17) de cas qui ont été faussement classifiés par les deux arbres de décision flou. La performance de classification combinée se trouvera donc entre 11.8 % et 23.6 % selon la méthode employée pour combiner les résultats des deux arbres de décision flous.

Afin de combiner les résultats des deux arbres de décision flous, nous calculons un degré de confiance pour chaque étiquette de classe à partir de la dissimilitude des deux valeurs d'appartenance.

L'erreur de classification employant les deux types d'arbres de décision flous (opacités circonscrites et opacités stellaires) s'élève donc à 17.6 % (3 échantillons sur 17 ont été classifiés de manière erronée). Cela n'est cependant pas une mesure très représentative, puisque la plupart des degrés de confiance cumulés sont très petits. La décision pour une des classes n'était donc pas prise avec beaucoup de confiance. Cela est surtout dû à la base de données qui n'est pas représentative de la population ciblée. Avec une base de données aussi limitée à notre disposition, les résultats de classification montrés ici peuvent seulement servir à démontrer la validité de notre approche. Une fois qu'une base de données plus importante sera disponible, nous devons exécuter une sélection d'attributs ainsi qu'une évaluation des performances statistiquement appropriées.

Afin d'obtenir une meilleure évaluation de l'approche proposée nous allons comparer les résultats de classification obtenus à un algorithme de CAD différent du nôtre. À cet effet nous avons choisi une approche qui ne s'appuie pas sur la théorie des ensembles flous.

Les résultats d'une classification par un arbre de décision classique en utilisant la méthode *leave-one-out* donnent une erreur de classification de 35.2 % (6 cas sur 17 mal classifiés) pour le modèle d'opacité circonscrite et une erreur de classification de 17.6 % (3 cas sur 17 mal classifiés) pour le modèle d'opacité stellaire. Comme dans le cas flou, nous traitons de multiples arbres de décision et devons ainsi extraire des informations supplémentaires afin d'évaluer correctement la performance de l'algorithme. Nous obtenons 41.2 % de cas contradictoires (7 sur 17) et 5.9 %

de cas (1 sur 17) qui ont été mal classifiés par les deux arbres de décision. La performance de classification combinée se trouvera donc entre 5.9 % et 47.1 % selon la méthode employée pour combiner les résultats des deux arbres de décision.

Les résultats obtenus sont très prometteurs. Particulièrement dans l'étape de classification où une décision finale est calculée, les degrés d'appartenance fournissent un avantage significatif.

I may not have gone where I intended to go, but I think I have ended up where I intended to be.

Douglas Adams

1

Introduction

Breast cancer continues to be one of the leading causes of cancer mortality among women (Imaginis 07). In France with about 42,000 new cases in 2000, breast cancer represents more than 35% of new cases of cancer (Remontet 03). The underlying causes for this disease remain unknown. Certain risk factors are known (family history, age at first full-term pregnancy, ionizing radiation, genetic factors, ...) but they account only for about 30% of all breast cancer cases (GAO. 91). There is no proven treatment to prevent breast cancer. Early screening is therefore the only known means to reduce mortality among the female population.

X-ray mammography is currently the primary method for detecting early breast cancers, reducing the mortality rate by about 30% for women 50 years and older (Kerlikowske 95). It plays a key role in detecting more cancers at an earlier stage (Olsen 01). Since the probability that a tumor produces metastases increases with its size, it is of primary importance to detect tumors at the earliest possible stage of development (Duffy 03).

However, about 30% of breast cancers are still missed by conventional screening mammography. One of the main reasons is the superimposition of tissue that obscures lesions in dense breasts (Holland 82). Digital mammography recently demonstrated superior performance in dense breast imaging compared to film screen mammography (Pisano 05). Digital Breast Tomosynthesis (DBT) (Dobbins III 03; Wu 03b) is a new three-dimensional (3D) limited-angle tomography breast imaging technique that will substantially overcome the superimposition problem for lesion detection. This thesis focuses on the processing and interpretation of the very recent images obtained with DBT.

Several projected views from different acquisition angles used to reconstruct slices parallel to the detector plane will potentially reduce the number of false positives (FP) caused by summation artifacts as well as the number of false negatives (FN) caused by the masking effect of overlying tissue. Furthermore, back projecting the tomographic views to reconstruct a 3D volume allows for an analysis of 3D lesion features such as volume, shape and spatial relations between candidates. At the same time we are still facing a number of challenges related to the image data, some of which are related to its tomographic nature. The dose per acquired image is significantly reduced in comparison to standard 2D mammograms, to maintain a comparable total patient dose per scan. This has a major impact on the Contrast to Noise Ratio (CNR) and thus on the detectability of breast lesions in the projected views. Working on a 3D volume provides an improved contrast when compared to tomography projections. The contrast for a

given object is however anisotropic (see Section 2.2.3) and the best contrast is obtained for a reconstructed slice that crosses the center of the object.

For identifying a potential breast lesion the radiologist still needs to inspect the entire set of reconstructed slices. A DBT acquisition typically contains between 40 and 80 slices per breast. The workload for radiologists is thus significantly increased when compared to standard 2D mammography.

At the same time, in about two-thirds of the lesions missed during the screening of 2D mammograms, the radiologist failed to detect the cancer that was evident retrospectively (Martin 79; Buchanan 83). Such misses might be due to the subtle nature of the visual findings, poor image quality, fatigue, or oversight by the radiologists. Double reading (by two radiologists) might increase sensitivity (Bird 90; Murphy Jr 90; Thurfjell 94). Thus, one goal of CAD is to increase the efficiency and effectiveness of breast cancer screening by using a computer system as second reader to aid the radiologist by pointing out suspect locations in mammograms (Vyborny 94). Screening mammograms is a repetitive task involving mostly normal images, which makes it well suited for computer analysis.

Another goal of CAD is to objectively analyze the characteristics of benign and malignant lesions seen on mammograms. This is expected to help the radiologist interpret lesions and thus decrease patient morbidity as well as the number of surgical biopsies performed and their associated complications. After a radiologist finds a suspicious lesion on a mammogram, he or she must then visually detect various radiographic characteristics. Analyzing such features, the radiologist then decides if the abnormality is likely to be malignant or benign and what course of action to recommend. Radiologists refer many patients for surgical biopsy on the basis of a radiographically detected mass lesion or cluster of microcalcifications. Although general rules for the differentiation between benign and malignant breast lesions exist (Bassett 87; Tabár 01), considerable variability occurs in lesion interpretation by radiologists with current radiographic techniques (Ciccone 92; Elmore 94).

Numerous studies have shown that CAD applied as an additional reader provides improved detection sensitivity compared to screening performed by human readers only (Freer 01; Karssemeijer 03; Gilbert 06). The objective of the present work is to conceive such a CAD algorithm for DBT data.

The three fundamental components of CAD are detection, analysis and decision. A detection step identifying the location of structures that may correspond to lesions, hereafter called candidates, is followed by an analysis step that extracts information about the candidates. In the final step a decision is taken based on the information extracted during the analysis. This decision leads to the differentiation between structures of interest, associated to suspicious lesions, and other detected candidates as well as to a differentiation between different types of breast lesions. As far as the first step is concerned, most approaches are able to achieve an almost perfect detection score, given that we accept a sufficiently high number of false alarms. Here, we will concentrate on the two other steps, where we try to eliminate false alarms based on information extracted from the image data.

The major challenge is then to master the ambiguities present in the images. Ambiguities may stem from numerous sources including: image noise, photon scatter, patient motion, superimposition of different structures, reconstruction artifacts, sampling and the diversity of the objects of interest. In this work we aim at modeling these ambiguities and making a decision through combination of several ambiguous information extracted from a given object. We use fuzzy set theory to model the ambiguity present in the images. We will show how the concept of gradual membership degree to a class can be used for this purpose.

In particular, we want to gather information over various consecutive processing steps. The final decision should then be based on the full range of gathered information. We need to avoid

elimination of candidates before all available information has been considered. This leads us to the basic structure of an automated detection algorithm. Three basic concepts for automated detection from tomographic data may be considered (Figure 1.1).

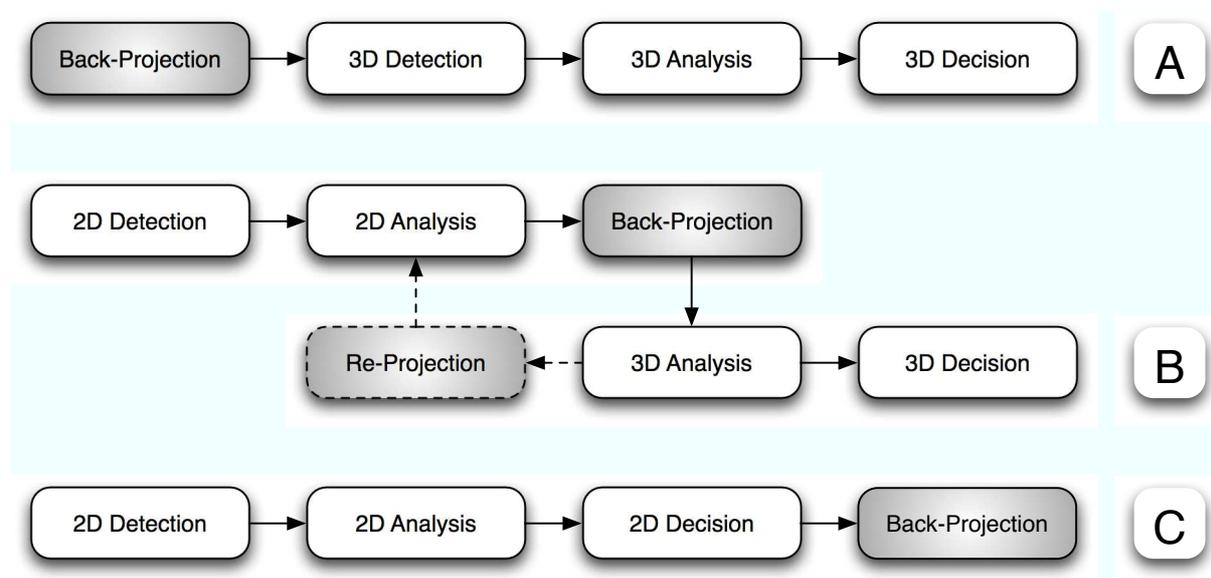


Figure 1.1: CAD Strategies: (A) Working directly on 3D data, (B) performing the detection and part of the analysis in 2D, backprojecting the intermediate results and performing further analysis and the decision step in 3D (this may be extended towards an iterative approach by adding a reprojection step), (C) performing the entire processing in 2D and only backprojecting the result into 3D.

In the first case (Figure 1.1.A) processing is performed exclusively on the 3D reconstructed volume. This approach is highly dependent on the algorithm used for reconstructing the 3D volume from the projected views. Since DBT is an emerging modality, 3D reconstruction algorithms for its particular geometry are still not fully optimized. The third approach (Figure 1.1.C) works on the projected view only and passes into 3D space only for presentation of the obtained results. However, we consider crucial to delay the detection decision for each candidate until information from each view can be jointly considered. Our approach therefore follows the concept of the approach shown in Figure 1.1.B. If part of the analysis is performed in 2D a simple backprojection may not be well adapted. Backprojection is per definition the combination of different projection images to reconstruct an image volume. If we wish to combine other information than intensity images, we will need to replace the usual backprojection step by a dedicated aggregation operator.

Finally an automated detection scheme should be able to identify any structure of interest present in the breast. In mammography, signs of a potential cancer may manifest in a variety of very different structures. We should aim to develop a high-level framework for automated detection that may easily apply for the detection of different types of breast structures and even to different modalities altogether.

The aim of this work is to derive such a high-level framework for computer-aided detection on DBT data.

We start by summarizing the context of digital mammography, digital breast tomosynthesis and CAD in Chapter 2. We briefly discuss anatomical details of the breast as well as the different radiological findings and their clinical indications. We introduce the physical principles of a mammography acquisition, and point out the particularities of digital breast tomosynthesis such as image dynamics, angular range, point spread function and acquisition artifacts. A short

overview of other breast imaging techniques is given. Finally, we present the state of the art of CAD algorithms in 2D, and present few recent publications concerning CAD for DBT. We present our approach to automated detection, situate the approach among the existing methods and motivate our choices.

The processing performed on the tomographic projection images are presented in Chapter 3. We discuss a detection operation based on multi-scale wavelet filter-banks that provides a marker for each detected object. We then concentrate on extracting fuzzy contours for the marked particles. A multi-thresholding approach is presented to compute fuzzy contours of microcalcification candidates. We introduce a new *hybrid active contour model* for mass segmentation. Based on this model we develop a scheme for extracting *multiple fuzzy contours* for mass candidates. Each fuzzy contour corresponds to a different hypothesis concerning the segmented particle. Finally a feature vector is extracted for each fuzzy contour. Using the extension principle we establish a set of fuzzy attributes based on this feature vector that provides additional information about the segmented particle.

In Chapter 4 we focus on the aggregation of information over the entire set of projected views. We introduce a *partial defuzzification* operation that provides a pixel-based representation of the fuzzy information extracted for each object, so-called *fuzzy particle maps*. Taking a priori knowledge about the acquisition geometry into account, we develop a back-projection/re-projection scheme based on these fuzzy particle maps. We show how this provides the necessary link between a given three-dimensional particle and the corresponding two-dimensional particles in the projected views. Using this information we introduce an aggregation strategy resulting in *cumulated fuzzy attributes* for each three-dimensional particle.

The computation of a final decision from the extracted data is described in Chapter 5. We present a fuzzy decision tree classifier able to process fuzzy input data. We show how the decision tree is working directly on the extracted cumulated fuzzy attributes and how this impacts the classification result. We insist on the fact that the use of fuzzy set theory allows us to model the ambiguity at each stage of the algorithm, and how we can avoid taking intermediate decisions. Instead, the entire range of extracted information is passed to the fuzzy decision tree. We talk about the different results obtained using the fuzzy decision tree, and how our earlier choices impact the interpretation of these results.

In Chapter 6 we illustrate an application example for the developed computer-aided detection framework. We discuss the available database and the issues concerning the clinical data for an emerging modality such as DBT. We demonstrate the behavior and performance of the individual processing steps and we provide quantitative as well as qualitative results of the application. A performance assessment of the complete processing chain is also being provided.

The contributions of our research work will be summarized and evaluated in Chapter 7. We will provide an outlook into future research paths and we will point out potential improvements for the presented methods.

The real voyage of discovery consists not in seeking new landscapes but in having new eyes.

Marcel Proust

2

Context and Motivation

This chapter aims to introduce the context of the presented work. The description of breast anatomy and breast pathologies given in the following sections is by no means exhaustive. It is merely intended to facilitate the understanding of the problem to be solved and the challenges we have to overcome. Numerous works have been published on different aspects of breast anatomy, breast pathologies and their translations in mammography images (Pisano 01; Travade 94; Kopans 89; Lanyi 87; Holland 82). However, not all knowledge in this field can be easily expressed in words. A decision taken by a radiologist, is mainly based on expert knowledge acquired over many years of professional experience, and is often difficult to express verbally. The task of integrating such knowledge into image processing algorithms therefore presents an even greater challenge, that we will try to address in the proposed approach.

When we conceive a CAD scheme, the motivation is not to replace the radiologist. On the contrary, the primary motivation for development of CAD systems is to assist the radiologist. We aim to develop a tool that facilitates the radiologists work and allows them to improve their performance (see Section 2.3).

2.1 The Breast

In this section we present the anatomy of the breast as well as the pathologies that may arise in the corresponding tissues in order to introduce the context for which our application has been developed.

2.1.1 Anatomy of the Breast

A simplified representation of the breast anatomy is given in Figure 2.1. The different structures the breast is composed of are:

- skin (10),
- Coopers ligaments (5),
- fatty tissue (6),

- glandular conjunctive tissues (11)
- lobules (7),
- ducts (8),

The breast is situated along the:

- pectoral muscle (4),
- chest wall / ribs (1).

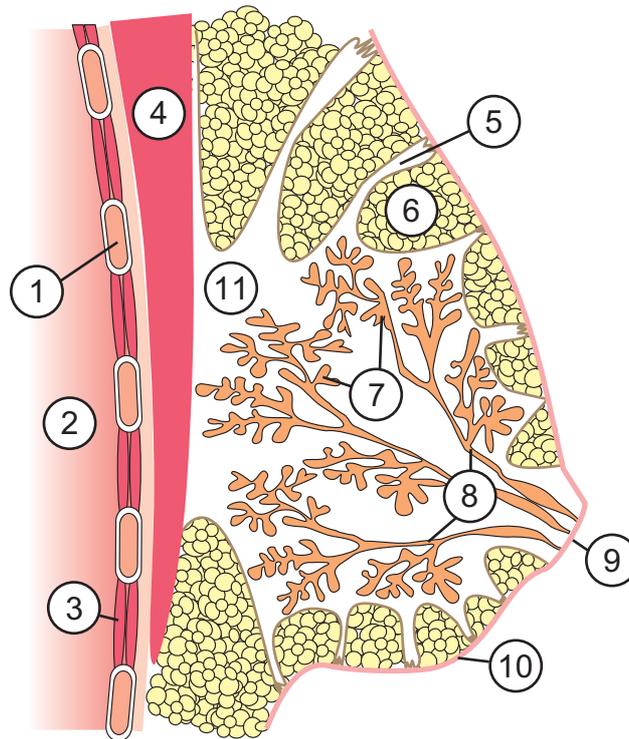


Figure 2.1: Simplified anatomy of the breast: (1) rib, (2) lung, (3) intercostal muscles, (4) pectoralis major muscle, (5) suspensory ligaments (of Cooper), (6) fat, (7) gland lobules, (8) lactiferous ducts, (9) nipple, (10) skin, (11) glandular conjunctive tissues.

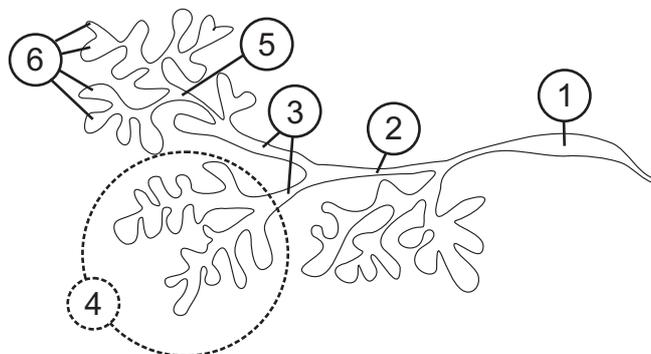


Figure 2.2: Schematic of a branching duct: (1) lactiferous sinus, (2) main duct, (3) extralobular terminal duct, (4) lobule, (5) intralobular part of the terminal duct, (6) ductules.

The mammary lactiferous system within the breast produce the milk. It consists in several lobules connected by ducts. Each breast has some 10-20 lactiferous ducts that drain milk from the lobules to the nipple, where each duct has an opening. For understanding the breast functionality we closer examine one of these ducts (Figure 2.2). Starting at the nipple, the duct widens to form the lactiferous sinus (1). Then, the main duct (2) narrows, branches and eventually forms the terminal ductal lobular unit, consisting of the extralobular terminal duct (3) and the lobule (4). From the intralobular part of the terminal duct (5), ductules (6) develop as blindly ending structures. 30-50 ductules are present in each lobule. Note that the lobules have varying degrees of development that have been classified as lobule types 1-3 (Russo 94).

2.1.2 Breast Cancer

Breast cancers develop out of an uncontrolled cell growth. When the malignant cells stay in the lactiferous ducts we speak of a cancer *in situ*. When they start to invade the surrounding tissues, the cancer is qualified as being *invasive*.

Not all possible degenerations of breast tissues automatically lead to development of a cancer. Examples of benign lesions are:

- fibrocystic changes
- fibroadenoma
- phyllodes tumors
- intraductal papilloma
- granular cell tumor
- fat necrosis
- mastitis
- duct ectasia

About 80% of all breast nodules detected during clinical screening are benign. The three basic steps of the progression toward cancer can be summarized as follows:

- Dysplasia (Latin for *bad form*) is the earliest form of pre-cancerous lesion recognizable in a biopsy by a pathologist. Dysplasia can be *low grade* or *high grade*. The risk of low-grade dysplasia transforming into high-grade dysplasia and, eventually, cancer is low. Treatment is usually easy.
- Carcinoma *in situ* is synonymous with high-grade dysplasia in most organs. The risk of transforming into cancer is high. Treatment is still usually easy, mainly based on minimally invasive surgery.
- Invasive carcinoma, is the final step in this sequence. It is a disease that, when left untreated, will invade the host and will probably kill her or him. It can be treated successfully when detected and diagnosed at an early stage.

Examples for malignant breast diseases are:

- carcinoma in situ (ductal and lobular),

- invasive carcinoma (ductal and lobular),
- inflammatory breast cancer.

Invasion and metastasis are the most insidious and life-threatening aspects of cancer. The capacity for invasion may not be expressed initially or in all tumors, hence the ability to cure most *in situ* lesions with a local intervention. However, most cancers unmask their invasive potential, thus progressing to frank malignancy from pre-existing carcinoma *in situ* (amongst others). Once a lesion becomes invasive, it can disseminate readily through whatever means or conduits are available to it, via the lymphatics and/or vascular channels. Invasion and metastases kill hosts through two processes: local invasion and distant organ injury. Local invasion can compromise the function of involved tissues by local compression, local destruction, or prevention of normal organ functioning. This is however not a life threatening concern in breast cancer since the breast does not constitute a vital organ. The most significant turning point in the disease is the establishment of metastasis. At this stage, the patient can no longer be cured by local therapy alone.

A malignant tumor becomes clinically detectable after a mean time of about 10 years and the process of forming metastases may occur prior to the stage where the tumor becomes detectable. The earlier a tumor is detected, the lower the risk that it has already formed metastases. Therefore, it is of paramount importance to detect the tumor in its earliest possible stage of development. For tumors that are detected while their diameter is still inferior to one centimeter, the survival rate rises as high as 98% (ACS 07).

Event though breast cancer continues to be one of the leading causes of cancer mortality among women, the underlying causes for this disease remain unknown. Some indicators of increased risk for breast cancer have however been identified. They include sex, age, breast density (Wolfe 76) and medical family history.

2.2 Breast Imaging

2.2.1 Mammography

Mammography provides a projection of the breast volume onto the image plane. In order to obtain a good contrast, images are acquired using relatively low-energy photons because the breast is only composed of soft tissues. Filters and anodes made of Molybdenum (Mo) or Rhodium (Rh) can be used to generate an X-ray spectrum of adequate energies (Muller 99). The detector captures those photons that are transmitted through the breast tissue. As a result, with digital detectors, the pixel intensity in the acquired images is proportional to the energy of photons arriving at the detector. These are so-called intensity images, where the signal intensity (assuming monoenergetic photons) is given by:

$$I = I_0 e^{-\int_0^L \mu(x,y,z) dl} \quad (2.1)$$

where μ is the linear attenuation coefficient of the penetrated tissue along the path of the X-ray and L its thickness (Figure 2.3).

The advantage of being able to visualize the entire range of breast tissue in a single image is at the same time the greatest weakness of mammographic imaging. The superimposition of different tissues that cross the path of a single X-ray beam causes them to be projected onto a single point on the detector (Figure 2.5). This superimposition of structures introduces uncertainty into the acquired X-ray images. In a projection image it becomes impossible to

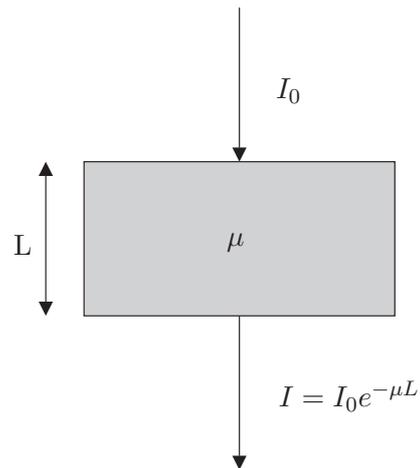


Figure 2.3: Attenuation of X-rays when passing through an object of thickness L and linear attenuation coefficient μ (here, as a simplification, for a constant μ and monochromatic X-rays)

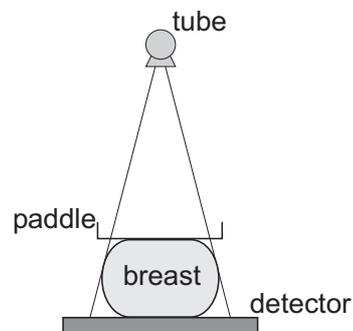


Figure 2.4: Principle of a mammographic acquisition. X-rays emitted by a tube (X-ray source) are attenuated by the breast volume before reaching the detector plane. The breast is compressed between the detector and a compression paddle.

distinguish between actual local high density structures and superimpositions of several lower density structures located at different heights in the compressed breast.

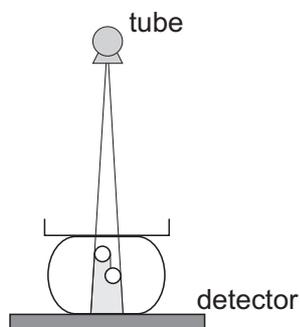


Figure 2.5: For a given tube position, two overlapping structures of breast tissue cannot be distinguished in the detector plane.

Mammography Examinations

Breast cancers take years to develop. Early in the disease, most breast cancers cause no symptoms. When a breast cancer is detected at a localized stage (it has not spread to the lymph nodes), the 5-year survival rate is 98%. If the cancer has spread to nearby lymph nodes (regional disease), the rate drops to 81%. If the cancer has spread (metastasized) to distant organs such as the lungs, bone marrow, liver, or brain, the 5-year survival rate is 26% (ACS 07).

A screening mammogram is an X-ray exam of the breast in a woman who has no symptoms. The goal of a screening mammogram is to find cancer when it is still too small to be felt by a woman or her doctor. Finding small breast cancers early by a screening mammogram greatly improves a woman's chance for successful treatment.

A screening mammogram usually takes two X-ray acquisitions of each breast (Figure 2.6). For some patients, more mammograms may be needed to include as much breast tissue as possible.

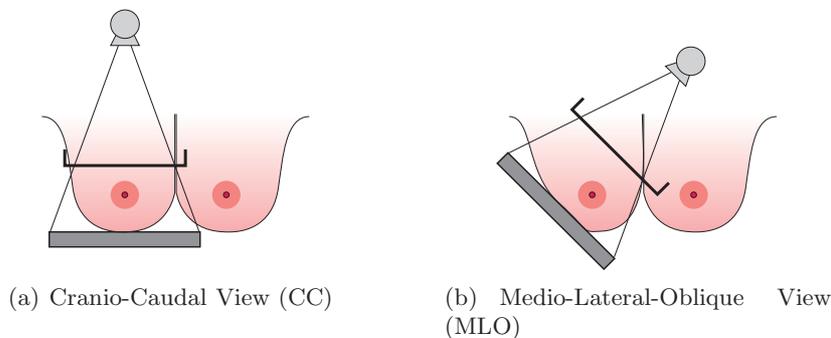


Figure 2.6: Mammographic views: examples of the two most commonly used views in mammographic screening.

A diagnostic mammogram is an X-ray exam of the breast in a woman who either has a breast complaint or has had an abnormality found during a screening mammogram. During a diagnostic mammography exam, more images are acquired to carefully study the breast condition. In most cases, special images involve magnification to bring more details on a small area of suspicious breast tissue making it easier to evaluate. Many other types of X-ray images can be obtained, depending on the type of problem and its location in the breast. These X-rays are tailored to the patient's needs.

Radiological Findings

In Section 2.1.2 we have introduced the main breast pathologies. We will now take a look at the signs of cancer visible in a mammogram. In general we distinguish three groups of cancer-related radiological findings:

- masses,
- microcalcifications,
- architectural distortions.

Areas of high density breast tissue are called *masses*. Different kinds can be observed in mammograms (circumscribed, spiculated, microlobulated, (D'Orsi 98)). In order to be able to

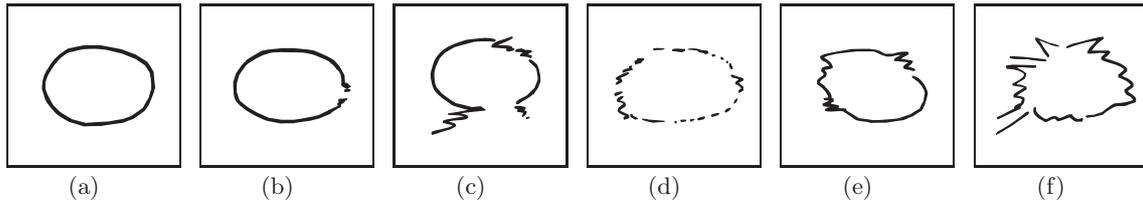


Figure 2.7: Studying the contours for different kinds of breast densities: (a) clear and well defined, (b) partial loss of parts of the contour, (c) regions of partial fuzziness and irregularity, (d) fuzzy, (e) irregular, (f) spiculated.

characterize a mass radiologists generally rely on its contour (Figure 2.7). Usually circumscribed masses are related to benign lesions while spiculated masses are related to malignant lesions. However, since a mass may be superimposed on other breast tissues, its margins are not always easily identified.

Another early indicator of a possible cancer consists of microcalcifications. These small calcium particles appear in many shapes and constellations (Figure 2.8). When assessing microcalcifications found in a mammogram, a number of characteristics need to be taken into account. Their size, shape, number and spatial distribution all bear information as to the potential breast lesion they may indicate.

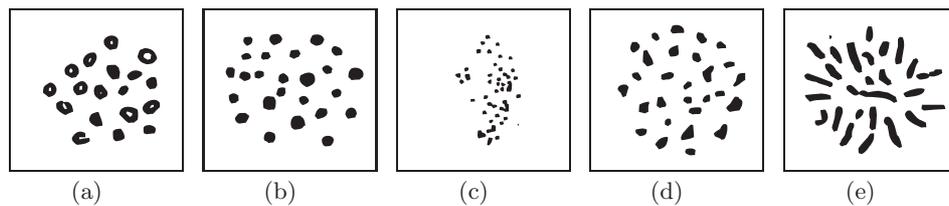


Figure 2.8: Morphology of microcalcifications with the corresponding classification as described in (Le Gal 84): (a) annular, (b) regularly punctiform, (c) too fine for precizing the shape, (d) irregularly punctiform, (e) vermicular.

Finally architectural distortions refer to a change in breast tissue while no visible change in density may be present. They are often characterized by converging, star-like, tissues. The breast tissues at that point seem to be *drawn in* to a central point. Regarding this aspect they are similar to spiculated breast masses, only that they are missing the dense breast tissue in the center of the lesion.

Figure 2.9 shows clinical examples for some of the radiological findings mentioned above. A detailed description of the different radiological findings along with a standardized rating system (benign - malignant) can be found in (D'Orsi 98).

2.2.2 Digital Mammography

Progress in electronic semiconductors over the last decades has allowed to produce large digital detectors that are covered with photo-diodes sensitive to photons in the visible spectrum and coated with needle structured scintillator material that converts X-ray photons into visible light. This technology led to the development of full field digital mammography (FFDM) systems (Muller 99). The first full field digital mammography system was approved by the U.S. Food and Drug Administration (FDA) in January 2000.

Here, the photons emitted by the X-ray source are captured by a digital detector composed

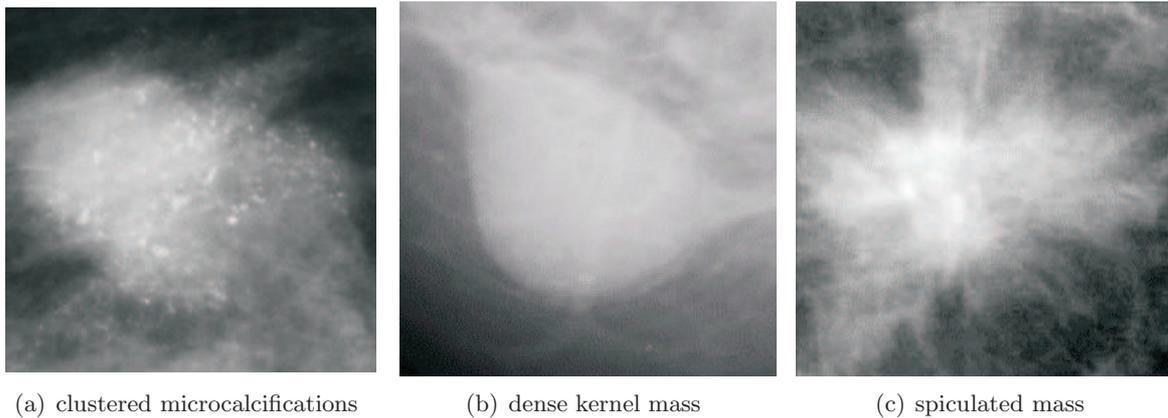


Figure 2.9: Different types of breast lesions.

of a scintillator and a matrix of photo diodes that cover an area equivalent to that of a standard mammographic film. Today, mammography detector sizes reach up to 24*32 cm and contain over 8 million detector elements (del) for a $100\mu m$ del size. Typical pixel sizes vary from 50 to $100\mu m$. Different studies have been performed to evaluate the impact of the pixel size on image quality and lesion detectability (Levy 00; Chan 01a), showing no clinical performance degradation using $100\mu m$ del size compared to film-screen technology.

Since their introduction digital mammography systems have become increasingly popular. About 16 % mammography systems sold in the United States in 2006 were digital systems. Recent studies have shown that performance of digital mammography is comparable to film screen mammography in terms of lesion detectability and hence clinical benefit (Pisano 05). FFDM systems have even been found to be superior to film screen mammography in some cases, such as for dense breasts. This is of particular interest, since dense breasts represent one of the major challenges in early cancer detection today.

Furthermore, FFDM provides advantages in terms of dose and image post-processing. The patient dose in an FFDM acquisition is generally lower than the dose used in film screen acquisitions to deliver images with the same image quality. In addition, the image is instantly available on a computer screen. No film development is necessary. Having the image available in electronic form is especially interesting for image processing tasks. CAD can then be applied directly after image acquisition. No more time is lost for development or film digitization improving the overall clinical workflow. Furthermore, a number of processing applications are made possible, that cannot be realized on digitized film. A large image dynamic allows for example, the contrast enhancement of image areas where the corresponding information on a film screen image is already lost due to under- or over-exposition.

2.2.3 Digital Breast Tomosynthesis

Digital Breast Tomosynthesis (DBT) is a new three-dimensional (3D) limited-angle tomography breast imaging technique that has the potential to substantially overcome the superimposition problem for lesion detection (Dobbins III 03; Wu 03b; Suryanarayanan 00; Niklason 97). Several X-ray projection images of the breast under compression are acquired from different X-ray source positions (Niklason 97; Wu 03b; Dobbins III 03; Ren 05) as shown in Figure 2.10. A tomographic reconstruction is then performed to produce a 3D data set of the breast tissue.

Typically 11 to 45 images are acquired evenly spaced over an angular range of ± 15 to ± 45 degrees. The total patient dose in DBT is comparable to current two-view (CC, MLO)

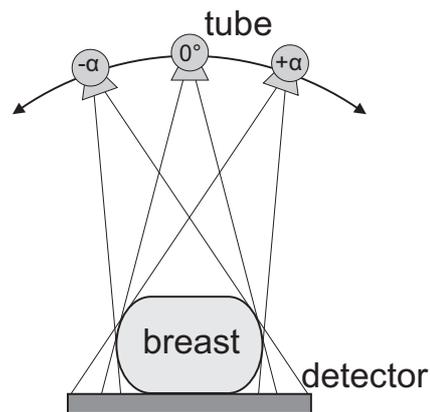


Figure 2.10: Principle of Digital Breast Tomosynthesis acquisition: Several X-ray projection images are acquired from different X-ray source positions.

mammography, but is distributed over many projections. The dose per acquired image is thus significantly reduced in comparison to standard 2D mammograms, to maintain a comparable total patient dose per scan. This has a major impact on the processing of the projections, as the characteristics of these images change dramatically, and algorithms developed for 2D mammograms cannot be generally applied without modification. Examples of the dynamic of pixel values in 2D digital mammography and DBT projected views are given in Figure 2.11 and Figure 2.12.

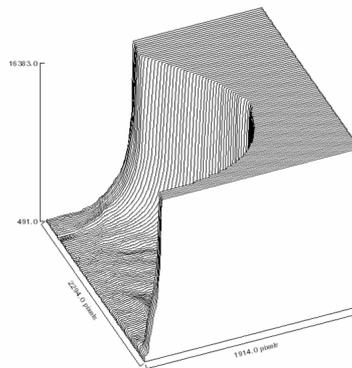


Figure 2.11: Image dynamic for a 2D standard digital mammogram (LMLO): gray-levels of the image are plotted over the image plane. The digital detector saturates in the image background. This facilitates the detection of the breast border.

The main advantage introduced by this new imaging modality lies in its capability to eliminate overlay of tissues and masking effect (Figure 2.13). The acquisition of projection images from different source positions enable to distinguish between two objects that may be indistinguishable from one another for a single given source position. In mammography images are acquired for a compressed breast. The patient's breast is compressed between the detector and a compression paddle (Figure 2.4). If we want to keep the breast under compression to allow a reduced patient dose, the detector needs to remain static and cannot rotate along with the X-ray source as in CT. Working with a detector of limited dimensions leads to a particular acquisition geometry that represents one of the major challenges in DBT.

The angular opening of the path of the X-ray source is limited by several factors (Figure

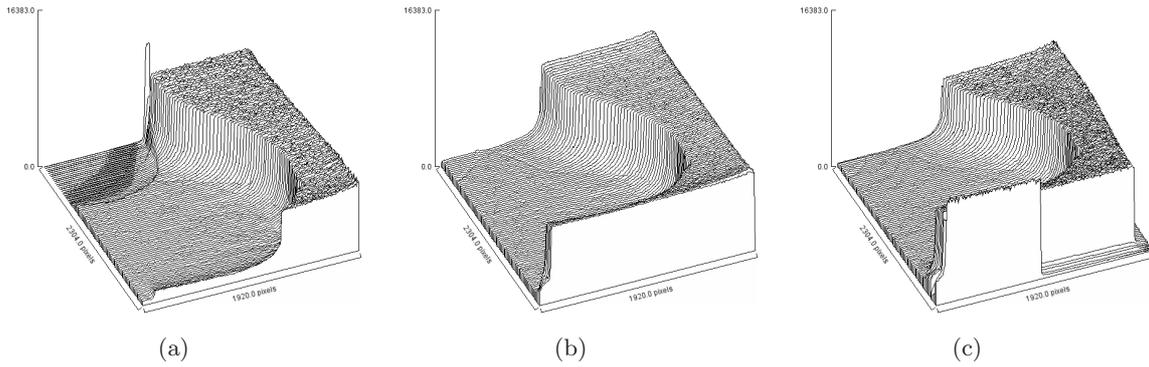


Figure 2.12: Image dynamic for DBT projected views: gray-levels of the images are plotted over the image plane for -30° (a), 0° (b), and 30° (c) projected views. Two particularities of these images can be seen. The digital detector is not saturated in the image background. Also, artifacts appear at the top (a) and the bottom (c) of the projected views. These are caused by the compression paddle entering the field of view (see Figure 2.14). In (a) we can see that the breast is truncated at the bottom, since parts of the breast are projected outside of the visual detector area.

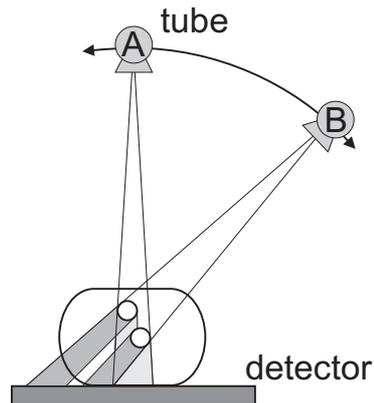


Figure 2.13: Elimination of overlaying and masking effects: two distinct objects that we were not able to distinguish using a single *standard* mammographic projection image (A). Using an angulated view (B), these two objects can now be clearly distinguished in the resulting tomographic projection image.

2.14). For tube positions far from the 0° central position, the compression paddle will enter the field of view of the detector and cause artifacts in the images (Figure 2.15). Furthermore, not the entire breast volume can be imaged on a static detector for all source positions. These constraints have led to limited angular range for DBT acquisitions to $\pm 15 - 45^\circ$, as mentioned above.

This limited angular opening introduces new challenges. The point spread function (PSF) of a tomographic system is mainly determined by two factors: total angular range and number of acquired projection images (Figure 2.16). The maximum resolution we can obtain in z -direction (perpendicular to the detector plane) is a function of the angular range of the acquisition system and the size of the object of interest (Figure 2.17 and Figure 2.18). This is however only a simplified description of the problem of PSF in tomosynthesis. For a more in depth analysis we refer the reader to (Avinash 06; Blessing 06).

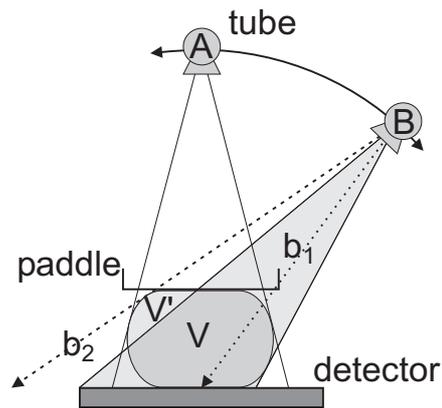


Figure 2.14: Limitation on the angular range for DBT: A projected view taken from a tube position far from the 0° center. Artifacts may occur caused by the compression paddle that enters the field of view (b_1). Furthermore, part of the breast volume cannot be projected onto the detector plane (b_2).

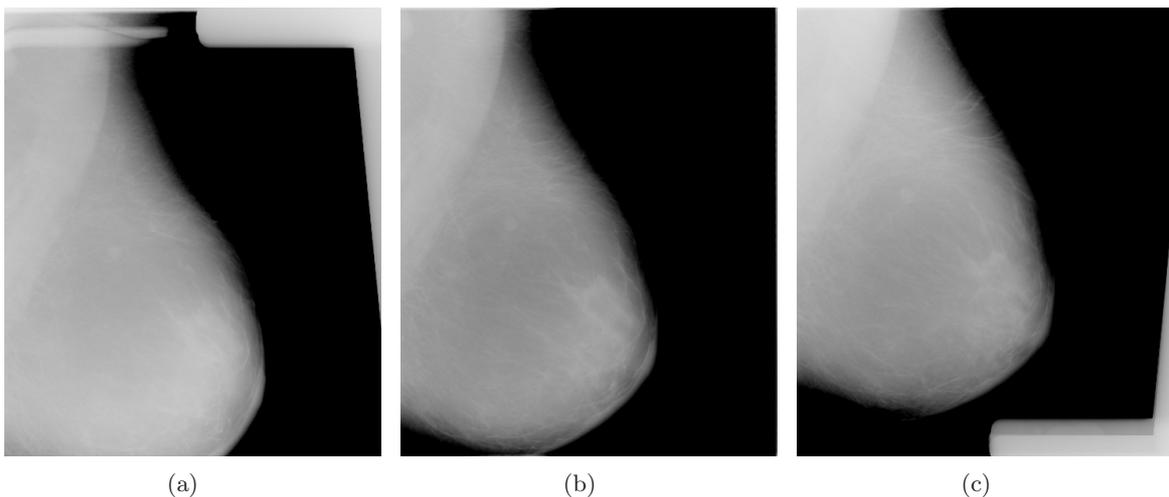
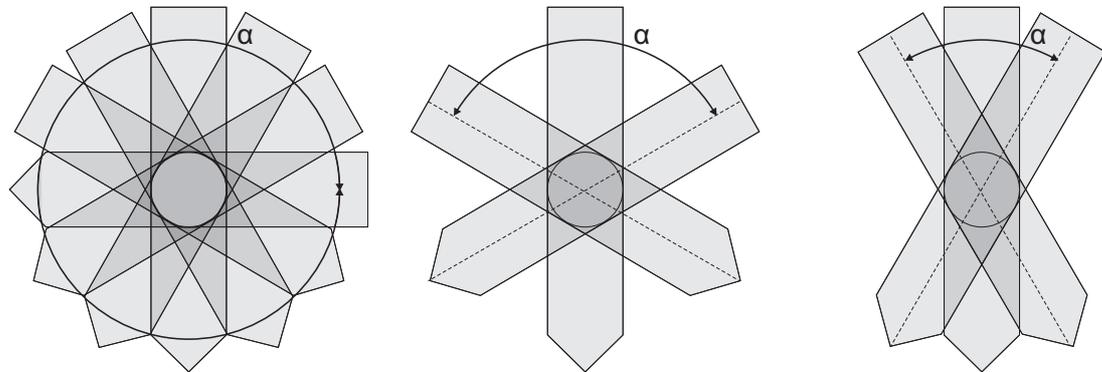


Figure 2.15: DBT projected views: (a) -30° projection, (b) 0° projection, (c) 30° projection. The artifacts in (a) (top right) and (c) (bottom right) are due to the compression paddle entering the field of view.

In DBT several tomographic projection images are acquired for different positions of the X-ray source (Figure 2.19). From this data, a three-dimensional breast volume is reconstructed using a dedicated reconstruction algorithm. However, for an average breast thickness and small objects of interest (microcalcifications), the reconstructed volume may contain up to 100 reconstructed slices. The increase of data volume, when passing from 4 views in 2D standard mammography to over 100 reconstructed slices in DBT, calls for a means to guide radiologists in their decision process. Since the number of images that need to be reviewed by a radiologist for a single DBT case, when compared to standard 2D mammography increases, radiologists may want to speed up lecture of individual images, to keep the time spent per patient constant. In order to do so without increasing the number of missed cancers, tools are needed to support the decision process. The workflow of the review of DBT images is essential for solving this problem, and this workflow should include elements that take CAD-type information into account.



(a) Full α angular range (CT-type geometry) (b) Wide α angular range for tomography (c) Narrow α angular range for tomography

Figure 2.16: Point spread functions for different acquisition geometries. We can see how the angular range of the different geometries affects the z-extension of an imaged object.

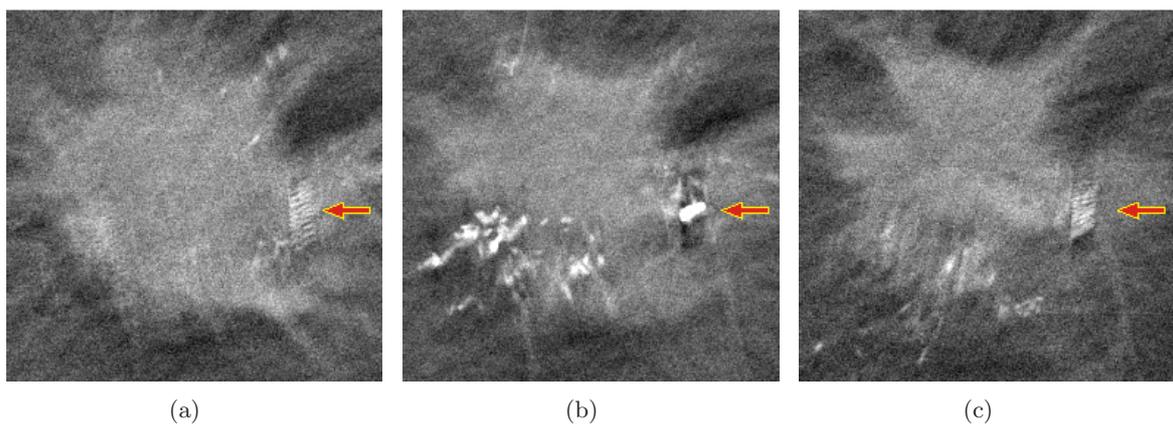


Figure 2.17: DBT reconstructed slices. (b) Slice that crosses the center of the object of interest (indicated by the arrow). (a) and (c) Slices that are located $4mm$ below and above the center slice respectively. The out-of-plane artifacts are visible as low-contrast *shadows* of a high-contrast object.

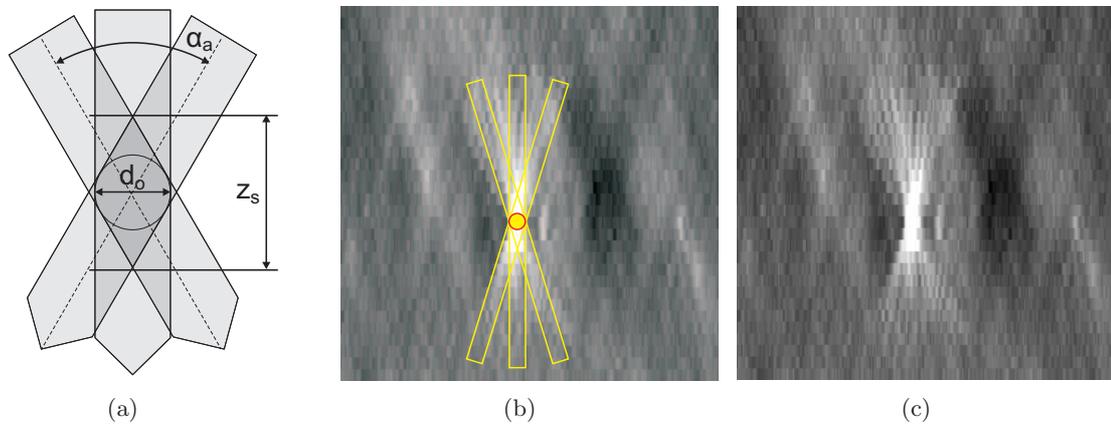


Figure 2.18: The maximal resolution we can obtain in z-direction depends on few parameters: the angular range, the expected size of the objects of interest, the source to image distance (SID), and the source to object distance (SOD). (a) Schematic of the geometry leading to the anisotropy in z-direction. (b) and (c) same phenomenon on the example of the object of interest from Figure 2.17

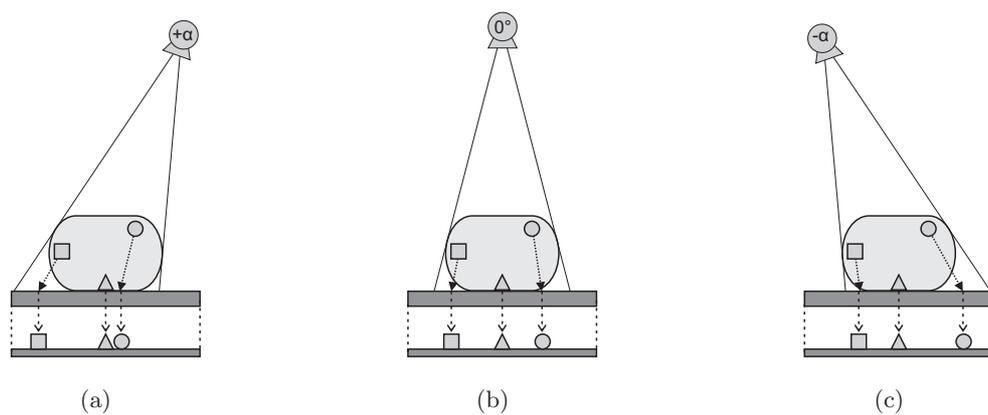


Figure 2.19: Tomographic projection images acquired for three different X-ray tube positions. For each tube position, a given object is projected onto the detector plane depending on its height above the detector plane.

2.2.4 Other Breast Imaging techniques

Breast Ultrasound

Ultrasound, also known as sonography, is an imaging method in which high-frequency sound waves are used to look inside a part of the body. A handheld instrument placed on the skin transmits the sound waves through the breast. Echoes from the sound waves are picked up and translated by a computer into an image that is displayed on a computer screen.

Breast ultrasound is sometimes used to evaluate breast problems that are found during a screening or diagnostic mammogram or on physical exam. Some studies have suggested that ultrasound may be a helpful addition to mammography when screening women with dense breast tissue, but the use of ultrasound instead of mammograms is not recommended in screening exams.

Ultrasound is useful for evaluating some breast masses especially to differentiate cysts from solid lesions. Breast ultrasound is also often used to help doctors guiding a biopsy needle into some breast lesions.

Ultrasound has become a valuable tool to use along with mammograms because it is widely available, non-invasive, and less expensive than other options. However, the effectiveness of an ultrasound test depends on the operators level of skill and experience.

Magnetic Resonance Imaging (MRI)

Breast MRI uses magnets and radio waves, instead of X-rays, to produce very detailed, volumetric images of the body. The most useful MRI exams for breast imaging use a contrast agent that is injected intravenously before or during the exam. This improves the ability of the MRI to clearly show breast tissue details.

MRI is most often used along with mammograms or breast ultrasound to detect breast cancer, particularly in women with very dense breasts. A few recent studies have shown that for younger women at very high risk of breast cancer, MRI screening finds more cancers than standard mammography alone. The MRI studies found many more abnormalities that were not cancers, which led to an increased number of unnecessary biopsy procedures. MRI is also more costly and time-consuming. Therefore, today MRI is not recommended by itself for the early detection of breast cancer.

Just as mammography uses X-ray machines designed especially to image the breasts, breast MRI also requires special equipment. Higher quality images are produced by dedicated breast MRI equipment than by machines designed for head, chest, or abdominal MRI scanning. However, most hospitals and imaging centers do not have dedicated breast MRI equipment available.

2.3 Computer-Aided Detection

In the preceding sections we have mentioned several motivating factors for developing computer-aided detection tools (also known as computer-aided diagnosis or computer-assisted detection). Cancers need to be detected in their earliest possible stage of development in order to maximize the chances for survival. Lesions may be very difficult to detect and contrast between a pathology and the surrounding breast tissues may be very faint so that lesions become hardly visible to the human eye without processing of the acquired image. Studies have shown that CAD has a significant impact in decreasing the number of missed cancers (Freer 01; Karssemeijer 03; Gilbert 06).

With these points as our key motivations we start out with a state of the art of 2D CAD algorithms. Different potential concepts for CAD approaches on DBT data are presented and recent publication for CAD algorithms for this emerging modality are discussed. We conclude with an introduction of the CAD approach that was developed during the course of our work.

2.3.1 CAD for 2D Mammography

Most CAD approaches, regardless if they be aimed at detecting microcalcifications, masses or other structures (Cheng 03; Cheng 06), follow the general scheme illustrated in Figure 2.20.

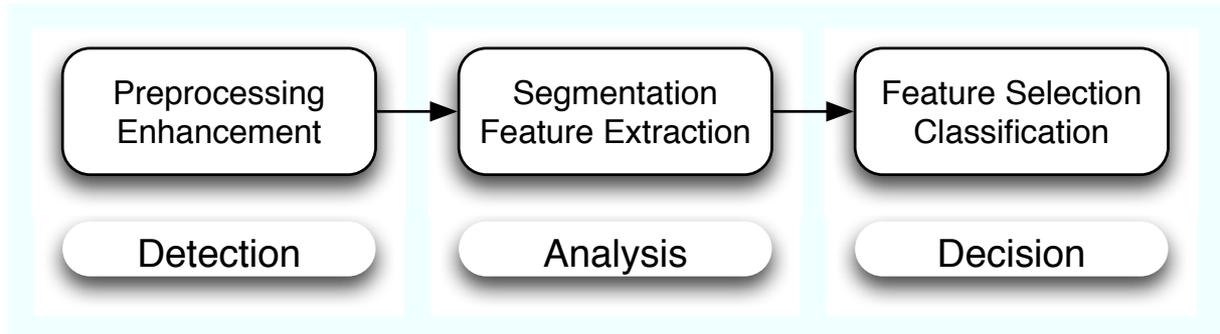


Figure 2.20: There are three basic modules to a CAD scheme: Detection, Analysis and Decision. These modules are generic and used independently of the object of interest.

We use this basic processing chain of computer-aided detection to structure the various methods used in different CAD approaches. Thus we will present the individual modules giving an overview over the most common techniques used for their implementation.

Preprocessing

The first step in CAD for mammography lies in the preprocessing of the images. The underlying principle of preprocessing is to increase the contrast between objects of interest and image background. This type of preprocessing is therefore also often referred to as image enhancement or contrast enhancement. Reasons for the need of preprocessing include the low contrast of mammographic images as well as the variation of intensities of the objects of interest and the underlying breast tissues. It is well known that if a region differs in luminance from its surroundings by less than 2%, it is indistinguishable to the human eye (Dengler 93).

The major problem with contrast enhancement algorithms is that for a given image, some regions may be under-enhanced while at the same time other regions may be over-enhanced. Under-enhancement can cause false negatives (FNs), and over-enhancement can lead to false positives (FPs). An object that is missed during this initial detection step (FN) can generally not be recovered in succeeding processing steps. A high sensitivity (low number of FNs) is therefore of paramount importance at this stage.

A commonly used approach is global histogram modification, where the intensity values of pixels are reassigned in order to make the new distribution of intensities uniform to the utmost extent. This may be achieved by histogram equalization or histogram stretching (Karssemeijer 93; Wilson 97; Gavrielides 02).

However, global histogram modification does very little for texture enhancement and is hence not well suited for mammogram enhancement. An improvement to this problem is provided by local or region based processing approaches. One method is based on nonlinear mapping (Braccialarghe 96). The implementation can be feature-based, including features such as edge

information, local statistic information and higher order correlations (Li 97; Kim 97; Gurcan 97). The local processing methods are quite effective in local texture enhancement.

An alternative to contrast enhancement is the removal of background from foreground. The resulting image is usually obtained by subtracting a low-pass filtered version of the image from the original (Zhou 89). Other global enhancement techniques include convolution with an unsharp masking filter (Chan 87) or an edge detector (e.g. Sobel Operator (Viton 96), Canny Edge Detector (Zheng 06)). These techniques aim at suppressing low frequency information while amplifying high frequency detail. They are therefore especially useful for microcalcification detection.

A number of popular methods for image enhancement are based on wavelet transformation (Mallat 92; Yoshida 96; Chang 97; Chen 97; Strickland 97; Pfisterer 99). The differences of implementing this kind of method lie in the mother wavelets used for the transformations and in the coefficients used. These methods function as a band-pass filter and can therefore be tuned to enhance different kinds of objects of interest (depending on their size and contrast).

Segmentation

According to (Cheng 06) there are four kinds of segmentation used in mammogram segmentation, each kind with a specific nature: classical techniques, bilateral image subtraction, multiscale technique and fuzzy techniques. They all aim to extract objects from background and thus, for our task of computer-aided detection, provide markers that indicate positions of potential breast pathologies, so-called *candidates*.

The classical methods include thresholding (global and local), region growing, edge detection, template matching and stochastic approaches.

Global thresholding has been widely used in image segmentation (Brzakovic 90; Gimenez 99). The thresholding is based on global values such as the histogram of the image. Local thresholding can refine the results obtained with global thresholding (Kallergi 92). In region growing, the basic idea is to grow iteratively starting from a seed pixels, and aggregate with the pixels that have similar properties (Woods 93; Lee 00). If the region stops growing, the final segmentation is obtained.

Edge detection is a traditional method for image segmentation with many different operators at hand: Roberts gradient, Sobel gradient, Prewitt gradient, Laplacian operator (Torre 86), Canny edge detector (Canny 86) etc. A combined method was developed in (Abdel-Mottaleb 96) to increase accuracy of the position of the detected edges. Another strategy for edge detection are active contours. Deformable models can be generally categorized as implicit and explicit models. They were introduced in (Kass 88). The so-called snake employs an energy minimization method to find the contour. The level set approach providing a geometric representation of an active contour was introduced in (Osher 88). For a more detailed description of active contour models, we refer the reader to Section 3.3.1.

Template matching uses prior information on mammograms to segment candidates using prototypes of the objects of interest (Tourassi 03). The prototypes are created based on the physical features of the objects of interest or based on the two-dimensional search function. Sub-regions are matched to the templates using a given criterion. Template matching usually results in a large set of candidates, a majority of which are FPs.

Stochastic and Bayesian methods have provided a general framework to model images and to express a priori knowledge. Markov random field models have been used to deal with the spatial relations between the labels obtained in an iterative segmentation process (Veldkamp 99; Li 01).

Bilateral image subtraction is also called the asymmetry approach, since it is based on

the normal symmetry between the left and the right breast. Left and right breast images are aligned and a subtraction image is computed (Méndez 98; Hadjarian 98). Regions of difference are located in this image.

One multiscale processing is using the multiresolution capability of the wavelet transform to distinguish between different sizes of objects. The image is convolved with a mother wavelet, that needs to be well adapted to the object of interest (Chen 97). For a more in depth description of wavelet-based segmentation techniques, refer to Section 3.2.

Because the contrast in mammograms is very low and the boundaries between objects and background often not well defined, traditional segmentation methods might not work well. Fuzzy set theory (Zadeh 65) has been used in medical image segmentation to model the ambiguities in the image (Kanzaki 92; Smits 94; Bothorel 97). Fuzzy techniques used in segmentation of mammograms include fuzzy thresholding, fuzzy region growing, and fuzzy contour extraction (Sameti 96; Cheng 98; Guliato 98). These are often extensions of traditional segmentation techniques. However, while crisp segmentation divides the image into a set of nonoverlapping regions. in fuzzy set theory, every pixel of the image may belong to several regions with a membership degree between 0 and 1. If a pixel belongs exclusively to one region, its membership degree to all other regions will thus be 0. For a more thorough discussion of fuzzy segmentation techniques, we refer the reader to Section 3.3.

Feature Extraction

After the segmentation of candidates has been performed, the next step consists in extracting features to gain more information about the candidate and its context. The feature space can be divided into four sub-spaces: intensity features, geometric features, texture features and cluster features.

Intensity features, also called spatial features, are the simplest among the four sub-spaces, as most of them represent simple statistics. They include contrast, intensity and noise measures as well as statistics on higher moments (Veldkamp 99; Huo 98). Also, many intensity features are based on histogram values for different regions (e.g. inside the candidate, the candidates contour, a given neighborhood of the candidate etc.). Key values of the histograms such as mean, median and standard deviation may be used as features (Huo 95).

Shape features are also known as morphological or geometric features. A large variety of shape features are used in mammograms (Veldkamp 99; Sahiner 01). They include area, circularity, perimeter, convexity, Feret ratio, and compactness. There are also several features based on the normalized radial length (NRL) or the normalized chord length (NCL) distribution (Kilday 93; Rangayyan 97). NRL is defined as the Euclidean distance of a given point to the centroid of the candidate (average of all points of the contour). NCL is the Euclidean distance of a pair of points on the candidates contour.

The third class of features are texture features. Since microcalcifications tend to be too small to derive useful texture measures inside their contour, texture features are used differently for detecting microcalcifications than for the detection of masses. For masses, features are generally extracted inside a given region (or for the pixels on the contour of that region) and in an ROI containing this region. For microcalcifications, this is performed either for the neighborhood of a single image pixel, or for analyzing the texture features of the entire image. The texture features measured in the two cases are however very similar.

One group of texture features is derived from spatial gray level dependency (SGLD) matrices or gray level co-occurrence matrices (GCM) (Wei 97; Sahiner 01). These matrices are used to measure the texture-context information. The images on which the SGLD matrices are calculated can be ROIs around the object or the entire image. For larger objects and especially

spiculated breast masses, the rubber-band straightening transform (RBST) image of the ROIs may be used for computing the SGLD matrices. This image is constructed by transforming a band of pixels surrounding the mass onto the 2D spatial domain (Dhawan 96; Sahiner 98).

The features based on gray level difference statistics (GLDS) are extracted from the GLDS vector of an image. The GLDS vector is the histogram of the absolute difference of pixel pairs separated by a given displacement vector (Weszka 76; Sahiner 96).

RLS features are based on the run lengths statistics of the image. A gray level run is a set of consecutive collinear pixels with the same gray level value. The run length then corresponds to the number of pixels in a given direction (Galloway 75; Kim 99).

After extracting features for individual microcalcifications, cluster features are used to group them into clusters. These include spatial features, morphology features and cluster description features. Cluster area, the number of microcalcifications and the distance between microcalcifications are the most popular cluster features since they do not only represent the features used by radiologists to characterize microcalcification clusters, but also due to their simplicity and effectiveness (Cheng 03).

Feature Selection

The feature space can be very large and complex due to the wide diversity of normal tissues and the variety of abnormalities. Nevertheless, only some of them are significant. Feature selection is the process of selecting an optimum subset of features from the large number of features available in a given problem domain determined after image segmentation. Feature extraction and selection is a key step in CAD since the performance of a CAD algorithm depends more on the optimization of the feature selection than the classification method (Cheng 06). In addition, relatively few features used in a classifier can keep the classification performance robust. Two major methods for feature selection have been used for CAD in 2D mammography. Stepwise feature selection is a heuristic procedure using statistical techniques based on Fisher's linear discriminant (Fisher 36). At the beginning, the selected feature pool is empty. At each following step, one available feature is either added to or removed from the feature pool by analyzing the features effect on a given selection criterion. Another common method for feature selection is genetic algorithms (GA). These are adapted heuristic search algorithms based on the principles of Darwinian evolution (Holland 92). In particular, GAs work very well on mixed combinatorial problems. However, they tend to be computationally expensive.

Classification

Classifiers play an important role in implementation of computer-aided detection algorithms. Once features of candidates have been extracted, these features (or a selected subset) are then input to a classifier. The classification result delivers a classification into normal breast tissue and breast pathologies. Depending on the general design of the CAD algorithms this may even include a classification into benign and malignant breast lesions. The most popular classifiers employed for this task include: Artificial Neural Networks (ANN), K-Nearest Neighbor classifier (KNN), Bayesian belief network (BBN) classifier, decision trees, kernel-based methods and combined classifiers.

ANNs are the collection of mathematical models that imitate the properties of biological nervous system and the functions of adaptive biological learning (Rumelhart 86; Lau 92). In (Chitre 93) a backpropagation neural network was used to classify microcalcifications by inputting a set of 10 SGLD texture features. The authors found an average of 0.6 under the receiver operating curve (ROC) for 191 *hard-to-diagnosis* cases. Another approach using back-

propagation neural networks incorporates the use of fuzzy logic to extract suspect regions, and then features are extracted from the regions (Verma 01). A comparison of the classification results of a neural network classifier with those of five radiologists was given in (Jiang 97). Experimental results showed that the neural network classifier had better performance than the radiologists in terms of area under the ROC curve (A_Z). ANNs have also been extensively used to classify masses in mammograms (Cheng 94; Floyd Jr 94; Fogel 98).

KNN classifier distinguishes unknown patterns based on their similarity to known samples (Cover 67). A comparison of the performances of ANN and KNN for the classification of microcalcifications was given in (Kramer 99). A set of shape-features and co-occurrence features as input to a KNN was described in (Zadeh 01). A comparison of a KNN, an ANN and human readers for the classification of mammographic masses has been described in (Arbach 03). They found $A_Z = 0.923$ for the BNN and $A_Z = 0.846$ for the expert radiologists. The KNN had a specificity of 85.7 % with a sensitivity of 84.6 %. The authors concluded that it was promising to use the BNN as a physician's assistant.

BBN is an optimal pattern recognition method, which uses a probabilistic approach to determine an optimal segmentation given a specific database (Pearl 88). In (Bankman 94) a Bayesian classifier was used to merge five features of microcalcifications. The algorithm obtained a sensitivity of 100% with 0.22 false-clusters per mammogram on a test of 9 mammograms. Seven classification methods including BNN and different ANNs have been compared on a set of 24 mammograms in (Woods 93). Each mammogram contained at least one biopsy-proven malignant microcalcification cluster. The result showed that Bayesian classifiers outperformed the ANNs. A two-level hierarchical scheme consisting of Bayesian classifiers for each level was used to classify masses in (Viton 96). The first level discriminates the spiculated masses from the non-spiculated masses. The second level separates masses with fuzzy boundaries from those with well defined boundaries. The same genetic algorithm and a common database was used in (Zheng 99) to optimize both Bayesian belief network and neural network. The results indicate that the performance of the two techniques converges towards the same level, indicating that feature extraction may be more crucial for CAD performance than the use of a specific classifier.

A decision tree is an ordered list of test operations on the feature values organized as a tree (Safavian 91). Compared to artificial neural networks the decision tree approach is much simpler and faster (Zheng 01). A test on 322 mammograms resulted in a sensitivity of 97.3 % and at 3.92 false positives per image. After the mammogram was segmented into regions with different gray levels and features, a binary decision tree was used to classify the ROIs into unsuspecting and suspicious mass classes. (Kegelmeyer Jr 94) implemented a computer-aided detection scheme including binary decision tree classification for screening of spiculated lesions. They compared four radiologists screening performance on 85 cases with and without CAD-markers on the images. The computer reports increased the average radiologist sensitivity by 9.7% ($P = 0.005$), moving from 80.6% to 90.3% with no decrease in average specificity. The CAD algorithms alone achieved a sensitivity of 100% with a specificity of 82%.

Fuzzy logic can improve the performance of decision trees (Bothorel 97). Fuzzy decision trees are discussed extensively in Section 5.3.

Support vector machines (SVM) are a group of supervised learning methods that can be applied to classification or regression (Vapnik 00). SVM based classification has been applied to mammogram classification in (El-Naqa 02). (Cao 04) proposes a Vicinal Support Vector Machine (VSVM) approach for mammographic mass detection including two steps. In the first step, abnormal cases are separated from normal ones by use of an SVM classifier. In the second step, the VSVM algorithm is used to decide whether a detected abnormal case is malignant or benign. A similar approach has been proposed by (Sun 04). Based on the discovery that the polynomial SVM is sensitive to microcalcification pixels and the linear SVM is sensitive to non-

microcalcification pixels, (Koutras 06) designed an adaptive threshold mechanism. They are able to exploit the complementary nature of linear and polynomial SVM to reduce false positive detection rate. In (Li 04) regions-of-suspicion (ROS) are diagnosed using SVMs with polynomial and radial basis function kernels. They obtain a recognition accuracy of above 87% for all types of abnormalities present in the breast.

Sequential or parallel combinations of the classifiers are used to improve the classification result. Each classifier may have its own region in feature space where it performs best. Different schemes for combining various classifiers have shown that combination of different classifiers can improve on the classification accuracy. There is however a tradeoff in terms of complexity and computational effort. A more detailed description of different decision making algorithms is discussed in Chapter 5.

Evaluation

It is important to notice that an objective comparison of the performances of different CAD methods is very difficult and even impossible due to the use of different databases. Moreover the proportion of subtle cases versus obvious cases also differs from one database to the other.

In addition to the databases, the methods of evaluation will also influence the performance of the computer-aided detection system. A receiver operating characteristic (ROC) curve is a plotting of true positive as a function of false positive. Higher ROC, approaching perfection at the upper left corner, indicates greater discrimination capacity. For evaluating true-positive detection, sometimes not only the existence but also the location of a tumor is required. A method better adapted for this purpose is the free-response operator characteristic (FROC) analysis which is a plot of the operating points showing the tradeoff between the TP fraction versus the average number of FP per image. Today, CAD algorithms are usually evaluated regarding the obtained sensitivity at one to tow FP per case (alternatively the obtained FP per case with a sensitivity superior to 90-95 %). Both FROC and ROC analysis suffer from certain limitations. For instance, they do not address the complexity of the images and it is difficult to establish a transformation between subjective measurements (radiologist's observations) and the objective ROC/FROC curves. The area under the curve is an important criterion for evaluating the diagnostic performance. It is usually referred to as A_Z index.

2.3.2 CAD for DBT

Basic Concepts

We start out by briefly recalling the basic processing steps of a CAD approach as shown in Figure 2.20 and described in Section 2.3.1. First a *detection* of candidate locations is performed. This may include any kind of filtering, smoothing, artifact correction and other preprocessing necessary to identify the objects of interest. Once candidates have been identified, additional information has to be gathered. This is done during the *analysis* step which may include segmentation and feature extraction for different hierarchical elements of the image (e.g. one object, a group of objects, the entire image). Finally this information is used during the *decision* step to classify the candidates. Here, feature selection, classification and hence candidate elimination may be performed.

In CAD for DBT when compared to CAD for standard mammography, the particularity is the three dimensional nature of the data. Some techniques used in 2D CAD may be applied without changes, others need to be modified to take into account the different nature of the data, and others again are not applicable at all. It is important to reflect the impact of the three dimensional nature of the data on CAD for DBT. We have therefore chosen to classify the

different CAD strategies regarding the way they deal with the 3D aspect of the data.

Given the three basic modules (Figure 2.20), and their order that should not be altered, one more component for establishing a complete 3D CAD scheme needs to be added. Working on tomographic data demands a module performing the transformation between 2D and 3D representation of the data. This is realized by performing backprojection of the 2D tomographic data into 3D space. With these four modules at hand, three strategies for CAD on DBT data have been identified. These are illustrated in Figure 2.21.

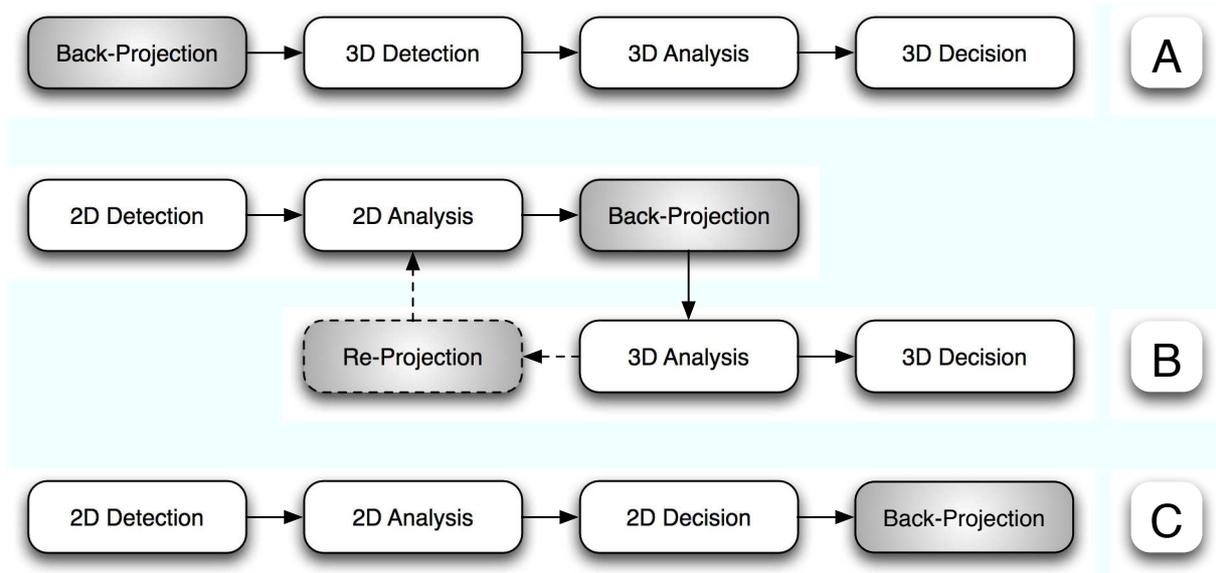


Figure 2.21: CAD strategies: (A) Working directly on 3D data, (B) performing the detection and part of the analysis in 2D, backprojecting the intermediate results and performing further analysis and the decision step in 3D (this may be extended towards an iterative approach by adding a reprojection step), (C) performing the entire processing in 2D and only backprojecting the result into 3D.

The first strategy depicted in Figure 2.21 (A) is working directly on the 3D representation of the data. Usually a dedicated 3D reconstruction algorithm (Wu 04a) is used to generate a 3D volume from the tomographic projection data. The different processing steps are then either performed on slices parallel to the detector plane, or directly on the 3D voxels. The advantage of this method lies in taking advantage of the increased contrast in the reconstructed slices as compared to the tomographic projection images. A major drawback is due to the anisotropy of the reconstructed volume caused by the limited angular range of the acquisition geometry. Furthermore, reconstruction algorithms for this new modality are not yet fully optimized. The properties of the reconstructed volume on which a CAD approach would be based is still susceptible to change during the early stages of development of reconstruction algorithms. Finally, processing on the 3D volume is computationally intensive due to the number of slices being generally more elevated than the number of projected views.

Figure 2.21 (B) shows another strategy for CAD on DBT data. Here, part of the processing is performed on the projection images. Then the result of that processing is backprojected into 3D space. This may be followed by a series of reprojection/backprojection steps, where each reprojection/backprojection step is followed by an analysis step in 2D/3D respectively. After this iterative process, a decision is taken. This is generally done in 3D space. This method combines several advantages. Working directly on the projection images presents a reduced computational effort compared to processing reconstructed slices that are generally much more numerous. Also, working in 2D, the approach is independent of a specific reconstruction technique. At the same

time, combining the results from different projection images allows taking into account the 3D nature of the data. However, when reprojection/backprojection iterations are performed as part of the approach, the computational effort may be high.

Finally, a third CAD strategy alternative for DBT data is depicted in Figure 2.21 (C). Here the entire processing is performed in 2D space, and the results are backprojected into 3D space. The advantage of this method is its low computational effort and the close relationship with standard 2D CAD approaches. It has therefore been used as preliminary study for CAD on DBT data (Chan 04). A strong limitation of this approach is however that it considers each projection image separately, and thus completely ignores the 3D nature of the data.

State of the Art

As DBT systems become available for clinical testing, different strategies for Computer-aided Detection (CAD) on DBT data are emerging. Here, we will present some recent works and classify them according to the 3D CAD approaches introduced in Figure 2.21.

Chan et al. have presented an approach applying automated mass detection for DBT on reconstructed slices (Chan 05). The approach is working entirely in 3D space, and thus corresponds to a type *A* approach (Figure 2.21). Working on a volume reconstructed with an iterative maximum likelihood algorithm (Wu 03a), they start out with candidate detection using 3D gradient field analysis. From these seed points, objects are segmented using region growing that is guided by the radial gradient magnitude. About twenty features were extracted of which seven were selected by stepwise feature selection. The most often selected features included object contrast, minimal gray level, volume change before and after 3D morphological opening, maximal perimeter, compactness and two run-length statistics texture features. A linear discriminant analysis classifier was then used for false-positive object reduction. On a database of 26 patients (15 biopsy-proven malignancies), a sensitivity of 85% with 2.2 false-positives per case was obtained using FROC analysis.

Another representative of a type *A* approach, but for calcification detection, has been presented in (Bernard 06). The breast volume has been reconstructed using simple backprojection reconstruction, in order to be able to easily compute CNR in the slices as a function of CNR in the projected views. Candidates are obtained by convolution of the images with Mexican hat wavelets. This is implemented as a multi-scale approach to account for the variation in size of microcalcifications. A segmentation is reached by thresholding the wavelet filter response relative to local noise level.

A method applying mass detection algorithms directly on the projected views (Type *B* approach) was presented in (Reiser 06). First, the projected views are median filtered and subsampled by a factor four. Mass detection in the individual projection images is based on a radial-gradient index (RGI) feature. Lesion candidates are obtained by thresholding the RGI feature image. Candidates positions are backprojected into 3D space. There, intersecting rays are evaluated and grouped with nearby 3D candidates. These candidates are weighted by a so-called candidate ray index (CRI) to account for the angular range over which a given candidate has been detected. After thresholding of the 3D feature volume, candidates are reprojected onto the detector plane for segmentation and feature extraction. Lesion boundaries are determined using RGI segmentation. For each candidate, 11 feature sets are obtained, for each of the projection images. Classification is achieved through the use of a linear discriminant analysis classifier. A sensitivity of 95% at 15 FP per case has been computed using FROC analysis.

A type *B* approach for microcalcification detection has been presented by Wheeler *et al* (Wheeler 06). High-pass filtering of the projected views leads to so-called *calcification residual* images. Through estimation of electronic and quantum noise, calcification signals are isolated

and combined through simple backprojection reconstruction. The obtained feature volume is then thresholded and serves to enhance the calcium signals in the 3D reconstructed volume.

An earlier work by Chan *et al* (Chan 04) is working exclusively in the 2D domain and can thus be categorized as a type *C* approach. In this preliminary study, a standard 2D CAD algorithm for mass detection is applied to tomographic projected views. The markers computed by the CAD in 2D are then backprojected to find marker positions in 3D space.

The Proposed Approach

For CAD, it is expected that the choice of the reconstruction algorithm, as well as its specific implementation, will affect algorithm performances. As DBT is an emerging technology, reconstruction algorithms for this modality are still not fully optimized. An implementation working on the projected views is straightforward because the unreconstructed image data is directly available and because DBT projection images are similar to conventional mammograms with lower SNR. Furthermore, clinical data for this modality still being very sparse, the development of new, three-dimensional features and especially their evaluation is difficult. It seems advised to start by using well established features from current 2D CAD approaches, and to adapt them for the use in the tomographic projection images. Also, working directly on the projection images, in combination with a dedicated aggregation operator, allows benefiting from additional information. After *standard* reconstruction from the projected views, each voxel of the reconstructed breast volume contains a single value. This value is generally the average of the contributions from each projected view. Additional information about the set of values contributing to this average, is lost (max, min, variance, ...). Finally, the number of projected views is generally much smaller than the number of reconstructed slices, which leads to a significantly lower computational complexity when applying CAD processing in the 2D domain.

For these reasons, we have chosen to develop an approach that follows the general strategy *B*. It involves processing in both the 2D and the 3D domains. This approach benefits from the advantages provided by the tomographic nature of the data. At same time it allows an implementation that stays independent of the DBT reconstruction algorithm. A schematic representation of the proposed CAD approach is given in Figure 2.22.

From the raw projected views we detect and mark candidates that correspond to potential breast lesions. These markers serve as initialization for a fuzzy segmentation approach where a set of candidate contours is obtained. We extract a number of features for each candidate contour that serve to compute fuzzy attributes characterizing the marked object. The combination of a partial defuzzification and a back-projection/re-projection step serve to establish the link between corresponding objects in different projection images. Based on this information we are able to aggregate the fuzzy attributes extracted in the different projected views for a given three-dimensional particle. A fuzzy decision tree classifier then takes a decision regarding each marked particle based on the entire information extracted by the different modules. The individual processing modules are described in detail in Chapters 3-5.

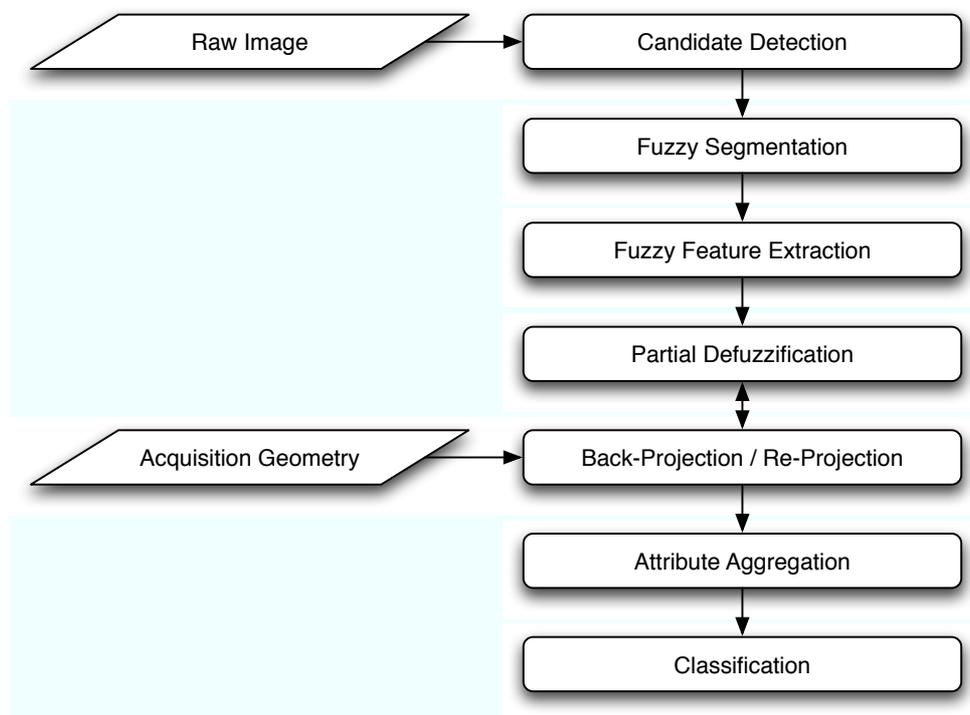


Figure 2.22: Algorithm scheme: order of the different processing modules in the proposed CAD concept.

Neurosis is the inability to tolerate ambiguity.

Sigmund Freud

3

Extracting Image Information

3.1 Introduction

In this chapter we present the first part of the processing chain for computer-aided detection. It concerns the extraction of all relevant information from the image data. First, structures corresponding to radiological signs of potential breast pathologies are detected and marked. Based on those markers a segmentation is performed to extract the boundaries of the structures leading to candidate particles. Finally, a set of features is extracted from the candidate particles, allowing us to characterize the structures. These modules are of utmost importance since all successive processing steps depend on their results.

A structure that is missed during the detection step cannot be recovered later on and will therefore be missed altogether. Sensitivity is hence attributed the highest priority during the detection of structures of interest. We accept to trade this high sensitivity for an elevated number of false alarms, which we aim to reduce during succeeding processing steps.

An important part of the information extracted for a given candidate particle is based on the boundaries of the object. A large variety of features used to characterize a candidate particle are hence based on boundary information. This concerns especially shape features, but also intensity features, that need to distinguish objects from background. Even a number of texture features rely on accurate boundary information in order to produce meaningful results. A robust and accurate localization of the object boundary is therefore of great importance.

The last section of this chapter discusses feature extraction. The majority of these features are based on boundary information obtained during the segmentation step. While it is very important to guarantee a high accuracy for the segmentation results, we need to assure at the same time that the feature extraction is robust with respect to minor variations in the contour extraction results.

In the following sections, we will give a brief overview of the state of the art in detection, segmentation and feature extraction and discuss in some detail the approaches we have chosen for extracting the relevant information about the different kinds of radiological findings.

3.2 Markers

In this section we present a type of preprocessing used to mark structures of interest in DBT data sets. We strive to localize structures of interest in order to provide markers that will serve as initialization to the following segmentation step. The underlying principle of this type of preprocessing is to enlarge the contrast between objects of interest and image background. This can be achieved by a number of different techniques. Among the most popular ones are histogram based methods and thresholding approaches (see Section 2.3.1). Another family of methods makes use of mathematical morphology operators to extract objects of predefined characteristics (Grimaud 91; Zhao 92; Vachier 95; Betal 97).

For our application, different considerations need to be taken into account. Firstly, as we have mentioned before, the DBT projected views exhibit a very low CNR. Secondly, the sizes of the objects of interest vary. When trying to detect microcalcifications we need to take into account the range of possible sizes (and shapes) we expect for this type of radiological finding. Furthermore, we are aiming at developing a single framework for detecting different kinds of breast anomalies. This introduces an additional variability in terms of properties of the objects of interest.

We have therefore chosen to perform the detection task by use of multi-scale wavelet filter-banks. For each projection, the multiscale wavelet filter responses are thresholded according to a model that includes information about noise. The results are then combined to obtain microcalcification and mass candidates. An important advantage of using wavelets compared to morphological filters like the top-hat operator (Rick 99) for lesion detection on X-ray mammograms is that we can quite easily predict the impact of linear filters on noise level.

Different banks of multiscale wavelets are used that are adapted to the pattern of different types of lesions. For microcalcification detection, we use Mexican hat wavelets that are second derivatives of Gaussians. For detection of masses with dense kernels, we apply modified Mexican hat wavelets on sub sampled images, benefiting from the large size of these patterns. For detection of stellate masses, directional filters are used. A multiscale approach allows us to better fit the wavelet to the pattern that may vary in a defined range of sizes.

3.2.1 Candidate Calcification Detection

Microcalcifications are small deposits of calcium that are expected to be more radio-opaque than their local environment. Their diameter ranges typically from 100 μm to 1 mm with a predominantly compact shape. The detection is performed on a tomographic projection image of 1800×2304 pixels, where each pixel corresponds to a square detector element with an edge length of 100 μm .

In the method described below, we suppose that the gray level observed at each pixel (i, j) incorporates the attenuation of a spherical shaped microcalcification of known radius r_c (the source to pixel X-ray path is assumed to cross the center of the sphere). Then we check whether the contrast measured in the image at (i, j) is sufficient for this point to be considered a calcium pixel given the sphere radius and noise level.

Since the projection of this sphere corresponds to a disk of radius r_c (neglecting the magnification factor), we compute the contrast using a Mexican hat wavelet with scale parameter equal to r_c . Therefore, the negative coefficients of the wavelet correspond to the area of this disk while the positive coefficients are mapping the background that is supposed to be uniform. A two-dimensional representation of a Mexican hat wavelet is given in Figure 3.1(a). By convolving an image with this filter, we obtain for each pixel position (i, j) the difference between the weighted mean gray level computed in the disk around (i, j) and the weighted mean value in

its neighborhood (Figure 3.1(b)). This wavelet provides a linear computation of local contrast for a given shape and size of objects of interest (Netsch 99). The wavelet's equation is given in polar coordinates by:

$$f(r) = -\left(1 - \frac{r^2}{r_c^2}\right)e^{-\frac{r^2}{r_c^2}} \quad (3.1)$$

If the measured contrast is not sufficient to be considered as originating from the attenuation of a microcalcification of size r_c or if the probability for this contrast to originate from quantum noise is too high, then pixel (i, j) is not considered a calcium pixel, using the following procedure (Bernard 06; Bernard 07).

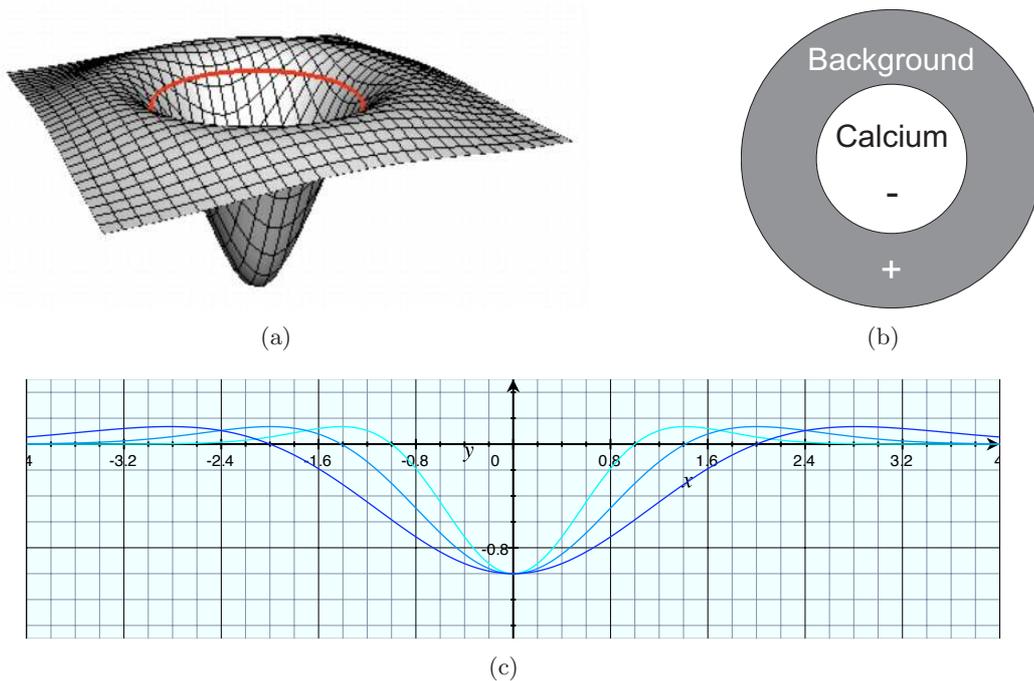


Figure 3.1: Mexican hat wavelet: (a) 2D graph, (b) front view and (c) 1D graph for different scale parameters r_c of a *Mexican hat* wavelet.

Now, we compute the expected attenuation $a_{r_c}(r)$ of a sphere of calcium of radius r_c in a uniform breast considering a mono-energetic X-ray spectrum characterized by the mean energy of the real spectrum:

$$a_{r_c}(r) = \begin{cases} I_b \cdot e^{-2\Delta\mu\sqrt{r_c^2-r^2}} & \text{if } r \in]-r_c, +r_c[\\ I_b & \text{otherwise} \end{cases} \quad (3.2)$$

where I_b is the background intensity in the microcalcification neighborhood and $\Delta\mu$ is the difference in the attenuation coefficients between breast tissue and microcalcifications.

Using Taylor's series we can derive, in low contrast situations (where $2r_c\Delta\mu \ll 1$), the contrast ΔI of the microcalcification measured with the wavelet:

$$\Delta I \simeq r_c \Delta\mu K I_b \quad (3.3)$$

where K is a constant. $\Delta\mu$ depends on breast composition, microcalcification composition and photon energy. It is difficult to predict $\Delta\mu$ since the exact composition of breast tissue and microcalcifications is unknown. We can nevertheless fix a lower bound $\Delta\mu_{min}$ for calcifications

that will be tuned according to performance assessment of the algorithm. Thus, we consider only those pixels to be calcium pixels that respect the following condition:

$$\Delta I > \alpha r_c I_b \quad \text{with} \quad \alpha = \Delta\mu_{min} K. \quad (3.4)$$

The background intensity I_b is measured in the image with the positive part of the wavelet coefficients (Figure 3.1).

Nevertheless, the condition expressed in Equation 3.4 is required but is not sufficiently selective. This is especially true in low Contrast to Noise Ratio (CNR) areas that often result from low dose acquisitions. In these regions, the probability for the measured contrast to originate from quantum noise is high. Thus, we need to add a second condition on the CNR in order to avoid marking areas corresponding to noisy background.

To compute the local noise standard deviation σ , we calculate at each pixel (i, j) the 50% fat/50% glandular equivalent breast thickness required to obtain the background gray level I_b using a simulation of the acquisition chain (Grosjean 06). From an input X-ray spectrum controlled by the acquisition parameters (kVp, mAs, anode material, filter material and thickness), we simulate the attenuated spectrum taking into account the different elements composing the image chain. The gray level is derived from the transformation of the attenuated spectrum into gray level performed by the detector (Rick 99). Changing breast thickness with a dichotomy method, the operation is repeated until the simulated gray level matches the measured gray level I_b . Once we found the equivalent breast thickness, the noise standard deviation is derived from the attenuated spectrum adding electronic noise, the effect of Modulation Transfer Function and wavelet filtrations.

Finally a pixel (i, j) is selected as a candidate calcium pixel if for one of the scales in an interval covering the microcalcification size range, the condition expressed in Equation 3.4 and a condition on contrast to noise ratio are respected, that is if :

$$\Delta I > \max(\alpha r_c I_b, \beta \sigma) \quad (3.5)$$

α and β are algorithm parameters that are tuned according to performance assessment on a biopsy truthed database.

3.2.2 Candidate Mass Detection

Masses constitute the second type of findings associated to breast cancer. Mass detection is often difficult because their contours are not always clearly defined and their contrast is often weak. Their diameter varies typically from 0.5 *cm* to 3 *cm*. Most suspicious masses are composed of a kernel (of variable density) possibly followed by spicules converging towards the kernel.

Dense Kernel Detection

Assuming that mass kernels are of spherical shape, we can apply the same process as for microcalcification detection (Equation 3.4), with adapted scale parameters. For mass detection, the second condition on contrast to noise ratio may not be required, since we integrate the signal on large areas reducing the noise level considerably. Figure 3.2(b) shows an example of intensity differences computed using Equation 3.3, where ΔI and I_b are measured using a single scale Mexican hat wavelet. Strong responses are not only obtained in the mass area but also at the border between pectoral muscle and breast, at the transition between glandular and fatty areas, as well as close to the breast border.

These strong responses outside the mass region come from a biased estimation of the background intensity. With Mexican hat wavelets, the background intensity I_b is computed as a

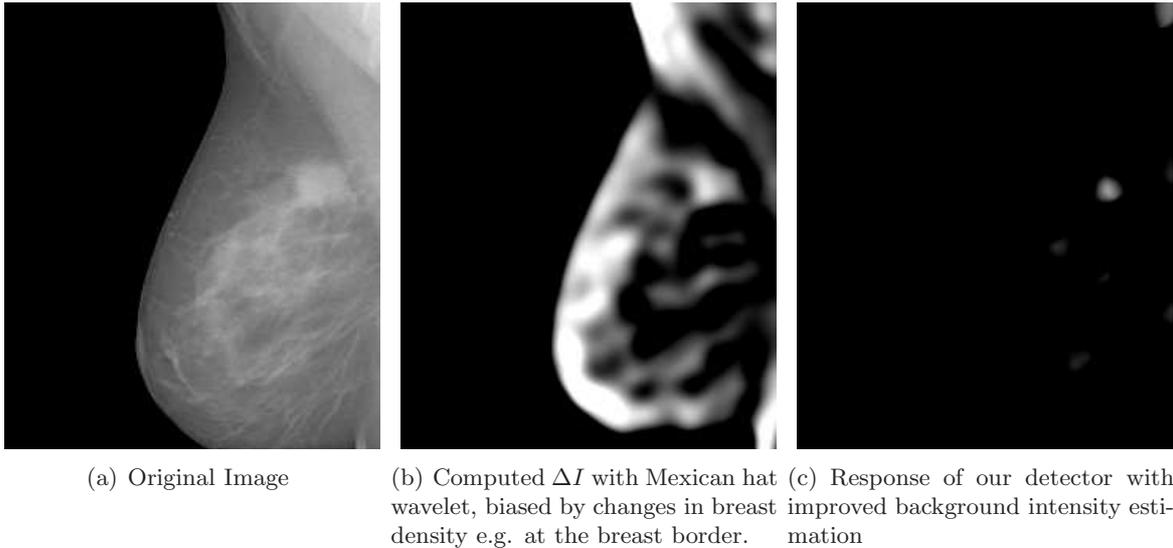


Figure 3.2: Dense kernel detector

weighted mean value around the mass, in a neighborhood proportional to the size of the mass. This generally makes sense for microcalcification detection since their sizes range from $100 \mu m$ to $1 mm$. As a consequence, neighborhood background variations due to breast thickness variations are small even close to the breast border. Masses however exhibit diameters from $0.5 cm$ to $3 cm$. In this case, the area covered by the projection of the object of interest is much larger and breast thickness variations in the mass neighborhood may be significant (see Figure 3.3). Thus, we have to face the problem of breast thickness decrease that leads to a high response of Laplacian of Gaussian close to the breast border. In (Karssemeijer 96; te Brake 99) the authors propose to artificially compensate the image for the breast fall-off before applying the filter. To be efficient, this requires a high precision of the compensation level and does not avoid the strong responses at the glandular/fat and pectoral muscle/breast transitions.

Since estimating background intensity as the weighted average value in the entire neighborhood of the supposed mass leads to an overestimation of the mass contrast, we prefer computing a lower bound of the contrast by:

- estimating the mass intensity I_m as the weighted mean value in the supposed mass area, where weights are decreasing with the distance from the center of the mass,
- considering the most attenuating part in the mass neighborhood as background intensity I_b .

The intensity I_b is obtained using a sliding window on a circular path around the mass kernel and computing the weighted mean value in the window (Figure 3.4), where weights are decreasing with the distance to the center of the window. Then, the smallest mean value among the obtained set is selected as background intensity. This results in a more selective response of the detector (Figure 3.2c) than with conventional filtering (Figure 3.2b).

Once we have estimated the background intensity I_b , we compute from equation (3.3):

$$\Delta I_{min} = r_m \Delta \mu_{min} K I_b \quad (3.6)$$

where K is a constant and r_m is the scale parameter that is equal to the radius of the expected mass. Only pixels for which $\Delta I > \Delta I_{min}$ for at least one scale r_m are considered to be dense

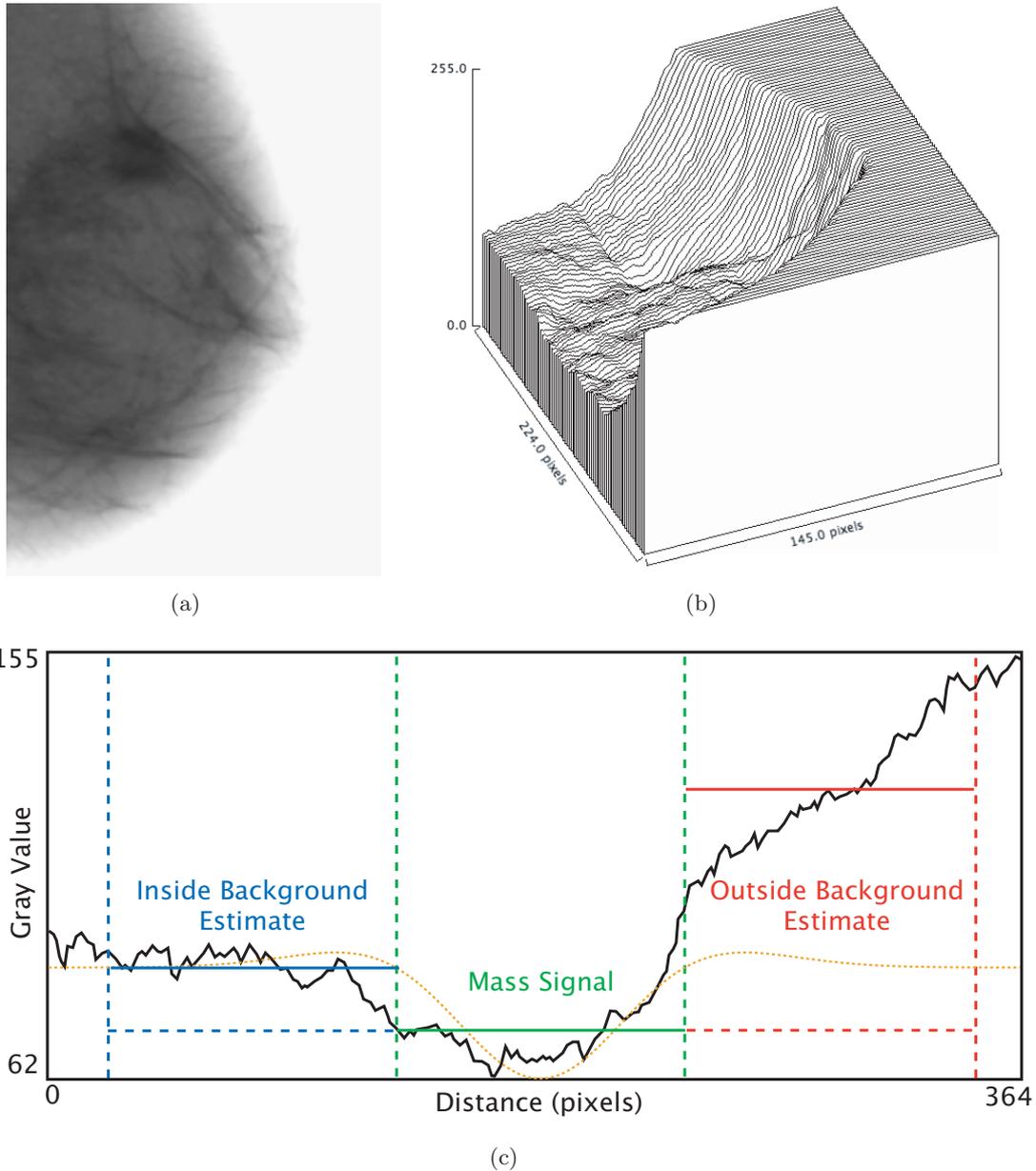


Figure 3.3: Clinical case containing a pathology: (a) Original breast ROI containing a mass, (b) the corresponding surface plot showing the gray level plotted over the image plane, and (c) a 1D gray level profile crossing the mass. The profile of the corresponding wavelet is depicted as dotted line. It is clearly visible, how the decrease in breast thickness towards the breast border impacts the background estimation for that area (outside region). Even in the absence of a pathology, the difference between the mean of the supposed mass and the outside background estimate will be significant. This leads to an elevated number of false alarms along the breast border, and needs to be compensated by the sliding window approach.

kernel pixels. $\Delta\mu_{min}$ is a threshold tuned according to a truthed database. This operation is repeated for a set of different scale values r_m representing the range of mass radii.

This method has however a drawback. If there is an object of equal (or higher) density in the neighborhood of a radiological finding, the contrast between the two objects will not be sufficient for them to be detected. This is especially true for masses that are partly obscured by dense breast tissue. Test on a truthed (2D-) database have shown that the sliding-window

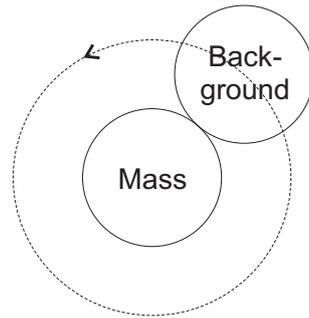


Figure 3.4: Sliding window process illustration: The background intensity is measured for a sliding window on a circular path around the center. The weighted mean value is computed for each position and then the smallest among these values is chosen to represent the background intensity.

technique provides superior performance when compared to computing the background estimate using the mean or the median. Further improvement of this method constitutes an interesting research topic to be investigated.

Stellate Mass Detection

Stellate lesions are composed of spicules converging towards a kernel of variable density. To detect stellate lesions, we detect spicule pixels, estimate their direction and compute an indicator of convergence.

We detect spicules by convolving the raw image with a bank of directional filters having different directions and scales. In (Karssemeijer 96; te Brake 99) the authors convolve the image with only three filter kernels that are second-order directional derivatives of a Gaussian kernel. Their directions are given by $\theta = \frac{\pi}{3}N_d$, with $N_d = 0, 1, 2$. The authors derive the response for an arbitrary direction from the responses of the three filters.

Estimating local direction from the response of only three directional filters may be imprecise. Using more filters can be computationally demanding since a convolution requires $O(N_i N_f)$ operations, N_i being the image size and N_f the filter size. Computing convolutions in the Fourier domain leads to a reduced computation time since it requires only $O(N_i \log(N_i))$ operations, taking advantage of the FFT algorithm. Nevertheless, the algorithm complexity remains considerable for large images. Convolving the image with a β -spline based wavelet (Unser 94; Unser 92) requires only $O(N_i)$ operations using a recursive implementation. These are approximations of Gabor wavelets that are compositions of a bandpass filter in the wave front and a lowpass filter in the orthogonal direction (see Figure 3.5):

$$g_{\omega, \theta, \sigma}(x, y) = \frac{1}{\sqrt{\pi\sigma}} e^{-\frac{u^2+v^2}{2\sigma^2}} e^{-i\omega u} \quad \text{with} \quad \begin{cases} u = x \cos(\theta) + y \sin(\theta) \\ v = -x \sin(\theta) + y \cos(\theta) \\ w = \frac{\pi}{2\sigma} \end{cases} \quad (3.7)$$

Thus β -spline based wavelets are direction and frequency selective as well. Since the computation time is reduced using β -spline based wavelet convolutions, we can trade this gain for an increase in the number of directional filters to obtain a better sampling of the direction domain. After convolving the image with the filter bank, we select the filter of maximal response in each pixel providing local direction. This selection leads to a directionally filtered image that is thresholded against background noise level in order to select only spicule pixels providing reliable directional information. The background noise level is estimated using the acquisition chain simulation as described earlier (Section 3.2.1).

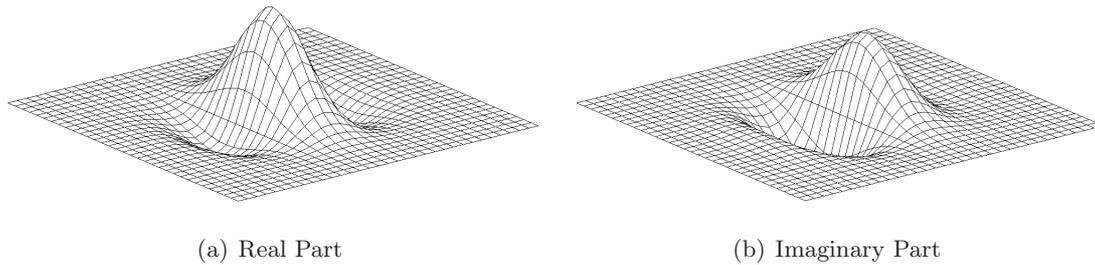


Figure 3.5: Gabor wavelet

Now, we suppose that a pixel (i, j) is the center of a stellate mass and we examine the configuration of the spicule pixel directions in its neighborhood. The neighborhood size is provided by the dense kernel detection process by considering the size of the filter of maximal response. We used the method of (Karssemeijer 96; te Brake 99) to compute an indicator of spicule convergence towards (i, j) . They use two features for detection of stellate patterns of straight lines. The first feature is based on the total number of pixels in a defined neighborhood with directions that point towards a given central area. The second feature is a measure for the angular distribution of these pixels. For each of k equally distributed bins the difference between the detected pixels and the statistical mean is computed.

Then, we decide on the presence of a stellate mass at (i, j) from the magnitude of the convergence indicator. For the detection of stellate masses containing also a dense kernel, the indicator of convergence may be combined with the response of the dense kernel detector using, for example, a learning algorithm or a rule-based system to provide a better decision.

3.3 Segmentation

The preceding detection step provides the position of potential radiological signs that need to be further investigated in order to decide whether they represent actual pathologies or false alarms. To this end we strive to extract the object boundaries in order to be able to characterize the object through feature extraction.

Many segmentation algorithms fail when trying to extract accurate contours for potential lesions in mammographic images. This is mainly due to two properties of these images: firstly, the background in mammograms can be heavily textured (depending on the composition and density of the breast), where the attenuation of the breast tissue tends to coincide with the attenuation of lesions. Secondly, a low patient dose, and consequently a low contrast to noise ratio (CNR), presents an additional challenge when trying to separate objects of interest from the background. This is especially true for DBT data, where the patient dose in individual projected views is significantly reduced as compared to 2D standard mammograms (Figure 3.6).

For any given segmentation task, the segmentation algorithm has to be chosen with respect to the object of interest as well as the properties of the image background. The properties of the main radiological signs in mammographic images exhibit distinct differences. When comparing mass lesions to microcalcifications, the obvious difference is regarding the objects sizes. Mass sizes typically range from 0.5 cm to 3 cm while microcalcifications typically exhibit a diameter between $100\ \mu\text{m}$ and 1 mm . Furthermore, the different composition of these objects leads to differences in terms of texture and contrast. Microcalcifications are typically very small and highly contrasted. A local thresholding approach therefore seems a method well adapted for segmentation of these objects. Masses however tend to be less contrasted and often exhibit partly obscure margins. A thresholding approach is likely to fail under these conditions (see

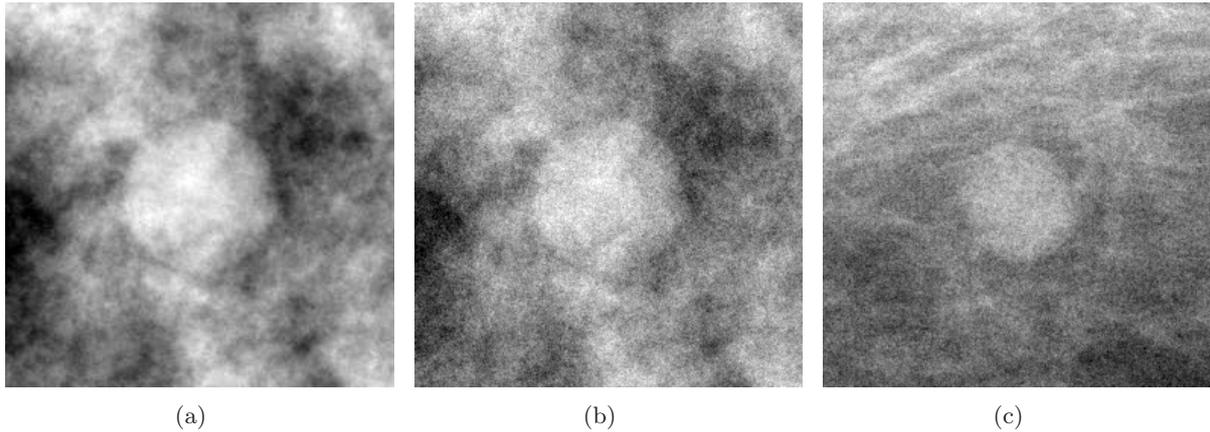


Figure 3.6: Challenges presented by the nature of the image data: (a) a simulated breast lesion for standard dose, (b) the same lesion, breast composition and acquisition technique but simulated for a typical dose of a tomographic projected view and (c) an actual ROI of a clinical tomographic projection. The difference in the background noise level and the contrast between object and background are clearly visible.

Figure 3.7). The need for closed contours and the desire to control the properties of the extracted contour lead us to investigate active contour models for the segmentation of mass lesions.

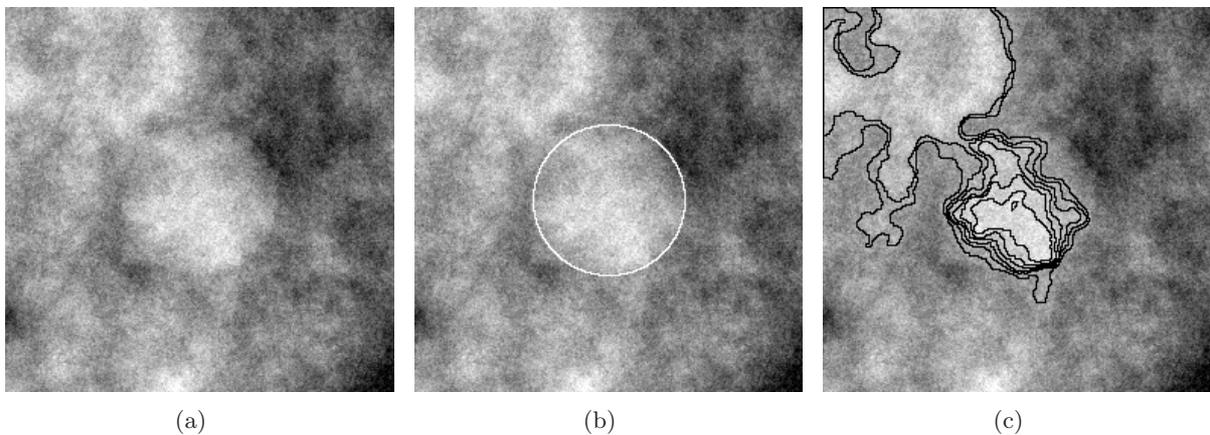


Figure 3.7: Limitations of the multi-thresholding approach: (a) ROI of a simulated circumscribed breast lesion, (b) corresponding contour of the simulated lesion, and (c) the set of contours extracted for this lesion by applying multi-thresholding. Some parts of the lesion are missed, while parts of the background have been included in the larger candidate contours.

The image properties of DBT projected views introduce additional ambiguity to the images and may be addressed using different strategies. One possibility to increase the robustness of the segmentation is to combine the information from several image features in the segmentation instead of taking just a single feature into account. This approach will be discussed in Section 3.3.1. Results are presented in Section 3.3.2. An alternative way to cope with the ambiguity in the image is to model this ambiguity during the segmentation step. This can be achieved by use of the fuzzy set theory, computing a number of contour candidates for each particle. This method will be presented in Sections 3.3.3 and 3.3.4.

3.3.1 Active Contour Models

As mentioned above, most segmentation algorithms exhibit significant difficulties when applied to DBT data. Figure 3.8 illustrates some of the reasons for these difficulties that stem from specific properties of the image data. It can be easily seen that gradient information is only available for part of the object boundaries. This is mainly due to overlying breast tissue and insufficient contrast between object and background. A straightforward solution to this problem consists in trying to connect these contour fragments in order to obtain a closed contour. This is however not an easy task.

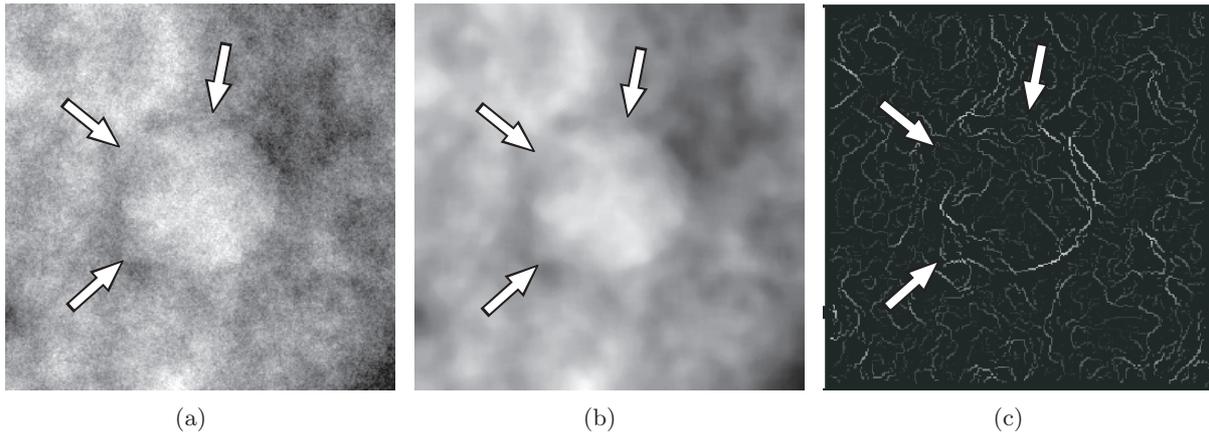


Figure 3.8: Problems in obtaining a closed contour from DBT data: (a) a raw simulated image, (b) a median filtered version of the same image and (c) the corresponding gradient image. The overlying breast tissue prevents the gradient operator from identifying the entire object boundary. Parts of the boundary are thus lost due to an insufficient contrast (indicated by the arrows).

Active contours have been extensively used for finding closed boundaries in medical images (Suri 02). The basic idea of active contours is to evolve a curve, subject to constraints from a given image I_{org} , in order to detect objects in that image. For instance, starting with a curve around the object to be detected, the curve moves towards its interior normal and has to stop on the boundary of the object.

Let Ω be a bounded open subset of \mathbb{R}^2 , with $\partial\Omega$ its boundary. Let $I_{org} : \Omega \rightarrow \mathbb{R}^2$ be a given image, and $C(s) : [0, 1] \rightarrow \mathbb{R}^2$ be a parametrized curve.

The curve evolution is controlled by two energy terms. The first term controls the internal energy of the curve (the elasticity and the rigidity) while the second term takes into account image based energy sources such as gradient information.

Contours produced by these models have several advantages. Most active contour models produce - by definition - closed contours, and their smoothness can be controlled. However, active contour models applied to tomographic projected views do not produce results of the desired accuracy, due to the image properties discussed above. One of the reasons lies in the fact that the majority of these algorithms are based on a single image feature.

A hybrid model seems therefore better adapted, since a method based on several image features will potentially increase the robustness and the accuracy of the segmentation.

In the classical snakes (Kass 88) and active contour models an edge detector is used, depending on the gradient of the image, to stop the curve at the boundary of the desired object.

In curve evolution problems, the level set method and in particular the motion by mean curvature (Osher 88) has been extensively used. The curve C is represented via a Lipschitz

function ϕ , by $C = \{(x, y) | \phi(x, y) = 0\}$, and the evolution of the curve is given by the zero-level curve function ϕ (Figure 3.9). The level-set formulation of geometric active contour models provides some important advantages over classical parametric active contours. It allows for topological changes and the discretization of the problem is solved on a fixed rectangular grid (Figure 3.10).

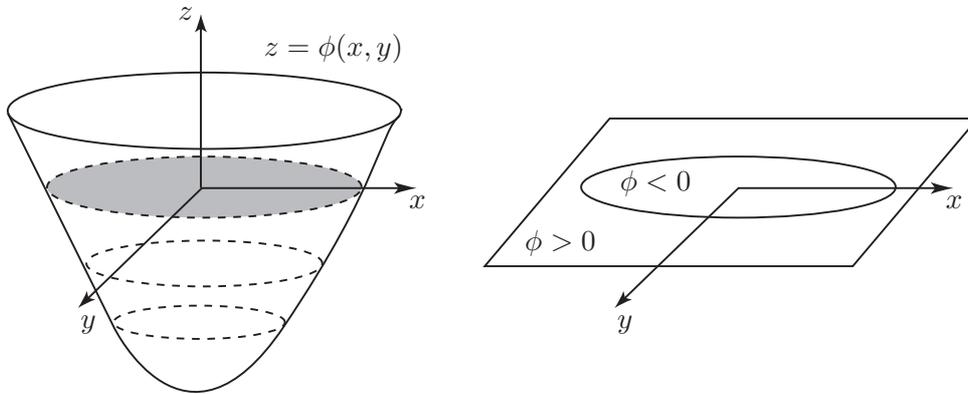


Figure 3.9: The function ϕ : In the geometric active contour models, the curve evolution is modeled by evolving a function ϕ of higher dimension. The 0-level-set of this function is the contour in the 2D image plane.

Another family of active contour models is based on the Mumford-Shah(Mumford 89) energy functional for segmentation:

$$E_{MS}(I, C) = \mu \cdot \text{Length}(C) + \lambda \int_{\Omega} |I_{org}(x, y) - I_{res}(x, y)|^2 dx dy + \int_{\Omega \setminus C} |\nabla I_{res}(x, y)|^2 dx dy. \quad (3.8)$$

The goal is to find an optimal partition of the original image I_{org} . The first term of the right hand side of the functional is the length of the contour and is hence a regularization term. The second term computes the homogeneity inside each segmented region, and the third term is the gradient along the extracted contours. The solution image I_{res} obtained by minimizing Equation 3.8 is therefore formed by smooth regions R_i with sharp boundaries C . Here, μ and λ are parameters of the algorithm. A particular case of this minimal partition problem was proposed by Chan and Vese (Chan 01b). In their work, they partition the image into homogeneous regions of nearly constant intensity values. The level set formulation for their two-phase (two resulting regions) approach is given by:

$$\begin{aligned} E(\bar{c}_1, \bar{c}_2, \phi) &= \mu \int_{\Omega} \delta(\phi(x, y)) |\nabla \phi(x, y)| dx dy \\ &+ \nu \int_{\Omega} H(\phi(x, y)) dx dy \\ &+ \lambda_1 \int_{\Omega} |I_{org}(x, y) - \bar{c}_1|^2 H(\phi(x, y)) dx dy \\ &+ \lambda_2 \int_{\Omega} |I_{org}(x, y) - \bar{c}_2|^2 (1 - H(\phi(x, y))) dx dy, \end{aligned} \quad (3.9)$$

where H is the Heaviside function, δ is the Dirac function, Ω is the image domain, \bar{c}_1 and \bar{c}_2 are the mean gray values inside and outside the contour respectively, and μ, ν (controlling the length of the contour and the area inside), λ_1 and λ_2 (weighting factors for the homogeneity of inside and outside region respectively) are parameters of the algorithm. The first two terms in

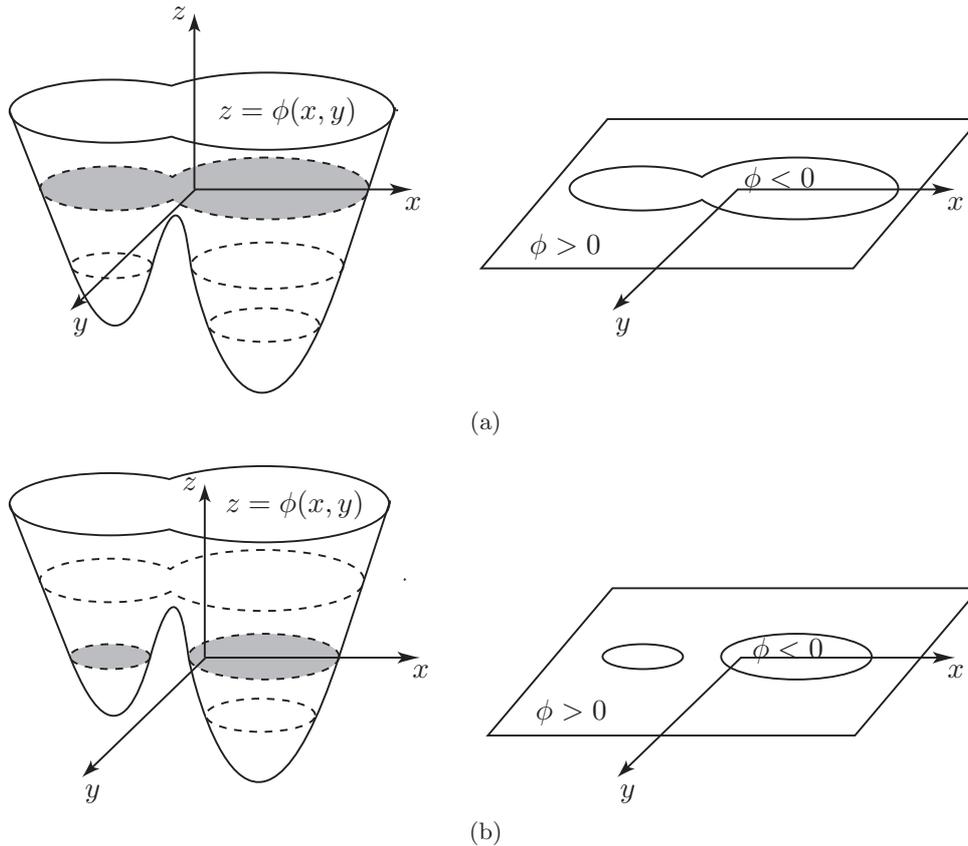


Figure 3.10: Advantages of the level-sets: One of the major advantages of the level-set formulation for geometric active contours is the ability to handle topological changes in the contour. Even though the contour in the 2D image plane may change its topology from a single closed contour (a) to several closed contours (b), this does not necessitate a topology change of the function ϕ and can be modeled in a straightforward way.

Equation 3.9 are designed to control the smoothness of the curve, even though the parameter ν is generally set to 0. The third and the fourth term control the homogeneity of the regions inside and outside of the contour respectively.

A Novel Hybrid Active Contour Model

Let us come back to the main challenges for segmenting digital breast tomosynthesis data. They have been identified to stem from the textured background and the low CNR (Cheng 06). The former renders a region-based approach difficult, while the latter has a significant impact on any edge-based technique. However, region-based techniques tend to cope fairly well with low CNR since they are based on comparing a neighboring outside pixel to a seed pixel with respect to a given criterion. The image may be filtered and smoothed prior to this operation without significant loss in performance. On the other hand, edge-based techniques are more robust regarding textured background. Low CNR may introduce an elevated number of false alarms, while filtering of the original image may lead to the deletion of weak edges. Figure 3.11 illustrates the issues raised by these techniques.

A number of image features may be taken into account when devising the external (image-based) energy term for an active contour model. The most popular features include the image gradient (edge information), pixel intensity (region homogeneity information), texture and color. Here, we have chosen two algorithms, each based on a single image feature as reference.

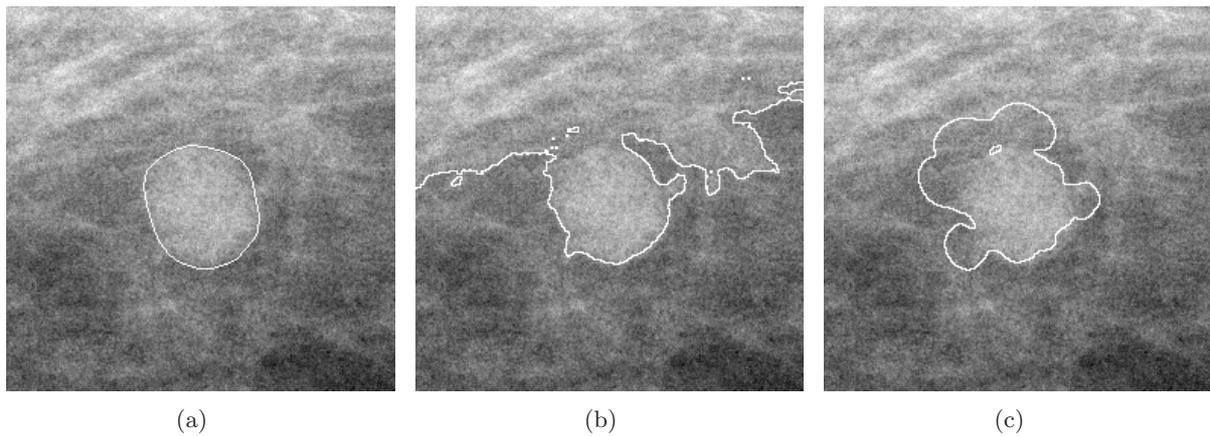


Figure 3.11: Drawbacks of different active contour models for a clinical example: (a) reference contour segmented by an expert, (b) segmentation result of a region-based active contour model and (c) segmentation result of an edge-based active contour model. The edge based model exhibits the classic limitations of this method. In regions of low or missing contour information, the contour bleeds into the background. The region based method tries to minimize the energy over the whole image, and fails due to the textured background.

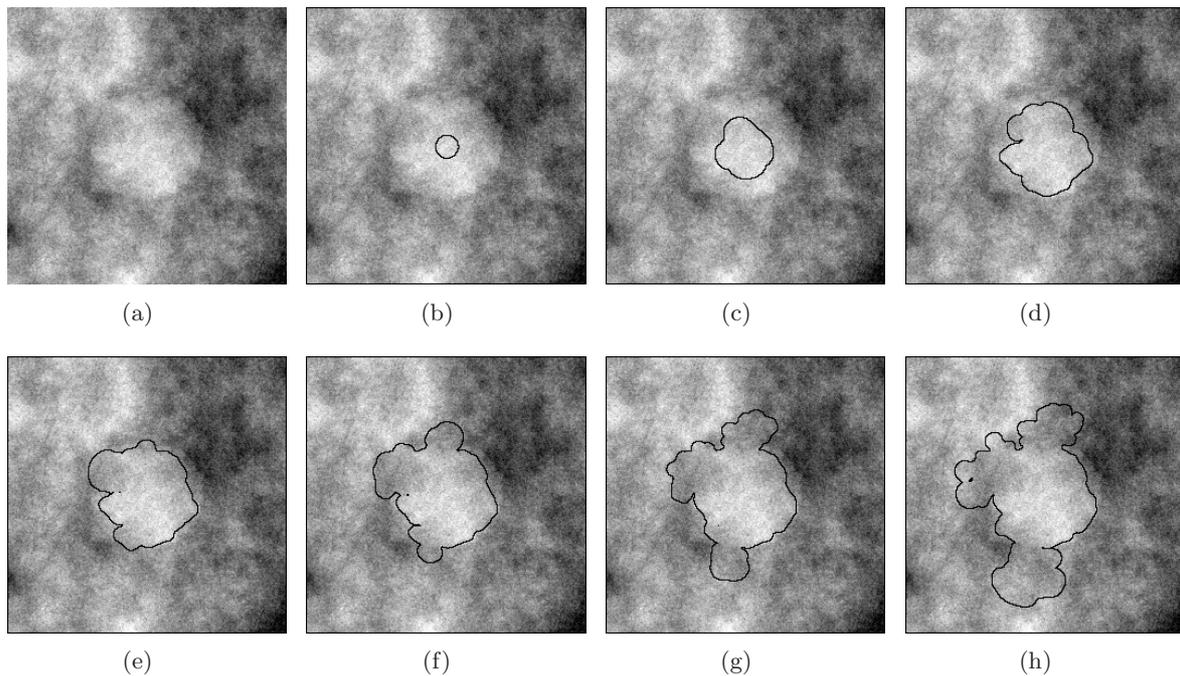


Figure 3.12: Example of a Fast-Marching Level-Set method (Sethian 99) (edge-based) applied on a simulated breast ROI (256×256 pixels): different stages during the iterative process are illustrated from iteration zero (a) to the convergence in iteration 1441 (h). The parameters of the algorithm used for this example were: $\sigma = 1.0$, $\alpha = -0.5$, $\beta = 3.0$, and a weight of curvature versus propagation scaling of 3.9. The parameters α and β define the transformation of the sigmoid image filter applied to intensify the differences between regions of low and high values in the speed image. It is clearly visible how the contour bleeds into the background where the boundary information is weak.

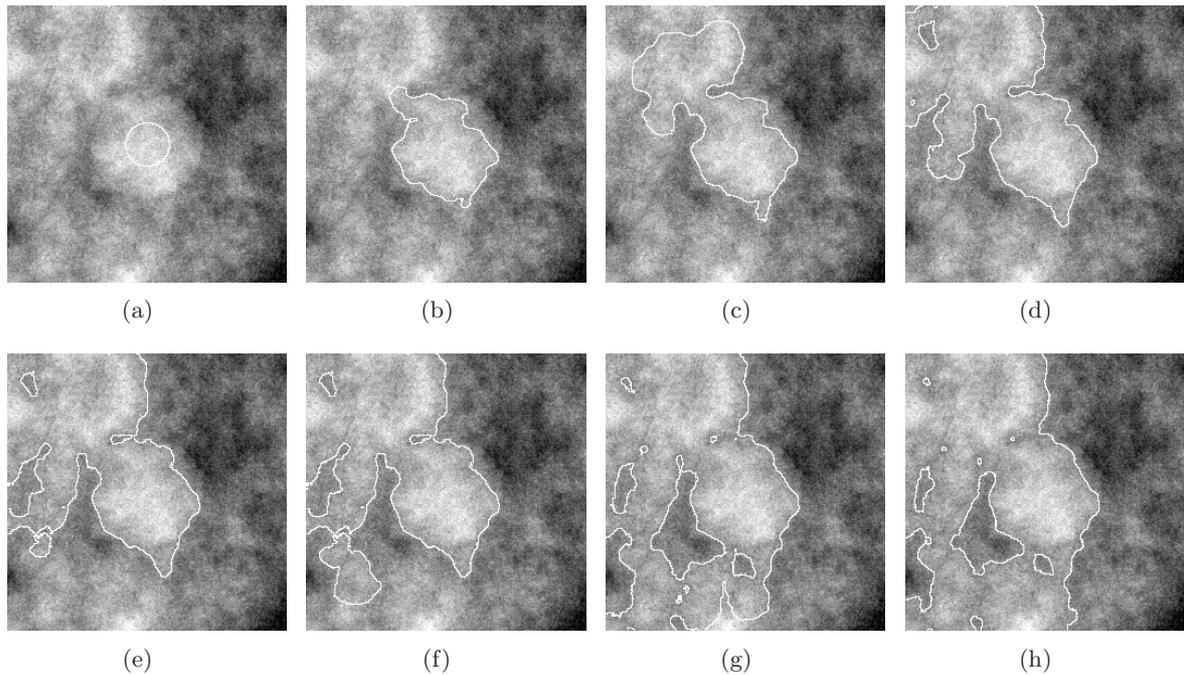


Figure 3.13: Example of the Chan-Vese method (region-based) processing a simulated breast ROI (256×256 pixels): different stages during the iterative process are illustrated from iteration zero (a) to the convergence in iteration 67 (h). The parameters of the algorithm used for this example where: $\mu = 1 \cdot 10^{-7}$, $\lambda_1 = \lambda_2 = 1.0$, and $\Delta t = 0.1$. Reinitialization was performed after each iteration. We can see that the contour evolution comes to a halt only after a global energy term (in the piecewise constant approximation) has converged to a local minimum. This results in a large region that is falsely segmented due to the breast tissue present in the image background.

They are used to illustrate the shortcomings of these kinds of approaches and to highlight the improvements provided by a hybrid approach. As a representative of the edge-based active contour models we have chosen a Fast-Marching implementation of the classical level-set approach (Osher 88; Malladi 95; Sethian 96). As a reference for region-based models we have chosen the piecewise constant approximation of the Mumford-Shah method proposed in (Chan 01b). The performance of the two reference algorithms on an exemplary simulated breast ROI containing a breast lesion is illustrated in Figure 3.12 and Figure 3.13.

Both approaches fail to produce an accurate segmentation result for the present data (Figure 3.14). Neither one of the image features (region homogeneity and gradient) provides sufficient information by itself. The edge-based method fails due to insufficient boundary information while the region-based approach cannot cope with the highly textured background. This raises the question, how we can integrate different image features, edge information and intensity information, into a single model in order to improve the results provided by the individual methods and to obtain an accurate description of the boundary of the object of interest.

In the last years, there has been a tendency towards algorithms which take advantage of the complementary nature of edge-based and region-based information (Haddon 90; Chakraborty 96; Freixenet 02). Several groups have worked on integrating these concepts into active contour models. Paragios et al. (Paragios 02) based their approach on geodesic active regions, while others approaches combined different active contour approaches with other segmentation techniques (Zhu 96; Ecabert 02; Munoz 03; Kim 05; Liu 06).

In order to obtain a formulation of a hybrid active contour model, we have chosen to start

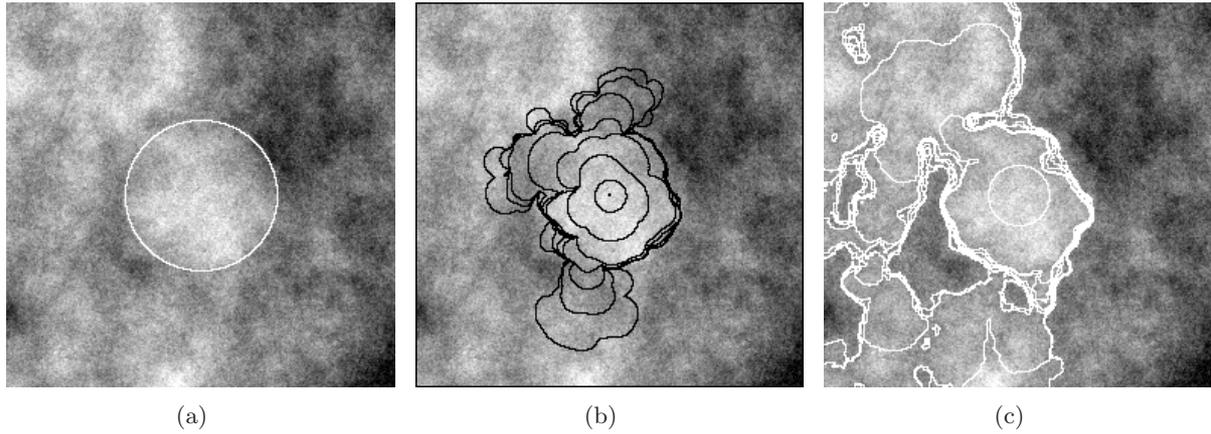


Figure 3.14: Comparison of active contour models based on a single image feature: (a) actual contour of the object in the simulated image, (b) segmentation result for the edge-based model and (c) segmentation result for the region based model. In (b) and (c) the multiple contours represent contours extracted at different times during the execution of the algorithm. These contours are indicated to convey a better understanding of the contour evolution from the initial contour to the final segmentation when convergence has been reached.

out from the approach based on maximizing region homogeneity (Equation 3.9). We then add an additional external force taking into account the edge information. The energy functional then becomes:

$$\begin{aligned}
 E(\phi) &= \mu \int_{\Omega} \delta(\phi(x, y)) |\nabla \phi(x, y)| dx dy \\
 &+ \nu \int_{\Omega} H(\phi(x, y)) dx dy \\
 &+ \alpha E_{region} + \beta E_{edge} + \gamma E_{pressure}.
 \end{aligned} \tag{3.10}$$

where μ , ν , α , β and γ are parameters of the algorithm. The energy term based on region homogeneity is given as in Equation 3.9, by (Chan 01b):

$$\begin{aligned}
 E_{region} &= \lambda_1 \int_{\Omega} |I_{org}(x, y) - \bar{c}_1|^2 H(\phi(x, y)) dx dy \\
 &+ \lambda_2 \int_{\Omega} |I_{org}(x, y) - \bar{c}_2|^2 (1 - H(\phi(x, y))) dx dy.
 \end{aligned} \tag{3.11}$$

The edge-based energy term is given by:

$$E_{edge} = - \int_{\Omega} \delta(\phi(x, y)) g(|I_{org}(x, y)|) dx dy, \tag{3.12}$$

where $g(|I_{org}(x, y)|)$ is a stopping function designed to slow down the contour in the vicinity of the edges, and is given by:

$$g(|I_{org}(x, y)|) = |\nabla(G_{\sigma}(x, y) * I_{org}(x, y))| \tag{3.13}$$

where $G_{\sigma} * I_{org}$, a smoother version of I_{org} , is the convolution of the image I_{org} with the Gaussian G_{σ} . The function $g(|I_{org}(x, y)|)$ is zero in homogenous regions and large at the edges.

Finally, the pressure term is given by:

$$E_{pressure} = \int_{\Omega} \delta(\phi(x, y)) dx dy, \tag{3.14}$$

which is a constant pressure term controlled by the parameter γ .

First, the original image is convolved with a median filter to limit the impact of noise. A Canny-Deriche edge detector (Deriche 87) is then applied to find the extrema of the gradient image. This edge map is thresholded in order to obtain a binary edge map. The threshold is determined for each image according to the density of the histogram of the edge map. Finally, the binary edge map is convoluted with a Gaussian kernel resulting in a smoothed version of the binary edge map. This processing is very popular with edge-based active contour models, since the smoothed edge map generally prevents edges to be overrun (see Figure 3.15).

Finally, γ is a constant term representing an inflation force. This term is needed to drive the contour towards the edges where the stopping term will compensate the inflation term and the contour evolution will come to a rest.

A major challenge when applying this model to image data lies in the choice of a well-adapted set of parameters μ , ν , α , β and γ .

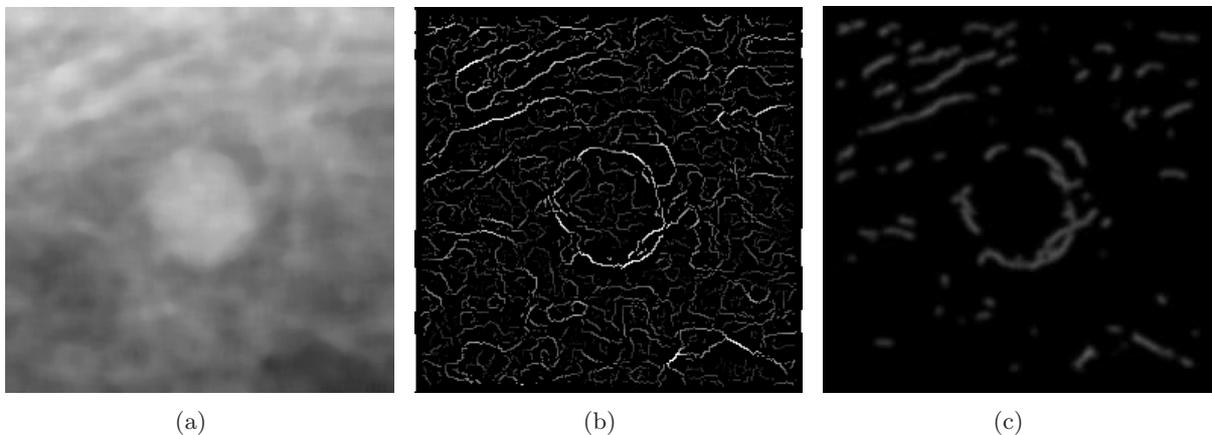


Figure 3.15: Examples of the preprocessing applied to the images for computing the external energy term: (a) median filtered image, (b) edge map produced by a Canny-Deriche edge detector and (c) the smoothed binary edge map obtained by adaptive thresholding and convolution with a Gaussian filter.

3.3.2 Experiments and Results

An accurate performance assessment of our hybrid active contour model regarding mass contour detection in clinical patient data cannot be performed directly since there is no ground truth available for this task. Three evaluation methods are used to quantify the detection results. Firstly, an expert has outlined contours of masses on mammograms (see Figure 3.16). Different criteria can be applied to quantify the accuracy of the segmentation result with respect to the reference contour (Timp 04). Secondly, applying the model to simulated images allows for a similar evaluation of the accuracy. Finally, a visual assessment of the segmentation results is performed by an expert.

The Clinical Database Since digital breast tomosynthesis is still an emerging modality, clinical data remain difficult to obtain. The lack of a truthed clinical database considerably narrows our ability to test and evaluate the proposed algorithm. The clinical database used here consists of few selected cases of visually detectable masses. The impact of lacking biopsy data corresponding to the images is however less severe than for the evaluation of overall detection algorithms. The truth, as far as segmentation is concerned, always needs to be defined by either

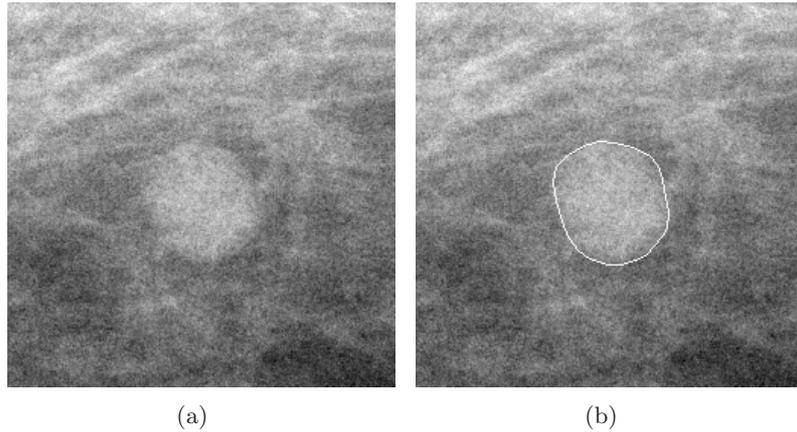


Figure 3.16: Manual segmentation by an expert: (a) ROI of a clinical image and (b) the reference contour extracted by an expert overlaid on the original image.

an expert or a priori knowledge about simulated images. Here, 8 actual breast lesions have been segmented manually by an expert for evaluation of the segmentation algorithm.

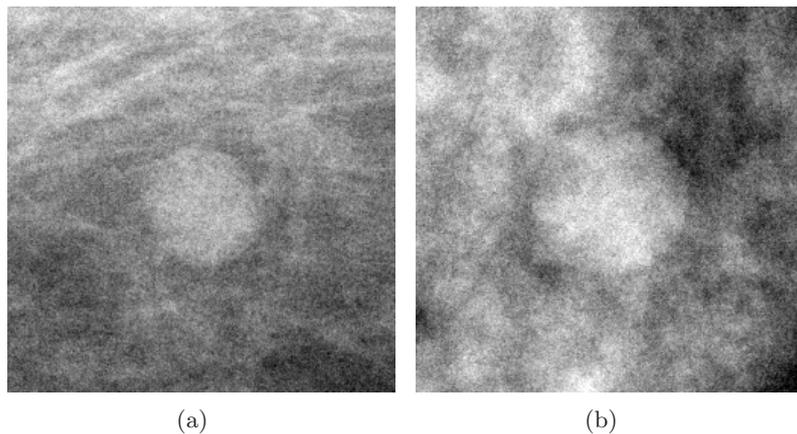


Figure 3.17: Examples of different kinds of data: (a) an actual breast lesion from a clinical image and (b) a simulated image for the same acquisition technique.

Simulated Data In order to be able to better test and evaluate our algorithm, we have generated a series of simulated images containing simulated breast lesions (see Figure 3.17). These images were computed for a typical dose found in digital breast tomosynthesis projection (about $\frac{1}{8}$ of a standard mammography dose) and for a breast thickness of 50 mm. Images were generated for different mass lesion sizes and several instances of texture and noise for each size. The database of simulated images contains 10 different instances of simulated DBT projection images. A detailed description of the processing chain used to generate the images can be found in (Grosjean 06). Attenuation coefficients used for masses come from composition of actual cancerous lesions (infiltrating ductal carcinoma), measured on actual specimen by Johns and Yaffe (Johns 87).

Evaluation Methods For evaluation purposes we have compared our hybrid active contour model to the two reference algorithms presented above: Fast-Marching model and Chan-Vese model. An expert has segmented the clinical database by hand in order to supply a reference for

comparing the different segmentation techniques (see Figure 3.16). For the simulated images, the dimensions of the simulated object are taken as reference. For each technique a number of measures were calculated comparing the segmentation result with the reference contour. These include the false positives and false negatives of the segmented contour, as well as similarity index, mean distance and Hausdorff distance between the two contours. False positives and false negatives are given as area (in pixels) relative to the area of the manually segmented object:

$$FP = \frac{a_{seg} - a_{seg \cap man}}{a_{man}}, \quad FN = \frac{a_{man} - a_{seg \cap man}}{a_{man}}, \quad (3.15)$$

where a_{seg} and a_{man} are the area inside the contour resulting from the segmentation and the area inside the reference contour manually segmented by an expert respectively. $a_{seg \cap man}$ is the area of the region resulting from the intersection of the regions enclosed by the two contours. The similarity index is the ratio between the intersection of the area enclosed by the two contours and the sum of their areas. While the mean distance is calculated averaging the distances of every pixel of contour C_{seg} to contour C_{man} and vice versa, the Hausdorff distance is a measure for the maximum amongst those distances, hence a very severe evaluation. A detailed description of the criteria of comparison is discussed in Appendix A.

Experimental Results

For all three compared models, the results were obtained from a convergence of the algorithm. This represents a local minimum of the energy functional. The contour evolution thus comes to a *natural* halt. A criterion needs to be defined to decide on whether a given minimum is significant, and should therefore stop the contour evolution. This criterion is often realized by computing the change in the function ϕ between two iterations. For the hybrid model, we have chosen to replace this criterion by a speed measure for the curve. If the movement of the curve is inferior to a given value (an additional parameter to the algorithm), we consider the local minimum to be sufficiently important to stop the contour evolution entirely.

In our numerical experiments, we generally chose the parameters of the hybrid model as follows: $\lambda_1 = 1$, $\lambda_2 = 0.5$, $\nu = 0$, $\mu = 0.05$, $h = 1$ (the step space), $\Delta t = 0.1$ (the time step), $\alpha = 1.0$ and $\beta = 200$.

Finding these parameter settings is not always an easy task. The impact of the external energy terms (edge-based energy and region-based energy) has been discussed in some detail above. The impact of variations in the parameter μ that controls the elasticity of the curve is illustrated in Figure 3.18.

For the Chan-Vese model, we chose $\mu = 1 \times 10^{-7}$ and reinitialized at each iteration in order to converge more quickly towards a stable segmentation. The goal for this approach was not to find the *best* possible segmentation (otherwise we could have chosen a higher μ), but to show the segmentation obtained by this method when the contour evolution comes to a rest, driven only by the region based term.

The edge-based Fast Marching Level-Set approach was performed with $\sigma = 1.0$, controlling the standard deviation of the Gaussian kernel used to convolute the image before application of the derivative operator. The parameters $\alpha = 0.5$ and $\beta = 3.0$ control the sigmoid used to map the dynamic of the gradient image to an interval $[0, 1]$ which is the input format for the speed image. The number of iterations performed was set to $i = 800$. Finally, we used a curvature/propagation ratio of 3.9, which allowed for a maximum smoothing of the curve in order to limit the bleeding into the background. This ratio controls the relationship between the internal energy of the curve and the propagation force. The initialization was chosen the same for the hybrid model and the region-based approach in order to allow for a fair comparison of

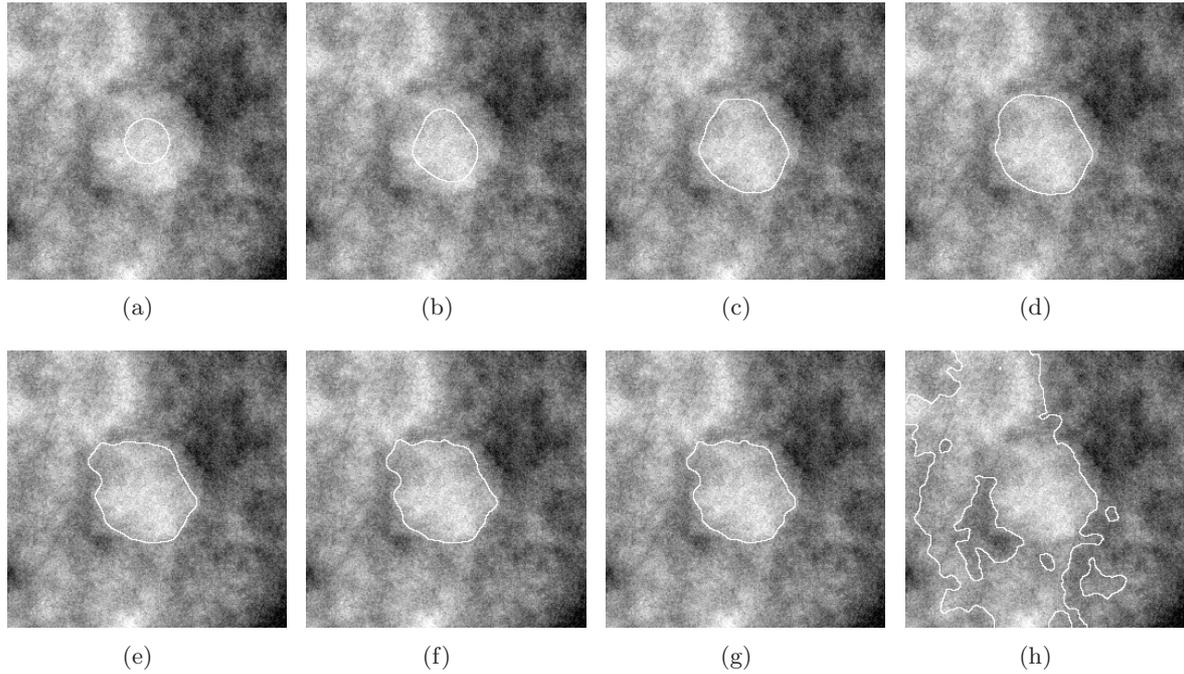


Figure 3.18: Example of parameter impact in processing a simulated breast ROI: the hybrid active contour model was applied with varying values for μ while keeping all other algorithm parameters constant ($\lambda = 1.0$, $\Delta t = 0.1$, $\alpha = 1.0$, $\beta = 300$, and no reinitialization). The values for μ that were used were: 1.0, 0.8, 0.5, 0.3, 0.1, 0.05, 0.03, and 0.01 for (a) to (h) respectively. In our model, the parameter μ controls the elasticity of the curve. We can see how larger values of μ increases the inner tension of the curve and thus smoothes the curve. For $\mu = 1.0$ (a) the inner tension becomes so strong that the external image based energies are no longer able to move the curve from its initial position. The lower the value for μ , the more irregular the contour.

the different methods. The edge-based fast-marching method is however implemented much like a region growing approach and thus takes a single seed point as initialization. We have chosen the center of gravity of the initialization of the two other approaches as initial seed.

We have compared the segmentation results obtained for each method to reference contours that have been manually segmented by an expert. This leads to a quantitative performance evaluation of the segmentation performance. Results are illustrated using different comparison metrics for clinical data in Table 3.1 and for simulated images in Table 3.2. A detailed description of the metrics used in Tables 3.1 and 3.2 can be found in Appendix A. We can see from the results that the edge-based model seems to work well on the simulated images while the hybrid model provides better performances for the clinical data. The region-based approach results in an elevated number of false positives since it tends to include the dense breast tissues in the image background in the segmented object. Overall, the hybrid model provides a good compromise between the two approaches that are based on a single image feature alone

Finally, a visual assessment of the contour extraction is performed. We compare the different segmentation approaches (region-based, edge-based, hybrid) and we also compared each of them to a reference contour segmented manually by an expert. This is illustrated for different clinical cases in Figure 3.19 - 3.21 and for a simulated breast lesion in Figure 3.22.

Figure 3.19 shows a ROI of a clinical case containing a circumscribed mass with partly obscured margin. This is a good illustration of the limitations of approaches based on a single image feature. The region-based approach fails due to the heavily textured background, while

Clinical Data	Region Based (Mean)	Region Based (σ)	Edge Based (Mean)	Edge Based (σ)	Hybrid Model (Mean)	Hybrid Model (σ)
False Positive	14.265	11.902	3.425	3.639	3,601	4.136
False Negative	0.040	0.071	0.158	0.182	0.198	0.410
Similarity Index	0.211	0.020	0.453	0.243	0.520	0.311
Mean Dist.	44.1	18.1	20.9	14.3	19.3	14.7
Hausdorff Dist.	138.9	39.9	43.6	16.8	53.3	37.3

Table 3.1: Results obtained for different segmentation algorithms on real data. All numbers are arithmetic means and standard deviation (σ) computed over the entire set of clinical data.

Simulated Data	Region Based (Mean)	Region Based (σ)	Edge Based (Mean)	Edge Based (σ)	Hybrid Model (Mean)	Hybrid Model (σ)
False Positive	5.557	6.240	0.846	1.438	1.204	2.141
False Negative	0.059	0.063	0.166	0.105	0.162	0.118
Similarity Index	0.441	0.283	0.729	0.023	0.711	0.225
Mean Dist.	30.7	12.6	9.1	4.7	11.3	7.5
Hausdorff Dist.	114.4	28.5	23.3	9.7	29.7	19.7

Table 3.2: Results obtained for different segmentation algorithms on simulated images. All numbers are arithmetic means and standard deviation (σ) computed over the set of simulated images.

the edge-based approach fails at the obscured part of the lesion boundary. The hybrid model however produces a smooth and more accurate contour.

In Figure 3.20, we demonstrate the performance of the different algorithms on the ROI of a clinical image containing a spiculated breast lesion. No dominant edge information is present in the image, and the edge-based method hence fails. Here, the result of the edge-based method stems from a good initialization and a dominant regularization term. The region-based method on the contrary detects the boundaries reasonably well. For this example, the difference between this method and the hybrid approach is notably due to the different regularization. It should be noted that the manual extraction of a reference contour for spiculated breast lesions is more subjective than for circumscribed lesions and may exhibit a significant inter-operator variability.

Figure 3.21 illustrates another classical example of highly textured background and ill-defined lesion margins. Again, the hybrid model manages to overcome the problem of leaking into the background and produces a smooth and accurate contour.

Finally, Figure 3.22 illustrates the segmentation results for a simulated image. The image contains a simulated, spherical, circumscribed mass, as well as typical breast tissues. As discussed earlier, the methods based on a single image feature fail due to different image characteristics. The contours resulting from the hybrid model do not exhibit these limitations.

It can be seen that the proposed method combines the advantages of the two reference algorithms and thus outperforms both. However, a validation on a larger database is required in order to obtain statistically significant results.

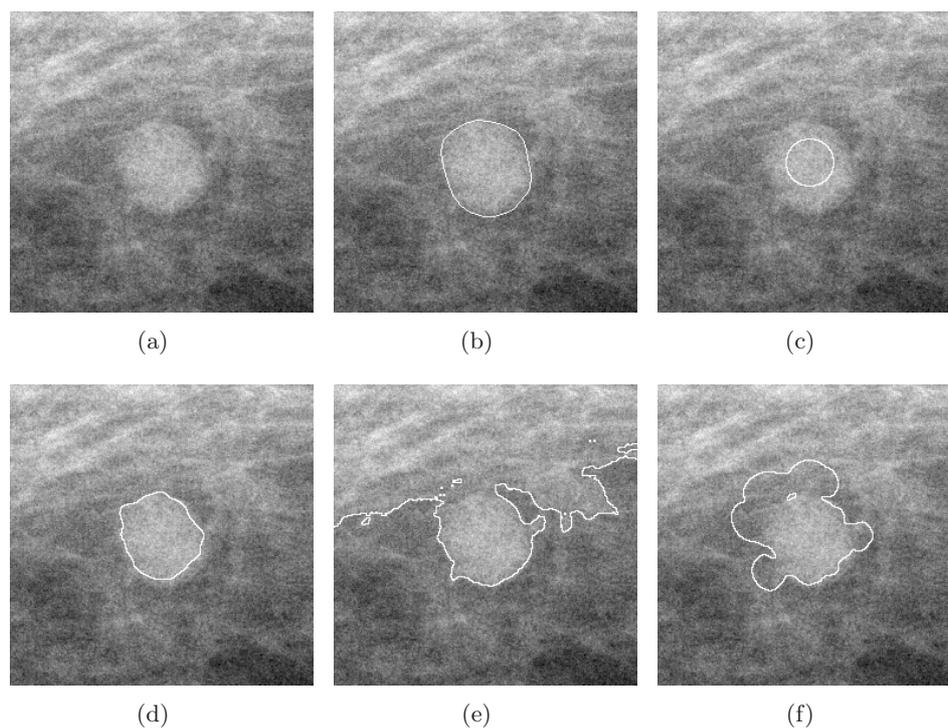


Figure 3.19: Results for an exemplary clinical case containing a circumscribed breast lesion: (a) original image, (b) reference contour overlaid on the original image, (c) initialization for the segmentation, (d) contour segmented by the hybrid model, (e) contour segmented by the region-based model, (f) contour segmented by the edge-based model.

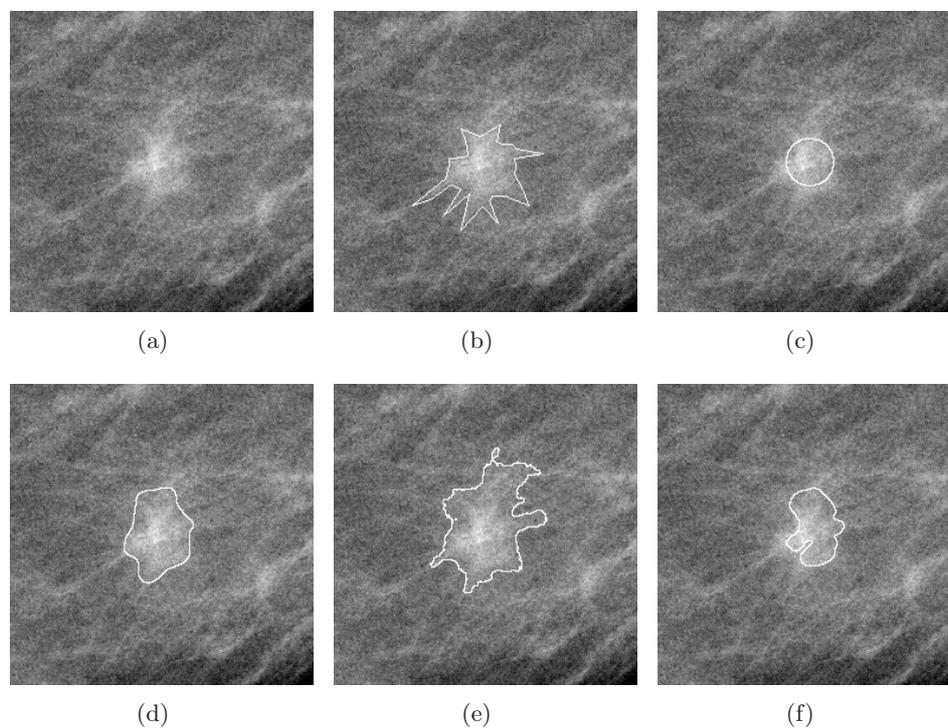


Figure 3.20: Results for an exemplary clinical case containing a spiculated breast lesion: (a) original image, (b) reference contour overlaid on the original image, (c) initialization for the segmentation, (d) contour segmented by the hybrid model, (e) contour segmented by the region-based model, (f) contour segmented by the edge-based model.

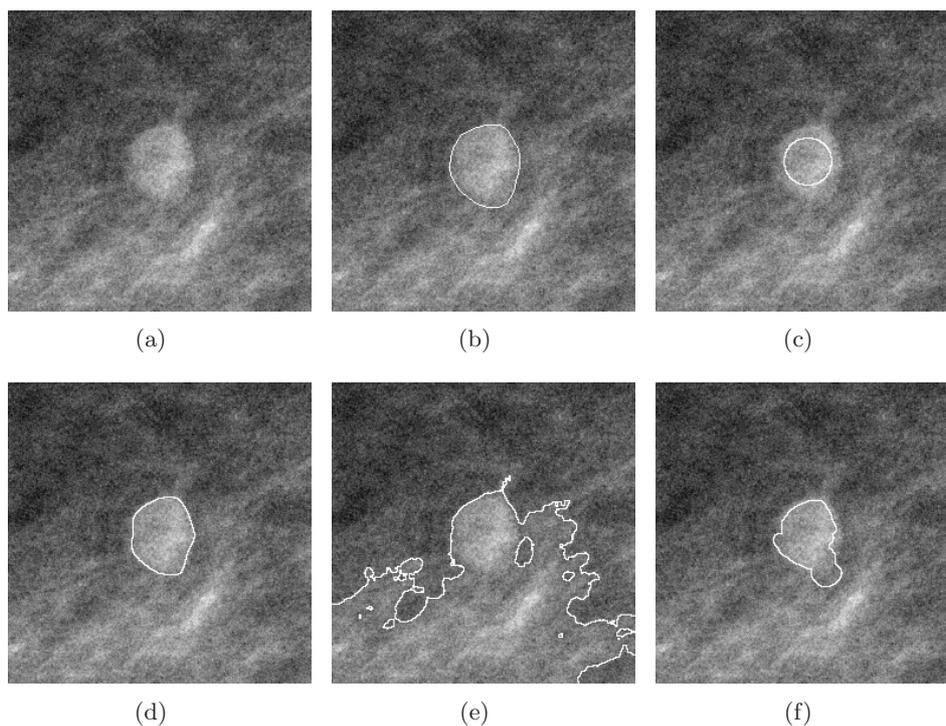


Figure 3.21: Results for an exemplary clinical case containing a circumscribed breast lesion: (a) original image, (b) reference contour overlaid on the original image, (c) initialization for the segmentation, (d) contour segmented by the hybrid model, (e) contour segmented by the region-based model, (f) contour segmented by the edge-based model.

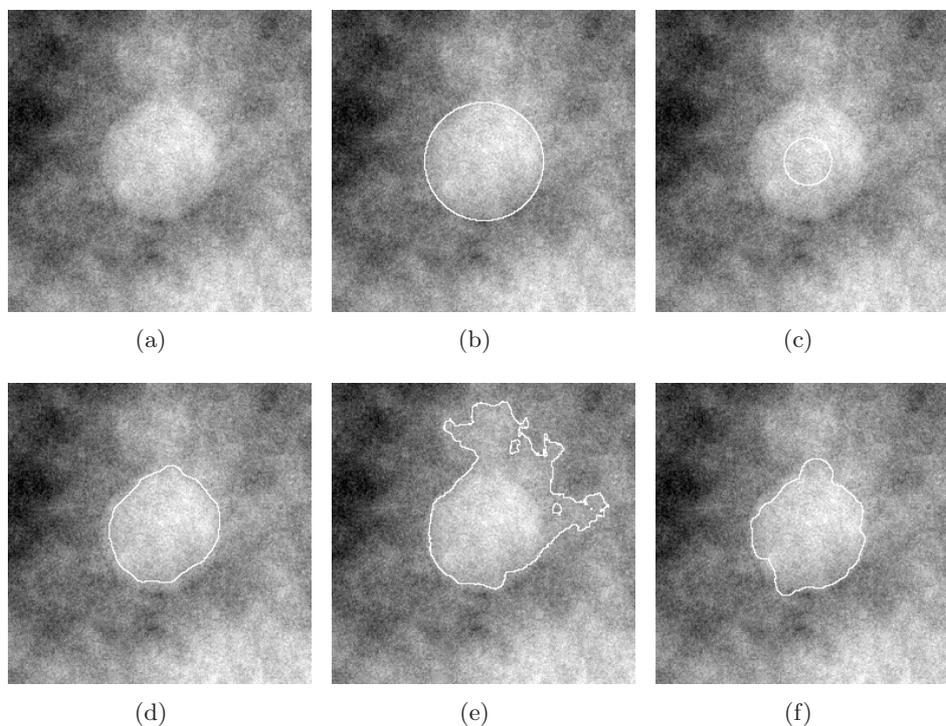


Figure 3.22: Results for an exemplary simulated image containing a simulated spherical breast lesion: (a) original image, (b) reference contour overlaid on the original image, (c) initialization for the segmentation, (d) contour segmented by the hybrid model, (e) contour segmented by the region-based model, (f) contour segmented by the edge-based model.

Conclusion

We have proposed a novel approach to segment masses in DBT datasets. Our approach exhibits several advantages. Working directly on the DBT projected views enables us to work independently of the reconstruction algorithm used to generate the 3D images. By introducing a hybrid active contour model, we take advantage of the available knowledge about different properties regarding breast masses for contour detection. This knowledge is translated in a combination of energies leading to an active contour model that provides a higher contour extraction accuracy with respect to the actual contour. Furthermore, our approach is more robust regarding a decreased CNR in the images, compared to classical active contour models based on a single image feature.

Potential improvements may be achieved by including additional image features in the energy functional. Incorporating a range of typical mass sizes or texture information are amongst the possibilities to be investigated. However, with an increasing number of parameters, finding a combination that is well adapted for a given detection task becomes increasingly difficult. An automation of this process is work in progress.

3.3.3 Fuzzy Contour Extraction

Fuzzy logic (Zadeh 65) can be used as a tool to manage imperfection in images. We can first deal with ambiguity at the segmentation level, then during the classification of the segmented particles using attributes measured from the contours. We usually distinguish between two types of imperfection: imprecision and uncertainty. At the segmentation level, imprecision characterizes particles we can identify with a high level of confidence but having contours difficult to localize with a high level of accuracy (Figure 3.23a). On the other hand, uncertainty characterizes particles having several potential contours for which it is difficult to determine the most relevant one (Figure 3.23b). This is the case when we have superimposition of objects, as in digital mammography. The imperfection of the contours propagates to the attributes measured from segmented particles and therefore to classification using those attributes. Based on previous developments (Bothorel 97; Rick 00), our approach presents the advantage to take into account imperfection in the image from the particle contours, introducing fuzzy contours, and to transfer this imperfection up to the classification step through fuzzy attributes.

Segmentation of Particles and Fuzzy Contours

If we consider images with particles that can be extracted using a thresholding technique, we can start from the regular thresholding pyramids approach (Kirsch 72) to describe thresholded images. This transformation consists in applying a set of thresholds corresponding to the gray levels available in the image. Each threshold determines connected subsets that can be linked one to the other using the usual inclusion operator. A fuzzy segmentation method was proposed by Kanzaki (Kanzaki 92) that can be considered to be the application of a regular thresholding pyramid to the result of the top-hat transform applied to an image. Particles are therefore considered to be membership functions to the class *contour*, each alpha-cut applied to these functions being a potential contour for a particle. Smits et al (Smits 94) proposed a similar approach improving computation of membership degrees and using a more sophisticated method for propagation of contours. In this approach, the image is processed at structure level and no more at pixel level. It is therefore possible to calculate membership degrees of pixels from attributes based on gray levels of the current and surrounding pixels but also based on geometrical properties of the particles themselves. Nevertheless, these two approaches cannot handle superimposed objects, as present in radiological images, since they are both based on the assumption

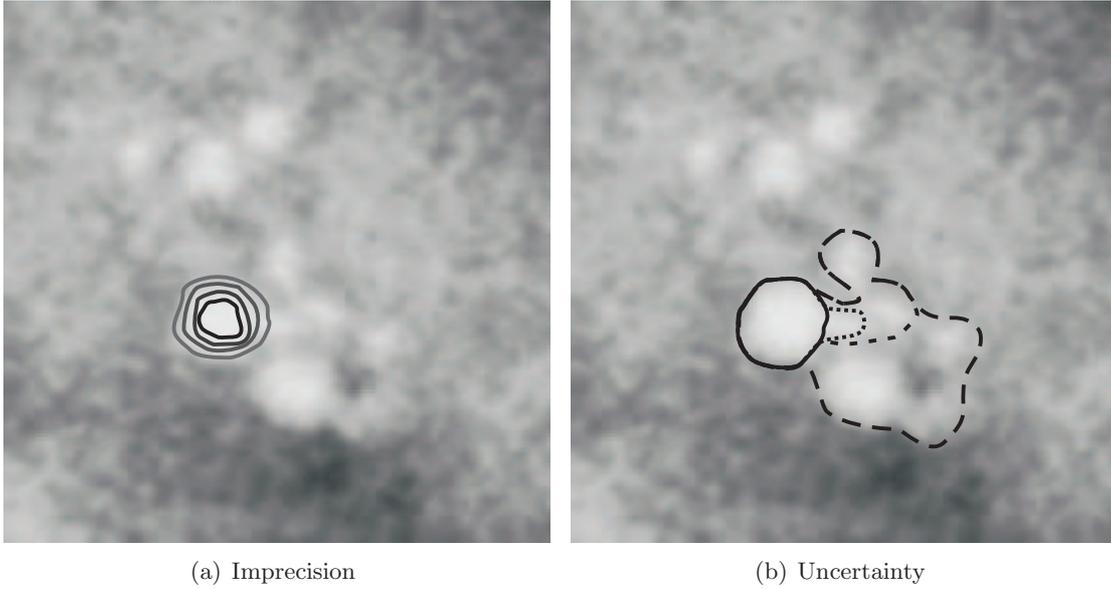


Figure 3.23: (a) A particle identified with a high level of confidence, but with contours difficult to localize with a high level of accuracy, presents imprecision. (b) A particle having several potential contours, for which it is difficult to determine the most relevant one, presents uncertainty.

that membership degrees increase with gray levels.

Our fuzzy segmentation approach is composed of three steps:

- provide a marker for each particle to be segmented;
- determine a set of candidate contours for each particle;
- build a fuzzy contour for each particle through the valuation of the candidate contours using their membership degrees to the class *contour*.

Handling imprecision and uncertainty of the particles at contour level, as opposed to pixel level, allows the reduction of the fuzzy segmentation dimensionality, since we pass from a two-dimensional description (region) to a one-dimensional description (contour).

Let the set of regional maxima of the intensity image I be defined as $Max(I)$. We then consider as markers the regional maxima $M_I \in Max(I)$ of the intensity image I determined in the support of the candidate particles detected using wavelet filtering.

Set of candidate contours For each particle marked by a marker M_I a set of candidate contours is obtained by region growing. The threshold of I at level t_i is defined as:

$$I_{t_i} = \{(x, y) \mid I(x, y) \geq t_i\} \quad (3.16)$$

Using this definition, we determine a set of contours \mathcal{C}_{M_I} using a multi-thresholding operator such as:

$$\mathcal{C}_{M_I} = \left\{ C_i = \partial R_{M_I}(I_{t_i}), t_1 = \max_{M_I}(I) > t_2 > \dots > t_{n_I} \right\} \quad (3.17)$$

with $\partial R_{M_I}(I_{t_i})$ being the border of the region R_{M_I} for the set $R_{M_I}(I_{t_i})_{i=1, \dots, n_I}$ that includes the marker M_I of the image I , and $t_1 = \max_{M_I}(I)$ being the maximum gray level of the image I in the marker M_I . This process is depicted in Figures 3.24a and 3.24b.

Each maximum of the image is processed independently. Therefore, some contours C_i can be duplicated and belong to several sets of contours built for particles marked with different markers (Figure 3.24c). This property allows handling superimposition in radiological images.

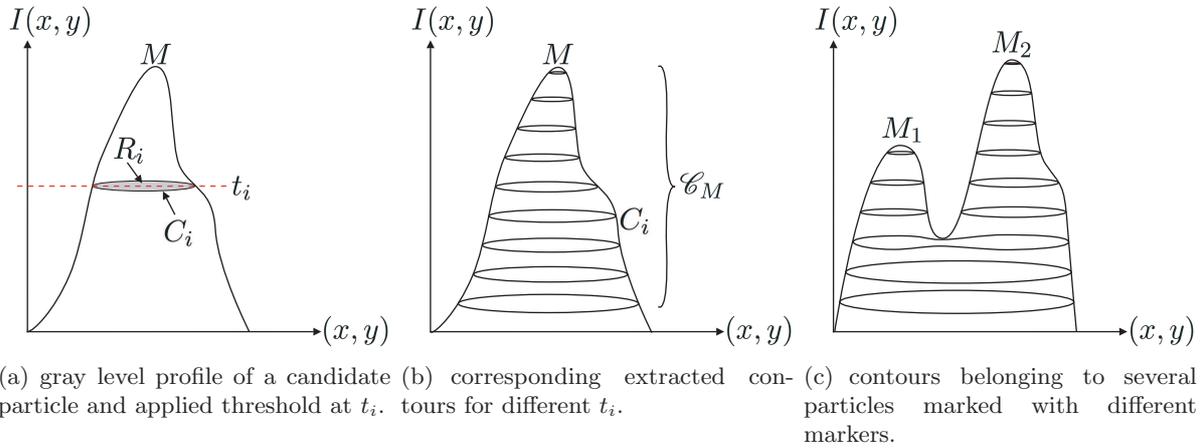
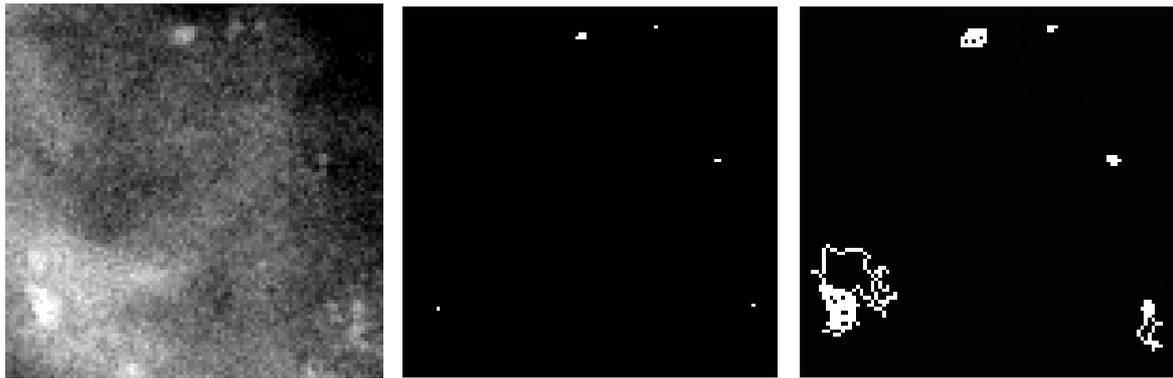


Figure 3.24: Set of candidate contours



(a) original breast ROI containing microcalcifications. (b) corresponding markers obtained from wavelet filtering. (c) fuzzy contours for the same ROI.

Figure 3.25: Fuzzy Contour Extraction

Fuzzy Contour Computing For a given particle marked with a local maximum M_I , the set of contours \mathcal{C}_{M_I} is considered the universe of discourse \mathcal{X} for the fuzzy logic approach. To each candidate contour C_i , we assign a numerical value corresponding to its membership degree to the class *contour* using the characteristics of the particle to be segmented (contrast, shape, size, texture index, localization, etc.). Therefore, a particle in the image I with a marker M_I is characterized by a fuzzy contour, fuzzy subset of \mathcal{X} , defined by a membership function $\mu_{\mathcal{C}/\mathcal{A}_1}$ such as:

$$\mu_{\mathcal{C}/\mathcal{A}_1} : \{C_1, \dots, C_{n_I}\} \rightarrow [0, 1] \quad C_i \mapsto \mu_{\mathcal{C}/\mathcal{A}_1}(C_i) \quad (3.18)$$

with $\mu_{\mathcal{C}/\mathcal{A}_1}(C_i)$ being the membership value for the candidate contour C_i and \mathcal{A}_1 a fuzzy subset characterizing a criterion that allows the calculation of the membership value.

The membership function to the class *contour* $\mu_{\mathcal{C}/\mathcal{A}_1}$ is built using a criterion characterized by a fuzzy subset \mathcal{A}_1 that represents a priori knowledge about the pertinence of the candidate contours (Figure 3.26). Let us assume that the most pertinent contours of a given particle have

the highest gradient mean values along the contour. Based on the knowledge we have about the images to process, we build the membership function $\mu_{\mathcal{A}_1}$ that characterizes the fuzzy subset \mathcal{A}_1 of the high gradients defined on the universe \mathcal{Y}_1 of the gradient values a_1 . For a given particle, we compute the arithmetic mean of the gradient values along each candidate contour C_i , determining an application ψ from the universe \mathcal{X} of the candidate contours C_i to the universe \mathcal{Y}_1 of the gradient values a_1 .

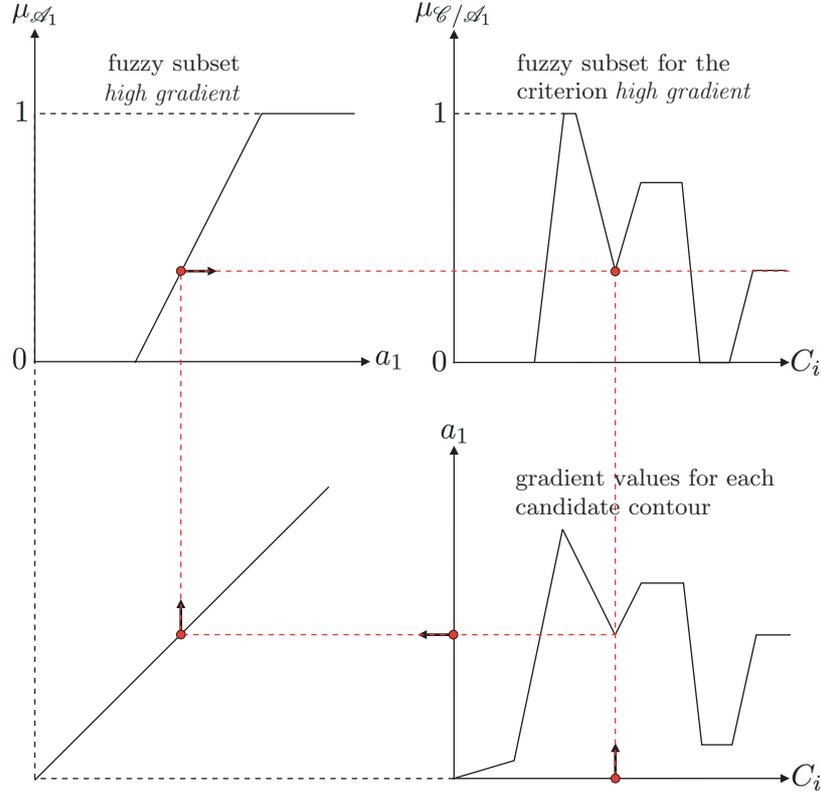


Figure 3.26: *A priori* knowledge about the images to process allows building the fuzzy subset *high gradient* (top left). For a given particle, the value of the gradient under each candidate contour is measured (bottom right). Therefore, the fuzzy subset *contour* relative to the criterion *high gradient* for the particle is derived (top right).

Using the fuzzy subset \mathcal{A}_1 on the universe \mathcal{Y}_1 allows defining the fuzzy subset $\mathcal{C}_{\mathcal{A}_1}$ on the universe \mathcal{X} :

$$\mu_{\mathcal{C}/\mathcal{A}_1}(C_i) = \mu_{\mathcal{A}_1}(\psi(C_i)), \quad (3.19)$$

where $\mu_{\mathcal{A}_1}$ acts as a lookup table on the attribute values.

When a combination of several criteria is considered (e.g., high gradient and small area for particles being potential microcalcifications), the valuation of the candidate contours C_i can be extended. Let us consider the fuzzy subsets $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_p$ defined on the universes $\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_p$ of the values of characteristics a_1, a_2, \dots, a_p respectively. The fuzzy subset \mathcal{C} on the universe \mathcal{X} relative to the criteria characterized by the fuzzy subsets $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_p$ is then the intersection of the fuzzy subsets $\mathcal{C}_{\mathcal{A}_1}, \mathcal{C}_{\mathcal{A}_2}, \dots, \mathcal{C}_{\mathcal{A}_p}$ (Figure 3.27). The membership function to the class *contour* $\mu_{\mathcal{C}/\mathcal{A}_1 \cap \mathcal{A}_2 \cap \dots \cap \mathcal{A}_p}$ based on the criteria a_1, a_2, \dots, a_p is defined by:

$$\forall C_i \in \mathcal{X}, \quad \mu_{\mathcal{C}/\mathcal{A}_1 \cap \mathcal{A}_2 \cap \dots \cap \mathcal{A}_p}(C_i) = \top (\mu_{\mathcal{C}/\mathcal{A}_1}(C_i), \mu_{\mathcal{C}/\mathcal{A}_2}(C_i), \dots, \mu_{\mathcal{C}/\mathcal{A}_p}(C_i)) \quad (3.20)$$

where \top denotes a conjunctive operation. The t-norm applied here is the original one introduced by Zadeh (Zadeh 65): the minimum.

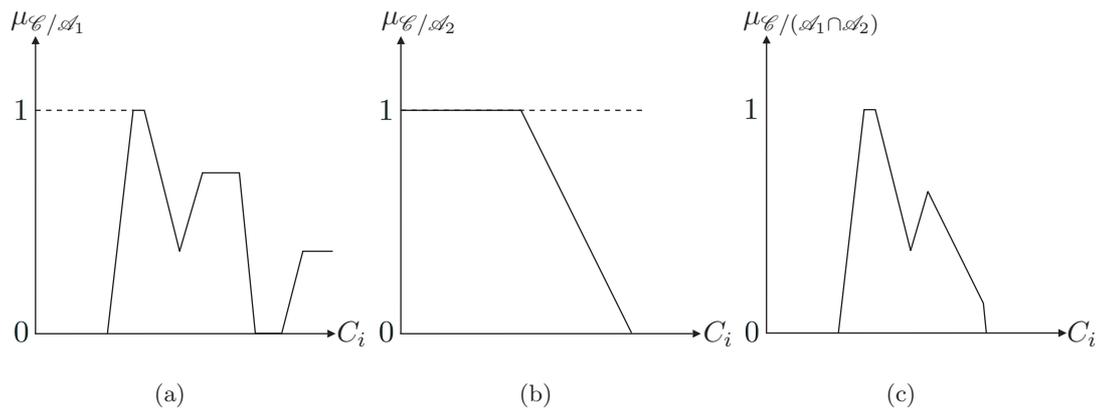


Figure 3.27: The fuzzy subset *contour* relative to the criteria *high gradient* and *small area* is obtained from the minimum between the fuzzy subset *contour* relative to the criteria *high gradient* and the fuzzy subset *contour* relative to the criteria *small area*.

3.3.4 A New Class of Fuzzy Contours: Fuzzy Active Contours

In the preceding section we have discussed the extraction of fuzzy contours using multi-thresholding. Several advantages of the use of fuzzy set theory have been pointed out. However, due to the properties of the clinical data, multi-thresholding is a powerful method only when used for microcalcification contour extraction (Bothorel 97). In this section, we will discuss a method for extracting fuzzy contours of mass lesions in DBT data based on active contour models.

Fuzzy contour extraction is used to model the ambiguity present in the image data. The parameters of an active contour model influence the path that the active contour will follow during its evolution in the image. The convergence criterion defines at which point during this evolution process the model will come to a halt. We may therefore choose to vary the convergence criterion or the parameter setting in order to obtain a set of candidate contours C_i . If we choose to vary the convergence criterion, the model may become more robust to local minima that stop the contour evolution before it can reach the actual object boundaries since we can indirectly define the importance of a minimum necessary to stop the contour evolution. However, if the model is not well adapted to the object of interest, the contour evolution process may not include good contours for the object. On the other hand, if we choose to vary the algorithm parameters, we may fit a wider range of objects of interest. For two different parameter settings $\{\mu_i, \alpha_i, \beta_i, \gamma_i\}$ and $\{\mu_j, \alpha_j, \beta_j, \gamma_j\}$ ($i \neq j$), we will obtain two corresponding candidate contours C_i and C_j (see Equation 3.10). However, this does not guarantee a convergence towards the appropriate minimum corresponding to the actual object boundaries. We therefore introduce a combination of the two approaches. First, we present a model resulting in a set of candidate contours for a given object of interest. We then discuss how this can be extended to multiple object classes.

A level-set approach

We start from the hybrid active contour model described in Section 3.3.1. We tune the parameters of the algorithm according to a priori knowledge about a given object of interest. For demonstration purposes we choose here a circumscribed breast mass as object of interest. A circumscribed mass is generally compact, with a sharp boundary. The corresponding parameters should therefore reflect a strong regularization (μ), and a strong weight on the edge-based energy term (β).

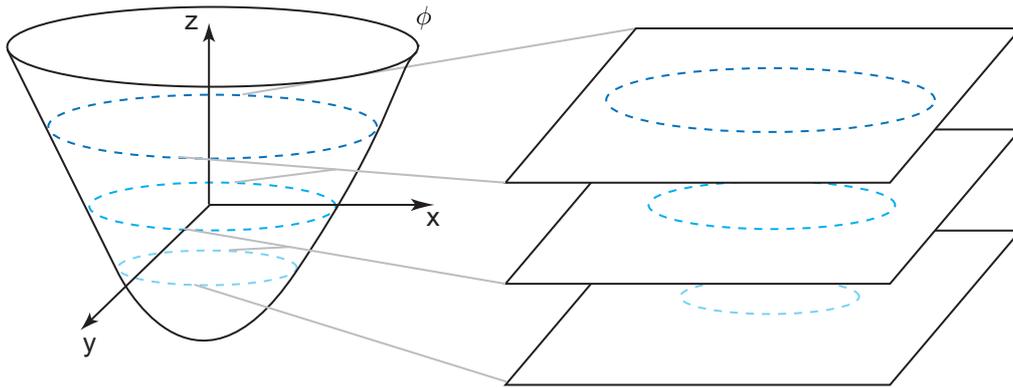


Figure 3.28: Extracting the level-sets: based on the function ϕ we are able to extract different candidate contours corresponding to level-sets extracted for different values of ϕ .

Once the model converges and the contour evolution comes to a rest, we have found our first candidate contour. This is the 0-level-set and equivalent to the contour extracted for the crisp case. Next we will take advantage of the properties of the level-set implementation for geometric active contours. Since we are evolving a three-dimensional function ϕ we may choose to extract level-sets of this function at levels where $\phi \neq 0$ (see Figure 3.28).

In this way, each extracted candidate contour corresponds to a level of the ϕ -function. We then need to answer three questions:

- How are the levels for extracting the candidate contours chosen?
- What modifications are necessary in order to obtain meaningful candidate contours at levels $\phi \neq 0$?
- How is the membership degree of each candidate contour to the class contour computed?

Candidate contours are obtained by extracting the level-set of the function ϕ at a given level ϕ_i . The choice of appropriate values for the ϕ_i will therefore determine the locations of the candidate contours. For best performance this should be subject to an analysis of ϕ . For reasons of simplicity we have chosen a more straightforward approach. We compute the levels ϕ_i evenly spaced over a narrow band around the 0-level-set.

In order to be able to extract meaningful contour candidates at levels $\phi \neq 0$ two considerations need to be taken into account: re-initialization and image domain. Re-initializing ϕ during the contour evolution process is generally achieved by use of the signed distance function to the 0-level-set. Re-initialization is used to avoid deformations of ϕ that may have a negative impact on the contour evolution process. In the case of fuzzy active contours however, the deformations of ϕ are the only information about the image-based energy terms at $\phi \neq 0$ and are therefore necessary for extracting a meaningful set of candidate contours. Re-initializing ϕ leads to candidate contours that are scaled, centered replica of the 0-level-set candidate contour (if we re-initialize using a signed distance function to the 0-level-set). Deformations caused by image based energy terms for $\phi \neq 0$ are erased. Re-initialization has therefore a negative impact on the set of candidate contours regarding the degree to which these candidate contours reflect the image data. For similar reasons, we prefer an approach that works on the entire ϕ in each iteration instead of just working on a narrow band around the 0-level-set. For example, if we apply a narrow band approach, only the values inside the narrow band around the 0-level-set are updated during an iteration of the contour evolution. Candidate contours should then only be extracted for ϕ_i corresponding to candidate contours inside the narrow band.

With our fuzzy active contours approach the contour extraction process is subject to constraints defining the object of interest. We therefore suggest to deduce the membership degree to the class *contour* (of a circumscribed mass) directly from the energy level of a given candidate contour. The lower the energy of a candidate contour, the better it can be considered to correspond to the class *contour*. We then need to define a function:

$$\mu_{\mathcal{C}/\mathcal{A}_E} : \{C_1, \dots, C_{n_I}\} \rightarrow [0, 1] \quad C_i \mapsto \mu_{\mathcal{C}/\mathcal{A}_E}(C_i) \quad (3.21)$$

with $\mu_{\mathcal{C}/\mathcal{A}_E}(C_i)$ the membership value for the candidate contour C_i and \mathcal{A}_E a fuzzy subset characterizing a criterion that allows the calculation of the membership value (in our case the energy of the candidate contour).

In fuzzy set theory, a fuzzy subset is often characterized by a trapezoid function used to translate attribute values into membership degrees. The function we use to translate the energy of a candidate contour into the membership degree to the class *contour* is given by:

$$\mu_{\mathcal{C}/\mathcal{A}_E}(C_i) = \begin{cases} 0 & \text{if } |E(C_i)| > E_{max} \\ 1 & \text{if } |E(C_i)| < E_{min} \\ g(E(C_i)) & \text{if } E_{min} \geq |E(C_i)| \geq E_{max} \end{cases} \quad \text{with } g(E(C_i)) = 1 - \frac{|E(C_i)| - E_{min}}{E_{max} - E_{min}} \quad (3.22)$$

where E_{min} and E_{max} are the lower and the upper bounds for the energy values of the candidate contours respectively and $E(C_i)$ is the energy value of the candidate contour C_i . The bounds of the energy values are fixed empirically. The resulting function is illustrated in Figure 3.29.

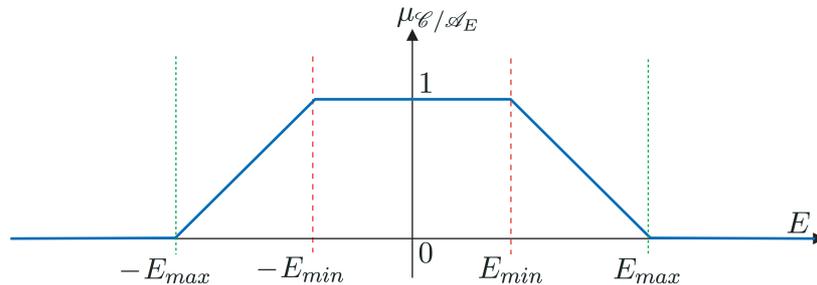


Figure 3.29: Membership function to the class *contour*: The fuzzy subset \mathcal{A}_E defines in what way the energy of a candidate contour characterizes the membership of the candidate contour to the class *contour*. The fuzzy subset \mathcal{A}_E has been realized as a trapezoid function (see Equation 3.22) defined by the parameters E_{min} and E_{max} that represent the lower and upper bounds of the energy values respectively.

Multiple Fuzzy Contours

One of the drawbacks of crisp segmentation techniques lies in the fact that it requires a decision to be taken. We want to delay this decision by use of fuzzy contour extraction in order to gather additional information before a decision is taken. However, the choice of parameters guiding the segmentation algorithms constitutes a decision as well (see (Bothorel 97)). This is especially true for an active contour approach like the one presented above.

In contrast to a multi-thresholding approach, active contour models rely on a priori knowledge about the object of interest. In fact, one of the major advantages of active contour models is that an a priori definition of the object of interest can be included in the energy functional in order to guide the contour evolution subject to these constraints. A problem arises when

the object of interest is unknown, or if we expect one of a set of objects that exhibit distinct differences regarding their properties.

There are two possibilities to address this issue. The first one consists in identifying the properties that are common to the entire set of objects of interest, and to define the active contour model in such a way that any of the objects may match the definition. We then need to design the active contour model based on a combination of the individual object class definitions. This presents a very general approach, allowing us to operate with a single object definition (resulting in a single fuzzy active contour per object). On the other hand, the more the object classes differ from each other, the lower the accuracy of the resulting segmentation. Since we are looking for a compromise between the *optimum* contour for each of the object classes, the accuracy of the compromise declines with growing object dissimilarities.

We have therefore chosen an alternative approach for obtaining accurate fuzzy active contours for different kinds of objects of interest. In the case of computer-aided detection, we do not know which kind of object we are dealing with. However, we generally know the set of objects that represent the possible alternatives. We then proceed in constructing a fuzzy active contour model for each of those objects. Each of the models will be applied to all of the detected objects. This process is illustrated in Figures 3.30 and 3.31 for two kinds of objects of interest: circumscribed masses and spiculated masses.

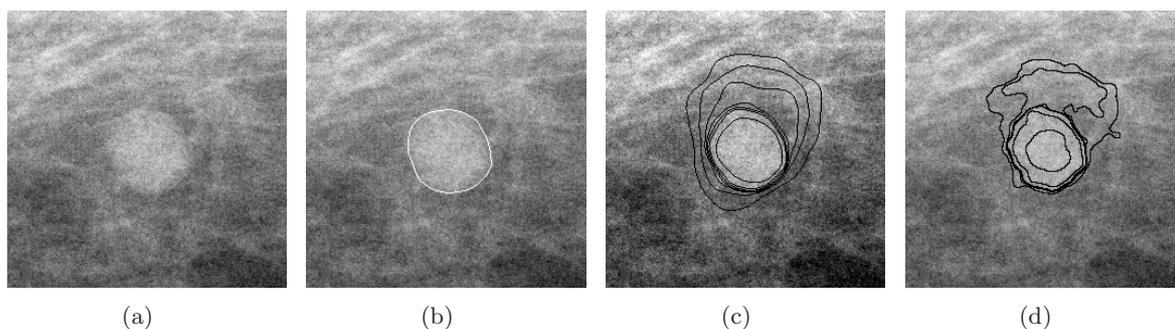


Figure 3.30: Example of different fuzzy active contour hypothesis on a circumscribed mass: (a) original breast ROI containing a circumscribed mass, (b) reference contour manually segmented by an expert, (c) fuzzy segmentation result using the circumscribed mass model and (d) fuzzy segmentation result using the spiculated mass model.

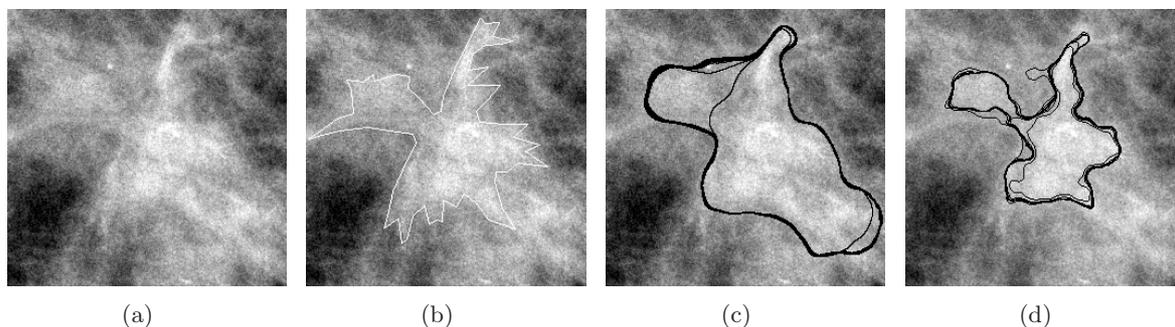


Figure 3.31: Example of different fuzzy active contour hypothesis on a spiculated mass: (a) original breast ROI containing a spiculated mass, (b) reference contour manually segmented by an expert, (c) fuzzy segmentation result using the circumscribed mass model and (d) fuzzy segmentation result using the spiculated mass model.

We can see from Figure 3.30 how the circumscribed active contour model is better suited

for the circumscribed mass present in the image. The strong regularization overcomes the irregularities in the contour and places where the contour information is corrupted (Figure 3.30 (c)). The spiculated model on the other hand leads to a fuzzy contour where many small irregularities have caused the contour evolution to deviate from the actual object boundary (Figure 3.30). In Figure 3.31 the opposite effect can be observed. The circumscribed model (Figure 3.31 (c)) leads to a segmentation that is too regular and does not follow the individual spicules of the mass. The spiculated mass model obtains a much better result and effectively follows the (main) spicules of the mass ((Figure 3.31 (d)).

Applying an active contour model that has been built based on a priori knowledge about a specific object of interest for segmentation of an unknown object can be interpreted as making a hypothesis regarding the unknown object. We consider the unknown object to correspond to the class of objects of interest that was used to build this fuzzy active contour model. We will then use succeeding processing steps to validate the hypothesis taken during the fuzzy segmentation. An illustration of this process for two exemplary classes (circumscribed mass and spiculated mass) is given in Figure 3.32.

Both fuzzy active contour models are applied to each detected object. The resulting fuzzy contours are then processed independently. For each fuzzy contour, feature extraction (Section 3.4.1), aggregation (Section 4.3) and classification (Section 5.4) are performed in order to obtain a confidence degree regarding the validity of the original hypothesis. These confidence degrees serve to make a final decision about the actual type of object at hand.

Here we have discussed the application of multiple fuzzy contours for the case of two models for different breast masses. The proposed approach can be easily extended to process an arbitrary number of different fuzzy contours, each corresponding to different assumptions made regarding the object of interest.

Experimental Results

The results in this section were obtained from convergence of the active contour models (see Section 3.3.2). In our numerical experiments, we have chosen the parameters of the fuzzy active contour model based on circumscribed masses as follows: $\lambda_1 = \lambda_2 = 1$, $\nu = 0$, $\mu = 1.5$, $h = 1$, $\Delta t = 0.1$, $\alpha = 1.0$, $\beta = 100$, and $\gamma = 2000$. No re-initialization was performed and we extracted seven candidates for each object. For this model the dominating terms correspond to the regularization (*compactness*) and the edge-based energy term (*sharp edges*).

For the model based on spiculated masses, we have chosen the following parameter setting: $\lambda_1 = \lambda_2 = 1$, $\nu = 0$, $\mu = 0.01$, $h = 1$, $\Delta t = 0.1$, $\alpha = 8.0$, $\beta = 100$, and $\gamma = 500$. No re-initialization was performed and we extracted seven candidates for each object. For the model based on spiculated masses the dominant energy term is the region-based term (*homogeneity*), while a smaller weight is assigned to the edge-based term and the regularization force.

Figures 3.33 to 3.36 show results for different models applied to different objects of interest. In each figure, (a) the original image, (b) the contour segmented by an expert and (c) the fuzzy segmentation result are shown. Figure 3.34 shows the level-set function ϕ corresponding to the clinical case in Figure 3.33 at convergence of the contour evolution. We can see how ϕ has been deformed over the entire image domain by the different image forces.

The fuzzy segmentation results show the set of candidate contours in white overlaid on the original image. The contour extracted for the 0-level-set is drawn in black. This helps visualizing the improvement provided by the fuzzy active contour approach. For the simulated lesions in Figure 3.36 we observe that the contour candidate at $\phi = 0$ (convergence of the crisp model) does not correspond to the best segmentation among the extracted candidates.

Figure 3.36 shows two good examples for the advantages of the fuzzy approach compared to

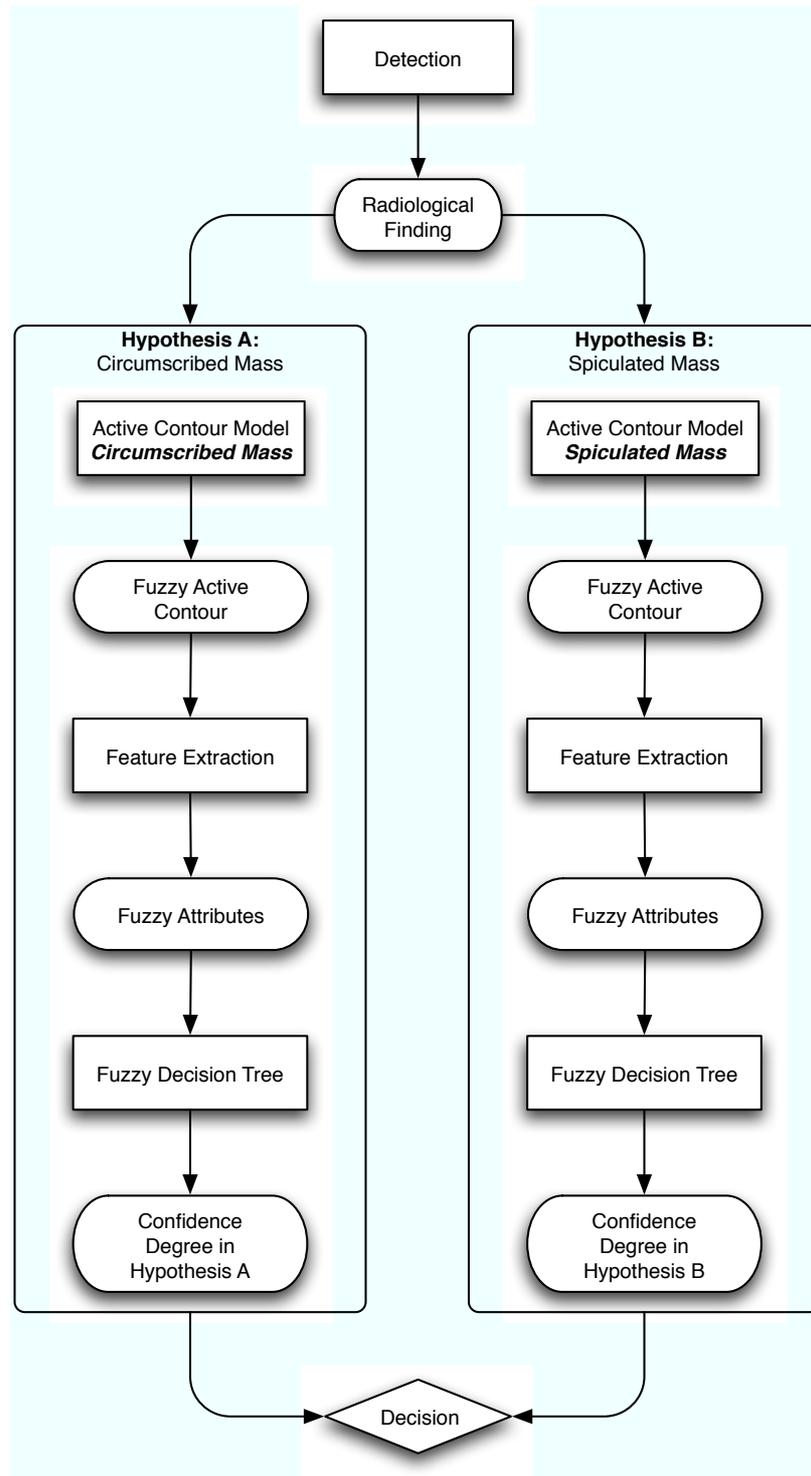


Figure 3.32: Hypothesis testing for a radiological finding. Two active contour models are applied. Each model takes into account a priori knowledge about an object of interest (circumscribed/spiculated mass). Each fuzzy contour is then processed separately. Feature extraction, aggregation, and classification provide a means to validate the initial hypothesis.

the crisp model. We can see that the model converges at a point where a considerable amount of background tissue is still included in the contour (black 0-level candidate contour). In this case, the next smaller candidate contour appears to be a much better approximation of the reference contour.

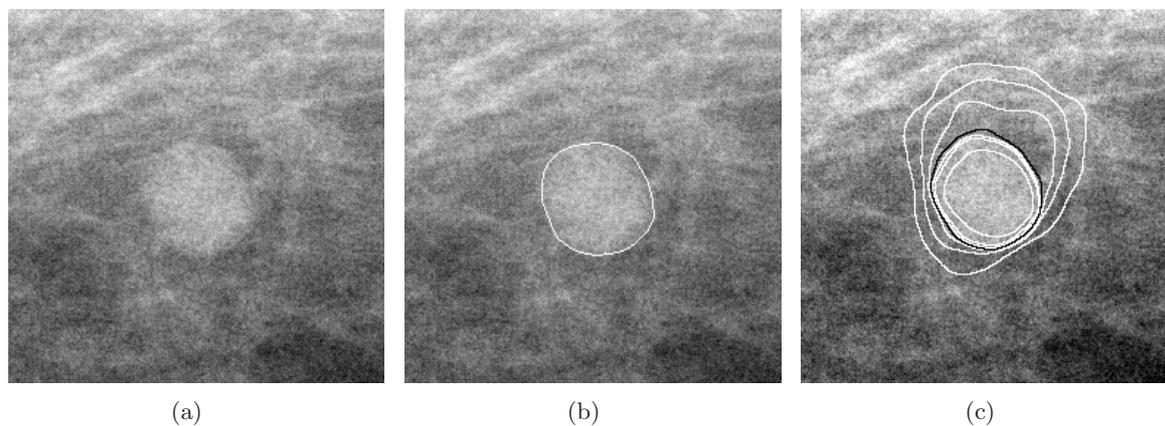


Figure 3.33: Results for the fuzzy active contour model (circumscribed mass model) applied to an exemplary clinical case containing a circumscribed breast mass: (a) original image and (b) reference contour overlaid on the original image. (c) the candidate contours extracted with the fuzzy active contour model are illustrated as an overlay on the original image. The black contour corresponds to the 0-level-set at convergence.

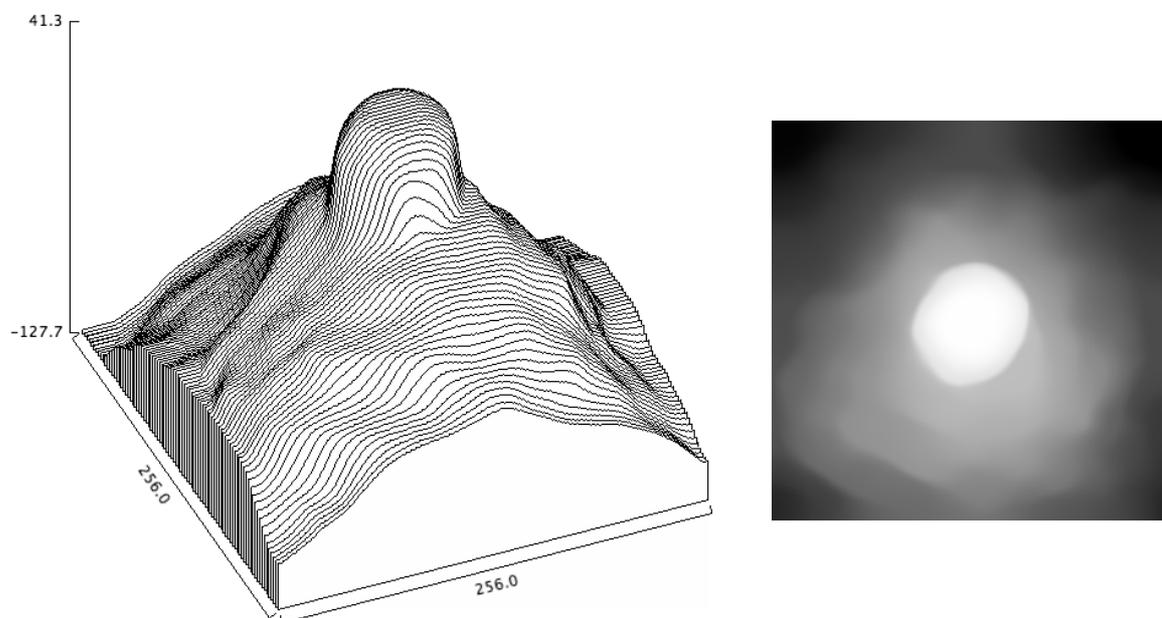


Figure 3.34: Level-set function ϕ : the value of ϕ is plotted over the image plane (left) and depicted as intensity image in top-view (right). We can see how the function has evolved and deviated from the original regular signed distance function. This deformation is due to the image based energy terms and allows extracting meaningful contour candidates at levels $\phi \neq 0$.

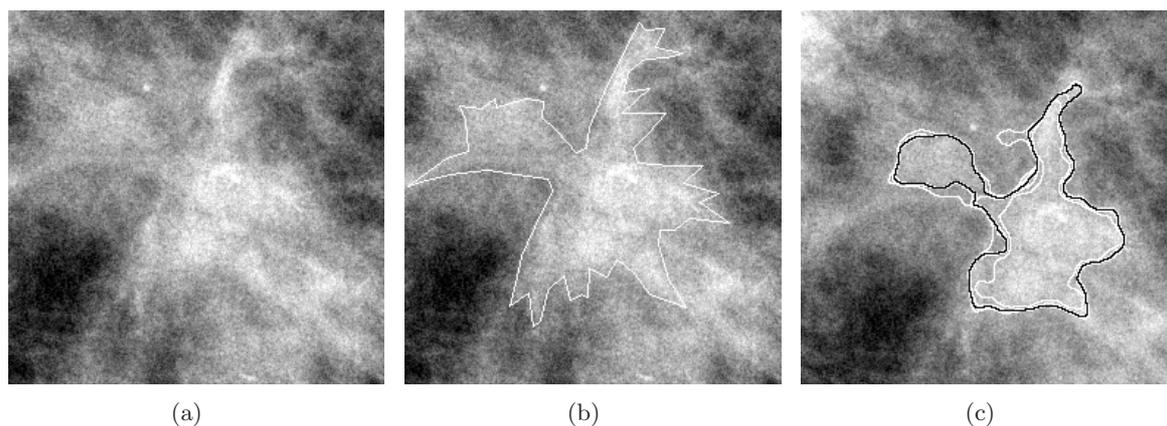


Figure 3.35: Results for the fuzzy active contour model (spiculated mass model) applied to an exemplary clinical case containing a spiculated breast mass: (a) original image and (b) reference contour overlaid on the original image. (c) the candidate contours extracted with the fuzzy active contour model are illustrated as an overlay on the original image. The black contour corresponds to the 0-level-set at convergence.

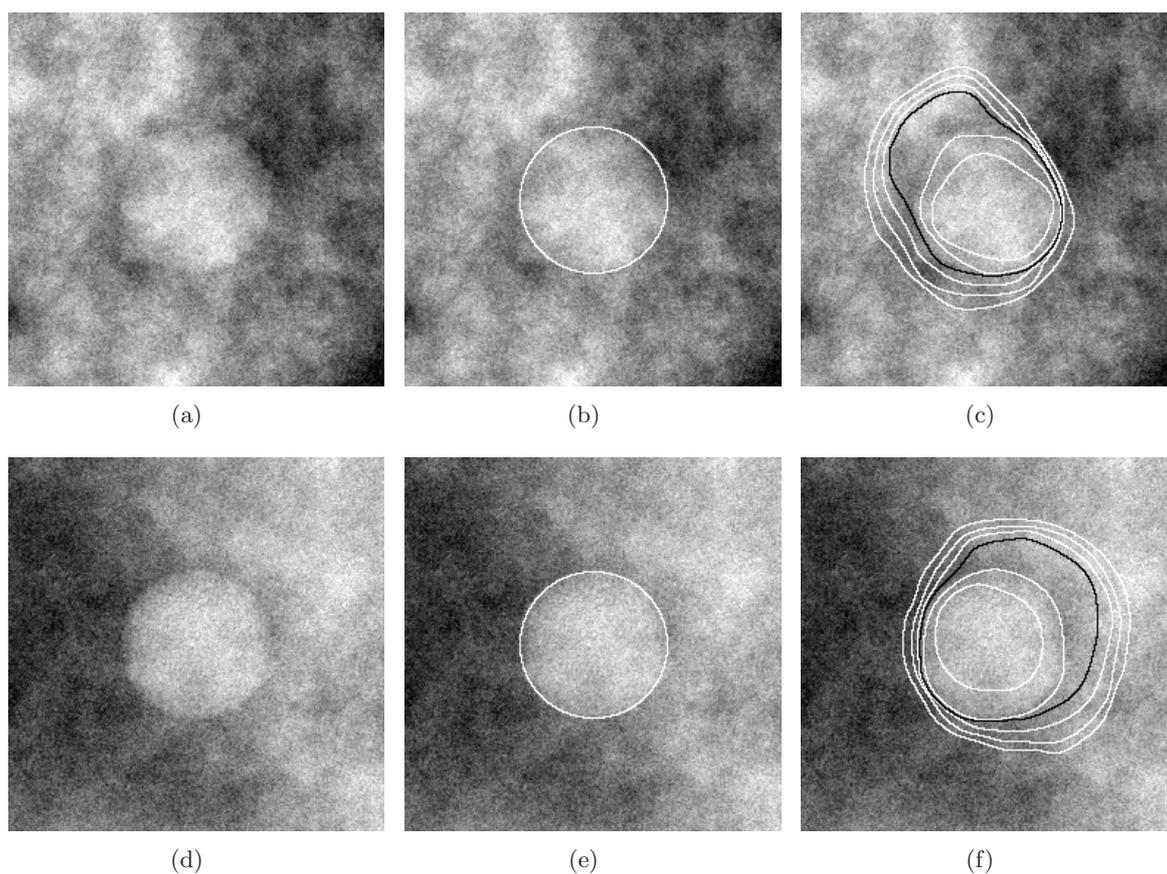


Figure 3.36: Results for the fuzzy active contour model (circumscribed mass model) applied to a simulated images containing circumscribed breast masses: (a,d) original image and (b,e) reference contour overlaid on the original image. (c,f) the candidate contours extracted with the fuzzy active contour model are illustrated as an overlay on the original image. The black contour corresponds to the 0-level-set at convergence.

3.4 Feature Extraction

Feature extraction has been extensively used in computer-aided detection on mammographic images. A wide selection of attributes has been tested for masses as well as for microcalcification detection (Bothorel 96; Cheng 03; Cheng 06). Feature extraction essentially aims at characterizing an object. Information about a given object is usually desired for classification purposes. We then aim at extracting a feature vector for a given object in order to compare this information with a given prototype, class or distribution. One of the problems of this technique stems from the fact that it is difficult to assess to which degree the extracted features actually characterize the object of interest. Ambiguity in the image translates to the segmentation result, which in turn may have a significant impact on the feature values. Fuzzy attributes have been introduced to model the ambiguity of the data and to be able to postpone the decision until a final decision is made (e.g. by a classifier). The present section is dedicated to fuzzy attributes and their properties.

3.4.1 Extraction of 2D Attributes

A number of features are extracted for each candidate contour in order to characterize the segmented object. Two groups of features are extracted: shape features and intensity/gradient features.

The shape features we are using are *area* (inside the contour), *perimeter* (of the contour), *Ferret ratio*, and *compactity*.

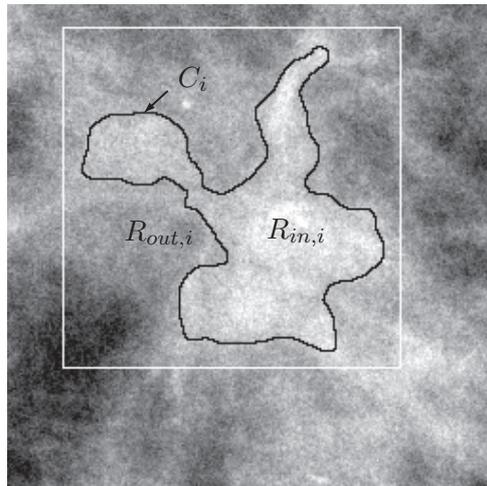


Figure 3.37: Computation of attributes. Grey-level and gradient based features are computed for three different regions of interest. The first region $R_{in,i}$ contains the pixels enclosed by the candidate contour (in black). The second region contains the pixels of the candidate contour C_i . The third region (white box) contains all the pixels in a region containing the contour ($R_{in,i} \cup R_{out,i} \cup C_i$).

Furthermore, eight histogram-based gradient and gray-level features are computed. Three regions are identified as illustrated in Figure 3.37. We compute the mean gradient along the contour as:

$$g_{C_i} = \frac{1}{N_{pt}} \sum_{n=1}^{N_{pt}} \nabla I(C_i(n)), \quad (3.23)$$

where N_p is the number of points in the contour and $\nabla I(C_i(n))$ is the gradient at the n th point

of C_i .

Another feature is based on the homogeneity of the intensity values in $R_{in,i}$. Furthermore, we compute the mean and the variance of the gradient direction in each of the three regions specified in Figure 3.37. These are features that have been widely used for characterization of masses since they have proven a good indicator of converging fibers (spicules).

An example of feature extraction for a set of candidate contours obtained from the fuzzy active contour model on a spiculated breast mass is illustrated in Figure 3.38. Feature values are drawn as a function over the candidate contours. The compactness seems to be a good attribute to characterize the spiculated form of the lesion, while the Ferret ratio seems poorly adapted for this task. This is not surprising since a very irregular contour of a spiculated mass may have identical values for the minimum and maximum Ferret. The interpretation of the other extracted attribute functions is less intuitive, especially since extracted values are not normalized and are difficult to analyze without comparison to values for other objects.

3.4.2 Computing 2D Fuzzy Attributes

Once the fuzzy contour of each particle marked with a marker M_I has been built, we are interested in differentiating particles that are potential radiological findings from particles that may correspond to normal breast structures. We may also need to distinguish between different radiological findings. We therefore suggest using fuzzy attributes (Bothorel 96) measured on the fuzzy contours as input of a classifier (e.g., a fuzzy decision tree) to achieve this classification.

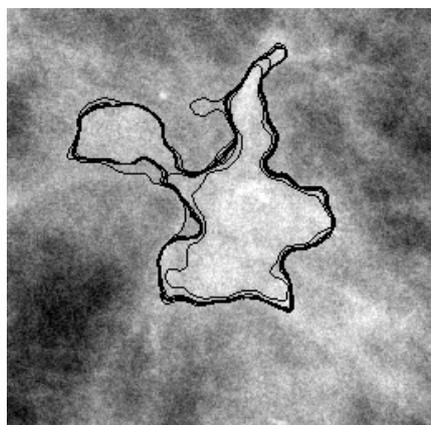
For each segmented particle in the image I , we have a fuzzy contour \mathcal{C} characterized by its membership function $\mu_{\mathcal{C}}$ defined on the universe \mathcal{X} of the candidate contours C_i . For each candidate contour C_i , we can therefore calculate the value $a_j(C_i)$ of the attribute a_j (e.g., compactness), defining an application φ from the universe \mathcal{X} of the candidate contours C_i to the universe \mathcal{Y}_j of the attribute values a_j . Finally, for each fuzzy contour \mathcal{C} , we build the fuzzy attribute $\mathcal{A}_{j,\mathcal{C}}$ of \mathcal{Y}_j characterized by the membership function $\mu_{\mathcal{A}_{j,\mathcal{C}}}$ using the extension principle (Figure 3.39):

$$\forall a_j \in \mathcal{A}_j, \quad \mu_{\mathcal{A}_{j,\mathcal{C}}}(a_j) = \begin{cases} \sup_{C_i \in \mathcal{X} | a_j = \varphi(C_i)} \mu_{\mathcal{C}/\mathcal{A}_k}(C_i) & \text{if } \varphi^{-1}(a_j) \neq \emptyset \\ 0 & \text{if } \varphi^{-1}(a_j) = \emptyset \end{cases} \quad (3.24)$$

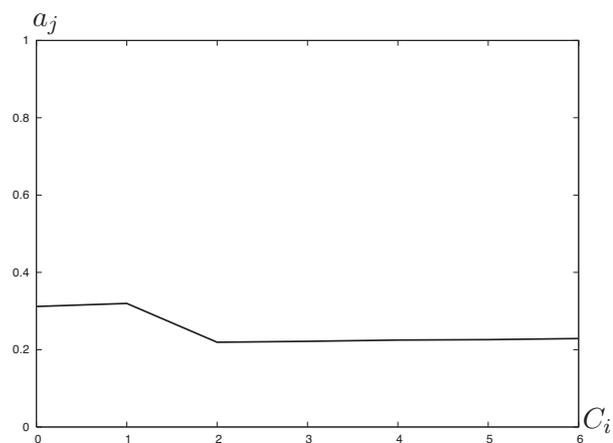
With fuzzy attributes, we keep the uncertainty and the imprecision of the particles in the values of the fuzzy attributes. The larger the support of the membership function, the larger the imprecision. The uncertainty leads to multiple lobes in the membership function.

Figure 3.40 shows an example for computing fuzzy attributes on a fuzzy active contour. The principle described in Figure 3.39 is applied to a clinical case containing a circumscribed breast mass. The membership value associated to each candidate contour results in a membership function for the fuzzy contour to the class *contour of a circumscribed breast mass* (Figure 3.40 (b)). The extracted attribute *area* is monotone and increasing as we would expect for a ordered set of contours such as the contours in Figure 3.40 (a). Using the extension principle (Equation 3.24) we derive a membership function for the fuzzy attribute *area*. This membership function can be interpreted as follows: for each attribute value (x -axis) we note how much this value is representative for the object characterized by the extracted fuzzy contour (membership degree on the y -axis).

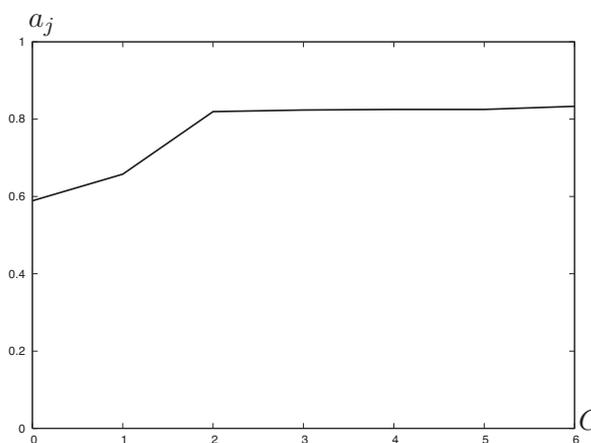
In Figure 3.40 the membership function of the attribute *area* has its maximum at $a = 4100$. This corresponds to the candidate contour C_2 which is characterized by a membership degree to the class *contour* $\mu_{\mathcal{C}} = 1$.



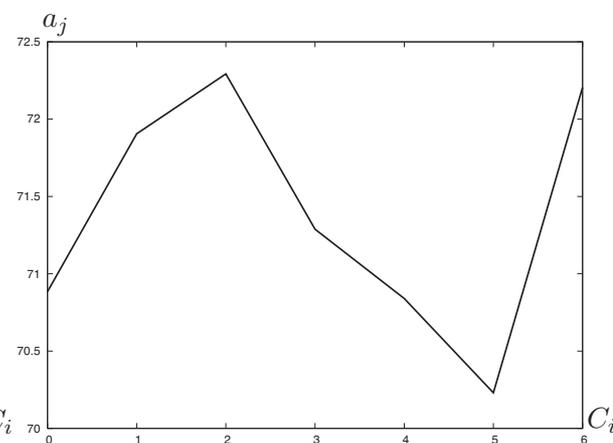
(a) fuzzy contours



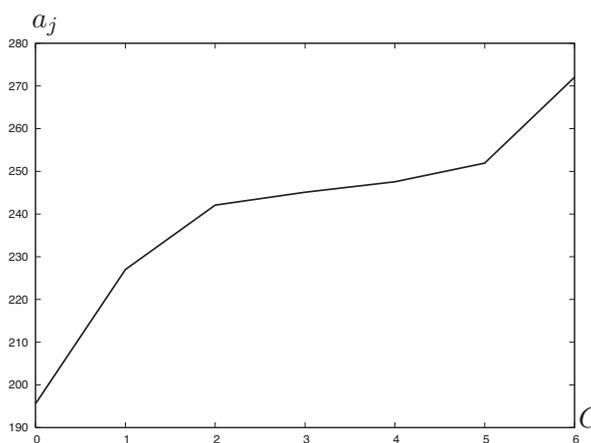
(b) attribute *compactness*



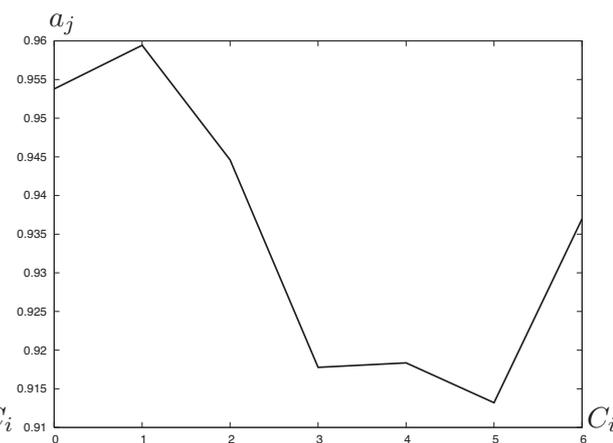
(c) attribute *Ferret ratio*



(d) attribute *mean gradient along contour*



(e) attribute *homogeneity*



(f) attribute *mean gradient direction along contour*

Figure 3.38: Example of feature extraction. Features are extracted separately for each candidate contour. For each feature the curve shows the value taken for each candidate contour (where C_0 corresponds to the innermost candidate contour and C_6 corresponds to the outermost candidate contour).

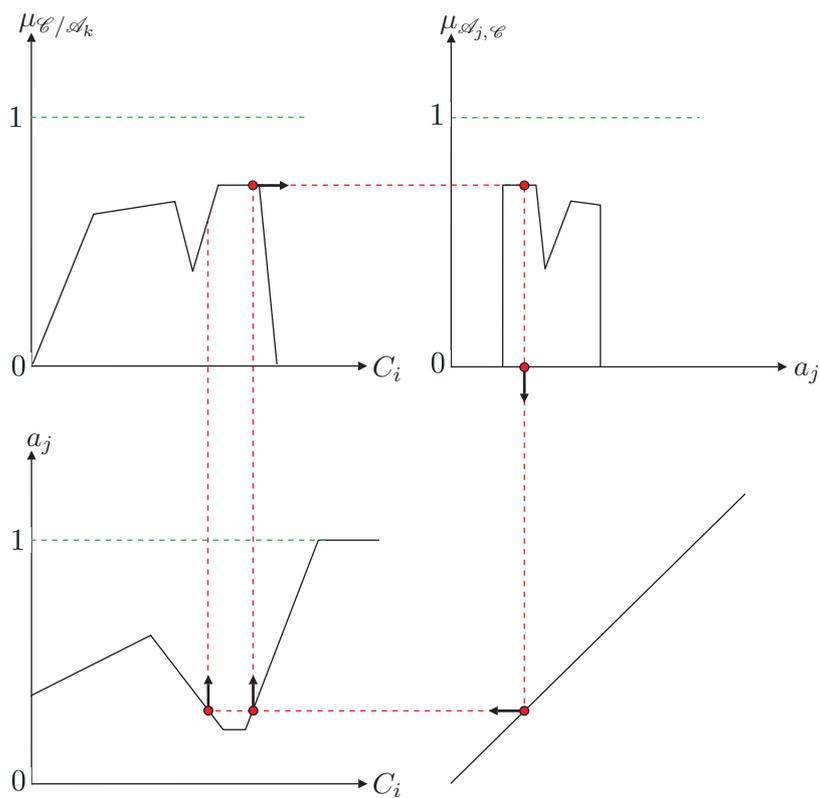


Figure 3.39: For a given particle, the value of the attribute (e.g., compactness) is measured for each candidate contour (bottom left). From the particle segmentation step, we know already the fuzzy subset *contour* relative to the criterion characterized by the fuzzy subset \mathcal{A}_k (top left). Once the value of the attribute has been measured for each candidate contour (bottom left), the extension principle is applied to derive the fuzzy attribute \mathcal{A}_j relative to the criterion characterized by the fuzzy subset \mathcal{A}_k for the current particle (top right).

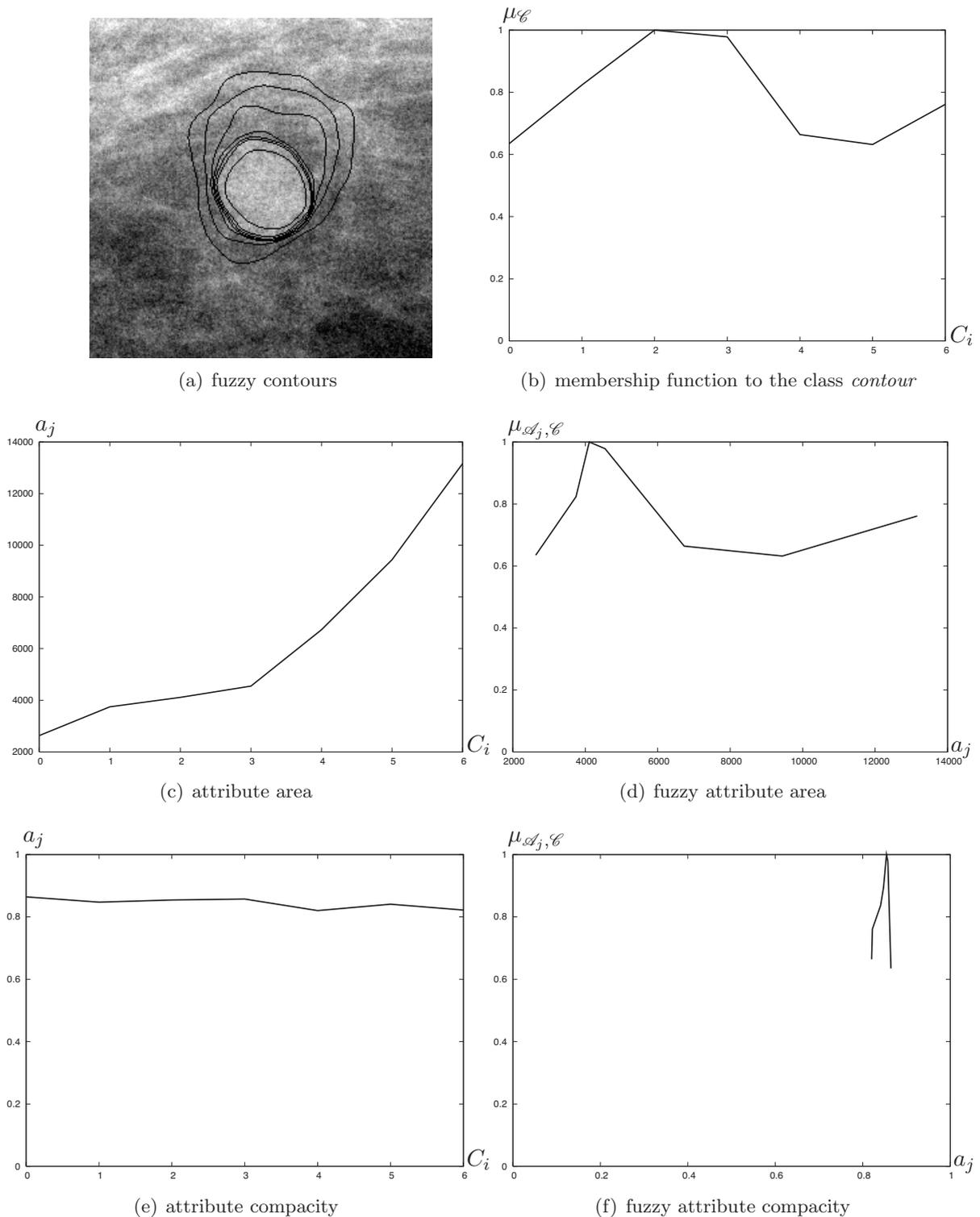


Figure 3.40: Example of fuzzy attribute extraction. Based on the energy of each candidate of the segmentation (a), we compute a membership degree to the class *contour* (b) (where C_0 corresponds to the innermost candidate contour and C_6 corresponds to the outermost candidate contour). The resulting membership function serves to compute fuzzy membership functions ((d) and (f)) for the attributes *area* and *compactness* ((c) and (e)). For the attributes ((c) and (e)) the feature value is drawn for each candidate contour. The fuzzy attributes ((d) and (f)) are drawn as membership values for each attribute value.

3.5 Conclusion

In this chapter we have presented a new hybrid active contour model for mass contour detection in digital breast tomosynthesis projection images. The combination of region-based and edge-based energy terms provides an improved segmentation performance when compared with active contour models based on a single image feature. To include additional image-based terms may further improve the performance of the algorithm. In particular, a texture-based term promises a significant impact due to the variety of textures present in mammographic images. Furthermore, the edge-based energy term may be refined. Today, this criterion is realized as a stopping term that has no impact outside of the close neighborhood of the edges. A vector diffusion scheme over the entire image domain (e.g. Gradient Vector Flow (Xu 97; Xu 98)) may be better adapted to be combined with the region homogeneity term that takes values in the entire image domain. Recently, research for detecting objects from partial contours has been published (Zhu 07) that seems an interesting approach for our data where object margins are often partially obscured by overlapping tissue or degraded by noise.

The hybrid active contour model has been extended for extracting fuzzy contours. An original approach for obtaining a set of candidate contours from a single evolving function ϕ has been developed. Our fuzzy segmentation technique effectively overcomes the problem that the active contour model converges at local minima before (or sometimes after) reaching the actual object boundary. Since the properties of the object of interest are included in the energy functional of the active contour model and the parameters are tuned accordingly, we can derive the membership degrees to the different candidate contours directly from the energy levels of the candidates. A method to automatically identify the levels ϕ_{l_i} for extraction of the candidate contours appears to be an interesting topic for further research. This could for example be based on a morphological analysis of the function ϕ . Furthermore, the computation of the membership degree to the class *contour* should be further investigated. Today, the bounds of the function $\mu_{\mathcal{C}/\mathcal{A}_E}$ are obtained empirically. With a larger clinical database at our disposal, this step may be replaced by a machine learning approach. The mammographic projection data is provided with a 16 bit image dynamic (14 bits used). For reasons of simplicity, the algorithm is only working on 8-bit images today. Furthermore, an extension to three dimensions would allow us to test the algorithm directly on the reconstructed breast volume.

Finally, a novel framework for extracting multiple fuzzy contours for an unknown object of interest has been introduced. Different fuzzy active contour models are applied to a given object of interest. Each model has been built based on a priori knowledge about a different kind of objects. Interesting extensions to this framework consist in a larger set of model classes (architectural distortions, breast fibers, . . .), and in refining the existing object definitions (where the same improvements as mentioned above for the hybrid active contour model apply).

*Falsehood has an infinity of combinations,
but truth has only one mode of being.*

Jean-Jacques Rousseau

4

Information Fusion

In the preceding chapter we have discussed the extraction of different kinds of information from the image. This information is extracted from each two-dimensional image of the projected views data set. In the present chapter we deal with finding an operator that allows us to pass from this two-dimensional representation of the DBT data to a three-dimensional representation.

Generally, in tomography or tomosynthesis, this is achieved by applying one of the reconstruction techniques that have been specifically designed for this purpose. We will give an overview of the most popular reconstruction techniques in Section 4.1. In our framework, feature extraction is performed on the projected views. This increases the dimensionality of the data to be fused compared to reconstruction of the intensity images. Classical reconstruction techniques cannot be applied without modification since they have been designed to process intensity images where a single value is assigned to each pixel. Furthermore, the presence of fuzzy contours needs to be taken into account when we combine the set of projected views. In Section 4.2 we present two novel approaches that propose solutions to this problem. *Partial defuzzification* aims at processing the projection images in a fashion that allows for back-projection-type operators to be used for passing into 3D space. The particle-based fusion of fuzzy data follows a different approach. We simply try to establish a correspondence between a given voxel and a set of pixels in the projection images. No volume to be displayed is computed. In fact, the obtained sets serve as input to a classifier that is described in Chapter 5.

4.1 Image Reconstruction

Image reconstruction is the basic operation performed in three-dimensional imaging. It allows combining a set of projected views in order to obtain a reconstruction of the original volume. Any preprocessing applied to the projected views should only serve to obtain an inversion of the projection process. In the following sections we discuss the processing of fuzzy data and feature vectors. Even though this represents a reconstruction-type operation, it is not a reconstruction in the strict sense, and we will therefore rather discuss aggregation strategies and back-projection-type operators.

Several comparisons of different reconstruction algorithms for tomosynthesis data have been conducted in the past (Dobbins 87; Wu 04a; Zhang 06). According to (Zhang 06), existing reconstruction methods for digital tomosynthesis, or equivalent limited-angle cone-beam tomog-

raphy, can be classified into four categories, including back-projection algorithms, transform algorithms, algebraic reconstruction techniques, and statistical reconstruction techniques.

4.1.1 Back-projection Methods

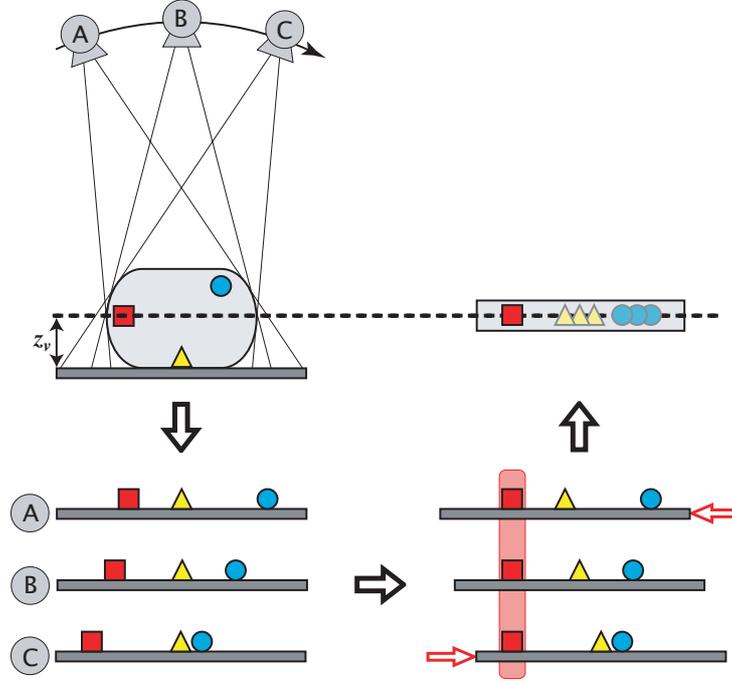


Figure 4.1: The principle of shift-and-add reconstruction: a series of N_p projection images is acquired (bottom left) for tube positions A , B , and C (top left). To reconstruct a slice of a given height z_v above the detector, each projection needs to be shifted by $\xi(z_v)$ and $\eta(z_v)$ which are the shift factors in x and y direction respectively (bottom right). The shifted projections are then summed to form the reconstructed slice.

The conventional tomosynthesis reconstruction algorithm is called *shift-and-add* (SAA). The basic principle of SAA is illustrated in Figure 4.1. The value for a voxel at position (x_v, y_v, z_v) is given by:

$$V(x_v, y_v, z_v) = \frac{1}{N_p} \sum_{k=1}^{N_p} I_{org,k}(x_v + \xi_k(z_v), y_v + \eta_k(z_v)) \quad (4.1)$$

where $V(x_v, y_v, z_v)$ is a voxel of the reconstructed volume, $I_{org,k}$ is the k th projection image, N_p is the number of projections and $\xi(z_v)$ and $\eta(z_v)$ are the shift factors in x and y direction respectively.

Shift-and-add does hence not include a filtering step and is a strictly linear operator. However, the SAA algorithm is valid only when the motion of the X-ray source is parallel to the detector (corresponding to a linear motion at a fixed height above the detector). Given a rotational tube motion as in DBT, different features in a selected plane cannot be simultaneously registered by simply shifting and adding the projections.

Instead, the shift required to register an object point depends on the in-plane location of this point. A back-projection (BP) algorithm, also called simple back-projection (SBP), accurately incorporates the imaging geometry no matter how the X-ray tube is moved. The equation for

this technique is very similar to Equation 4.1. The difference lies in the additional scaling factor $s_k(z_v)$. The equation thus becomes:

$$V(x_v, y_v, z_v) = \frac{1}{N_p} \sum_{k=1}^{N_p} I_{org,k}(s_k(z_v) \cdot x_v + \xi_k(z_v), s_k(z_v) \cdot y_v + \eta_k(z_v)) \quad (4.2)$$

If the motion of the focal spot is parallel to the detector, the scaling factor $s_k(z_v)$ is the same for all projections. When we compare Equation 4.1 and Equation 4.2, it is apparent that as the motion of the X-ray focal spot is parallel to the detector, the SAA reconstruction is equivalent to the SBP except for a constant scaling factor.

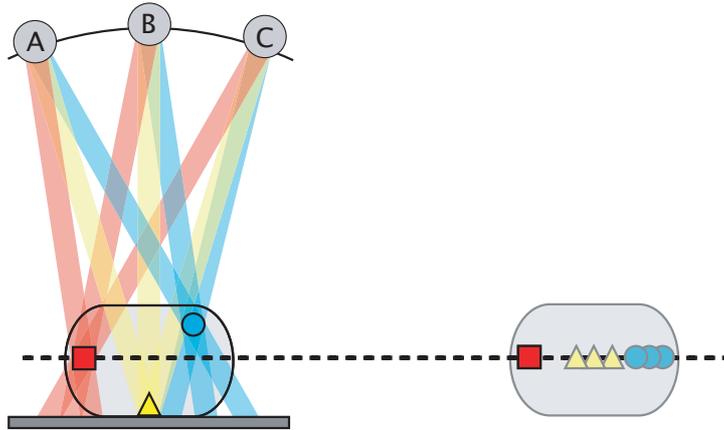


Figure 4.2: Limitations of the acquisition geometry: the limited angular range available in digital breast tomosynthesis results in artifacts in the reconstructed volume (see Section 2.2.3). An object is *in focus* only in the slice that runs through the center of the object. In other slices, the object will be *out-of-plane* and will thus appear as multiple low-contrast copies (right). These streak artifacts (left) through the reconstructed volume are a consequence of the limited angular range of the acquisition system.

The SAA method is thus a simplified version of the SBP method. Some additional strategies have been employed with the SBP method to reduce the streak artifacts (out-of-plane artefacts, interplane artefacts), which are typically very strong in SBP reconstructed images (see Figure 4.2). Especially high-contrast features tend to produce strong artifacts. One of the strategies applied to reduce these artifacts include the use of order statistics operators (Claus 02). Instead of averaging the pixel values from the projection images that contribute to a specific voxel position, these values are ordered first. Then, during the averaging, the minimum and maximum values (one or several) are not taken into account. This allows to efficiently eliminate a considerable amount of streak artifacts. An illustration of this technique is given in Figure 4.3.

4.1.2 Transform Algorithms

Transform algorithms include the filtered back-projection (FBP) method and other transfer function methods. In the FBP method (Lauritsch 98; Mertelmeier 06), the Fourier slice theorem plays a fundamental role and projection images are transformed to the spatial frequency domain. With a parallel-beam approximation, the 2D Fourier transform of a projection image acquired at a given angle corresponds to one slice sample in the spatial frequency domain at this angle. Thus, multiple projection images can be combined to obtain a discrete sampling of the whole imaged volume spectrum.

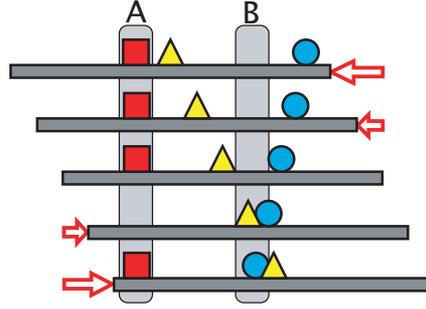


Figure 4.3: Order statistics based back-projection: OSBP is an attempt to cope with the limitations of the limited angular range in DBT. This reconstruction technique is based on SBP. However, instead of computing the average for each pixel after the shift operation, an order statistics operator is applied. A ranking filter sorts the values corresponding to a given voxel and the N_{max} highest as well as the N_{min} lowest values are not taken into account when computing the average. This way, projections where the signal of the object is corrupted by noise (A) or where out-of-plane structures appear (B), are not taken into account. This non-linear operation quite effectively reduces the streak artifacts in the resulting volume.

For 3D cone-beam breast tomosynthesis, the Feldkamp algorithm (FDK) (Feldkamp 84), an approximation of the the cone-beam FBP algorithm, has been investigated by (Wu 04a), but the authors in (Wu 04a) found the reconstruction results to be very noisy and the structure details to be poorly visible.

In other transfer function methods, impulse response functions or point spread functions are specifically designed based on the imaging geometry and the application. One example of this type of approach is matrix inversion tomosynthesis (MITS) (Dobbins 87) that has been studied for different X-ray modalities.

4.1.3 Algebraic Reconstruction Techniques

In algebraic reconstruction techniques, the tomographic inverse problem is formulated as solving a large-scale system of linear equations of the following type:

$$\begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_N \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} \rightarrow \mathbf{A}\mathbf{x} = \mathbf{y} \quad (4.3)$$

where \mathbf{A}_n is the projection matrix for the n th projected view, \mathbf{y}_n is the corresponding vector of the projection data, and \mathbf{x} is the matrix of (unknown) linear attenuation coefficients of the voxels. This is the linear system model for the tomosynthesis reconstruction, and it is the basis for the algebraic reconstruction techniques. Each element of the coefficient matrix \mathbf{A} contains for example the intersection path length of one X-ray within one specific voxel. The reconstruction is accomplished by iteratively updating the unknown linear attenuation coefficients by minimizing the error between the measured and the calculated projection data (see Figure 4.4).

The original method in this family of algebraic reconstruction techniques is known by the same name - ART (Gordon 70). In ART, the linear attenuation coefficients are updated in a *ray-by-ray* manner. All voxels along the ray under consideration are updated by the difference between the detected and the computed value. Since only one projection value is used to update the linear attenuation coefficient at a time, ART has fast convergence speed but will converge

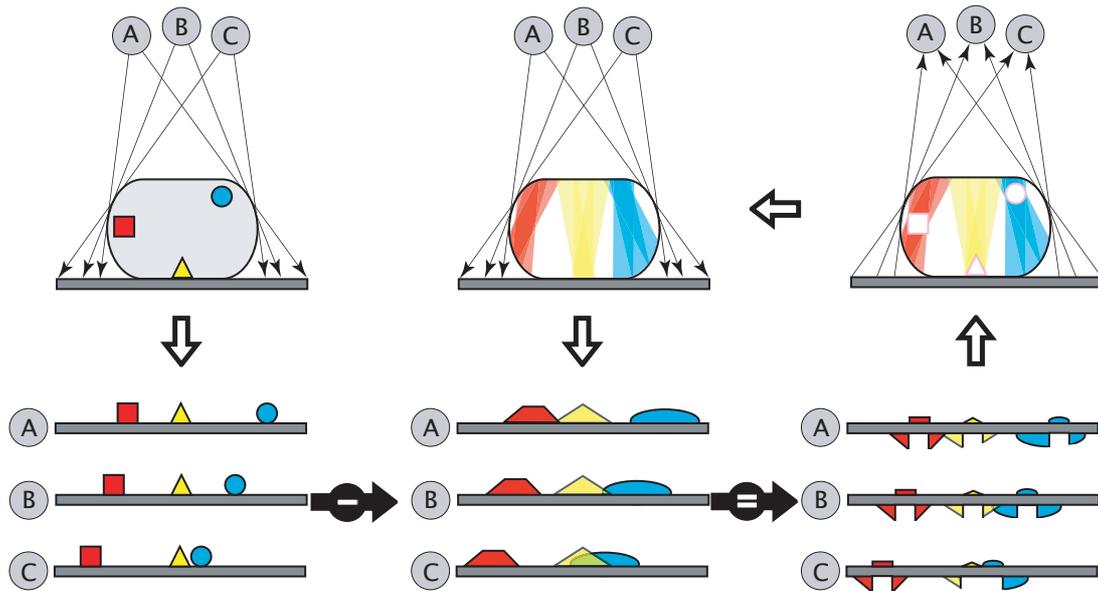


Figure 4.4: Principle of iterative reconstruction techniques. Here, the original acquisition is depicted on the left (original volume on the top and projected views on the bottom). The volume of the attenuation coefficients is usually initialized with a SBP reconstruction result (top center). A single iteration consists of three steps: back-projection (top right), re-projection (top center) and computation of the reconstruction error (bottom row). After each back-projection/re-projection operation the resulting projection images (bottom center) are compared to the original tomographic projected views (bottom left). From this comparison difference images are computed (bottom right) that serve to update the reconstructed volume. This iterative process is performed until the algorithm converges and a given stopping criterion is met.

to a (least squares) solution which can be very noisy for severely ill-posed inverse problems such as limited-angle tomosynthesis reconstruction (Zhang 06).

To improve the ART method, variations on its implementation have been proposed. In SART (Andersen 84), the linear attenuation coefficient of each voxel is updated based on all rays traversing this voxel. This has proven to be a good trade-off between a noisy solution and a slow convergence speed.

4.1.4 Statistical Reconstruction Techniques

All reconstruction algorithms discussed so far are deterministic methods. On the other hand, statistical reconstruction methods tackle the inverse problem from a statistical point of view, assuming the unknown attenuation coefficients as a random variable following some specific probability distribution function. An example of this type of method is the maximum likelihood method (ML). Since there is no analytic solution to the logarithm likelihood function, it is difficult to search the entire space of unknown sets to find the ML solution. Various algorithms have been investigated to maximize the log-likelihood function iteratively, such as the expectation maximization (EM) algorithm (Wu 03a), convex algorithm (Wu 04b), and gradient algorithm (Lange 95).

4.2 Aggregation of Fuzzy Data - A New Pixel-Based Approach

The present section is dedicated to finding a fusion operator that allows for combining the entire range of information extracted in the processing steps as described in Chapter 3. This problem is illustrated in Figure 4.5. In the preceding section, we have discussed methods that are specifically designed to process intensity images and to find a solution to an unknown matrix of attenuation coefficients. In our case, any pixel may be associated not only to a single intensity value but to an entire vector of values. This results from the fact that any pixel in a projection image may be enclosed by one or several candidate contours. For each of these contours a feature vector has been extracted. This represents additional information about the image that the fusion operator needs to take into account (see Figure 4.6).

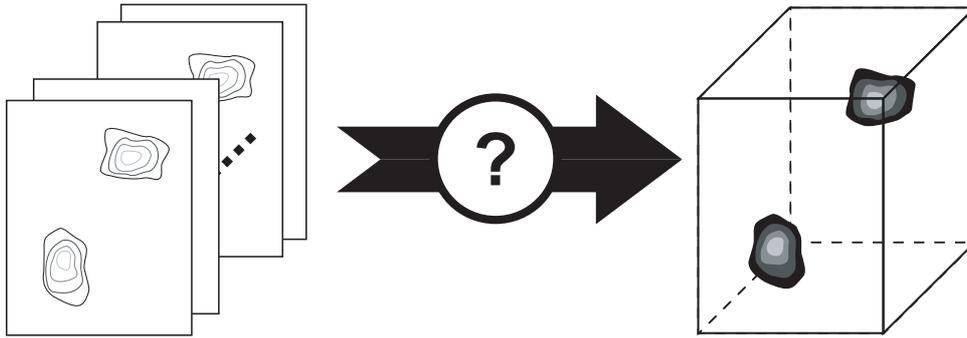


Figure 4.5: Fusion of fuzzy image data: An operator is needed to pass from a two-dimensional representation of fuzzy particles (fuzzy contours) to a three-dimensional representation. Fuzzy contours are illustrated with a gray level corresponding to their membership degree to a given class of objects (left). The objects in the volume (right) are intended to represent 3D fuzzy particles resulting from the fusion.

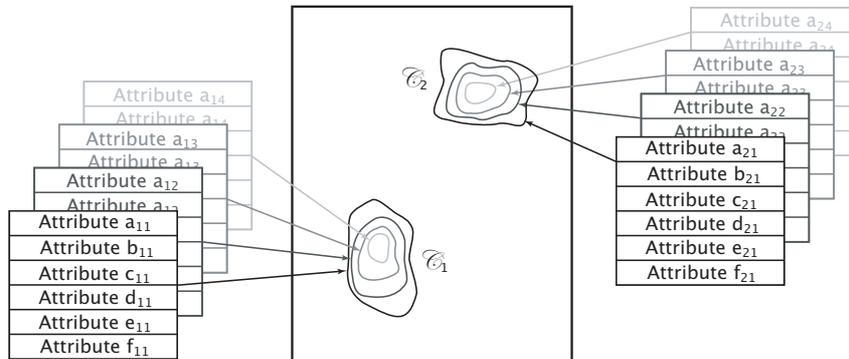


Figure 4.6: Dimensionality of fuzzy images: in an image containing fuzzy contours a set of values may be attributed to each pixel or set of pixels (e.g. contour). Each of these values represents a characteristic associated with the pixel or the set. These may be feature values extracted for a contour. Furthermore, the membership value of a contour to a given class may be stored.

In the following we will present two approaches for coping with this situation. In Section 4.2.1 we develop a method that performs a transformation on the projected views so that each pixel position is associated with a single value. Even though these values do not correspond to an intensity image, this operation allows us to apply a (basic) back-projection-type technique that results in a three-dimensional image volume (Section 4.2.2).

The second method (Section 4.3) follows a different approach. We use information about the geometry of the acquisition system to establish a correspondence between candidate particles

in the different projection images. The combined information from the entire set of projection images is then (potentially together with additional information extracted from a reconstructed volume) fed to a classifier. At that point a decision (e.g. about a given radiological finding corresponding to a cancer) can hence be based on the full set of extracted information, without the need to take intermediate decisions.

4.2.1 Partial Defuzzification

At this point we dispose of a set of images where each pixel may exhibit a large number of associated values. We want to aggregate these images in order to obtain a 3D representation of the extracted information. In the projected views each fuzzy contour consists of a set of candidate contours and their associated membership values to the class *contour*. Aggregating this information directly is not an easy task. If we aim for an equivalent representation in 3D space (e.g. a set of candidate surfaces) we encounter topological difficulties that demand a very complex aggregation strategy. In order to overcome these topological challenges we have chosen to introduce a method to simplify the representation of the fuzzy particles. This step can be seen as a *partial defuzzification*.

In order to transform the available image information into images suitable for a back-projection-type method, we propose to convert the actual contour description into a pixel description (see Figure 4.7). Our aim is to reduce the complexity of the fusion operation. We therefore choose to reduce the complexity of the particle representation to a point where each pixel holds exactly one value. This value should be a summary of all information extracted so far for this position.

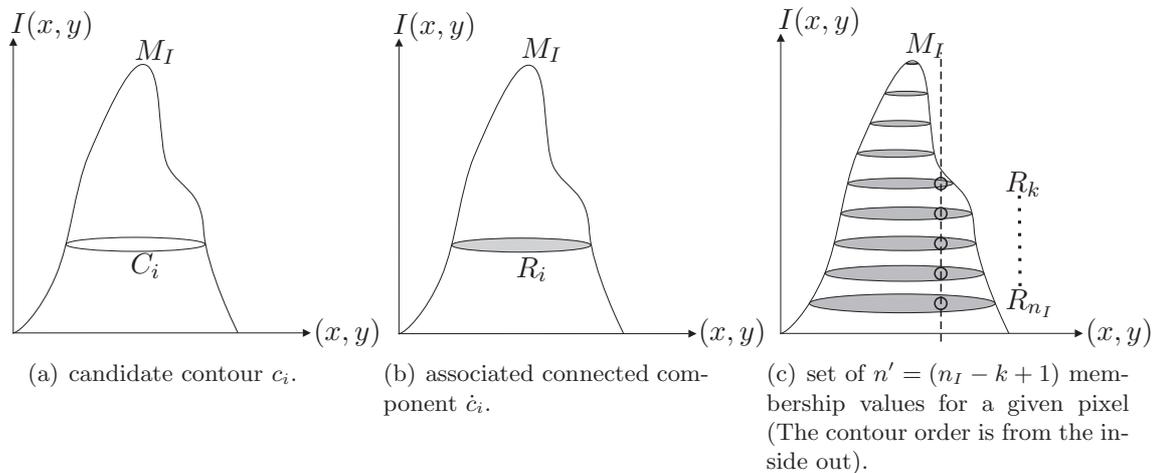


Figure 4.7: Conversion of the contour description into a pixel description.

The transformation we introduce here consists in defuzzifying some but not all aspects of the fuzzy contours, thus reducing the degree of fuzziness of the candidate particles. In Chapter 3 we have defined $Max(I)$ as the set of regional maxima of the intensity image I . The set of contours of a particle marked by the marker $M_I \in Max(I)$ in the intensity image I is given by Equation 3.17. The set of points belonging to the particle is then simply defined as:

$$\mathcal{R}_{M_I} = \left\{ R_i = R_{M_I}(I_{t_i}), t_1 = \max_{M_I}(I) > t_2 > \dots > t_{n_I} \right\} \quad (4.4)$$

where $R_{M_I}(I_{t_i})$ is the region for the set $R_{M_I}(I_{t_i})_{i=1, \dots, n_I}$ that includes the marker M_I of the image I , and I_{t_i} the threshold of I at level t_i .

This set is crisp, and the next step consists in assigning attribute values to each point, based on the information available on the contours. Here, we choose to represent this information by using the membership value of the given candidate contour to the class contour. This membership value is given by Equation 3.21. This value will now be attributed to all pixels of the corresponding connected component (Figure 4.7(b)). Each $R_{M_I}(I_{t_i})_{i=1,\dots,n_I}$ becomes a fuzzy set with the following membership function:

$$\forall (x, y) \in R_i, \quad \mu_{R_i}(x, y) = \mu_{\mathcal{C}}(C_i). \quad (4.5)$$

We have attributed a membership value to all pixels of the connected component that corresponds to a given candidate contour. Therefore, we have for each pixel (x, y) a set of possible membership values $\mathcal{M}_{M_I}(x, y)$ as shown in Figure 4.7(c). The number of elements of $\mathcal{M}_{M_I}(x, y)$ is equal to the number of connected components that include the pixel (x, y) :

$$\exists i \mid (x, y) \in R_i \quad \text{with} \quad R_i \in \mathcal{R}_{M_I}, \quad \mathcal{M}_{M_I}(x, y) = \{\mu_{\mathcal{C}}(C_k), \dots, \mu_{\mathcal{C}}(C_{n_I})\} \quad (4.6)$$

where $n' = (n_I - k + 1)$ is the number of connected components R_i satisfying the condition:

$$(x, y) \in R_i. \quad (4.7)$$

In order to obtain a single value for each pixel position, we need to define an operator that combines the set of membership values $\mathcal{M}_{M_I}(x, y)$ into a single membership value $\mu_{\mathcal{C}}$ per pixel. Since we have no prior knowledge about the *best* contour, we choose to proceed in a cautious manner, which leads us to a disjunctive operator. This type of operator, i.e. a t-conorm, performs a union of all pieces of information (Dubois 80).

Using a t-conorm denoted as \perp , we derive a single membership value for each pixel of a given candidate particle using:

$$\forall (x, y) \in \mathcal{I}, \quad I_{M_I}^f(x, y) = \perp_i \mu_{R_i}(x, y) \quad (4.8)$$

where \mathcal{I} is the set of pixels contained in a tomosynthesis projection image, $I^f(x, y)$ is the resulting two-dimensional fuzzy particle map relative to the marker M_I . It appears obvious that the resulting membership value should not be higher than the highest membership value in the set. We have therefore chosen the max operator, being the smallest t-conorm (see Figure 4.8).

However, since we have allowed contours to belong to different markers at the same time (see Section 3.3.3), we cannot simply combine the fuzzy particles into a resulting image. The same pixel may receive different resulting membership values with respect to different candidate particles (Figure 4.8). In fact, we have to apply the max operator not only to decide between the membership values of competing candidate contours of a given particle, but at the same time to decide between the membership values of a given candidate contour to different candidate particles. Taking this into account, Equation 4.8 becomes:

$$\forall (x, y) \in \mathcal{I}, \quad I^f(x, y) = \perp_m I_m^f(x, y) = \perp_m \perp_i \mu_{R_{m,i}}(x, y) \quad (4.9)$$

where $R_{m,i}(x, y)$ is the i th connected component of the candidate particle marked by the m th marker that contains the pixel (x, y) and $I^f(x, y)$ is the resulting fuzzy particle map. We have thus obtained a representation that is well suited for an aggregation step, that combines information acquired in different tomosynthesis views using a back-projection-type method.

Examples of partial defuzzification applied to clinical data are given for a cluster of microcalcifications in Figure 4.9(b) and for a circumscribed breast mass in Figure 4.10. In Figure 4.9 the entire particles appear in black. This is simply due to the fact, that multi-thresholding was

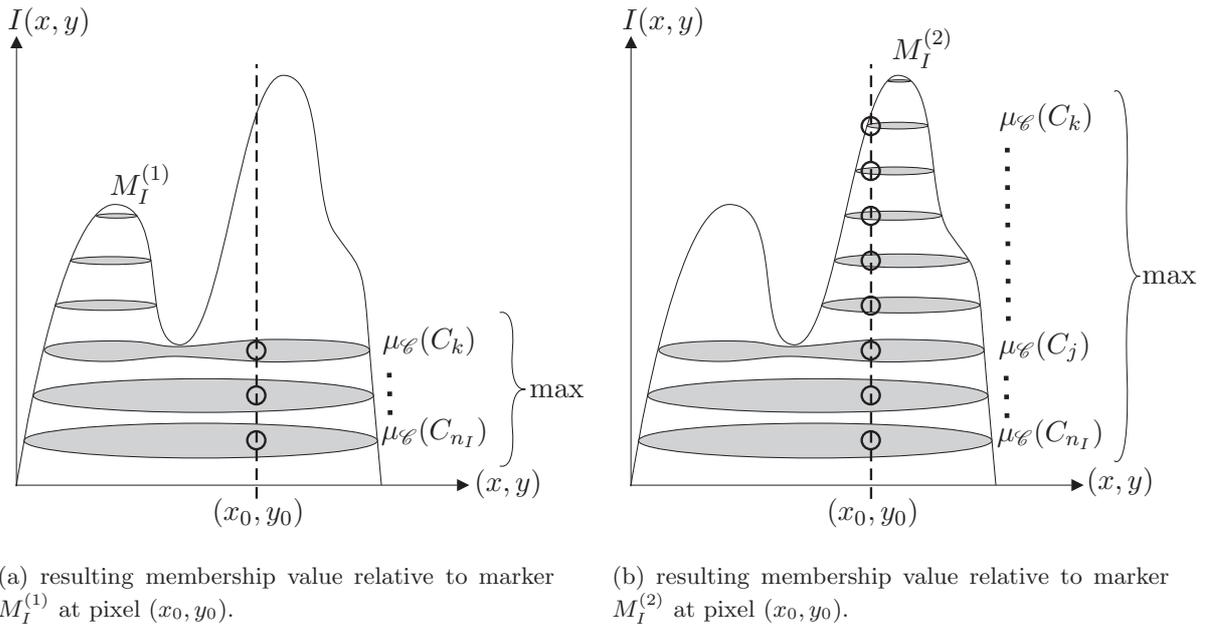


Figure 4.8: Problem of overlapping for contours marked by two different markers $M_I^{(1)}$ and $M_I^{(2)}$.

applied for fuzzy contour extraction. In this case the microcalcifications are so small that almost all the pixels of the particle belong to one of the candidate contours (marked in black).

In Figure 4.10(c) we remark that only two different (non-zero) levels of membership are present in the fuzzy particle map. Since for each pixel we take the max of all membership values of the contours containing the pixel (see Equation 4.8) not all contours will have an impact on the fuzzy particle map. Any contour that is enclosed by a contour with a higher membership degree to the class *contour* will be eliminated by the max operator. In Figure 4.10(c) the candidate contour with the highest membership value was C_3 and the candidate contour (not enclosed in C_3) with the next highest membership degree was C_6 . C_0 - C_2 have therefore been covered by C_3 while C_4 and C_5 have been covered by C_6 .

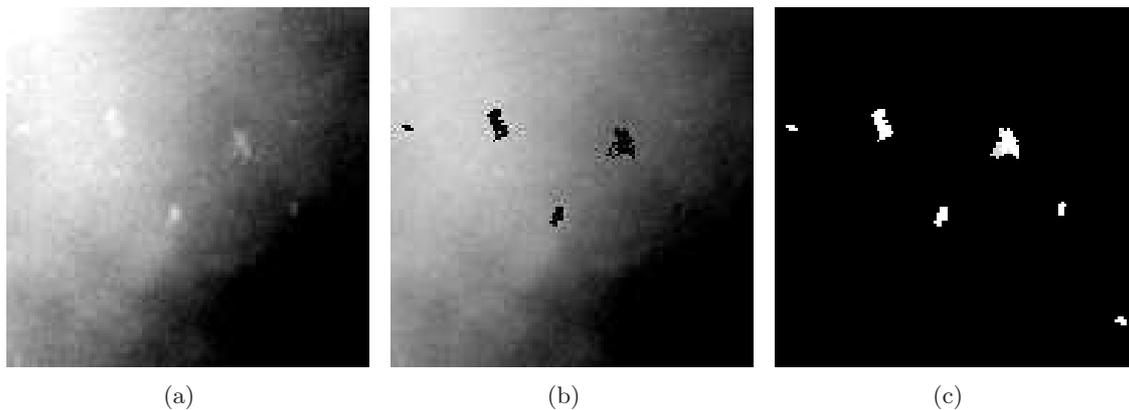


Figure 4.9: (a) selected region of a DBT projected view, (b) extracted fuzzy contours, and (c) the corresponding fuzzy particle map for microcalcification detection. In the fuzzy particle map, higher grey levels correspond to higher membership degrees (black: $\mu_{\mathcal{C}} = 0$, white: $\mu_{\mathcal{C}} = 1$).

Examples of clinical ROI in different projected views (each corresponding to the same 3D ROI) and the set of corresponding fuzzy particle maps are given in Figure 4.11 for microcalci-

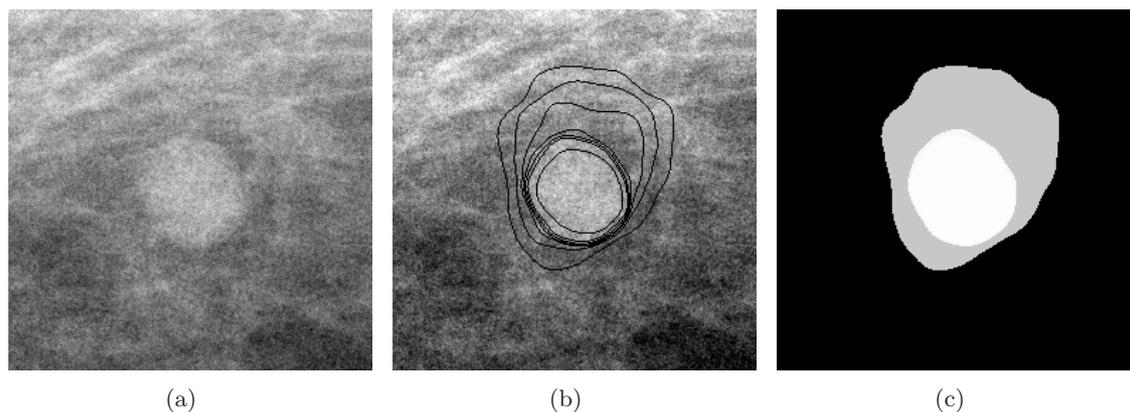


Figure 4.10: (a) selected region of a DBT projected view, (b) extracted fuzzy contours, and (c) the corresponding fuzzy particle map for mass detection.

fication detection and in Figure 4.12 for mass detection. These are good examples for working on different projection images. We can see the impact of the different tube positions on the resulting projected views. The tissue that superposed the objects of interest depends on the view angle. The resulting fuzzy particle maps for different projection images reflect these properties of the data.

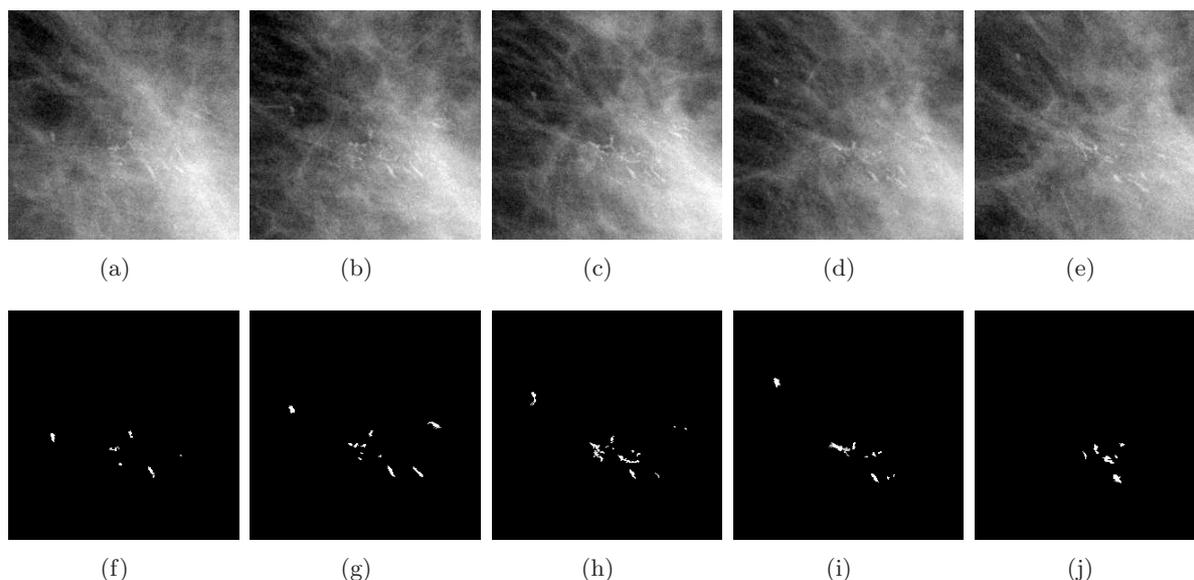


Figure 4.11: Selected region of a DBT acquisition: (a-e) corresponding ROI in different projected views, and (f-j) the corresponding fuzzy particle maps for microcalcification detection.

4.2.2 Aggregating 2D Fuzzy Particle Maps

After performing separate fuzzy detections in each of the N_p projection images of the DBT acquisition, we have generated a so-called fuzzy particle map corresponding to each projected view. In this section we will show how to aggregate the information given by this set of two-dimensional fuzzy particle maps in order to obtain what we call a *fuzzy particle volume*. The spatial correspondence between the set of fuzzy particle maps and the resulting fuzzy particle volume is established through the use of a priori knowledge about the acquisition geometry.

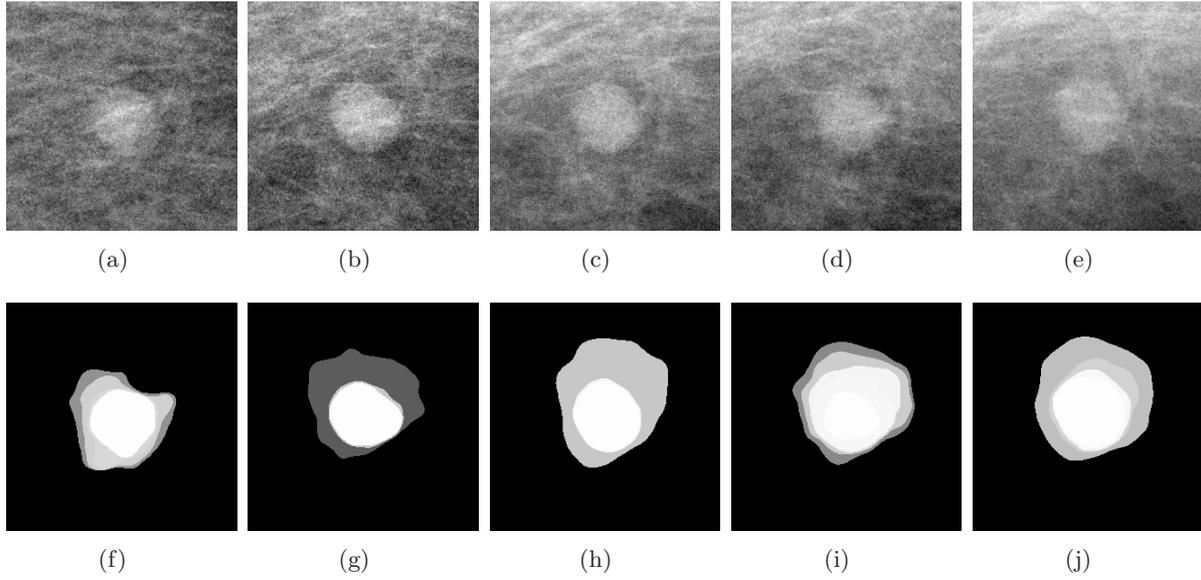


Figure 4.12: Selected region of a DBT acquisition: (a-e) corresponding ROI in different projected views, and (f-j) the corresponding fuzzy particle maps for mass detection.

Our goal is to find for each three-dimensional voxel the corresponding information in all of the N_p fuzzy particle maps that were created. Each voxel (x_v, y_v, z_v) is projected onto the detector plane for each of N_p source positions, resulting in a pixel x_k, y_k in the k th projected view such as:

$$\forall (x_v, y_v, z_v) \in V^f, \quad \Lambda_k : (x_v, y_v, z_v) \mapsto (x_k, y_k), \quad k = \{1, \dots, N_p\} \quad (4.10)$$

where V^f is the fuzzy particle volume and Λ_k is the geometric transformation linking the voxel (x_v, y_v, z_v) with the pixel (x_k, y_k) defined by:

$$(x_k, y_k) = \Lambda_k(x_v, y_v, z_v) = (s_k(z_v) \cdot x_v + \xi_k(z_v), s_k(z_v) \cdot y_v + \eta_k(z_v)) \quad (4.11)$$

with $\xi_k(z_v)$ and $\eta_k(z_v)$ the shift factors in x_v and y_v direction and $s_k(z_v)$ the scaling factor, computed from a priori knowledge about the acquisition geometry.

The aggregation of information gathered in the fuzzy particle maps for a given voxel is then expressed as:

$$V^f(x_v, y_v, z_v) = \Psi_{k=1}^{N_p} \left[I_k^f(x_k, y_k) \right] = \Psi_{k=1}^{N_p} \left[I_k^f(\Lambda_k(x_v, y_v, z_v)) \right] \quad (4.12)$$

where $V^f(x_v, y_v, z_v)$ is the voxel intensity at position (x_v, y_v, z_v) of the fuzzy particle volume, $I_k^f(x_k, y_k)$ is the pixel intensity at position (x_k, y_k) of the k th fuzzy particle map (see Equation 4.9), corresponding to the projection of position (x_v, y_v, z_v) , and Ψ is the aggregation operator. Figure 4.13 illustrates this aggregation operation.

We now have to carefully choose the aggregation operator Ψ . A classification of data fusion operators with respect to their behavior was proposed in (Bloch 96). Three classes of operators are distinguished: Context Independent Constant Behavior (CICB) operators, Context Independent Variable Behavior (CIVB) operators, and Context Dependent (CD) operators. CD operators are not only computed from the information issued by the sensors, but also depend on global knowledge about the sources to be fused. The behavior of CIVB operators is context independent, but their behavior depends on the information issued from the sensors. Since in our case all sources (the different projected views) are equitable, and because we expect a behavior that is the same whatever the values to combine, we have chosen an operator of the CICB

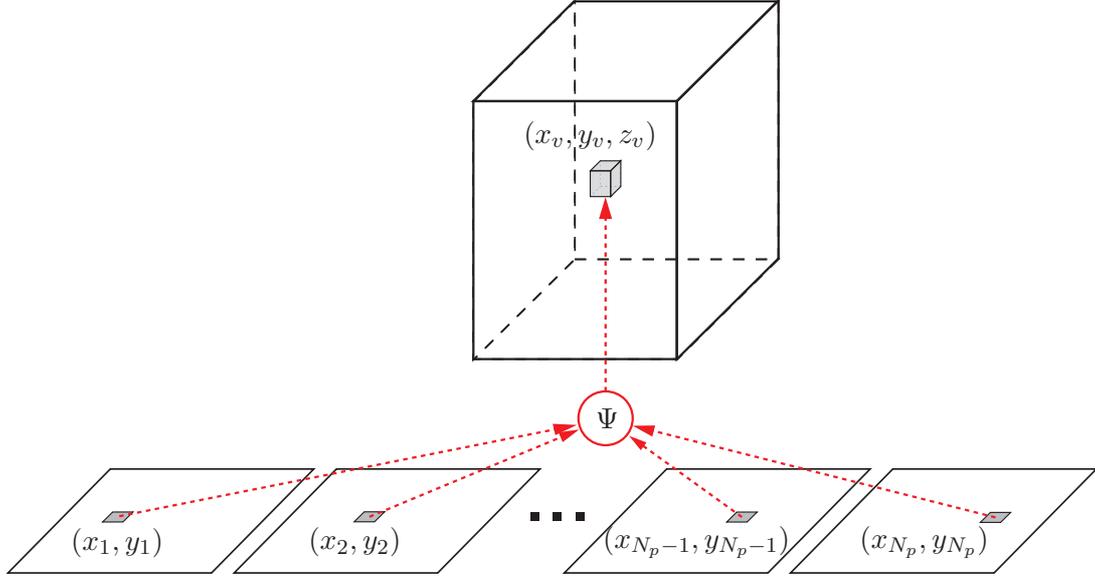


Figure 4.13: Information aggregation strategy: for a given voxel (x_v, y_v, z_v) the information from all corresponding pixels (x_k, y_k) is aggregated using the operator Ψ . The position of the pixel (x_k, y_k) corresponding to the projection of a given voxel (x_v, y_v, z_v) in the k th fuzzy particle map is computed using a priori knowledge about the acquisition geometry.

class. Three types of operators used in fuzzy sets theory are CICB: triangular norms (t-norms), triangular conorms (t-conorms) and mean operators. Within these types of operators, the mean operators are the only ones that always behave like a compromise. For reasons of simplicity, we have chosen the arithmetical mean to be used as aggregation operator in combining the set of fuzzy particle maps. Using this aggregation operator, Equation 4.12 can be rewritten as:

$$\begin{aligned} V^f(x_v, y_v, z_v) &= \frac{1}{N_p} \sum_{k=1}^{N_p} [I_k^f(x_k, y_k)] \\ &= \frac{1}{N_p} \sum_{k=1}^{N_p} I_k^f(s_k(z_v) \cdot x_v + \xi_k(z_v), s_k(z_v) \cdot y_v + \eta_k(z_v)) \end{aligned} \quad (4.13)$$

with the shift factors $\xi_k(z_v)$ and $\eta_k(z_v)$ and the scaling factor $s_k(z_v)$. These factors are known since they are directly related to the acquisition geometry.

Examples of this aggregation method for the clinical ROIs of Figures 4.11 and 4.12 are illustrated in Figures 4.14 and 4.15 respectively. For these results we show the slice of the fuzzy particle volume where the structure of interest is in focus as well as two neighboring slices in each direction to illustrate the evolution in z -direction. For the ROI containing microcalcification we observe the impact of the limited angular view. Each microcalcification is only in focus for a very limited number of reconstructed slices. In the other slices they appear as low contrast copies, so-called streak-artifacts.

Additional clinical examples are depicted in Figures 4.16 to 4.19. The processing in 2D (projected views and fuzzy particle maps) is given alongside the processing performed in 3D space (reconstructed volume and fuzzy particle volume). The displayed ROI in the projections correspond to the same region as the reconstructed slices. Different examples illustrate the processing for microcalcification detection, detection of circumscribed breast masses as well as detection of spiculated breast masses.

From the fuzzy particle volumes we can see the impact of the aggregation operator. Even

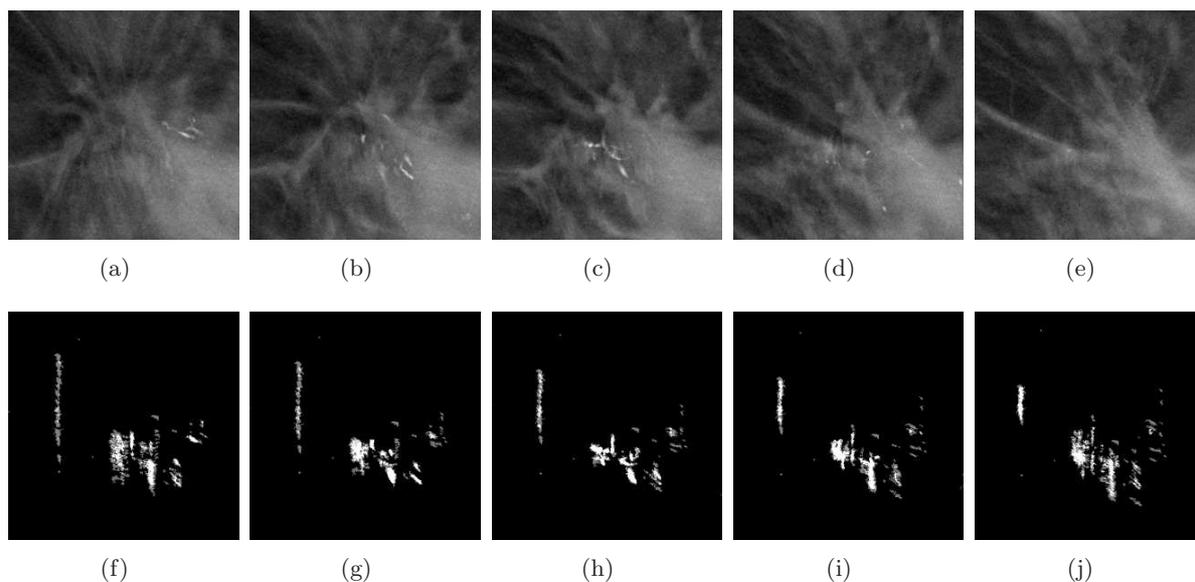


Figure 4.14: Selected region of a DBT acquisition: (a-e) ROI of different slices parallel to the detector plane reconstructed with SART and (f-j) the corresponding ROI from the slices through the fuzzy particle volume for microcalcification detection.

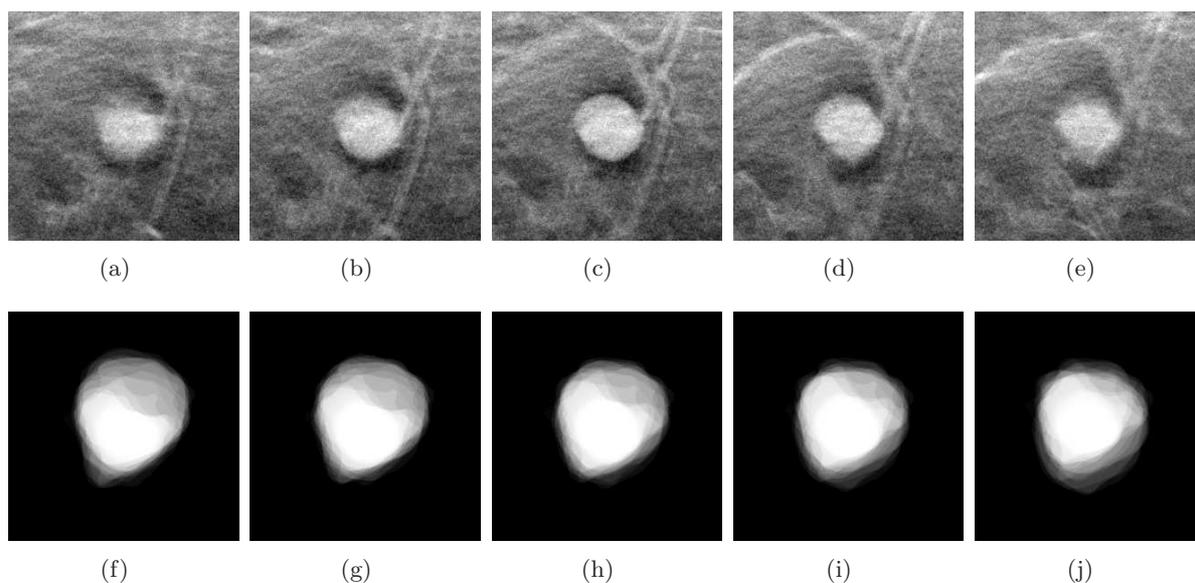


Figure 4.15: Selected region of a DBT acquisition: (a-e) ROI of different slices parallel to the detector plane reconstructed with SART and (f-j) the corresponding ROI from the slices through the fuzzy particle volume for mass detection.

though the segmentation may not be accurate in all of the projected views, the aggregation using the arithmetic mean produces results that correspond well to the reconstructed breast volume.

We have thus established a framework where the geometry of the acquisition system is taken into account and used to back-project the fuzzy particle maps into 3D space. We are however quite flexible regarding the actual aggregation operator. When using the arithmetic mean as in Equation 4.13 we obtain a classical SBP-type operator (see Equation 4.2).

The use of different aggregation operators provides interesting alternatives. Figure 4.20 illustrate results for three different basic aggregation operators: the arithmetic mean, the min

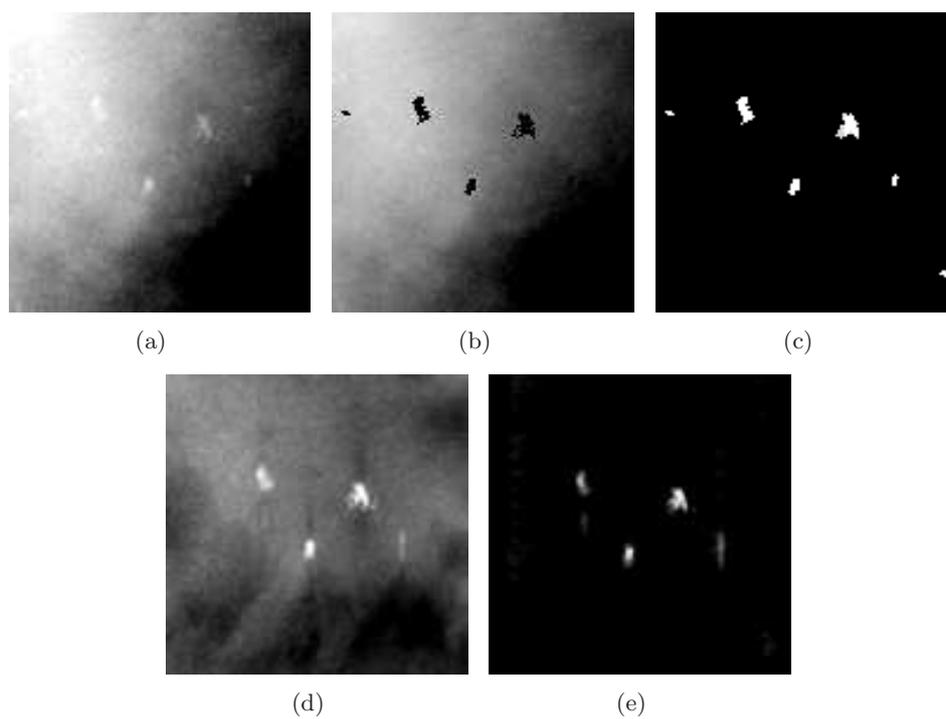


Figure 4.16: (a) selected region of a DBT projected view containing microcalcifications, (b) extracted fuzzy contours, (c) fuzzy particle map, (d) a slice of the reconstructed volume, and (e) the corresponding slice of the fuzzy particle volume.

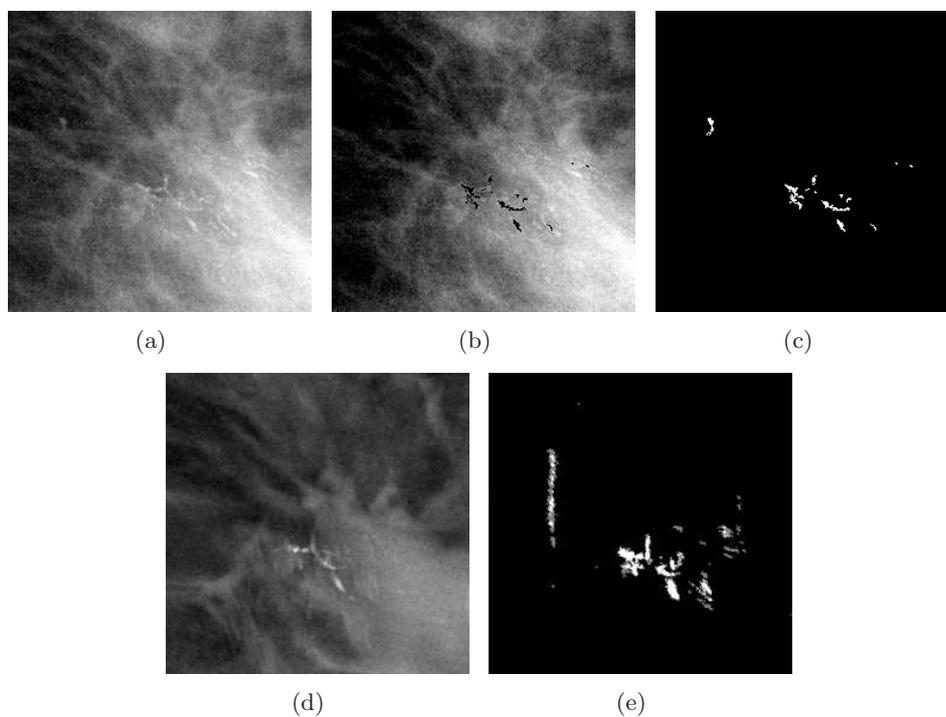


Figure 4.17: (a) selected region of a DBT projected view containing microcalcifications, (b) extracted fuzzy contours, (c) fuzzy particle map, (d) a slice of the reconstructed volume, (e) and the corresponding slice of the fuzzy particle volume.

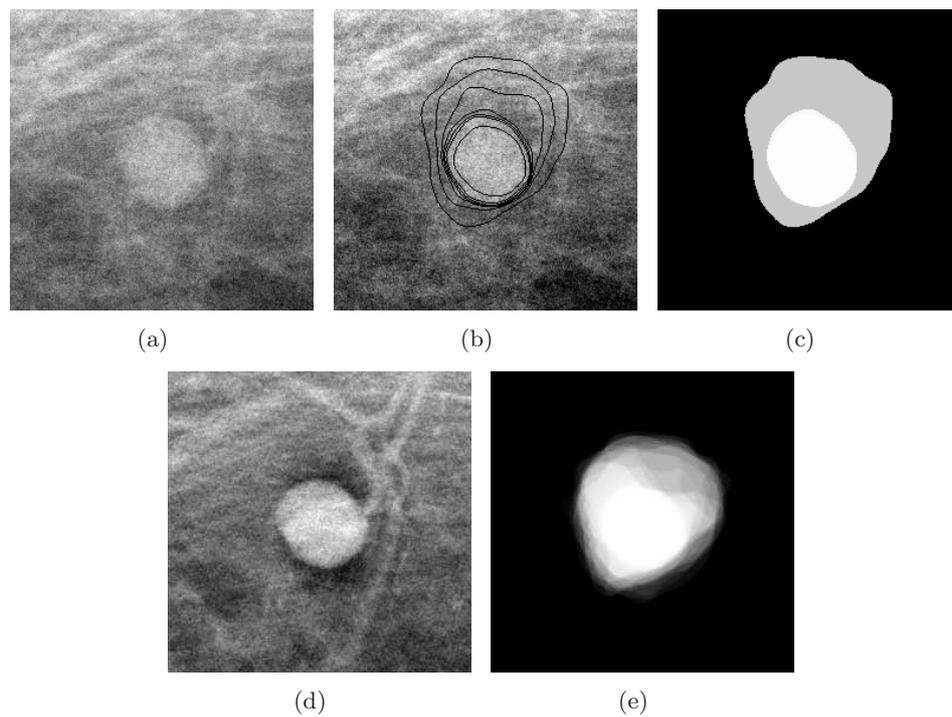


Figure 4.18: (a) selected region of a DBT projected view containing a circumscribed breast mass, (b) extracted fuzzy contours, (c) fuzzy particle map, (d) a slice of the reconstructed volume, (e) and the corresponding slice of the fuzzy particle volume.

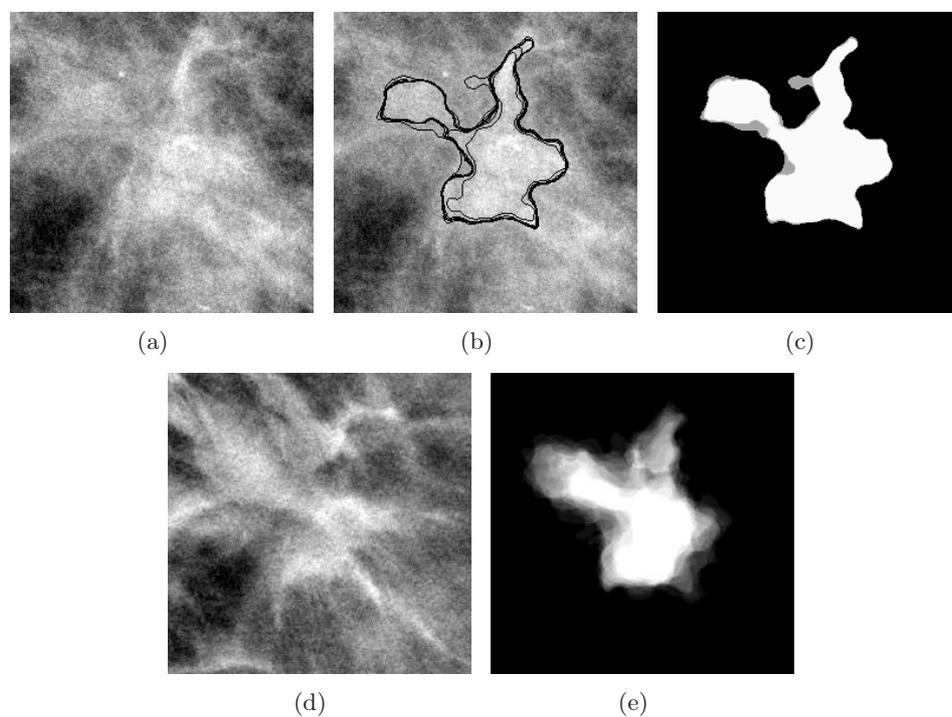


Figure 4.19: (a) selected region of a DBT projected view containing a spiculated breast mass, (b) extracted fuzzy contours, (c) fuzzy particle map, (d) a slice of the reconstructed volume, and (e) the corresponding slice of the fuzzy particle volume.

and the max. As we can see, the choice of aggregation operator has a significant impact on the resulting 3D fuzzy particle map. The max operator presents the most cautious alternative, where no information is neglected. The min represents the opposite end of the spectrum. Using the min operator for combining the fuzzy particle maps, missing information from any one projected view regarding a given object leads to its elimination from the corresponding voxels of the fuzzy particle volume. As we have mentioned above, the arithmetic mean constitutes a robust compromise between these behaviors.

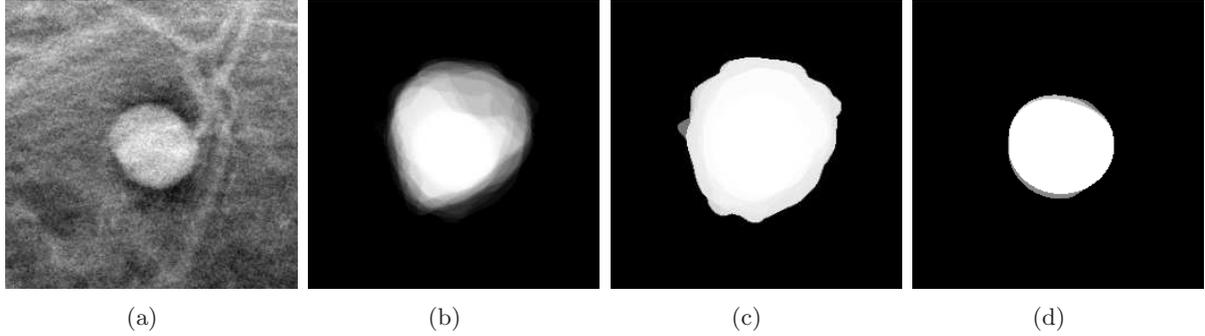


Figure 4.20: Comparison of different aggregation operators: (a) ROI of a slice parallel to the detector plane reconstructed with SART and (b) resulting ROI for a slice of the fuzzy particle volume using the arithmetic mean, (c) the max and (d) the min. A high grey level corresponds to a high membership value. It can be seen that the arithmetic mean provides a good compromise. The central parts of the particle show the strongest response, while the intensity decreases in the surrounding region. The max provides less specificity in terms of object boundaries. However, this operator exhibits properties (very cautious behavior) that make it an interesting alternative for linking particles (see Section 4.3.1). In this example, the min operator provides a good result, showing only the strongest response for the mass. In general this operator is however poorly adapted for our approach. Firstly, the gradual membership is almost completely lost. Secondly, if a particle is not detected in all of the projected views, it will be eliminated from the fuzzy particle volume.

4.3 Aggregation of Fuzzy Data - A New Particle-Based Approach

In this section we present an approach for fusion of fuzzy particles that does not incorporate a pixel-based representation of the particles. In fact, our goal is to obtain a combination of the information extracted for a given particle over the entire set of projection images. To this end, as for the pixel-based approach (see Section 4.2), we use a priori knowledge about the acquisition geometry to find in each projection the two-dimensional particle that corresponds to a given three-dimensional particle (Figure 4.21). The central difference is that the pixel-based approach is only used to establish the set of 2D fuzzy particles in the projected views that correspond to a given 3D particle. The computed fuzzy particle volume is not passed on for further processing. Only the set of 2D fuzzy particles and the associated feature vectors will be preserved.

4.3.1 Linking 2D Fuzzy Particles

We are looking at establishing a set of 2D particles for every 3D particle (even though this 3D particle is never actually computed or visualized). This is equivalent to finding for each 2D particle of a given projected view the corresponding particles in the remaining projected views.

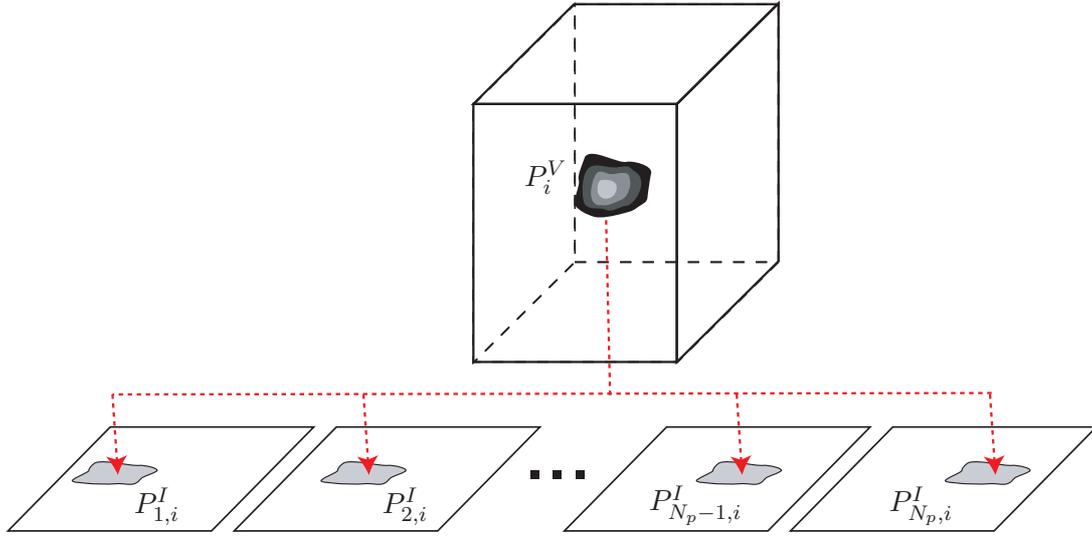


Figure 4.21: Strategy for aggregation of fuzzy data: for a given three-dimensional particle P_i^V the information from all corresponding two-dimensional particles $P_{k,i}^I$ is aggregated. The position of a particle $P_{k,i}^I$ corresponding to the projection of a given particle P_i^V in the k th fuzzy particle map is computed using *a priori* knowledge about the acquisition geometry.

We can then pass all information extracted for a given particle as a single data-set to the next processing step. Consequently, the final decision can be more easily based on the entire set of available information.

Let P_i^V be the i th object of interest in the 3D volume and $P_{k,i}^I$ be the 2D fuzzy particle in the k th of N_p fuzzy particle maps corresponding to P_i^V . We are then looking for a transformation Λ such as:

$$P_i^V = \Lambda_k(P_{k,i}^I), \quad k = \{1, \dots, N_p\} \quad (4.14)$$

The general geometric framework derived from the acquisition geometry has been described in Equation 4.11. However, unlike in Section 4.2 we are not looking to establish a link between pixel positions in the projected views. In fact, we want to establish a link between different candidates $P_{k,i}^I$ detected in the projected views that correspond to the same three-dimensional particle P_i^V .

Considering an implementation on clinical patient data, the solution to this problem is not straightforward. Due to noise and overlying tissue, detection responses for a given particle may be of varying amplitude in different projection images. In some projection images the particle may not have been detected at all while in others the position of the detected particle deviates from the projection of the 3D particle.

To cope with these ambiguities in the sets of detected 2D particles we take advantage of the partial defuzzification technique and the corresponding aggregation strategy described in Section 4.2.1. Once we have computed the fuzzy particle volume, we project the 3D particles (one at a time) onto the 2D fuzzy particle maps. We may then be confronted with a number of *special* cases that need to be addressed (Figure 4.22). A straightforward (crisp) approach would be to simply perform a binary check of the overlap between the 2D particle and the re-projected 3D particle. If the projection of the three-dimensional object overlaps with a single 2D fuzzy particle (a) or with no particle at all (b) the result is obvious. In the case of multiple overlaps (case (c) and (d) in Figure 4.22) additional tests need to be performed in order to decide which particle $P_{k,i}^I$ to associate with the three-dimensional object P_i^V .

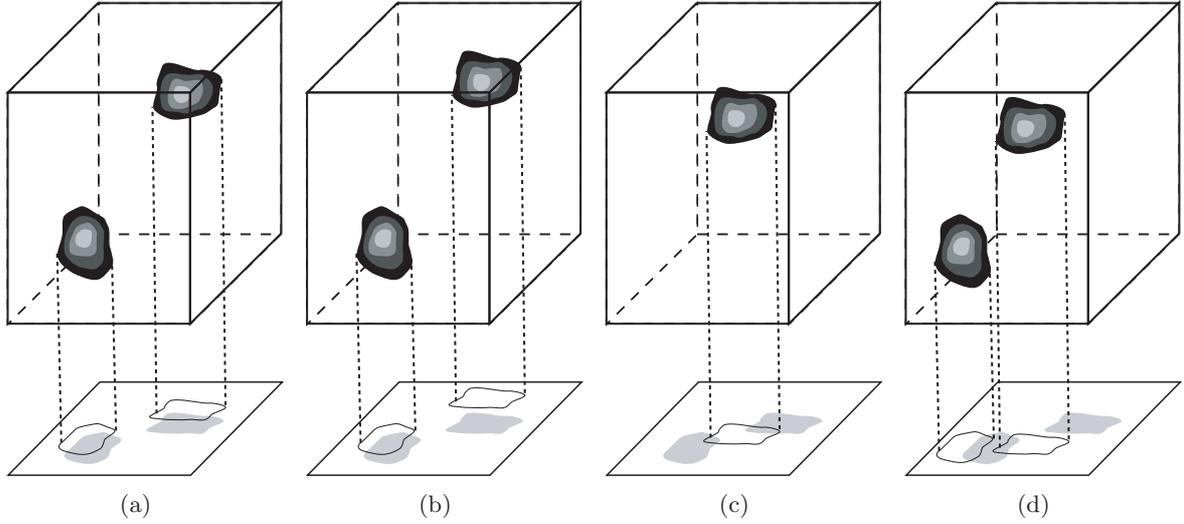


Figure 4.22: Different overlap scenarios for finding correspondence between a three-dimensional particle in the volume and its two-dimensional counterparts in the projection images: (a) overlap with a single two-dimensional particle, (b) no overlap with a two-dimensional particle, (c) a single 3D particle overlapping with several 2D particles, and (d) several 3D particles overlapping with the same 2D particle.

However, performing a simple test of the overlap does not take the fuzzy nature of the data into account. Since we base the link between the different particles on the fuzzy contours (and the associated fuzzy particle maps), we can profit from this additional information to establish the correspondence between the different particles.

Comparing the re-projection of a 3D particle P_i^V with a 2D particle $P_{k,i}^I$ is an operation using similarity measures between two fuzzy subsets. Let $P_{k,i}^R$ be the re-projection of the particle P_i^V relative to the k th source position. The more $P_{k,i}^I$ is similar to $P_{k,i}^R$, the more $P_{k,i}^R$ will also be similar to $P_{k,i}^I$. We are hence looking for a symmetric operator. A metric satisfying these conditions is the resemblance of fuzzy objects (Rifqi 96). The measure of resemblance applied in our case has been described in (Dubois 80). The resemblance $S(A, B)$ is then defined as:

$$S(A, B) = \frac{M(A \cap B)}{M(A \cup B)} \quad (4.15)$$

where M is a measure of information, and A and B are fuzzy subsets. In our case A and B correspond to the 2D fuzzy particle $P_{k,i}^I$ and the re-projected 3D fuzzy particle $P_{k,i}^R$. An example for the computation of the resemblance on clinical data is given in Figure 4.23.

A 2D particle $P_{k,i}^I$ is associated to a 3D particle P_i^V if it satisfies the condition:

$$S(P_{k,i}^I, P_{k,i}^R) \geq t_S, \quad (4.16)$$

where t_S is a threshold value that has been determined empirically. We have now obtained for each 3D particle P_i^V the set of corresponding 2D fuzzy particles \mathcal{P}_i^I defined as:

$$\mathcal{P}_i^I = \{P_{n,i}^I\}_{N_D} \quad \text{with} \quad N_D \leq N_p, \quad (4.17)$$

where N_D is the number of projected views where the particle was detected.

The final step of the aggregation process consists in aggregating the fuzzy attributes of each particle set. The complete algorithm scheme for aggregation on particle level is depicted in Figure 4.24.

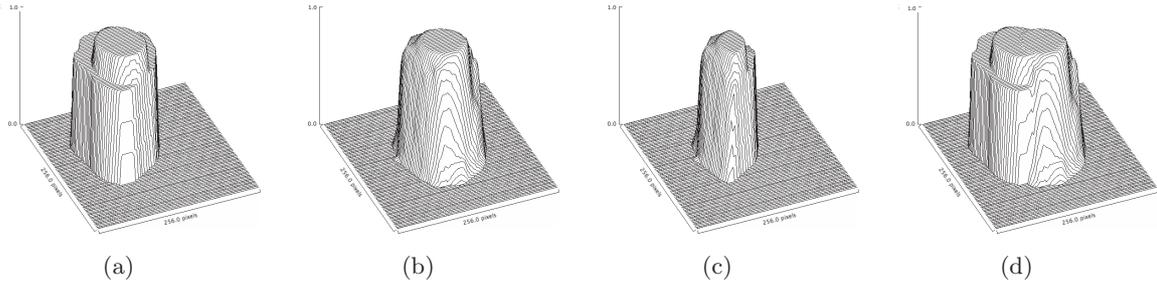


Figure 4.23: Resemblance between two fuzzy subsets. Example for the case of Figures 4.16(c) and 4.16(e): (a) two-dimensional fuzzy particle and (b) re-projection of the three-dimensional fuzzy particle (the membership values are on the third axis). The resemblance is impacted by differences in shape as well as the respective positions of the particles. The intersection and the union of (a) and (b) are depicted in (c) and (d) respectively.

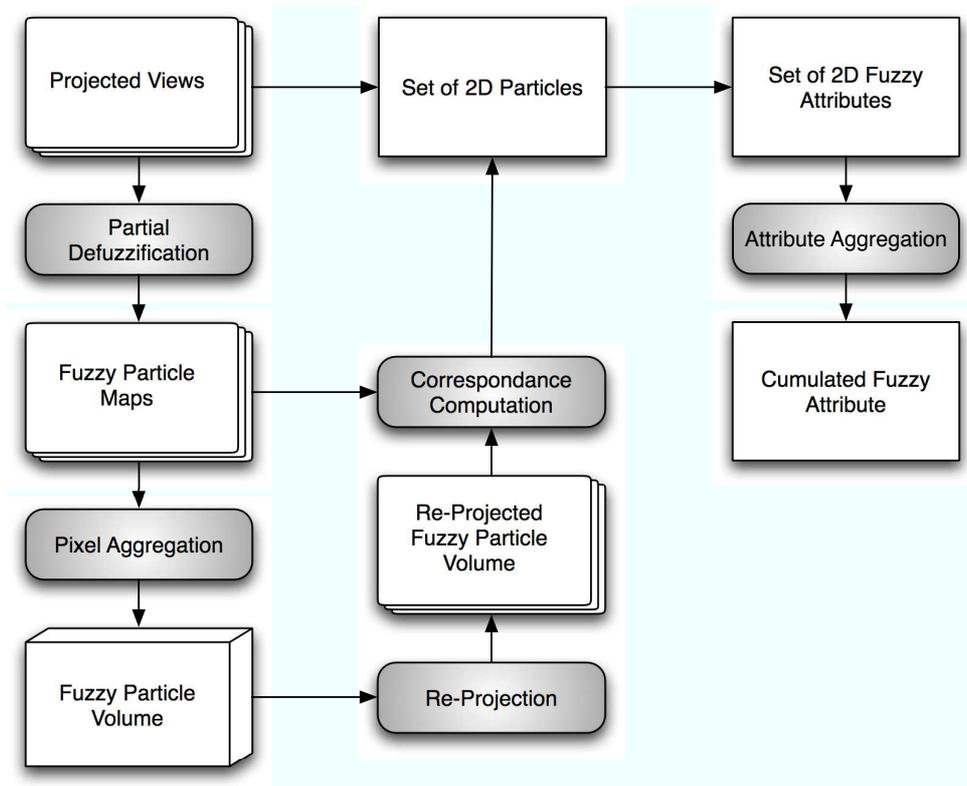


Figure 4.24: Algorithm scheme for aggregation on particle level.

4.3.2 A Novel Approach for Aggregating Fuzzy Attributes

The aggregation step is complete when a comprehensive description of each 3D object has been established that takes into account information extracted over the entire set of projected views. In the preceding section, we have obtained a set \mathcal{P}_i^I of 2D particles that correspond to a given 3D particle P_i^V . Now, our aim is to combine the features of the set of 2D particles in order to obtain a set of features that characterizes the 3D particle based on the entire set of projected views. In the following paragraphs we will discuss the aggregation of fuzzy attributes over the set of projected views for a given particle. For reasons of simplicity we will omit the index i and consider that we are dealing with a single 3D particle and the corresponding 2D particles in the projected views.

Cumulated Fuzzy Attributes A 2D fuzzy particle P_k^I is characterized by a set of fuzzy attributes $\{\mathcal{A}_{j,\mathcal{C}}\}_{j=1,\dots,N_a}$ extracted from the fuzzy contour of the particle (see Section 3.4.1), N_a being the number of fuzzy attributes measured from the fuzzy contour \mathcal{C} . We are looking for an operator that combines the membership functions of a given fuzzy attribute $\mathcal{A}_{j,\mathcal{C}}$ over the entire set of projections. The resulting membership function needs to represent the *cumulated fuzzy attribute* $\mathcal{A}_{j,\mathcal{C}}^c$ for the 3D particle P^V .

In order to be able to choose an operator suited for this task, we need to define the kind of data that we want to combine. According to Section 3.4.1 the membership function of a fuzzy attribute indicates to which degree a given attribute value is characteristic for the particle. The attribute value of a contour that exhibits a high membership degree to the class *contour* will therefore be considered highly characteristic for the considered particle.

How can we then interpret a case, where the membership functions of a fuzzy attribute in different projection images differ significantly? There are many possible reasons for this kind of phenomenon (noise, overlaying tissue, properties of the three-dimensional object, ...). In any case, it translates to the fact that the object of interest - seen under different view angles - is characterized by different attribute values.

One approach to deal with this kind of situation would be to attribute more importance to values that have been observed multiple times. Using e.g. a mean operator, values represented under multiple view angles would be preferred over those that have only been observed few times. This is an occurrence-based approach that seems to correspond well to an intuitive solution.

However, another approach should be considered as well. Let us consider a situation where we have acquired a number of projected views for a given patient. It may then occur that due to noise and overlying parenchyma tissue, a breast pathology is only visible on a single projected view. Then, we only have a single projection where the borders of the tumor can be clearly distinguished. A single *outlier* may sometimes represent the critical information needed to obtain the correct clinical interpretation. Since we are developing an application for breast cancer detection, where missed cancers represent a considerable risk to patient's health, we prefer to proceed in a cautious manner. We have therefore chosen a possibilistic approach, where any information is considered as a possibly correct observation. This leads us to a combination of the different fuzzy attributes via a t-conorm. We have chosen the *max* operator (smallest t-conorm, fuzzy union) for combining the membership functions of a given attribute over the entire set of projection images.

The *cumulated fuzzy attribute* of a 3D object P_i^V is then defined by:

$$\forall a_j \in \mathcal{A}_j, \quad \mu_{\mathcal{A}_{j,\mathcal{C}}^c}(a_j) = \frac{1}{k} \mu_{\mathcal{A}_{j,k,\mathcal{C}}}(a_j), \quad k = \{1, \dots, N_p\} \quad (4.18)$$

Examples for the described method applied on a clinical ROI containing a circumscribed breast mass and for a clinical ROI containing a spiculated breast mass are illustrated in Figures 4.25 and 4.26 respectively. For the circumscribed mass a high compactness (around 0.8) has been measured for most candidate contours in the projected views. This information is translated to the cumulated fuzzy attribute where values around 0.8 for the compactness of the particle are indicated to be most representative for the particle. For the spiculated mass in Figure 4.25 lower compactness values have been computed. Furthermore, the variance of the different projected views is higher than in Figure 4.26. This leads to a higher ambiguity in the cumulated fuzzy attribute function. Values between 0.2 and 0.45 seem to be equally representative of the particle. The obtained cumulated fuzzy attributes express that a high compactness is characteristic for the circumscribed breast mass in Figure 4.25 while a low compactness is characteristic for the spiculated breast mass in Figure 4.26. This corresponds well to the intuitive interpretation of the respective examples.

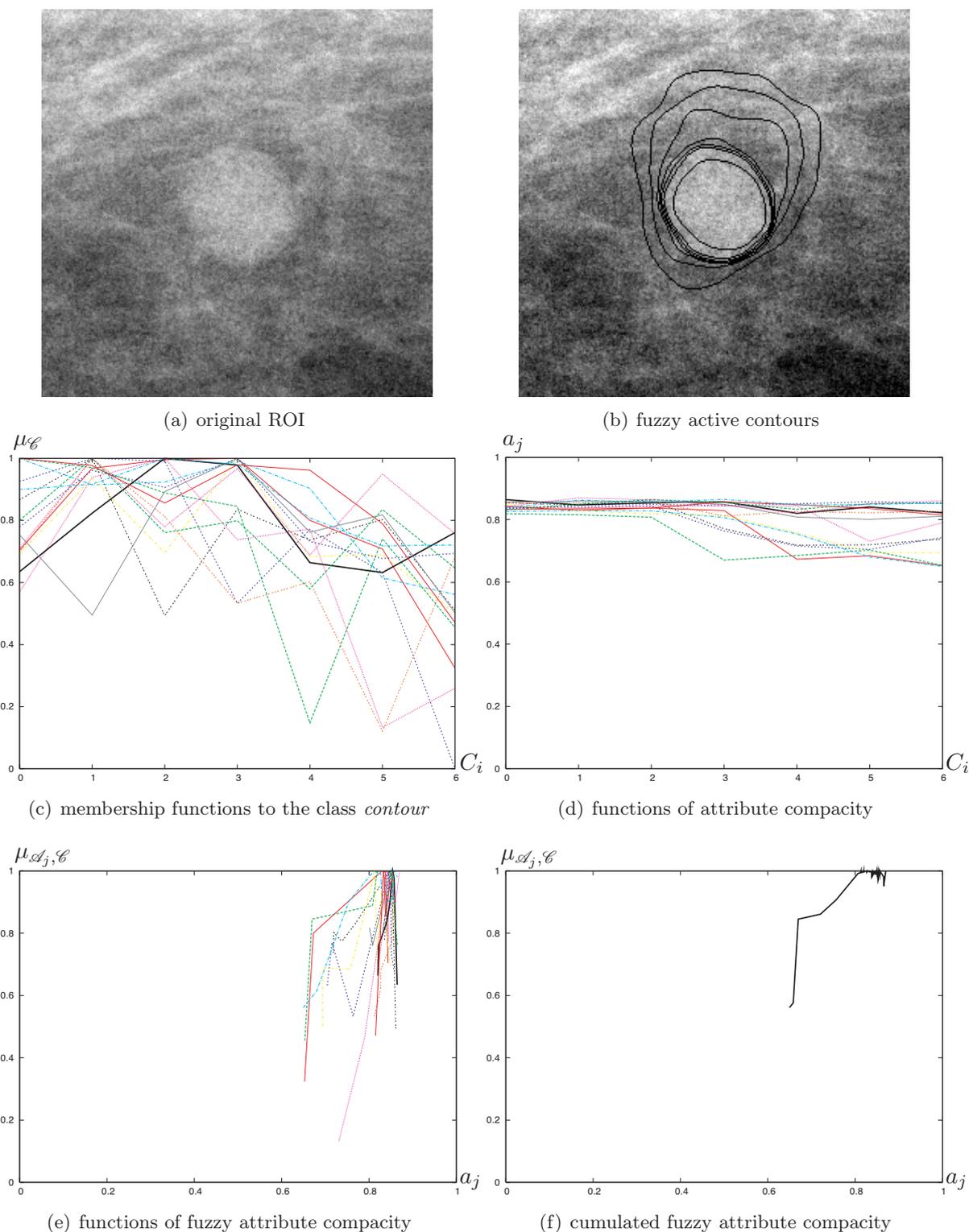


Figure 4.25: Aggregation of fuzzy attributes: (a) ROI of a DBT projected view containing a circumscribed mass lesion, (b) candidate contours extracted using a fuzzy active contour model, (c) corresponding membership functions to the class *contour*, (d) attribute values for the set of candidate contours for the attribute *area*, (e) superposition of the N_p membership functions of the fuzzy attribute and (f) the resulting membership function for the cumulated fuzzy attribute *area*. In (c-e) the different curves represent values for different projected views. The solid black line corresponds to the 0-degree projection depicted in (a) and (b).

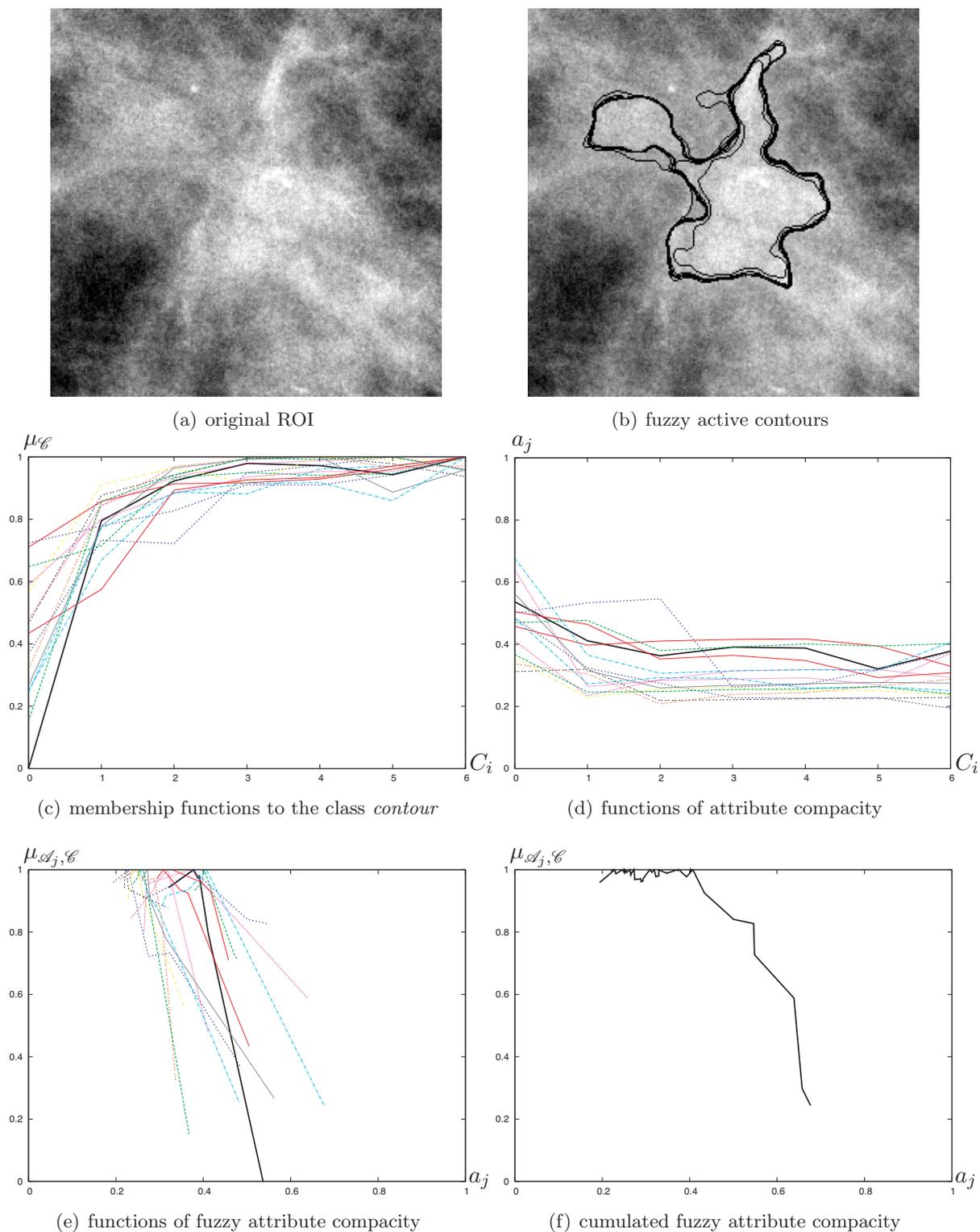


Figure 4.26: Aggregation of fuzzy attributes: (a) ROI of a DBT projected view containing a spiculated mass lesion, (b) candidate contours extracted using a fuzzy active contour model, (c) corresponding membership functions to the class *contour*, (d) attribute values for the set of candidate contours for the attribute *area*, (e) superposition of the N_p membership functions of the fuzzy attribute and (f) the resulting membership function for the cumulated fuzzy attribute *area*. In (c-e) the different curves represent values for different projected views. The solid black line corresponds to the 0-degree projection depicted in (a) and (b).

4.4 Conclusion

In this chapter we have introduced two novel methods for aggregation of fuzzy information from a set of tomographic projection images. First, a pixel-based approach has been developed. We have shown how *partial defuzzification* can be used to obtain a pixel-based fuzzy representation of the objects from the fuzzy contours. A dedicated aggregation operator is applied to obtain a *fuzzy particle volume* from the set of *fuzzy particle maps*. Today, the values of the fuzzy particle maps are averaged to obtain a pixel in the fuzzy particle volume. Other aggregation operators have been briefly investigated. However, a more in depth investigation of alternative aggregation operators appears an interesting field for further research.

A new particle-based approach has been developed as well. Here, the pixel-based approach is applied only to establish the link between a given three-dimensional particle and the corresponding two-dimensional particle in each projected view. This information is the used to compute *cumulated fuzzy attributes* for each three-dimensional particle. The cumulated attributes serve as input to a classification step. An improvement to the attribute aggregation strategy may be provided by adapting the aggregation operator to the properties of each individual attribute. Attributes that are expected to have the same values in all projected views would be aggregated differently than those whose values are expected to vary in between different projections.

Now that we have all this useful information, it would be nice to do something with it. (Actually, it can be emotionally fulfilling just to get the information. This is usually only true, however, if you have the social life of a kumquat.)

Unix Programmer's Manual

5

Decision Making

In this chapter we aim to make a decision from the information extracted in the preceding processing steps. In our application, this decision is used to help radiologists in identifying cancers in mammographic images. The decision thus refers to associating each candidate particle or group of candidate particles with a predefined *class* of findings. We want to obtain an indicator as to the type of breast pathology present, and the level of malignancy associated with it. This knowledge can then be used to attract the radiologist's attention to those objects, that represent a potential danger to the patient's health.

We start out with an overview of popular classification techniques and their application in CAD. We then discuss the ID3 method for classification using crisp binary decision trees. Tree construction as well as prediction of unknown samples will be illustrated. The main part of this chapter is dedicated to fuzzy decision tree classifiers. A fuzzy decision tree may be used to process crisp or fuzzy input data. We have applied a fuzzy decision tree to make a decision from the cumulated fuzzy attributes computed in Section 4.3. The presented approach was developed in (Bothorel 96). In Section 5.4 we introduce several original contributions for improving the fuzzy decision tree working on fuzzy input data.

5.1 Introduction

We generally distinguish between two types of classifiers: supervised and unsupervised. With supervised learning, the class L_c of each sample of the training data set is known along with the feature vector x . We thus know the restriction of the decision function onto a part of the feature space. The learning process strives to find an extension of this function to \mathbb{R}^d . Examples of classifiers based on supervised learning are K-Nearest Neighbor (KNN) (Cover 67), Artificial Neural Networks (ANN) (Rumelhart 86; Pearl 88; Lau 92), Support Vector Machines (SVM) (Vapnik 00), and Decision Trees (Quinlan 86; Safavian 91). On the other hand, with unsupervised learning, the class labels associated with the training data set are unknown. Hence, classes have to be determined automatically, after a clustering of neighboring points has been performed according to given mathematical criteria. K-Means, Dynamic Clouds (Diday 76) and Self-Organizing Maps (SOM) (Kohonen 82) are examples of unsupervised machine learning algorithms.

Numerous classifiers have been used for CAD on mamographic images such as ANN (Chitre 93;

Bankman 94; Floyd Jr 94; Cheng 94), BBN (Zheng 99), KNN (Arbach 03), decision trees (Kegelmeyer Jr 94; Iyer 00; Zheng 01) and SVM (El-Naqa 02; Sun 04). Among them, decision trees exhibit some properties, that are especially well suited for our problem.

The basic ID3 methodology has been extended to C4.5 in (Quinlan 93). When constructing a decision tree, one can now deal with training sets that have unknown attribute values. Another extension consists of the capability to classify samples from a test database that have unknown attribute values. This is achieved by estimating the probability of the various possible results. A comprehensive review of decision tree technology can be found in (Safavian 91; Zighed 00).

Decision trees have been widely used for classification of mammograms. A binary decision tree was used to classify texture features in (Kegelmeyer Jr 94). These features were computed from a map of local edge orientation for the detection of stellate lesions in mammograms. In (Iyer 00) the authors applied a two-step classification for the feature-based interpretation of mammograms. The image is decomposed using the discrete wavelet transform, and a clustering algorithm is used to initiate the segmentation at the subband of the decomposition. The classifier is then applied to decide whether a given region is suspicious for cancer or not. A binary decision tree was used to define a certain separability at a coarse level. The authors pointed out that a crisp classifier cannot classify perfectly because of the non separability of some features. And since the (theoretically existent) perfect set of features is usually extremely difficult to obtain, fuzzy techniques may provide a solution to this seemingly *hard-to-solve* problem. Another approach made use of a binary decision tree for classification of extracted regions based on some feature set in a tumor detection scheme in screening mammography (Zheng 01). The classification by use of a decision tree was chosen due to its simplicity and low computational overhead. A fuzzy binary decision tree classifier is applied for tumor detection in (Li 95). The fuzzy decision tree was chosen over ANN and kNN since better performances could be achieved using different subsets of features at various decision levels instead of a single *best* set of features in a one-step decision. The classification performance was assessed using a free response receiver operating characteristic (FROC), yielding a sensitivity of 100% at the expense of about six false positive (FP) detections per image. A combination of an ID3 decision tree classifier and a kNN classifier is realized in (Oliver 05). Each classifier is used to compute a fuzzy membership. The average combination of these membership values is used to produce the final result. It is pointed out that the ID3 decision tree classifier has a higher efficiency compared to the kNN classifier. The underlying cause has been identified as being that the decision tree classifier contains a feature selection discrimination process.

The advantages of decision tree classifiers thus include:

- Flexibility regarding size of training set: they still produce consistent results even if the number of examples in the training database is very small. In our case where few clinical cases are available today, use of a decision tree classifier allows us to develop a first prototype from the limited available data sets.
- Easy to include expert rules: the translation of rules given by an expert (e.g. in our case a radiologist) in natural language (e.g. *irregular contour*) is possible without major difficulty. There are some rules for differentiation between benign and malignant masses (Bassett 87; Tabár 01). We can choose to base the node tests directly on these rules.
- Processing of symbolic and numerical attributes is possible (see Figure 5.3). If a calcification has been labeled as *round* by an expert we do not need to translate this into an attribute value (e.g. compacity).
- No a priori knowledge about attributes is needed (e.g. the objects of interest are *round*) to perform the construction of the decision tree. We will present an approach where the

node tests of the decision tree are derived from histograms of attribute values computed over the samples of the training data sets.

- Results from a classification by a decision tree are easily translated into natural language. A combination of different leaves and their respective paths through the tree leads directly to a comprehensive description of the obtained results. In our case this is especially helpful, since the set of rules obtained from training the tree can be compared to and validated with expert knowledge. Furthermore we may derive new comprehensive rules from training on a large database.

5.2 Decision Trees

In this section we describe the basic principles and terminology of decision trees. We will demonstrate how a decision tree can be constructed, and how such a tree is used for classifying samples of unknown class membership.

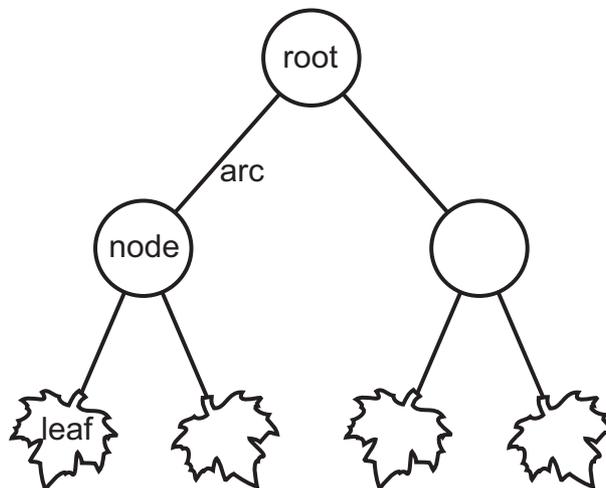


Figure 5.1: Basic Terminology of Decision Trees.

5.2.1 Introduction

A schematic of a decision tree is given in Figure 5.1. A tree is made up of a *root* or *root node* that connects via *arcs* to its siblings. These may be internal *nodes* or terminal *nodes*, so-called *leaves*. Each node is associated with an attribute and each arc originating from that node is associated with one of the categories or values that this attribute can take (Figure 5.2). A *node test* is applied to each sample that passes a given internal node. The result of this test determines which branch the sample will pass into next. Decision trees can process numerical or symbolic data (see Figure 5.3).

5.2.2 Tree Construction

There are numerous approaches for constructing decision trees from training data. A well known and widely used technique is the Iterative Dichotomiser 3 (ID3) first introduced in (Quinlan 86). The algorithm is based on Occam's razor (Blumer 87): it prefers smaller decision trees (simpler theories) over larger ones.

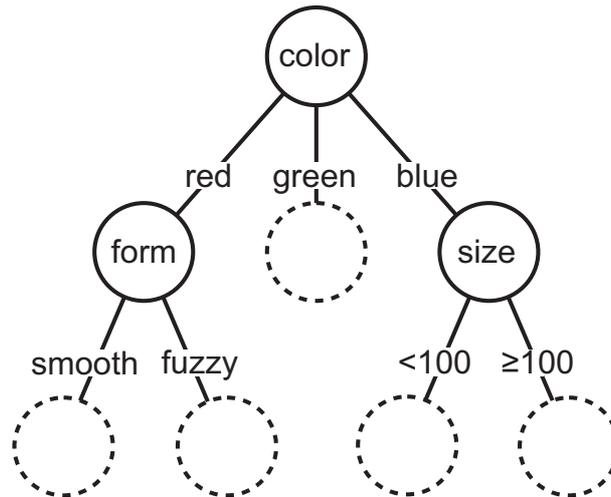


Figure 5.2: Example of attributes and their corresponding values in a decision tree. Nodes are associated with attributes while the arcs originating from such a node are associated with the values taken by the corresponding attribute.

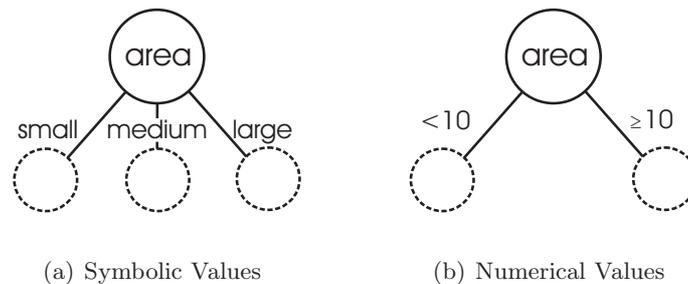


Figure 5.3: Two examples of node tests for the attribute *area*. On the left we have three classes of symbolic values. On the right a threshold is used to divide the range of numerical values into two distinct sets.

The ID3 method realizes the construction of the decision tree in a single pass (as opposed to methods that apply an iterative approach (Quinlan 86; Utgoff 89)). The basic steps of tree construction by ID3 are:

- Select the attribute that best classifies the samples.
- Divide the data into sets of fixed attribute value (one set per attribute value).
- Build a tree with the number of branches equal to the number of sets.
- For each subtree: repeat this process.
- If a given stopping criterion is met, then stop.

When constructing a decision tree we thus start out with a root node. In order to find the node test for the root node, we have to identify the attribute providing the *best* classification of the training data set. This ability is the so-called *discriminating power* of an attribute which is computed by means of the measure of discrimination H_d (Marsala 06). To this end we compute a number of threshold values for each attribute, resulting in the desired number of classes. The threshold value may be derived from a priori knowledge about the attributes (e.g. expert rules).

An alternative, automatic method includes computation of histograms for each class of results (Figure 5.4(a)). We then need to determine a threshold such that the error of classification is minimal for the given attribute (Figure 5.4(b)), using for example a Bayesian decision process (Hartmann 82). This leads to the construction of membership functions for each attribute.

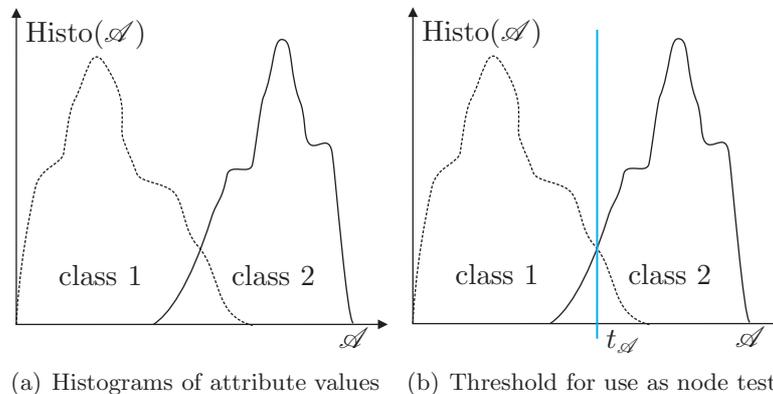


Figure 5.4: Automatic node test computation

Once these membership functions have been computed, a metric is needed to compare their capabilities to classify the data correctly. A large array of measures of discrimination have been used for this task (Hartmann 82; Blumer 87; Aczel 06). In the ID3 algorithm, the *Shannon entropy* (Shannon 49) is generally used for this purpose. This entropy is given by:

$$H_S(A) = - \sum_{i=1}^{N_A} P(A, i) \log P(A, i) \quad (5.1)$$

where A stands for a given attribute, N_A is the total number of values that this attribute can take, and $P(A, i)$ denotes the a priori probability of the attribute A having the value i .

After selection of the attribute exhibiting the highest discriminating power, the entire training data set passes the node test based on this attribute. A subtree is created for each class that results from this test. Each sample of the training data passes into the subtree corresponding to the class it was associated with. This process is repeated for each subtree until a given stopping criterion is met. This may include criteria based on the number of samples per leaf, the purity of the leaves, the number of nodes in the tree or the number of attributes used for the classification. A complete decision tree and its associated node tests are illustrated in Figure 5.5. The membership functions to the two classes are complements to each other in the case of crisp decision trees. For reasons of simplicity we will therefore display only one of these functions from here on.

Obtaining a Final Decision. The last step in constructing a decision tree consists in deriving rules for a final decision. Once every sample has passed through the tree and arrived in a leaf, we need to formulate rules on which a decision can be based. For a given leaf this may be realized by a majority vote (the leaf is attributed the class to which belong the majority of the samples in the leaf). Furthermore, leaves belonging to the same class may be regrouped to reduce the complexity of the tree. This is a large field of research by itself, and numerous works have been dedicated to this subject (Safavian 91; Hüllermeier 05).

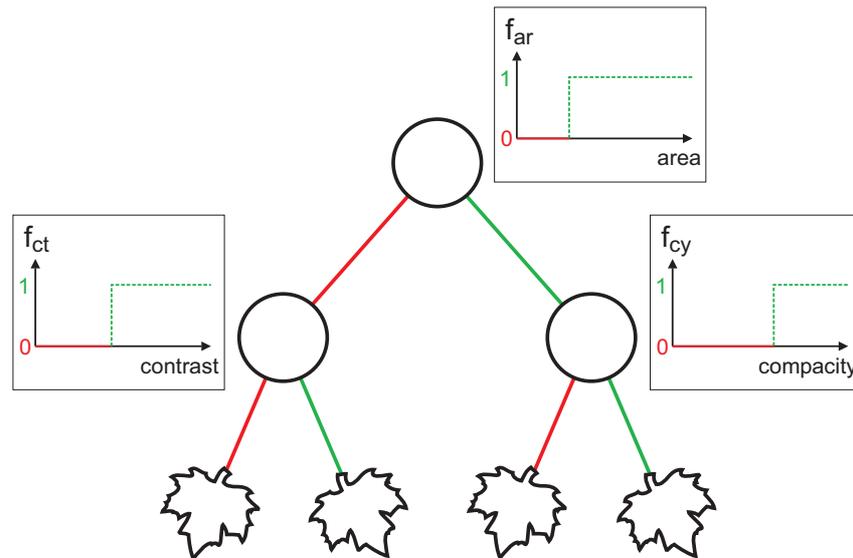


Figure 5.5: A decision tree and corresponding node tests.

5.2.3 Classification of Test Data

After the decision tree has been constructed, it can be used to classify samples from a test database. When a sample passes through the tree, its attribute values are tested at the nodes to decide which path the sample will follow. Once the sample arrives in a leaf, it will be attributed to the class corresponding to this leaf (see Section 5.2.2 and Figure 5.6). When the entire training data set has passed through the tree and all samples have been attributed to a class the classification is complete.

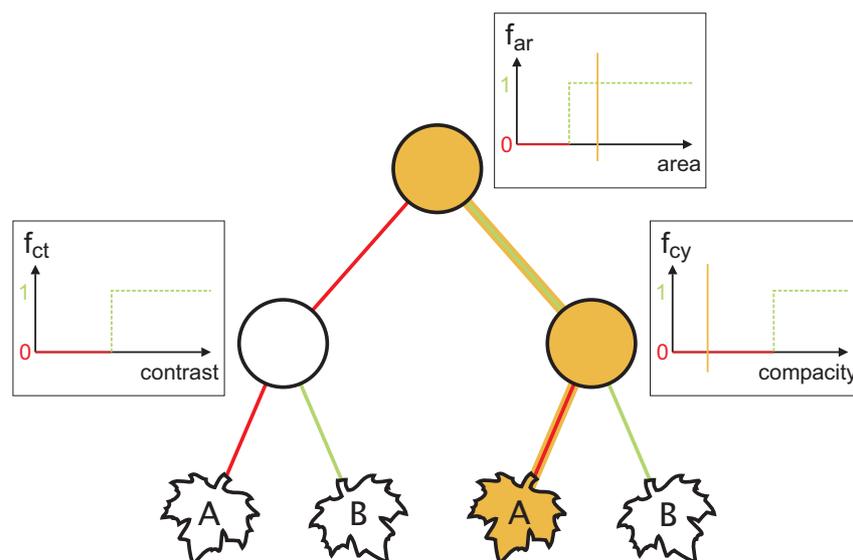


Figure 5.6: Example of a decision tree processing a sample from a test database. The attribute values of the sample and its path through the tree are highlighted. The sample is classified as belonging to class A.

5.3 Fuzzy Decision Trees

Today, many classification techniques are trying to take advantage of the benefits of fuzzy processing.

Incorporation of fuzzy rules and fuzzy processing exists for most popular classifiers such as KNN (Keller 85), ANN (Kosko 92; Lee 94; Chen 00), decision trees (Janikow 98) and Support Vector Machines (Lin 02). A comprehensive review has been given in (Hüllermeier 05). In addition there have been classifiers especially built for taking advantage of fuzzy set theory like the one presented in (Sun 93).

However, there are only very few approaches where classifiers actually work on fuzzy data. Such implementations have notably been published for decision trees (Yuan 95; Bothorel 96; Janikow 98; Marsala 99; Chiang 02) and neural networks (Lee 94; Ishibuchi 00). Based on the properties of decision trees presented in Section 5.1 we choose to perform a classification by use of a decision tree. We then need to adapt this methodology to take into account the fuzzy nature of the data. In the following we describe the basic principles of classification using fuzzy decision trees and we give an overview of existing approaches.

5.3.1 Crisp Input Data

In this section we describe the use of fuzzy set theory to improve the predictive power of crisp decision trees. Processing of fuzzy attribute values will be discussed in Section 5.3.2.

The basic idea of introducing fuzzy set theory into the concept of classical (crisp) decision trees is to compensate for one of the central shortcomings of these trees: the severity with which the tested values are thresholded at each tree node. Examples having attribute values close to each other, but on different sides of the node threshold, will take completely different paths through the classical decision tree. In order to overcome this limitation, we use one of the well known properties of fuzzy set theory: the gradual membership to a given class or set.

The fuzzy decision tree therefore accepts examples that follow to some degrees different paths through the tree (Figure 5.7). In crisp decision trees, each example finds its way to a single leaf of the tree. When working with fuzzy decision trees, this is no longer valid. In fact, each example will follow all possible descending paths through the tree to arrive in all tree leaves. The degree to which a given example arrives in a given leaf reflects its path through the tree, and the node tests passed along the way.

The objective is to replace the node tests of the crisp case with node tests that allow for a gradual membership to one or several of the available classes. The conception of such a structure raises two main questions:

- How to define node tests?
The node tests for crisp decision trees consist of a simple threshold value that separates the classes. In the case of fuzzy node tests, we now need to define separator functions for the different classes.
- How to obtain a decision from results in different leaves?
Deriving a final decision from the results of a crisp decision tree is straightforward. In the fuzzy case however, where each sample arrives in all leaves with a membership degree between 0 and 1 and each leaf belongs to each class to a certain degree, a method for obtaining a final decision from the needs to be investigated.

In order to construct our fuzzy decision tree we apply once again the basic steps of tree construction known from the ID3 algorithm (see Section 5.2.2).

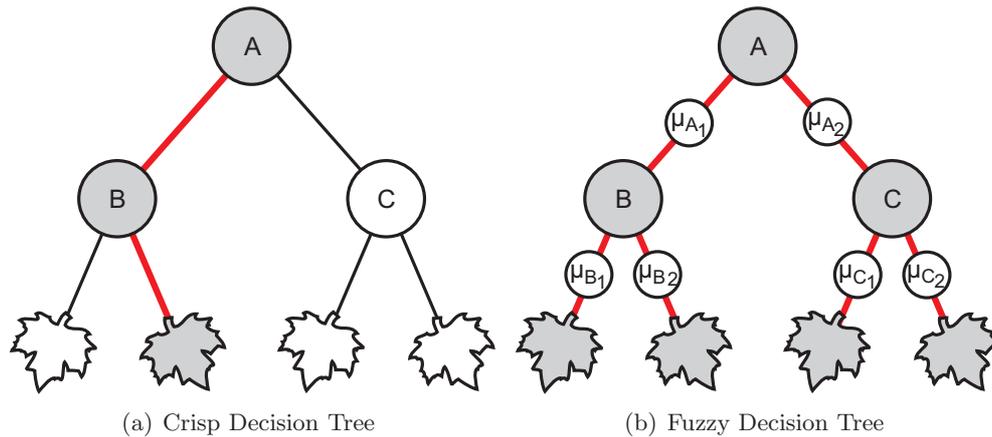


Figure 5.7: Paths taken by a given sample that passes through the nodes of a decision tree. In (a) the case of a crisp decision tree is shown, where the sample passes nodes A and B before it arrives in a single leaf. In the case of a fuzzy decision tree (b), the sample passes all nodes of the tree, each one to some degree, therefore arriving in each leaf.

Fuzzy Node Tests. In a fuzzy decision tree the attribute test $attribute > threshold$ is replaced by the test $fuzzy\ attribute = value$, where value is represented by a membership function.

Example: Instead of testing whether $particle\ size > 50\ pixels$, we will test whether the particle is *small* and whether it is *large*. If $size < 20$ the example will follow the branch *small size*. If $size > 80$ the example follows the branch *large size*. However, if $20 < size < 80$, both branches will be followed, to some degree.

There are two major groups of approaches to define the membership functions replacing the thresholds for the test at the nodes: the one based on a priori knowledge about the attributes and the automatic methods that do not depend on such knowledge. One example of the first group consists in having an expert determining the membership functions based on a priori knowledge about the features. We note that even when all separator functions are determined using a priori knowledge (there is no learning from a test database needed to establish the separator functions), we still need measures of discrimination in order to build an optimally small tree.

The separator functions that constitute the node test for a given attribute are obtained from the histogram of the test samples over the attribute values (see Figure 5.8). A measure of discrimination is needed to choose the *best* among all attributes - the attribute providing the best discrimination of the test samples between the classes.

Then, the separator functions used for the node test are not identical with the attribute histograms used to compute the contrast measure. Using the attribute histograms directly and without modifications as separator functions for the node tests comprises a number of limitations that need to be addressed. Firstly, the method does not allow for computation of membership functions outside of the histograms domain. Secondly, we observe a high sensibility to noise especially for attribute values that are rarely represented in the training data set. A number of processing steps are performed on the fuzzy attribute histograms in order to derive more robust node tests with a higher power of generalization. This is necessary since the size of the database we are working with is quite limited. The preprocessing applied to the fuzzy attribute histograms is illustrated in Figure 5.9. These preprocessing operations are described in more detail in (Bothorel 96).

The different methods for constructing fuzzy decision trees can essentially be distinguished by the measure of discrimination applied to choose the attributes for the node tests. There is a number of measures of discrimination to choose from. The so-called Star-entropy (entropy of

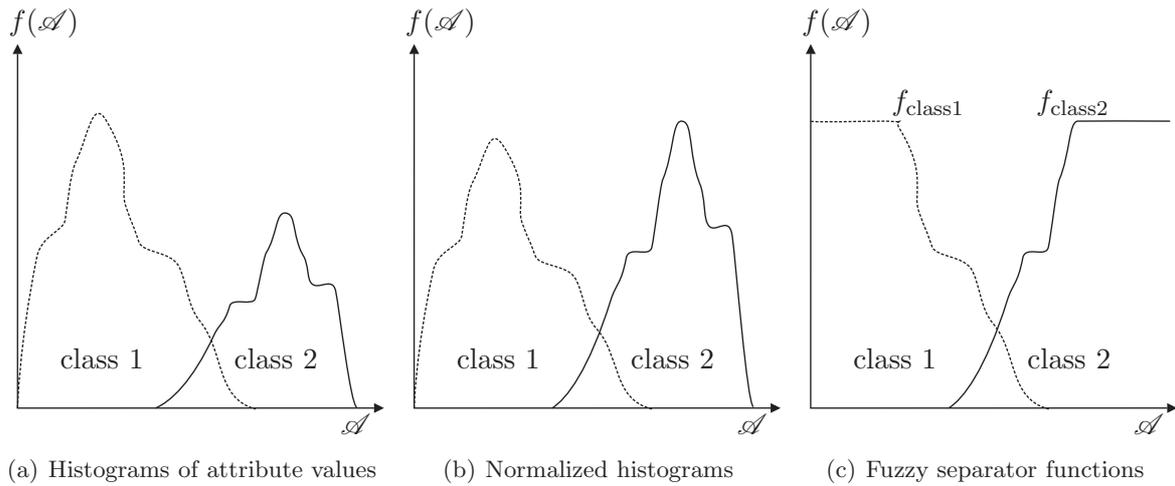


Figure 5.8: (a) For a given attribute we compute histograms for each class. (b) These histograms are then normalized with respect to the area under the curve. (c) Finally, separator functions for use as node tests are computed using marginal probability laws. Several processing steps are applied to generalize the attribute histograms (see Section 5.4.2).

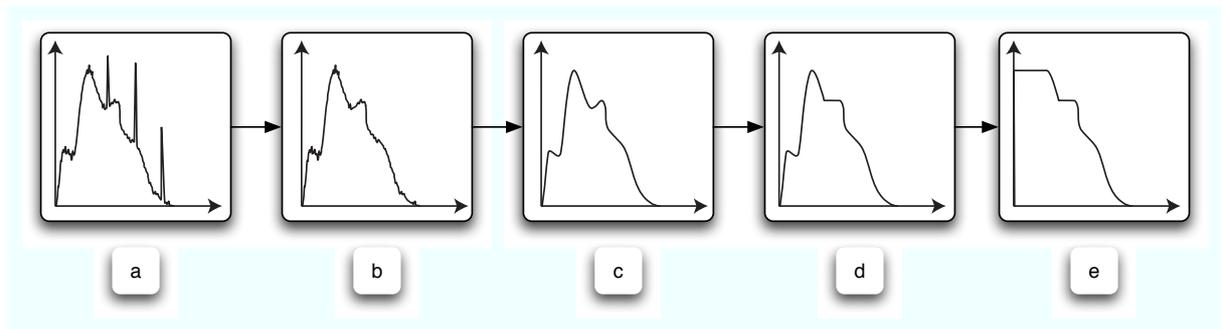


Figure 5.9: Processing applied to the fuzzy attribute histograms: original histogram (a), median filter output (b), mean filter output (c), reconstructed function (d), and final separator function where the maximum has been extended to the limit of the domain (e).

fuzzy events), a measure derived from the Shannon-entropy has been extensively used for this purpose (Umanol 94; Ramdani 94; Ichihashi 96; Janikow 98):

$$H_S^*(C) = - \sum_{i=1}^{N_A} P^*(A, i) \log(P^*(A, i)), \quad (5.2)$$

where P^* is a probability of fuzzy events. The probability of fuzzy events generally used has been introduced in (Zadeh 68).

Furthermore, other types of entropies have been introduced and have been applied for discrimination of attributes in fuzzy decision trees (DeLuca 72; Kosko 86; Fan 99; Fan 02).

A different measure of discrimination has been proposed in (Yuan 95). It is based on each attribute's contribution in reducing the *classification ambiguity*. The definition of classification ambiguity is based on a measure of satisfiability (see also Section 5.3.2).

Another method is based on the contrast between two fuzzy subsets (Bothorel 96):

$$C(\mathcal{A}_1, \mathcal{A}_2) = \frac{\int \min(\mu_{\mathcal{A}_1}, \mu_{\mathcal{A}_2})}{\int \max(\mu_{\mathcal{A}_1}, \mu_{\mathcal{A}_2})} \tag{5.3}$$

where \mathcal{A}_k is the fuzzy subset that corresponds to the class c_k (with the membership function $\mu_{\mathcal{A}_k}$) and $C(\mathcal{A}_1, \mathcal{A}_2)$ is a contrast measure between the two fuzzy subsets \mathcal{A}_1 and \mathcal{A}_2 . The integrals are computed over the definition domains of the fuzzy subsets. The attribute that minimizes $C(\mathcal{A}_1, \mathcal{A}_2)$ is considered the *best* one (see Figure 5.10).

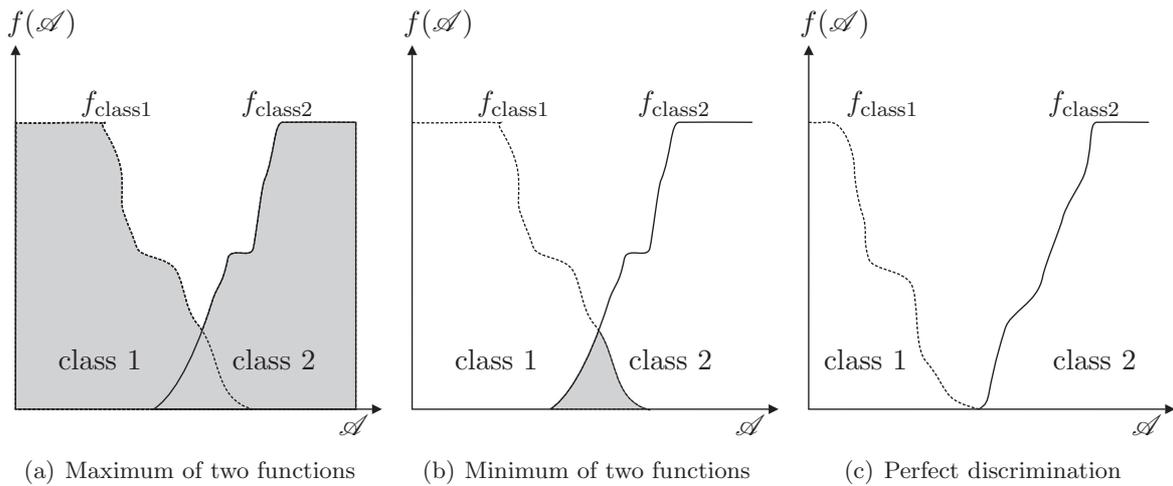


Figure 5.10: Choice of most discriminating attribute by measuring the contrast between the separator functions: (a) The area under the curve for the maximum of the fuzzy subsets f_{A_1} and f_{A_2} . (b) The area under the curve for the minimum of the fuzzy subsets f_{A_1} and f_{A_2} . (c) A set of separator functions that represent a perfect discrimination between two classes.

A means of comparing different measures of discrimination has been described in (Marsala 98) and an overview of different measures of discrimination that allow for the ranking of attributes is given in (Marsala 06).

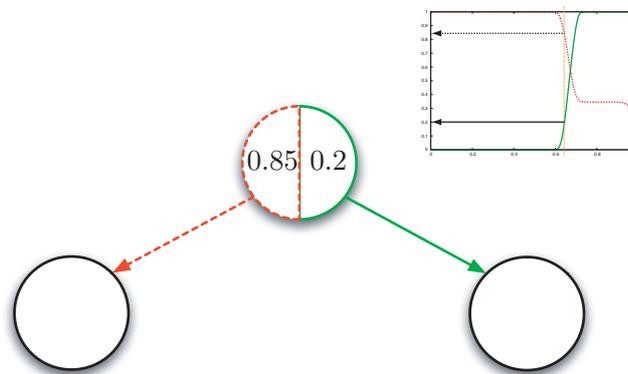


Figure 5.11: Construction of a fuzzy decision tree. A sample arriving at the root node is tested with the corresponding node test. This results in a membership degree to each of the two classes (class A and class B, left half and right half of the root node respectively).

Once we have identified the *best* attribute, the root node having this attribute's separator functions as node test is created. Then, the entire training data set is tested by this node. For

Class Associations of the Leaves During construction of the fuzzy decision tree each training sample will arrive in each leaf with a given cumulated degree of membership. We can then compute the class density for class L_c in the i th leaf by:

$$n_i^{L_c} = \sum_{j=1}^{\aleph_{L_c}} \mu_{i,j}^c, \quad (5.4)$$

where \aleph_{L_c} is the cardinality of the training data set for class L_c and $\mu_{i,j}^c$ is the cumulated degree of membership of sample j at leaf i .

This process is illustrated in Figure 5.13.

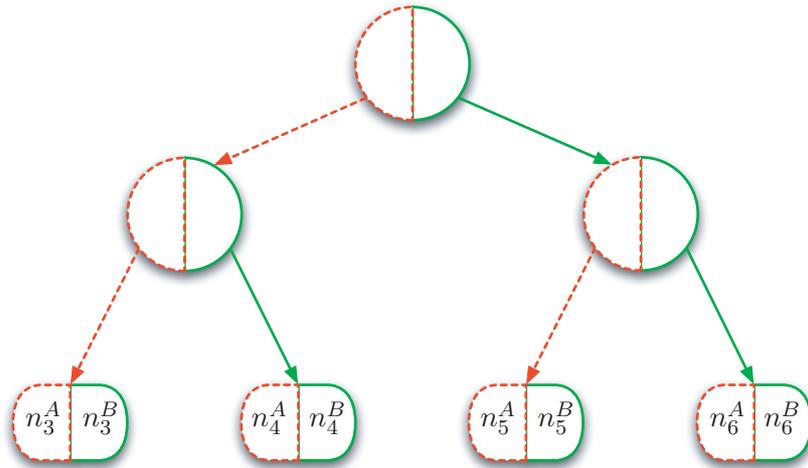


Figure 5.13: Class densities for different leaves. When all samples of the training database have been processed, we need to find the associated classes for each leaf of the tree. For each class we compute the sum of the cumulated degrees in a given leaf of all samples belonging to this class. We then compute a ratio of class densities for each leaf, representing the distribution of samples arriving in the leaf.

Obtaining a Final Decision Until now, the tree has been used to classify a set of data into a number of leaves. Since the number of leaves is generally much larger than the number of desired output classes, we look for a transformation that produces the desired final result while taking into account the classification given by the set of leaves.

In the crisp case, this operation is straightforward, since each sample arrives in a single leaf, and is thus automatically attributed the class associated with this leaf. In the case of a fuzzy decision tree this process needs to be adapted to the fact that each sample is present in each leaf after tree construction. In general, fuzzy decision trees do not exhibit *pure* leaves. During the construction of the fuzzy decision tree each leaf is attributed to each class with a certain degree. After a test sample has been processed by all nodes of the fuzzy decision tree, it has arrived in all of the leaves with a certain cumulated membership degree. We then have to find an operator for combining the set of cumulated membership degrees into a single degree for each class. Each path through the tree can be interpreted as a set of rules describing an object that belongs to one of the classes. One class may then contain very different object definitions. Here, we want to derive the membership degree to the class for a given sample. Since it is sufficient that the sample satisfies one of the object definitions in order to be considered a member of the class, we need to perform a fuzzy union of the different membership degrees to the class. This leads us to applying a t-conorm to each set of cumulated membership degrees. We thus obtain a membership degree to each class for each test sample as illustrated in Figure 5.14.

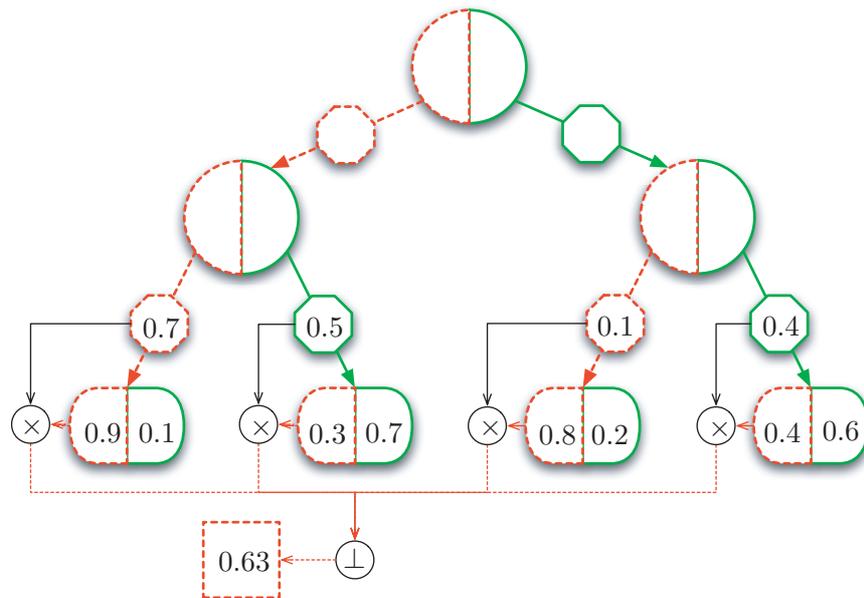


Figure 5.14: Obtaining a final decision. A full fuzzy decision tree is shown processing a test sample. The associated membership degrees at each node (inside the nodes) as well as the associated cumulated degrees (octagons) are given. The values inside the leaves correspond to the distribution of samples of the two classes (class densities). The total membership degree to the class A is computed by weighting the cumulated degree in each leaf with the corresponding class distribution for class A. The resulting values are then combined over the full set of leaves using a t-conorm. In our example, the max has been used for this purpose.

This does not necessarily lead to a crisp decision. The final result could be represented by the obtained membership degrees to different classes for each test sample. Alternatively, the membership degrees could be compared or thresholded to obtain a crisp decision. Once the rules for obtaining a final decision from the leaves have been computed, the construction of the fuzzy decision tree is complete. A fully constructed fuzzy decision tree is depicted in Figure 5.15.

Classification of Test Data

After the fuzzy decision tree has been constructed, it can be used to classify samples from a test database. When a sample passes through the tree its attribute values are tested at the nodes and membership degrees to the different classes are computed. For samples that pass consecutive internal nodes cumulated membership degrees are computed (see Figure 5.16). When the entire training data set has passed through the tree and each sample has been attributed to a class with some membership degree, the classification is complete.

5.3.2 Fuzzy Input Data

Introduction

After having established the methodology to construct a fuzzy decision tree and its application to a set of crisp data, we now aim to extend this behavior towards processing of fuzzy input data. When processing feature vectors that contain fuzzy values, additional considerations need to be addressed:

- How to evaluate the degree to which the fuzzy value satisfies the node test?

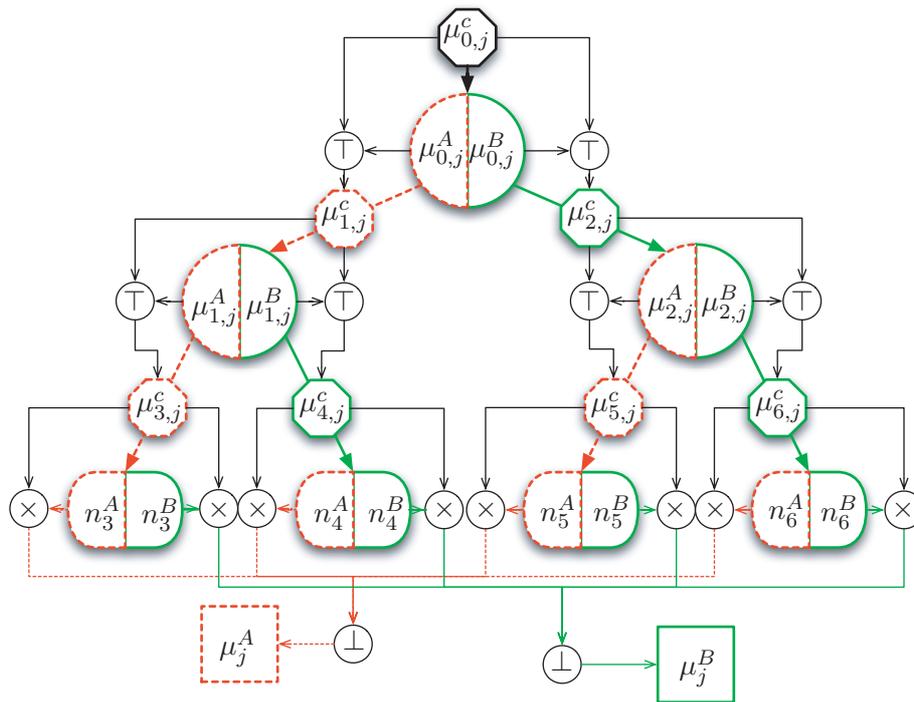


Figure 5.15: Fully constructed fuzzy decision tree: Inside the nodes are the membership degrees $\mu_{i,j}^A$ and $\mu_{i,j}^B$ to each class computed at the corresponding node test. The octagons contain the cumulated degrees of membership $\mu_{i,j}^c$ for the test sample. Each leaf contains the class densities n_i^A and n_i^B for the two classes.

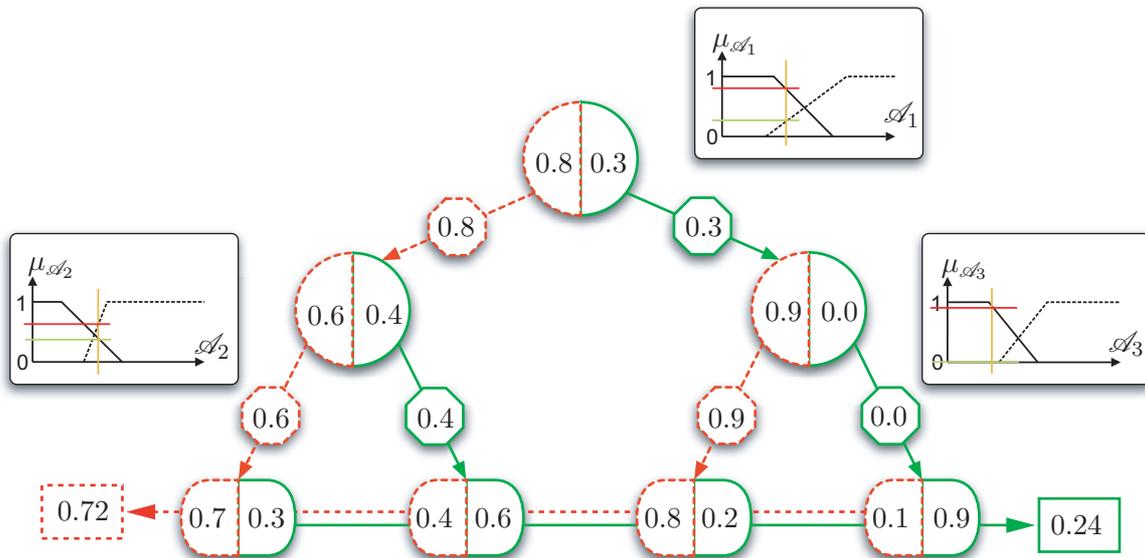


Figure 5.16: Example of a crisp test sample being processed by a fuzzy decision tree. The values indicated inside the nodes are the membership degrees to each class computed at the corresponding node test. The values in the octagons are the cumulated degrees of membership for the test sample. The values in each leaf correspond to the class densities of that leaf. The membership degrees for each class are computed by applying the max operator.

When working on crisp data, computation of the membership degree to a given class is straightforward. A given pair of membership function and attribute value directly result in the corresponding membership value. In the case of fuzzy attribute values, two functions (membership function and fuzzy attribute value) have to be taken into account in order to derive the membership value.

- How to find the node tests from the fuzzy values of the examples of the training data set? In Section 5.3 we have pointed out the need for construction of a fuzzy node test. Either based on the input data or on a priori knowledge. This now has to be extended to fuzzy node tests for evaluating fuzzy values.

Tree Construction

When building a fuzzy decision tree for processing fuzzy data we follow the same general scheme as described in Section 5.3.1. In order to take into account the fuzzy nature of the samples, modifications regarding the computation of membership degrees and the construction of the separator functions for the node tests have to be applied.

Computation of Membership Degree. In the case of crisp data we had to compute the membership degree of a given attribute value to a class, represented by a fuzzy subset (Figure 5.17(a)). In the case of fuzzy input data the attribute values of each sample are themselves fuzzy subsets. We therefore need to compute the membership degree of one fuzzy subset to another fuzzy subset (Figure 5.17(b)). This is a fairly common problem in fuzzy set theory and is usually expressed as a degree of *satisfiability* S of one fuzzy subset regarding another:

$$S(\mathcal{A}, \mathcal{B}) = \frac{\int \mu_{\mathcal{A} \cap \mathcal{B}}(x) dx}{\int \mu_{\mathcal{B}}(x)} \quad (5.5)$$

where \mathcal{A} and \mathcal{B} are fuzzy subsets and the integrals are computed over the definition domains of the fuzzy subsets.

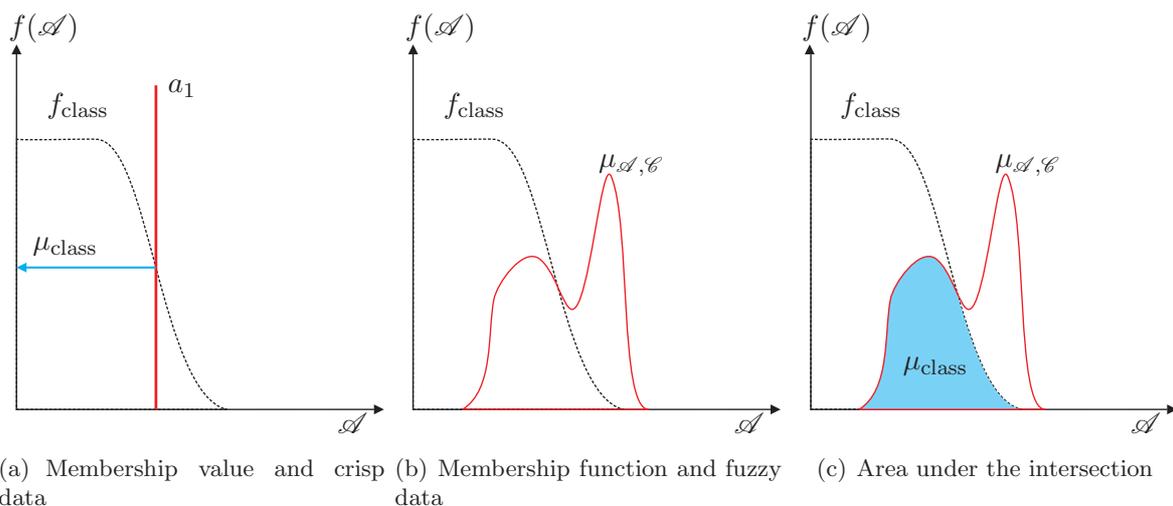


Figure 5.17: Satisfiability of a fuzzy value to a membership function

Fuzzy Node Tests from Fuzzy Data. Building a separator function from crisp attribute values is realized by computing a histogram of the test samples for each attribute. Now that we

wish to process fuzzy attribute values, we need to derive a method for constructing a histogram from fuzzy values.

A *fuzzy histogram* is computed analogously to a crisp histogram. During the construction process we simply have to add a function for each sample instead of a mere value. Before we add the attribute value of a given sample to the histogram, we normalize the attribute by dividing by the area under the membership function (see Figure 5.18). This entails two consequences:

- The contribution of the more precise attribute values is emphasized,
- The area of the histogram corresponds to the number of attribute values used for its construction.

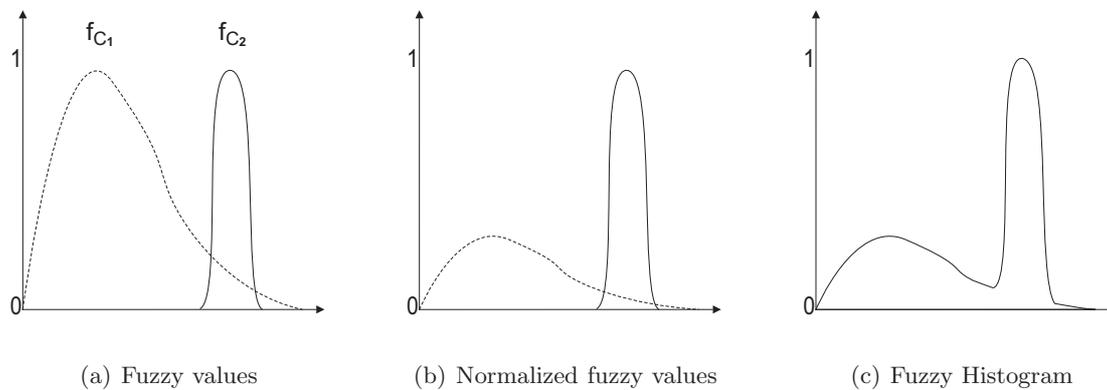


Figure 5.18: Computation of a fuzzy histograms

Finally, the area of the fuzzy attribute histograms are also normalized to 1. This is equivalent to saying that we impose for the histograms to contain an equal number of attribute values in order to be able to compare them (entering the same attribute values multiple times in the *smaller* histogram).

At each node the fuzzy histogram is computed for each attribute. The contribution of each sample is weighted by the cumulated degree of membership of the samples at that node (see Figure 5.19). The contrast between the histograms for the two classes is then computed for each attribute. The comparison of these contrast measures serves to identify the minimum contrast attribute at the node. This is the attribute with the highest discriminating power (Figure 5.20).

Classification of Test Data

We now dispose of all necessary components for building a fuzzy decision tree capable of working with fuzzy data. An example of a sample exhibiting fuzzy attribute values passing through the tree is illustrated in Figure 5.21.

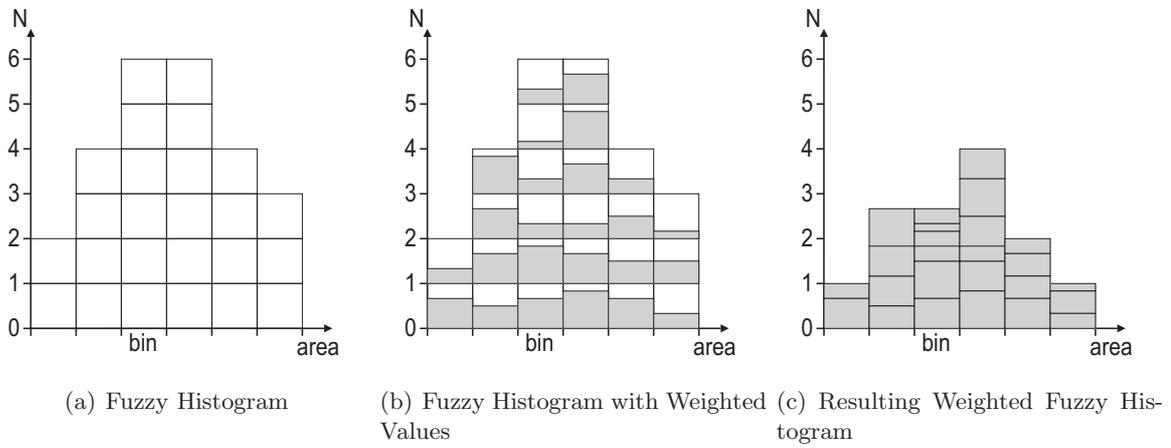


Figure 5.19: Computation of a weighted fuzzy histogram: (a) The fuzzy histogram constructed for a given fuzzy attribute \mathcal{A}_1 using the entire training database. Each empty square corresponds to a test sample with a cumulated degree equal to 1 (true for all samples at the root node). (b) The contribution of each sample (white squares) has been weighted by the respective membership degrees. This results in a weighted contribution (grey rectangles) for each sample. (c) Resulting weighted fuzzy attribute histogram at a the current node.

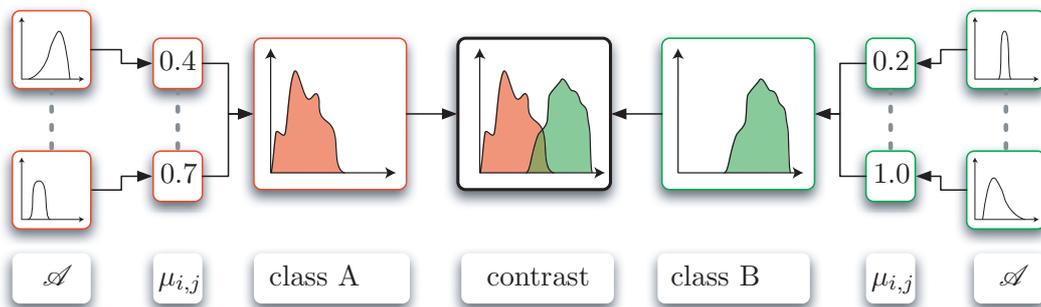


Figure 5.20: Contrast between two classes for a given attribute. For each class (class A to the left and class B to the right) the attribute histogram is computed from the weighted sample of the corresponding class. The contrast is then obtained from the histograms of the two classes for a given attribute.

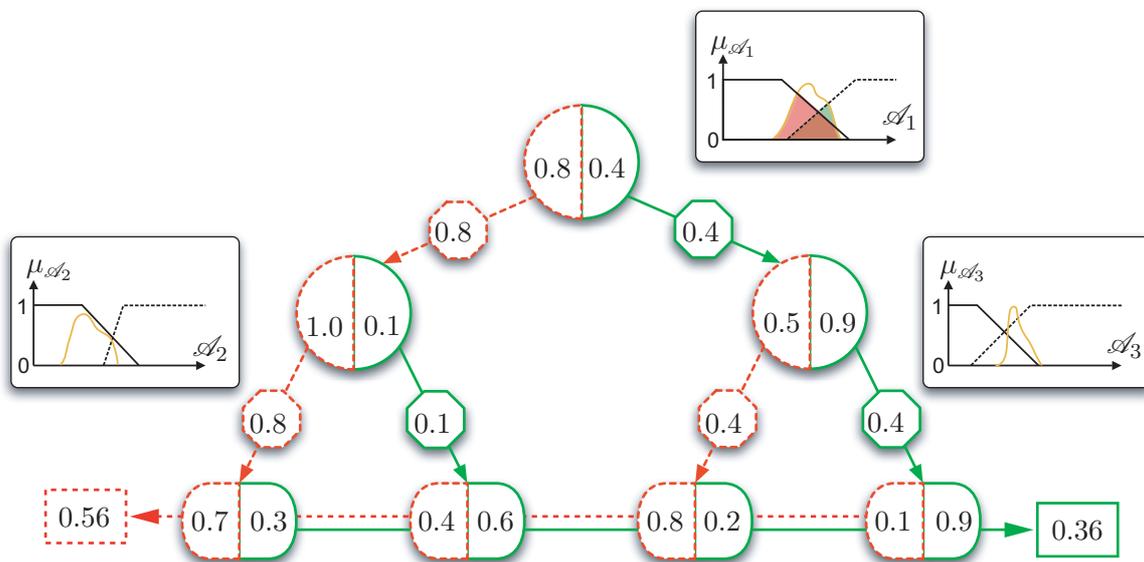


Figure 5.21: A fuzzy decision tree processing a given fuzzy example of the database. For each node the separator functions are given along with the fuzzy attribute values. The intersection between a fuzzy attribute value and the membership functions to the different classes are shaded to represent the satisfiability of the value to each class.

5.4 The Proposed Approach: Extensions and Improvements

In this section we will take a closer look at the specificities of our implementation, and the details of some issues that need to be addressed. The approach we have used for classification of the fuzzy attribute values computed in Section 3.4.1 is based on the approach developed in (Bothorel 96). There are however some particularities concerning the nature of our data and the structure of the tree that need to be taken into account. Details of the presented approach concern:

- the databases of synthetic training data and test data used to validate the algorithm.
- the measure of discrimination used for ranking the attributes,
- the construction of the membership functions for the node tests (also called separator functions) and how these are connected to a priori assumptions about the attributes,
- the possibility of using an attribute multiple times in the tree,
- the computation of the cumulated degree of membership,
- how to obtain a final decision,

5.4.1 Synthetic Data

In order to validate the correct and expected behavior of our implementation, we have created a number of synthetic data sets. There are two main reasons for us to discuss these data in some detail. Firstly, creating well adapted synthetic fuzzy training data and test data is not a trivial task, and may present a scientific interest by itself. Secondly, evaluation of the behavior of the fuzzy decision tree when working on synthetic data is used for illustration. The fuzzy decision tree algorithm used is quite complex. In order to understand the behavior of the fuzzy decision tree in different situations we will control as many parameters as possible. It is much easier to interpret the behavior of the tree for a case where we know what kind of result to expect.

Crisp Data A straightforward approach for implementing crisp training data sets results from using only two attributes. The training data as well as the classification result can then be represented by a two-dimensional image. For the training data, each sample is positioned in a coordinate system representing the two-dimensional feature space.

Fuzzy Data When working with fuzzy input data, the training samples cannot simply be represented as points in the feature space. In fact, each sample has an associated membership function for each of the attributes. In the simplified case of trapezoidal membership functions that translates to a given fuzzy interval in each dimension of the feature space.

In description of fuzzy values for continuous numeric attributes, trapezoidal fuzzy numbers are widely used since they can sufficiently represent fuzzy values and they are simple to describe and process (Zimmermann 01). We have therefore used trapezoidal fuzzy numbers for the synthetic data presented in this section. For purposes of simplicity, the samples of the two classes have been represented as crosses and rectangles respectively. The extensions of these symbols in the two dimensions of the feature space correspond to the support of the respective trapezoidal fuzzy numbers. Examples for this approach are given for different training data sets in Figure 5.22.

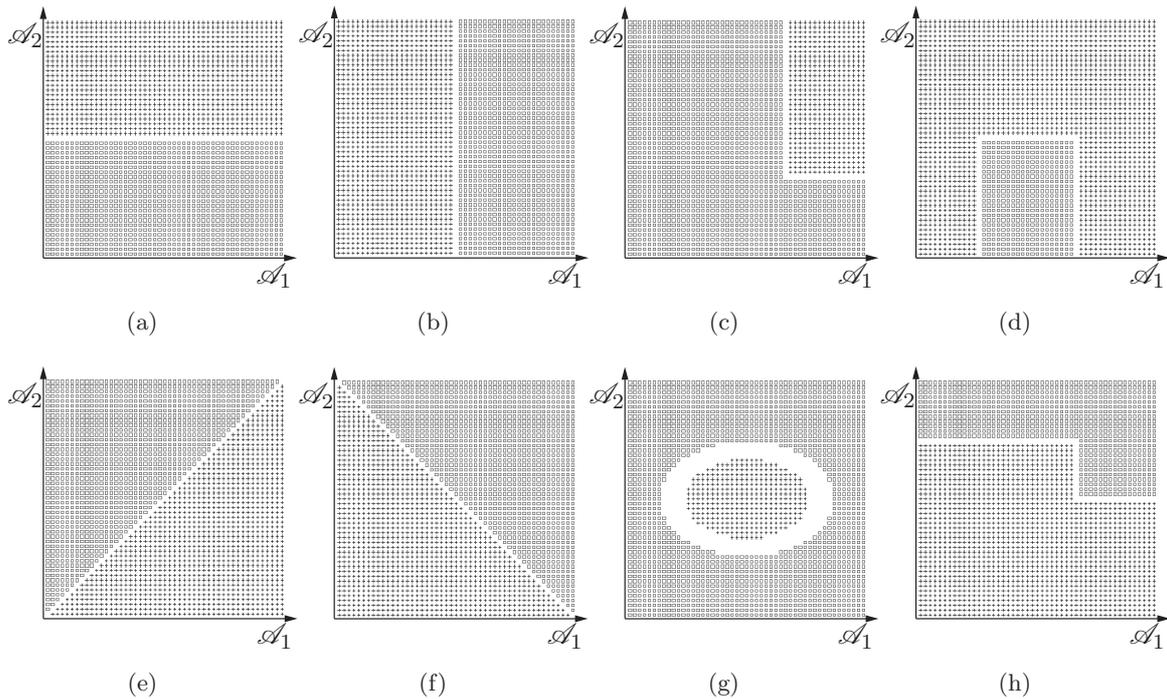


Figure 5.22: Synthetic fuzzy training data sets for the fuzzy decision tree. The two classes are marked as crosses (A) and boxes (B). For each sample the corresponding class symbol is drawn in the feature space according to the values of attribute \mathcal{A}_1 and \mathcal{A}_2 . The first row (a-d) shows examples of training data sets with independent attributes. In (a) and (b) one attribute alone provides perfect discrimination while the other one has no discriminating power at all. A case where both attributes are needed for perfect discrimination between the classes is shown in (c). The attributes are independent since a simple conjunction of their associated thresholds results in the correct generalization. The same is true for example (d). However, attribute \mathcal{A}_1 is not monotone in this example (the attribute value needs to be inside/outside a given interval to discriminate between the two classes). The data sets depicted in the second row (e-h) are not independent and thus difficult to correctly classify with the described fuzzy decision tree.

Data Parameters In addition to the distribution of the training samples and the form of the trapezoidal fuzzy numbers there are some parameters concerning the synthetic data that may be modified. These concern the distance between the two classes (the gap in the images that separates the two data sets), the number of training samples in a given interval of attribute values, and the noise in the data sets. An example for different noise levels is illustrated in Figure 5.23.

5.4.2 Node Tests

Obtaining the a priori knowledge necessary for building separator functions is a considerable challenge. A very large number of cases would need to be studied for an expert to be able to even acquire the necessary knowledge. For DBT new considerations have to be taken into account, and assumptions that were made for 2D standard mammography cannot generally be applied. Furthermore, since the DBT system providing the input data is still an emerging modality, there are at present only few experts starting to get familiar with the image data produced by these systems. In addition, there may be some new features visible in these images to characterize breast lesions, that have yet to be discovered. All these premises have led us to adopt an approach where the construction of the fuzzy decision tree is performed by learning

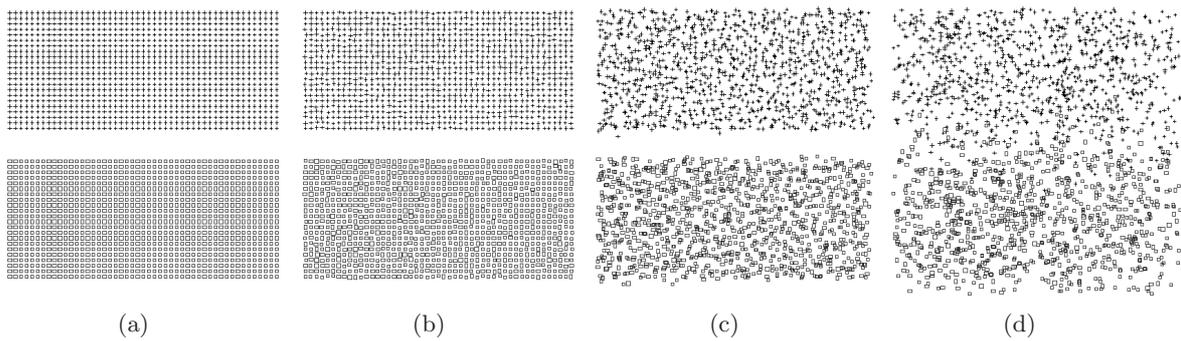


Figure 5.23: Impact of noise on the training data: no noise (a) and different levels of noise (b - d). Noise was realized as random variables in the two dimensions of the feature space that have been added to the respective fuzzy attribute values.

from a database of test samples (as described in Section 5.3.2).

Attribute Properties It is important to note that some a priori assumptions about the attributes have been made that justify the applied processing. These assumptions include some key properties of the attributes that characterize the objects of interest:

- Monotony
- Regularity
- Monolobe

These conditions are true for the majority of the attributes used in computer aided detection in mammographic images. Examples include *gradient*, *compacity*, *elongation*, *homogeneity*, and many histogram based gradient and gray level features. An example for an attribute that does not satisfy this condition is the *area* of a mass lesion. Contrary to microcalcifications (that are small objects whose area is thus below a *single* given threshold), masses typically range in a certain *interval* of area values. If the used attributes do not satisfy the conditions, the processing needs to be adapted to the properties of the attributes, and cannot be applied as are. This problem is illustrated in Figure 5.24.

The processing of the separator functions described above entail another important limitation to the algorithm: attributes cannot be used multiple times in the same path through the tree (from the root node to a given leaf). An example for this property is given in Figure 5.25.

Since we cannot use attributes multiple times in a given path through the tree, only independent attributes can be reliably separated. Here, when we are talking about *independent* attributes we are referring to the relation between the different attributes used to characterize the object (e.g. in a class definition). If we are looking for an object that is *round* AND *small* then the corresponding attributes for the *compacity* and the *area* are considered to be independent since they are combined by a simple conjunction only. On the other hand we have attributes where each attribute value directly depends on the corresponding value of another attribute. This may correspond to an object definition where the mean gradient along the contour needs to be at least twice as high as the mean gradient inside the contour. This definition of independent attributes does therefore allow redundant attributes to be used. Even though their values are not independent, the relation between their values does not contribute to the discrimination between the classes. An example would be an object that is *round* AND not *elongated*. This could be represented by a small compacity and a small Ferret ratio. Even

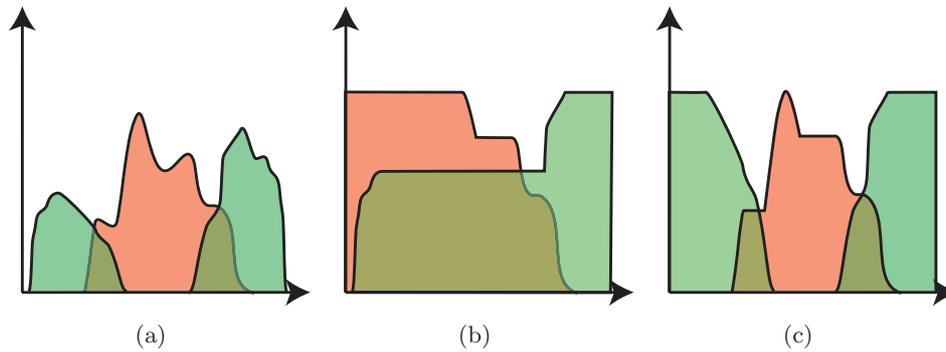


Figure 5.24: Example of a non-monotone fuzzy attribute. The original fuzzy attribute histogram is illustrated in (a). Applying the same processing as for monotone attributes leads to the node tests depicted in (b). This is clearly not a desirable generalization of the training data set. In fact, we need to adapt the processing applied to the fuzzy histograms (see Figure 5.9). For non-monotone attributes the processing needs to be modified according to a priori knowledge about the attribute. For example if the attribute is the area of a mass lesion, then we know that masses exhibit areas in a certain interval of values. The lobe of this interval can therefore be reconstructed from both sides up to the maximum. The two other lobes correspond to the objects that are either too small (left lobe) or too big (right lobe) to be a mass lesion. These lobes are treated independently according to Figure 5.9.

though these attributes are clearly related and most objects should behave very similarly for the two attributes, they are not related in the class definition. Their connection is therefore not important for the discrimination between the classes, and does suffer from the limitation of the decision tree.

If the attributes are not independent from one another, the classification result obtained from the fuzzy decision tree does not always produce a good generalization of the training data. Because this is an inherent property of the algorithm, it cannot be recovered by well adapted parameter settings. An illustration of this issue on a simple example of synthetic data is given in Figure 5.26. A popular approach to overcome this type of problem in principal component analysis (PCA) also called Karhunen-Loève transform (Fukunaga 90).

However the limitations mentioned above (independent attributes, each attribute used only once), do not concern our application. All attributes extracted from the fuzzy contours (Section 3.4.1) are independent according to our definition. If this is ensured, there is no need for multiple attribute use. If we consider a path through the fuzzy decision tree to represent a series of conjunctions of rules (decisions on attribute values), then it does not make sense to look for objects that are *round* AND *small* AND *round*.

Measure of Discrimination When choosing a measure of discrimination well adapted for a given problem the following points have to be taken into account:

- The entropy measures are generally better suited for use with symbolic attributes than when processing numerical attributes. Furthermore, the choice of a threshold value by use of an entropy measure may be challenging on a noisy database.
- It is always difficult to compare attributes having different numbers of possible values. Especially when using entropy based attributes, it is necessary to adapt binary and multi-valued attributes in order to guarantee a fair comparison.
- Every measure of discrimination has its specific properties, and we need to take into

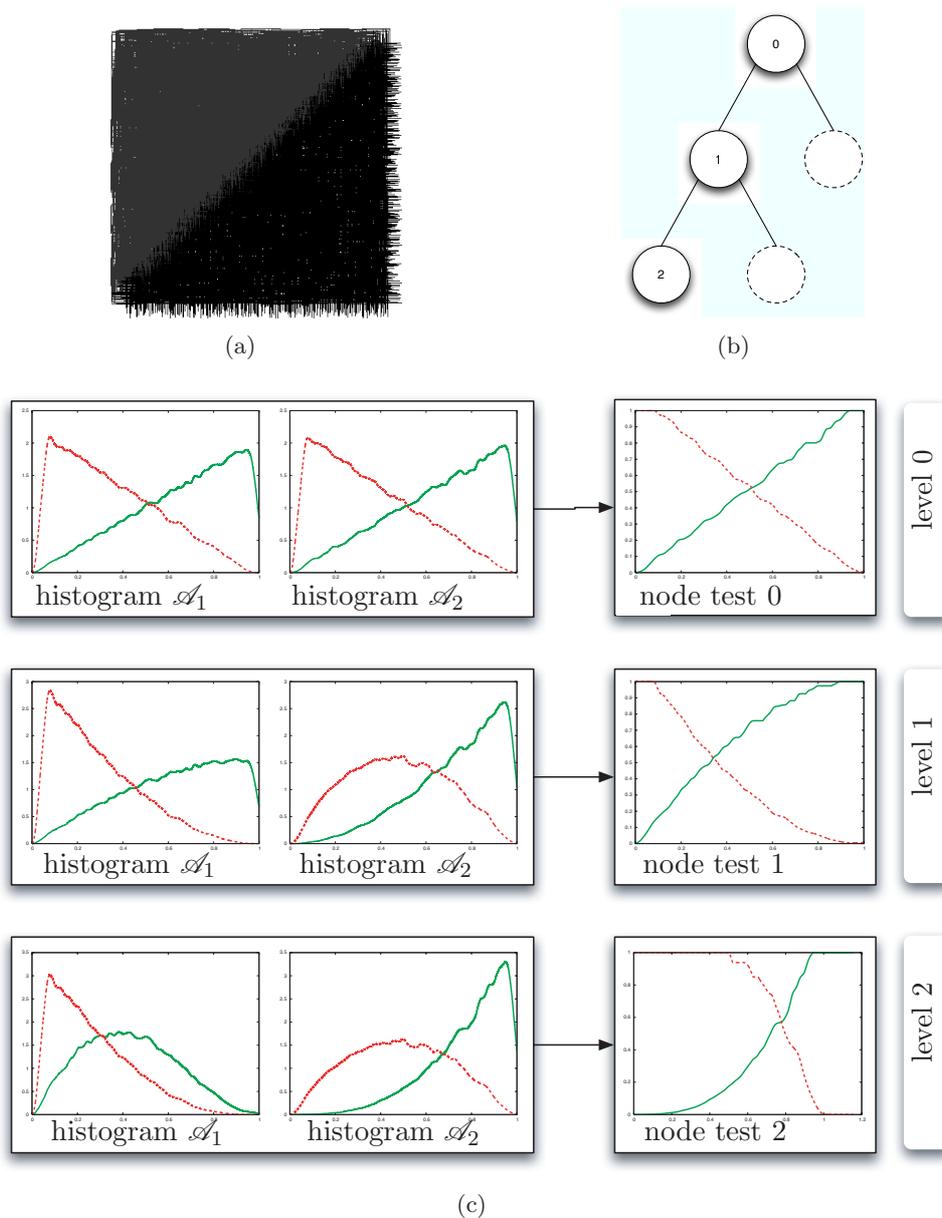


Figure 5.25: Problems with multiple attribute use. Illustrated are a set of fuzzy training data (with dependent attributes) (a), a part of the three constructed for this training data set (b), and the corresponding attribute histograms and node tests (c). The attribute histograms for the two attributes (c:left) and the resulting node test based on the most discriminating attribute (c:right) are illustrated for three consecutive levels of the tree. At the third level, both attributes have already been used. A number of issues can be observed. Firstly, the test at the root node does not allow to completely eliminate any of the samples ($\mu_{i,j}^A = 0 \vee \mu_{i,j}^B = 0$). Secondly, the processing applied to the fuzzy attribute histograms results in a node test at level 1 (build from the histogram of \mathcal{A}_1) that is very similar to the test at the root node (level 0). Consequently, the histogram of \mathcal{A}_2 changes very little between level 1 and level 2. When \mathcal{A}_2 is re-used at level 2, the resulting node test is again quite similar to the one at the root node. The problem lies in the processing described in Figure 5.9. In the case of multiple attribute use, this processing (and especially the normalization and the extension of the maxima to the limits of the domain) eliminates the little discrimination obtained at each level of the tree.

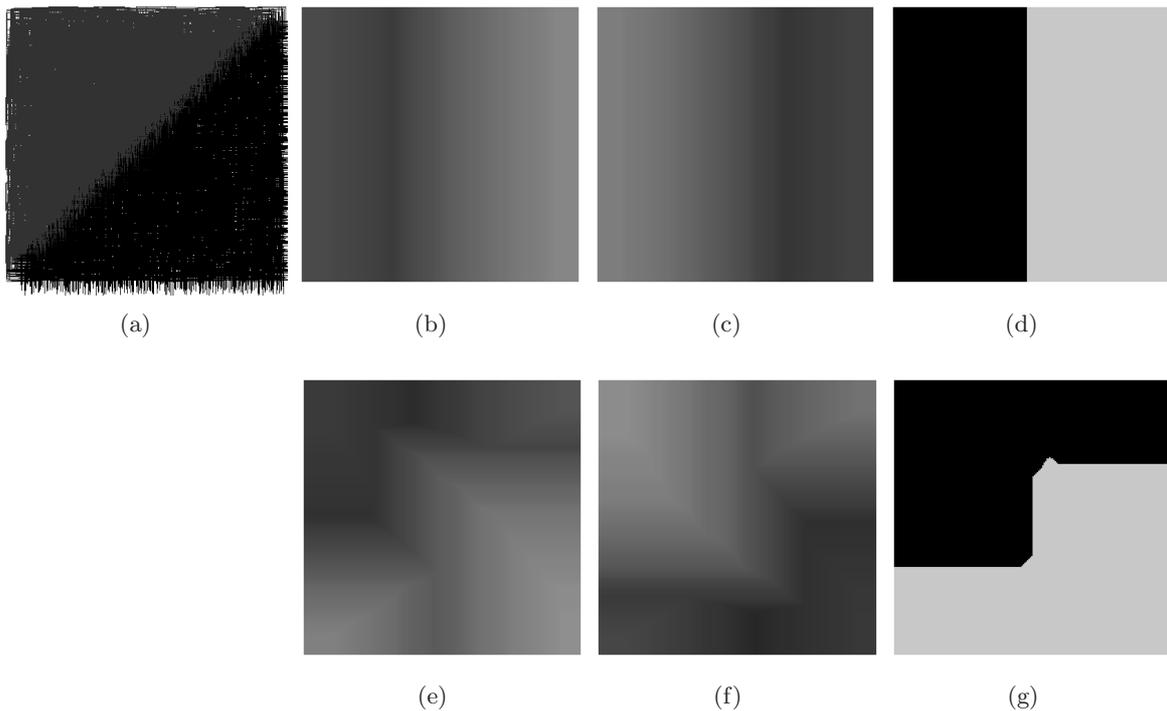


Figure 5.26: Processing of dependent attributes. The training data is shown in (a). The other images in the first row show the separation generated by the test at the root node. Given are the membership degrees to the two classes ((b) and (c)), and the defuzzified classification result (d). The second row shows the images corresponding to the final classification results ((e)-(g)). Since we are working with a limited feature space of only two attributes the maximum number of nodes in the fuzzy decision tree is equal to three (if each attribute may only be used once). The root node will use one attribute for the node test and the remaining attribute may be used by the two children of the root node. The root node provides a separation of the feature space perpendicular to the feature axis corresponding to the node test. The two resulting regions are subsequently divided by the two children nodes. Using each of the two attributes only once, the closest approximation of the diagonal separation of the feature space present in the training data is given in (d).

account the nature of the available database. A measure that favors the construction of small leaves (containing few samples) is poorly adapted to a noisy database.

The measure of discrimination used here is based on a contrast measure between two fuzzy subsets as described in Section 5.3.1. It is better adapted to deal with a small and noisy database than most other measures (Bothorel 96).

Satisfiability When looking for a well adapted measure of satisfiability we have to take three properties into account. These concern the importance of the maximum, the processing of uncertainty and the translation of uncertainty in the fuzzy attribute values. The satisfiability discussed in Section 5.3.2 represents a compromise that delivers good results corresponding generally well to an intuitive approach. A more in-depth discussion of the subject can be found in (Rifqi 96).

Cumulated Degree of Membership We have chosen the min operator (t-norm by Zadeh (Zadeh 65)) for computing the cumulated degree of membership since it allows for the cumulated

degree to be independent of the number of node passed and thus of the length of the path through the tree. Another advantage of the min operator lies in the fact that it allows to process even some cases where the attributes values are not quantified. For attributes where only the order of the values is known the min operator still produces valid results.

Once a test sample passes through the fuzzy decision tree, we combine its cumulated degree in each leaf with the membership degrees to the different classes. As we have pointed out before, the relationship between the different membership degrees and a given class are lost when several membership degrees are being combined into a single cumulated degree of membership. The cumulated degree of membership of each sample at a given node is used for building the corresponding node test. The contribution of each sample to the fuzzy histogram is thus weighted by its cumulated membership degree at that node (see Figure 5.19).

Obtaining a Final Decision Working on computer-aided detection, missed cancers are our main concern. We therefore tend to adapt a very cautious methodology where we do not neglect any information that has been extracted. In the case where we are looking to obtain a final decision from the set of leaves of the fuzzy decision tree, we consider the following situation: If there exists a combination of rules that leads us to the conclusion that the inspected radiological finding (the sample) corresponds to a cancer, then we should consider this the most important information. Proceeding in such a cautious way leads us to adopt the max operator for computing the final membership degrees. The membership degree of a given test sample to each class is then the max of all membership degrees in all leaves of this sample to this class.

By combining the results of this operation over the full set of leaves, we obtain for each test sample a resulting membership degree to each class. This process is illustrated in Figure 5.27, using the max operator as the indicated t-conorm.

Application Example Finally we have at our disposal all necessary parts for construction as well as prediction using a fuzzy decision tree. The performance of the presented algorithm on a set of synthetic data is illustrated in Figure 5.28 and Figure 5.29.

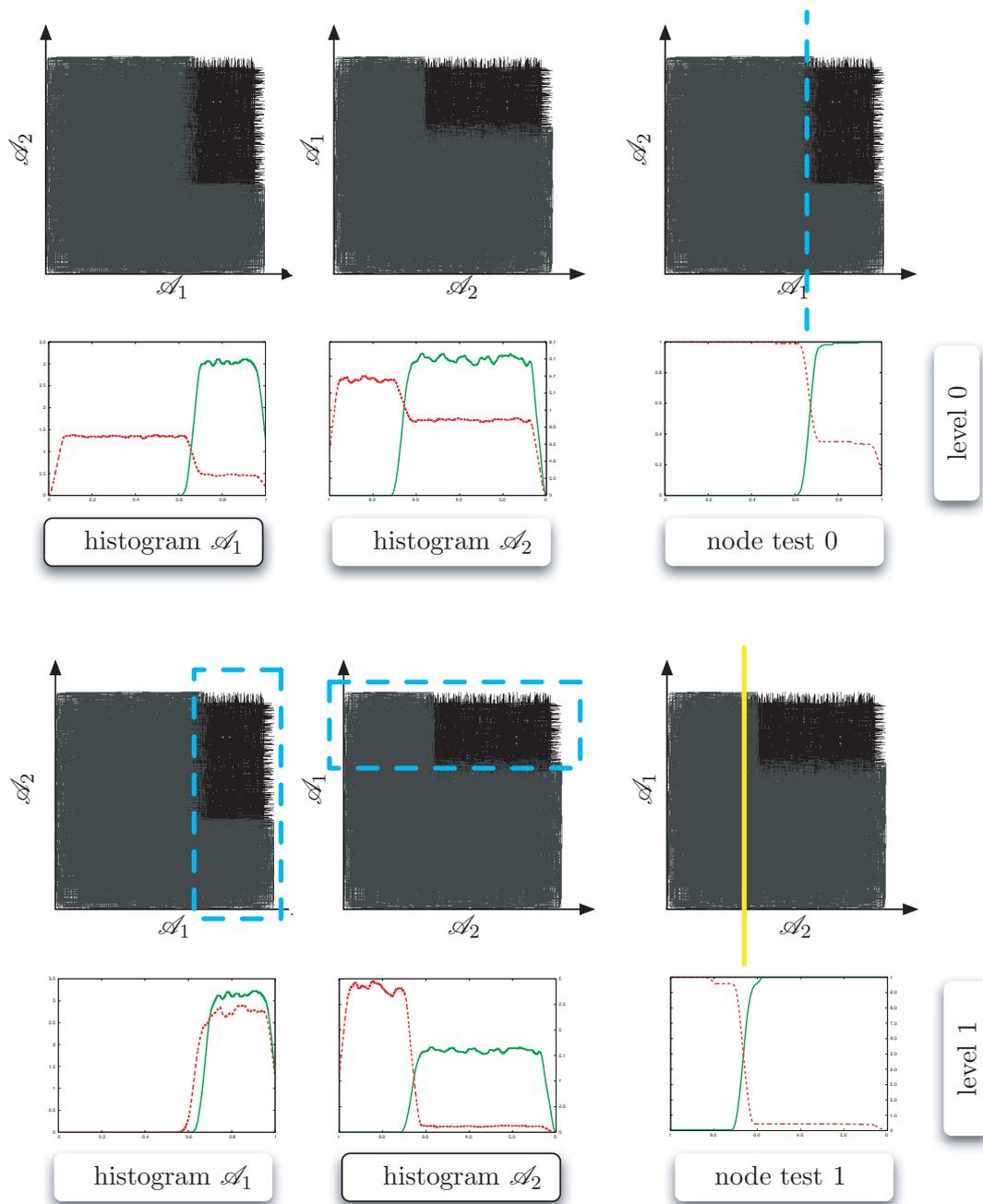


Figure 5.28: Example of classification of synthetic data using the fuzzy decision tree. The first line illustrates the processing performed at the root node. For visual support, the training data has been depicted on top of each plot to illustrate the obtained partition of the feature space. A histogram is computed for each attribute in order to find the most discriminating attribute. The histogram corresponding to this attribute (outlined in black) is then processed to obtain the node test. The satisfiability to this test determines the path of each sample (the degree associated to each path). The second row shows the corresponding procedure for the left child of the root node. The samples that arrive in this node with a cumulated membership degree $\mu_{i,j}^c \neq 0$ are indicated by the dashed box. Attribute \mathcal{A}_1 has already been used and does not provide further discriminating power. Attribute \mathcal{A}_2 is thus chosen for the node test. The final classification results for this example are given in Figure 5.29.

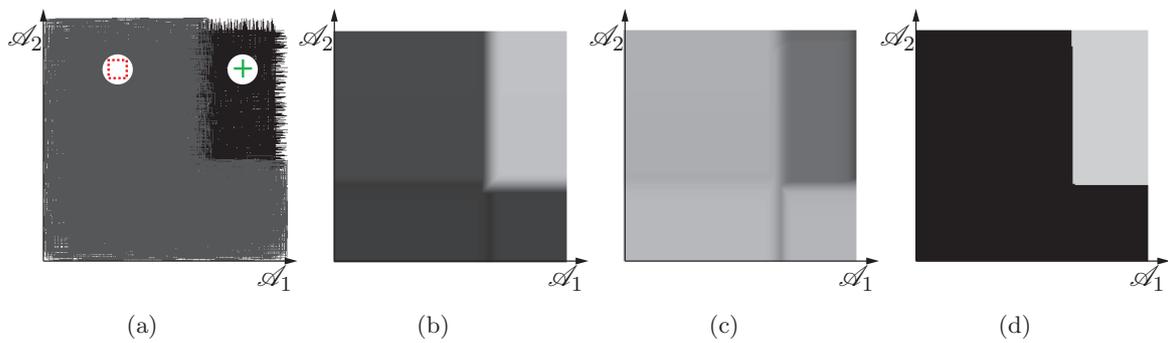


Figure 5.29: Classification results for the example of synthetic data described in Figure 5.28. Training data (a), membership degrees (higher grey level indicates a higher membership degree) to the two classes ((b) and (c)), and the defuzzified classification result (d) obtained from the max of (b) and (c).

5.5 Classifying Mass Attributes from DBT Projection Data

In this section we demonstrate the classification of clinical patient data using the fuzzy decision tree. We will limit ourselves to the construction of very simple fuzzy decision trees to illustrate the methodology described in the preceding sections. An actual classification of the clinical database is demonstrated in Chapter 6.

The Database A database of 17 breast masses has been processed by our algorithm resulting in a set of cumulated fuzzy attributes for each mass. We now apply the fuzzy decision tree methodology for feature selection in order to identify the most discriminating attributes.

The size of the training database has a significant impact on the generalization power of the constructed fuzzy decision tree. Furthermore there are some important algorithmic limitations that need to be taken into account when working on a small database.

Firstly, the presented approach for constructing a fuzzy decision tree is based on attribute histograms of the training samples. The smaller the training database, the less it will be representative of the actual population. In addition to that, the database is divided into subsets at each node. The fuzzy attribute histograms are then computed only considering the contribution of each sample to the histogram weighted by the cumulated degree of each sample at that node. This is generally a desired effect for combining the rules represented by the nodes and even necessary for finding the attribute that best discriminates the remaining subset of training samples. It may however have a negative impact in the case of a small training database. In fact, this method may lead to attribute histograms that are less and less representative of the population. For a very limited database we may therefore choose not to weight the attribute histograms by the cumulated degree of each sample. Instead, we will use the attribute histograms constructed over the entire training database at all nodes of the tree. Even though this will potentially introduce errors in the classification result, we consider that the benefits outweigh this drawback in the case of a very limited training database.

A second consideration regarding the training database needs to be taken into account. If we are dealing with a training database that cannot be considered to be representative of the population, the results of the tree need to be normalized for the number of sample of each class contained in the training database. During tree construction, the samples arriving in each leaf are used to compute class densities for each leaf (see Section 5.3.1). However, when computing these values, we need to take the composition of the training data into account. If we dispose of two times as many training samples belonging to class A than samples belonging to class B , it seems advised to weight the class densities of each leaf by these values. The adjusted class density $n_i^{A'}$ is then computed using:

$$n_i^{A'} = \frac{\aleph_B \cdot n_i^A}{\aleph_B \cdot n_i^A + \aleph_A \cdot n_i^B}, \quad n_i^{B'} = \frac{\aleph_A \cdot n_i^B}{\aleph_A \cdot n_i^B + \aleph_B \cdot n_i^A}, \quad (5.6)$$

where n_i^A and n_i^B are the densities of samples of class A and B in leaf i respectively. \aleph_A and \aleph_B denote the cardinalities of the training data sets for classes A and B respectively. This normalization leads to a case where we compare the percentage of samples of class A (relative to the total number of samples of class A) to the percentage of samples of class B . Without normalization the comparison would be directly performed between the number of samples of each class and would thus not take into account the total number of sample of the respective classes.

This adjustment of the class densities in each leaf is not necessarily required when working on a representative database. In that case a larger number of samples for a given class represents a natural phenomenon that we may wish to take into account (if spiculated breast masses occur

ten times more often than circumscribed breast masses we may choose to have the decision tree take this elevated probability for the occurrence of one class into account). On the other hand, if we aim to analyze the object exclusively based on its inherent properties and follow more of a possibilistic approach, we should normalize the class densities in each leaf for any size of database.

Experiments and Results The above database has been processed independently by two approaches based on different assumptions about the objects of interest (see Figure 3.32). Here, we will only show classification results for the characterization obtained from applying the circumscribed mass model. The classification result then corresponds to answering the question whether the initial hypothesis (the object corresponds to a circumscribed breast mass) is in fact valid or not.

Figures 5.30 to 5.32 show results obtained for six exemplary fuzzy decision trees. Each tree contains a single node (the root node). These examples illustrate well the need for a feature selection step prior to the construction of the full decision tree. The first two examples (Figure 5.30) show attributes with an elevated discriminating power that should be used in the fuzzy decision tree. The third and the fourth example (Figure 5.31) represent attributes with a lower discriminating power that will probably not lead to a significant improvement of the classification result. The most interesting examples are illustrated in Figure 5.32. The node tests of the fuzzy decision tree are designed to provide a generalization of the fuzzy attribute histograms. However, they assume two histograms that do not exhibit lobes at the same values (since that represents a very high contrast and thus a low discriminating power). Good attributes therefore lead to discriminating node tests, bad attributes on the other hand may lead to arbitrary and false generalizations. These attributes may still - by coincidence - lead to a low classification error (especially on a small database), as it is the case in Figure 5.32. The contrast is thus a much better indicator of the discriminating power than the resulting classification error rate. On the contrary, the node tests do not reflect the training database, and the use of these attributes in the fuzzy decision tree may actually degrade the classification results. A feature selection step is therefore especially important when working with a limited database.

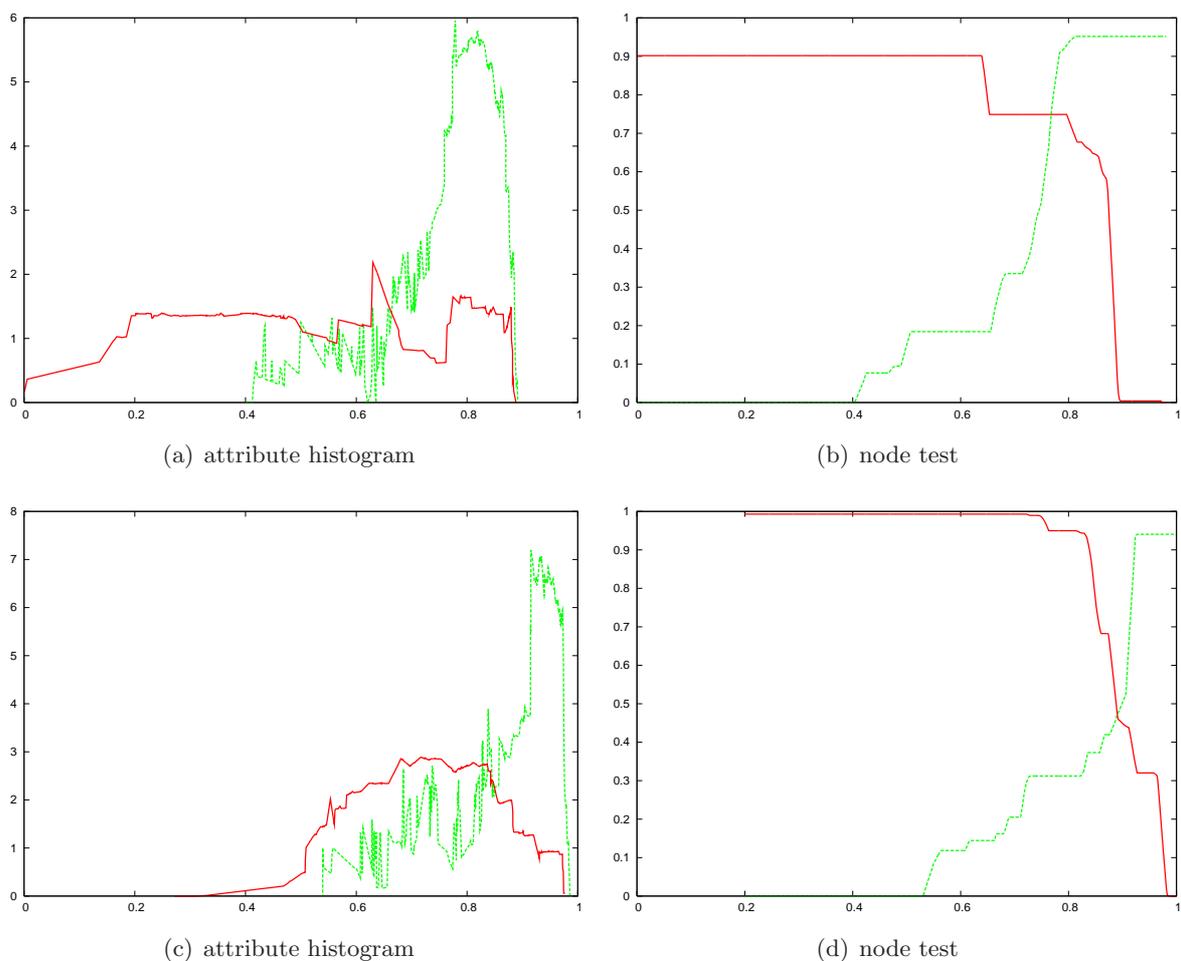


Figure 5.30: Example of fuzzy tree construction for the attributes *compactness* (a,b) and *Ferret ratio* (c,d). The dashed line corresponds to the class *circumscribed mass* while the solid line correspond to the class *not circumscribed mass*. The contrast of the attribute *compactness* is $C = 0.286$ and the classification error is $E = 23\%$. The contrast of the attribute *Ferret ratio* is $C = 0.389$ with a classification error of $E = 29\%$. These are examples of an attributes with elevated discriminating powers. The contrasts are low, and the node tests provide good generalizations of the fuzzy attribute histograms.

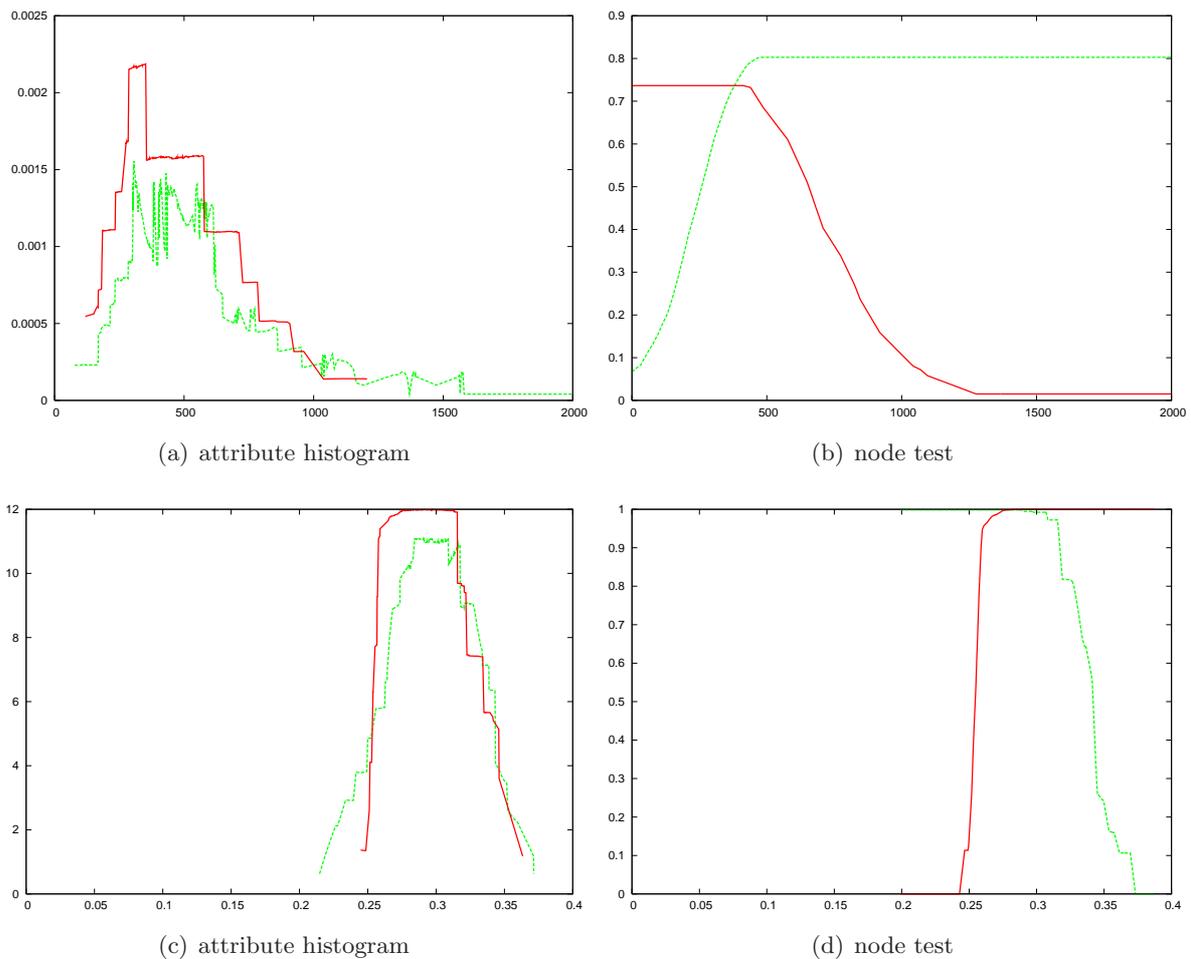


Figure 5.31: Example of fuzzy tree construction for the attributes *homogeneity* (a,b) and *variance of the gradient direction along the contour* (c,d). The contrast of the attribute *homogeneity* is $C = 0.512$ and the classification error is $E = 47\%$. The contrast of the attribute *variance of the gradient direction along the contour* is $C = 0.766$ with a classification error of $E = 58\%$. These are examples of attributes with low discriminating powers. The contrasts are elevated, and the node tests do not provide good generalizations of the fuzzy attribute histograms.

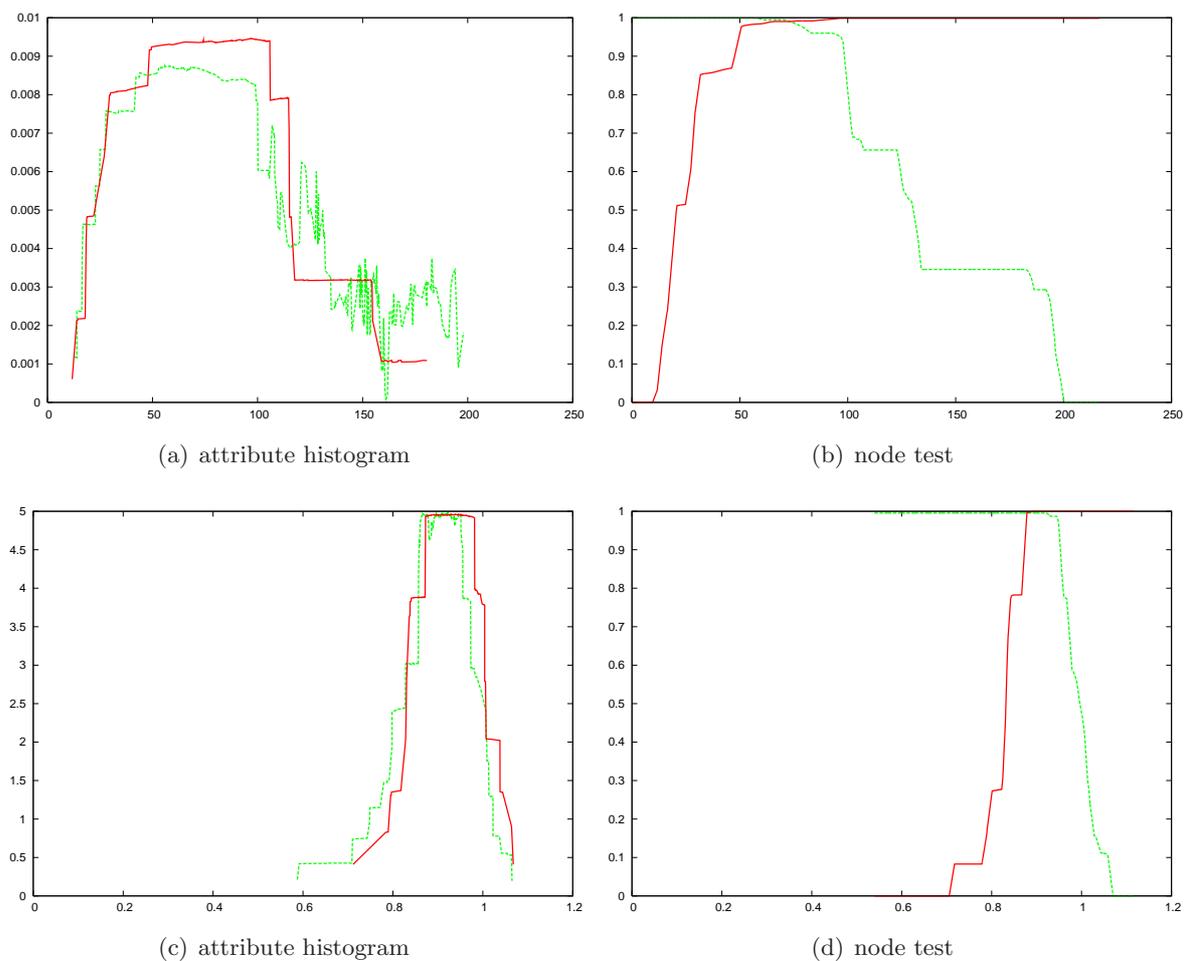


Figure 5.32: Example of fuzzy tree construction for the attributes *Ferret min* (a,b) and *mean gradient direction along the contour* (c,d). The contrast of the attribute *Ferret min* is $C = 0.804$ and the classification error is $E = 41\%$. The contrast of the attribute *mean gradient direction along the contour* is $C = 0.765$ with a classification error of $E = 29\%$. These are examples of an attributes with very low discriminating powers. The contrasts are elevated, and the node tests do not provide good generalizations of the fuzzy attribute histograms.

5.6 Performance Evaluation and Comparison

Since the classification result of the fuzzy decision tree is expressed as a confidence degree to each class, a defuzzification step is needed in order to obtain a crisp prediction L_p of the class label. This can be achieved by applying the max operator to the obtained membership degrees. In the case of two classes A and B the predicted class label is:

$$L_{p,j} = \begin{cases} A & \text{if } \mu_j^A \geq \mu_j^B \\ B & \text{if } \mu_j^B > \mu_j^A \end{cases} \quad (5.7)$$

where $L_{p,j}$ is the predicted class label for the j th sample.

In the proposed framework, multiple fuzzy decision trees are applied to test different a priori assumptions about a given particle. In order to obtain a decision for a particle from multiple fuzzy decision trees we need to combine their respective results. However, the results obtained from different fuzzy decision trees may be conflicting. We then seek additional information to be able to decide which result we can trust the most. To this end we apply the confidence degree proposed in (Bothorel 96).

For a given degree of membership μ^A to the class A we will trust the degree of membership μ^B to the class B more if its distance to μ^A is elevated. For example, for $\mu^A = 0.8$, we will have more trust in the result if $\mu^B = 0.1$ than for $\mu^B = 0.7$.

The confidence degree is thus defined as:

$$D_c = \frac{|\mu^A - \mu^B|}{\mu^A + \mu^B} \quad (5.8)$$

where D_c is the confidence degree, and μ^A and μ^B are the degrees of membership to the classes A and B respectively.

In our case, where we have different object classes, each with μ^A the membership degree to the class A and $\mu^{\bar{A}}$ the membership degree to the complement of the class A , Equation 5.8 becomes:

$$D_c = \frac{|\mu^A - \mu^{\bar{A}}|}{\mu^A + \mu^{\bar{A}}} \quad (5.9)$$

In our framework, we deal with multiple object classes and thus multiple resulting degrees of membership. We therefore propose an extension of the confidence degree to N_o object classes:

$$\mathcal{D}_c = \frac{\sqrt[N_o]{\sum_{i=0}^{N_o} D_c^i \cdot \mu^i}}{\sum_{i=0}^{N_o} D_c^i \cdot \mu^i}, \quad (5.10)$$

where \mathcal{D}_c is the *cumulated confidence degree*, D_c^i is the confidence degree for class i , μ_i is the membership degree to the class i and N_o is the number of object classes. The cumulated confidence degree expressed the confidence in the decision obtained from multiple fuzzy decision trees (membership degrees to different classes L_p). In a CAD application we may choose to include information about the confidence in a decision when presenting the result to the radiologist. For example, we may choose to enhance the contrast of a particle relative to the confidence degree obtained for this particle. An application of these metrics is given in Chapter 6.

5.7 Conclusion

In this chapter we have presented a fuzzy decision tree classifier working on fuzzy data. The fuzzy decision tree is based on the work by (Bothorel 96). Several new improvements and extensions

have been proposed. The original fuzzy decision tree has been applied to minimal trees (one node and two leaves) only. An extension to more complex trees has been studied. To this end we have investigated the multiple use of attributes in a path through the tree. Furthermore we have introduced a new method to process attributes that are not monotone.

The main contribution regarding fuzzy decision trees lies in considerations concerning the database. Firstly, a synthetic database has been constructed, and the behavior of the fuzzy decision tree has been illustrated for different synthetic examples. The processing of clinical patient data has lead us to consider further improvements of the algorithm. For small databases, we propose to take into account a weighting factor for the class densities in each leaf. A feature selection step based on the fuzzy decision tree has been proposed. This is especially important for small databases where the dimensionality of the extracted feature vector is close to the cardinality of the database. Finally, we have introduced the *cumulated confidence degree* as a measure for the confidence in the classification result of a single object that has been processed according to different a priori assumptions about the object.

Furthermore, alternative measures for identifying the most discriminating attribute should be investigated. The use of a different method may increase the generalization power of the fuzzy decision tree and may resolve the problems of multiple attribute use, since today the main limitation for using attributes multiple times lies in the generalizations made during the construction of the node tests. Finally, different t-norms and t-conorms for the aggregation of the membership degrees may provide a positive impact on the classification performance.

There are three principal means of acquiring knowledge: observation of nature, reflection, and experimentation. Observation collects facts; reflection combines them; experimentation verifies the result of that combination.

Denis Diderot

6

CAD Results for Mass Detection on Clinical DBT Data

In the preceding chapters we have discussed the individual processing steps of the proposed CAD framework. Here, we show the application of the algorithm for mass detection on clinical patient data. We present the data sets of clinical DBT cases used for evaluation of the algorithm. Then, a quick overview of the different algorithm steps and some implementation details are provided. We discuss the importance of feature selection and the relationship with the size of the database and a priori knowledge about the objects of interest. Finally, we conduct a performance assessment of the algorithm. Due to the limited size of the database we apply the leave-one-out method to be able to use the same database for training of the fuzzy decision tree and for testing. We provide an interpretation of the results as well as a comparison to a similar approach that does not make use of fuzzy set theory.

6.1 The Database

The DBT data sets used in this study have been acquired with a GE Polaris DBT prototype at the Massachusetts General Hospital (MGH), Boston, MA, USA. For each breast, 15 projected views have been acquired over an angular range of ± 30 degrees. To guarantee an overall patient dose that is not superior to the dose delivered in standard digital mammography, the dose per projection image is reduced to approximately 10 % of the dose delivered to the patient during a standard digital mammography examination. All acquisitions have been performed in an medio-lateral oblique (MLO) view angle.

No clinical ground truth was available for the image data (neither radiologists reports nor histology results). We have therefore performed a screening of selected cases with several medical imaging experts. The screening of the clinical cases was first performed on slices reconstructed with SART. We have then identified the corresponding ROIs in the tomographic projection images.

The set of 17 breast masses have been identified from 9 different DBT data sets. The database consists of 10 spiculated breast masses (Figure 6.1) and 7 circumscribed breast masses (Figure 6.2). Examples of the breast masses contained in the database are given for a number of ROIs extracted from the central (0 degree) projected views. Because of the reduced dose in

the projected views, some of the masses are very difficult to distinguish in these images. For the same reason, the present database presents a considerable challenge for testing our CAD framework.

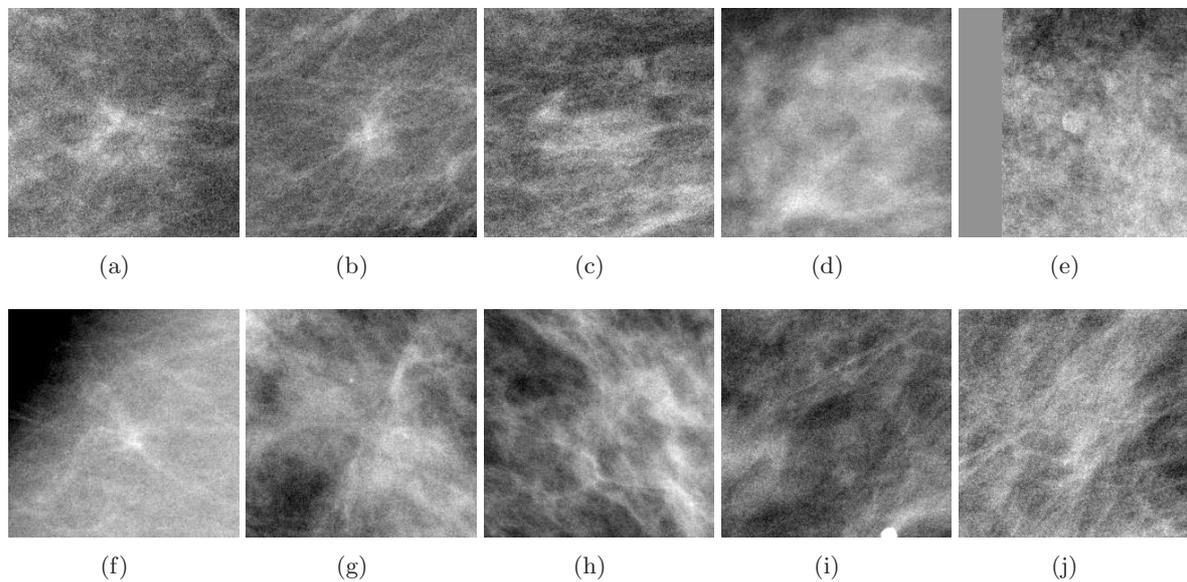


Figure 6.1: Database examples for the DBT CAD framework. A number of exemplary clinical cases containing spiculated breast masses are given. Due to the nature of the tomographic projection data some of the masses are very difficult to visually distinguish from the surrounding breast tissue.

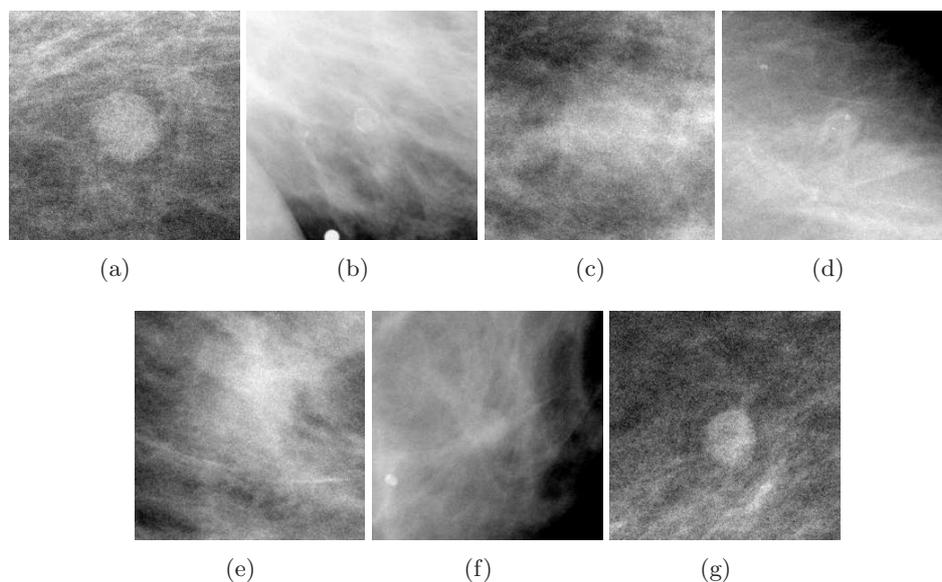


Figure 6.2: Database examples for the DBT CAD framework. A number of exemplary clinical cases containing circumscribed breast masses are given. Due to the nature of the tomographic projection data, some of the masses are very difficult to visually distinguish from the surrounding breast tissue.

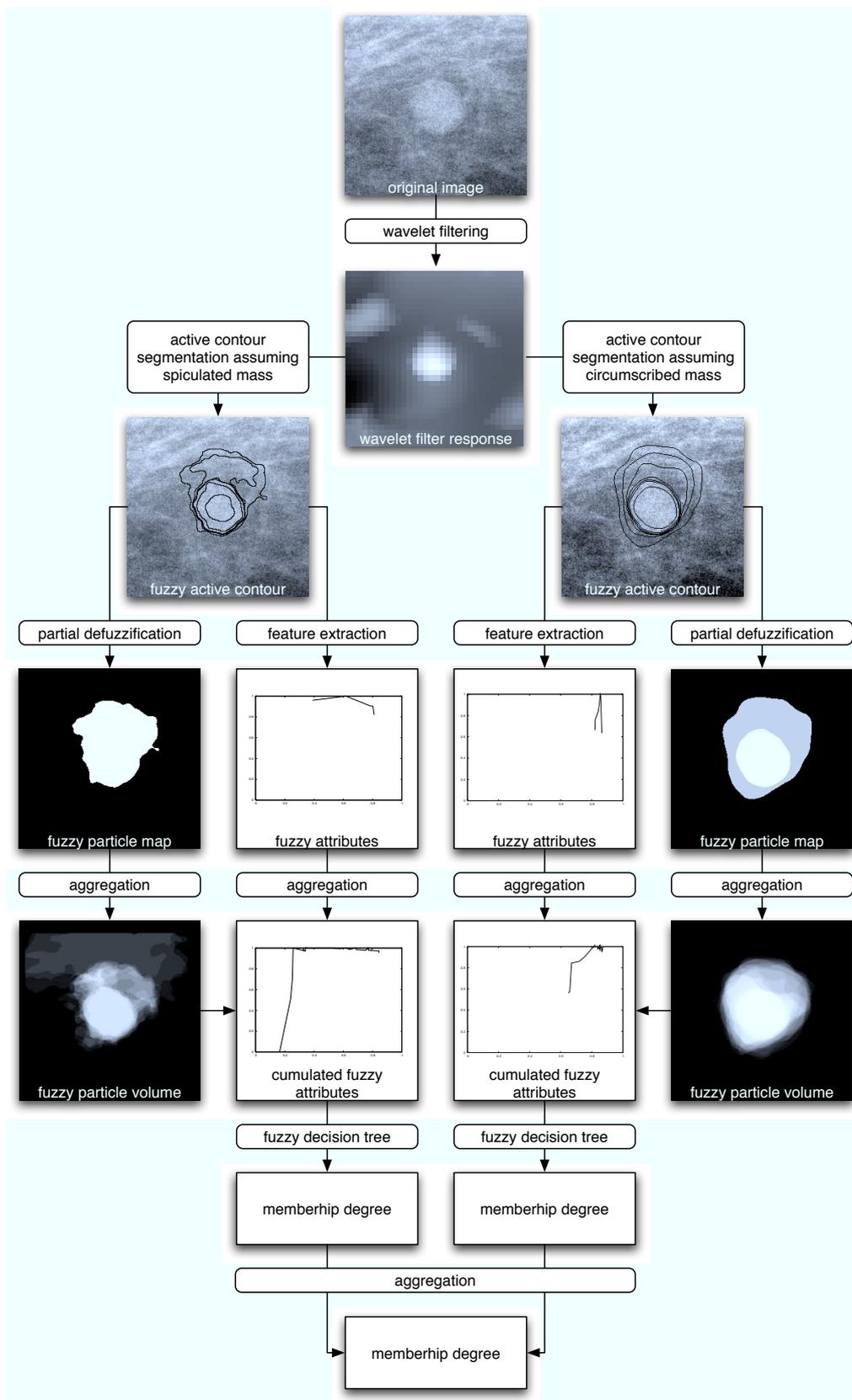


Figure 6.3: One example processed by the entire algorithm chain. Image results of the different processing modules for a clinical case containing a circumscribed mass. From top to bottom: original raw image, wavelet filter response (sub-sampled), fuzzy active contours, 2D fuzzy particle map, a slice of the fuzzy particle volume. A fuzzy attribute extracted for each model as well as the correspondent cumulated fuzzy attribute are depicted alongside the fuzzy particle map/volume.

6.2 The Algorithm

Figure 6.3 illustrates image results obtained for one specific ROI at different stages of the proposed CAD framework. For each breast mass, 15 projection images need to be processed. First, a multi-scale Mexican hat wavelet filter is applied to the raw projection images. Since masses are generally objects that exhibit a diameter superior to $5mm$, and in order to suppress noise and improve the computation time of the algorithm, the raw image is sub-sampled by a factor of 8. The wavelet filter response provides markers of structures that correspond potentially to radiological findings.

In the next step, two fuzzy active contour models are applied to the image in order to extract a set of candidate contours. This processing is performed on the full resolution projection images and initialized based on the markers provided by the wavelet filtering. The fuzzy active contour model that assumes a circumscribed mass has been implemented using the following parameter settings: $\lambda_1 = \lambda_2 = 1$, $\nu = 0$, $\mu = 1.5$, $h = 1$, $\Delta t = 0.1$, $\alpha = 1.0$, $\beta = 100$, $\gamma = 2000$, and no re-initialization. For this model the dominating terms correspond to the regularization (*compactness*) and the edge-based energy term (*sharp edges*). The model that assumes a spiculated mass, was applied using the following parameter setting: $\lambda_1 = \lambda_2 = 1$, $\nu = 0$, $\mu = 0.01$, $h = 1$, $\Delta t = 0.1$, $\alpha = 8.0$, $\beta = 100$, $\gamma = 500$, and no re-initialization. For this model the dominant energy term is the region-based term (*homogeneity*), while the edge-based term and the regularization force are assigned a smaller weight. The membership of each candidate contour to the class *contour* has been computed based on the energy of the candidate contour.

Then, we apply feature extraction to the set of candidate contours. 15 attributes are measured on the candidate contours. These are *area*, *perimeter*, *compactness*, *Ferret ratio*, *Ferret min*, *Ferret max*, *mean gradient along the contour*, and *homogeneity*. Furthermore we compute *mean gradient orientation* and *variance of the gradient orientation* for the contour, the region inside the contour, and an ROI containing the contour. These attributes have been chosen based on a priori knowledge about the objects of interest and have been widely used in the literature for characterization of breast masses (Cheng 06). Using the extension principle we obtain from these attributes a set of fuzzy attributes characterizing the segmented object.

In an aggregation step, we establish the link between fuzzy contours in different projected views that correspond to the same three-dimensional particle. To this end we apply a partial defuzzification to each fuzzy contour. This can be seen as a simplification of the fuzzy representation where each pixel contains a single value corresponding to its membership to the particle. The resulting fuzzy particle maps are aggregated to obtain a fuzzy particle volume. By re-projection of the fuzzy particle volume and comparison with the fuzzy particle maps we obtain a set of two-dimensional particles that correspond to the same three-dimensional particle. The fuzzy attributes from different projection images are then aggregated based on this information resulting in cumulated fuzzy attributes (one fuzzy membership function per attribute for each three-dimensional particle).

For each applied fuzzy active contour model a separate classifier is needed to validate the assumptions made about the object of interest. We have constructed a fuzzy decision tree for each model that results in two membership degrees for each particle. The first is the membership degree μ^A to the class A that expresses the validity of the zero-hypothesis, while the second is the membership degree $\mu^{\bar{A}}$ to the complement of the class \bar{A} , that expresses an invalid zero-hypothesis. For each fuzzy decision tree we may obtain a decision through the max of the two membership degrees. Furthermore a confidence degree is computed that expresses the confidence in the obtained result. Since we are dealing with several fuzzy decision trees, their results then need to be aggregated in order to obtain one decision about the particle. A complete flow-chart of the algorithm is depicted in Figure 6.4.

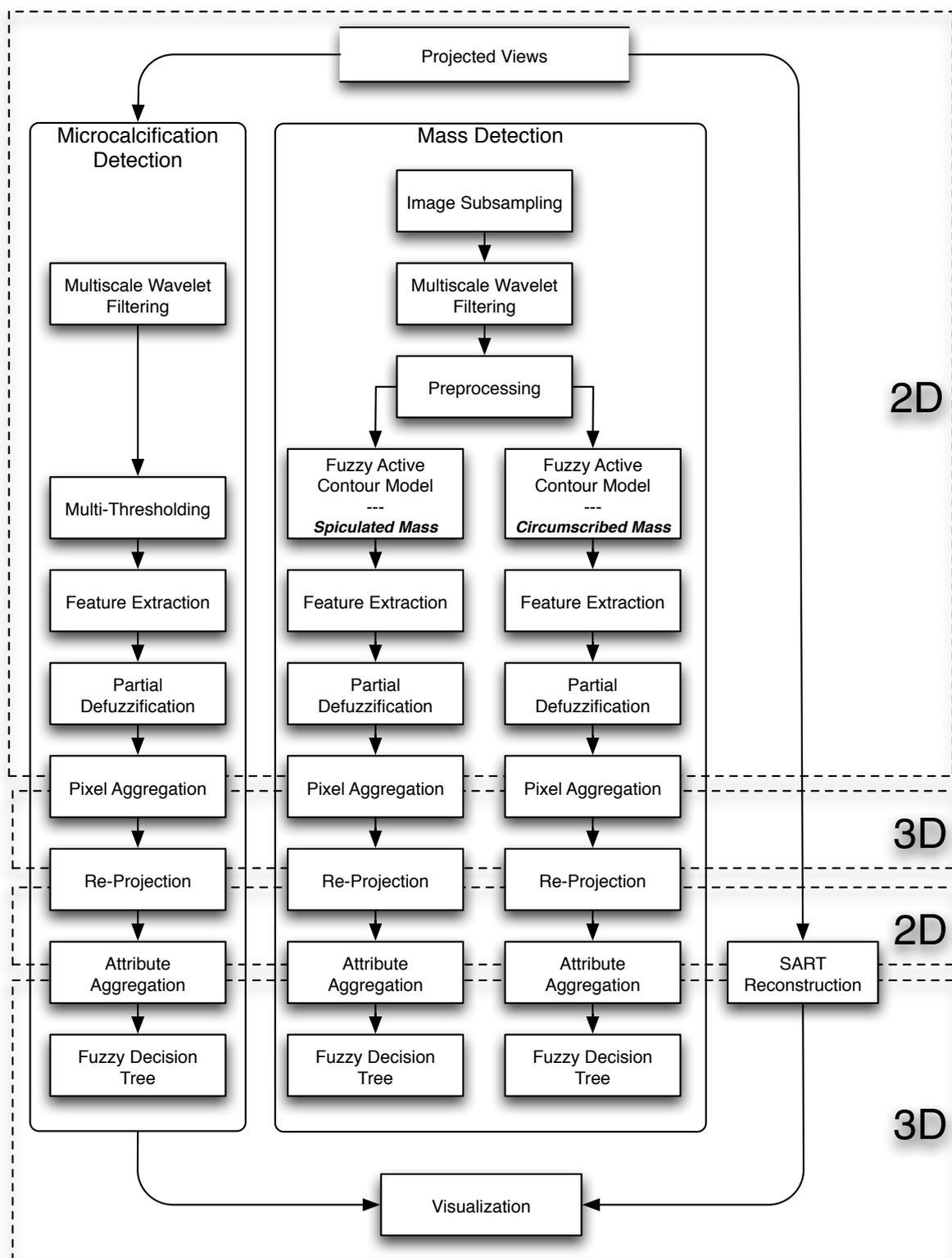


Figure 6.4: Flow-chart of the complete CAD processing framework for DBT data. Working on the projected views, three main branches are followed: microcalcification detection (left), mass detection (center), and reconstruction for visualization (right). The results of the three branches may be combined in a visualization module.

6.3 Algorithm Performance

The aim of this section is to obtain an evaluation of the proposed CAD approach for DBT data sets. A CAD algorithm is intended as a tool to help radiologists in their detection and diagnosis tasks. In order to evaluate such a tool we therefore need to create a situation representative for its intended use and measure the performance. Ideally this performance should then be compared to other CAD applications in order to be able to evaluate the obtained results.

In order to evaluate a CAD algorithm, we are looking to compute certain key performance values. These are usually expressed in terms of false positives (FP) and false negatives (FN). The FN are often substituted by the sensitivity (Se) of the algorithm. Since most algorithms can be tuned to produce different value pairs of Se and FP, these values need to be jointly considered. Alternatively, several such pairs of values for Se and FP are computed and a curve is drawn to connect the points. This curve is generally known as Receiver Operating Characteristics (ROC) curve (Metz 78; Metz 86). A popular value for comparison of CAD algorithms is the area A_z under the ROC curve.

The main problem regarding these evaluation techniques lies in the availability of a clinical database. In order to evaluate the performance in terms of the FP and FN of an algorithm a clinical database and the corresponding biopsy-proven ground-truth is needed. For statistically relevant performance measurements, this database should be sufficiently large and representative of the population targeted by the application. In our case, even though we dispose of a large number of clinical images, the clinical ground-truth is missing entirely. We are therefore not able to evaluate the performance of the proposed algorithm according to classical evaluation schemes at the moment.

However, a performance assessment of the proposed CAD framework is necessary in order to justify the chosen approach, and a visual evaluation does not seem sufficient for this task. We have therefore performed a review of selected DBT cases (only *diagnosis* cases and no *screening* cases) with a group of mammography experts in order to identify mass lesions. We will assess the performance of the algorithm regarding the capability to distinguish between the different kinds of breast masses (circumscribed breast masses and spiculated breast masses) that have been detected during the expert review.

This approach towards an evaluation of the proposed CAD framework has been chosen for two reasons. On the one hand, without corresponding ground-truth, it is impossible to exclude the presence of a breast lesion for a given region of interest. Hence, evaluation of FP and FN in the classical sense is not possible. On the other hand, the capability to distinguish between different types of breast lesions demonstrates several important properties of the framework. In order to distinguish between different kinds of lesions, we need to be able to obtain a valid characterization of each object based on the segmentation and the extracted features. Furthermore, we need a decision step that is able to derive valid conclusion from the obtained data. The proposed evaluation strategy is well suited to assess the performance of our approach regarding these key points.

Finally, we will compare our results with results obtained for another approach applied to the same clinical database. We are using fuzzy logic in many different applications of our framework to model the ambiguity in the images. In order to demonstrate the improvements obtained by applying these fuzzy techniques, we compare the proposed approach to an approach that does not make use of fuzzy logic.

6.3.1 Feature Selection

Feature selection is applied in order to identify a subset of the extracted feature vector that includes only the most discriminating attributes and eliminates at the same time attributes that are redundant. In our case there are additional reasons for applying a feature selection step. Firstly, we have extracted a feature vector that is of the same order of magnitude as our clinical database. If all feature would be used for constructing the fuzzy decision trees, we risk an over-fitting of the fuzzy decision trees. This results in a loss of generalization power which we want to avoid. Secondly, the limited size of the database leads to attribute histograms that are sometimes poorly representative of the targeted population. In fact, a number of the obtained attribute histograms and corresponding node tests result in a degradation of the classification. We therefore aim to reduce the impact of these corrupt attribute histograms by applying a feature selection step.

Attribute	Contrast (circumscribed model)	Contrast (spiculated model)
gradient	0.593	0.608
area	0.565	0.565
compacty	0.286	0.268
Ferret ratio	0.389	0.376
Ferret min	0.804	0.637
Ferret max	0.703	0.480
perimeter	0.727	0.519
gray variance (mask)	0.512	0.523
mean gradient (contour)	0.694	0.680
gradient direction variance (contour)	0.766	0.756
gradient direction variance (mask)	0.696	0.654
gradient direction variance (ROI)	0.506	0.472
gradient direction mean (contour)	0.765	0.726
gradient direction mean (mask)	0.712	0.659
gradient direction mean (ROI)	0.638	0.643

Table 6.1: Discriminating power of the cumulated fuzzy attributes.

To this end we have built a minimal fuzzy decision tree for each attribute. During the construction of these fuzzy decision trees contrast values for each attribute are computed that reflect the attributes discriminating power. The results for the two object models are given in Table 6.1.

A lower contrast corresponds to a higher discriminating power for the attribute. Sometimes, the most discriminating attribute at a given point in the fuzzy decision tree is denoted as *minimum contrast attribute*. We can see from Table 6.1 that the *compacty* and the *Ferret ratio* are the best attributes for both tree models in terms of the discriminating power. Examples of attributes with a small discriminating power are the variances of the gradient directions in the different regions (contour, mask, and ROI; see Section 3.4.1). An interesting case is the attribute *Ferret max* which appears to be much more discriminating in the case of the spiculated mass model than for the circumscribed model. This is easily explained by the different weighting of the regularization terms for the two models.

Usually, these attributes would be ranked according to their discriminating power in order to choose the best ones (eliminate the attributes with the highest contrast values) in order to obtain a good subset of the feature vector. However, for a limited database this process is not entirely

automatic. A priori knowledge about the objects of interest indicates that some attributes offering good discrimination capabilities are corrupted. For example most of the spiculated masses in our database are large while the majority of the circumscribed masses contained in the database are small. Consequently, the attribute *area* leads to a low classification error for distinguishing between these objects of interest. Nevertheless we know that the two types of breast masses cannot generally be distinguished by their size. We therefore chose, based on this knowledge, not to include the attribute *area* when constructing the full fuzzy decision tree.

6.3.2 Results Using the Leave-One-Out Method

Since we only have a very limited number of clinical cases at our disposal, we had to perform the training as well as the prediction with the same database. We have applied the leave-one-out method, where a sample is tested on a tree that was build with the entire training database except for the sample to be tested. This process is repeated for each of the samples in the database. The result of the leave-one-out method for trees using only a single attribute are given in Table 6.2.

Attribute	Classification Error (circumscribed model)	Classification Error (spiculated model)
gradient	70%	41%
area	35%	35%
compacity	23%	23%
Ferret ratio	29%	29%
Ferret min	41%	23%
Ferret max	41%	35%
perimeter	41%	23%
gray variance (mask)	47%	41%
mean gradient (contour)	52%	52%
gradient direction variance (contour)	58%	58%
gradient direction variance (mask)	47%	41%
gradient direction variance (ROI)	35%	35%
gradient direction mean (contour)	29%	35%
gradient direction mean (mask)	35%	47%
gradient direction mean (ROI)	58%	58%

Table 6.2: Classification results for a single node using the leave-one-out method.

The results from applying the leave-one-out method for classification using a single attribute are well suited to demonstrate the limitations of the clinical database. In the case of a clinical database that is representative of the targeted population, the obtained error rates should correspond to the contrast values for each attribute. For such a database, the attribute histograms and consequently the node tests provide a good generalization of the training database. Therefore, the error rates for the classification with the obtained trees should be proportional to the contrast measured for each attribute. However, with the database at hand, this is not true for all attributes. Even though the attributes with the lowest contrast values (*compacity* and *Ferret ratio*) result in relatively low error rates, there are other attributes that do not exhibit such a consistent behavior.

If we look for example at the *mean gradient direction along the contour* we find a high contrast for both models. The corresponding classification errors are however relatively low (compared to the other attributes). The logical explanation for this behavior is the lack of generality of

the used database. When we look at the error rates we need to keep in mind that a random classification of the data should statistically leads to an error rate of 50 %. Attributes that exhibit a classification error superior to 50 % may still provide an interesting discrimination if they are used somewhere else in the tree than at the root node. However, since we are working with few training samples and consequently will limit the tree construction to few nodes (to avoid over-fitting) these attributes are more likely to degrade the classification performance.

6.3.3 Obtaining a Final Decision

Now that we have computed the discriminating power for each attribute of the feature vector, we can construct a full tree for each model (circumscribed and spiculated). The fuzzy decision tree for the spiculated mass model comprises two attributes (*compactness* and *Ferret ratio*), while the tree for the circumscribed mass model was build from three attributes (*compactness*, *Ferret ratio*, and *homogeneity*). Since we are applying the leave-one-out method we need to build a new tree for each test sample (trained with the remaining samples of the database). We therefore do not dispose of a single tree that can be interpreted. Instead, we have chosen a representative example for a tree of each model and we will give some key values for the set of constructed fuzzy decision trees (for the circumscribed model 5.8 nodes per tree on average). Two exemplary fuzzy decision trees build for classifying a circumscribed mass (case ID = 1) are illustrated in Figure 6.5.

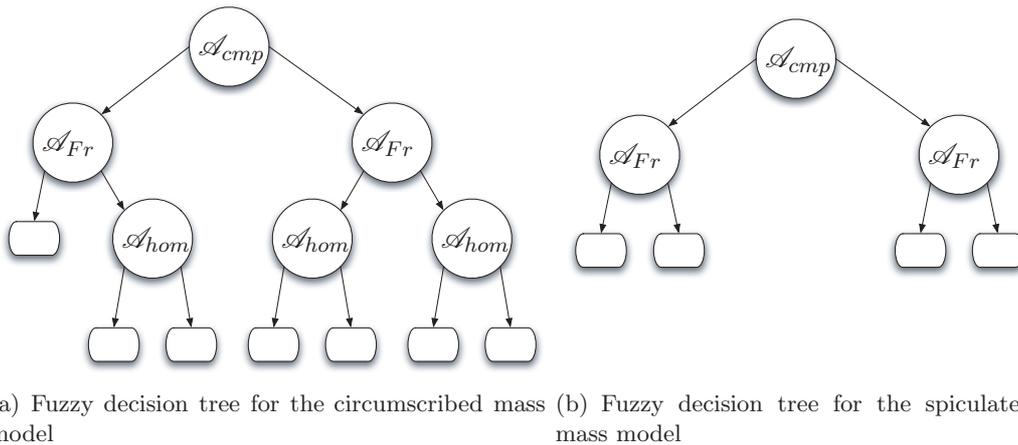


Figure 6.5: Examples of fuzzy decision trees constructed using the leave-one-out method for a circumscribed mass (ID=1) of the database. The attributes indicated at each node correspond to *compactness* (\mathcal{A}_{cmp}), *Ferret ratio* (\mathcal{A}_{Fr}), and *homogeneity* (\mathcal{A}_{hom}).

Classification of the database with the fuzzy decision trees using the leave-one-out method results in two trees (circumscribed model and spiculated model) for each example of the database and thus four associated membership degrees (μ^{crc} , $\mu^{\overline{crc}}$, μ^{spc} , and $\mu^{\overline{spc}}$). In order to compute the classification error we will need to combine these degrees to obtain a decision regarding the example. We then need to defuzzify the classification results of the fuzzy decision tree. For each tree we consider the max of the two membership values (e.g. for the circumscribed model: $\max(\mu^{crc}, \mu^{\overline{crc}})$) to correspond to the resulting class label of the tested example. The membership degrees to each class for the two models as well as the assigned class label resulting from the defuzzification are given in Table 6.3.

From Table 6.3 we obtain a classification error of 17.6 % (3 of 17 cases falsely classified) for the circumscribed mass model and a classification error of 17.6 % (3 of 17 cases falsely classified) for the spiculated mass model. Since we are dealing with multiple fuzzy decision trees, these values alone do not suffice to assess the classification performance. In fact we need to investigate

Case ID	L_c	μ^{crc}	$\mu^{\overline{crc}}$	L_p^{crc}	μ^{spc}	$\mu^{\overline{spc}}$	L_p^{spc}	Comment
1	crc	0.61	0.35	crc	0.45	0.68	crc	ok
2	spc	0.39	0.41	spc	0.50	0.36	spc	ok
3	spc	0.38	0.55	spc	0.51	0.38	spc	ok
4	spc	0.35	0.43	spc	0.55	0.42	spc	ok
5	spc	0.66	0.35	crc	0.35	0.65	crc	error
6	crc	0.64	0.32	crc	0.32	0.62	crc	ok
7	crc	0.59	0.24	crc	0.29	0.55	crc	ok
8	spc	0.46	0.40	crc	0.44	0.43	spc	conflict
9	spc	0.41	0.57	spc	0.54	0.41	spc	ok
10	spc	0.35	0.52	spc	0.54	0.40	spc	ok
11	crc	0.45	0.43	crc	0.60	0.39	spc	conflict
12	crc	0.42	0.55	spc	0.58	0.39	spc	error
13	crc	0.60	0.25	crc	0.29	0.59	crc	ok
14	crc	0.66	0.27	crc	0.27	0.70	crc	ok
15	crc	0.57	0.34	crc	0.49	0.51	crc	ok
16	crc	0.58	0.35	crc	0.44	0.64	crc	ok
17	crc	0.55	0.21	crc	0.26	0.62	crc	ok

Table 6.3: Classification results for a full fuzzy decision tree using the leave-one-out method.

the combination of the classification results. We obtain 11.8 % (2 of 17) conflicting cases and 11.8 % (2 of 17) of cases that were falsely classified by both fuzzy decision trees. The combined classification performance will therefore lie between 11.8 % and 23.6 % depending on the method used to combine the results from the two fuzzy decision trees.

In order to combine the results from the two fuzzy decision trees, we compute a confidence degree for each class label based on the dissimilarity of the two membership values (see Section 5.6). If the class labels from the two trees do not conflict with one another ($crc \wedge \overline{spc}$ or $\overline{crc} \wedge spc$) we have obtained the class label L^c for the tested example. In the case where the obtained class labels are conflicting (e.g. the circumscribed model concludes a circumscribed mass and the spiculated models concludes a spiculated mass) we need to compare the obtained labels from each tree. Therefore, we weight each maximum membership degree with its corresponding confidence degree D_c (see Section 5.6). The maximum of the two results in the class label for the tested example. This relation is expressed as:

$$L^c = \begin{cases} A & \text{if } \left(\max(\mu^A, \mu^{\overline{A}}) \cdot D_c^A \right) > \left(\max(\mu^B, \mu^{\overline{B}}) \cdot D_c^B \right) \\ B & \text{if } \left(\max(\mu^B, \mu^{\overline{B}}) \cdot D_c^B \right) > \left(\max(\mu^A, \mu^{\overline{A}}) \cdot D_c^A \right) \end{cases} \quad (6.1)$$

The confidence degrees and the resulting class labels for the entire clinical database computed using the leave-one-out method are given in Table 6.4. The total classification error using both types of fuzzy decision trees (circumscribed and spiculated mass model) is at 17.6% (3 of 17 sample have been falsely classified). This is however not a very representative measure. As we can see from Table 6.4, most of the cumulated confidence degrees are quite low. The decisions for one of the classes was therefore not a very confident one. This is mostly due to the database that is not representative of the targeted population.

With such a small database at our disposal, the classification results shown here, can only serve as to prove the validity of our approach. Once a larger database will be available, we will need to perform a statistically relevant feature selection and performance evaluation.

Case ID	L_c	μ^{crc}	$\mu^{\overline{crc}}$	μ^{spc}	$\mu^{\overline{spc}}$	D_c^{crc}	D_c^{spc}	\mathcal{D}_c	L_p
1	crc	0.61	0.35	0.45	0.68	0.27	0.20	0.17	crc
2	spc	0.39	0.41	0.50	0.36	0.03	0.16	0.08	spc
3	spc	0.38	0.55	0.51	0.38	0.18	0.14	0.07	spc
4	spc	0.35	0.43	0.55	0.42	0.11	0.13	0.07	spc
5	spc	0.66	0.35	0.35	0.65	0.31	0.30	0.20	crc
6	crc	0.64	0.32	0.32	0.62	0.33	0.32	0.21	crc
7	crc	0.59	0.24	0.29	0.55	0.42	0.31	0.25	crc
8	spc	0.46	0.40	0.44	0.43	0.07	0.01	0.03	crc
9	spc	0.41	0.57	0.54	0.41	0.16	0.13	0.07	spc
10	spc	0.35	0.52	0.54	0.40	0.19	0.15	0.08	spc
11	crc	0.45	0.43	0.60	0.39	0.02	0.20	0.12	crc
12	crc	0.42	0.55	0.58	0.39	0.14	0.20	0.11	spc
13	crc	0.60	0.25	0.29	0.59	0.42	0.35	0.25	crc
14	crc	0.66	0.27	0.27	0.70	0.42	0.45	0.28	crc
15	crc	0.57	0.34	0.49	0.51	0.25	0.02	0.15	crc
16	crc	0.58	0.35	0.44	0.64	0.25	0.19	0.14	crc
17	crc	0.55	0.21	0.26	0.62	0.44	0.40	0.24	crc

Table 6.4: Classification results for a full fuzzy decision tree using the leave-one-out method. Errors are indicated in bold.

6.3.4 The Contribution of Fuzzy Logic

In order to get a better evaluation of the proposed CAD approach we want to compare the obtained classification results to a different CAD algorithm. To this end we have chosen an approach that does not make use of fuzzy set theory. Comparing the classification results of the two approaches will hence serve two purposes. Firstly, we provide a comparative performance assessment for the two algorithms working on the same clinical database. Secondly, comparing our approach to a purely crisp CAD approach will enable us to quantify the contribution of using fuzzy set theory to model that ambiguity present in the images.

The reference algorithm differs mainly in the use of fuzzy set theory from the approach presented above. For each active contour model, only the 0-level-set is extracted. Feature extraction is then performed for this contour. The features are aggregated over the entire set of projected views using the arithmetic mean of the attribute values. Feature selection is performed in a way equivalent to the method described in Section 6.3.1 (only that in this case, the entropy is used for measuring the discriminating power of the attributes instead of the contrast). A crisp binary decision tree classifier is then used to obtain a decision for each of the two zero-hypotheses that was made by the active contour models. A simplified algorithm scheme for the crisp reference algorithm is depicted in Figure 6.6.

We make use of the leave-one-out method for performance evaluation of the crisp decision tree. Examples of crisp decision trees obtained for testing a case containing a circumscribed mass (case ID = 1) is illustrated in Figure 6.7.

The results of the crisp decision tree classification using the leave-one-out method are given in Table 6.5. We obtain a classification error of 35.2 % (6 of 17 cases falsely classified) for the circumscribed mass model and a classification error of 17.6 % (3 of 17 cases falsely classified) for the spiculated mass model. As in the fuzzy case, we are dealing with multiple decision trees and thus need to extract additional information in order to properly assess the performance of the

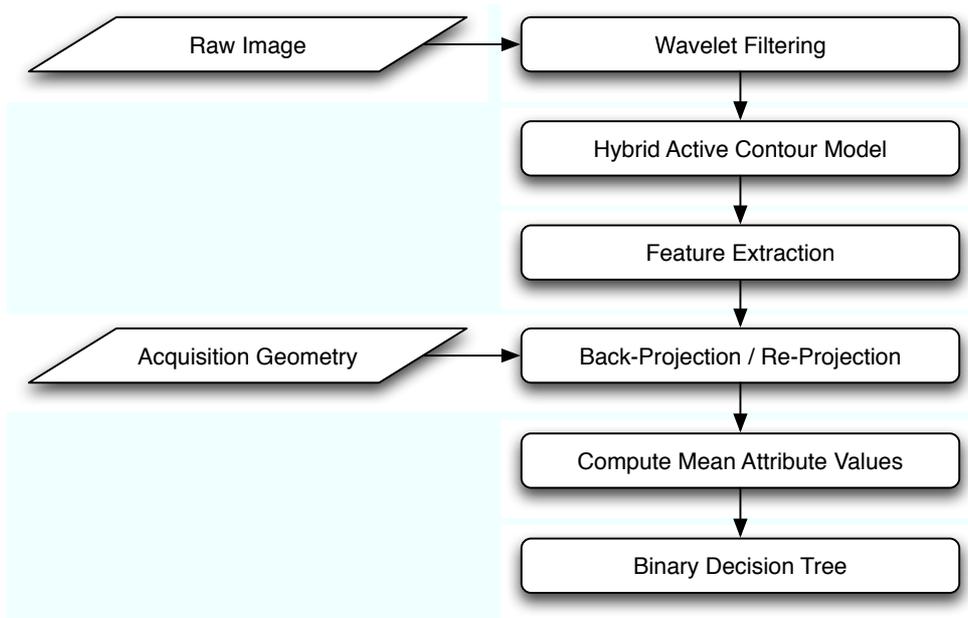
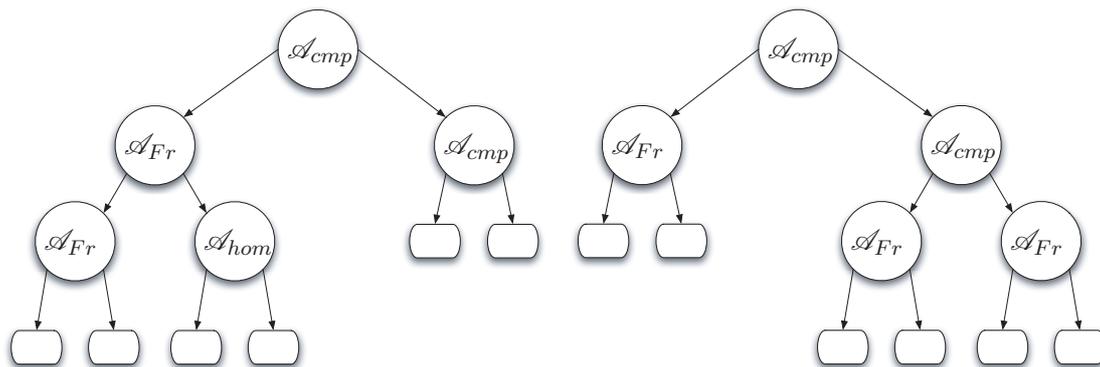


Figure 6.6: Flow-chart of the crisp CAD algorithm.



(a) Binary decision tree for the circumscribed mass model
 (b) Binary decision tree for the spiculated mass model

Figure 6.7: Examples of binary decision trees constructed using the leave-one-out method for a circumscribed mass (ID=1) of the database. The attributes indicated at each node correspond to *compactness* (\mathcal{A}_{cmp}), *Ferret ratio* (\mathcal{A}_{Fr}), and *homogeneity* (\mathcal{A}_{hom}).

algorithm. From Table 6.5 we get 41.2 % conflicting cases (7 of 17) and 5.9 % of cases (1 of 17) that were falsely classified by both decision trees. The combined classification performance will therefore lie between 5.9 % and 47.1 % depending on the method used to combine the results from the two crisp decision trees.

Finding a method for combining the results of the crisp decision trees is however limited to fewer choice than in the case of fuzzy decision trees since the available information is sparse. The only means to decide in cases of conflicting results is to base the decision on a priori knowledge (which may not always be available). For example, if we know the probability for the occurrence of an example of each class we may assign a higher priority to the class with the higher probability. Another strategy may be based on the targeted application. If we consider spiculated breast masses to be more likely to correspond to a cancer, then we may choose to proceed in a cautious manner (aiming at a high sensitivity) and assign all conflicting cases to spiculated masses. This will however result in an elevated number of false alarms. Tuning

Case ID	L_c	L_p^{crc}	L_p^{spc}	Comment
1	crc	crc	crc	ok
2	spc	spc	crc	conflict
3	spc	spc	crc	conflict
4	spc	crc	spc	conflict
5	spc	crc	spc	conflict
6	crc	crc	crc	ok
7	crc	crc	crc	ok
8	spc	crc	crc	error
9	spc	crc	spc	conflict
10	spc	crc	spc	conflict
11	crc	crc	crc	ok
12	crc	crc	crc	ok
13	crc	crc	crc	ok
14	crc	crc	crc	ok
15	crc	spc	crc	conflict
16	crc	crc	crc	ok
17	crc	crc	crc	ok

Table 6.5: Classification results for a full binary decision tree using the leave-one-out method.

the operating point of the crisp decision trees according to a clinical database appears to be a difficult task.

6.3.5 Conclusion

Today, the main issue for further development and testing of the fuzzy decision trees lies in the limited database that is available. The results presented in this chapter are intended as a proof of concept rather than a statistically relevant evaluation. When a larger database becomes available, the classification performance and the generalization power of the fuzzy decision tree will need to be validated in order to further improve the method.

This said, the obtained results look very promising. The DBT projection data acquired at reduced dose per scan represent a great challenge for CAD. Furthermore we have chosen to rather demonstrate a meaningful application of our CAD framework than to optimize the classification results when the available database is very limited. Better results (lower error rates) can be obtained for more complex fuzzy decision trees, but the size of the database limits the number of attributes we can use without risking over-fitting. The comparison with a crisp CAD approach illustrated well the impact of using fuzzy set theory in our framework. The classification results obtained for the proposed CAD framework are superior to the results obtained for the crisp reference algorithm. This comparison is however biased by the limited database. The more important conclusion lies therefore in comparing the methodologies. Especially in the classification step where a final decision is computed the gradual membership degrees provide a significant advantage.

Prediction is very difficult, especially if it's about the future.

Nils Bohr

7

Conclusion

7.1 Main Contributions

In this work, we have developed a new framework for computer-aided detection on digital breast tomosynthesis data sets. X-ray mammography is currently the primary method for detecting early breast cancers that constitute one of the leading causes of cancer mortality among women. DBT is a new three-dimensional breast imaging modality where a series of X-ray projection images are acquired over a limited angular range. A breast volume is reconstructed from the projected views to overcome the problems for detecting radiological findings which are caused by overlying breast tissue. A primary goal of CAD is to increase the efficiency and effectiveness of breast cancer screening. Especially in DBT data sets where the amount of data that a radiologist needs to inspect is considerably increased when compared to standard 2D mammography.

The main contributions of our work are threefold. Firstly, DBT is an emerging modality with few clinical data sets available, and research concerning CAD for these data is still in the early stages of development. Secondly, we propose a high-level reconstruction independent framework for detection and characterization in medical images, that may be applied to different radiological findings or different imaging modalities altogether. Working directly on the DBT projection images provides advantages in terms of flexibility and processing time for a modality where reconstruction algorithms are not yet fully optimized. Thirdly, we make use of the fuzzy set theory to model the ambiguity present in the image data. This allows us to gather information over several successive processing steps and to delay the decision until all extracted information can be jointly considered.

We will quickly recall the principal steps of the proposed CAD framework: First, a multi-scale Mexican hat wavelet filter is applied to the raw projection images providing markers of structures that correspond potentially to radiological findings. Then, fuzzy contours are computed for each marked particle. For microcalcification detection this is realized through a multi-thresholding operation. In the case of mass detection, multiple fuzzy active contours are extracted for each particle, each associated to a different assumption about the particle. Next, we compute a fuzzy feature vector for each fuzzy contour. In the following aggregation step, the extracted information is aggregated over the entire set of projected views. To this end a partial defuzzification is applied to establish the link between corresponding particles in different projected views. The resulting vector of cumulated fuzzy attributes is used as input to a set of

fuzzy decision trees. Each tree provides the validity for one of the assumptions made during the fuzzy contour extraction. Finally, the classification results from the different trees are combined to obtain a final decision regarding each particle.

We have presented a new hybrid active contour model for contour detection that we applied to the segmentation of masses in digital breast tomosynthesis projection images (Peters 07). The combination of region-based and edge-based energy terms provides an improved segmentation performance when compared with active contour models based on a single image feature. The hybrid active contour model has been extended for extracting fuzzy contours. An original approach for obtaining a set of candidate contours from a single evolving function ϕ has been developed. Our fuzzy segmentation technique effectively overcomes the problem that the active contour model converges towards a local minimum before (or sometimes after) reaching the actual object boundary. Since the properties of the object of interest are included in the energy functional of the active contour model and the parameters are tuned accordingly, we can derive the membership degrees to the different candidate contours directly from the energy levels of the candidates. A novel framework for extracting multiple fuzzy contours for an unknown object of interest has been introduced. Different fuzzy active contour models are applied to a given object of interest. This allows to consider a number of possible assumptions concerning a given particle and has proven a powerful tool for characterization of different types of breast masses. Each model has been build based on a priori knowledge about a different kind of object. Feature extraction, aggregation and classification are performed for each model separately, resulting in two membership degrees for each model. A final decision regarding a given particle is then based on the individual membership degrees and the corresponding confidence degrees obtained from different fuzzy decision trees.

We have introduced two novel methods for the aggregation of fuzzy information from a set of tomographic projection images. First, a pixel-based approach has been developed. We have shown how *partial defuzzification* can be used to obtain a pixel-based fuzzy representation of the objects from the fuzzy contours. This provides a simplification and a good compromise between a complex fuzzy representation and a crisp representation lacking the depth of the fuzzy information. A dedicated aggregation operator has been applied to obtain a *fuzzy particle volume* from the set of *fuzzy particle maps*. Today, the values of the fuzzy particle maps are averaged to obtain a pixel in the fuzzy particle volume (Peters 05; Peters 06b). A great flexibility regarding the aggregation operators has been demonstrated. A new particle-based approach has been developed as well. Here, the pixel-based approach is applied only to establish the link between a given three-dimensional particle and the corresponding two-dimensional particle in each projected view. This information is the used to compute *cumulated fuzzy attributes* for each three-dimensional particle. The full information extracted in the projected views is therefore aggregated into a three-dimensional characterization of the particle. Then the cumulated attributes serve as input to a classification step. We are thus able to delay the decision and to preserve the full range of extracted information until a final decision is obtained from a dedicated classification step.

We have discussed a fuzzy decision tree classifier working on fuzzy data proposed in a former work (Bothorel 96). Several original improvements and extensions have been proposed. The original fuzzy decision tree has been applied to minimal trees (one node and two leaves) only. The extension to more complex trees has been studied. To this end we have investigated the multiple use of attributes in a path through the tree. We have introduced a new method to process attributes that are not monotone. We took special attention to considerations concerning the database. The behavior of the fuzzy decision tree has been illustrated for different synthetic examples. The processing of clinical patient data has lead us to consider further improvements of the algorithm. For small databases, we have computed weighting factors for the class densities in each leaf. A feature selection step based on the fuzzy decision tree has been proposed that

allows to identify an optimized subset of the feature vector. Finally, we have introduced the cumulated confidence degree as a measure for the confidence in the classification result of a single object that has been processed taking into account several different a priori assumptions about the object.

We have developed a novel framework for reconstruction-independent CAD on DBT projected views. Our choice to follow a reconstruction-independent approach was based on three main considerations, where the novelty of the modality had a important impact on our decision. The first consideration was regarding reconstruction algorithms. For a CAD framework working on the reconstructed breast volume, the algorithm used to reconstruct this volume will define the properties of the image data and thus impact the applied CAD processing. Since reconstruction algorithms for DBT are not yet fully optimized, we have preferred to base our framework on the projected views providing a stable and controlled base for our approach. Secondly, working on the projection images allows to make use of operators as well as a priori knowledge from standard mammography, since these images are closely related. Thirdly, processing of the projected views promised an improvement in terms of the computational effort, since the slices of the reconstructed volume are generally much more numerous than the projected views.

The proposed approach exhibits several advantages especially for a CAD approach at this early stage of DBT development. However, with more systems installed every year, knowledge about properties of the reconstructed breast volume and radiologist understanding of the three-dimensional nature of the radiological findings will grow. And as reconstruction algorithms continue to improve, algorithms designed to work directly on the reconstructed breast volume become increasingly interesting. Furthermore, the considerations regarding the processing time need to be further investigated. When comparing the number of projected views to the number of reconstructed slices, we assume that the applied processing is equivalent. However, the lower CNR in the projected views leads to an elevated number of false alarms which in return increases the processing time per projection.

We have made use of fuzzy set theory to model the ambiguity present in the images (Peters 06a). This has proven to be a powerful tool, especially when applied to the very noisy low-dose images of the projected views that constitute a significant challenge to segmentation algorithms. The ability to delay the decision about a given particle until the information from all projected views can be taken into account is a key property of the proposed framework. We consider this ability to be of central importance for any processing performed on the projection images. If the decision about a particle is made before the aggregation, only a fraction of the information for this particle is available and can be taken into account. At the same time, use of fuzzy set theory often increases the dimensionality of the processing and consequently leads to an increased complexity of the algorithm.

We acknowledge that an investigation on a larger clinical database is needed. Especially the lack of biopsy-proven clinical ground-truth constitutes a limitation for the performance assessment of the proposed CAD framework. We will then be able to compare our results to other recent CAD approaches for DBT. These shortcomings should be overcome in the future, when clinical data becomes more readily available and will hopefully confirm our preliminary results.

7.2 Suggestions for Further Work

Clinical Database Today, the main issue for evaluating a computer-aided detection framework for DBT lies in the availability of clinical data sets. Digital Breast Tomosynthesis is an emerging modality with very few systems installed worldwide. Even-though more and more systems are being installed, the modality is still at the stage of clinical validation with no com-

mercial product available today. With very few cases available and no corresponding clinical, biopsy-based truth a statistically relevant quantitative performance assessment of a CAD algorithm for DBT data sets is not available today. Even though we aimed at compensating this shortcoming by use of simulated images, expert reviews and qualitative illustrations of the algorithm performance, the lack of a clinical database still represents the major weakness of the presented results. When a biopsy-truthed clinical database becomes available, a finer tuning of the entire system will become possible and the performance of the developed framework will need to be reevaluated.

Mass Detection Detection of breast masses in radiographic projection images has been intensively studied and works by many different groups have been made public on the subject (Cheng 06). The aim of our work was to study in particular the three-dimensional nature of the data and the related issues. We have therefore passed little time trying to find or optimize a well adapted mass detection algorithm for our data. This is especially true for spiculated masses and architectural distortions, that represent an important part of the malignant lesion population. In order to obtain a fully automated detection and diagnosis tool, some work will need to be performed in order to identify a suitable detection algorithm for the data at hand. With reconstruction algorithms constantly improving, the detection may even be performed in 3D space.

Microcalcification Clusters In our work we have concentrated on processing of breast lesions rather than microcalcifications. Firstly, microcalcification detection is a task that is well established and provides good performances in clinical applications with up to 98.3% sensitivity with a false positive rate of 0.3 clusters per image (Cheng 03). This is not true for mass detection algorithms that still struggle to obtain equivalent performances (R2 Technologies 06). Secondly, the introduction of DBT was hailed by the medical community as a means to resolve the superimposition problem for screening of mass lesions. Many physicians are quite satisfied with the improved microcalcification visibility in recent digital mammography systems (Fischmann 05). For mass lesions that are not visible in standard mammograms today and are consequently missed during the screening process DBT is considered to resolve the problems of superimposition and masking effects..

However, to complete our framework, an additional processing step for microcalcification detection is needed. For the detection of microcalcifications, it is of great interest to extract more information than the one based on individual particles only. Important information may be associated to a group or specific configuration of particles and their relations to each other. In order to identify a group of microcalcifications associated to a breast pathology, an additional description level should be introduced. If we consider an entire cluster of microcalcifications as a single object, we can process this object in a manner equivalent to individual particles. First, we construct a graph where each particle is associated to a node and the edges of the graph correspond to the distances between the particles (Figure 7.1). Here, the distances need not necessarily be Euclidean distances between the spatial coordinates of the particles. In fact, such a distance may include measures of dissimilarity between the objects or measures of homogeneity of the graph. A *fuzzy cluster* can be extracted by use of multi-thresholding applied to the arcs of the graph. Fuzzy attributes can be extracted that characterize the cluster object. Figure 7.2 and Figure 7.3 illustrate the importance of form and orientation of a microcalcification cluster.

Reconstruction vs. Aggregation The framework developed during our work is independent of the reconstruction technique applied for visualization of the breast volume. This choice has been made for three reasons. Firstly, DBT is a new modality and reconstruction algorithms for

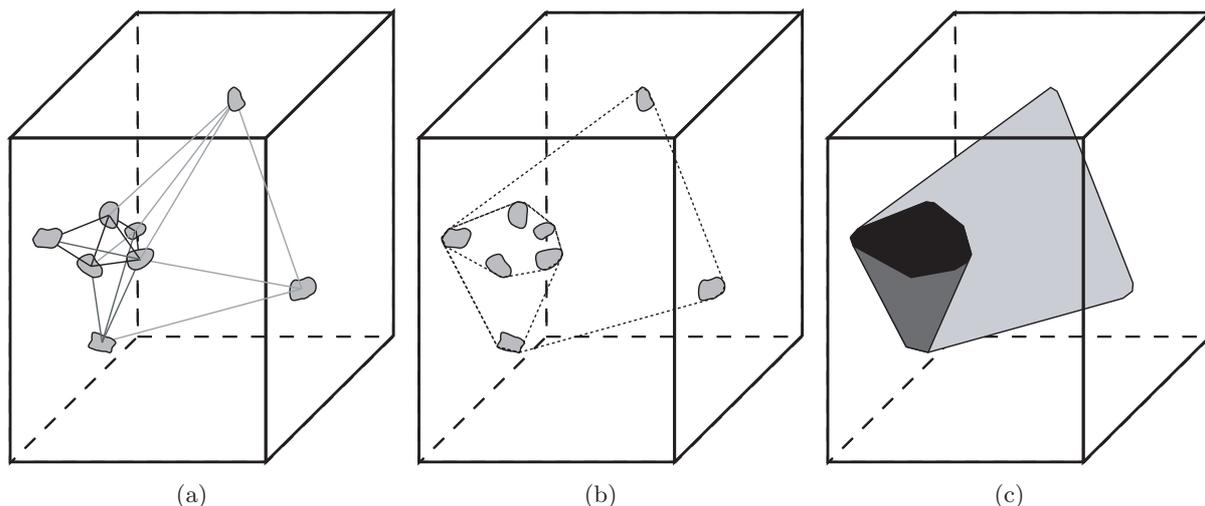


Figure 7.1: Fuzzy microcalcification cluster detection. In order to detect a fuzzy cluster we first construct a graph of all particles. Each particle is associated a node. Arcs are constructed to connect the nodes. When the distances between the objects are computed, we can attribute a weight to each arc of this graph. This weight is equivalent to a given distance between the two objects connected by the arc (a). A multi-thresholding operation results in a set of candidate clusters (b). For each of these clusters we extract the delimiting surface and the corresponding cluster features (c). We thereby obtain a fuzzy description of the microcalcification cluster.

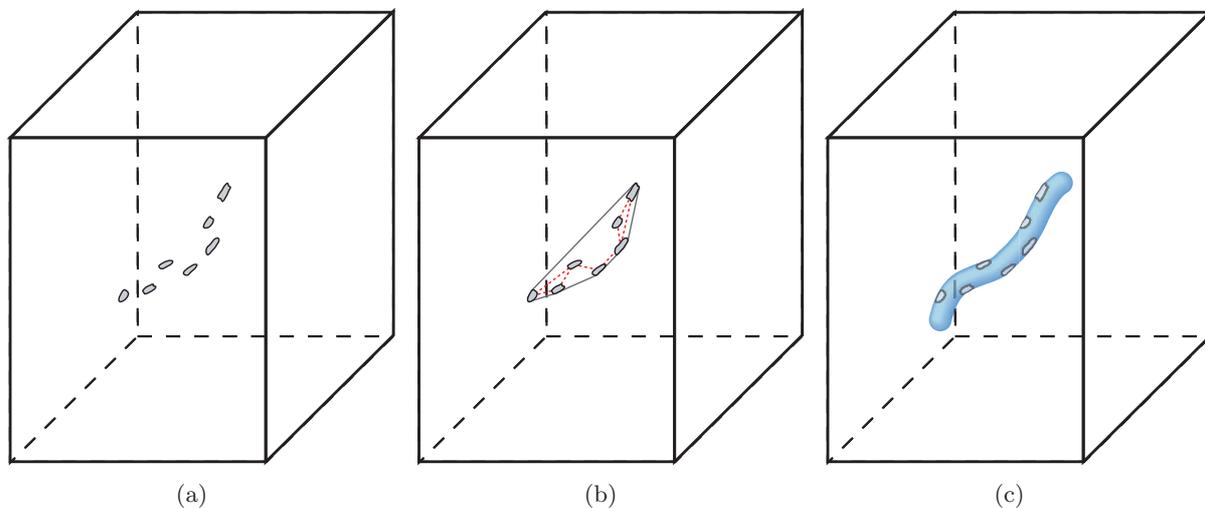


Figure 7.2: The importance of shape features for microcalcification clusters. For a set of detected particles (a) we compute the corresponding cluster (b). Feature extraction on this cluster object may provide additional information. For example, an elevated value of the elongation of the object may indicate that the cluster corresponds to vascular calcifications (c).

this task are not yet fully optimized. Due to the limited angular range of the tomosynthesis acquisition, reconstruction of the breast volume presents an ill-posed problem. This results in reconstruction artifacts that are most prominent in the space dimension perpendicular to the X-Ray detector. Secondly, the number of reconstructed slices is generally much more elevated than the number of projection images. A significant performance improvement may thus be obtained by working on the projection images. Thirdly, working on the projection images allows us to consider each projection of a given three-dimensional object individually. Once a 3D volume has

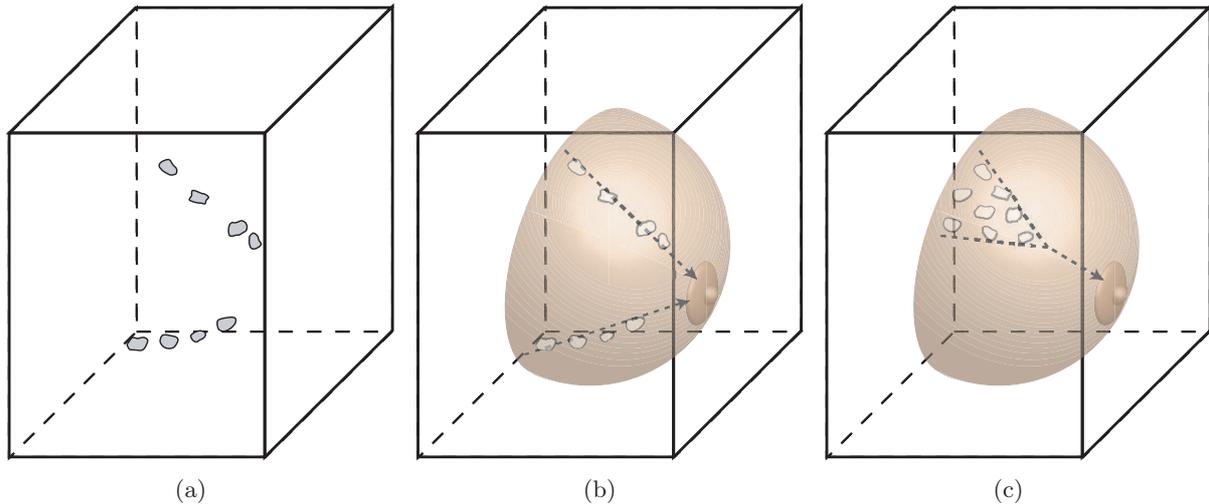


Figure 7.3: The orientation of the cluster object. Another clinically relevant cluster feature is the orientation of the cluster object. If the cluster exhibits a distinct elongation, the orientation of the principal axis of the cluster is measured. We can obtain additional clinically relevant information from the fact, whether or not the cluster is oriented towards the nipple.

been reconstructed, the information about the individual contributions to the value of a given voxel has been lost.

One area that appears interesting for future research is the compensation of reconstruction artifacts. Better adapted reconstruction algorithms may already reduce the artifacts present in reconstructed breast volumes today. Additionally a de-convolution of the data (either projected or in the volume) may serve to obtain the original, undistorted breast volume. These are however very complex tasks, and intensive research is being performed in this area.

Another possible improvement comes from the comparison between working on the projected views and working on the reconstructed volume. With better reconstruction algorithms, working directly on the reconstructed breast volume becomes increasingly appealing. A possible drawback is still the high computational effort. Today, we only compare the number of projection images to the number of reconstructed slices. This is however a very simplistic view. Since the CNR in the projected views is much lower than in the reconstructed slices, the same sensitivity in both will lead to a higher number of false alarms in the projection images. But an elevated number of false alarms leads to a higher computational effort (even if we consider that we will be able to eliminate these false alarms in succeeding processing steps). It might be interesting to try and derive a relationship between the acquisition parameters on one hand and the computational effort for processing in 2D as well as 3D on the other hand. This relation would need to take into account the number of projections as well as the dose per scan and the noise model. The result of this model could serve to take an educated decision as to what representation of the tomographic data (projected views or reconstructed volume) could be processed with the lowest computational effort.

Visualization The last step missing in our processing chain in order to provide a clinical tool to aid radiologists in their diagnostic and screening tasks lies in the presentation of the obtained results. The visualization of the data and the workflow of the image and volume presentation are vast fields of research by themselves. In our specific application, the fact that a gradual membership degree to a given class or reference is computed, rather than a binary decision, may come in handy at this stage. We could imagine a system that enhances the contrast of a given

structure depending on the membership degree of the corresponding fuzzy particle. Colors could indicate different levels of membership. A radiologist may even visualize corresponding attribute values for a suspect particle to help in the decision process. This seems to be an interesting research topic, and should be developed in close relation with the medical community.

*As a math atheist, I should be excused
from this.*

Calvin



Criteria to Compare Volumes and Surfaces

Comparison of objects is always attached to subjective criteria. However, in order to quantify the quality of results, there exist several measures which are typically used to compare an object A (automatically segmented) and an object M (manually segmented). Some of these criteria and some examples are described in this Appendix.

A.1 Volume comparison

Here A and M represent the filled volumes of the segmented objects.

- False positive:

$$FP(M, A) = \frac{|A| - |M \cap A|}{|M|}$$

and false negative:

$$FN(M, A) = \frac{|M| - |M \cap A|}{|M|}$$

It is important to notice that the normalization is calculated with respect to M , which is the reference volume, the manual segmentation (the ground truth). Thus, FP represents the voxels of the automatic segmentation that are detected as being in the object but that, in fact, do not belong to it. FN represents the voxels that are in the object but are not detected by the automatic segmentation. Depending on the application, the relevance of FP or FN varies. For example, if the goal is to detect the extension of a tumor for treatment, FN should be very close to 0.

FP and FN can also be normalized with respect to A . In that case, for example, a small FP_A indicates a good accuracy of the segmentation:

$$FP_A(M, A) = \frac{|A| - |M \cap A|}{|A|}$$

- Percentage error with respect to M :

$$PE_M(M, A) = \frac{|M \cup A| - |M \cap A|}{|M|} \times 100, \quad \text{or}$$

$$PE_M(M, A) = [FP(M, A) + FN(M, A)] \times 100$$

and with respect to A :

$$PE_A(M, A) = \frac{|M \cup A| - |M \cap A|}{|A|} \times 100$$

- Intersection-union ratio between both volumes:

$$IUR(M, A) = \frac{|M \cap A|}{|M \cup A|}$$

- Similarity index between both volumes:

$$S(M, A) = \frac{2|M \cap A|}{|M| + |A|}$$

As explained in (Zijdenbos 94) the similarity index S is sensitive to variations in shape, size and position (as illustrated in Section A.3) and a value of $S > 0.7$ indicates a strong agreement.

A.2 Surface comparison

Here A and M represent the surfaces of the segmented objects.

- Mean distance between the surfaces:

$$D_{mean}(M, A) = \frac{1}{2}[d_{mean}(M, A) + d_{mean}(A, M)]$$

with

$$d_{mean}(M, A) = \frac{1}{|M|} \sum_{m \in M} D(m, A)$$

where $D(m, A) = [\min_{a \in A} d(m, a)]$ and d is the Euclidean distance.

- RMS (root mean square) distance between the surfaces:

$$D_{RMS}(M, A) = \sqrt{\frac{1}{2}[d_{RMS}(M, A)^2 + d_{RMS}(A, M)^2]}$$

with

$$d_{RMS}(M, A) = \sqrt{\frac{1}{|M|} \sum_{m \in M} D(m, A)^2}.$$

- Hausdorff distance between the surfaces:

$$D_H(M, A) = \max(d_H(M, A), d_H(A, M))$$

with

$$d_H(M, A) = \max_{m \in M} D(m, A).$$

As this measure is a maximum distance, it is a particularly severe evaluation which is sensitive to “peaks” in the segmentation (see Section A.3 for some examples).

A.3 Illustrative examples

Figure A.1 illustrates several examples of comparison between objects (in 2D for the sake of readability). The detailed descriptions of these objects are in Table A.1.

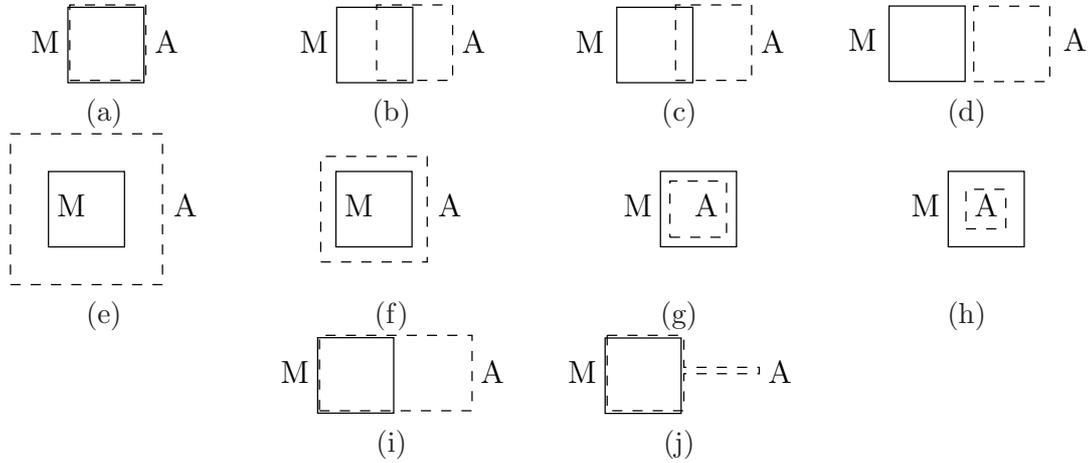


Figure A.1: Illustrative examples for volume/surface comparison. M (continuous line) represents the manually segmented object and A (dashed line) represents the automatically segmented object.

	Minimum x	Maximum x	Minimum y	Maximum y
Object M	50	100	75	125
Object A in example (a)	50	100	75	125
Object A in example (b)	75	125	75	125
Object A in example (c)	87	137	75	125
Object A in example (d)	105	155	75	125
Object A in example (e)	25	125	50	150
Object A in example (f)	40	110	65	135
Object A in example (g)	57	93	82	118
Object A in example (h)	63	88	88	113
Object A in example (i)	50	150	75	125
Object A in example (j)	50	100	75	125
(two rectangles)	100	150	98	102

Table A.1: Description of the examples in Figure A.1. Units are mm.

Table A.2 shows the results of the comparisons of objects A and M using the different criteria defined in Sections A.1 and A.2. Here we present a brief interpretation of the criteria and these

results:

- The ideal case with a perfect segmentation is illustrated in (a). Here, all values are 0 except IUR and S which are equal to 1. In general, the lower (closer to 0) values for FP , FN , PE_M , PE_A and the distances D_{mean} , D_{RMS} and D_H , the better. And the higher (closer to 1) values for IUR and S , the better.
- When the automatic segmentation A contains the manual segmentation M , FN takes value 0 (examples (e), (f), (i) and (j)). When M contains A , FP is 0 (examples (g) and (h)).
- The difference between FP and S is that S is more sensitive to variations in shape, size and position. For example, if object A completely overlaps object M which is one half of A (case (i)), FP is almost 1, which means “incorrect detection”, whereas S value is 0.67, which means “not so bad detection” (as the value is close to 0.7).
- IUR and S behave slightly differently. IUR takes lower values than S and it decreases faster than S when the differences between A and M increase (see examples (a) to (d)). Therefore, IUR is stricter than S .
- One very bad result is illustrated in (d). Here the automatic segmentation A is not at all superimposed on the manual segmentation M . Thus, FP and FN are equal to 1, and PE_M and PE_A have maximum values. IUR and S are 0. However, the values of the distances (mean, RMS or Hausdorff) are comparable with the results for other cases (for example, cases (e), (i) or (j)). We can conclude that the quality of a segmentation cannot be only evaluated by taking into account the distances criteria D_{mean} , D_{RMS} and D_H .
- In particular, Hausdorff distance D_H is a very severe criterion. It measures the maximum distance between one point in M and one point in A . Thus the values of D_H for cases (i) and (j) are the same. The value of D_H in case (d) is very similar to the previous ones, however the segmentation in (d) is clearly worse than in (i) or (j). The conclusion is that the information furnished by D_H has to be completed with other criteria.

Which is the best segmentation among the examples illustrated in Figure A.1?

With respect to the volume measures it is clear that the best case is (j). FP and FN have very small values or 0, error percentages PE_M and PE_A are very small, and IUR and S are very close to 1. The mean distance D_{mean} is very small too. However D_{RMS} and D_H are not the best. The best distance measures are those of case (g), but the volume criteria are not specially good. In particular, the value of FN could be too high for some applications (for example, if M is a tumor and the automatic segmentation does not detect it completely).

In conclusion, the selection of a criterion (or a group of criteria) in order to compare two objects depends on the particular application and on the type of differences to be highlighted by the user. Thus, in some measure, comparison still remains subjective.

Example	FP	FN	PE_M (%)	PE_A (%)	IUR	S	D_{mean} (mm)	D_{RMS} (mm)	D_H (mm)
(a)	0.000	0.000	0.0	0.0	1.00	1.00	0.0	0.0	0
(b)	0.490	0.490	98.0	98.0	0.34	0.51	12.5	16.1	25
(c)	0.725	0.725	145.1	145.1	0.16	0.27	18.5	23.2	37
(d)	1.000	1.000	200.0	200.0	0.00	0.00	30.0	36.3	55
(e)	2.922	0.000	292.2	74.5	0.25	0.41	26.2	26.3	35
(f)	0.938	0.000	93.8	48.4	0.52	0.68	10.2	10.3	14
(g)	0.000	0.474	47.4	90.0	0.53	0.69	7.1	7.2	9
(h)	0.000	0.740	74.0	284.8	0.26	0.41	13.0	13.1	18
(i)	0.980	0.000	98.0	49.5	0.50	0.67	9.9	19.3	50
(j)	0.096	0.000	9.6	8.8	0.91	0.95	4.5	12.5	50

Table A.2: Results of the comparisons for the examples in Figure A.1.

List of Patents and Publications

- Publication of a chapter in a book:

G. Peters, S. Muller, S. Bernard and I. Bloch. *Wavelets and Fuzzy Contours in 3D CAD for Digital Breast Tomosynthesis*. In M. Nachttegaal, D. Van der Weken, E. Kerre and W. Philips, editors, *Soft Computing and Image Processing: Recent Advances*. Springer Verlag, 2006

- International conferences with refereed papers in conference proceedings:

G. Peters, S. Muller, S. Bernard, R. Iordache and I. Bloch. *A Hybrid Active Contour Model for Mass Detection in Digital Breast Tomosynthesis*. In *Medical Imaging: Proceedings of the SPIE*, 2007

S. Bernard, S. Muller, G. Peters and R. Iordache. *Fast microcalcification detection on digital breast tomosynthesis datasets*. In *Medical Imaging: Proceedings of the SPIE*, 2007

S. Bernard, S. Muller, G. Peters and R. Iordache. *Microcalcification detection on simple back-projection reconstructed slices using wavelets*. In *Proceedings of the International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, volume 1, pages 84–86, Osaka, Japan, 2006

G. Peters, S. Muller, S. Bernard, R. Iordache and I. Bloch. *Reconstruction-Independent 3D CAD for Mass Detection in Digital Breast Tomosynthesis Using Fuzzy Particles*. In *Medical Imaging: Proceedings of the SPIE*, 2006

F.W. Wheeler, A.G.A. Perera, B.E. Claus, S.L. Muller, G. Peters and J.P. Kaufhold. *Micro-calcification detection in digital tomosynthesis mammography*. In *Medical Imaging: Proceedings of the SPIE*, volume 6144, pages 671–682, 2006

G. Peters, S. Muller, S. Bernard, R. Iordache, F. Wheeler and I. Bloch. *Reconstruction-Independent 3D CAD for Calcification Detection in Digital Breast Tomosynthesis Using Fuzzy Particles*. In Springer, editor, *Proceedings of the 10th Iberoamerican Conference on Pattern Recognition (CIARP)*, pages 400–408, 2005

- Patents:

Five patents have been submitted during the course of our work that are currently awaiting registration with French and U.S. American patents offices.

Bibliography

- [Abdel-Mottaleb 96] M. Abdel-Mottaleb, C.S. Carman, C.R. Hill and S. Vafai. *Locating the boundary between the breast skin edge and the background in digitized mammograms*. In K. Doi, L.L. Giger, R.M. Nishikawa and R.A. Schmidt, editors, *Digital Mammography: Proceedings of the International Workshop on Digital Mammography (IWDM)*, pages 467–470. Elsevier, Amsterdam, 1996.
- [ACS 07] ACS. *Cancer Facts & Figures 2007*. <http://www.cancer.org>, 2007.
- [Aczel 06] J. Aczel. *Entropies, characterizations, applications and some history*. Elsevier, 2006.
- [Andersen 84] A. H. Andersen and A. C. Kak. *Simultaneous Algebraic Reconstruction Technique (SART): A Superior Implementation of the ART Algorithm*. *Ultrasonic Imaging*, vol. 6, no. 1, pages 81–94, 1984.
- [Arbach 03] L. Arbach, J.M. Reinhardt, D.L. Bennett and G. Fallouh. *Mammographic masses classification: comparison between backpropagation neural network (BNN), K nearest neighbors (KNN), and human readers*. In *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, volume 3, pages 1441–1444, 2003.
- [Avinash 06] G. Avinash, K. Israni and B. Li. *Characterization of point spread function in linear digital tomosynthesis: a simulation study*. In *Medical Imaging: Proceedings of the SPIE*, volume 6142, pages 1669–1676, 2006.
- [Bankman 94] IN Bankman, J. Tsai, DW Kim, OB Gatewood and WR Brody. *Detection of microcalcification clusters using neural networks*. In *Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 590–591, 1994.
- [Bassett 87] L.W. Bassett and R.H. Gold. *Breast Cancer Detection: Mammography and Other Methods in Breast Imaging*. Grune & Stratton, 1987.
- [Bernard 06] S. Bernard, S. Muller, G. Peters and R. Iordache. *Microcalcification detection on simple back-projection reconstructed slices using wavelets*. In *Proceedings of the International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, volume 1, pages 84–86, Osaka, Japan, 2006.
- [Bernard 07] S. Bernard, S. Muller, G. Peters and R. Iordache. *Fast microcalcification detection on digital breast tomosynthesis datasets*. In *Medical Imaging: Proceedings of the SPIE*, 2007.

- [Betal 97] D. Betal, N. Roberts and G.H. Whitehouse. *Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology*. British Journal of Radiology, vol. 70, no. 837, pages 903–917, 1997.
- [Bird 90] R.E. Bird. *Professional quality assurance for mammography screening programs*. Radiology, vol. 177, no. 2, page 587, 1990.
- [Blessing 06] M. Blessing, D. Godfrey, F. Lohr and F. Yin. *Analysis of the Point Spread Function of Isocentric Digital Tomosynthesis (DTS)*. Medical Physics, vol. 33, page 2266, 2006.
- [Bloch 96] I. Bloch. *Information Combination Operators for Data Fusion: A Comparative Review with Classification*. IEEE Transactions on Systems, Man, and Cybernetics, vol. 26, no. 1, pages 52–67, 1996.
- [Blumer 87] A. Blumer, A. Ehrenfeucht, D. Haussler and M.K. Warmuth. *Occam's razor*. Information Processing Letters, vol. 24, no. 6, pages 377–380, 1987.
- [Bothorel 96] S. Bothorel. *Analyse d'image par arbre de décision floue - Application à la classification sémiologique des amas de microcalcifications en mammographie numérique*. PhD thesis, Université Paris 6, 1996.
- [Bothorel 97] S. Bothorel, B. Bouchon-Meunier and S. Muller. *A Fuzzy Logic Based Approach for Seminological Analysis of Microcalcifications in Mammographic Images*. International Journal for Intelligent Systems, vol. 12, pages 819–848, 1997.
- [Braccialarghe 96] D. Braccialarghe and G.H. Kaufmann. *Contrast enhancement of mammographic features: a comparison of four methods*. Optical Engineering, vol. 35, no. 1, pages 76–80, 1996.
- [Brzakovic 90] D. Brzakovic, X.M. Luo and P. Brzakovic. *An approach to automated detection of tumors in mammograms*. IEEE Transactions on Medical Imaging, vol. 9, no. 3, pages 233–241, 1990.
- [Buchanan 83] J.B. Buchanan, J.S. Spratt and L.S. Heuser. *Tumor growth, doubling times, and the inability of the radiologist to diagnose certain cancers*. Radiological Clinics of North America, vol. 21, no. 1, pages 115–26, 1983.
- [Canny 86] J. Canny. *A computational approach to edge detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pages 679–698, 1986.
- [Cao 04] A. Cao, Q. Song, X. Yang, S. Liu and C. Guo. *Mammographic mass detection by vicinal support vector machine*. In Proceedings of the IEEE International Joint Conference on Neural Networks, volume 3, pages 1953–1958, 2004.
- [Chakraborty 96] A. Chakraborty, L.H. Staib and J.S. Duncan. *Deformable boundary finding in medical images by integrating gradient and region information*. IEEE Transactions on Medical Imaging, vol. 15, no. 6, pages 859–870, 1996.
- [Chan 87] H.-P. Chan, C.J. Vyborny, H. MacMahon, C.E. Metz, K. Doi and E.A. Sickles. *Digital mammography. ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications*. Interventional Radiology, vol. 22, no. 7, pages 581–9, 1987.

- [Chan 01a] H.-P. Chan, M.A. Helvie, N. Petrick, B. Sahiner, D.D. Adler, C. Paramagul, M.A. Roubidoux, C.E. Blane, L.K. Joynt, T.E. Wilson *et al.* *Digital mammography: observer performance study of the effects of pixel size on the characterization of malignant and benign microcalcifications*. *Academic Radiology*, vol. 8, no. 6, pages 454–66, 2001.
- [Chan 01b] T.F. Chan and L.A. Vese. *Active contours without edges*. *IEEE Transactions on Image Processing*, vol. 10, no. 2, pages 266–277, 2001.
- [Chan 04] H.-P. Chan, J. Wei, B. Sahiner, E. A. Rafferty, T. Wu, M. A. Roubidoux, R. H. Moore, D. B. Kopans, L. M. Hadjiiski and M. A. Helvie. *Computerized Detection of Masses on Digital Tomosynthesis Mammograms - A Preliminary Study*. In Springer, editor, *Digital Mammography: Proceedings of the International Workshop on Digital Mammography (IWDM)*, Chapel Hill, NC, 2004.
- [Chan 05] H.-P. Chan, J. Wei, B. Sahiner, E.A. Rafferty, T. Wu, M.A. Roubidoux, R.H. Moore, D.B. Kopans, L.M. Hadjiiski and M.A. Helvie. *Computer-aided Detection System for Breast Masses on Digital Tomosynthesis Mammograms: Preliminary Experience 1*. *Radiology*, vol. 237, no. 3, pages 1075–1080, 2005.
- [Chang 97] C.M. Chang and A. Laine. *Enhancement of mammograms from oriented information*. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 524–527, 1997.
- [Chen 97] C.H. Chen and G.G. Lee. *On Digital Mammogram Segmentation and Microcalcification Detection Using Multiresolution Wavelet Analysis*. *Graphical Models and Image Processing*, vol. 59, no. 5, pages 349–364, 1997.
- [Chen 00] J.-L. Chen and J.-Y. Chang. *Fuzzy Perceptron Neural Networks for Classifiers with Numerical Data and Linguistic Rules as Input*. *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pages 730–745, 2000.
- [Cheng 94] S.N.C. Cheng, H.P. Chan, M.A. Helvie, M.M. Goodsitt, D.D. Adler and D.C. St Clair. *Classification of mass and nonmass regions on mammograms using artificial neural networks*. *Journal of Imaging Science and Technology*, vol. 38, no. 6, pages 598–603, 1994.
- [Cheng 98] H. D. Cheng, Y. M. Lui and R. I. Freimanis. *A Novel Approach to Microcalcification Detection Using Fuzzy Logic Technique*. *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pages 442–450, 1998.
- [Cheng 03] H.D. Cheng, X. Cai, X. Chen, L. Hu and X. Lou. *Computer-aided Detection and Classification of Microcalcifications in Mammograms: A Survey*. *Pattern Recognition*, vol. 36, no. 12, pages 2967–2991, 2003.
- [Cheng 06] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai and H.N. Du. *Approaches for Automated Detection and Classification of Masses in Mammograms*. *Pattern Recognition*, vol. 39, no. 4, pages 646–668, 2006.
- [Chiang 02] I.J. Chiang and J.Y. Hsu. *Fuzzy classification trees for data analysis*. *Fuzzy Sets and Systems*, vol. 130, no. 1, pages 87–99, 2002.

- [Chitre 93] Y. Chitre, A.P. Dhawan and M. Moskowitz. *Artificial neural network based classification of mammographic microcalcifications using image structure features*. In Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, volume 1, pages 50–52, 1993.
- [Cicccone 92] G. Cicccone, P. Vineis, A. Frigerio and N. Segnan. *Inter-observer and intra-observer variability of mammogram interpretation: a field study*. European Journal of Cancer, vol. 28, no. 6-7, pages 1054–8, 1992.
- [Claus 02] B.E.H. Claus and J.W. Eberhard. *A new method for 3D reconstruction in digital tomosynthesis*. In Medical Imaging: Proceedings of the SPIE, pages 814–824. SPIE, 2002.
- [Cover 67] T.M. Cover and P.E. Hart. *Nearest Neighbor Pattern Classification*. IEEE Transactions on Information Theory, vol. 13, no. 1, pages 21–27, 1967.
- [DeLuca 72] A. DeLuca and S. Termini. *A definition of a non-probabilistic entropy in the setting of fuzzy sets*. Information and Control, vol. 20, pages 301–12, 1972.
- [Dengler 93] J. Dengler, S. Behrens and J.F. Desaga. *Segmentation of microcalcifications in mammograms*. IEEE Transactions on Medical Imaging, vol. 12, no. 4, pages 634–642, 1993.
- [Deriche 87] R. Deriche. *Using Canny’s criteria to derive a recursively implemented optimal edge detector*. International Journal of Computer Vision, vol. 1, no. 2, pages 167–187, 1987.
- [Dhawan 96] A.P. Dhawan, Y. Chitre, C. Kaiser-Bonasso *et al.* *Analysis of mammographic microcalcifications using gray-level image structure features*. IEEE Transactions on Medical Imaging, vol. 15, no. 3, pages 246–259, 1996.
- [Diday 76] E. Diday and J.C. Simon. *Clustering analysis*. Digital Pattern Recognition, vol. 10, pages 47–94, 1976.
- [Dobbins III 03] J.T. Dobbins III and D.J. Godfrey. *Digital X-Ray Tomosynthesis: Current State of the Art and Clinical Potential*. Physics in Medicine and Biology, vol. 48, pages R65–R106, 2003.
- [Dobbins 87] JT Dobbins, AO Powell and YK Weaver. *Matrix Inversion Tomosynthesis: Initial image reconstruction*. In RSNA 73rd Scientific Assembly (Chicago, IL), 1987.
- [D’Orsi 98] C.J. D’Orsi, L.W. Bassett, S.A. Feiget *et al.* *Breast imaging reporting and data system (BI-RADS)*. Reston, VA: American College of Radiology, 1998.
- [Dubois 80] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New-York, 1980.
- [Duffy 03] S.W. Duffy, L. Tabar, B. Vitak, N.E. Day, R.A. Smith, H.H. Chen and M.F. Yen. *The relative contributions of screen-detected in situ and invasive breast carcinomas in reducing mortality from the disease*. European Journal of Cancer, vol. 39, no. 12, pages 1755–60, 2003.

- [Ecabert 02] O. Ecabert and J.P. Thiran. *Variational image segmentation by unifying region and boundary information*. In Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02), volume 2, pages 885–888, 2002.
- [El-Naqa 02] I. El-Naqa, Y. Yang, M.N. Wernick, N.P. Galatsanos and R.M. Nishikawa. *A support vector machine approach for detection of microcalcifications*. IEEE Transactions on Medical Imaging, vol. 21, no. 12, pages 1552–63, 2002.
- [Elmore 94] J.G. Elmore, C.K. Wells, C.H. Lee, D.H. Howard and A.R. Feinstein. *Variability in Radiologists' Interpretations of Mammograms*. The New England Journal of Medicine, vol. 331, no. 22, page 1493, 1994.
- [Fan 99] J. Fan and W. Xie. *Distance measure and induced fuzzy entropy*. Fuzzy Sets and Systems, vol. 104, no. 2, pages 305–314, 1999.
- [Fan 02] J.L. Fan and Y.L. Ma. *Some new fuzzy entropy formulas*. Fuzzy Sets and Systems, vol. 128, no. 2, pages 277–284, 2002.
- [Feldkamp 84] L.A. Feldkamp, L.C. Davis and J.W. Kress. *Practical cone-beam algorithm*. J. Opt. Soc. Am. A, vol. 1, no. 6, pages 612–619, 1984.
- [Fischmann 05] A. Fischmann, KC Siegmann, A. Wersebe, CD Claussen and M. Muller-Schimpfle. *Comparison of full-field digital mammography and film-screen mammography: image quality and lesion detection*. British Journal of Radiology, vol. 78, no. 928, page 312, 2005.
- [Fisher 36] R.A. Fisher. *The use of multiple measures in taxonomic problems*. Ann. Eugenics, vol. 7, pages 179–188, 1936.
- [Floyd Jr 94] C.E. Floyd Jr, J.Y. Lo, A.J. Yun, D.C. Sullivan and P.J. Kornguth. *Prediction of breast cancer malignancy using an artificial neural network*. Cancer, vol. 74, no. 11, pages 2944–8, 1994.
- [Fogel 98] D.B. Fogel, E.C. Wasson, E.M. Boughton and V.W. Porto. *Evolving artificial neural networks for screening features from mammograms*. Artificial Intelligence in Medicine, vol. 14, no. 3, pages 317–326, 1998.
- [Freer 01] T.W. Freer and M.J. Ulissey. *Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center1*. Radiology, vol. 220, pages 781–786, 2001.
- [Freixenet 02] J. Freixenet, X. Muñoz, D. Raba, J. Martí and X. Cufí. *Yet another survey on image segmentation: region and boundary information integration*. In Proceedings of the 7th European Conference on Computer Vision-Part III, volume 2352, pages 408–422. Springer-Verlag London, UK, 2002.
- [Fukunaga 90] K. Fukunaga. Introduction to Statistical Pattern Recognition. Elsevier, 1990.
- [Galloway 75] MM Galloway. *Texture classification using gray level run lengths*. Computer Graphics and Image Processing, vol. 4, pages 172–179, 1975.
- [GAO. 91] US GAO. *Breast Cancer, 1971-1991: Prevention, Treatment and Research*. 1991.

- [Gavrielides 02] M.A. Gavrielides, J.Y. Lo and C.E. Floyd Jr. *Parameter optimization of a computer-aided diagnosis scheme for the segmentation of microcalcification clusters in mammograms*. Medical Physics, vol. 29, page 475, 2002.
- [Gilbert 06] F.J. Gilbert, S.M. Astley, M.A. McGee, M.G.C. Gillan, C.R.M. Boggis, P.M. Griffiths and S.W. Duffy. *Single Reading with Computer-aided Detection and Double Reading of Screening Mammograms in the United Kingdom National Breast Screening Program*. Radiology, vol. 241, pages 47–53, 2006.
- [Gimenez 99] V. Gimenez, D. Manrique, J. Rios and A. Vilarrasa. *Iterative method for automatic detection of masses in digital mammograms fro computer-aided diagnosis*. In Medical Imaging: Proceedings of the SPIE, volume 3661, pages 1086–1093. SPIE, 1999.
- [Gordon 70] R. Gordon, R. Bender and G.T. Herman. *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography*. Journal of Theoretical Biology, vol. 29, no. 3, pages 471–481, 1970.
- [Grimaud 91] M. Grimaud. *La geodesie numérique en morphologie mathématique. Application à la détection automatique de microcalcifications en mammographie numérique*. PhD thesis, Ecole des Mines de Paris, 1991.
- [Grosjean 06] B. Grosjean, S. Muller and H. Souchay. *Lesion detection using an a-contrario detector in simulated digital mammograms*. Medical Imaging: Proceedings of the SPIE, vol. 6146, 2006.
- [Guliato 98] D. Guliato, R.M. Rangayyan, W.A. Carnielli, J.A. Zuffo and J.E.L. Desautels. *Segmentation of breast tumors in mammograms by fuzzy region growing*. Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE, vol. 2, 1998.
- [Gurcan 97] M.N. Gurcan, Y. Yardimci, A.E. Cetin and R. Ansari. *Detection of microcalcifications in mammograms using higher orderstatistics*. Signal Processing Letters, IEEE, vol. 4, no. 8, pages 213–216, 1997.
- [Haddon 90] J.F. Haddon and J.F. Boyce. *Image segmentation by unifying region and boundary information*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 10, pages 929–948, 1990.
- [Hadjarian 98] A. Hadjarian, J. Bala, S. Gutta, S. Trachiots and P. Pachowicz. *The fusion of supervised and unsupervised techniques for segmentation of abnormal regions*. In N. Karssemeijer, M. Thijssen, J. Hendriks and L. Erning, editors, Digital Mammography: Proceedings of the International Workshop on Digital Mammography (IWDM), pages 299–302. Kluwer, Amsterdam, 1998.
- [Hartmann 82] C. Hartmann, P. Varshney, K. Mehrotra and C. Gerberich. *Application of information theory to the construction of efficient decision trees*. IEEE Transactions on Information Theory, vol. 28, no. 4, pages 565–577, 1982.
- [Holland 82] R. Holland, M. Mravunac, J. H. Hendriks and B. V. Bekker. *So-called Interval Cancers of the Breast. Pathologic and Radiologic Analysis of Sixty-Four Cases*. Cancer, vol. 49, pages 2527–2533, 1982.

- [Holland 92] J.H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT Press, 1992.
- [Hüllermeier 05] E. Hüllermeier. *Fuzzy methods in machine learning and data mining: Status and prospects*. *Fuzzy Sets and Systems*, vol. 156, no. 3, pages 387–406, 2005.
- [Huo 95] Z. Huo, M.L. Giger, C.J. Vyborny, U. Bick, P. Lu, D.E. Wolverton and R.A. Schmidt. *Analysis of spiculation in the computerized classification of mammographic masses*. *Medical Physics*, vol. 22, page 1569, 1995.
- [Huo 98] Z. Huo, M.L. Giger, C.J. Vyborny, D.E. Wolverton, R.A. Schmidt and K. Doi. *Automated computerized classification of malignant and benign masses on digitized mammograms*. *Academic Radiology*, vol. 5, no. 3, pages 155–68, 1998.
- [Ichihashi 96] H. Ichihashi, T. Shirai, K. Nagasaka and T. Miyoshi. *Neuro-fuzzy ID3: a method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning*. *Fuzzy Sets and Systems*, vol. 81, no. 1, pages 157–167, 1996.
- [Imaginis 07] Imaginis. *General information on breast cancer*. <http://imaginis.com/breasthealth>, 2007.
- [Ishibuchi 00] H. Ishibuchi and M. Nii. *Neural networks for soft decision making*. *Fuzzy Sets and Systems*, vol. 115, no. 1, pages 121–140, 2000.
- [Iyer 00] N.S. Iyer, A. Kandel and M. Schneider. *Feature-based fuzzy classification for interpretation of mammograms*. *Fuzzy Sets and Systems*, vol. 114, no. 2, pages 271–280, 2000.
- [Janikow 98] C.Z. Janikow. *Fuzzy decision trees: issues and methods*. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 28, no. 1, pages 1–14, 1998.
- [Jiang 97] Y. Jiang, R.M. Nishikawa, D.E. Wolverton, C.E. Metz, R.A. Schmidt and K. Doi. *Computerized classification of malignant and benign clustered-microcalcifications in mammograms*. *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pages 521–523, 1997.
- [Johns 87] P.C. Johns and M.J. Yaffe. *X-ray characterisation of normal and neoplastic breast tissues*. *Phys. Med. Biol*, vol. 32, no. 6, pages 675–695, 1987.
- [Kallergi 92] M. Kallergi, K. Woods, L.P. Clarke, W. Qian and R.A. Clark. *Image segmentation in digital mammography: comparison of local thresholding and region growing algorithms*. *Comput Med Imaging Graph*, vol. 16, no. 5, pages 323–31, 1992.
- [Kanzaki 92] K. Kanzaki. *The Use of Morphology and Fuzzy Set Theory in FLIR Target Segmentation and Classification*. PhD thesis, Polytechnic University of New York, 1992.
- [Karssemeijer 93] N. Karssemeijer. *Adaptive Noise Equalization and Image Analysis in Mammography*. In *Proceedings of the 13th International Conference on Information Processing in Medical Imaging*, pages 472–486. Springer-Verlag London, UK, 1993.

- [Karssemeijer 96] N. Karssemeijer and G. M. te Brake. *Detection of Stellate Distortions in Mammograms*. IEEE Transactions on Medical Imaging, vol. 15, no. 5, pages 611–619, 1996.
- [Karssemeijer 03] N. Karssemeijer, J.D.M. Otten, A.L.M. Verbeek, J.H. Groenewoud, H.J. de Koning, J.H.C.L. Hendriks and R. Holland. *Computer-aided Detection versus Independent Double Reading of Masses on Mammograms*. Radiology, vol. 227, pages 192–200, 2003.
- [Kass 88] M. Kass, A. Witkin and D. Terzopoulos. *Snakes: Active contour models*. International Journal of Computer Vision, vol. 1, no. 4, pages 321–331, 1988.
- [Kegelmeyer Jr 94] W.P. Kegelmeyer Jr. *Computer-aided mammographic screening for spiculated lesions*. Radiology, vol. 191, no. 2, pages 331–337, 1994.
- [Keller 85] J.M. Keller, M.R. Gray and J. Givens Jr. *Fuzzy K-nearest neighbor algorithm*. IEEE Transactions on Systems, Man, and Cybernetics, vol. 15, no. 4, pages 580–584, 1985.
- [Kerlikowske 95] K. Kerlikowske, D. Grady, S. M. Rubin, C. Sandroock and V. L. Ernster. *Efficacy of Screening Mammography. A Meta-Analysis*. The Journal of the American Medical Association, vol. 273, no. 2, pages 149–154, 1995.
- [Kilday 93] J. Kilday, F. Palmieri and MD Fox. *Classifying mammographic lesions using computerized image analysis*. IEEE Transactions on Medical Imaging, vol. 12, no. 4, pages 664–669, 1993.
- [Kim 97] J.K. Kim, J.M. Park, K.S. Song and H.W. Park. *Adaptive mammographic image enhancement using first derivative and local statistics*. IEEE Transactions on Medical Imaging, vol. 16, no. 5, pages 495–502, 1997.
- [Kim 99] J.K. Kim and H.W. Park. *Statistical Textural Features for Detection of Microcalcifications in Digitized Mammograms*. IEEE Transactions on Medical Imaging, vol. 18, no. 3, page 231, 1999.
- [Kim 05] S. Kim. *A hybrid level set approach for efficient and reliable image segmentation*. In Proceedings of the IEEE International Symposium on Signal Processing and Information Technology., volume 1, pages 743–748, 2005.
- [Kirsch 72] R. A. Kirsch. *Resynthesis of Biological Images from Tree Structured Decomposition Data*. In IFIP Working Conference on Computer Graphic Languages, pages 1–19, 1972.
- [Kohonen 82] T. Kohonen. *Self-organized formation of topologically correct feature maps*. Biological Cybernetics, vol. 43, no. 1, pages 59–69, 1982.
- [Kopans 89] D.B. Kopans. *Asymmetric breast tissue*. Radiology, vol. 171, no. 3, pages 639–643, 1989.
- [Kosko 86] B. Kosko. *Fuzzy entropy and conditioning*. Information Sciences, vol. 40, no. 2, pages 165–174, 1986.
- [Kosko 92] B. Kosko. *Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1992.

- [Koutras 06] A. Koutras, I. Christoyianni, G. Georgoulas and E. Dermatas. Computer Aided Classification of Mammographic Tissue Using Independent Component Analysis and Support Vector Machines, volume 4132. Springer Berlin / Heidelberg, 2006.
- [Kramer 99] D. Kramer and F. Aghdasi. *Texture analysis techniques for the classification of microcalcifications in digitised mammograms*. In Proceedings of the 5th IEEE AFRICON conference, volume 1, pages 395–400, 1999.
- [Lange 95] K. Lange and J.A. Fessler. *Globally convergent algorithms for maximum a posteriori transmission tomography*. IEEE Transactions on Image Processing, vol. 4, no. 10, pages 1430–1438, 1995.
- [Lanyi 87] M. Lanyi. *Diagnosis and Differential Diagnosis of Breast Calcifications*. Springer-Verlag, 1987.
- [Lau 92] C. Lau. *Neural networks: theoretical foundations and analysis*. IEEE Press, 1992.
- [Lauritsch 98] G. Lauritsch and W. Haerer. *Theoretical framework for filtered back projection in tomosynthesis*. In Medical Imaging: Proceedings of the SPIE, volume 3338, pages 1127–1137, 1998.
- [Le Gal 84] M. Le Gal, G. Chavanne and D. Pellier. *Diagnostic value of clustered microcalcifications discovered by mammography (apropos of 227 cases with histological verification and without a palpable breast tumor)*. Bulletin du Cancer, vol. 71, no. 1, pages 57–64, 1984.
- [Lee 94] H.-M. Lee and W.-T. Wang. *A Neural Network Architecture for Classification of Fuzzy Inputs*. Fuzzy Sets and Systems, vol. 63, no. 2, pages 159–173, 1994.
- [Lee 00] Y.J. Lee, J.M. Park and H.W. Park. *Mammographic mass detection by adaptive thresholding and region growing*. International Journal of Imaging Systems and Technology, vol. 11, no. 5, pages 340–346, 2000.
- [Levy 00] L.D. Levy, S.L. Muller, K. Priday and A. Rick. *Impact of Pixel Size for the Differentiation of Benign and Malignant Microcalcification Clusters in Digital Mammography*. Presented at the Annual Meeting of the Radiological Society of North America (RSNA), 2000.
- [Li 95] H.D. Li, M. Kallergi, L.P. Clarke, V.K. Jain and R.A. Clark. *Markov random field for tumor detection in digital mammography*. IEEE Transactions on Medical Imaging, vol. 14, no. 3, pages 565–576, 1995.
- [Li 97] H. Li, KJ Liu, S.C.B. Lo, O.T. Inc and MD Jessup. *Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms*. IEEE Transactions on Medical Imaging, vol. 16, no. 6, pages 785–798, 1997.
- [Li 01] H. Li, Y. Wang, KJ Liu, SC Lo and MT Freedman. *Computerized radiographic mass detection—part I: Lesion site selection by morphological enhancement and contextual segmentation*. IEEE Transactions on Medical Imaging, vol. 20, no. 4, pages 289–301, 2001.

- [Li 04] Y. Li and J. Jiang. *Combination of SVM Knowledge for Microcalcification Detection in Digital Mammograms*, volume 3177. Springer Berlin / Heidelberg, 2004.
- [Lin 02] C.F. Lin and S.D. Wang. *Fuzzy support vector machines*. IEEE Transactions on Neural Networks, vol. 13, no. 2, pages 464–471, 2002.
- [Liu 06] S. Liu and J. Li. *Automatic medical image segmentation using gradient and intensity combined level set method*. In Proceedings of the IEEE EMBS Annual International Conference., volume 1, pages 3118–3121, 2006.
- [Malladi 95] R. Malladi, J.A. Sethian and B.C. Vemuri. *Shape modeling with front propagation: a level set approach*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 2, pages 158–175, 1995.
- [Mallat 92] S. Mallat and S. Zhong. *Characterization of signals from multiscale edges*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 7, pages 710–732, 1992.
- [Marsala 98] C. Marsala. *Apprentissage inductif en présence de données imprécises: Construction et utilisation d'arbres de décision flous*. PhD thesis, Université Paris 6, 1998.
- [Marsala 99] C. Marsala and B. Bouchon-Meunier. *An Adaptable System to Construct Fuzzy Decision Trees*. In 18th International Conference of the North American Fuzzy Information Processing Society (NAFIPS), pages 223–227, 1999.
- [Marsala 06] C. Marsala and B. Bouchon-Meunier. *Ranking Attributes to Build Fuzzy Decision Trees: a Comparative Study of Measures*. In Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ IEEE), pages 1777–1783, 2006.
- [Martin 79] J.E. Martin and M. Moskowitz. *Breast cancer missed by mammography*. American Journal of Roentgenology, vol. 132, no. 5, pages 737–739, 1979.
- [Méndez 98] A.J. Méndez, P.G. Tahoces, M.J. Lado, M. Souto and J.J. Vidal. *Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms*. Medical Physics, vol. 25, page 957, 1998.
- [Mertelmeier 06] T. Mertelmeier, J. Orman, W. Haerer and M.K. Dudam. *Optimizing filtered backprojection reconstruction for a breast tomosynthesis prototype device*. In Medical Imaging: Proceedings of the SPIE, volume 6142, page 61420F. SPIE, 2006.
- [Metz 78] C.E. Metz. *Basic principles of ROC analysis*. Semin Nucl Med, vol. 8, no. 4, pages 283–98, 1978.
- [Metz 86] C.E. Metz. *ROC methodology in radiologic imaging*. Interventional Radiology, vol. 21, no. 9, pages 720–33, 1986.
- [Muller 99] S. Muller. *Full-Field Digital Mammography Designed as a Complete System*. European Journal of Radiology, vol. 31, no. 1, pages 25–34, 1999.
- [Mumford 89] D. Mumford and J. Shah. *Optimal approximation by piecewise smooth functions and associated variational problems*. Comm. Pure Appl. Math, vol. 42, no. 4, pages 577–685, 1989.

- [Munoz 03] X. Munoz, J. Freixenet, X. Cufi and J. Marti. *Active regions for colour texture segmentation integrating region and boundary information*. In Proceedings of the International Conference on Image Processing., volume 3, 2003.
- [Murphy Jr 90] W.A. Murphy Jr, J.M. Destouet and B.S. Monsees. *Professional quality assurance for mammography screening programs*. Radiology, vol. 175, no. 2, pages 319–20, 1990.
- [Netsch 99] T. Netsch and H.-O. Peitgen. *Scale-Space Signatures for the Detection of Clustered Microcalcifications in Digital Mammograms*. IEEE Transactions on Medical Imaging, vol. 18, no. 9, pages 774–786, 1999.
- [Niklason 97] L.T. Niklason. *Digital tomosynthesis in breast imaging*. Radiology, vol. 205, no. 2, pages 399–406, 1997.
- [Oliver 05] A. Oliver, J. Freixenet and R. Zwigelaar. *Automatic Classification of Breast Density*. IEEE International Conference on Image Processing (ICIP), vol. 2, 2005.
- [Olsen 01] O. Olsen and P. Gotzsche. *Cochrane review on screening for breast cancer with mammography*. The Lancet, vol. 358, no. 9290, pages 1340–1342, 2001.
- [Osher 88] S. Osher and J.A. Sethian. *Fronts propagating with curvature-dependent speed- Algorithms based on Hamilton-Jacobi formulations*. Journal of Computational Physics, vol. 79, no. 1, pages 12–49, 1988.
- [Paragios 02] N. Paragios and R. Deriche. *Geodesic active regions and level set methods for supervised texture segmentation*. International Journal of Computer Vision, vol. 46, no. 3, pages 223–247, 2002.
- [Pearl 88] J. Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988.
- [Peters 05] G. Peters, S. Muller, S. Bernard, R. Iordache, F. Wheeler and I. Bloch. *Reconstruction-Independent 3D CAD for Calcification Detection in Digital Breast Tomosynthesis Using Fuzzy Particles*. In Springer, editor, Proceedings of the 10th Iberoamerican Conference on Pattern Recognition (CIARP), pages 400–408, 2005.
- [Peters 06a] G. Peters, S. Muller, S. Bernard and I. Bloch. *Wavelets and Fuzzy Contours in 3D CAD for Digital Breast Tomosynthesis*. In M. Nachtgael, D. Van der Weken, E. Kerre and W. Philips, editors, Soft Computing and Image Processing: Recent Advances. Springer Verlag, 2006.
- [Peters 06b] G. Peters, S. Muller, S. Bernard, R. Iordache and I. Bloch. *Reconstruction-Independent 3D CAD for Mass Detection in Digital Breast Tomosynthesis Using Fuzzy Particles*. In Medical Imaging: Proceedings of the SPIE, 2006.
- [Peters 07] G. Peters, S. Muller, S. Bernard, R. Iordache and I. Bloch. *A Hybrid Active Contour Model for Mass Detection in Digital Breast Tomosynthesis*. In Medical Imaging: Proceedings of the SPIE, 2007.

- [Pfisterer 99] R.S. Pfisterer and F. Aghdasi. *Hexagonal wavelets for the detection of masses in digitized mammograms*. In *Medical Imaging: Proceedings of the SPIE*, volume 3813, pages 966–977. SPIE, 1999.
- [Pisano 01] E.D. Pisano, C. Kuzmiak, M. Koomen and W. Cance. *What every surgical oncologist should know about digital mammography*. *Seminars in Surgical Oncology*, vol. 20, no. 3, pages 181–186, 2001.
- [Pisano 05] E.D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J.K. Baum, S. Acharyya, E.F. Conant, L. L. Fajardo, L. Bassett, C. D’Orsi, R. Jong, M. Rebner and Digital Mammographic Imaging Screening Trial (DMIST) Investigators Group. *Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening*. *New England Journal of Medicine*, vol. 353, no. 17, pages 1773–1783, 2005.
- [Quinlan 86] J.R. Quinlan. *Induction of decision trees*. *Machine Learning*, vol. 1, no. 1, pages 81–106, 1986.
- [Quinlan 93] J.R. Quinlan. *C 4. 5: Programs for Machine Learning*. Morgan Kaufmann, 1993.
- [R2 Technologies 06] R2 Technologies. www.r2tech.com, 2006.
- [Ramdani 94] M. Ramdani. *Système d’Induction Formelle à Base de Connaissances Imprécises*. PhD thesis, Université Pierre et Marie Curie, 1994.
- [Rangayyan 97] R.M. Rangayyan, N.M. El-Faramawy, J.E.L. Desautels and O.A. Alim. *Measures of Acutance and Shape for Classification of Breast Tumors*. *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, page 799, 1997.
- [Reiser 06] I. Reiser, R.M. Nishikawa, M.L. Giger, T. Wu, E.A. Rafferty, R. Moore and D.B. Kopans. *Computerized mass detection for digital breast tomosynthesis directly from the projection images*. *Medical Physics*, vol. 33, no. 2, pages 482–491, 2006.
- [Remontet 03] L. Remontet, J. Esteve, AM Bouvier, P. Grosclaude, G. Launoy, F. Mene-goz, C. Exbrayat, B. Tretarre, P.M. Carli, A.V. Guizardet *al.* *National estimates: trends in incidence and mortality due to cancer in France during 1978 to 2000*. *Bulletin Epidemiologique Hebdomadaire*, no. 41/42, pages 190–193, 2003.
- [Ren 05] B. Ren, C. Ruth, J. Stein, A. Smith, I. Shaw and Z. Jing. *Design and performance of the prototype full field breast tomosynthesis system with selenium based flat panel detector*. In *Medical Imaging: Proceedings of the SPIE*, volume 5745, pages 550–561, 2005.
- [Rick 99] A. Rick, , S. Muller, S. Bothorel and M. Grimaud. *Quantitative Modeling of Microcalcification Detection in Digital Mammography*. In *MIC-CAI’99, Second International Conference*, Cambridge, England, pages 32–41, September 1999.
- [Rick 00] A. Rick, S. Bothorel, B. Bouchon-Meunier and S. Muller. *Fuzzy Techniques in Mammographic Image Processing*. In E. Kerre and M. Nachtgael, editors, *Studies in Fuzziness and Softcomputing: Fuzzy Techniques in Image Processing*. Springer, New-York, 2000.

- [Rifqi 96] M. Rifqi. *Mesures de comparaison, typicalité et classification d'objets flous : théorie et pratique*. PhD thesis, Université Paris 6, 1996.
- [Rumelhart 86] D.E. Rumelhart, G.E. Hintont and R.J. Williams. *Learning representations by back-propagating errors*. *Nature*, vol. 323, pages 533–536, 1986.
- [Russo 94] J. Russo. *Toward a physiological approach to breast cancer prevention*. *Cancer Epidemiology Biomarkers & Prevention*, vol. 3, no. 4, pages 353–364, 1994.
- [Safavian 91] S.R. Safavian and D. Landgrebe. *A survey of decision tree classifier methodology*. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pages 660–673, 1991.
- [Sahiner 96] B. Sahiner, H.-P. Chan, N. Petrick, D. Wei, M.A. Helvie, D.D. Adler and M.M. Goodsitt. *Classification of mass and normal breast tissue: a convolutionneural network classifier with spatial domain and texture images*. *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pages 598–610, 1996.
- [Sahiner 98] B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie and M.M. Goodsitt. *Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis*. *Medical Physics*, vol. 25, page 516, 1998.
- [Sahiner 01] B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie and L.M. Hadjiiski. *Improvement of mammographic mass characterization using spiculation measures and morphological features*. *Medical Physics*, vol. 28, page 1455, 2001.
- [Sameti 96] M. Sameti and R.K. Ward. *A fuzzy segmentation algorithm for mammogram partitioning*. In K. Doi, L.L. Giger, R.M. Nishikawa and R.A. Schmidt, editors, *Digital Mammography: Proceedings of the International Workshop on Digital Mammography (IWDM)*, pages 471–474. Elsevier, Amsterdam, 1996.
- [Sethian 96] J.A. Sethian. *A fast marching level set method for monotonically advancing fronts*. *Proceeding of the National Academy of Sciences of the United States of America (PNAS)*, vol. 93, no. 4, pages 1591–1595, 1996.
- [Sethian 99] JA Sethian. *Level Set Methods and Fast Marching Methods*. *Evolving interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Univ. Press, Cambridge, 1999.
- [Shannon 49] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949.
- [Smits 94] P. Smits, M. Mari, A. Teschioni, S. Dellapiane and F. Fontana. *Application of Fuzzy Methods to Segmentation of Medical Images*. In *Proceedings of the 5th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 910–915, 1994.
- [Strickland 97] RN Strickland and H.I. Hahn. *Wavelet transform methods for object detection and recovery*. *IEEE Transactions on Medical Imaging*, vol. 6, no. 5, pages 724–735, 1997.

- [Sun 93] C.T. Sun and J.S. Jang. *A neuro-fuzzy classifier and its applications*. Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ IEEE), pages 94–98, 1993.
- [Sun 04] X. Sun, W. Qian and D. Song. *Three-class classification in computer-aided diagnosis of breast cancer by support vector machine*. In Medical Imaging: Proceedings of the SPIE, volume 5370, pages 999–1007, 2004.
- [Suri 02] J.S. Suri, K. Liu, S. Singh, S.N. Laxminarayan, X. Zeng and L. Reden. *Shape recovery algorithms using level sets in 2-D/3-D medical imagery: a state-of-the-art review*. IEEE Transactions on Information Technology in Biomedicine, vol. 6, no. 1, pages 8–28, 2002.
- [Suryanarayanan 00] S. Suryanarayanan, A. Karellas, S. Vedantham, S.J. Glick, C.J. D’Orsi, S.P. Baker and R.L. Webber. *Comparison of tomosynthesis methods used with digital mammography*. Academic Radiology, vol. 7, no. 12, pages 1085–97, 2000.
- [Tabár 01] L. Tabár, P.B. Dean and T. Tot. Teaching Atlas of Mammography. Thieme, 2001.
- [te Brake 99] G. M. te Brake and N. Karssemeijer. *Single and Multiscale Detection of Masses in Digital Mammograms*. IEEE Transactions on Medical Imaging, vol. 18, no. 7, 1999.
- [Thurfjell 94] E.L. Thurfjell, K.A. Lernevall and A.A.S. Taube. *Benefit of Independent Double Reading in a Population-based Mammography Screening Program*. Radiology, vol. 191, pages 241–244, 1994.
- [Timp 04] S. Timp and N. Karssemeijer. *A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography*. Medical Physics, vol. 31, page 958, 2004.
- [Torre 86] V. Torre and T. Poggio. *On edge detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 2, pages 147–163, 1986.
- [Tourassi 03] G.D. Tourassi, R. Vargas-Voracek, D.M. Catarious Jr and C.E. Floyd Jr. *Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information*. Medical Physics, vol. 30, page 2123, 2003.
- [Travade 94] A. Travade, A. Isnard and H. Gimbergues. Imagerie de la pathologie mammaire. Masson, 1994.
- [Umanol 94] M. Umanol, H. Okamoto, I. Hatono, H. Tamura, F. Kawachi, S. Umedzu and J. Kinoshita. *Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems*. In Proceedings of the Third IEEE Conference on Fuzzy Systems. IEEE World Congress on Computational Intelligence., pages 2113–2118, 1994.
- [Unser 92] M. Unser, A. Aldouri and M. Eden. *On the Asymptotic Convergence of β -Spline Wavelets to Gabor Functions*. IEEE Transactions on Information Theory, vol. 38, no. 2, pages 864–872, 1992.
- [Unser 94] M. Unser. *Fast Gabor-Like Windowed Fourier and Continuous Wavelet Transforms*. IEEE Signal Processing Letters, vol. 1, no. 5, pages 76–79, 1994.

- [Utgoff 89] P.E. Utgoff. *Incremental Induction of Decision Trees*. Machine Learning, vol. 4, no. 2, pages 161–186, 1989.
- [Vachier 95] C. Vachier. *Extraction de Caractéristiques, Segmentation d'Image et Morphologie Mathématique*. PhD thesis, Ecole des Mines de Paris, 1995.
- [Vapnik 00] V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 2000.
- [Veldkamp 99] W.J.H. Veldkamp and N. Karssemeijer. *Improved method for detection of microcalcification clusters in digital mammograms*. In *Medical Imaging: Proceedings of the SPIE*, volume 3661, pages 512–522. SPIE, 1999.
- [Verma 01] B. Verma and J. Zakos. *A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques*. *IEEE Transactions on Information Technology in Biomedicine*, vol. 5, no. 1, pages 46–54, 2001.
- [Viton 96] J.L. Viton, M. Rasigni, G. Rasigni and A. Llebaria. *Method for characterizing masses in digital mammograms*. *Optical Engineering (Bellingham, Washington)*, vol. 35, no. 12, pages 3453–3459, 1996.
- [Vyborny 94] C.J. Vyborny and M.L. Giger. *Computer vision and artificial intelligence in mammography*. *American Journal of Roentgenology*, vol. 162, no. 3, pages 699–708, 1994.
- [Wei 97] D. Wei, H.-P. Chan, N. Petrick, B. Sahiner, M.A. Helvie, D.D. Adler and M.M. Goodsitt. *False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis*. *Medical Physics*, vol. 24, page 903, 1997.
- [Weszka 76] J. Weszka, CR Dyer and A. Rosenfeld. *A Comparative Study of Texture Measures for Terrain Classification*. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, pages 269–285, 1976.
- [Wheeler 06] F.W. Wheeler, A.G.A. Perera, B.E. Claus, S.L. Muller, G. Peters and J.P. Kaufhold. *Micro-calcification detection in digital tomosynthesis mammography*. In *Medical Imaging: Proceedings of the SPIE*, volume 6144, pages 671–682, 2006.
- [Wilson 97] M. Wilson, S. Mitra, G.H. Roberson and Y.Y. Shieh. *Automated microcalcification detection in mammograms using statistical variable-box-threshold filter method*. In *Medical Imaging: Proceedings of the SPIE*, volume 3165, page 195. SPIE, 1997.
- [Wolfe 76] JN Wolfe. *Risk for breast cancer development determined by mammographic parenchymal pattern*. *Cancer*, vol. 37, no. 5, pages 2486–92, 1976.
- [Woods 93] K. Woods, C. Doss, K. Bowyer, J. Solka, C. Priebe and P. Kegelmeyer. *Comparative Evaluation of Pattern Recognition Techniques for Detection of Microcalcifications in Mammography*. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 6, pages 1417–1436, 1993.
- [Wu 03a] T. Wu, R. Moore, D. Kopans, A. Stewart, W. Phillips, M. Stanton, J. Eberhard, B. Opsahl-Ong, L. Niklason and M. Williams. *Tomosynthesis mammography reconstruction using a maximum likelihood method*. *Medical Physics*, vol. 30, no. 6, page 1332, 2003.

- [Wu 03b] T. Wu, A. Steward, M. Stanton, T. McCauley, W. Phillips, D. B. Kopans, R. H. Moore, J. W. Eberhard, B. Opsahl-Ong, L. Niklason and M. B. Williams. *Tomographic Mammography Using a Limited Number of Low-Dose Cone-Beam Projection Images*. Medical Physics, vol. 30, no. 3, pages 365–380, 2003.
- [Wu 04a] T. Wu, R. H. Moore, E. A. Rafferty and D. B. Kopans. *A Comparison of Reconstruction Algorithms for Breast Tomosynthesis*. Medical Physics, vol. 31, no. 9, pages 115–126, 2004.
- [Wu 04b] T. Wu, J. Zhang, R. Moore, E. Rafferty, D. Kopans, W. Meleis and D. Kaeli. *Digital tomosynthesis mammography using a parallel maximum-likelihood reconstruction method*. In Medical Imaging: Proceedings of the SPIE, volume 5368, pages 1–11, 2004.
- [Xu 97] C. Xu and J.L. Prince. *Gradient Vector Flow: A New External Force for Snakes*. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pages 66–71, 1997.
- [Xu 98] C. Xu and J.L. Prince. *Generalized gradient vector flow external forces for active contours*. Signal Processing, vol. 71, no. 2, pages 131–139, 1998.
- [Yoshida 96] H. Yoshida, K. Doi, RM Nishikawa, ML Giger and RA Schmidt. *An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms*. Academic Radiology, vol. 3, no. 8, pages 621–7, 1996.
- [Yuan 95] Y. Yuan and M.J. Shaw. *Induction of fuzzy decision trees*. Fuzzy Sets and Systems, vol. 69, no. 2, pages 125–139, 1995.
- [Zadeh 65] L. A. Zadeh. *Fuzzy Sets*. Information and Control, vol. 8, no. 3, pages 338–353, 1965.
- [Zadeh 68] L.A. Zadeh. *Probability measures of fuzzy events*. Journal of Mathematical Analysis and Applications, vol. 23, no. 10, pages 421–427, 1968.
- [Zadeh 01] HS Zadeh, SP Nezhad and FR Rad. *Shape-based and texture-based feature extraction for classification of microcalcification in mammograms*. In Medical Imaging: Proceedings of the SPIE, volume 4322, pages 301–310, 2001.
- [Zhang 06] Y. Zhang, H.-P. Chan, B. Sahiner, J. Wei, M.M. Goodsitt, L.M. Hadjiiski, J. Ge and C. Zhou. *Tomosynthesis reconstruction using the simultaneous algebraic reconstruction technique (SART) on breast phantom data*. In Medical Imaging: Proceedings of the SPIE, volume 6142, page 614249. SPIE, 2006.
- [Zhao 92] D. Zhao, M. Shridhar and D.G. Daut. *Morphology on detection of calcifications in mammograms*. In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), volume 3, pages 129–132, 1992.
- [Zheng 99] B. Zheng, Y.-H. Chang, X.-H. Wang and W.F. Good. *Comparison of Artificial Neural Network and Bayesian Belief Network in a Computer-Assisted Diagnosis Scheme for Mammography*. In IEEE International Joint Conference on Neural Networks (IJCNN), volume 6, pages 4181–4185, 1999.

- [Zheng 01] L. Zheng and A.K. Chan. *An Artificial Intelligent Algorithm for Tumor Detection in Screening Mammogram*. IEEE Transactions on Medical Imaging, vol. 20, no. 7, pages 559–567, 2001.
- [Zheng 06] J. Zheng. *CAD System for Detecting Clustered Microcalcifications in Digital Mammograms using Independent Component Analysis and Neural Network*. International Journal of Computer Assisted Radiology and Surgery, vol. 1, pages 517–527, 2006.
- [Zhou 89] XH Zhou and R. Gordon. *Detection of early breast cancer: an overview and future prospects*. Crit Rev Biomed Eng, vol. 17, no. 3, pages 203–55, 1989.
- [Zhu 96] S.C. Zhu and A. Yuille. *Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 9, pages 884–900, 1996.
- [Zhu 07] W. Zhu and T. Chan. *A Variational Model for Capturing Illusory Contours Using Curvature*. Journal of Mathematical Imaging and Vision, vol. 27, no. 1, pages 29–40, 2007.
- [Zighed 00] D.A. Zighed and R. Rakotomalala. *Graphes d'induction: apprentissage et data mining*. Hermès Science Publications, 2000.
- [Zijdenbos 94] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin and A.C. Palmer. *Morphometric analysis of white matter lesions in MR images: method and validation*. IEEE Transactions on Medical Imaging, vol. 13, no. 4, pages 716–724, 1994.
- [Zimmermann 01] H.J. Zimmermann. *Fuzzy Set Theory-And Its Applications*. Kluwer Academic Publishers, 2001.