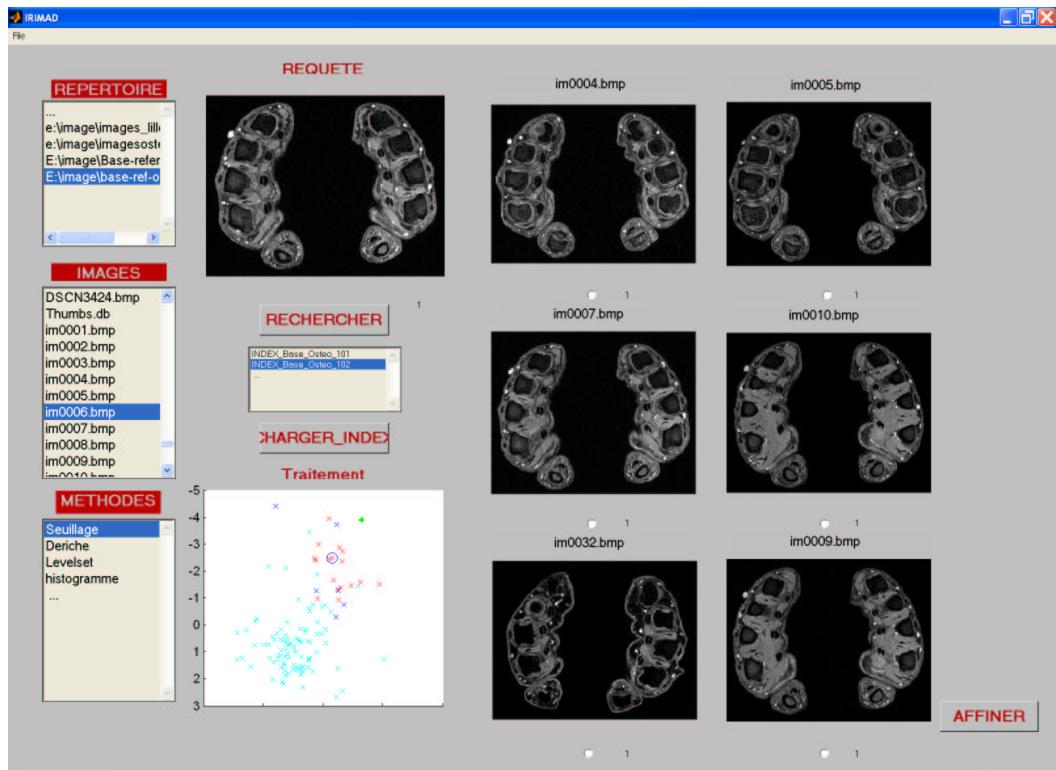


Modèles de mélange pour la recherche d'images par le contenu : applications aux pathologies ostéo-articulaires

Micheline NAJJAR— Thèse de doctorat



Soutenance le 18 Mai 2004 devant le jury composé de :

Mesdames Blanc-Feraud L. (Rapporteur)
 Guérin-Dugué A. (Rapporteur)
 Cotten A.
Messieurs Pauwels E.
 Cocquerez J-P. (Directeur de thèse)
 Ambroise C. (Directeur de thèse)

Laboratoire Heudiasyc UMR CNRS 6599
Université de Technologie de Compiègne

MODÈLES DE MÉLANGE POUR LA RECHERCHE D'IMAGES PAR LE CONTENU : APPLICATIONS AUX PATHOLOGIES OSTÉO-ARTICULAIRES

RÉSUMÉ

L'indexation et la recherche d'images par le contenu (CBIR) demeurent des problèmes importants qui trouvent des applications potentielles en médecine, étant donné la place que l'image numérique médicale occupe à présent dans le quotidien des hôpitaux.

Cette thèse présente un système original d'indexation et de recherche d'images médicales pour une application aux images IRM des mains. Il s'adresse au service de Radiologie Ostéo-Articulaire à l'hôpital Roger Salengro du CHRU de Lille. De tels systèmes existent aujourd'hui quand les images sont annotées par des mots clés. La recherche est plus complexe lorsque la requête s'effectue uniquement à partir des images.

La mise en œuvre d'un algorithme performant de segmentation automatique basée sur les "ensembles de niveau" permet d'obtenir des représentations pertinentes de l'image. Un index basé sur les attributs des régions est ainsi fabriqué pour coder les connaissances des médecins afin de diagnostiquer une image.

Une nouvelle approche, appelée **EMiner**, de recherche par boucle de pertinence est introduite. Elle est basée sur une classification semi-supervisée de la base des index utilisant les modèles de mélange et l'algorithme EM. Une nouvelle technique de sélection de variables est associée pour améliorer les résultats de la recherche des images similaires.

La qualité des résultats obtenus pour la recherche d'images pathologiques et non pathologiques valide le système proposé aussi évalué par l'équipe médicale du service.

MOTS-CLÉS :

Fouille d'images par le contenu
Images médicales
Classification semi-supervisée
Modèle de mélange
Algorithme EM
Sélection de variables
Segmentation d'image
Théorie des ensembles de niveau

MIXTURE MODELS FOR CONTENT BASED IMAGE RETRIEVAL: OSTEO-ARTICULAR MR IMAGING APPLICATION

ABSTRACT

The development of computer power and the rise of information technologies have strongly supported the numerical medical image production attached to various methods like the MR, CT Imaging, etc. These images are stored in huge databases. Their processing requires the development of effective procedures making it possible to structure them, and to retrieve a given image among structured data. Our work is focused on the study of a content-based indexing and retrieval (CBIR) system for osteo-articular MR imaging application, aiming at intervening in the students and doctors diagnosis help.

Nowadays, such systems exist when images are labelled by keywords. Research is however more difficult when based only on the image contents. The state of the art of a powerful algorithm for automatic segmentation based on the "Level Set Theory" makes it possible to obtain relevant representations of the image. An index based on region features is calculated.

An original approach, called **EMiner**, for searching by similarity is introduced. It is based on an index database classification using mixture models and the EM algorithm. A new feature selection method is developed, in the context of semi-supervised learning, to ameliorate searching results.

The system presented is evaluated and validated using a medical image database labelled by an expert. It is now implemented in the Osteo-articular radiology service of Lille CHRU hospital. A testing phase, performed by physician, will take place in the near future to evaluate the system subjective efficiency.

KEYWORDS:

Content based Image Retrieval
Medical Imaging
Mixture Model
EM Algorithm
Semi-Supervised Learning
Feature Selection
Image Segmentation
Levelset Theory

La vie est une perle, conservons-la
précieusement
La vie est un doux rêve, réalisons-le
entièrement
La vie est la vie, défendons-la inlassablement

A mes parents,

à ma nièce Joy Mary,

Remerciements

Je tiens tout d'abord à exprimer ma gratitude à mon directeur de thèse Jean-Pierre Cocquerez. Tout au long de ma thèse, il m'a consacré une grande disponibilité, pour discuter des questions scientifiques comme pour m'aider dans les aspects pratiques. Il a su m'épauler, aux moments opportuns, pour ses conseils, ses réflexions et sa patience dans l'avancement de ce travail. Nos discussions m'ont donné le plaisir de comprendre, chercher et formuler des idées nouvelles et concrètes.

Mes vifs remerciement à mon co-directeur de thèse Christophe Ambroise. Il m'a transmis ses connaissances avec beaucoup de patience tout au long de mon travail, et m'a apporté une aide déterminante pour bien accomplir la thèse. Je le remercie beaucoup pour l'étroite collaboration que nous avons eu depuis mon arrivée à Heudiasyc. Son soutien, sa disponibilité et ses qualités pédagogiques et humaines sont références.

Je suis reconnaissante envers Anne Guérin-Dugué et Laure Blanc-Feraud, pour avoir accepté de rapporter sur mon travail de thèse et pour leur contribution décisive au manuscrit final.

Je remercie vivement Eric Pauwels, d'avoir bien voulu participer au jury de thèse, malgré une soutenance en langue française.

Un grand merci à Anne Cotten, présidente du jury, pour toute l'aide qu'elle m'a apporté pendant 3 ans pour définir l'application médicale de mon travail de recherche.

Je suis particulièrement reconnaissante à Nathalie Boutry, Patrick Potteau pour la création de la base d'images médicales expertisées.

Je remercie spécialement Bernard Debuisson, directeur de la recherche à L'UTC, pour le soutien logistique qu'il m'a accordé au début de la thèse.

Je remercie également Ali Charara, professeur à l'UTC, pour son enthousiasme et sa disponibilité à créer des collaborations entre l'UTC et l'Université Libanaise. C'est grâce à ses efforts que j'ai eu l'occasion de préparer ma thèse au Laboratoire Heudiasyc.

Je remercie tout particulièrement Rogelio Lozano, Directeur du Laboratoire Heudiasyc, pour m'avoir permis d'effectuer mon travail au sein du laboratoire dans les meilleures conditions. Ses qualités scientifiques et humaines sont exemplaires.

Je remercie Magali Collignon, Nathalie Alexandre, Nathalie Laboureur et Dominique Porras pour leur gentillesse et leur diligence face à toutes mes requêtes. Je remercie aussi l'équipe informatique du laboratoire Heudiasyc, en particulier Corinne Boscolo, Véronique Moisan, Paul Crubilié, David Zowolinski et Jean-Claude Escande, pour leur aide efficace et leur amabilité.

Je tiens à remercier chaleureusement mes amis et mes collègues, pour m'avoir aidé et encouragé au cours de ma thèse dans les bons et moins bons moments. Je pense en particulier à Pedro Castillo, Alexandro Dzul, Benoit Fouconnier, Pascal Derache, Aurélien Flipo, Bouchra Abboud, Wala Touhami, Rabih Boueri, Marguerita Eid, Imad Sallit, Georges Abou Jaoudé et Elie Bou Chakra. Aux membres de mon équipe de Basket, Rogelio Lozano, Hugo Romero, Anand Sanchez, Dominique Gibrael, je vous remercie beaucoup des agréables moments qu'on a passé ensemble.

Éternelle reconnaissance à ma famille. Ma mère Salma, c'est grâce à toi que j'ai réalisé

mes rêves en veillant sans cesse sur moi *en silence*. Mon frère Elie, je n'oublie pas ton aide ni mesurable ni quantifiable pour aller jusqu'au bout dans mes études. Ma grande soeur Ibtissam, tu es mon ange gardien, je t'aime beaucoup. A ma petite soeur Sanaa, un grand merci d'avoir m'accompagné pendant deux ans de ma thèse, tu étais très proche de moi dans les plus durs moments, en me préparant à manger surtout.

Je n'oublie pas mon beau frère Tony Chahwan, ma nièce Joy Mary, tu m'as appris la joie de la vie. Un grand merci à mon oncle Nicolas, qui m'a accompagné tout au long de mes chemins à Roumieh.

Mon amour éternelle à mon père qui est parti beaucoup trop tôt...

De tout mon coeur, merci mon Dieu...

Table des matières

| | | |
|----------|---|-----------|
| 1 | Introduction générale | 1 |
| 2 | La recherche d'images par le contenu en médecine | 5 |
| 2.1 | Les bases d'images médicales en médecine | 5 |
| 2.2 | Recherche dans une base d'images : généralités | 6 |
| 2.3 | Systèmes de recherche d'images par le contenu | 7 |
| 2.3.1 | Principe | 7 |
| 2.3.2 | Objectifs de la recherche d'images par le contenu | 7 |
| 2.3.3 | Impact de la nature du domaine d'application de l'imagerie médicale | 8 |
| 2.3.4 | Fossé sémantique | 10 |
| 2.3.5 | Applications potentielles de la recherche d'images par le contenu en médecine | 10 |
| 2.4 | Description du contenu des images | 11 |
| 2.4.1 | Attributs | 12 |
| 2.4.2 | Points d'intérêt | 15 |
| 2.4.3 | Segmentation | 16 |
| 2.5 | Indexation | 17 |
| 2.6 | Similarité | 18 |
| 2.7 | Interactivité | 18 |
| 2.8 | Présentation de quelques systèmes de fouilles d'images médicales | 19 |
| 2.8.1 | Système IGDS | 19 |
| 2.8.2 | Système ASSERT | 20 |
| 2.8.3 | Système IRMA | 22 |
| 2.8.4 | Système à base de connaissance pour la recherche d'images médicales | 24 |
| 2.8.5 | Synthèse | 25 |
| 2.9 | Notre système, nos choix | 26 |
| 3 | Etat de l'art de la recherche par boucle de pertinence | 27 |
| 3.1 | Introduction | 27 |
| 3.2 | Fonctionnement des systèmes de recherche interactive | 27 |
| 3.2.1 | Termes et notations | 28 |
| 3.3 | Approches géométriques de bouclage de pertinence | 29 |

| | | |
|----------|---|-----------|
| 3.3.1 | Méthodes heuristiques | 30 |
| 3.3.2 | Méthodes optimales | 30 |
| 3.4 | Approches statistiques de bouclage de pertinence | 34 |
| 3.4.1 | Système PicHunter : Modélisation bayésienne de pertinence | 34 |
| 3.4.2 | Modélisation non-paramétrique de pertinence | 35 |
| 3.4.3 | Approche par inférence bayésienne | 37 |
| 3.4.4 | Approches par classification | 38 |
| 3.4.5 | Approche par une classification semi-supervisée | 41 |
| 3.5 | Conclusion | 41 |
| 4 | Apprentissage semi-supervisé | 43 |
| 4.1 | Notations et terminologies | 44 |
| 4.2 | Classification semi-supervisée | 46 |
| 4.2.1 | Approches géométriques de la semi-supervision | 47 |
| 4.2.2 | Machines à vecteurs de support semi-supervisées | 48 |
| 4.2.3 | Approches paramétriques de la semi-supervision | 48 |
| 4.3 | Apprentissage avec les modèles de mélange et l'algorithme EM | 52 |
| 4.3.1 | Hypothèses | 52 |
| 4.3.2 | Algorithme EM | 53 |
| 4.4 | Comparaison entre classification supervisée et semi-supervisée | 54 |
| 4.4.1 | Données synthétiques | 56 |
| 4.4.2 | Base d'images ANN | 59 |
| 4.5 | Conclusion | 63 |
| 5 | EMiner pour la recherche interactive d'images | 65 |
| 5.1 | EMiner pour la recherche interactive d'images | 65 |
| 5.1.1 | Description de la recherche interactive | 67 |
| 5.2 | Evaluation des performances | 67 |
| 5.2.1 | Protocole d'évaluation | 68 |
| 5.2.2 | Résultats de recherche par EMiner | 70 |
| 5.2.3 | Comparaison entre EMiner et SVM | 72 |
| 5.3 | Sélection de variables | 83 |
| 5.3.1 | Sélection de variables dédiée pour la classification supervisée | 84 |
| 5.3.2 | Sélection de variables pour une classification non supervisée | 87 |
| 5.4 | Notre approche | 88 |
| 5.4.1 | Critère | 90 |
| 5.4.2 | Algorithme | 90 |
| 5.4.3 | Remarque 1 | 92 |
| 5.4.4 | Remarque 2 | 93 |
| 5.4.5 | Expérimentations | 93 |
| 5.5 | Conclusions | 97 |

| | |
|---|------------|
| 6 Application aux pathologies ostéo-articulaires | 99 |
| 6.1 Introduction | 99 |
| 6.2 Description de l'application | 100 |
| 6.2.1 Contenu de la base | 100 |
| 6.2.2 Contenu des images : indices visuels | 100 |
| 6.3 Présentation du système | 103 |
| 6.4 Segmentation | 105 |
| 6.4.1 Prétraitement | 105 |
| 6.4.2 Détection du contour par l'utilisation de la théorie des ensembles de niveaux | 108 |
| 6.5 Extraction des caractéristiques | 111 |
| 6.5.1 Attributs de forme: les moments invariants de Hu | 113 |
| 6.5.2 Attributs de texture utilisant les masques de Laws | 117 |
| 6.6 Indexation | 118 |
| 6.6.1 Calcul de l'index | 119 |
| 6.6.2 Classification non supervisée des régions | 120 |
| 6.7 Expérimentations et validation réelle | 121 |
| 6.7.1 Protocole d'évaluation | 122 |
| 6.7.2 Résultats de recherche | 124 |
| 6.7.3 Résultats d'évaluation | 125 |
| 6.7.4 Recherche par sélection de variables | 134 |
| 6.8 Bilan | 137 |
| 7 Conclusion générale | 139 |
| 7.1 Récapitulatif du travail | 139 |
| 7.2 Perspectives d'applications et de recherche | 141 |
| A Algorithme EM et modèles de mélange | 145 |
| A.1 Hypothèses | 145 |
| A.2 Algorithme EM | 146 |
| A.3 Apprentissage non-supervisé | 146 |
| A.3.1 Cas d'un mélange gaussien | 147 |
| A.4 Prise en compte de labels complètement connus | 148 |
| A.5 Prise en compte de labels partiellement connus | 148 |
| A.6 Formules finales | 150 |
| A.6.1 Modèle de mélange gaussien | 150 |
| B Les Machines à vecteurs de support | 153 |
| B.1 Détermination d'une frontière linéaire | 154 |
| B.1.1 Cas de deux classes séparables par un hyperplan | 154 |
| B.1.2 Cas de deux classes non séparables linéairement | 155 |
| B.2 Détermination d'une frontière non linéaire | 156 |

| | |
|---|------------|
| C Théorie des courbes de niveaux | 157 |
| C.1 Evolution d'une courbe dans le plan | 157 |
| C.1.1 Formulation Lagrangienne | 157 |
| C.2 Théorie des courbes de niveaux "Level set" | 159 |
| C.2.1 Définitions | 159 |
| C.2.2 Equation d'évolution | 159 |
| C.2.3 Avantages de la méthode | 161 |
| C.3 Schémas numériques | 162 |
| C.3.1 Equation d'Hamilton-Jacobi | 162 |
| C.3.2 Schéma monodimensionnel | 162 |
| C.3.3 Schéma bidimensionnel | 163 |
| C.4 Front d'onde pour obtenir les contours d'une image | 164 |
| C.4.1 Extension de k_i | 164 |
| C.4.2 Elaboration de l'extension de k_i | 165 |
| C.5 Algorithmes de résolution de l'équation d'évolution | 165 |
| C.5.1 La méthode Narrow Band | 166 |
| C.5.2 La méthode Fast Marching | 167 |
| C.5.3 La méthode Hermes | 167 |
| C.6 Résultats | 169 |
| Bibliographie | 171 |

Chapitre 1

Introduction générale

Cadre du travail

Cette thèse a été préparée dans le cadre du projet de recherche "Système de communication pour la formation en médecine basé sur l'indexation d'images" inscrit au contrat de plan Etat Région Picardie pour la période 2000-2006. Le but est de produire un **démonstrateur d'un système interactif** reliant bases de données d'images médicales et utilisateurs (médecins et étudiants) à des fins de formation et d'aide au diagnostic.

Ce projet correspond à une attente des médecins qui a été identifiée par l'Observatoire Régional de la Santé (ORS) de Picardie. En effet, une immense quantité d'images existe en stock sous forme de papier argentique, ou est produite régulièrement par les appareils récents d'imagerie médicale. La gestion de ces archives, très variable suivant les services cliniques, est souvent liée au degré de pénétration des technologies de l'information et de la communication dans l'hôpital. Cela peut aller de l'empilement des clichés argentiques au serveur d'images, en passant par l'étagère de CD ROM. Dans tous les cas, ces gisements sont largement sous exploités.

L'activité des médecins, en particulier les radiologues, est fortement basée sur la recherche d'indices visuels dans les images. Ainsi notre projet s'appuie sur la recherche d'images par le contenu et plus particulièrement sur la recherche dans des bases d'images identiques à une image requête. Les applications qui en découlent sont de plusieurs natures :

- si l'image requête est typique d'une catégorie, l'identification des images semblables à celle-ci permet à organiser les bases d'images (classification),
- si la base contient des images expertisées, un étudiant peut formuler un diagnostic sur une image requête exercice et le système de fouille d'images renverra des images semblables. L'étudiant pourra alors vérifier la validité de son diagnostic et affiner sa connaissance grâce aux commentaires attachés aux images renvoyées,
- si la base contient un ensemble de dossiers de patients contenant des images, le médecin peut rechercher tous les cas similaires à partir d'une image requête et améliorer

ainsi son suivi thérapeutique.

Les partenaires du projet sont :

- les laboratoires HEUDIASYC (UMR 6589) et BMGBM (UMR 6600) de l'Université de Technologie de Compiègne,
- le Centre Hospitalier Régional Universitaire (CHRU) de Lille,
- le Centre Hospitalier Universitaire d'Amiens,
- les entreprises Planet HC, (spécialisée dans l'édition de solutions informatiques pour les professionnels de la santé) et Archimed (spécialisée dans la conception et l'intégration de systèmes d'information multimédia),
- l'Observatoire Régional de la Santé de Picardie.

Ce projet a vocation de se rattacher au Consortium UMVF (Université Médicale Virtuelle Francophone) correspondant à un projet du RNTS (Réseau National de Technologie de la Santé).

Problématique

Notre application médicale a pour mission de mettre en œuvre, au service de Radiologie Ostéo-Articulaire à l'hôpital Roger Salengro du CHRU de Lille, un système de recherche par le contenu des IRM (Image à Résonance Magnétique) des mains. L'importance de ce type d'images réside dans la détection précoce de la polyarthrite rhumatoïde. Cette pathologie est un rhumatisme inflammatoire susceptible d'entraîner des destructions articulaires très importantes mais dont le pronostic précoce est rendu fondamental par le développement de nouveaux traitements susceptibles de stopper l'évolution de la maladie.

Ainsi, le but de notre travail consiste à aider le personnel dans ce service d'organiser le stockage des images numériques, produites d'une façon continue et extensive par les caméras IRM, afin d'y accéder d'une façon "intelligente".

Le seul moyen rapide actuel d'accéder à un dossier patient est la fouille dans des films archivés dans les armoires. Les données numériques sont mal organisées. Le médecin est obligé de fouiller dans des CDs sauvegardés par le technicien du service (voir figure 1.1).

Organisation du mémoire

A part la structuration des bases d'images médicales qui est indispensable, la recherche d'images par le contenu en médecine trouve son intérêt dans plusieurs applications. Dans le chapitre 2 de ce mémoire, nous analysons ce thème de recherche. Après une étude bibliographique des systèmes médicaux existants, nous décrivons l'architecture de notre système en justifiant nos choix. La potentialité principale du système consiste à permettre l'interaction répétée de l'utilisateur au cœur du système, afin d'intégrer ses connaissances dans la procédure de recherche dans le but de répondre à ses besoins. Le système est donc caractérisé par la présence d'une boucle de pertinence.

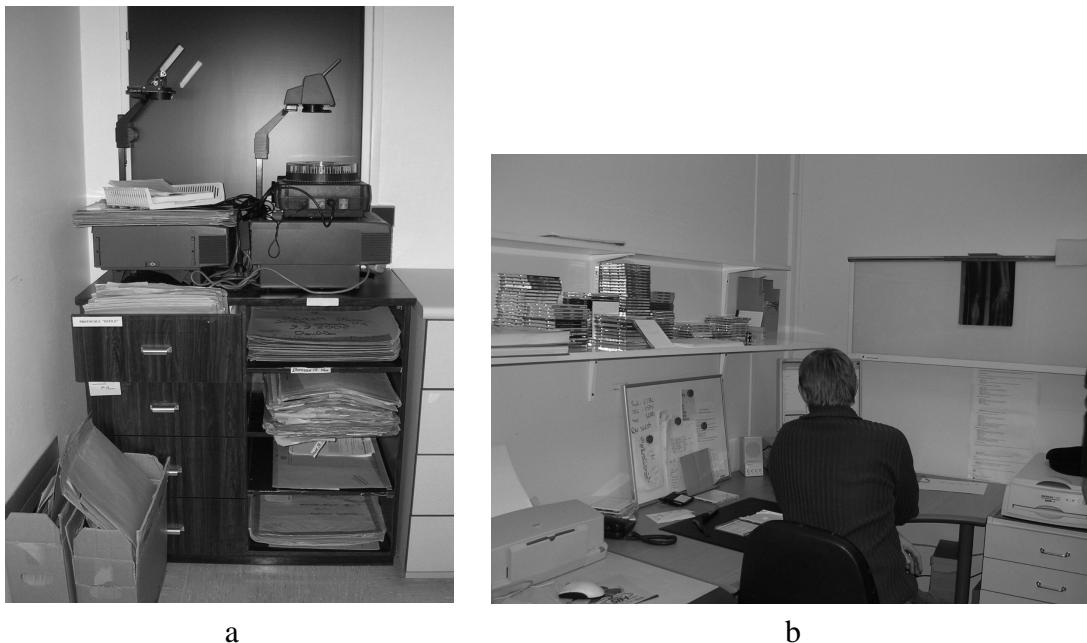


FIG. 1.1 – Archivage actuel des dossiers patients : a- des films archivés dans les armoires et b- des CDs sur des étagères. Photos prises avec la permission du responsable du service.

Ce choix est effectué pour accomplir la mission demandée au système : aider les médecins spécialisés à accéder aux bases d’images intelligemment pour chercher des cas présentant des symptômes similaires, et aider à la formation des étudiants en médecine.

Dans le chapitre 3 nous dressons un état de l’art des méthodes de recherche par boucle de pertinence. Nous distinguons entre les approches géométriques et les approches statistiques.

La méthode de recherche interactive, que nous proposons, est basée sur une approche statistique. C’est une approche d’apprentissage semi-supervisé qui a pour objectif de séparer l’espace des index des images en deux classes : images pertinentes et images non pertinentes. L’ensemble d’apprentissage est constitué des images annotées, pertinentes et non pertinentes, par l’utilisateur et celle non annotées de la base. Notre approche semi-supervisée est basée sur les modèles de mélange et l’algorithme EM. L’apport de l’apprentissage semi-supervisé par rapport à l’apprentissage supervisé est prouvé dans le chapitre 4.

Dans le chapitre 5, nous présentons notre algorithme de recherche par boucle de pertinence nommé **EMiner**. Nous comparons ensuite **EMiner** à une autre technique basée sur les machines à vecteurs de support. Nous montrons que notre approche est efficace pour chercher des catégories d’images multimodales. Dans ce chapitre 5, nous introduisons aussi une nouvelle méthode de sélection de variables dédiée à l’apprentissage semi-supervisé dans le but d’améliorer les résultats de la recherche interactive.

Le chapitre 6 est consacré à l’application médicale. Nous présentons d’abord la méthode d’indexation proposée pour fabriquer les index des images. Cette représentation

d'images est basée sur les indices visuels sur lesquels se base le médecin pour diagnostiquer une image. Ensuite, nous évaluons notre système d'indexation et de recherche interactive par l'intermédiaire d'une base d'images étiquetée créée par les médecins et l'informaticien du service.

Nous concluons enfin en récapitulant les principaux points du travail effectué et en évoquant les perspectives d'applications et de recherche.

Chapitre 2

La recherche d'images par le contenu en médecine

Ce chapitre présente le cadre général des systèmes de *recherche d'images par le contenu* (CBIR : Content Based Image Retrieval) dédiés aux applications médicales.

Le principe des systèmes de CBIR est expliqué afin de dégager quelques aspects importants dans le domaine de l'imagerie médicale. Nous dressons ensuite un état de l'art des systèmes médicaux existants. Enfin nous justifions le choix de notre système destiné aux images ostéo-articulaires.

2.1 Les bases d'images médicales en médecine

Le développement exponentiel de la puissance de calcul des ordinateurs et des capacités de stockage a favorisé fortement la production d'images médicales numériques bidimensionnelles ou volumiques attachées à diverses modalités comme l'IRM, l'endoscopie, la scintigraphie, etc. Les images produites, fixes ou vidéo sont stockées sous différentes formes qui peuvent aller de la base de données structurées à l'étagère de CD ROM. Plusieurs problèmes se posent alors :

1. la structuration des bases d'images (organisation, indexation),
2. la recherche d'une image particulière (fouille),
3. la compression et le transfert des images entre utilisateurs (codage, sécurisation).

Le premier problème n'est pas nécessairement réglé dans les services hospitaliers et concerne davantage l'archivage des images. L'accès à ces images via les réseaux de télécommunication est de qualité très variable. Cependant, il existe des systèmes d'archivage et d'échange d'images médicales opérationnels (Huang et Taira, 1992). Le troisième problème est crucial quand les images sont de grandes dimensions. Le comportement des méthodes de fouille d'images sur des données compressées fait l'objet de travaux de recherche (Jiang et al., 2002).

Dans ce mémoire, nous n'aborderons que le second problème ci-dessus.

2.2 Recherche dans une base d'images : généralités

Il est possible d'accéder de plusieurs manières aux images d'une base. Distinguons les méthodes suivantes par ordre de difficultés décroissantes :

1. interrogation en langage naturel,
2. interrogation à l'aide d'image(s) "requête",
3. interrogation à l'aide de mots-clés.

L'interrogation à l'aide de mots-clés est répandue, les systèmes n°1 opérationnels sont souvent basés sur des moteurs de recherche standards mais qui nécessitent l'emploi d'un langage unifié comme UMLS (Unified Medical Language System).

L'interrogation en langage naturel demeure un thème de recherche en communication "homme-machine". Si des systèmes opérationnels existent pour des domaines où le vocabulaire est restreint, il est clair qu'en médecine, de tels systèmes ne sont envisageables qu'à moyen terme sachant que le vocabulaire est complexe et que la sémantique des images recouvre une connaissance que les médecins acquièrent après plusieurs années d'étude. Cependant, nous présentons dans ce chapitre (paragraphe 2.8.4) un système à base de connaissance pour la recherche d'images médicales qui constitue un exemple d'étape intermédiaire vers l'interrogation en langage naturel.

Entre les deux modes d'interrogation précédents se situe l'utilisation d'images requête. L'idée de base est que l'image en elle même contient les informations pertinentes. Ainsi, plutôt que de décrire l'image pour formuler la requête, celle-ci est envoyée directement au système. C'est le principe des systèmes de recherche d'images par le contenu dont l'objectif principal est de retrouver une image semblable visuellement à une image requête.

Cette présentation est séduisante mais elle masque une difficulté majeure provenant de la polysémie d'une image. L'image requête possède des sens différents selon les utilisateurs, voire pour un même utilisateur. Sa signification peut varier suivant le contexte propre de la perception. Ainsi, d'une manière très générale, une image peut :

1. représenter la réalité : un phénomène physique, un objet, une scène, un organe, une zone pathologique, etc....,
2. évoquer un souvenir, un sentiment, une maladie,
3. illustrer un concept complexe, par exemple : la grève, la souffrance.

Comment cette polysémie est elle traitée dans les systèmes d'analyse d'images en vision par ordinateur ou dans le système visuel humain ?

Que ce soit dans l'un ou l'autre, l'image captée n'est pas utilisée dans son état brut, mais subit un traitement qui élabore une *représentation*. Cette dernière peut être considérée comme un codage plus ou moins sophistiqué de l'information pertinente contenue dans l'image. On conçoit même l'existence de plusieurs représentations. Ainsi, dans le système visuel humain, le contexte de l'observateur conditionne la perception et l'interprétation : c'est le mécanisme de *probabilisation* dont on peut considérer que le principe est de favoriser une représentation mentale donnée. De même, dans un système informatique, comme

il est inconcevable d'avoir une représentation universelle, il importe de disposer, dans les applications complexes, de plusieurs représentations issues de différentes modélisations de l'univers sur lequel on travaille.

Avec les images médicales, le problème de polysémie se pose d'un point de vue pratique, essentiellement dans le cas n° 1 ci dessus. Même dans ce cadre plus réduit, une radiographie d'un organe peut donner lieu à plusieurs interprétations en l'absence de contexte clinique.

Dans ce contexte, il est légitime de s'interroger sur l'architecture d'un système de fouilles d'images acceptant des images requêtes provenant des différentes modalités et concernant différentes pathologies et organes. Ceci constitue un thème de recherche en soi. Cependant, on peut admettre qu'indépendamment de l'architecture, il y aura un problème de sélection de modèles pour élaborer la représentation adaptée à ce que recherche l'utilisateur.

Les systèmes de recherche d'images par le contenu font l'objet de nombreux travaux de recherche dont on peut trouver un inventaire dans (VeltKamp et Tanase, 2001). Les travaux de cette thèse se situe dans ce cadre.

2.3 Systèmes de recherche d'images par le contenu

2.3.1 Principe

Comme nous l'avons indiqué précédemment, la recherche d'image par le contenu nécessite l'élaboration d'une représentation de l'image ou d'une signature (appelée index) contenant l'information pertinente. La structure générale d'un système de recherche d'image par le contenu est fournie figure 2.1.

Le système comprend un bloc "hors ligne" et "en ligne". Le bloc "hors ligne" calcule tous les index de la base d'images et construit une base d'index. Au passage, signalons que, dans notre travail, nous n'aborderons pas le problème d'organisation des bases d'images et d'index et du fonctionnement des requêtes. Pendant le fonctionnement "en ligne", l'utilisateur fournit une image requête, le système calcule l'index, le confronte à la base d'index et affiche un nombre donné d'images d'index voisins de l'index requête. L'utilisateur annote les images retournées comme pertinentes ou non. Le système modifie son comportement en fonction des réponses et fait une nouvelle proposition.

En reprenant une partie de l'analyse faite par (Smeulders et al., 2000), nous pouvons dégager quelques aspects importants des systèmes de la recherche d'images par le contenu présentés dans les sous sections suivantes.

2.3.2 Objectifs de la recherche d'images par le contenu

On distingue plusieurs types d'objectifs :

- Recherche d'une image spécifique.

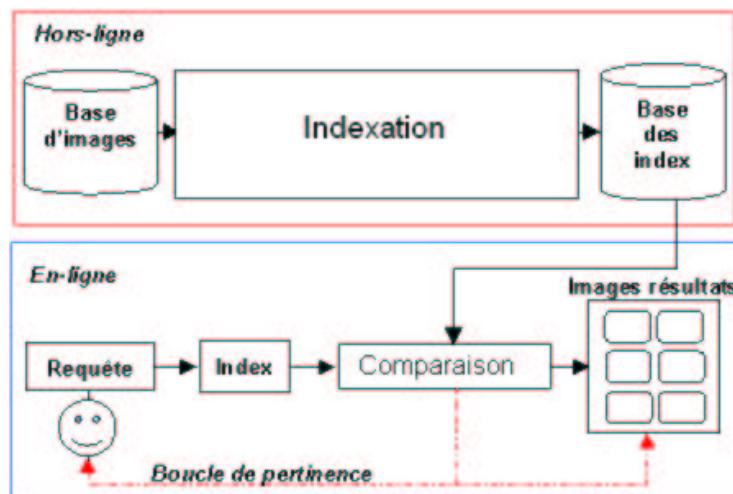


FIG. 2.1 – Schéma général d'un système de recherche d'images basée sur le contenu.

L'utilisateur recherche une image précise dans la base. L'image cible est identifiée par une présentation interactive au système d'une suite d'exemples (Cox et al., 2000). L'objectif peut être aussi de retrouver l'image d'un objet différent de celle que possède l'utilisateur.

- Recherche d'une image semblable à une requête.
La requête peut être une image ou une caricature de ce que recherche l'utilisateur.
- Construction de catégories.
Le principe d'interrogation du système est le même que le précédent, mais l'objectif est de structurer une base ou de l'étiqueter.

2.3.3 Impact de la nature du domaine d'application de l'imagerie médicale

Le domaine d'application intervient essentiellement par :

- son étendue,
- le protocole d'acquisition des images,
- le contexte patient,
- la sémiologie de la pathologie.

Etendue du domaine d'application

Une image médicale peut représenter un organe, un ensemble d'organes, une coupe du corps, etc. Si les bases d'images contiennent des images acquises suivant différentes modalités et concernent différentes pathologies, elles sont de grande dimension, le domaine concerné est donc très vaste, il y a une grande variété dans l'apparence et le contenu de l'image. Les images risquent fort d'être polysémiques.

Dans le cadre de cette thèse, les images sont des images IRM d'une pathologie ostéo-articulaire : la polyarthrite rhumatoïde. Le domaine est restreint, la polysémie des images est réduite. Que le domaine soit restreint ou pas, les images s'adressent à des spécialistes. Ces derniers doivent avoir une connaissance fine de la sémiologie des pathologies pour comparer des images.

On peut résumer l'impact de la nature du domaine sur la fouille d'images par le tableau 2.2 dérivé de celui de (Smeulders et al., 2000) mais adapté à l'imagerie médicale.

| | domaine restreint | domaine large |
|--------------------------------|---------------------------|--|
| diversité des images | faible | importante |
| connaissance | spécifique | générale spécifique |
| champ sémantique | homogène | hétérogène |
| vérité terrain | oui, accessible | oui, difficile à acquérir |
| description du contenu | objective | objective |
| nature de l'application | expert | multi-expert |
| approche | modèle | multi modèle |
| domaine scientifique principal | reconnaissance des formes | recherche d'information reconnaissance des formes |
| attributs image | invariants spécifiques | invariants génériques invariants spécifiques |
| interactivité | limité | complexe |
| évaluation | quantitative, qualitative | quantitative, qualitative |
| architecture du système | classique | hiérarchique |
| taille de la base d'images | moyenne | grande à très grande |

FIG. 2.2 – *Influence de la nature du domaine en fouille d'images.*

Le tableau 2.2 indique des tendances générales car, bien sûr, il existe un grand nombre de situations intermédiaires.

Quand le domaine est large, il est évident que l'architecture du système de fouille est complexe (cf. paragraphe 2.8.3).

Contexte d'acquisition de l'image

En médecine, l'image est souvent acquise suivant une modalité plus ou moins complexe. Par exemple, en IRM, la prise d'images se fait à différents moments après injection d'un produit de contraste.

Par ailleurs, l'appareil d'acquisition est réglé d'une certaine manière et dans des conditions données. Cela induit une variabilité dans la représentation d'un même objet ou d'un même phénomène.

Ces informations sur le protocole ne sont pas accessibles à la lecture de l'image seule, elles sont indispensables pour l'interprétation.

Connaissances sur le domaine

D'une manière générale, l'interprétation d'une image fait appel à des données extraites directement de l'image et des connaissances du domaine d'application. Ces dernières ne sont pas nécessairement représentées explicitement dans le système, mais leur absence peut rendre difficile voire impossible la comparaison de deux images.

D'ailleurs, il est plus juste de considérer que les connaissances du domaine sont exploitées au moins de deux manières : 1) pour extraire les informations pertinentes dans l'image, 2) pour comparer deux images. Par exemple, en imagerie médicale dans les images de la pathologie que nous traitons dans cette thèse, les connaissances du domaine interviennent bien à ces deux niveaux :

- la synovite se manifeste par des zones claires autour des articulations métacarpiennes. Il conviendra donc de trouver des traitements d'images qui extraient et caractérisent automatiquement ces zones.
- deux images de synovites peuvent être différentes d'aspect, d'un point de vue topologique, mais correspondent à un même état de la pathologie si cette dernière se manifeste sur des articulations différentes du métacarpe.

En plus des connaissances sur la sémiologie d'une pathologie, des informations sur le patient sont parfois indispensables pour affiner le diagnostic.

Ces lacunes entre l'image seule et l'image associée à son protocole d'acquisition et au contexte du patient constituent ce que Smeulders appelle le "sensory gap".

2.3.4 Fossé sémantique

Un des problèmes importants en recherche d'images par le contenu vient du fait que l'utilisateur recherche une similarité sémantique et que le système propose une similarité par traitement d'images et de données. C'est ce que plusieurs auteurs appelle le **fossé sémantique** et qui est défini comme l'écart entre l'information que l'on peut extraire d'une image et l'interprétation qu'en a un observateur dans un contexte donné et au regard des connaissances qu'il possède.

2.3.5 Applications potentielles de la recherche d'images par le contenu en médecine

Dans le cadre du projet "Système de communication pour la formation en médecine basé sur l'indexation d'images"¹, nous avons identifié les applications potentielles suivantes correspondant à une attente des médecins et qui nécessitent la création d'une base d'images expertisées :

1. Formation

Ce contexte concerne essentiellement la radiologie. En se limitant, dans un premier temps, à une pathologie et à une modalité d'examen données, il convient de

1. sous programme 4 HTSC

construire une base d'images expertisées et annotées par le médecin représentative des cas rencontrés². Le principe d'un système de formation basé sur la fouille d'images est alors le suivant :

- (a) une image "exercice" est proposée à l'étudiant,
- (b) l'étudiant formule un avis sur l'image,
- (c) il envoie l'image vers le système qui lui propose en retour N_r images qu'il considère comme semblables.
- (d) l'étudiant examine les images retournées et positionne pour chacune un indicateur de pertinence,
- (e) le système modifie son analyse en fonction des annotations de l'étudiant,
- (f) retour en (1d) pour quelques itérations,
- (g) quand l'étudiant est satisfait, il accède au texte associé à la meilleure image retournée et le compare à son analyse sur l'image requête.

2. Recherche d'exemples pathologiques semblables

Le médecin spécialiste en Centre Hospitalier Universitaire, recherche dans une base de dossier patient, ceux qui contiennent des images semblables à l'image requête. En consultant les dossiers patients contenant les images retournées, il complète sa connaissance d'une pathologie.

3. Aide au diagnostic

Le médecin souhaite un enrichissement de son diagnostic en consultant des cas correspondant à des images semblables. Cette application est voisine de la précédente. Le système de fouille d'images est un outil d'amélioration de la connaissance.

2.4 Description du contenu des images

Cette phase ne peut pas se faire sans le médecin. Le problème est d'élaborer une représentation pertinente de l'image et de la coder sous forme d'un index. Il est donc normal de retrouver la plupart des méthodes de traitement et d'analyse d'images qui se situent en amont de la reconnaissance des formes. En plus de la pertinence de la représentation, l'autre propriété importante est son invariance. L'objet de cette section n'est pas d'être exhaustive mais de rappeler quelques aspects fondamentaux.

En médecine, la plupart des images médicales sont en niveau de gris. Pour cela, nous nous concentrons dans cette étude à ce type d'images. Il est important de noter que toutes les procédures de traitement d'image couleur (segmentation, détection de contour, etc...) tridimensionnelles ne sont qu'une extension de celles des images en niveau de gris bidimensionnelles. Le lecteur intéressé par l'image couleur peut consulter avec profit le livre de Trémeau et al. (2004) qui rassemble les méthodes de traitement ainsi que le problème d'indexation de bases d'images couleur en plus d'autres applications.

2. la constitution d'une telle base est un problème en soi

2.4.1 Attributs

Après traitement de l'image, des attributs sont calculés sur l'image entière ou sur des parties de l'image obtenues par exemple par segmentation en région, par détection de contours (fermés ou pas) ou autour des points caractéristiques.

Attributs de niveau de gris

Nous distinguons des attributs de "bas niveau" et de "haut niveau".

Les attributs de niveau de gris de "bas niveau" sont les moments et les moyennes d'espace (voir définitions Cocquerez et Philipp, 1995).

En effet, l'histogramme est l'attribut d'intensité de gris couramment utilisé en indexation. Il consiste en l'accumulation de l'information de niveau de gris des pixels sur toute (ou une partie) de l'image. Il est à rappeler que l'histogramme normalisé est la moyenne d'espace du premier ordre. L'histogramme des cooccurrences est la moyenne d'espace du second ordre.

Les moments d'espace centrés et non centrés caractérisent aussi l'information de niveau de gris. La moyenne de niveau de gris (moment d'espace du premier ordre) et la variance (moment d'espace centré de second ordre) ainsi que le troisième moment, sont souvent employés en indexation.

Les attributs de "haut niveau" sont fabriqués après une transformation ou une analyse statistique sur un ensemble d'images observées. Une méthode consiste à représenter l'image entière par ses composantes principales extraites après l'analyse en composantes principales (ACP) sur l'ensemble d'images de la base. Cette méthode a été introduite par Kirby et Sirovich (1990) pour la reconnaissance des visages. (Nastar et al., 1998) utilisent cette approche en indexation pour la détection des visages (Turk et Pentland, 1991) et des objets 3D (Murase et Nayar, 1995).

Texture

Dans le domaine du traitement de l'image, il n'existe pas de définition unique de la texture. Elle se décrit en termes linguistiques (Rao, 1990) tels que la rugosité, le contraste, etc., que les chercheurs se sont efforcés de caractériser depuis plusieurs années. Une texture peut avoir un aspect périodique ou aléatoire.

En raison de son importance et de son utilité dans la reconnaissance des formes et dans la vision par ordinateur, des travaux de recherche très riches sur l'analyse de la texture ont été réalisés pendant les trois dernières décades. Nous renvoyons le lecteur à trois textes qui couvrent assez bien l'évolution du domaine (van Gool et al., 1985; Reed et Bur, 1993; Randen et Husoy, 1999). Egalement, la texture joue aujourd'hui un rôle important dans la recherche d'images par le contenu.

L'analyse des textures repose sur diverses approches. Comme les présente (Maître, 2003, chapitre 12), nous distinguons :

1. Les approches par mesures statistiques :

Dans ces approches, on mesure les propriétés statistiques autour de chaque pixel $p(i,j)$ de l'image dans une fenêtre W_{ij} de taille et de forme déterminées. Nous citons :

- les statistiques de premier ordre : les moments, centrés et non centrés, l'énergie, l'entropie, le contraste, la dynamique, etc...,
- la matrice de cooccurrence et les indices de texture de second ordre introduites par Haralick et al. (1973),
- les statistiques de points particuliers : on cherche dans ce cas, dans la fenêtre W_{ij} , la densité moyenne de certains points d'intérêt. Les points d'intérêts les plus utilisés sont, soient les maxima locaux de l'intensité, soient les points de contour (après application d'un détecteur et de seuillage).

2. La détection d'énergie dans des canaux fréquentiels :

L'idée sous-jacente à toutes ces techniques est d'extraire l'énergie portée par le signal (image) dans les bandes de fréquences diverses. De nombreux filtres sont proposés, ils sont conçus d'une façon heuristique. Citons entre autres :

- les filtres de Law (Laws, 1979),
- les filtres de Gabor,
- les filtres en anneau et en coin (Coggins et Jain, 1985),
- les ondelettes.

3. Les filtres optimisés :

Les filtres heuristiques donnent de bons résultats en reconnaissance de texture. Mais le problème réside dans le nombre élevé d'attributs extraits, qui par conséquence influencent les méthodes de classification qui peuvent être lourdes et hasardeuses. Il est possible de mettre en place des techniques adaptées pour les seules textures que l'on recherche. Ce qui permet de se concentrer sur quelques mesures seulement de l'espace fréquentiel. Il s'agit donc de trouver les filtres qui optimisent un critère lié au problème de la classification de textures donné. Les filtres de Gabor, qui disposent d'un petit nombre de paramètres seulement, se prêtent assez bien à ce type d'optimisation (Unser, 1986).

4. Les modélisations autorégressives :

De la même façon que les signaux monodimensionnels sont analysés par des modélisations autorégressives unidimensionnelles (AR et ARMA) afin de mettre en évidence les périodicités qu'ils contiennent, les signaux bidimensionnels peuvent être analysés par des modèles autorégressifs bidimensionnels (Mao et Jain, 1992).

5. Les approches par champs markoviens.

Il n'y a pas de méthodes d'analyse de textures qui soient à ce jour complètement efficaces. A chaque type de textures ou d'images correspond une méthode pertinente. Des conseils d'utilisation sont données par Randen et Husoy (1999) après une étude comparative très bien présentée.

Les attributs de texture en indexation font leur apparition dans QBIC (Niblack et al., 1993). Les caractéristiques en question sont le grain (coarseness en anglais), le contraste

et l'orientation (Tamura et al., 1978) (approche statistique). Les travaux de Haralick et al. (1973) sur la texture et en particulier les matrices de cooccurrences ont servi de base pour l'indexation (Aksoy et Haralick, 1998). Les modèles autorégressifs sont également couramment utilisés (Vasconcelos et Lippman, 2000b), au même titre que les filtres de Gabor (Fournier et al., 2001) et les ondelettes (Wang et Wiederhold, 2001).

Description de formes

Les descripteurs (ou paramètres) de forme sont des nombres qui représentent chaque forme (d'un objet) et permettent de les classer. Ces paramètres ont été abondamment proposés dans la littérature (Coster et Chermant, 1985; Pratt, 1991; Nadler et Smith, 1992). Comme le propose Mehtre et al. (1997), un schéma récapitulatif des techniques de description de formes est donné dans la figure 2.3. Ces techniques peuvent être classées en deux catégories : méthodes basées frontières et méthodes basées régions. La première catégorie des techniques caractérise seulement le contour ou la frontière de l'objet en ignorant son intérieur. Le deuxième type de méthodes prend en compte les détails internes comme par exemple le nombre de trous à l'intérieur de l'objet.

Nous trouvons des descripteurs simples à calculer comme la surface et le périmètre qui se calcule en comptant les pixels du contour. L'inconvénient de ce type d'attribut est qu'ils sont liés à l'échelle et la taille de l'image. Nous ne pouvons pas comparer les objets utilisant ces descripteurs sans normalisation des échelles des images. Les travaux en reconnaissance des formes sont plutôt dirigés à la recherche des attributs robustes aux transformations géométriques telles que la translation, la rotation et le changement d'échelle. Ainsi les moments invariants de Hu (1962) sont utilisés pour caractériser l'intégralité de la forme d'une région. Une seconde approche basée frontière fait classiquement référence aux descripteurs de Fourier (Persoon et Fu, 1977) et porte sur une caractérisation des contours de la forme.

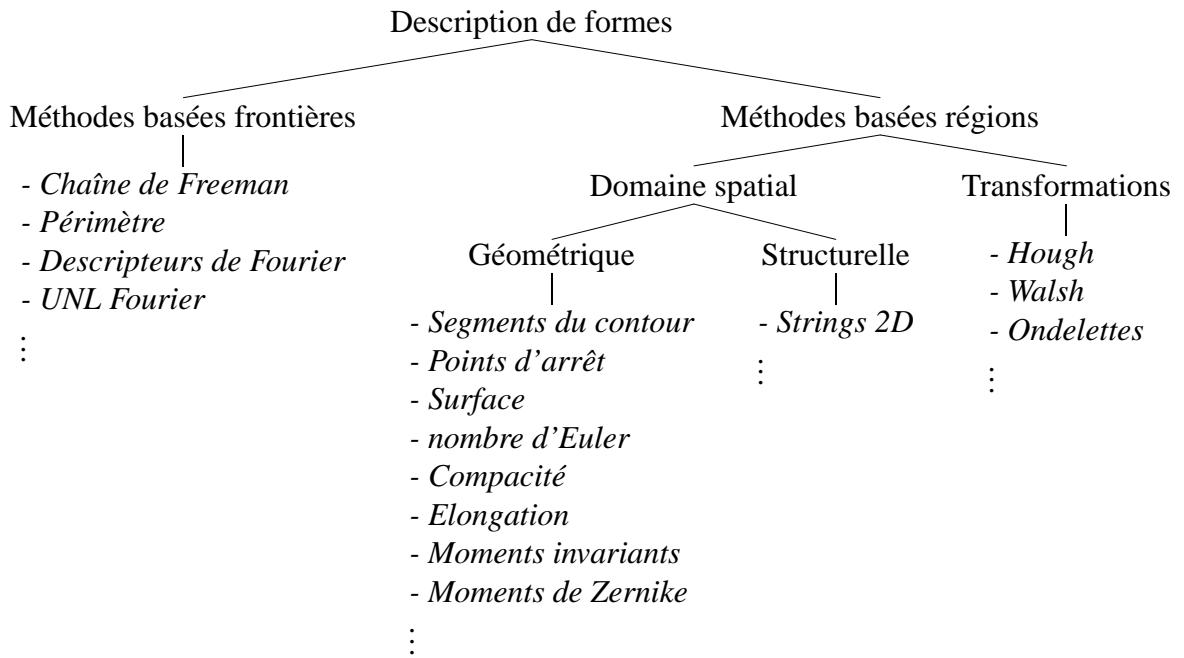


FIG. 2.3 – Un aperçu général des techniques de description de formes

Aujourd’hui, les descripteurs de forme prennent une place importante en recherche d’images par le contenu (Rui et al., 1996; Derrode et al., 1999; Derrode, 1999). Ils sont complémentaires de la description de texture et de couleur même parfois essentiels et surtout en imagerie médicale, où le diagnostic s’effectue par caractérisation de la forme des régions pathologiques.

Une comparaison des attributs de forme pour la recherche des images par le contenu est réalisée par Mehtre et al. (1997). La combinaison des deux descripteurs : moments invariants de Hu et UNL³ Fourier (Rauber et Steiger-Garçao, 1992) donnent des résultats prometteurs.

2.4.2 Points d’intérêt

Un point d’intérêt est communément défini par un changement bidimensionnel important du signal (Schmid et Mohr, 1997). Les points qui correspondent le mieux à cette définition sont les coins et les jonctions en T, mais aussi les endroits de forte variation de texture. L’hypothèse proposée réside sur la robustesse de cette information face aux modifications importantes que pourraient subir l’image.

L’utilisation des points d’intérêt est liée principalement à deux domaines, celui de l’apariement d’images pour la stéréovision et celui de l’indexation. Dans les deux cas, l’extraction des points d’intérêt n’est que la base des développements, la tâche principale étant

3. Universidade Nova de Lisboa

de faire correspondre le point d'intérêt adéquat dans les deux images à appairer. Pour cela plusieurs techniques sont possibles (graphes, carte de distances géométriques), mais les plus utilisées définissent un vecteur de caractéristiques locales pour chaque point. D'autres voies d'utilisation existent, Bas (2000) propose notamment d'utiliser ces points pour composer un partitionnement spécifique à l'image, afin d'assurer des points d'ancre robuste pour insérer un marquage de l'image.

Pour que les caractéristiques locales soient utiles (en appariement pour la stéréovision, indexation, ...), elles doivent être invariantes aux transformations de l'image : rotation, translation, changement d'échelle ou d'illuminant. Schmid et Mohr (1997) proposent le calcul d'invariants différentiels autour de points d'intérêt. Cette description locale du signal a été étendue aux images couleurs par Gros et al. (1997). De manière similaire, Tuytelaars et Gool (1999) utilisent les invariants couleurs comme caractéristiques locales en indexation.

2.4.3 Segmentation

Conceptuellement, la segmentation consiste à extraire de l'image des primitives, soit de type contour, soit de type région. Ces primitives seront exploitées ultérieurement pour effectuer la reconnaissance des formes, le recalage, la mise en correspondance, etc. La segmentation n'est donc pas un objectif en soi mais une étape située en amont de dispositifs d'interprétation d'images.

La littérature foisonne de méthodes de segmentation, mais aucune n'est valable pour tout type d'image. Le choix d'une méthode dépend souvent des connaissances *a priori* sur l'image. Un aperçu général des approches de segmentation d'image est illustré dans la figure 6.7 (Cocquerez et Philipp, 1995).

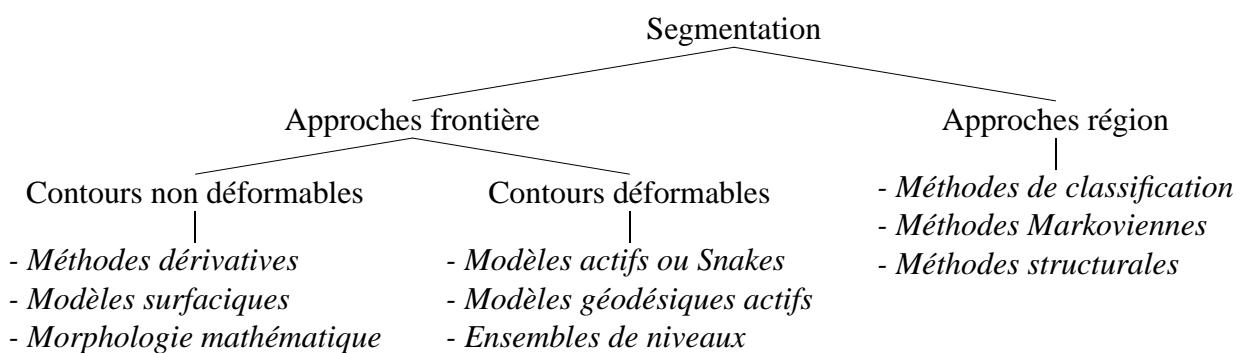


FIG. 2.4 – Un aperçu général des techniques de segmentation d'image

La segmentation est très importante pour l'indexation et la recherche d'images par le contenu (Winter et Nastar, 1999; Pauwels et Frederix., 1999; Carson et al., 1999; Wang et

Wiederhold, 2001), spécialement en imagerie médicale où nous sommes obligés d'extraire les régions d'intérêts et de les caractériser (niveau de gris, forme, texture) afin de trouver des régions similaires dans la base. En outre, la segmentation des images nous permet d'obtenir une description spatiale (topologie) des régions de l'image. Il est important de noter que la segmentation totalement automatique des images médicales est un problème difficile à résoudre sans l'intervention d'un expert (Olabarriaga et Smeulders, 2001; Pham et al., 1998).

2.5 Indexation

Après une introduction aux principales caractéristiques d'images utilisées en indexation⁴, nous nous intéressons maintenant aux méthodes de structuration de ces informations pour la formation de la signature appelée encore index.

La fabrication d'un index doit résulter d'un compromis entre la pertinence de l'information pour chaque image et les contraintes de temps de recherche. Les index doivent être choisis en fonction de l'application visée, au même titre que les caractéristiques primaires dont ils sont issus.

Comme cela est introduit dans le paragraphe précédent (*cf.* section 2.4.1), il existe plusieurs types de primitives pour l'indexation : le pixel, le bloc de pixel, le point d'intérêt, la région, le contour, etc. Deux approches de structurations sont alors possibles : l'approche globale et l'approche partielle.

La première consiste en l'accumulation de l'information sur toute l'image pour la formation d'une signature dite globale, ce qui restreint la recherche aux images globalement similaires. Deux structures de l'index global sont alors possibles :

1. structure à plat,
2. structure hiérarchique.

La première structure consiste à concaténer les vecteurs d'attributs de couleur, de forme, de texture, etc. pour construire le vecteur index. Dans la deuxième structure, les vecteurs d'attributs ne sont pas concaténés dans le but de définir une mesure de similarité qui correspond à chacun de ces vecteurs.

Dans l'approche partielle, l'image est représentée par un ensemble de régions obtenues après segmentation et caractérisée chacune par un vecteur d'attributs. Dans ce cas, il s'agit de rechercher une région ou un sous-ensemble de régions similaires dans la base d'images. La recherche peut s'effectuer ainsi par comparaison de toutes les combinaisons possibles des régions du sous-ensemble donné (requête) et les régions des images de la base (Carson et al., 1999).

Pour représenter la topologie de l'image, il s'agit d'indexer l'image par son graphe d'adjacence. Les graphes sont utilisés depuis longtemps en traitement et analyse d'images pour essayer de représenter la topologie de l'image. C'est un outil d'analyse complet mais complexe à gérer. Classiquement, les coûts combinatoires varient exponentiellement avec

4. Le mot indexation signifie plutôt la fabrication des signatures appelées encore index. On ne traite pas l'indexation de la base des données (structuration) extraites des images.

la taille des graphes pour des opérations de recherche notamment (Eshera et Fu, 1984; Shapiro et Haralick, 1981). En indexation et recherche d'images, l'utilisation des graphes est limitée vu les contraintes de temps de recherche. Philipp et Lekkak (2003) utilisent l'algorithme d'appariement "inexacte" de graphes proposé par (Cordella et al., 1998) pour rechercher des objets similaires représentés par un sous-graphe du graphe de l'image.

2.6 Similarité

Après avoir abordé le problème d'indexation par le contenu, intéressons-nous à la question de la recherche par similarité au sein des bases d'images.

L'analyse de la similarité entre deux images est un thème important en traitement d'images. En général, la comparaison est effectuée entre les signatures images précédemment calculées. L'ensemble de mesures entre les vecteurs index définit une similarité inter-images.

Comme nous l'avons déjà défini, l'index est une structure compacte des attributs extraits des images. La définition d'une mesure similarité entre les index implique le choix d'une mesure de similarité entre attributs. Encore une fois, il y a tout l'héritage de la reconnaissance des formes en ce qui concerne le calcul de similarité. Nous citons les distances les plus couramment utilisées : distance euclidienne, distance de Mahalanobis, les distances de Minkowski L_p . Des mesures plus spécifiquement dédiées à la comparaison d'histogrammes sont également employées.

Les approches statistiques de classification ainsi que de la théorie de décision peuvent être aussi appliquées pour trouver les images appartenant à une catégorie d'images donnée (Vasconcelos et Lippman, 2000c).

2.7 Interactivité

Les index calculés hors ligne ne permettent pas toujours de caractériser efficacement la sémantique des images. Les images renvoyées par le système sont similaires à la requête du point de vue de ces index objectifs, mais leur sémantique n'a bien souvent rien à voir avec celle de l'image requête.

Cette différence, dénommée fossé sémantique (voir section 2.3.4), limite grandement les performances des systèmes sans interaction.

L'intervention de l'utilisateur dans la recherche est proposée pour réduire voire combler ce fossé sémantique. De plus, l'Homme dans la boucle constitue un changement majeur dans l'approche de reconnaissance, dans la mesure de similarité notamment. Il doit guider le système, et les solutions sont dépendantes de son intervention.

Ce problème de recherche interactive est appelée "recherche par boucle de pertinence". Il demeure un thème de recherche très important que nous traitons en détails dans le chapitre suivant.

2.8 Présentation de quelques systèmes de fouilles d'images médicales

Nous présentons quatre systèmes de recherche d'images médicales. Le domaine des deux premiers systèmes (IGDS et ASSERT) est restreint. Celui des deux derniers peut être classé comme large (voir tableau 2.2).

Nous avons détaillé volontairement la présentation de ces systèmes pour permettre au lecteur une vue globale d'un système de recherche par le contenu et pour lui donner un exemple de points traités dans les sections précédentes. Après une synthèse de cette étude nous présentons notre système en justifiant nos choix.

Le lecteur intéressé par la liste, presque complète mais peu détaillée, des travaux réalisés en recherche d'images par le contenu pour des applications médicales peut consulter l'article de Müller et al. (2004).

2.8.1 Système IGDS

Nous présentons ce système pour donner une exemple d'application des systèmes de recherche par le contenu pour l'aide au diagnostic.

Comaniciu et al. (1999) ont développé IGDS (Image Guided Decision Support) qui supporte la prise de décision en pathologie clinique. La tâche du système est de localiser, rechercher, afficher à l'écran des cas morphologiques conformes à l'image requête, et de suggérer le diagnostic le plus pertinent en se basant sur la majorité des cas retrouvés dans les images similaires.

La base de données contient les images des spécimens microscopiques de cellules de sang (cf. figure 2.5). Ces cellules appartiennent à quatre classes dont trois sont des classes anormales plus une classe saine. La base contient 261 images de référence dont la vérité terrain est connue, afin d'être utilisée lors de la recherche en ligne de cellules similaires.

Les cellules sont distinguées par la forme, la couleur, la surface et la texture du noyau. Ainsi, les auteurs ont utilisé le descripteur invariant de Fourier pour décrire la forme : 10 harmoniques ou encore 40 coefficients de Fourier ont été utilisés pour construire le vecteur forme. La dis-similarité entre deux vecteurs forme est mesurée par la distance euclidienne.

L'analyse de la texture est basée sur le modèle autorégressif MRSAR (multiresolution simultaneous autoregressive model). Ensuite, Un vecteur de dimension 15 est obtenu en accord avec ce modèle pour décrire la texture. La distance entre deux noyaux en termes de texture est ainsi indiquée par la distance de Mahalanobis.

Vu que l'attribut de couleur n'avait pas le pouvoir discriminant pour distinguer les trois classes, les auteurs ont utilisé l'information couleur pour séparer le noyau du fond de l'image et non pas pour distinguer les cellules.

La surface du noyau est prise proportionnelle au nombre de pixels à l'intérieur de la région du noyau, car les images dans la base de données sont toutes prises à la même échelle. La dis-similarité mesurée entre les surfaces de deux noyaux est la racine de leur différence quadratique.

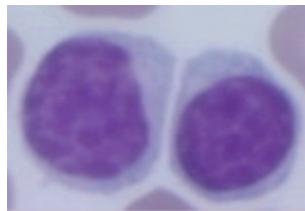


FIG. 2.5 – *Une image d'un spécimen microscopic de sang. Cette image exemple présente deux cellules du sang.*

La mesure de dis-similarité totale entre deux noyaux est choisie comme une combinaison linéaire des distances des trois attributs (texture, couleur, surface). Les poids de cette distance ont été calculés hors ligne, en se servant des 261 images de référence, telle que la somme des probabilités conditionnelles de la décision correcte soit maximale.

Une session typique de recherche commence par le chargement de l'image requête à partir du microscope. Ensuite, le physicien choisit une région rectangulaire qui contient la cellule d'intérêt. La région est alors segmentée automatiquement en utilisant un algorithme basé sur une classification (non supervisée) non paramétrique de la couleur des pixels.

Par défaut, le système retourne les huit plus proches voisines images en calculant la similitude entre l'image requête et chacune des images dans la base de données. La classification suggérée de l'image requête est basée sur la majorité votant parmi les huit images retournées.

2.8.2 Système ASSERT

Shyu et al. (1999) ont conçu le système ASSERT⁵ de recherche basé sur le contenu, pour des images médicales. Il se caractérise par l'intégration du physicien dans la phase d'indexation.

La base de données se compose des images HRCT (High Resolution Computed Tomography : tomodensitométrie de haute résolution) du poumon (voir figure 2.6). La recherche d'image basée sur le contenu est particulièrement nécessaire pour ce domaine parce que le diagnostic se fait actuellement par consultation d'un atlas édité des pathologies de poumon, pour les cas qui ne sont pas immédiatement reconnaissables. Le système a été réalisé pour sauver le radiologue de cette tâche laborieuse de la pagination de l'atlas en recherchant une image qui ressemble à la pathologie de leur patient.

Pour archiver une image, un médecin entoure une ou plusieurs régions pathologiques (appelé PBR : Pathology Bearing regions) dans l'image en question. Le système recherche alors les images les plus visuellement semblables de la base de données en utilisant un index constitué d'une combinaison des descripteurs locaux des PBR.

5. C'est un travail de collaboration entre le département de radiologie à l'université de Indiana, de l'école de médecine à l'université de Wisconsin, et des laboratoires de "Machine Learning" et "Robot Vision" à l'école de génie informatique et électrique à l'université de Purdue.

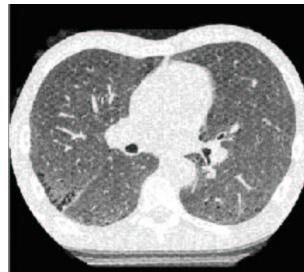


FIG. 2.6 – une image HRCT du poumon.

En effet, chaque image est représentée par un ensemble de régions appelées LFS (Lobular Feature Set). Une région LFS correspond à un lobe ou une combinaison de lobes du poumon. La frontière des lobes est délimitée par le praticien. Chaque LFS est décrite par l'ensemble des attributs de sa région pathologique PBR et de celle adjacente à sa frontière.

Extraction des attributs d'un PBR

Les attributs sont extraits comme suit :

- Un physicien marque les régions d'intérêts pathologiques.
- Extraction automatiquement de la région du poumon en utilisant des routines d'analyse d'images binaires.
- Extraction des attributs locaux de chaque PBR. Il y a 255 attributs extraits, citons par exemple :
 - a- Texture : en appliquant l'approche statistique basée sur la matrice de Cooccurrence de niveau de gris : un ensemble d'attributs de texture est calculé, tel que l'énergie, l'entropie, la homogénéité, le contraste et la corrélation.
 - b- Propriétés de niveau de gris : un histogramme de niveau de gris comme attribut local. L'écart type et la moyenne de niveau de gris des pixels sont calculés pour le reste du poumon (attribut global).
 - c- Forme : l'axe le plus long et l'axe le plus court, l'orientation et la mesure de complexité de la forme en utilisant des descripteurs et des moments de Fourier.
 - d- Contour : utilisation de l'opérateur de contour de Sobel pour calculer la distribution des contours : rapport du nombre de pixel du contour sur le nombre de pixel dans la région pour différents contours détectés après plusieurs seuillages.
 - e- Structure de variations de niveau de gris dans le PBR : Calculer le nombre de régions segmentées dans un PBR, l'histogramme des surfaces pour toutes ces régions, et les statistiques de niveau de gris pour chaque région.
 - f- Distance entre le centre de la surface d'un PBR marqué et la frontière la plus proche du poumon.
 - g- Position de PBR dans le poumon.
 - h- Rapport de la surface de PBR sur la surface du poumon.

La dimension de l'ensemble d'attributs est réduite à 12 en appliquant la méthode de "sélection séquentielle ascendante" de variables (sequential forward selection SFS) de Kitler (1978), qui trouve un ensemble d'attributs pertinents.

Indexation

Un classifieur à table de hachage multidimensionnel (multidimensional hash table) (Grewe et Kak, 1995) est proposé pour accélérer la procédure de recherche des images similaires. Cependant, il est utilisé pour classer les LFS efficacement. La classification effectuée est supervisée. Il existe 46 classes de LFS. L'idée principale est de construire un arbre de décision basé sur la minimisation de l'entropie moyenne. La distribution statistique de chaque classe de LFS est approximée, dans l'espace des attributs, par une gaussienne.

Similarité

Les images similaires à une image requête sont retrouvées après deux étapes successives :

1. retrouver les images de la base qui correspondent à la classe des LFSs de l'image requête,
2. ensuite, retourner à l'écran parmi les images retrouvées, celles qui contiennent les plus proches voisines PBRs à l'image requête. La similarité entre deux PBRs est mesurée par une distance euclidienne.

2.8.3 Système IRMA

Face à la connaissance médicale complexe, au moins trois niveaux sémantiques additionnels⁶ d'abstraction⁷ sont nécessaires selon Lehmann et al. (2000) pour construire les systèmes à des applications médicales. IRMA⁸ coupe le procédé de recherche en sept étapes consécutives :

1. catégorisation (utilisant des attributs globaux),
2. recalage (en géométrie et contraste),
3. extraction d'attributs (des attributs locaux),
4. sélection d'attributs (dépend de la catégorie et la requête),
5. indexation (représentation en une structure de "Blobs" hiérarchique et multi-échelle),

6. par rapport aux systèmes standard de CBIR

7. Démarche de l'esprit qui consiste, au cours d'un raisonnement, à éliminer les aspects les moins pertinents de la réflexion pour ne considérer que ceux qui sont essentiels. La capacité d'abstraction est une des principales caractéristiques de l'intelligence humaine. Certains systèmes d'intelligence artificielle ont la capacité d'utiliser l'abstraction pour résoudre des problèmes. Par extension, le terme "abstraction" désigne également le résultat (le modèle) issu de la démarche d'abstraction.

8. "Image Retrieval in Medical Applications" est un projet de coopération entre le département de radiologie, le département de l'informatique médicale, de la division de traitement d'images médicales et "the Chair of Computer Science VI" à "Aachen University of Technology (RWTH Aachen)" en Allemagne.

6. identification (lie la connaissance *a priori* du contenu des images aux Blobs),
7. recherche (traitée au niveau des Blobs).

Chaque étape représente un niveau plus élevé d'abstraction de l'image, reflétant un niveau croissant de la compréhension du contenu de l'image.

L'étape "catégorisation" vise à déterminer pour chaque image entrant dans la base de donnée sa modalité, son orientation aussi bien que la région examinée du corps et le système fonctionnel. Les dispositifs DICOM (Digital Imaging and Communications in Medicine) modernes pour la sauvegarde des images, comportent toute l'information exigée par l'étape "catégorisation" du système IRMA. Pour cela, un code hiérarchique détaillé a été développé (Güld et al., 2002), il excède la complexité des étiquettes existantes dans les images en format DICOM standard. Néanmoins, des procédures automatiques de "catégorisation" sont proposées (Keyser et al., 2003) pour l'archivage rapide des images numérisées qui ont été acquises par des modalités basées "film" ou des dispositifs numériques autre que DICOM. À cette première partie du traitement, aucune information sur le contenu d'image n'est disponible. Par conséquent, des descripteurs caractérisant l'image entière doivent être employés pour la "catégorisation". La "catégorisation" automatique de IRMA est basée sur une base de données de référence de 10.000 images choisies et manuellement classées par les radiologues expérimentés. Cette base de données intègre la connaissance médicale à un niveau bas d'abstraction.

Ainsi, la "catégorisation" automatique des images requêtes est effectuée en combinant l'information d'en-tête de DICOM et des attributs globaux de l'image. Cependant, la "catégorisation" dans IRMA n'est pas exclusive. Les étapes suivantes du traitement sont appliquées sur les catégories les plus probables.

Le "recalage" en géométrie (rotation, translation et mise à l'échelle) et en contraste produit un ensemble de paramètres de transformation qui sont stockés pour chaque image et chacune de ses catégories probables. En accord avec Tagare et al. (1997), l'enregistrement est basé sur des prototypes qui sont manuellement définis pour chaque catégorie. Encore une fois la connaissance d'un expert médical est incorporée dans le système IRMA. Notez que la transformation n'est pas exécutée explicitement à cette étape de traitement. Les paramètres produits sont utilisés dans les couches : "identification" et "recherche".

L'étape "extraction d'attributs" donne des descriptions locales des images, c'est-à-dire génère un ensemble d'attributs calculés pour chaque pixel. Ils peuvent être, soient indépendants des catégories (par exemple résulter de la détection de contour ou de l'analyse de texture) ou spécifiques à des catégories, comme l'application d'un modèle actif de forme qui emploie explicitement la connaissance *a priori* de la catégorie.

La séparation de l'étape "sélection d'attributs" de l'étape "extraction d'attributs" laisse intégrer la catégorie de l'image et le contexte de la requête dans le processus d'abstraction. Par exemple, la même radiographie pourrait être sujette à une fracture ou à un examen de cancer, ayant pour résultat une combinaison des attributs basés contour ou texture. Afin d'éviter le calcul exhaustif pendant la recherche en ligne, ces ensembles d'attributs sont pré-calculés pour chaque image dans chacune de ses catégories probables.

Jusqu'ici, le stockage d'une image dans la base de données inclut la catégorisation, la détermination des paramètres de recalage pour chaque catégorie, l'extraction d'attributs de

l'image ainsi que le choix de quelques ensembles distincts d'attributs dans chaque catégorie. Dans le processus de recherche, cette quantité d'informations doit être rigoureusement réduite. Pour cela et selon l'ensemble d'attributs sélectionné, l'image est segmentée en régions appropriées par classification automatique des attributs (Thies et al., 2003). Toutes les régions sont décrites par des moments invariants ayant pour résultat des structures appelées "Blobs". Par exemple, les Blobs sont égales à des ellipses si seulement le premier et le deuxième moment sont employés pour la description. Ensuite, la représentation en Blobs de l'image est ajustée suivant les paramètres déterminés dans l'étape de recalage. Cette représentation permet à l'utilisateur de rechercher des images entières aussi bien que des régions d'intérêt (ROIs).

L'étape "identification" permet l'intégration des connaissances médicales à certaines Blobs produites pendant l'étape "indexation" afin de les identifier si c'est possible. Elle se fonde sur des prototypes définis pour chaque catégorie, et marqués par des experts médicaux. Ainsi, l'identification est la base fondamentale qui introduit la compréhension de haut niveau de l'image en analysant les relations spatiales et temporelles entre les Blobs et permet d'avoir une structure prototype de Blobs.

La "recherche" est effectuée en fouillant les structures hiérarchiques des Blobs. Dans IRMA, une requête peut être construite de l'une ou de toutes les composantes suivantes :

- une liste des catégories possibles des images de la base de données,
- un exemple de structure de Blobs à l'échelle optimale pour traiter la requête,
- un ensemble d'attributs locaux qui décrivent mieux les propriétés significatives de la requête (par exemple, un ensemble d'attributs de texture ou de forme).

Notez que seulement l'étape "recherche" exige des calculs en ligne tandis que toutes les autres étapes peuvent être exécutées automatiquement en lots, au temps d'entrée d'une image dans la base de données.

2.8.4 Système à base de connaissance pour la recherche d'images médicales

Nous présentons ce système développé par l'UCLA⁹ (Chu et al., 1998) car il utilise un modèle sophistiqué de représentation d'images qui comprend quatre couches :

- (Raw Data Layer : RDL), cette couche établit les liens avec la base d'images brutes et le système. En particulier, elle convertit l'image brute qui peut avoir différents formats et modes de compression, en un format canonique utilisable par le système.
- (Feature and Content Layer : FCL), cette couche extrait et mémorise les primitives d'une image (les contours, les régions) et les relations spatiales qui les lient. Le choix des primitives pertinentes a été fait par un expert.
- (Schema Layer : SL), cette couche établit les relations entre les entités médicales (par exemple : lésions, ventricule droit,...) et les objets de la couche FCL.

9. University of California, Los Angeles

- (Knowledge Layer), cette couche correspond à une représentation sous forme hiérarchique :
 - des liens entre des termes qualitatifs médicaux comme (sombre, opaque, granuleux,...) et les données quantitatives associées aux objets de la couche SL.
 - des liens entre une ontologie des relations spatiales en médecine et les relations spatiales entre objets de SL.

L'interrogation du système se fait à l'aide d'un langage de requête visuel. L'utilisateur dispose de plusieurs images d'entités visuelles qu'il sélectionne et place dans une fenêtre de requête. Il a la possibilité de délimiter des zones dans ces images et de les étiqueter à l'aide de mots-clés portant sur leur aspect et leur position. Ensuite, un processeur de requête interprète le contenu de la fenêtre et formule une requête sous forme de prédictats. Ainsi, l'utilisateur peut construire une requête qui en langage quasi naturel serait du genre :

"Trouver un patient qui a une tumeur ayant le même aspect et la même localisation que celle qui est présentée".

2.8.5 Synthèse

Comme nous l'avons vu à la section 2.3.3, la connaissance de l'expert est indispensable pour concevoir un système de recherche d'images médicales basée sur le contenu.

Prenons le système IGDS qui est conçu pour l'aide au diagnostic. La connaissance du domaine de l'application est intégré au niveau du calcul de la similarité entre images. Cela est réalisé pendant une phase hors ligne d'apprentissage sur une base de référence étiquetée par des spécialistes.

ASSERT se base sur les annotations du médecin pour détecter les régions pathologiques des images (phase d'extraction des attributs). En plus, Il se sert d'une base étiquetée pour classer les LFSs durant l'étape d'indexation.

Le système développé par l'UCLA s'intitule "Système à base de connaissance pour la recherche d'images médicales". Le choix de primitives pertinentes de l'image a été fait par un expert.

Finalement, IRMA incorpore la connaissance d'un expert à chaque étape d'abstraction du contenu de l'image.

Aucun de ces systèmes ne laisse à l'expert le choix d'interpréter les résultats de recherche *en ligne* et de donner son avis au vu des images similaires retrouvées par le système. L'affinement de la requête par boucle de pertinence n'existe pas.

En effet, que peut-on faire si le système se trompe ? Autrement dit, s'il retourne au spécialiste des images qui ne ressemblent pas à l'image requête. Cela est toujours possible, car même si le système s'enrichit en apprenant d'une base étiquetée, un taux d'erreur de reconnaissance de 0% est impossible à atteindre. Par ailleurs, la création d'une base d'apprentissage est un problème en soi, qui demande beaucoup de travail et de temps (parfois des années surtout en médecine) pour archiver tous les prototypes ou les cas possibles.

2.9 Notre système, nos choix

Nous avons choisi de concevoir un système de recherche d'images basé sur le contenu selon le schéma général de la figure 2.1. Ce schéma est assez général, il peut être utilisé pour d'autres applications médicales dont le domaine est restreint.

Nous avons veillé à la conception d'une méthode de recherche par boucle de pertinence qui intègre l'avis de l'utilisateur dans la mesure de similarité entre images afin de s'adresser soit au spécialiste soit à l'étudiant en formation, dans le but :

1. d'affiner la recherche des images similaires pathologiques. Dans ce cas le système s'adresse aux spécialistes (voir section 2.3.5).
2. de créer, à long terme, des bases structurées. Cela peut se réaliser par la sauvegarde des annotations des spécialistes à chaque fois qu'ils utilisent le système (voir section 2.3.2).
3. de se servir d'un tel système pour l'aide à la formation des étudiants en médecine (voir section 2.3.5).

La méthode de recherche par boucle de pertinence proposée fait l'objet du chapitre 5.

La phase d'indexation ne peut être que spécifique pour chaque application, pour extraire les informations pertinentes de chaque image. Il est difficile, voire impossible d'extraire des attributs de bas niveau de l'image qui représentent le contenu de l'image, sans avoir la connaissance sur le domaine de l'application (*cf.* section 2.3.3). Nous avons choisi de fabriquer une représentation des images ostéo-articulaires, qui soit :

1. totalement automatique,
2. basée sur les indices visuels interprétés par le spécialiste pour comprendre le contenu d'une image et faire le diagnostic.

La fabrication des signatures d'images est choisie automatique pour deux raisons :

1. simplifier la tâche de l'utilisateur au cours d'une recherche en ligne dans le cas où le système est employé par des spécialistes,
2. permettre l'utilisation de ce système pour la formation des étudiants qui ont peu d'informations sur la sémiologie de la pathologie.

Le chapitre 6 de cette thèse est consacrée à notre méthode d'indexation des images IRM d'une pathologie ostéo-articulaire : la polyarthrite rhumatoïde.

Chapitre 3

Etat de l'art de la recherche par boucle de pertinence

L'objectif de ce chapitre est de présenter le fonctionnement des systèmes de recherche interactive d'images par le contenu et de réaliser une synthèse des principales approches rencontrées dans la littérature.

3.1 Introduction

Dans le chapitre précédent nous avons montré que la recherche d'images par le contenu exploitait des méthodes et des techniques du traitement des images et de la reconnaissance des formes. Ainsi, dans un premier temps, le traitement d'image offre les moyens d'extraire des attributs pertinents pour élaborer une représentation de l'image sous forme de signature (index) ou de structure (graphe). Ces attributs sont identifiés avec l'aide de l'utilisateur expert.

Ensuite l'index étant choisi, et la base d'images étant indexée, le problème est de trouver un jeu d'images semblables à l'image requête. A ce niveau, on emploie des méthodes issues de la reconnaissance des formes, en particulier des mesures de similarité.

Cependant le contexte d'utilisation d'un système CBIR fait que l'on quitte les approches classiques de reconnaissance des formes car on veut prendre l'avis de l'utilisateur sur les images renvoyées. Cet avis peut alors être utilisé par le système pour effectuer une nouvelle recherche. On est donc en présence d'un système bouclé : appelé système avec boucle de pertinence.

3.2 Fonctionnement des systèmes de recherche interactive

L'algorithme de recherche interactive joue le rôle d'une interface entre un système informatique de grande capacité calculatoire mais de faible capacité d'interprétation et un

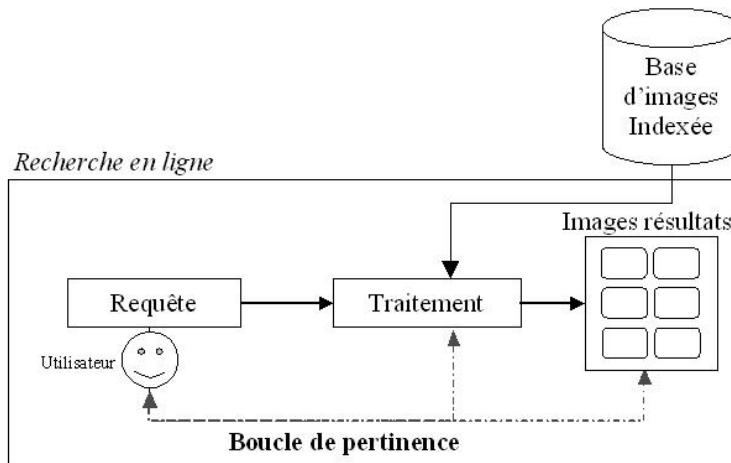


FIG. 3.1 – Schéma illustrant la boucle de pertinence

opérateur humain présentant les caractéristiques inverses. Ces systèmes fonctionnent sur le principe du bouclage, par répétition de la séquence d'événements suivante (*cf. fig 3.1*) :

1. le système recherche et présente un ensemble de résultats à l'utilisateur;
2. l'utilisateur expertise ces résultats;
3. le système intègre l'expertise utilisateur pour mettre à jour ses critères de recherche;
4. le processus est itéré (retour en 1) si l'utilisateur demeure insatisfait.

Plusieurs méthodes de bouclage de pertinence sont présentées dans la littérature. Comme le proposent Vasconcelos et Lippman (2001), nous distinguons l'approche *géométrique* de l'approche *statistique*.

3.2.1 Termes et notations

Dans ce chapitre, l'image requête sera représentée par l'index r élaboré à partir des valeurs numériques d'attributs extraits de l'image. Toutes les images sont indexées, c'est à dire représentées par le même index. La recherche se fait donc sur les index par définition d'une mesure de similarité.

Toute image cible de la base est représentée par son index x formé par p variables. On rencontre deux structures de l'index dans la littérature :

1. Structure à plat : dans ce cas les vecteurs d'attributs (de couleur, de forme ou de texture etc.) sont concaténés pour former le vecteur index :

$$\mathbf{x} = (x_j; j = 1, \dots, p);$$

2. Structure hiérarchique : les vecteurs d'attributs ne sont pas concaténés, x est noté :

$$\mathbf{x} = (x_j^{(m)}; j = 1, \dots, p_m; m = 1, \dots, M);$$

C'est une structure hiérarchique à M vecteurs d'attributs et $p = \sum_{m=1}^M p_m$.

Par ailleurs, la base contient N images indexées par $\{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N\}$.

Avant d'expliquer les différentes techniques de recherche interactive, notons que sous la même notion de "bouclage de pertinence", plusieurs méthodes ont été développées en se basant sur différentes hypothèses :

- Quelques unes supposent que l'utilisateur cherche une cible particulière (*Target search* en anglais) (Cox et al., 2000), alors que beaucoup d'autres supposent que l'utilisateur recherche des images similaires à une image requête (Fournier et al., 2001; Rui et al., 1997; Rui et Huang, 2000; Rui et al., 1997).
- Des algorithmes mettent à jour leur critère de recherche par les exemples annotés pertinents (positifs) et non-pertinents (négatifs) (Meilhac et Nastar, 1999; Tieu et Viola, 2000) tandis que d'autres par les exemples positifs seulement (Chen et al., 2001; Ishikawa et al., 1998). Wu et al. (2000) proposent un algorithme d'apprentissage des images annotées et non-annotées.

Le système PicHunter est connu par le "jugement relatif" : une image est annotée plus proche de l'image cible cherchée qu'une autre.

- Les attributs extraits des images sont représentés par une structure hiérarchique dans (Fournier et al., 2001; Rui et Huang, 2000; Laaksonen et al., 1999) pour gérer la compétition inter-attributs et intra-attributs, alors qu'une structure à plat est adoptée dans (Ishikawa et al., 1998; Zhou et Huang, 2001).

3.3 Approches géométriques de bouclage de pertinence

Dans le cadre des approches géométriques du bouclage de pertinence, la recherche repose sur deux entités, *la requête* et *la fonction de similarité*. La requête est un point d'ancrage dans l'espace de recherche auquel la fonction de similarité permet de comparer n'importe quelle image.

Nous pouvons distinguer deux types de méthodes géométriques de bouclage de pertinence. La première correspond à la mise à jour de la requête alors que la seconde correspond à la modification de la mesure de similarité (à partir des images annotées par l'utilisateur). Ces deux approches sont parfois utilisées conjointement comme dans (Rui et Huang, 2000).

La mise à jour de la requête consiste dans le cas d'une requête simple à la déplacer dans l'espace de recherche, ou à la transformer en une requête multiple en prenant des images retournées par l'utilisateur (Wu et al., 2000; Fournier et al., 2002). La modification de la fonction de similarité permet de changer le voisinage de recherche autour de la requête.

Parmi les méthodes géométriques, nous distinguons deux classes de méthodes pour affiner la requête et la fonction de similarité : les méthodes heuristiques et les méthodes optimales.

3.3.1 Méthodes heuristiques

Les méthodes heuristiques sont basées sur une formulation heuristique ou empirique de l'affinement de la requête et de la fonction de similarité.

Le système Mars (Rui et al., 1997) est parmi les premiers systèmes qui ont introduit la boucle de pertinence dans le domaine de recherche d'image. Il propose deux techniques indépendantes pour affiner la requête puis la fonction de similarité. Ainsi, en donnant plus de poids aux images pertinentes qu'aux images non-pertinentes annotées par l'utilisateur, le vecteur requête affiné est obtenu après addition des index des images pertinentes et soustraction des index des images non-pertinentes à l'index de la requête initiale.

De point de vue de l'adaptation de la fonction de similarité, le problème se ramène à modifier la pondération des différentes composantes de la signature. Mars propose que le poids des attributs soit fixé comme l'inverse de l'écart type de ses valeurs sur l'ensemble des images pertinentes. L'idée étant qu'une composante discriminante de l'index est caractérisée par une faible variance sur les images pertinentes.

Heinrichs et al. (2000) introduisent aussi une approche heuristique pour l'adaptation de la fonction de similarité. Elle est basée sur la notion du rang : un attribut discriminant tend à classer les images pertinentes parmi les résultats les plus proches à la requête et inversement pour les images non-pertinentes. Cependant, le poids associé à chaque attribut est choisi inversement proportionnel au rang des images pertinentes et proportionnel au rang des images non-pertinentes. Soit $\tilde{r}_P^{(m)}$ (respectivement $\tilde{r}_N^{(m)}$) le rang moyen calculé sur l'ensemble des images pertinentes annotées (respectivement non-pertinentes), alors le poids $\beta^{(m)}$ associé au m^e attribut est mis à jour comme suit :

$$\beta^{(m)*} = \beta^{(m)} \times \frac{\tilde{r}_N^{(m)}}{\tilde{r}_P^{(m)}}. \quad (3.1)$$

Ces méthodes heuristiques sont généralement simples et intuitives, mais elles n'ont pas de base théorique solide. D'où la proposition des méthodes *optimales*. Cette appellation est due au fait que la requête et la fonction de similarité sont mises à jour pour l'optimisation d'un critère choisi à priori. Nous présentons ci-après les approches optimales de (Peng et al., 1999; Ishikawa et al., 1998; Rui et Huang, 2000; Fournier et al., 2001).

3.3.2 Méthodes optimales

Avec la technique proposée par Peng et al. (1999), un premier pas vers l'optimalité est réalisé. Il s'agit d'une technique probabiliste pour capturer les variables pertinentes d'un index pendant la recherche par similarité. Le système retourne les K plus proches images voisines de la requête suivant une distance euclidienne pondérée. Les poids de cette distance sont adaptés en fonction d'une mesure de pertinence locale des variables. Cette mesure est estimée à partir d'un critère de réduction de l'erreur de classification bayésienne. L'ensemble d'apprentissage est formé par les images pertinentes et non-pertinentes annotées par l'utilisateur.

Approche optimale du système MindReader

Le système MindReader développé par Ishikawa et al. (1998), intègre les deux techniques indépendantes de mise à jour de la requête et de la fonction de similarité de Mars (Rui et al., 1997) en un seul algorithme d'apprentissage basé sur un critère d'optimalité qui repose sur la minimisation des distances entre le vecteur requête et les vecteurs des exemples positifs. La métrique utilisée est une distance quadratique généralisée. Si \mathbf{r} est le vecteur index de la requête et \mathbf{x}_j est celui d'une image cible j , la distance quadratique généralisée $d_j(\mathbf{x}_j)$ vaut :

$$d_j(\mathbf{x}_j) = (\mathbf{r} - \mathbf{x}_j)^T W (\mathbf{r} - \mathbf{x}_j), \quad (3.2)$$

où W est une matrice pleine symétrique permettant d'intégrer la corrélation entre les composantes de l'index. La structure de l'index est à plat.

Supposons que l'utilisateur identifie P images pertinentes parmi celles retrouvées par le système et leur associe un degré de pertinence (π_j pour la j ième image pertinente). Les valeurs de pertinence et de distance sur cet ensemble de P images sont rassemblées sous forme vectorielle, soit $\boldsymbol{\Pi} = [\pi_1, \dots, \pi_P]^t$ et $\mathbf{d} = [d_1, \dots, d_P]^t$. Le problème se ramène ensuite à la minimisation du critère suivant :

$$\mathcal{J} = \boldsymbol{\Pi}^t \times \mathbf{d}, \quad (3.3)$$

sous la contrainte :

$$\det(W) = 1. \quad (3.4)$$

L'utilisation des multiplicateurs de Lagrange a permis aux auteurs de trouver les solutions optimales à ce problème et de déterminer la fonction de mise à jour de W et du vecteur requête \mathbf{r} .

Approche optimale du système Mars

Rui et Huang (2000) proposent une reformulation complète de système Mars sur le même principe d'optimalité du système MindReader. Contrairement à la structure à plat des index adoptée par MindReader, la méthode de Rui repose sur une structure hiérarchique séparant les différents vecteurs attributs. En notant $d^{(m)}(\mathbf{x})$ la distance quadratique généralisée obtenue entre le vecteur requête $\mathbf{r}^{(m)}$ et le vecteur image $\mathbf{x}^{(m)}$ pour le m -ième vecteur attributs, la distance hiérarchique totale $D(\mathbf{x})$ vaut :

$$D(\mathbf{x}) = U(d^{(m)}(\mathbf{x})) \quad \text{avec} \quad d^{(m)}(\mathbf{x}) = (\mathbf{r}^{(m)} - \mathbf{x}^{(m)})^T W^{(m)} (\mathbf{r}^{(m)} - \mathbf{x}^{(m)}). \quad (3.5)$$

où $U(\cdot)$ est la fonction "inter-attribut" combinant les différents vecteurs attributs de nombre M . Rui et al. choisit d'utiliser une combinaison linéaire des distances, c'est à dire que $U(\cdot)$ est caractérisée par le vecteur de paramètres suivants $\mathbf{u} = [u^{(1)}, \dots, u^{(M)}]$. Les $W^{(m)}$ sont les matrices "intra-attribut" symétriques.

Les valeurs de pertinence et de distance sur l'ensemble de P images annotées pertinentes seront $\boldsymbol{\Pi} = [\pi_1, \dots, \pi_P]^t$ et $\mathbf{D} = [D_1, \dots, D_P]^t$, et le problème se ramène ensuite à la minimisation du critère suivant :

$$\mathcal{J} = \Pi^t \times \mathbf{D}, \quad (3.6)$$

sous les contraintes :

$$\sum_{m=1}^M \frac{1}{u^{(m)}} = 1 \text{ et } \det(W^{(m)}) = 1; \quad (1 \leq m \leq M). \quad (3.7)$$

La solution optimale pour le calcul de la matrice de pondération intra-attribut $W^{(m)}$, est identique pour Mars et MindReader. Elle nécessite l'inversion de la matrice de covariance (notée $C^{(m)}$) entre images annotées positives et requête. Dans les conditions normales de fonctionnement de ces systèmes, le nombre d'exemples positifs est inférieur à la taille p_m de l'attribut m . $C^{(m)}$ n'est donc pas inversible. MindReader propose une solution sous-optimale par passage à la pseudo-inverse, toutefois, Rui et Huang remarquent que cette solution conduit à de très faibles performances. Plus particulièrement, ils estiment qu'il est préférable de n'utiliser qu'une matrice de pondération diagonale $W^{(m)}$ correctement estimée, plutôt qu'une matrice pleine mal estimée.

Approche optimale du système RETIN

RETIN (REcherche Et Traque INteractive d'images) est un système développé par Fournier et al. (2001) pour l'indexation et la recherche d'images dans les bases généralistes. L'approche est basée sur la définition d'une requête et d'une fonction de similarité affinées à partir d'interactions avec l'utilisateur.

Deux méthodes ont été développées (Fournier et al., 2002), l'une basée sur une requête simple où seule la fonction de similarité est mise à jour, et l'autre où la requête et la fonction de similarité sont adaptées au cours de la recherche interactive. Les méthodes de mise à jour proposées sont basées sur des règles d'optimisation.

Requête simple La recherche dans ce cas est une recherche-par-l'exemple. L'utilisateur fournit une seule image exemple (requête) au système. Ensuite, à chaque itération (bouclage) de la recherche, le système affiche un ensemble de résultats parmi lesquels l'utilisateur annote les images pertinentes et non-pertinentes. Ceci permet, au fil des itérations, d'accumuler un ensemble d'exemples pertinents et non pertinents.

Le système fonctionne par comparaison d'une image requête à une image cible de la base. La similarité est calculée comme une double somme pondérée :

$$S(\mathbf{r}, \mathbf{x}) = \sum_{m=1}^M \beta^{(m)} \sum_{j=1}^{p_m} \alpha_j^{(m)} f(r_j^{(m)}, x_j^{(m)}) = \sum_{m=1}^M \beta^{(m)} s^{(m)}(\mathbf{r}, \mathbf{x}), \quad (3.8)$$

avec f la fonction de similarité de base pour la comparaison de deux nombres réels et $s^{(m)}(\mathbf{r}, \mathbf{x})$ la similarité intra-attribut m . Les poids $\beta^{(m)}$ (resp. $\alpha_j^{(m)}$) sont des paramètres réels qui gèrent la compétition inter-attribut (resp. intra-attribut).

La mise à jour des paramètres inter-attribut (resp. intra-attribut) en boucle de pertinence est basée sur une optimisation. Pour une image indexée \mathbf{x}_j , le critère à optimiser est l'erreur quadratique entre la similarité désirée y_j (qui vaut 1 si l'image annotée est pertinente et 0 sinon) et la similarité réelle :

$$Err_j = \frac{1}{2}(y_j - S(\mathbf{r}, \mathbf{x}_j))^2. \quad (3.9)$$

L'optimisation du critère est réalisée par une descente de gradient :

$$\beta^{(m)*} = \beta^{(m)} - \mu \frac{\partial Err_j}{\partial \beta^{(m)}}, \quad (3.10)$$

$$\alpha_j^{(m)*} = \alpha_j^{(m)} - \mu \frac{\partial Err_j}{\partial \alpha_j^{(m)}}, \quad (3.11)$$

pour $1 \leq j \leq p_m$ et $1 \leq m \leq M$ et $\mu \in I\!\!R^+$.

Requête multiple La requête multiple (indexée par \mathcal{P}) est composée d'un ensemble d'images "types" représentatives du but de l'utilisateur. En supposant cet ensemble de taille L , on a : $\mathcal{P} = \{\mathbf{r}_l, 1 \leq l \leq L\}$. La requête multiple est aussi appelée *requête composée* ou *prototype*.

Pour une image cible (indexée par \mathbf{x}) et pour une recherche basée sur M vecteurs d'attributs, la mesure de similarité à la requête multiple est calculée comme suit :

$$S(\mathcal{P}, \mathbf{x}) = \sum_{m=1}^M \beta^{(m)} s^{(m)}(\mathcal{P}, \mathbf{x}), \quad (3.12)$$

où $s^{(m)}(\mathcal{P}, \mathbf{x})$ est la similarité requête-cible pour le m -ième espace d'attributs.

Afin de respecter la multimodalité éventuelle de la requête multiple au sein des espaces d'attributs, les auteurs considèrent que chaque image exemple est une source d'information pour le calcul de similarité. Ils excluent toute refonte des signatures des images du prototype sous la forme d'un vecteur requête unique et affiné, contrairement à l'approche géométrique classique de bouclage de pertinence. Ils proposent, en revanche, la fusion des similarités de la cible à chacune des images du prototype (Bloch et Maître, 1994). Dans le cas où l'image cible est proche d'un des exemples du prototype, la fusion favorise la similarité à ce dernier. Si la cible est éloignée de toutes les images du prototype, une similarité consensuelle, de type moyenne, est préférée. Afin d'assurer un tel comportement, les similarités données par chaque image du prototype sont fusionnées, au sein de chaque espace d'attributs, par l'opérateur barycentre défini par Huet et Philipp (1998) :

$$s^{(m)}(\mathcal{P}, \mathbf{x}) = \frac{\sum_{l=1}^L a_{l,m} \cdot s^{(m)}(\mathbf{r}_l, \mathbf{x})}{\sum_{l=1}^L a_{l,m}}, \quad (3.13)$$

où les poids $a_{l,m}$ correspondent aux similarités, soit :

$$s^{(m)}(\mathcal{P}, \mathbf{x}) = \frac{\sum_{l=1}^L s^{(m)}(\mathbf{r}_l, \mathbf{x})^2}{\sum_{k=1}^L s^{(k)}(\mathbf{r}_k, \mathbf{x})}. \quad (3.14)$$

Dans le cadre de la requête multiple, seuls les poids inter-attributs $\beta^{(m)}$ sont modifiés par boucle de pertinence.

3.4 Approches statistiques de bouclage de pertinence

3.4.1 Système PicHunter : Modélisation bayésienne de pertinence

PicHunter (Cox et al., 2000) utilise une règle de Bayes pour prédire l'image cible cherchée par l'utilisateur sachant son action. La recherche n'est pas amorcée par une image exemple.

En effet, la cible est supposée une variable aléatoire notée T (comme *Target*), qui peut prendre comme valeur chaque image \mathbf{x}_i de la base $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. La loi de probabilité associée à T peut être estimée et mise à jour après plusieurs bouclages de pertinence.

Pour toute image \mathbf{x}_i de la base et d'après la loi de Bayes, la probabilité *a posteriori* d'être la cible sachant l'historique est formulée comme suit :

$$P(T = \mathbf{x}_i | H_t) = \frac{P(A_t | T = \mathbf{x}_i, \mathcal{D}_t, H_{t-1}) P(T = \mathbf{x}_i | H_{t-1})}{\sum_{l=1}^N P(A_t | T = \mathbf{x}_l, \mathcal{D}_t, H_{t-1}) P(T = \mathbf{x}_l | H_{t-1})}, \quad (3.15)$$

avec

- A_t action de l'utilisateur à l'instant t ,
- \mathcal{D}_t l'ensemble de N_D images affichées à l'écran à l'instant t ,
- et l'historique $H_t = \{\mathcal{D}_1, A_1, \dots, \mathcal{D}_t, A_t\}$ formé par les images affichées au passé ainsi que les actions de l'utilisateur.

Le processus de calcul de la probabilité *a posteriori* est incrémental. À l'étape initiale, toutes les images sont supposées équiprobables ($P(T = \mathbf{x}_i) = 1/N$).

Le cœur de cette méthode est le terme $P(A_t | T = \mathbf{x}_i, \mathcal{D}_t, H_{t-1})$ appelé *modèle utilisateur* car son objectif est de prédire l'action de l'utilisateur étant donné l'historique \mathcal{D}_t, H_{t-1} et l'hypothèse que \mathbf{x}_i soit la cible qu'il cherche. Une simplification est faite dans (Cox et al., 2000) et le modèle utilisateur est supposé $P(A_t | T = \mathbf{x}_i, \mathcal{D}_t)$. Chaque utilisateur est censé fournir les mêmes réponses dans les mêmes circonstances.

Une particularité de l'approche par boucle de pertinence de PicHunter est la prise en compte de "jugements relatifs" : si deux images \mathbf{x}_1 et \mathbf{x}_2 sont affichées à l'écran ($\mathcal{D}_t = \{\mathbf{x}_1, \mathbf{x}_2\}$), l'utilisateur précise laquelle est la plus proche de son but. Il n'y a donc que deux actions possibles : choisir la première image (action $A = 1$) ou choisir la seconde image ($A = 2$). Le modèle utilisateur s'écrit alors comme suit :

$$P_{sigmoid}(A = 1 | T = \mathbf{x}_i, \{\mathbf{x}_1, \mathbf{x}_2\}) = \frac{1}{1 + \exp((d(\mathbf{x}_1, \mathbf{x}_i) - d(\mathbf{x}_2, \mathbf{x}_i)) / \sigma)}. \quad (3.16)$$

La distance entre deux images \mathbf{x}_1 et \mathbf{x}_2 est une distance pondérée :

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sum_{m=1}^M w^{(m)} d^{(m)}(\mathbf{x}_1, \mathbf{x}_2), \quad (3.17)$$

avec $d^{(m)}$ est une distance L_1 pour le m^e attribut. Les poids $w^{(m)}$ et σ sont fixés grâce à des tests psychophysiques.

A partir de la probabilité *a posteriori* construite sur la base entière, le système sélectionne les images à afficher. Plusieurs options sont alors possibles. En premier lieu, le système peut choisir les images les plus probables. Cette stratégie est sous optimale car elle ne permet pas d'explorer la base, et se concentre, au contraire dans une zone particulière de cette dernière. En second lieu, un schéma de sélection ("Most-Informative Display Updating scheme") sous contrainte de minimisation de l'entropie est introduit afin d'explorer toute la base. Le but du schéma est d'obtenir le plus d'informations possible de l'utilisateur pendant l'exploitation de ces informations pour terminer rapidement la recherche.

3.4.2 Modélisation non-paramétrique de pertinence

Meilhac et Nastar (1999) proposent de modéliser séparément la pertinence et la non-pertinence des images de la base, à partir des images respectivement annotées pertinentes et non-pertinentes par l'utilisateur.

Pour chaque image \mathbf{x} de la base, deux densités de probabilité non-paramétriques (une pour la pertinence $f(\mathcal{P}/\mathbf{x})$ et une pour la non-pertinence $f(\mathcal{N}/\mathbf{x})$) sont estimées par la méthode des fenêtres de Parzen.

En effet, soient :

- $\mathbf{x} = (x_j, j = 1, \dots, p)$ le vecteur index de p composantes d'une image de la base (structure à plat),
- P_l est l'ensemble des indices i des images étiquetées pertinentes par l'utilisateur,
- N_l est l'ensemble des indices des images étiquetées non-pertinentes.

La fonction de lissage de l'estimateur de Parzen choisie est gaussienne, c'est à dire :

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \quad (3.18)$$

La densité $f(\mathbf{x}/\mathcal{P})$ de Parzen sur chaque composante j est :

$$G_{\theta_j^{P_l}}(x) = \sum_{i \in P_l} g_{\sigma_j(i)}(x - \mu_j(i)), \quad (3.19)$$

avec $\theta_j^{P_l} = (\mu_j(i), \sigma_j(i))_{i \in P_l}$ est l'ensemble de paramètres des images pertinentes suivant la composante j .

En supposant que les composantes sont indépendantes entre elles, nous aurons :

$$f(\mathbf{x}/\mathcal{P}) = \prod_{j=1}^p G_{\theta_j^{P_l}}(x). \quad (3.20)$$

De la même façon, la densité $f(\mathbf{x}/\mathcal{N})$ de Parzen :

$$f(\mathbf{x}/\mathcal{N}) = \prod_{j=1}^p G_{\theta_j^{N_l}}(x_j), \quad (3.21)$$

avec $\theta_j^{N_l}$ est l'ensemble de paramètres des images non-pertinentes suivant l'attribut j .

Une règle de décision est ensuite appliquée afin de décider de la pertinence globale de l'image :

$$J(\mathbf{x}) = -\log(f(\mathcal{P}/\mathbf{x})) + \log(f(\mathcal{N}/\mathbf{x})). \quad (3.22)$$

En appliquant les règles de Bayes sous certaines hypothèses, les auteurs ont pu utiliser cette règle de décision :

$$\begin{aligned} I_\theta(\mathbf{x}) &= -\log(f(\mathbf{x}/\mathcal{P})) + \log(f(\mathbf{x}/\mathcal{N})) \\ &= \sum_{j=1}^p (-\log(G_{\theta_j^{P_l}}(x_j)) + \log(G_{\theta_j^{N_l}}(x_j))). \end{aligned} \quad (3.23)$$

Pour la méthode d'estimation des paramètres $\theta = (\theta_i^{P_l}, \theta_i^{N_l})$, chaque image étiquetée est considérée comme centre d'une gaussienne :

$$\forall i \in P_l \cup N_l, \mu_j(i) = x_{ij}. \quad (3.24)$$

Ceci crée une zone d'influence autour de l'exemple. Dans les espaces des composantes, les images voisines d'un exemple pertinent voient leur pertinence globale augmenter et inversement pour les images voisines d'un exemple non-pertinent.

En outre, le paramètre de largeur de la gaussienne choisi (l'écart type σ) est inversement proportionnel au nombre d'images étiquetées de la même classe :

$$\forall i \in P_l, \sigma_j(i) = \frac{\sigma_j^{\mathcal{DTB}}}{\log(\text{cardinal}(P_l))}, \quad (3.25)$$

$$\forall i \in N_l, \sigma_j(i) = \frac{\sigma_j^{\mathcal{DTB}}}{\log(\text{cardinal}(N_l))}, \quad (3.26)$$

$\sigma_j^{\mathcal{DTB}}$ étant l'écart type de la composante j calculée sur toute la base.

Durant les premières itérations, ce nombre est faible, σ est grand et les densités de probabilité sont peu informatives. Au fil de la recherche, le nombre d'exemples croît, σ diminue et la densité s'affine en laissant émerger ce que l'algorithme identifie comme les résultats pertinents.

Une particularité de cette méthode est que les densités de probabilité construites sont multimodales. Dans le cas des densités uni-modales gaussiennes, I_θ aura cette expression simple (avec la même hypothèse d'indépendance entre composantes) :

$$I_\theta(\mathbf{x}) = \frac{1}{2} \sum_{j=1}^p \left(\left(\frac{x_j - \mu_j^P}{\sigma_j^P} \right)^2 - \left(\frac{x_j - \mu_j^N}{\sigma_j^N} \right)^2 \right), \quad (3.27)$$

où $\mu^P = (\mu_j^P, j = 1, \dots, p)$ et $\sigma^P = (\sigma_j^P, j = 1, \dots, p)$ sont respectivement les vecteurs moyenne et écart-type de la gaussienne modélisant la pertinence (μ^N et σ^N de la gaussienne modélisant la non-pertinence).

3.4.3 Approche par inférence bayésienne

Après une étape de *représentation de l'image* par un mélange de gaussiennes, Vasconcelos et Lippman (2000b) traitent le problème de recherche d'image comme un problème d'inférence bayésienne.

En effet, ayant un espace de représentation \mathcal{F} des images de la base formé par K classes,

$$\begin{aligned} g : \mathcal{F} &\longrightarrow M = \{1, \dots, K\} \\ \mathbf{x} &\longrightarrow y, \end{aligned}$$

le système de recherche consiste à trouver la classe y de l'image requête de signature \mathbf{x} . La fonction de décision est la vraisemblance *a posteriori* (classifieur bayésien):

$$g^*(\mathbf{x}) = \arg \max_k P(\mathbf{x}|y = k). \quad (3.28)$$

Dans le cas d'un mélange de C_k gaussiennes, $P(\mathbf{x}|y = k)$ s'écrit :

$$P(\mathbf{x}|y = k) = \sum_{c_k=1}^{C_k} \pi_{c_k} \frac{1}{\sqrt{(2\pi)^p |\Sigma_{c_k}|}} e^{-\frac{1}{2}(\mathbf{x}-\mu_{c_k})^T \Sigma_{c_k}^{-1} (\mathbf{x}-\mu_{c_k})}. \quad (3.29)$$

Le calcul des paramètres $(\mu_{c_k}, \Sigma_{c_k})$ des gaussiennes est réalisé hors ligne pendant la phase d'indexation. Ensuite la boucle de pertinence a pour rôle d'augmenter la performance de l'inférence bayésienne. En utilisant l'ensemble des images annotées positives par l'utilisateur $\mathbf{x}^t = \{\mathbf{x}_1, \dots, \mathbf{x}_t\}$ comme une séquence de t requêtes (t est un laps de temps), la fonction de similarité est mise à jour comme suit :

$$g^*(\mathbf{x}^t) = \arg \max_k \{\alpha \log P(\mathbf{x}_t|y = k) + (1 - \alpha) \log P(y = k|\mathbf{x}_1, \dots, \mathbf{x}_{t-1})\}, \quad (3.30)$$

où $\alpha \in (0, 1]$ est intégré dans cette règle de mise à jour dans le but d'affecter un degré de pertinence à l'image annotée à l'instant t plus grand que celui des images annotées jusqu'à l'instant $t - 1$.

Pour tenir compte des images annotées négatives une nouvelle version de l'équation 3.30 est introduite dans (Vasconcelos et Lippman, 2000a).

3.4.4 Approches par classification

L'approche par classification aborde le problème de bouclage de pertinence sous l'angle de la séparation des images pertinentes et non-pertinentes dans l'espace des attributs. Plus généralement, étant donné un certain nombre d'exemples pertinents ou non-pertinentes, il s'agit de construire le classifieur permettant de grouper au mieux les données pertinentes et de les discriminer des données non-pertinentes.

Classification par quantification vectorielle (LVQ)

Les travaux de Wood et al. (1998) sont parmi les premiers à traiter le problème du bouclage de pertinence sous l'angle de classification. Leur système travaille sur des régions (la signature est formée par un vecteur d'attributs calculés sur une région) et se base sur une requête simple.

Une quantification vectorielle est effectuée afin de déterminer automatiquement des classes (clusters) au sein de l'espace des signatures des images de la base. Ensuite, une séparation des classes pertinentes et non-pertinentes est effectuée à partir des images annotées pertinentes ou non-pertinentes par l'utilisateur.

Ce système de classification est relativement complexe et ne prend en compte qu'indirectement les exemples annotées puisque la quantification vectorielle est réalisée d'une manière non-supervisée.

Suivant le même principe de classification, mais cette fois-ci avec une carte auto-organisatrice (self-organizing map SOM), Laaksonen et al. (1999), dans leur système Pic-SOM, utilisent le TS (Tree-Structured)-SOM pour supporter plusieurs vecteurs d'attributs de l'image en parallèle (structure hiérarchique de l'index).

Technique Boosting

Tieu et Viola (2000) utilisent plus de 45,000 "attributs fortement sélectifs" et une technique boosting pour estimer une fonction de discrimination dans cet espace de variables.

Ainsi, pour chaque variable, un classifieur "faible" (weak classifier) est formulé en supposant un modèle gaussien pour les exemples positifs, annotés par l'utilisateur, et les exemples négatifs choisis au hasard. L'attribut pour lequel l'hypothèse gaussienne des deux classes est la plus efficace (taux d'erreur de classification est minimal) est sélectionné. Cet algorithme d'apprentissage est répété $T = 20$ fois. Ensuite le classifieur "fort" (strong classifier) est une combinaison linéaire pondérée des T classificateurs "faibles" comme dans la technique d'apprentissage AdaBoost (Freund et Schapire, 1999).

Techniques par les machines à vecteurs de support

Depuis quelques années, la technique des machines à vecteurs de support (SVM) (Vapnik, 1995), connue en analyse discriminante, est au centre des préoccupations des chercheurs du domaine pour ses bonnes capacités de généralisation dans le cas de classification dans un espace de grande dimension à partir de peu d'exemples.

Hong et al. (2000) et Zhang et al. (2001) intègrent la technique SVM dans la recherche d'images par bouclage de pertinence. Ils utilisent les SVM afin de déterminer la frontière entre les images pertinentes et non-pertinentes. Si nous appelons $\{(\mathbf{x}_i, y_i), i = 1, \dots, l\}$ l'ensemble d'apprentissage ($y_i \in \{-1, 1\}$ suivant que l'image est non-pertinente ou pertinente resp.), alors la mesure d'appartenance J d'une image de signature \mathbf{x} à la classe des images pertinentes s'écrit (voir annexe B) :

$$J(\mathbf{x}) = \sum_{i=1}^l \alpha_i y_i \mathcal{K}(\mathbf{x}_i, \mathbf{x}) + \beta_0, \quad (3.31)$$

où β_0 est une constante et les poids α_i sont calculés par une optimisation quadratique sous contraintes et sont déterminés pendant la phase d'apprentissage. \mathcal{K} est une fonction appelée *noyau* permettant des séparations non-linéaires des données. Zhang et al. (2001) ont utilisé le noyau gaussien ($\mathcal{K}(\mathbf{x}, \mathbf{x}_i) = e^{-\|\mathbf{x} - \mathbf{x}_i\|^2/2\sigma^2}$ appelé encore *Radial Basis function RBF*), et dans (Hong et al., 2000) le noyau polynomial a été choisi ($(\mathbf{x}^t \cdot \mathbf{x}_i + 1)^d, d = 1$).

Vu le petit nombre d'exemples annotés par l'utilisateur, l'ensemble d'apprentissage ne peut pas représenter la vraie distribution des deux classes des images pertinentes et non-pertinentes (Chen et al., 2001). Si l'hypothèse d'existence d'une structure de classe, par exemple gaussienne, est acceptable pour les images pertinentes, il n'en est pas de même pour les images non-pertinentes. En effet, les contre-exemples se répartissent sur plusieurs classes et le nombre d'échantillons disponibles est trop limité pour espérer apprendre correctement la structure sous-jacente. C'est la raison pour laquelle certains auteurs présentent la question comme un problème de classification biaisée (Zhou et Huang, 2001).

Cependant, Chen et al. (2001) contournent le problème des exemples négatifs en proposant une version mono-classe des SVM. L'idée est de modéliser la classe des images pertinentes au lieu de résoudre le problème de séparation des données pertinentes et non-pertinentes. En effet, les auteurs construisent le plus petit support (hyper-sphère minimale) qui contient la plupart des images pertinentes dans l'espace index après transformation non-linéaire.

Si l'ensemble d'apprentissage formé par les images annotées pertinentes par l'utilisateur est $\{\mathbf{x}_1, \dots, \mathbf{x}_l\}$, le problème d'optimisation est traduit par les équations suivantes :

$$\min_{\rho, \xi_i \in \mathbb{R}, \mathbf{c}} \rho^2 + \frac{1}{v.l} \sum_i \xi_i, \quad (3.32)$$

sous les contraintes :

$$\|\phi(\mathbf{x}_i) - \mathbf{c}\|^2 \leq \rho^2 + \xi_i, \quad \xi_i \geq 0 \text{ pour } i = 1, \dots, l. \quad (3.33)$$

$v \in [0, 1]$ et ρ est le rayon de l'hyper-sphère. \mathbf{c} est le centre de l'hyper-sphère. ξ_i est l'erreur permise pour relâcher les contraintes sur la frontière de l'hypersphère (voir annexe B). ϕ est une transformation non-linéaire de l'espace : $\mathbf{x} \rightarrow \phi(\mathbf{x})$.

La mesure d'appartenance J d'une image de signature \mathbf{x} à la classe des images pertinentes s'écrit alors :

$$J(\mathbf{x}) = \rho^2 - \sum_{i,j} \alpha_i \alpha_j \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) + 2 \sum_i \alpha_i \mathcal{K}(\mathbf{x}_i, \mathbf{x}) - \mathcal{K}(\mathbf{x}, \mathbf{x}), \quad (3.34)$$

Les poids α_i ainsi que ρ sont calculés après résolution du problème d'optimisation quadratique (cf. équations 3.33,3.32) par les multiplicateurs de Lagrange. Dans le cas non-linéaire la fonction gaussienne \mathcal{K} est choisie comme noyau. Les tests réalisés montrent que le classifieur SVM mono-classe s'adapte aux distributions *multimodales* des données positives.

Analyse Discriminante Biaisée

Pour ne pas exclure des images annotées non-pertinentes, Zhou et Huang (2001) introduisent une technique appelée Biased Discriminant Analysis (BDA - Analyse Discriminante Biaisée). Le problème est de déterminer l'espace discriminant optimal dans lequel les images pertinentes sont regroupées ensemble le plus loin possible des images non-pertinentes. Mathématiquement parlant, le problème posé est de trouver la transformation optimale tel que le rapport de "la matrice de dispersion des exemples négatifs vis-à-vis du centre des exemples positifs" sur "la matrice de dispersion intra-classe des exemples positifs" soit maximum. D'où l'appellation d'une analyse discriminante biaisée vu le traitement effectué autour des exemples positifs.

Ainsi, si $\{\mathbf{x}_i, i \in P_l\}$ et $\{\mathbf{x}_i, i \in N_l\}$ sont respectivement les exemples positifs et négatifs annotées par l'utilisateur, le critère biaisé défini est écrit comme suit :

$$W^* = \arg \max_W \left| \frac{W^t S_{N_l} W}{W^t S_{P_l} W} \right|, \quad (3.35)$$

où

$$S_{N_l} = \sum_{i \in N_l} (\mathbf{x}_i - \mathbf{m}_{P_l})(\mathbf{x}_i - \mathbf{m}_{P_l})^t, \quad (3.36)$$

$$S_{P_l} = \sum_{i \in P_l} (\mathbf{x}_i - \mathbf{m}_{P_l})(\mathbf{x}_i - \mathbf{m}_{P_l})^t. \quad (3.37)$$

\mathbf{m}_{P_l} est le vecteur moyenne des exemples positifs. La solution optimale est calculée de la même façon que la technique classique de l'analyse discriminante de Fisher.

A cause de la faible taille de l'ensemble d'apprentissage (images annotées) par rapport à la dimension p de l'index, les matrices de dispersion des équations 3.36 et 3.37 sont généralement biaisées. Les auteurs proposent une régularisation de ces matrices comme l'a introduit Friedman (1989). En effet, il s'agit d'ajouter une faible quantité à la diagonale des matrices :

$$S_{P_l}^r = (1 - \mu)S_{P_l} + \frac{\mu}{p} \text{trace}(S_{P_l})I, \quad (3.38)$$

$$S_{N_l}^d = (1 - \gamma)S_{N_l} + \frac{\gamma}{p} \text{trace}(S_{N_l})I, \quad (3.39)$$

avec p est la dimension de l'espace d'origine et I la matrice identité. μ et γ sont respectivement les facteurs de régularisation et de "discounting". Le choix de la combinaison (μ, γ) n'est pas indiqué dans (Zhou et Huang, 2001).

Les tests comparatifs permettent aux auteurs de montrer l'apport de leur méthode par rapport à la technique de l'analyse discriminante de Fisher.

Pour tenir en compte la non-linéarité des données, une approche basée sur les noyaux de Kernel \mathcal{K} (voir annexe B) est aussi proposée (KBDA : Kernel-Based Biased Discriminant Analysis). Si ϕ est une transformation non-linéaire de l'espace : $\mathbf{x} \rightarrow \phi(\mathbf{x})$, le critère de l'équation 3.35 devient :

$$W^* = \arg \max_W \left| \frac{W^t S_{N_l}^\phi W}{W^t S_{P_l}^\phi W} \right|. \quad (3.40)$$

L'exposant ϕ est utilisé pour dénoter les quantités dans le nouvel espace.

Une comparaison entre KBDA et BDA a été aussi faite (utilisant la même fonction RBF de kernel) pour tester la capacité des noyaux à traiter la distribution non-linéaire des données. En outre, des tests comparatifs entre KBDA et SVM, ont montré l'apport de l'approche KBDA dans le cas d'un petit nombre d'exemples négatifs annotés.

3.4.5 Approche par une classification semi-supervisée

Wu et al. (2000) traitent la boucle de pertinence dans le cadre d'un apprentissage semi-supervisé. Ils introduisent un algorithme EM modifié, appelé D-EM (Discriminant- EM), qui a comme ensemble d'apprentissage les images annotées (positives et négatives) ainsi que les images non-annotées par l'utilisateur en bouclage de pertinence.

D-EM exécute l'analyse discriminante de Fisher à l'intérieur des itérations de l'algorithme EM afin de choisir le sous-espace de variables telle que l'hypothèse des deux classes sur les distributions de données ait un meilleur support.

Ainsi, l'algorithme D-EM classe les données après les avoir projetées dans le sous-espace de dimension $(K - 1)$, où $K = 2$ est le nombre de classes considérées. Wu et al. (2000) avancent que cette projection permet de s'abstraire de l'hypothèse gaussienne des classes dans l'espace de départ.

Les auteurs montrent via leurs tests expérimentaux que la combinaison des exemples annotés et non-annotés réduit l'erreur de classification.

Il est à noter que, dans cette approche, les index sont projetés sur une droite ce qui simplifie donc à l'extrême la structure de la base d'index. Ainsi, ce type de projection ne permet pas de modéliser des structures de classes complexes telle qu'une configuration où la classe des images pertinentes serait entourée par la classe des images non pertinentes.

3.5 Conclusion

Les méthodes géométriques reposent sur le choix *a priori* d'une fonction de similarité et d'un critère d'optimisation pour la mise à jour de cette fonction en boucle de pertinence. En outre, l'approche géométrique est moins adaptée à la recherche d'une catégorie

d'images *multimodales* sauf dans le cas de l'utilisation d'une requête multiple (Chakrabarti et al., 2000; Fournier et al., 2002). Par ailleurs, la séparation des deux mises à jour de la requête et la fonction de similarité est une tâche compliquée.

Par contre, la recherche d'une catégorie d'images par bouclage de pertinence est un problème de classification particulier. Il est caractérisé par un ensemble d'apprentissage de petite taille (exemples pertinents et non-pertinents). Cependant, le faible nombre des exemples non-pertinents ne permet pas d'estimer correctement la distribution de la classe correspondante. Les approches statistiques présentées dans le paragraphe 3.4 utilisent les exemples annotés seulement, sauf Wu et al. (2000) qui intègrent les images non-annotées dans la phase d'apprentissage. Mais comme nous avons déjà expliqué (*cf.* sous section 3.4.5) leur méthode n'est pas adaptée pour des classes de structure complexe.

Dans le chapitre suivant, nous présentons l'avantage de l'intégration des exemples non classés dans l'apprentissage et l'apport de la classification semi-supervisée à celle supervisée dans le cas d'un faible nombre d'exemples étiquetés, grâce à une méthode d'apprentissage semi-supervisée basée sur les modèles de mélange et l'algorithme EM. Dans le chapitre 5 nous appliquons cette approche au problème de recherche d'images par boucle de pertinence. Nous la comparons par la suite à une autre approche supervisée basée sur les machines à vecteurs de support.

Chapitre 4

Apprentissage semi-supervisé avec les modèles de mélange et l'algorithme EM

La théorie de l'apprentissage statistique joue un rôle très important dans de nombreux domaines tels : les bases de données, la reconnaissance des formes, l'intelligence artificielle.

Dans le contexte de la reconnaissance des formes, apprendre d'un ensemble de couples (x,y) , c'est chercher à prédire y à partir de x ; y pouvant être une valeur qualitative (comme par exemple normal/anormal pour un organe) ou quantitative (la quantité de pluie du mois prochain). Dans la littérature, le problème de prédiction est appelé "régression" quand la valeur à prédire est quantitative et "classification" si cette valeur est qualitative (catégorie). Ces deux problèmes distincts peuvent être vus comme un problème d'estimation d'une fonction $f(x,y)$ (Hastie et al., 2001).

Dans ce chapitre, nous étudions le problème de l'apprentissage semi-supervisé d'un ensemble de données supposées générées par un modèle de mélange, pour effectuer une classification après estimation des densités de probabilité du modèle par l'algorithme EM.

La première section est consacrée à un état de l'art des méthodes de classification semi-supervisée. Ensuite, nous décrivons notre approche paramétrique et l'algorithme d'estimation associé (une application de l'algorithme EM). Dans la dernière section, l'apprentissage semi-supervisé est comparé à l'apprentissage supervisé dans le cas de deux classes. Nous montrons que lorsqu'un petit nombre de données classées (étiquetées) est disponible, l'intégration des exemples non classés dans l'apprentissage améliore la classification.

4.1 Notations et terminologies

Soit \mathcal{X} l'ensemble d'apprentissage constitué de N observations et noté $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. \mathbf{x}_i est un vecteur de p éléments réels, noté comme suit : $\mathbf{x}_i = (x_{ij}, j = 1, \dots, p)$. Notons \mathbf{x}_i^t le transposé de \mathbf{x}_i .

Notre problème consiste à estimer les paramètres d'un classifieur \mathcal{C} qui classe les observations \mathcal{X} en K classes :

$$\mathcal{C} : \mathcal{X} \longrightarrow \mathcal{Y},$$

où $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$. $\mathbf{y}_i = (y_{ik}, k = 1, \dots, K)$ est un vecteur binaire indiquant la classe de \mathbf{x}_i tel que :

$y_{ik} \in \{0, 1\}$ et $\sum_{k=1}^K y_{ik} = 1$. $y_{ik} = 1$ signifie que \mathbf{x}_i provient de la $k^{i\text{ème}}$ classe (notée C_k).

Dans ce chapitre nous noterons \mathbf{X} la matrice de dimension $N \times p$ formée par les N vecteurs \mathbf{x}_i de \mathcal{X} :

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^t \\ \mathbf{x}_2^t \\ \vdots \\ \mathbf{x}_N^t \end{pmatrix} \text{ ou encore } \mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Np} \end{pmatrix}.$$

La matrice $\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1^t \\ \vdots \\ \mathbf{y}_N^t \end{pmatrix}$ définie par les N vecteurs \mathbf{y}_i est appelée matrice de classification .

Distinguons quatre problèmes d'apprentissage :

1. apprentissage supervisé,
2. apprentissage non-supervisé,
3. apprentissage semi-supervisé,
4. apprentissage partiellement supervisé.

Comme son nom indique, l'apprentissage supervisé implique l'estimation des paramètres du classifieur \mathcal{C} à partir de l'ensemble \mathcal{X} des données déjà classées ($\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ connu). Le but est de prédire la classe d'une nouvelle observation \mathbf{x} . Un exemple numérique est présenté dans le tableau 4.1.

En pratique, il n'est pas toujours possible d'avoir les classes des N observations $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. Dans le cas où les classes des N observations sont inconnues, on se trouve devant un problème d'apprentissage non-supervisé (appelé en anglais clustering) où on doit estimer les paramètres du classifieur ainsi que $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ (cf. table 4.2).

L'apprentissage semi-supervisé se sert d'un ensemble de données formé par deux sous-ensembles. Un sous-ensemble de m observations de classe connue (\mathbf{y}_i connue) et un autre de $N - m$ individus non classés comme illustré dans le tableau 4.3.

| X | | | | Y | | | |
|----------|-----|-----|-----|----------|---|---|----------------------------------|
| 3.5 | 2.3 | 0.3 | 4.2 | 0 | 0 | 1 | 1 |
| 2.2 | 1.4 | 2.9 | 1.3 | 0 | 1 | 0 | |
| 4.2 | 1.7 | 2.2 | 1.1 | 0 | 0 | 1 | : |
| 2.5 | 2.3 | 0.3 | 4.2 | 0 | 0 | 1 | |
| 9.2 | 2.4 | 2.9 | 1.3 | 1 | 0 | 0 | |
| 6.2 | 1.2 | 2.2 | 1.1 | 1 | 0 | 0 | <i>N</i> |
| 4.4 | 2.4 | 2.1 | 1.0 | ? | ? | ? | ← nouvelle observation à classer |

TAB. 4.1 – Exemple d'apprentissage supervisé: cas de trois classes

| X | | | | Y | | | |
|----------|-----|-----|-----|----------|---|---|----------|
| 3.5 | 2.3 | 0.3 | 4.2 | ? | ? | ? | 1 |
| 2.2 | 1.4 | 2.9 | 1.3 | ? | ? | ? | |
| 4.2 | 1.7 | 2.2 | 1.1 | ? | ? | ? | : |
| 2.5 | 2.3 | 0.3 | 4.2 | ? | ? | ? | |
| 9.2 | 2.4 | 2.9 | 1.3 | ? | ? | ? | |
| 6.2 | 1.2 | 2.2 | 1.1 | ? | ? | ? | <i>N</i> |

TAB. 4.2 – Exemple d'apprentissage non-supervisé

| X | | | | Y | | | |
|----------|-----|-----|-----|----------|---|---|--------------|
| 3.5 | 2.3 | 0.3 | 4.2 | 0 | 0 | 1 | 1 |
| 2.2 | 1.4 | 2.9 | 1.3 | ? | ? | ? | ← non classé |
| 4.2 | 1.7 | 2.2 | 1.1 | 0 | 0 | 1 | |
| 2.5 | 2.3 | 0.3 | 4.2 | ? | ? | ? | : |
| 9.2 | 2.4 | 2.9 | 1.3 | ? | ? | ? | |
| 6.2 | 1.2 | 2.2 | 1.1 | 1 | 0 | 0 | <i>N</i> |

TAB. 4.3 – Exemple d'apprentissage semi-supervisé

Finalement, ayant un ensemble d'observations où m individus parmi les N sont de classe connue et l individus sont connus *pour ne pas appartenir à certaines classes*. Les $N - l - m$ observations sont non classées. L'apprentissage est ainsi appelé partiellement supervisé (voir exemple dans la table 4.4). A noter que ce type d'apprentissage est applicable quand on a au moins trois classes.

| X | | | | Y | | | |
|-----|-----|-----|-----|---|---|---|---|
| 3.5 | 2.3 | 0.3 | 4.2 | 0 | 0 | 1 | 1 |
| 2.2 | 1.4 | 2.9 | 1.3 | 0 | ? | ? | ← n'appartient pas à la première classe |
| 4.2 | 1.7 | 2.2 | 1.1 | 0 | 0 | 1 | ← classé |
| 2.5 | 2.3 | 0.3 | 4.2 | ? | ? | ? | ⋮ ← non classé |
| 9.2 | 2.4 | 2.9 | 1.3 | ? | ? | ? | |
| 6.2 | 1.2 | 2.2 | 1.1 | 1 | 0 | 0 | N |

TAB. 4.4 – *Exemple d'apprentissage partiellement supervisé*

4.2 Classification semi-supervisée

Dans cette section, un état de l'art des méthodes de classification semi-supervisée est présenté. La semi-supervision intervient lorsque l'on dispose à la fois d'un ensemble de données étiquetées et non étiquetées. Généralement, les données non supervisées sont disponibles en grand nombre car peu coûteuses à produire. Au contraire, les données supervisées qui nécessitent l'expertise humaine sont plus rares mais également plus riches en informations. On distingue deux approches duales de la semi-supervision suivant le point de départ où on se place :

- utilisation d'un grand volume de données non étiquetées pour améliorer un apprentissage supervisé,
- utilisation de la donnée supervisée pour aider un algorithme de classification automatique (clustering).

L'objectif de notre travail se situe dans le cadre du premier point. En effet, nous disposons d'un très petit nombre d'exemples étiquetés et nous souhaitons améliorer notre classification par l'intégration des données non étiquetées disponibles.

Concernant le deuxième point, on peut considérer que l'on ne dispose originellement d'aucune donnée étiquetée. Dans ce cas, on est contraint d'utiliser un algorithme de classification non supervisé et l'information supervisée est intégrée au fur et à mesure de sa disponibilité. Saint-Jean et Frélicot (2001) proposent une méthode paramétrique de classification semi-supervisée avec rejet afin d'obtenir une classification automatique robuste. L'intérêt de cette méthode est illustré dans le cas d'un jeu de données de type XOR (OU-exclusif) ou contenant des données aberrantes (outliers). Demiriz et al. (1999) ont développé une approche de classification automatique (clustering) semi-supervisée utilisant des algorithmes génétiques.

4.2.1 Approches géométriques de la semi-supervision

Pedrycz (1985) a introduit la notion de supervision partielle en modifiant l'algorithme des C-moyennes floues (Fuzzy C-means FCM). Le principe de la méthode des C-moyennes floues est de minimiser itérativement un critère \mathcal{J} . Ce critère est une fonction du degré d'appartenance d'un exemple à une classe, de sa distance à cette dernière, et d'un coefficient de flou :

$$\begin{aligned} \mathcal{J}_0 &= \sum_{k=1}^K \sum_{i=1}^N u_{ik}^m d_{ik}^2, \\ \text{avec } \sum_{k=1}^K u_{ik} &= 1 \forall i \in \{1, \dots, N\}. \end{aligned} \quad (4.1)$$

u_{ik} et d_{ik} désignent respectivement le degré d'appartenance et la distance de x_i à la classe C_k . m est le coefficient de flou.

La version semi-supervisée de cet algorithme proposée par Pedrycz (ssFCM) consiste à ajouter un terme de supervision \mathcal{J}_s à ce critère. Dans cette version, le coefficient de flou est fixé à 2 :

$$\mathcal{J} = \alpha_0 \mathcal{J}_0 + \alpha_s \mathcal{J}_s = \alpha_0 \sum_{k=1}^K \sum_{i=1}^N u_{ik}^2 d_{ik}^2 + \alpha_s \sum_{k=1}^K \sum_{i=1}^N (u_{ik} - f_{ik} b_i)^2 d_{ik}^2. \quad (4.2)$$

Le second terme \mathcal{J}_s de ce critère représente l'information liée à la semi-supervision. Le facteur b_i désigne si x_i est étiqueté ($b_i = 1$) ou non ($b_i = 0$). Dans le cas où le point x_i est étiqueté, l'élément f_{ik} est le degré d'appartenance fixé par l'expert. En revanche, il est impossible de prendre en compte l'information inverse: "Cet exemple n'appartient pas à cette classe". En effet, on peut considérer que l'information de supervision pour x_i n'est complète que si tous les f_{ik} sont entièrement donnés par l'expert. Les coefficients α_0 et α_s permettent d'effectuer un compromis entre les points supervisés et non supervisés. Pedrycz propose de fixer les valeurs de α_0 et α_s par $1/n_u$ et $1/n_d$ respectivement où n_u et n_d désignent le nombre de points non supervisés et supervisés.

Le point important de la ssFCM de Pedrycz est qu'elle recherche une partition (matrice des degrés d'appartenance) de l'ensemble des données, y compris sur des données supervisées. Bensaid et al. (1996) proposent une variante semi-supervisée de FCM. Ils proposent de restreindre la recherche de partition aux données non supervisées. La matrice de partition des données supervisées est fixée par l'expert. Cette méthode a été proposée pour améliorer la segmentation par classification non supervisée d'images IRM. Une autre approche semi-supervisée de classification avec sur-partitionnement a été développée (Bensaid et Bezdek, 1998; Amar et al., 1997) afin de produire une partition possibiliste. L'idée sous-jacente de cette méthode consiste à associer, à chaque point de l'ensemble, la classe du point supervisé le plus proche.

4.2.2 Machines à vecteurs de support semi-supervisées

Bien que les machines à vecteurs de support soient initialement fondées sur un apprentissage totalement supervisé, plusieurs versions semi-supervisées (S^3VM) ont été proposées ces dernières années (Bennett et Demiriz, 1998; Fung et Mangasarian, 2001; Demiriz et Bennett, 2001). Ces versions ont été développées seulement dans le cadre d'une classification binaire (cas de deux classes) linéaire.

L'idée de Bennett et Demiriz (1998) est de construire la meilleure machine pour les données supervisées tout en minimisant le nombre de données non étiquetées dans la marge (entre les deux hyperplans). Pour un exemple \mathbf{x}_i , être dans la marge implique (voir annexe B) $|\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0| < 1$. Soient τ_i et v_i les termes d'erreurs qui permettent aux exemples non supervisés placés respectivement dans la classe positive et négative d'être mal placés (par exemple à l'intérieur de la marge, voir figure B.1.b de l'annexe B) :

- si le point \mathbf{x}_i est supposé dans la classe positive, il est en dehors de la marge si : $(\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0) + \tau_i \geq 1$,
- si le point \mathbf{x}_i est supposé dans la classe négative, il est en dehors de la marge si : $-(\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0) + v_i \geq 1$.

L'objectif est de minimiser ces termes d'erreurs. Le critère à minimiser peut s'écrire alors :

$$\min_{\boldsymbol{\beta}, \beta_0, \xi, \tau, v} \|\boldsymbol{\beta}\| + \gamma \left[\sum_{i=1}^{n_d} \xi_i + \sum_{i=n_d+1}^{n_d+n_u} \min(\tau_i, v_i) \right],$$

sous la contrainte :

$$\begin{cases} y_i(\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0) + \xi_i \geq 1, & \xi_i \geq 0, \quad i = 1, \dots, n_d, \\ \mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0 + \tau_i \geq 1, & \tau_i \geq 0, \quad i = n_d + 1, \dots, n_d + n_u, \\ -(\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0) + v_i \geq 1, & v_i \geq 0, \quad i = n_d + 1, \dots, n_d + n_u. \end{cases} \quad (4.3)$$

y_i est le label de d'un exemple supervisé \mathbf{x}_i , il est égal à +1 pour les exemples positifs et -1 pour les exemples négatifs. n_u et n_d désignent le nombre de points non supervisés et supervisés respectivement.

La principale difficulté de mise en œuvre des machines à vecteurs de support se situe dans le temps de calcul nécessité par l'optimisation quadratique sous un grand nombre de contraintes. En plus, il n'est pas sûr que l'introduction de données non supervisées et les contraintes qui leur sont associées permette d'améliorer la convergence de la méthode d'optimisation. Bennett et Demiriz (1998) ont montré, sur 10 différentes expérimentations, que S^3VM réalise de meilleures performances sur 3 d'entre elles (par rapport à l'algorithme SVM classique) et ne perturbe pas les résultats des 7 autres expériences.

4.2.3 Approches paramétriques de la semi-supervision

L'idée de faire apprendre des classificateurs par combinaison des données étiquetées et non étiquetées est vieille dans la communauté de statistiques. Au moins dès 1968, elle a été suggérée par Hartley et Rao (1968) pour construire des classificateurs par maximisation

de la vraisemblance en examinant toutes les affectations de classes possibles. L'article sur lequel se base tout développement ultérieur est de (Day, 1969) et présente une approche itérative semblable à l'algorithme EM pour estimer les paramètres d'un mélange de deux lois normales avec des covariances connues seulement à partir de données non étiquetées. Des algorithmes itératifs similaires ont été aussi proposées (McLachlan, 1975; Titterington, 1976) pour construire des classificateurs par maximum de vraisemblance des données étiquetées et non étiquetées, principalement pour des mélanges des distributions normales.

Ganesalingam et McLachlan (1978) ont étudié le problème de l'estimation de la fonction discriminante linéaire de Fisher :

$$h(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^t \mathbf{x}, \quad (4.4)$$

c'est à dire l'estimation de β_0 et $\boldsymbol{\beta}$ par la maximisation de vraisemblance des données étiquetées et non étiquetées pour un problème de deux lois normales mono-variables de variances égales et connues. Ils ont montré, via le calcul de l'efficacité asymptotique relative de la fonction discriminante linéaire estimée, que les données non étiquetées utilisées contiennent des informations pertinentes. O'Neill (1978) est allé une étape plus loin, il a traité le cas de distributions normales multivariées avec des matrices de covariance égales et connues. Cette efficacité "asymptotique" n'est pas nécessairement valable dans le cas où la taille N de l'échantillon des données observées est petite par rapport à la dimension p du vecteur \mathbf{x} (Day, 1969) d'où les travaux de Ganesalingam et McLachlan (1979) pour étudier ce cas.

Dempster et al. (1977) ont présenté l'algorithme Expectation-Maximisation (EM), formalisant plusieurs des techniques itératives précédemment suggérées pour la maximisation de vraisemblance avec des données manquantes. Son applicabilité pour l'estimation des paramètres des modèles de mélange par maximum de vraisemblance (ou maximum a posteriori) avec des données étiquetées et non étiquetées (Murray et Titterington, 1978) a été reconnue immédiatement. Ensuite Little (1977) a employé ceci pour la classification. Depuis lors, cette approche continue à être employée et étudiée (McLachlan et Ganesalingam, 1982; Shahshahani et Landgrebe, 1994; Ganesalingam, 1989). Trois excellents aperçus de l'histoire de l'algorithme EM et son application aux modèles de mélange sont les livres de McLachlan et Basford (1988), McLachlan et Krishnan (1997) et McLachlan et Peel (2000).

Le travail de Nigam et al. (2000) est le premier à explorer la combinaison des données étiquetées et non étiquetées pour la classification des textes. En effet, la classification supervisée des documents textuels, nécessite l'obtention de documents étiquetés qui sont chers alors que de grandes quantités de documents non étiquetés sont aisément disponibles. Dans le but d'améliorer la performance de la classification, Nigam et al. (2000) ont présenté un algorithme pour apprendre des documents étiquetés et non étiquetés. Il est basé sur la combinaison de l'algorithme EM classique avec un mélange de densités et d'un classifieur naïf de Bayes, qui supposent l'indépendance de variables. L'algorithme forme d'abord le classifieur naïf de Bayes qui estime le vecteur Φ des paramètres par maximum a posteriori en utilisant l'ensemble des documents supervisés disponibles \mathcal{X}_d :

$$\hat{\Phi} = \arg \max_{\Phi} p(\mathcal{X}_d / \Phi) \cdot p(\Phi), \quad (4.5)$$

et étiquette (calcule la probabilité d'appartenance) l'ensemble des documents non supervisés \mathcal{X}_u . Il forme ensuite un nouveau classifieur en utilisant les étiquettes de tous les documents, et le réitère (étapes E et M de l'algorithme EM) jusqu'à la convergence. Si \mathcal{D} est l'ensemble des observations supervisées et non supervisées $\mathcal{X}_d \cup \mathcal{X}_u$, la log-vraisemblance L à maximiser par l'algorithme EM sera écrite :

$$L(\Phi; \mathcal{D}) = L(\Phi; \mathcal{X}_d) + L(\Phi; \mathcal{X}_u). \quad (4.6)$$

En effet, cet algorithme fonctionne bien quand les données se conforment avec le modèle de mélange génératif choisi *a priori*. Cependant ce n'est pas tout le temps le cas en pratique, et une dégradation de performance peut en résulter. Les auteurs ont proposé deux extensions à l'algorithme EM qui améliorent la performance de classification dans ces conditions :

1. Un facteur de poids λ pour moduler la contribution des données non étiquetées.

Par analogie avec l'équation 4.6, la log-vraisemblance à maximiser est :

$$L(\Phi; \mathcal{D}) = L(\Phi; \mathcal{X}_d) + \lambda L(\Phi; \mathcal{X}_u). \quad (4.7)$$

2. L'utilisation de plusieurs composantes de mélange par classe.

Le nombre de composantes par classe ainsi que λ sont choisis par validation croisée. les auteurs ont prouvé, via les résultats expérimentaux obtenus avec trois différentes bases de documents réelles, que l'utilisation des données non étiquetées réduit l'erreur de classification jusqu'à 30% .

Récemment Larsen et al. (2002) ont proposé un algorithme de classification semi-supervisée hiérarchique fondé sur des modèles de mélange. Cet algorithme peut se décomposer en trois étapes :

- Etape 1 : Sur-partitionner les données en R clusters (composantes). R doit être supérieure au nombre final de classes (K).
- Etape 2 : Effectuer une classification hiérarchique agglomérative à partir des R clusters jusqu'à en obtenir un seul.
- Etape 3 : Couper la hiérarchie pour obtenir la partition finale selon le nombre de classes optimal pour un critère ou fixé par l'utilisateur.

Une fois l'étape 1 effectuée, on dispose d'une estimation du vecteur de paramètres du modèle de mélange. L'estimation des paramètres est faite par un algorithme EM modifié, appelé USGGM (Unsupervised/Supervised Generalizable Gaussian Mixture model) et qui est une extension de l'algorithme EM modifié proposé par Miller et Uyar (1997), pour apprendre des données étiquetées et non étiquetées et tel que chaque classe est modélisée par plusieurs composantes du mélange gaussien. En effet, soit $y \in \{1, \dots, K\}$ le label indiquant la classe de \mathbf{x} . La densité du couple entrée sortie (\mathbf{x}, y) est supposée mélange de gaussienne (Miller et Uyar, 1997) :

$$f(y, \mathbf{x} | \Phi) = \sum_{r=1}^R P(y|r) f(\mathbf{x}|r) \pi_r, \quad (4.8)$$

avec

$$f(\mathbf{x}|r) = \frac{(2\pi)^{-\frac{p}{2}}}{\det|\Sigma_r|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_r)^t \boldsymbol{\Sigma}_r^{-1} (\mathbf{x}-\boldsymbol{\mu}_r)}, \quad (4.9)$$

et R est le nombre de composantes. $f(\mathbf{x}|r)$ est la densité gaussienne de proportion π_r ($\sum_{r=1}^R \pi_r = 1$). $P(y|r)$ est la probabilité *a posteriori* de la classe y par rapport à la composante r , $\sum_{y=1}^K P(y|r) = 1$. La r 'ième gaussienne est décrite par le vecteur moyenne $\boldsymbol{\mu}_r$ et la matrice de covariance $\boldsymbol{\Sigma}_r$. Φ est le vecteur de tous les paramètres du modèle : $\Phi = \{P(y|r), \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r, \pi_r : \forall r, y\}$. L'objectif est d'estimer Φ en utilisant la combinaison des exemples étiquetés et non étiquetés $\mathcal{D} = \mathcal{X}_d \cup \mathcal{X}_u$ supposés indépendantes. La log-vraisemblance L à maximiser par l'algorithme EM est :

$$L(\Phi; \mathcal{D}) = \sum_{\mathbf{x} \in \mathcal{X}_d} \log \sum_{r=1}^R P(y|r) f(\mathbf{x}|r) \pi_r + \lambda \sum_{\mathbf{x} \in \mathcal{X}_u} \log \sum_{r=1}^R f(\mathbf{x}|r) \pi_r. \quad (4.10)$$

Les détails de cet algorithme USGGM sont expliqués clairement dans (Larsen et al., 2002).

Dans l'étape 2, la fusion entre deux clusters est décidée à l'aide d'une mesure de similarité probabiliste entre les vecteurs paramètres correspondents. Trois mesures de similarité ont été comparées: (1) la mesure de dissimilarité L_2 , (2) la confusion entre clusters et (3) la divergence de Kullback-Leibler.

Cet algorithme est très important par le fait qu'il permet de déterminer automatiquement le nombre de composantes de chaque classe. Le seul paramètre à choisir est le nombre total de composantes R . Par contre, il ne permet pas la classification partiellement supervisée¹. Par ailleurs, Ambroise et Govaert (2000) ont introduit une approche paramétrique toujours basée sur les modèles de mélange et l'algorithme EM, pour apprendre des labels partiellement connus. Dans la section 4.3 suivante, nous allons présenter une extension de cette approche. Nous allons l'appliquer par la suite pour retrouver des images similaires en recherche interactive d'images basée sur le contenu.

Modélisation d'une classe par un mélange

Modéliser une classe par plusieurs composantes de mélange n'est pas nouveau. Hastie et Tibshirani (1996) ont introduit cette idée dans le cadre de nouvelle méthode de classification *supervisée* appelée "Analyse Discriminante de mélange" ou en anglais "Mixture discriminant analysis".

Chaque classe C_k est supposée un mélange de R_k (valeur est fixée *a priori*) composantes gaussiennes notées c_{kr} . La densité de la k ème classe s'écrit alors :

$$f_k(\mathbf{x}|\Phi_k) = \sum_{r=1}^{R_k} \pi_{kr} f_{kr}(\mathbf{x}|\boldsymbol{\theta}_{kr}), \quad (4.11)$$

où

- \mathbf{x} désigne la variable aléatoire générant x_i ;

1. voir définition dans le paragraphe 4.1

- π_{kr} est la proportion de la composante c_{kr} , ($0 < \pi_{kr} < 1$ et $\sum_{r=1}^{R_k} \pi_{kr} = 1$);
- $\Phi_k = (\pi_{k1}, \dots, \pi_{kR_k}, \boldsymbol{\theta}_{k1}, \dots, \boldsymbol{\theta}_{kR_k})$ et $\boldsymbol{\theta}_{kr} = (\boldsymbol{\mu}_{kr}, \boldsymbol{\Sigma})$;

Contrairement à notre méthode, les paramètres de chaque classe sont estimés séparément, par l'algorithme EM, à part la matrice de variance-covariance $\boldsymbol{\Sigma}$ qui est supposée la même pour les composantes gaussiennes de toutes les classes. Cette hypothèse a été adoptée afin d'aboutir à une discrimination et réduction du nombre de paramètres à estimer comme dans l'approche LDA (Analyse Discriminante Linéaire) de Fisher.

Remarque Nous montrons que les paramètres de toutes les classes peuvent être estimés ensemble par traduction du problème à une estimation d'un mélange gaussien de $R = \sum_{k=1}^K R_k$ composantes.

4.3 Apprentissage avec les modèles de mélange et l'algorithme EM

Le but de notre travail est de (1) trouver un algorithme applicable à toute sorte d'apprentissage (*cf. section 4.1*) et (2) ayant des classes de formes complexes (par exemple une classe entourée par une autre ou bien multimodale). D'où l'idée de l'apprentissage avec les modèles de mélange et l'algorithme EM. Ainsi chaque classe est supposée formée de plusieurs composantes.

4.3.1 Hypothèses

Les données \mathcal{X} sont supposées générées par un modèle de mélange à R composantes. Ainsi la densité de probabilité de l'observation \mathbf{x}_i s'écrit :

$$f(\mathbf{x}_i | \Phi) = \sum_{r=1}^R \pi_r f_r(\mathbf{x}_i | \boldsymbol{\theta}_r), \quad (4.12)$$

avec π_r la proportion de la composante c_r , ($0 < \pi_r < 1$ et $\sum_{r=1}^R \pi_r = 1$). Le vecteur Φ est l'ensemble de paramètres à estimer, $\Phi = (\pi_1, \dots, \pi_R, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R)$.

Supposant que chaque classe C_k ² est formée par plusieurs composantes du mélange, on aura :

$$f(\mathbf{x}_i | \Phi) = \sum_{r \in C_1} \pi_r f_r(\mathbf{x}_i | \boldsymbol{\theta}_r) + \dots + \sum_{r \in C_K} \pi_r f_r(\mathbf{x}_i | \boldsymbol{\theta}_r). \quad (4.13)$$

Soit R_k le nombre de composante formant la classe C_k .

Notons que chaque classe est cependant un modèle de mélange. En effet, soit Π_k la probabilité *a priori* de la classe C_k , on aura :

2. la lettre majuscule C signifie une classe et la lettre minuscule c représente une composante

$$\Pi_k \triangleq \sum_{r \in C_k} \pi_r, \quad (4.14)$$

et $f(\mathbf{x}_i|\Phi)$ peut être écrite:

$$f(\mathbf{x}_i|\Phi) = \Pi_1 \sum_{r \in C_1} \frac{\pi_r}{\Pi_1} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r) + \dots + \Pi_K \sum_{r \in C_K} \frac{\pi_r}{\Pi_K} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r), \quad (4.15)$$

ou encore

$$f(\mathbf{x}_i|\Phi) = \Pi_1 g_1(\mathbf{x}_i|\Phi_1) + \dots + \Pi_K g_K(\mathbf{x}_i|\Phi_K). \quad (4.16)$$

Ainsi chaque classe C_k est un modèle de mélange puisque sa densité vérifie :

$$g_k(\mathbf{x}_i|\Phi_k) = \sum_{r \in C_k} \frac{\pi_r}{\Pi_k} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r), \quad (4.17)$$

tel que

$$\sum_{r \in C_k} \frac{\pi_r}{\Pi_k} = 1. \quad (4.18)$$

Φ_k est le vecteur paramètre de la classe C_k : $\Phi_k = \left(\left\{ \frac{\pi_r}{\Pi_k}, \boldsymbol{\theta}_r \right\}; r \in C_k \right)$.

Prenons par exemple, le cas de trois classes, la densité de \mathbf{x}_i s'écrira :

$$f(\mathbf{x}_i|\Phi) = \sum_{r \in C_1} \pi_r f_r(\mathbf{x}_i|\boldsymbol{\theta}_r) + \sum_{r \in C_2} \pi_r f_r(\mathbf{x}_i|\boldsymbol{\theta}_r) + \sum_{r \in C_3} \pi_r f_r(\mathbf{x}_i|\boldsymbol{\theta}_r), \quad (4.19)$$

ou encore :

$$f(\mathbf{x}_i|\Phi) = \sum_{r=1}^{R_1} \pi_r f_r(\mathbf{x}_i|\boldsymbol{\theta}_r) + \sum_{r=R_1+1}^{R_1+R_2} \pi_r f_r(\mathbf{x}_i|\boldsymbol{\theta}_r) + \sum_{r=R_1+R_2+1}^{R_1+R_2+R_3} \pi_r f_r(\mathbf{x}_i|\boldsymbol{\theta}_r). \quad (4.20)$$

Ainsi, le problème d'estimation des paramètres du classifieur ($\mathcal{C} : \mathcal{X} \longrightarrow \mathcal{Y}$) revient à estimer les paramètres des R composantes du modèle de mélange.

4.3.2 Algorithme EM

Une méthode classique pour estimer les paramètres d'un modèle de mélange par maximisation de vraisemblance est l'algorithme EM (Dempster et al., 1977), qui possède la propriété de convergence vers un maximum local (Ambroise, 1996; Dang, 1998).

Quelque soit le problème d'apprentissage (*cf. section 4.1*), l'algorithme EM peut être exprimé de la même façon (voir annexe A pour la démonstration). En effet, codons le vecteur \mathbf{y}_i , qui indique de quelle classe est issu \mathbf{x}_i , par \mathbf{z}_i comme suit (sans perte de généralité prenons le cas de trois classes):

- $\mathbf{z}_i = (\underbrace{1, \dots, 1}_{R_1 \text{ fois}}, \underbrace{0, \dots, 0}_{R_2 \text{ fois}}, \underbrace{0, \dots, 0}_{R_3 \text{ fois}})$, pour \mathbf{x}_i appartenant à la classe C_1 c'est à dire \mathbf{y}_i connu et vaut $(1, 0, 0)$,

- $\mathbf{z}_i = (\underbrace{1, \dots, 1}_{R_1 \text{ fois}}, \underbrace{1, \dots, 1}_{R_2 \text{ fois}}, \underbrace{1, \dots, 1}_{R_3 \text{ fois}})$, pour \mathbf{y}_i inconnue ($\mathbf{y}_i = (?, ?, ?)$), ou autrement dit \mathbf{x}_i est non classé,
- $\mathbf{z}_i = (\underbrace{0, \dots, 0}_{R_1 \text{ fois}}, \underbrace{1, \dots, 1}_{R_2 \text{ fois}}, \underbrace{1, \dots, 1}_{R_3 \text{ fois}})$, pour \mathbf{x}_i n'appartenant pas à la classe C_1 : $y_{i1} = 0$, y_{i2} et y_{i3} sont inconnues c'est à dire $\mathbf{y}_i = (0, ?, ?)$.

Un exemple illustratif est donné dans la table suivante 4.5.

| | X | | | | Y | | Z | | | | | | |
|-----|----------|-----|-----|---|----------|---|----------|---|---|---|---|---|---|
| 3.5 | 2.3 | 0.3 | 4.2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 2.2 | 1.4 | 2.9 | 1.3 | 0 | ? | ? | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| 4.2 | 1.7 | 2.2 | 1.1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 2.5 | 2.3 | 0.3 | 4.2 | ? | ? | ? | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 9.2 | 2.4 | 2.9 | 1.3 | ? | ? | ? | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 6.2 | 1.2 | 2.2 | 1.1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |

TAB. 4.5 – Exemple de plusieurs composantes de mélange par classe. Cas de $R_1 = R_2 = R_3 = 3$.

Les deux étapes de l'algorithme EM sont :

- Étape E :

Pour $i = 1, \dots, N$ et $r = 1, \dots, R$, calculer

$$p(r|\mathbf{x}_i, \mathbf{z}_i, \Phi^{(q)}) = c_{ir}^{(q)} = \frac{z_{ir}\pi_r^{(q)}f_r(\mathbf{x}_i|\boldsymbol{\theta}_r^{(q)})}{\sum_{l=1}^R z_{il}\pi_l^{(q)}f_l(\mathbf{x}_i|\boldsymbol{\theta}_l^{(q)})}. \quad (4.21)$$

- Étape M :

Calculer $\Phi^{(q+1)}$ qui maximise

$$Q(\Phi|\Phi^{(q)}) = \sum_{i=1}^N \sum_{r=1}^R c_{ir}^{(q)} \log \pi_r f_r(\mathbf{x}_i|\boldsymbol{\theta}_r). \quad (4.22)$$

La probabilité d'appartenance à la classe C_k est :

$$P(\mathbf{x}_i \in C_k | \mathbf{x}_i, \mathbf{z}_i; \hat{\Phi}) = \sum_{r \in C_k} p(r|\mathbf{x}_i, \mathbf{z}_i; \hat{\Phi}). \quad (4.23)$$

$\hat{\Phi}$ est l'estimation de Φ par l'algorithme EM.

La table 4.6 détaille l'algorithme EM dans le cas d'un mélange de gaussiennes.

4.4 Comparaison entre classification supervisée et semi-supervisée

Deux questions émergent : pourquoi apprendre en semi-supervisé ? autrement dit, est-ce que l'apprentissage semi-supervisé aide et dans quel cas ? et (2) qu'est ce que nous apportent comme informations les exemples non classés ?

Entrées :

Les données $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$,
les labels $\{z_1, \dots, z_N\}$ connus, inconnus, ou partiellement connus.

Initialisation

$$q \leftarrow 0$$

Initialiser les proportions, centres et variances à une valeur arbitraire $\Phi^{(0)}$

Répéter

- Etape E : calculer les probabilités d'appartenance $c_{ir}^{(q+1)}$
pour i de 1 à N , r de 1 à R **faire**

$$c_{ir}^{(q+1)} \leftarrow \frac{z_{ir}\pi_r^{(q)}f_r(\mathbf{x}_i|\boldsymbol{\theta}_r^{(q)})}{\sum_{l=1}^R z_{il}\pi_l^{(q)}f_l(\mathbf{x}_i|\boldsymbol{\theta}_l^{(q)})};$$

- Etape M : recalculer les paramètres $\Phi^{(q+1)}$
pour r de 1 à R **faire**

$$\begin{aligned} n_r^{(q+1)} &\leftarrow \sum_{i=1}^N c_{ir}^{(q+1)}; \\ \pi_r^{(q+1)} &\leftarrow \frac{n_r^{(q+1)}}{N}; \\ \boldsymbol{\mu}_r^{(q+1)} &\leftarrow \frac{\sum_{i=1}^N c_{ir}^{(q+1)} \cdot \mathbf{x}_i}{n_r^{(q+1)}}; \\ \Sigma_r^{(q+1)} &\leftarrow \frac{1}{n_r^{(q+1)}} \sum_{i=1}^N c_{ir}^{(q+1)} \cdot (\mathbf{x}_i - \boldsymbol{\mu}_r^{(q+1)}) (\mathbf{x}_i - \boldsymbol{\mu}_r^{(q+1)})^t; \end{aligned}$$

$$q \leftarrow q + 1;$$

jusqu'à $\mathbf{c}^{(q+1)} \simeq \mathbf{c}^{(q)}$ ou $q = Q_{max}$

Sortie :

$$\hat{\Phi} \leftarrow \Phi^{(q)}$$

TAB. 4.6 – Algorithme EM : cas d'un mélange gaussien et de labels partiellement connus

Nous allons, dans ce qui suit, essayer de répondre à ces questions en comparant les deux approches d'apprentissage utilisant deux jeux de données : des données synthétiques et un ensemble d'index d'une base d'images.

4.4.1 Données synthétiques

Les données utilisées dans ces expérimentations sont générées par un mélange de trois gaussiennes et divisées en deux classes : la classe C_1 formée par deux composantes et la classe C_2 par une seule. La figure 4.1 montre un échantillon de 1000 observations.

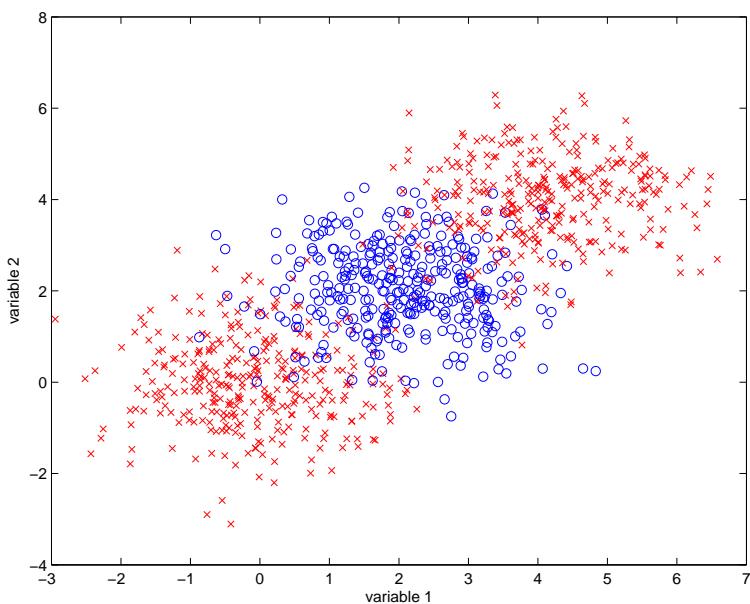
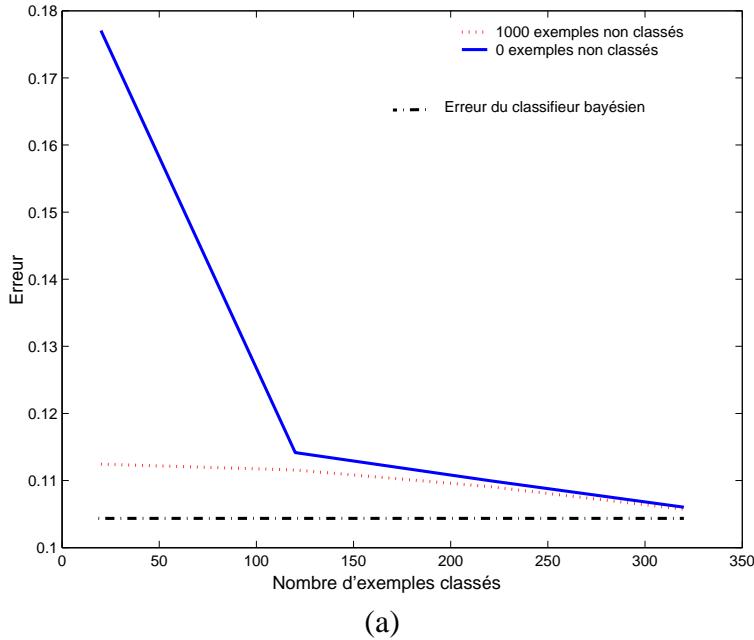


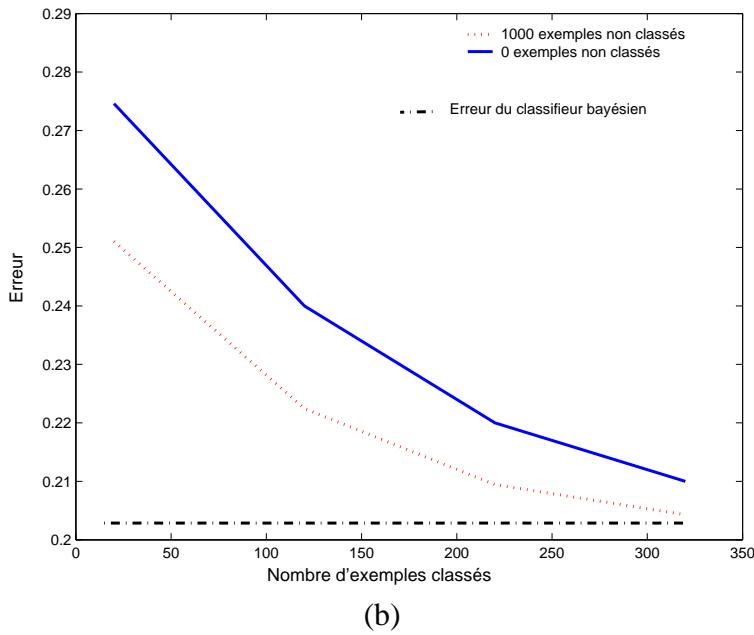
FIG. 4.1 – *Données synthétiques : la première classe "x" est formée par deux nuages circulaires et séparés par la deuxième classe en "o".*

La figure 4.2 montre l'effet de l'incorporation des données non classées en apprentissage. L'axe vertical indique l'erreur de classification calculée sur un ensemble de test et l'axe horizontal indique le nombre d'exemples classés utilisés en apprentissage (pour plus de détails voir table 4.7). Il est bien clair que l'intégration des exemples non classés améliore l'estimation des paramètres du classifieur et d'autant plus que les données classées sont rares.

Dans la figure 4.3 nous considérons l'effet du nombre d'exemples non classés en apprentissage. Pour 2 échantillons différents de classe inconnue nous varions sur l'axe horizontal le nombre d'exemples non classés. Naturellement, avoir plus de données non classées aide d'autant plus quand les données classées sont peu nombreuses.



(a)



(b)

FIG. 4.2 – Erreur de classification après apprentissage supervisé et semi-supervisé (1000 exemples non classés). Avec un petit nombre d’observations de classe connue, l’estimation des paramètres utilisant l’algorithme EM est meilleur en apprentissage semi-supervisé. Les deux algorithmes tendent à converger l’un de l’autre quand on a un nombre suffisant d’exemples classés. (a) l’erreur du classifieur bayésien est de 10% environ, (b) 20%.

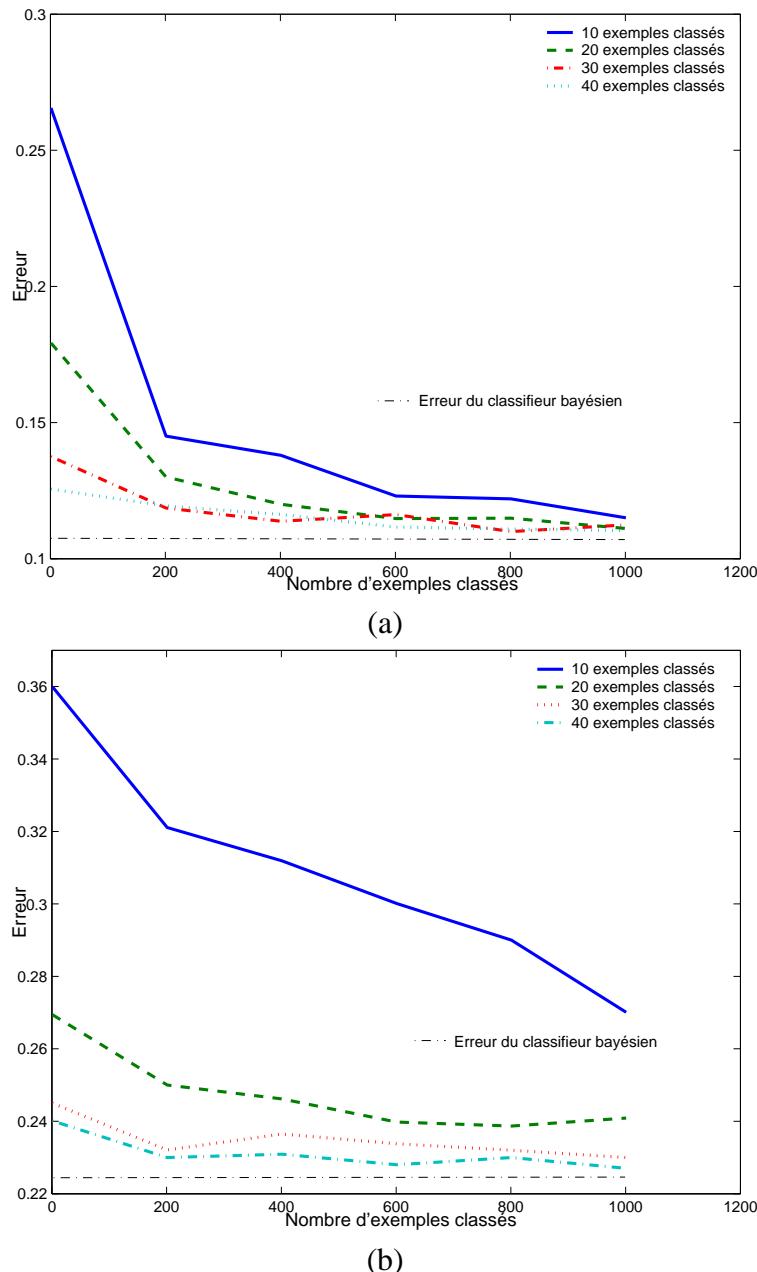


FIG. 4.3 – Erreur de classification en fonction du nombre d'exemples non classés. La prise en compte de données non classées permet de diminuer l'erreur. (a) l'erreur du classifieur bayésien est de 10% environ, (b) 20%.

1. Initialisation :
 - Nc nombre d'exemples classés;
 - Nnc nombre d'exemples non classés;
2. Simuler Nc exemples supposés de classe connue $Eclasses$;
3. Simuler Nnc exemples de classe inconnue $Enonclasses$;
4. Simuler un ensemble Test de Nt exemples $Etest$ servant pour calculer l'erreur;
5. Estimer les paramètres du mélange *parametres* par EM utilisant exemples classés $Eclasses$ et non classés $Enonclasses$;
6. Classer les exemples de l'ensemble $Etest$ utilisant les paramètres estimés *parametres* et calculer l'erreur de classement *Erreur*;
7. Incrémenter Nc (cas de la figure 4.2) ou Nnc (cas de la figure 4.3) et retourner à l'étape 2;

TAB. 4.7 – Procédure utilisée pour tracer la courbe de l'erreur de classification sur des données synthétiques en fonction du nombre d'exemples classés et non classés utilisés en apprentissage. Cette procédure est répétée 40 fois et la valeur moyenne de l'erreur est tracée. 50 essais de l'algorithme EM sont réalisés à chaque appel de cette procédure. L'essai qui a la plus grande vraisemblance est sélectionné.

4.4.2 Base d'images ANN

Les expérimentations présentées dans ce paragraphe sont appliquées sur la base d'images ANN de l'université de Washington qui contient 493 photos de 11 catégories différentes. Les index extraits de ces images sont fournis par J.Fournier et al.³ grâce à l'approche expliquée dans (Fournier et al., 2001) du système RETIN de recherche d'images similaires basée sur le contenu.

Le problème posé consiste à classer la base en deux classes : (1) classe C_1 des images pertinentes similaires à une image requête (ici les images appartenant à la catégorie de la requête) et (2) classe C_2 des images non pertinentes supposées ici celles qui n'appartiennent pas à la catégorie cherchée.

Le problème principal de la recherche interactive d'images est le fait d'avoir un très petit nombre d'images annotées (ou classées) par rapport à la taille de la base. D'où l'intérêt de l'intégration des images non annotées (non classées) dans l'apprentissage.

L'estimation de l'erreur de classification dans ce cas est non pertinente. Ayant un grand nombre d'images non pertinentes, un bon classifieur peut être obtenu en classant toutes les images comme non pertinentes. D'où l'utilisation des 2 critères classiques qui sont le rappel et la précision :

³. Équipe de Traitement des Images et du Signal, Université de Cergy-Pontoise / ENSEA, 6 av. du Ponceau 95014 Cergy-Pontoise Cedex, France

$$\text{Precision} = \frac{\text{nombre d'images pertinentes retournées}}{\text{nombre d'images retournées}}, \quad (4.24)$$

$$\text{Rappel} = \frac{\text{nombre d'images pertinentes retournées}}{\text{nombre d'images pertinentes dans la base}}. \quad (4.25)$$

Les images retournées correspondent aux images les plus pertinentes trouvées par le système de recherche et retournées à l'utilisateur. Dans ces expérimentations, le nombre d'images retournées est choisi tel que le rappel soit égal à la précision, c'est à dire il est égal au nombre d'images pertinentes dans la base.

Dans le but de voir l'effet de l'intégration des images non classées en apprentissage, nous avons tracé la courbe du rappel (précision) en fonction du nombre d'images non classées. La procédure utilisée pour tracer ces courbes est expliquée en détails dans la table 4.8.

La figure 4.4 montre le graphe du rappel tracé pour 3 différents petits nombres (10, 20, 30) d'images classées. Les catégories cherchées⁴ sont "arborgreens" (courbes en haut) et "campusinfall" (courbes en bas). Le modèle du mélange choisi est formé d'une seule gaussienne par classe. On constate qu'en augmentant le nombre d'images non classées le rappel croît jusqu'à un maximum (50 images non classées environ) puis il commence à diminuer.

La dégradation des performances obtenue peut être expliquée par le fait qu'avec un petit nombre d'images de classe connue, l'estimation des paramètres en apprentissage semi-supervisé peut tomber sur des distributions gaussiennes loin de celles des exemples classés, et l'intégration des images non classées n'a pas pu aider à la classification.

Deux solutions sont proposées par K.Nigam et al. (2000) pour régler ce problème :

1. modéliser la classe par un mélange de composantes plus nombreuses,
2. utiliser un facteur de poids λ pour moduler la contribution des données non étiquetées.

Vu la structure de la classe C_2 des images non pertinentes qui englobe 10 catégories, elle peut être modélisée par plusieurs gaussiennes. Les graphes de la figure 4.5 illustrent les courbes de rappel vs. le nombre d'images non classées pour différents modèles de la classe C_2 et pour 20 images classées. La classe C_1 est toujours estimée par une seule gaussienne. On voit clairement l'amélioration des performances en augmentant le nombre de gaussiennes. Bien évidemment, augmenter le nombre de gaussiennes par classe augmente le nombre de paramètres à estimer et par conséquence un grand nombre d'exemples classés est nécessaire pour un apprentissage supervisé. D'où la diminution du taux de rappel pour 0 images non annotées.

Cozman et Cohen (2002) ont montré que la dégradation des performances en utilisant les données non étiquetées est obtenue quand le modèle choisi par hypothèse est différent du modèle réel qui génère les données. Ils ont présenté plus récemment dans (Cozman et al., 2003), une analyse mathématique de cette dégradation. Elle est due au fait que l'estimateur par maximum *a posteriori* est parfois biaisé.

4. voir vérité terrain de la base

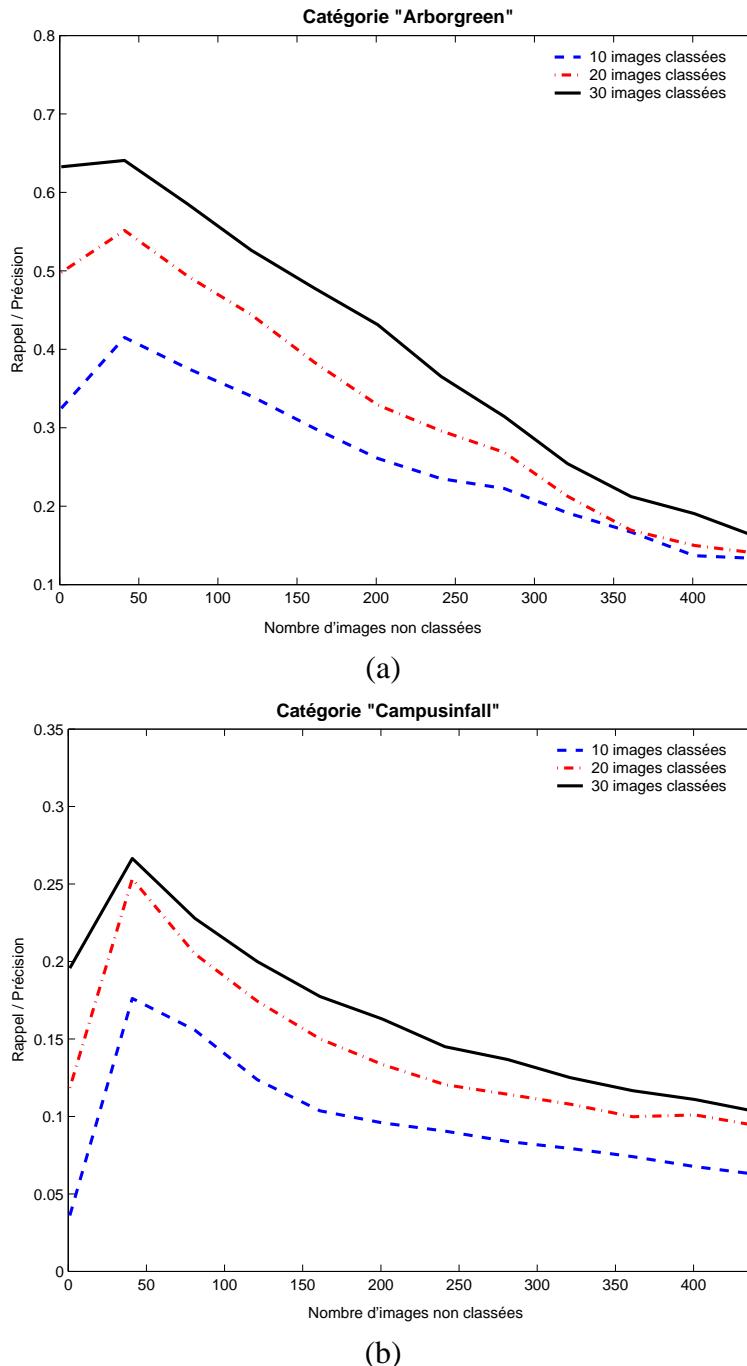


FIG. 4.4 – Rappel versus nombre des images non classées intégrées dans l’ensemble d’apprentissage pour 3 différents nombres d’images annotées. (a) catégorie “arbogreens” (b) catégorie “campusinfall”. La classe des images pertinentes ainsi que des images non pertinentes, sont modélisées par une seule gaussienne. Une dégradation des performances est constatée.

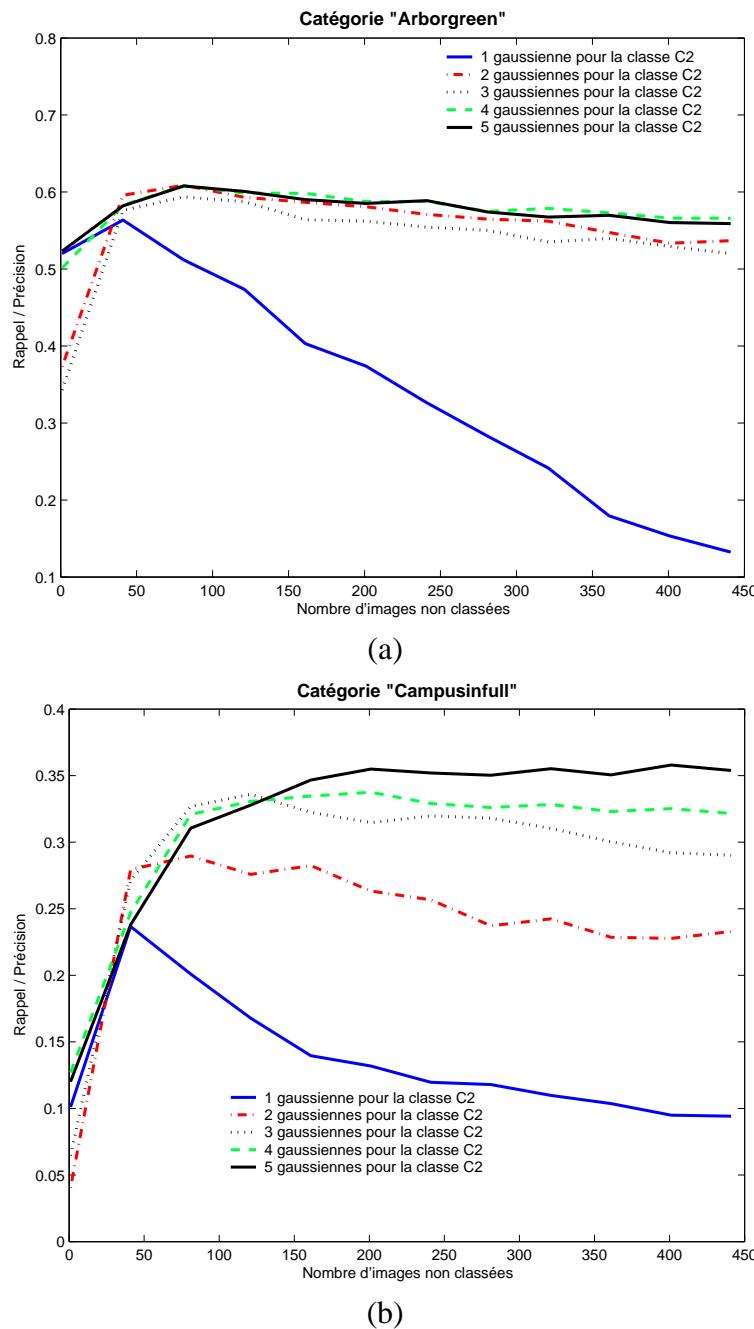


FIG. 4.5 – *Rappel versus le nombre des images non classées intégrées dans l'ensemble d'apprentissage. Les images annotées sont au nombre de 20. Les courbes sont tracées pour différents nombres de gaussiennes pour la classe des images non pertinentes. La classe des images pertinentes est toujours modélisée par une seule gaussienne.* (a) catégorie "arborgreens" (b) catégorie "campusinfall". Plusieurs composantes par classe améliore les performances.

1. Initialisation
 - Choisir une catégorie comme classe des images pertinentes C_1 . le reste est supposé appartenant à la classe C_2 (vérité terrain);
 - N_{nc} nombre d'images non classées;
 - N_{c1} (N_{c2}) le nombre d'images de classe connue C_1 (C_2);
2. Tirer au hasard N_{c1} et N_{c2} images de classe connue $E_{classes}$;
3. Tirer au hasard N_{nc} images de classe inconnue $E_{nonclasses}$;
4. Estimer les paramètres du mélange *paramètres* par EM utilisant images classées $E_{classes}$ et non classées $E_{nonclasses}$;
5. Calculer le rappel *Rappel* de toutes les images de la base utilisant *paramètres*;
6. Incrémenter N_{nc} et retourner à l'étape 2;

TAB. 4.8 – Procédure utilisée pour tracer la courbe de l'erreur (rappel) de classification de la base ANN (problème de deux classes) en fonction du nombre d'exemples non classés utilisés en apprentissage semi-supervisé. Cette procédure est répétée 200 fois et la valeur moyenne est tracée.

4.5 Conclusion

Dans ce chapitre, nous avons présenté l'intérêt de la semi-supervision en classification. Dans une première partie, nous avons dressé un état de l'art des méthodes actuelles. Ensuite nous avons introduit une méthode paramétrique d'apprentissage semi-supervisé.

Nous avons comparé les deux classifications semi-supervisée et supervisée. Nous pouvons conclure que l'apprentissage en semi-supervisé est très important dans le cas où les labels sont rares.

Dans le chapitre suivant nous appliquons cette méthode d'apprentissage semi-supervisé basé sur les lois de mélange et l'algorithme EM pour la recherche interactive d'images basée sur le contenu.

Chapitre 5

EMiner : les modèles de mélange pour la recherche interactive d'images

L'objectif de ce chapitre est la présentation d'une technique originale de recherche interactive d'images basée sur le contenu. La méthode relève de l'approche statistique et plus précisément de l'apprentissage statistique. Nous considérons ici une approche semi-supervisée appelée "EMiner", utilisant les modèles de mélange et l'estimation de leurs paramètres par l'algorithme EM.

Les particularités de l'indexation d'images par rapport à une classification standard sont : le très faible nombre d'images annotées en début de recherche (petite taille de l'ensemble d'apprentissage), la dimension souvent importante de l'index (nombre important de variables) ainsi que la définition incrémentale de l'ensemble d'apprentissage. De plus les paramètres du classifieur doivent être remis à jour "rapidement" après chaque nouveau bouclage de pertinence. C'est dans ce cadre que notre travail se situe.

Une méthode de sélection de variables est aussi introduite dans la boucle de pertinence pour éliminer les variables redondantes et rendre la recherche interactive plus adaptable aux besoins de l'utilisateur.

5.1 EMiner pour la recherche interactive d'images

Nous nous plaçons dans le cas d'une recherche d'une catégorie d'images. La recherche peut être initiée par une image exemple (requête simple). Ainsi à chaque itération de la recherche ou encore à chaque bouclage, le système affiche un ensemble d'images résultat sensées être proches de l'image cherchée par l'utilisateur. Ce dernier annote les images en pertinentes et non-pertinentes, ce qui permet, au fil des itérations, d'accumuler un ensemble d'exemples pertinents et non-pertinents. La recherche peut aussi être initiée par une requête composée de plusieurs images (requête multiple). Ces exemples sont alors fournis directement par l'utilisateur ou bien sont accumulés durant la recherche interactive.

L'approche présentée consiste ainsi à estimer les paramètres d'un classifieur qui distingue "au mieux" les deux classes C_1 et C_2 d'images respectivement pertinentes et non-pertinentes et attribue un degré de pertinence à chaque image de la base.

L'une des caractéristiques de la recherche interactive d'images est la rareté des observations (images) annotées. L'approche semi-supervisée basée sur les modèles de mélange (voir chapitre 4) semble adaptée : utiliser l'ensemble des images (annotées et non annotées) de la base pour estimer les paramètres.

Considérons une base composée de N images et notons \mathbf{x}_i l'index de l'image I_i . $\mathbf{x}_i = (x_{ij}, j = 1, \dots, p)$ est un vecteur de p éléments réels. L'ensemble des N index est noté $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. Ils sont supposés générés par un modèle de mélange à R composantes.

Pour permettre (1) la recherche de catégories de formes complexes et multimodales et (2) la modélisation de la classe C_2 des images non-pertinentes englobant différentes catégories, nous avons choisi de modéliser les classes C_1 et C_2 par un mélange de R_1 et R_2 composantes gaussiennes respectivement. Il est donc évident que $R = R_1 + R_2$. La densité \mathbf{x}_i s'écrit :

$$f(\mathbf{x}_i | \Phi) = \sum_{r=1}^{R_1} \pi_r f_r(\mathbf{x}_i | \boldsymbol{\theta}_r) + \sum_{r=R_1+1}^{R_1+R_2} \pi_r f_r(\mathbf{x}_i | \boldsymbol{\theta}_r). \quad (5.1)$$

La probabilité d'appartenance à la classe C_1 s'exprime comme :

$$P(\mathbf{x}_i \in C_1 | \mathbf{x}_i, \mathbf{z}_i; \Phi) = \sum_{r=1}^{R_1} p(r | \mathbf{x}_i, \mathbf{z}_i; \Phi), \quad (5.2)$$

et celle de la classe C_2 :

$$P(\mathbf{x}_i \in C_2 | \mathbf{x}_i, \mathbf{z}_i; \Phi) = \sum_{r=R_1+1}^{R_1+R_2} p(r | \mathbf{x}_i, \mathbf{z}_i; \Phi). \quad (5.3)$$

Soient Π_1 et Π_2 les probabilités a priori des classes C_1 et C_2 respectivement,

$$\Pi_1 = \sum_{r=1}^{R_1} \pi_r, \quad (5.4)$$

$$\Pi_2 = \sum_{r=R_1+1}^{R_1+R_2} \pi_r. \quad (5.5)$$

Les annotations de l'utilisateur seront représentées par les vecteurs annotations ou étiquettes $\mathbf{z}_i \in \{0,1\}^{R_1+R_2}$ qui sont codés comme suit :

- $\mathbf{z}_i = (\underbrace{1, \dots, 1}_{R_1 \text{ fois}}, \underbrace{0, \dots, 0}_{R_2 \text{ fois}})$, pour \mathbf{x}_i annoté pertinente c'est à dire appartenant à la classe C_1 ,
- $\mathbf{z}_i = (\underbrace{0, \dots, 0}_{R_1 \text{ fois}}, \underbrace{1, \dots, 1}_{R_2 \text{ fois}})$, pour \mathbf{x}_i annoté non-pertinente (appartenant à la classe C_2),

- $\mathbf{z}_i = (\underbrace{1, \dots, 1}_{R_1+R_2 \text{ fois}})$ si \mathbf{x}_i est non annoté (classe inconnue).

Les paramètres du mélange sont estimés par l'algorithme EM comme expliqué dans le chapitre précédent (*cf.* algorithme 4.6 du chapitre 4), à la différence que les résultats sont capitalisés itération à itération.

5.1.1 Description de la recherche interactive

Lorsque l'utilisateur a choisi son image requête, les paramètres des lois attachées aux classes C_1 et C_2 sont estimés en utilisant l'image requête comme seule image annotée pertinente. Toutes les autres images de la base restent non annotées mais sont utilisées dans l'estimation. Ensuite, les probabilités conditionnelles d'appartenances des images à la classe des images pertinentes sont calculées et les meilleures N_r images résultat sont affichées à l'écran par ordre décroissant d'intérêt au sens du rapport des probabilités conditionnelles :

$$J(\mathbf{x}) = \frac{P(\mathbf{x} \in C_1 | \mathbf{x}, \mathbf{z}; \Phi)}{P(\mathbf{x} \in C_2 | \mathbf{x}, \mathbf{z}; \Phi)}. \quad (5.6)$$

L'utilisateur annote les m images qui lui semblent pertinentes et non pertinentes et laisse non annotées les images dont il doute de la ressemblance. Chaque bouclage de pertinence affine la requête en ré-estimant les paramètres du modèle avec de nouvelles annotations.

Bien entendu, nous pouvons toujours trouver le nombre optimal de composantes du modèle de mélange. Mais par souci de rapidité de recherche interactive en ligne, le nombre de composantes du modèle de mélange a été fixé *a priori* de telle sorte que le nombre de composantes de la classe des images pertinentes soit égal au nombre d'images annotées positives. Le nombre de composantes de la classe des images non-pertinentes est fixé à un nombre relativement élevé (20 par exemple). Nous avons aussi considéré des modèles de variance-covariance simples. Cette solution permet de réduire considérablement le nombre de paramètres à estimer.

Dans la suite nous utiliserons le nom EMiner pour dénommer cette approche de la recherche interactive d'images basée sur les modèles de mélange. Le tableau 5.1 récapitule les étapes de l'algorithme EMiner.

5.2 Evaluation des performances

L'objectif de cette section est d'expérimenter, d'évaluer et de valider notre approche de recherche d'images similaires. Cette évaluation montre la capacité de notre méthode à apprendre et à interagir avec l'être humain en ligne, par boucle de pertinence, afin de répondre à ses besoins et retourner les images qui lui semblent similaires.

Dans un premier temps, nous évaluons les performances de notre technique EMiner de recherche interactive et la comparons dans une deuxième étape à une autre technique basée sur les machines à vecteurs de support.

1. Choisir la ou les image(s) requête(s);
2. Mettre à jour les étiquettes de toutes les images de la base $\{z_1, \dots, z_N\}$;
3. Estimer le vecteur des paramètres Φ des gaussiennes des deux classes C_1 et C_2 par l'appel de l'algorithme **EM** (donné dans le tableau 4.6 du chapitre 4). Les images annotées positives sont choisies comme centres des gaussiennes de la classe pertinente.
4. Classer les images de la base par l'ordre décroissant de :

$$J(\mathbf{x}) = \frac{P(\mathbf{x} \in C_1 | \mathbf{x}, \mathbf{z}; \Phi)}{P(\mathbf{x} \in C_2 | \mathbf{x}, \mathbf{z}; \Phi)};$$

5. Afficher à l'écran les N_r premières images (les plus similaires à la requête);
6. Récupérer les nouvelles annotations de l'utilisateur;
7. Retourner à l'étape 2 (si l'utilisateur n'est pas satisfait);

TAB. 5.1 – *Algorithme EMiner de recherche d'images par boucle de pertinence.*

Les tests sont effectués sur la base *ANN* composée d'environ 500 images partitionnées en 11 catégories de 25 à 50 images. Deux attributs, de dimension 25 chacun, sont calculés (Couleurs Lab et Filtres de Gabor) par le système RETIN (Fournier, 2002) pour former les distributions représentant chaque image. Ainsi chaque image est représentée par un vecteur index de dimension $p = 50$.

5.2.1 Protocole d'évaluation

Dans le but de quantifier l'efficacité de recherche des images similaires, nous avons utilisé deux critères classiques de qualité pour l'évaluation de notre système (Smeulders et al., 2000) : la précision et le rappel. Ils sont définis de la manière suivante :

$$\text{Précision} = \frac{N_{rp}}{N_r}, \quad (5.7)$$

$$\text{Rappel} = \frac{N_{rp}}{N_c}, \quad (5.8)$$

avec :

- N_c est le nombre des images pertinentes de la base,
- N_r le nombre des images retrouvées et retournées à l'utilisateur,
- N_{rp} le nombre des images retrouvées pertinentes parmi les N_r .

Précisons que pour réaliser l'évaluation, il est nécessaire de disposer d'une vérité-terrain sur la base d'images afin d'automatiser le processus. Cette vérité-terrain est construite manuellement en regroupant les images en quelques catégories génériques.

Lors de l'évaluation des performances d'une recherche, toute image retrouvée appartenant à la catégorie de l'image requête est comptée comme correcte. La vérité-terrain conditionne donc l'évaluation et en particulier le type de recherche que nous prétendons évaluer par un tel procédé. Notre référence est cohérente avec la recherche de catégories d'images, toutefois, la construction de cette vérité terrain reste ambiguë, subjective et liée à notre propre expertise.

Puisqu'il n'y a aucune valeur optimale du cardinal de l'ensemble des images retrouvées¹, les deux critères sont habituellement présentés en fonction du nombre d'images retrouvées N_r . La courbe de la précision selon le rappel (appelée PR) est également significative. Par exemple, une basse précision pour un rappel élevé indique que le système capture difficilement des images de la classe de l'image requête. Un système de recherche idéal est caractérisée par la courbe précision/rappel idéale tracée dans la figure 5.1.

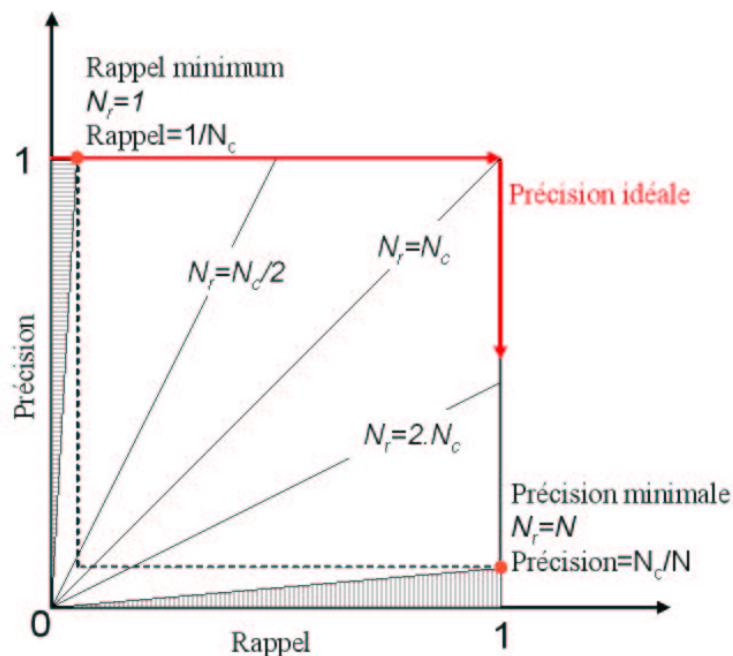


FIG. 5.1 – Courbe de précision/rappel PR d'un système idéal de recherche d'images. N_c est le nombre des images d'une classe donnée de la base de taille N , et N_r est le nombre des images retournées à l'utilisateur. Les points d'intersection de la courbe PR avec les droites tracées pour a) précision=rappel c'est à dire ($N_r = N_c$), b) précision=2*rappel ($N_r = N_c/2$) et c) précision=rappel/2 ($N_r = 2.N_c$) sont significatifs. Par exemple, pour savoir la valeur de la précision pour $N_r = N_c/2$, il suffit de trouver le point d'intersection de la courbe PR et la droite précision=2*rappel.

Notons que ces deux critères, précision et rappel, sont beaucoup plus performants pour l'évaluation d'un système de recherche interactive que l'erreur de classification. Une erreur

1. Ce sont les images retrouvées similaires et affichées à l'écran par le système

très basse peut être obtenue en prévoyant toujours la classe non pertinente, puisque nous avons quelques images pertinentes et de très nombreuses images non pertinentes.

Cependant, il n'y a aucun protocole commun d'évaluation de qualité entre les équipes de recherche travaillant dans le raffinement de la recherche d'images par boucle de pertinence. Dans le tableau 5.2 nous expliquons le protocole d'évaluation que nous proposons pour évaluer nos travaux.

Soient :

- N la taille de la base (le nombre total des images);
- b le nombre de boucles de pertinence effectuées;
- m le nombre des images annotées à chaque boucle de pertinence;
- lot l'ensemble des images parmi lesquelles sont tirées les meilleures,
 $N - b \times m$ est le nombre des images du lot.

Ainsi pour m (égal par exemple à 5) et pour une catégorie donnée :

Pour toutes les images de la catégorie, *faire* :

o Initialisation :

- image requête \leftarrow image courante;
- classifier les images de la base.

o Pour $b = 1$ à b_{max} , *faire* :

1. sélection des m meilleures images parmi le *lot*;
2. annotation des m images non précédemment annotées;
3. les images annotées sont retirées du lot pour le prochain tirage;
4. classification des images de la base;
5. calcul des critères pour N_r images retrouvées allant de 1 à N .

La précision et le rappel calculés sont moyennés sur toutes les images requêtes de la catégorie.

TAB. 5.2 – *Protocole d'évaluation*.

5.2.2 Résultats de recherche par EMiner

La figure 5.2 présente les performances de recherche par boucle de pertinence. Les différentes courbes représentent la précision/rappel PR d'une simulation sur la catégorie donnée pour 1, 2, 3, 4 et 5 bouclages à raison de $m = 5$ annotations par bouclage.

Nous observons une augmentation des performances grâce au bouclage de pertinence. Après chaque itération la courbe de précision/rappel pour n'importe quelle classe évolue pour se rapprocher de la courbe idéale (*cf.* figure 5.1). Cette évolution reste évidemment très dépendante de la structure ou de la complexité de chaque classe.

Pour savoir comment lire une courbe, prenons par exemple la catégorie "arborgreens" formée par $N_c = 47$ images.

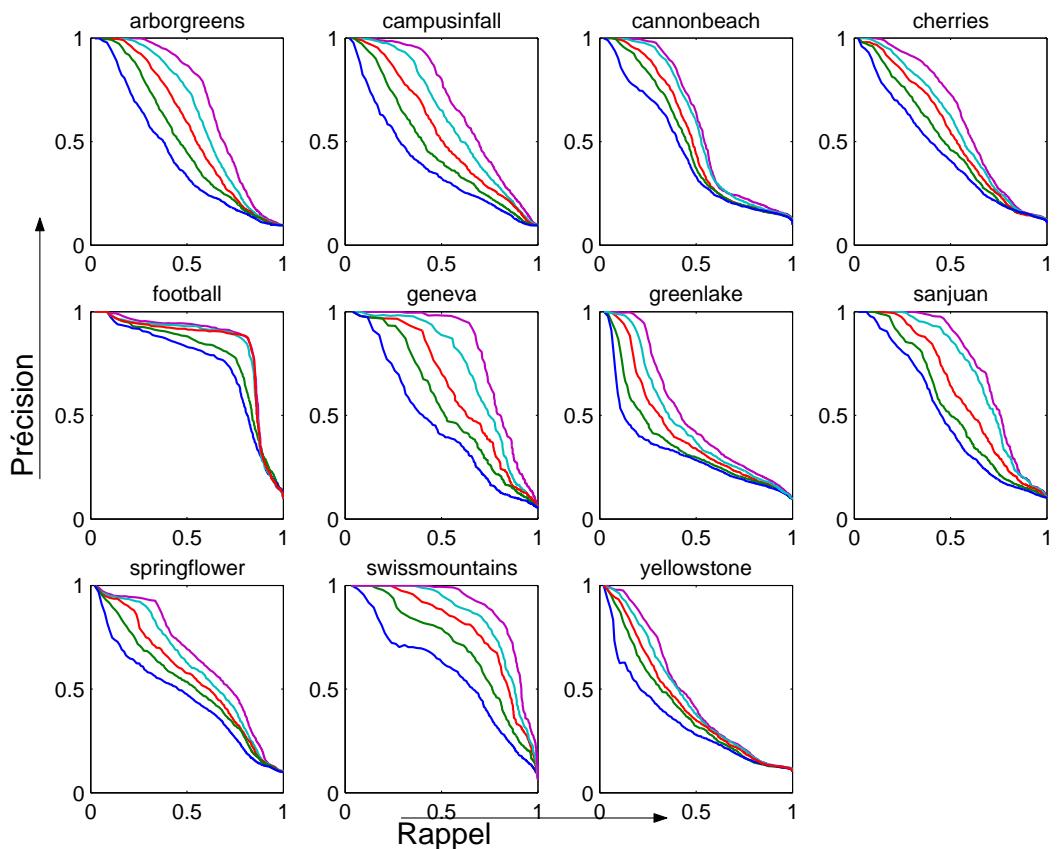


FIG. 5.2 – *Evaluation de EMiner. Les courbes PR sont tracées pour 1,2,3,4 et 5 bouclages de pertinence. 5 images ($m = 5$) sont annotées à chaque itération.*

La droite "précision=1" est importante pour l'interprétation des résultats. Observons la figure 5.3, ainsi pour 1, 2, 3, 4 et 5 bouclages de pertinence nous avons un rappel de 0.05, 0.1, 0.15, 0.18 et 0.27 respectivement. Cela signifie que si nous retournons à l'utilisateur $\lceil N_r = 0.05 * 47 \rceil = 3$ images, alors ces 3 images sont toutes pertinentes. Ceci est obtenu pour 5 images annotées seulement parmi 500 images environ non annotées de la base. Ce nombre augmente à 5, 8, 9 et 13 images qui sont toutes pertinentes retournées à l'utilisateur pour 10, 15, 20 et 25 images annotées parmi 500 images environ de la base.

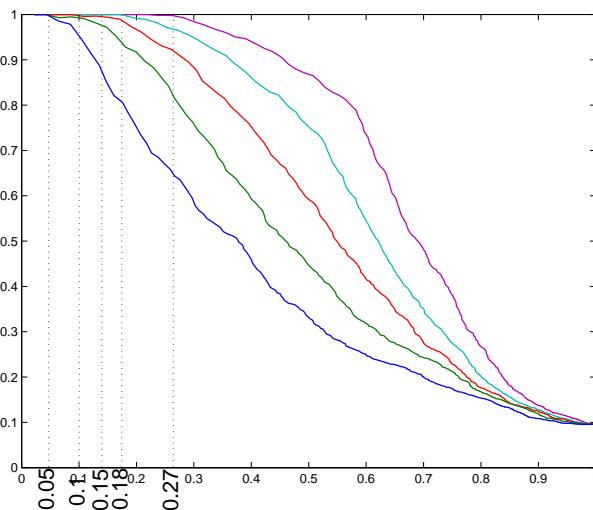


FIG. 5.3 – Intersection de la droite "précision=1" avec les courbes PR de la catégorie "arborgreens".

Passons à la droite "précision=rappel" de la figure 5.4. L'intersection de celle-ci avec les courbes PR pour 1, 2, 3, 4 et 5 bouclages de pertinence nous donne un rappel=précision de 0.425, 0.475, 0.54, 0.578 et 0.64 respectivement. Prenons la première boucle de pertinence c'est à dire pour 5 images annotées, un nombre d'images retrouvées pertinentes $N_{rp} = \lceil 0.425 * 47 \rceil = 20$ parmi $N_r = N_c = 47$ images retrouvées est constaté. Ensuite ce nombre d'images retrouvées pertinentes augmente à 23, 26, 28 et 31 respectivement parmi 47 images retrouvées pour 2, 3, 4 et 5 itérations respectivement.

Nous pouvons dire au vu des résultats, que les performances de notre système de recherche sont satisfaisantes mais il semble important de le comparer à une autre approche.

5.2.3 Comparaison entre EMiner et SVM

Les machines à vecteurs de support ont démontré leurs capacités dans le domaine de la reconnaissance de formes, et connaissent aujourd’hui un intérêt croissant dans la communauté de recherche d’images par le contenu (voir section 3.4.4 du chapitre 3).

L’objectif de cette section est de comparer notre algorithme EMiner de recherche interactive à une autre approche basée sur un classifieur SVM présentée ci-après et réalisée par

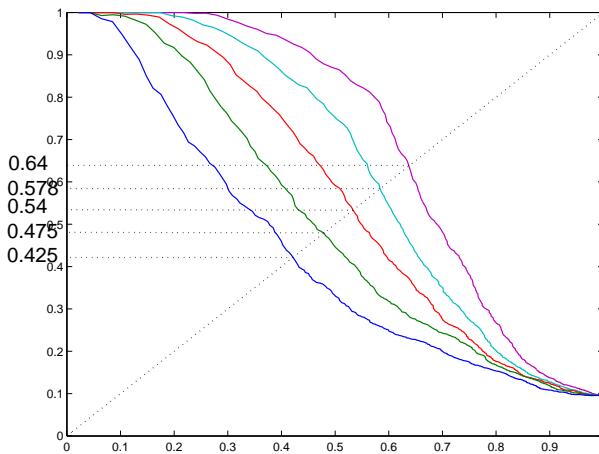


FIG. 5.4 – Intersection de la droite "précision=rappel" avec les courbes PR de la catégorie "arborgreens".

P. Gosselin et M. Cord² (Gosselin et al., 2004). Le protocole d'évaluation adopté est celui du tableau 5.2. Les tests sont établis sur la base ANN toujours avec les mêmes signatures de la section 5.2.2 fabriquées par le système RETIN de recherche d'images dans des bases généralistes.

Recherche interactive par des techniques SVM

Nous avons déjà expliqué dans la section 3.4.4 du chapitre 3 la technique de recherche interactive par SVM à deux classes. Elle consiste à trouver la frontière du classifieur qui sépare les images pertinentes des images non pertinentes utilisant les images annotées par l'utilisateur comme ensemble d'apprentissage pour estimer les paramètres de la frontière du classifieur. L'estimation des paramètres de la frontière est détaillée dans l'annexe B.

Pour se rappeler, soient $\{\mathbf{x}_i\}_{i \in \{1, \dots, l\}}$, les vecteurs index représentant les images annotées, et $\{y_i\}_{i \in \{1, \dots, l\}}$, $y_i \in \{-1, 1\}$ leurs annotations respectives (1 = pertinent, -1 = non pertinent). La mesure de similarité à la requête ou à la catégorie recherchée est :

$$J(\mathbf{x}) = \sum_{i=1}^l \alpha_i y_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i) + \beta_0.$$

Il existe une méthode SVM très proche de celle à deux classes est la technique SVM à une classe (Scholkopf et al., 1999) qui permet d'estimer le support d'une distribution d'un ensemble $\{\mathbf{x}_i\}_{i \in \{1, \dots, l\}}$ de vecteurs appartenant à la même classe. Nous avons déjà vu (section 3.4.4 du chapitre 3) l'application de cette méthode à la recherche interactive d'image où le problème revient à estimer le support des images pertinentes appartenant à

². Équipe de Traitement des Images et du Signal, University of Cergy-Pontoise / ENSEA, 6 av. du Ponceau 95014 Cergy-Pontoise Cedex, France.

la classe de l'image requête. Cependant la fonction de distribution de la classe pertinente peut être écrite (*cf.* équation 3.34) sous cette forme :

$$J(\mathbf{x}) = \sum_{i=1}^l \alpha_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i) + \beta_0.$$

L'algorithme SVM proposé dans ces expérimentations utilise ces deux approches SVM. Dans le cas où les deux types d'annotations (pertinentes et non pertinentes) sont disponibles c'est la technique SVM à deux classes qui est appelée. Dans le cas contraire, c'est à dire si seulement des images annotées pertinentes sont retournées c'est la technique SVM à une classe qui est exécutée.

Ainsi pendant la recherche interactive, le classifieur est entraîné avec la signature de l'image requête, puis les images avec la plus grande valeur de $J(\mathbf{x})$ sont affichées. L'utilisateur annote m images en fonction de leur appartenance à la classe recherchée. Par la suite, seules les annotations de l'utilisateur sont prises en compte pour déterminer la frontière qui sépare les images pertinentes des images non pertinentes (cas de SVM à deux classes) ou la densité des images pertinentes (cas de SVM à une classe). Le tableau 5.3 résume les étapes de la recherche interactive basée SVM.

1. Choisir la ou les image(s) requête(s);
2. Mettre à jour la liste des étiquettes des images annotées $\{y_1, \dots, y_l\}$;
3. Estimer les paramètres $\{\alpha_i\}_{i \in \{1, \dots, l\}}$ et β_0 du classifieur :
 - SVM à deux classes si les annotations sont positives et négatives,
 - SVM à une classe si les annotations sont toutes positives.
4. Classer les images de la base par l'ordre décroissant de :

$$J(\mathbf{x}) = \sum_{i=1}^l \alpha_i y_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i) + \beta_0;$$

5. Afficher à l'écran les N_r premières images (les plus similaires à la requête);
6. Récupérer les nouvelles annotations de l'utilisateur;
7. Retourner à l'étape 2 (si l'utilisateur n'est pas satisfait);

TAB. 5.3 – *Algorithme SVM de recherche d'images par boucle de pertinence.*

Le noyau choisi dans la technique SVM est gaussien : $\mathcal{K}(\mathbf{x}, \mathbf{x}_i) = e^{-\|\mathbf{x} - \mathbf{x}_i\|^2 / 2\sigma^2}$; tel que $\sigma = 1$.

Résultats de comparaison

Nous traçons les différentes courbes *PR* d'une simulation de EMiner et SVM sur toutes les catégories données pour 1, 2, 3, 4 et 5 bouclages à raison de $m = 5$ annotations par

bouclage. Au vu des résultats, chacune des deux méthodes semble avoir son domaine de préférence.

En effet et pour mettre en cause les résultats obtenus, nous avons projeté les index des images de la base dans un espace de dimension deux (2-D) suivant la méthode de Sammon (1969). Cet espace est obtenu pour une valeur de Stress (critère à minimiser) égale à 0.14. La connaissance d'une vérité terrain (*cf. section 5.2*) sur notre base nous a permis de séparer les différentes catégories d'images et d'observer spécifiquement leur répartition parmi le reste de la base. Les images de la classe étudiée sont repérées par des "o" rouges et celles du reste de la base sont représentées par des "." bleus. Cette méthode d'exploration donne une idée de la répartition des données (source d'inspiration) mais il est délicat d'en tirer des conclusions généralisables.

Nous avons observé quatre cas distincts :

1. Cas de la figure 5.5.a, EMiner est plus performant que SVM.

Dans l'espace 2-D (voir figure 5.5.b), les individus de chacune de ces catégories sont dispersés et mélangés avec le reste de la base. Ce qui explique "peut être" les résultats obtenus.

2. Cas de la figure 5.6.a, SVM est plus performant que EMiner. Pour les deux catégories cannonbeach et springflower, les courbes *PR* de EMiner sont meilleures que celles de SVM pour un rappel grand.

Dans la figure 5.6.b les individus de chaque catégorie ne sont pas mélangés avec d'autres individus d'autres catégories. On retrouve des points isolés dans le cas des catégories cannonbeach et springflower. Ce qui explique "peut être" le fait que ces individus sont retrouvés par EMiner avant SVM c'est à dire pour un N_r grand.

3. Cas de la figure 5.7.a, EMiner est plus performant pour N_r petit, ensuite SVM est plus performant pour un nombre N_r un peu plus élevé. EMiner est plus efficace à la fin pour N_r grand.

Si nous observons la répartition dans l'espace 2-D (*cf. figure 5.7.b*), nous trouvons que la plupart des individus sont groupés ensemble loin de quelques uns dispersés dans l'espace des attributs. Cela justifie "peut être" le comportement de EMiner et SVM.

4. Cas de la figure 5.8.a, les deux algorithmes donnent à peu près les mêmes résultats.

Dans la figure 5.8.b, les individus de la catégorie sont groupés ensemble et séparés du reste de la base.

En effet, le SVM travaille sur la frontière qui sépare la classe des images pertinentes de la classe des images non pertinentes, alors que le EMiner se concentre davantage sur les centres des deux classes. Le classifieur SVM découpe l'espace en deux morceaux, et si le nombre d'annotations est faible, cette séparation peut être inattendue (pour l'utilisateur). Par exemple, sur la catégorie "geneva" (voir des images exemples de la catégorie 5.9), rien n'empêche le fait que lors des premières annotations, certains modes de la classe recherchée soient classés comme non pertinents. L'EMiner, au contraire, puisqu'il se concentre sur les centres des classes, n'a pas ce genre de problème. Il recherche des images proches du centre et non des images proches d'une frontière mal établie par le SVM lors des pre-

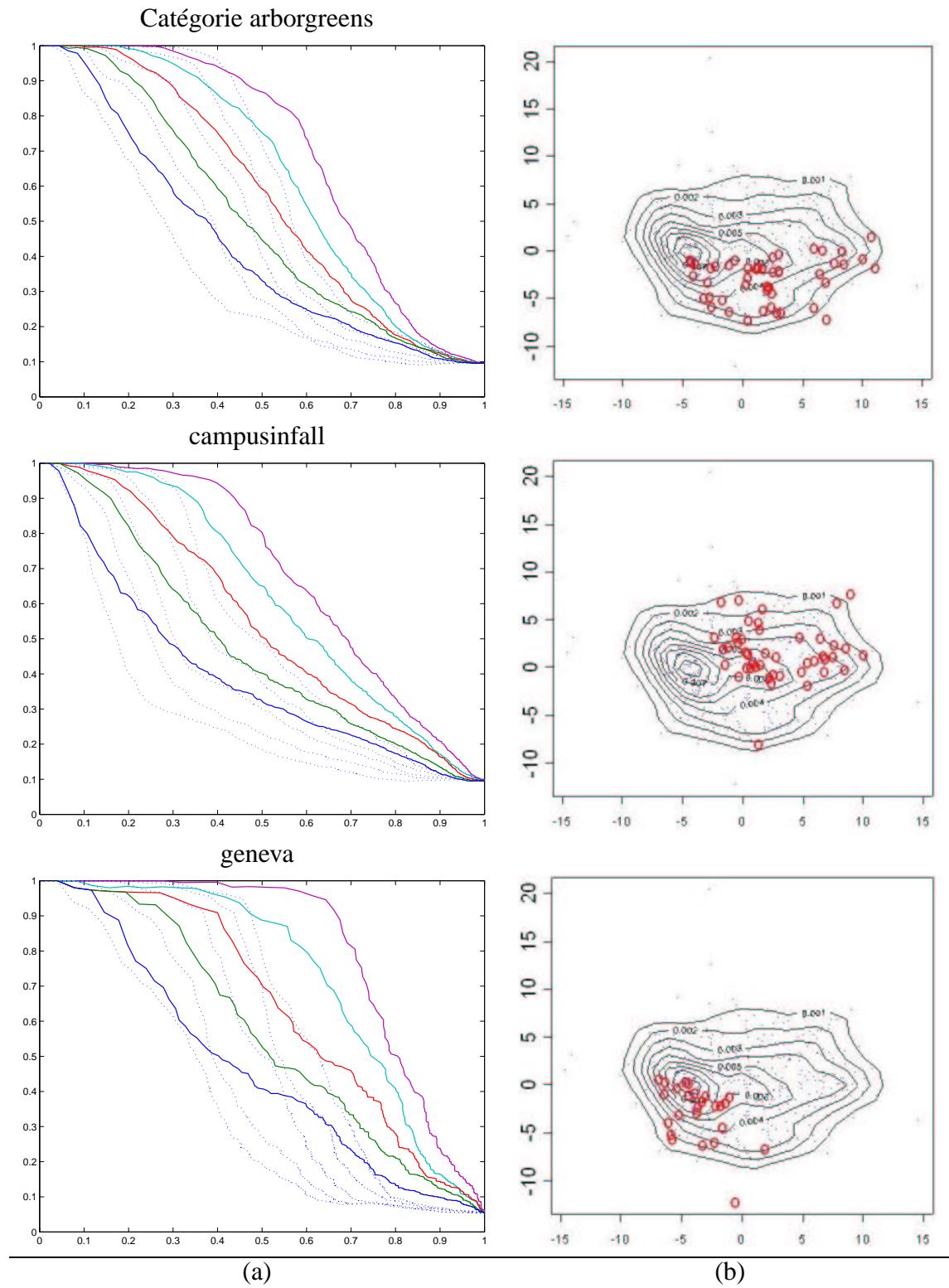


FIG. 5.5 – A gauche et de haut en bas, comparaison des courbes PR pour 1, 2, 3, 4, 5 bouclages de pertinence sur les images des trois catégories arbogreens, campusinfall et geneva. Les courbes obtenues par EMiner sont en trait plein et les courbes obtenues par SVM en pointillées. A droite et de haut en bas, répartition des images de la catégorie (en "o" rouges) parmi l'ensemble des images de la base ANN ("." bleus) au sein de l'espace 2-D obtenu par la méthode de Sammon (Stress = 0.14). EMiner est plus performant que SVM.

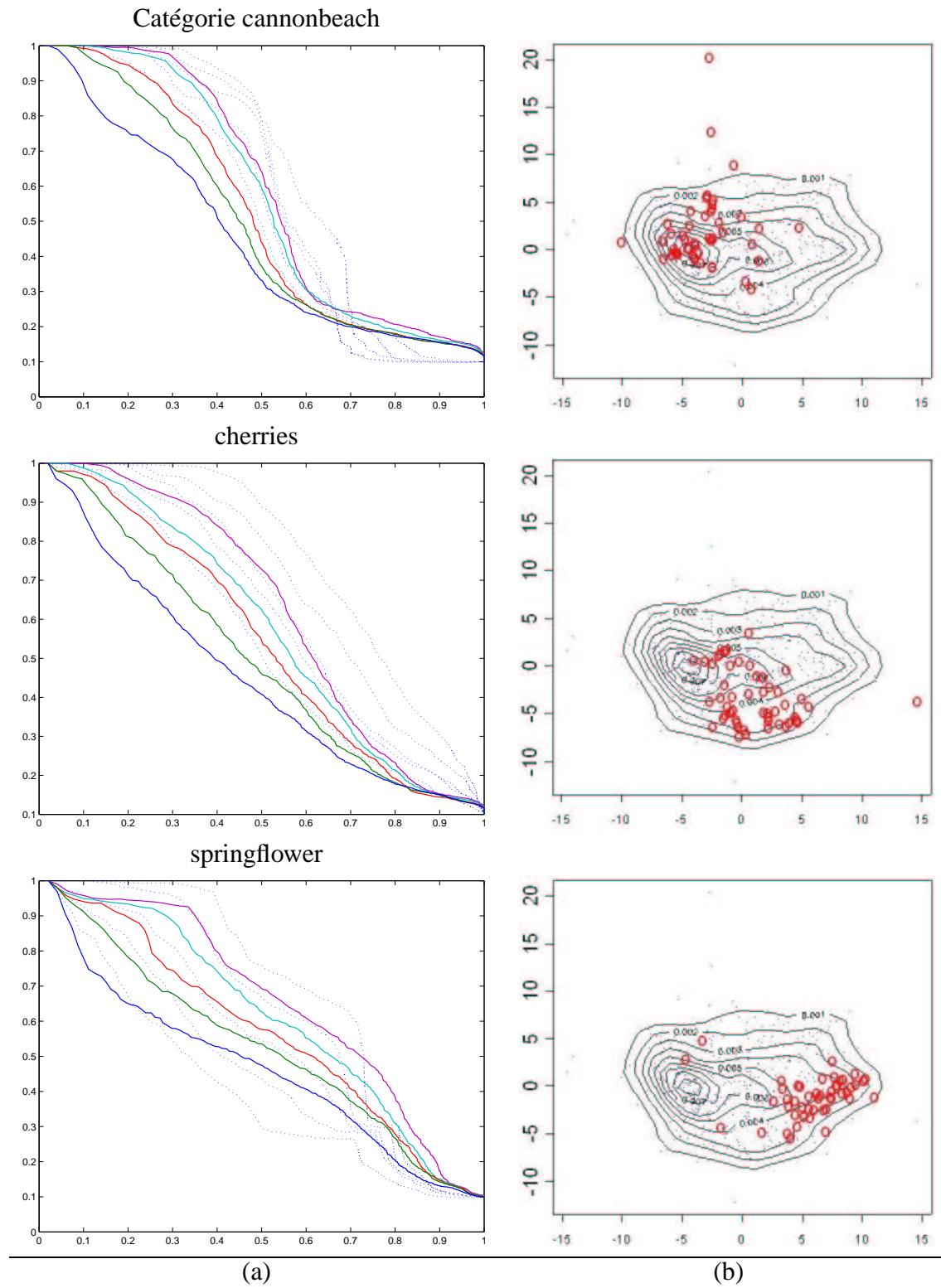


FIG. 5.6 – A gauche et de haut en bas, comparaison des courbes PR pour 1, 2, 3, 4, 5 bouclages de pertinence sur les images des trois catégories cannonbeach, cherries et springflower. Les courbes obtenues par EMiner sont en trait plein et les courbes obtenues par SVM en pointillées. A droite et de haut en bas, répartition des images de la catégorie (en "o" rouges) parmi l'ensemble des images de la base ANN ("." bleus) au sein de l'espace 2-D obtenu par la méthode de Sammon (Stress = 0.14). Les courbes PR de EMiner sont mieux que celles de SVM pour un rappel grand.

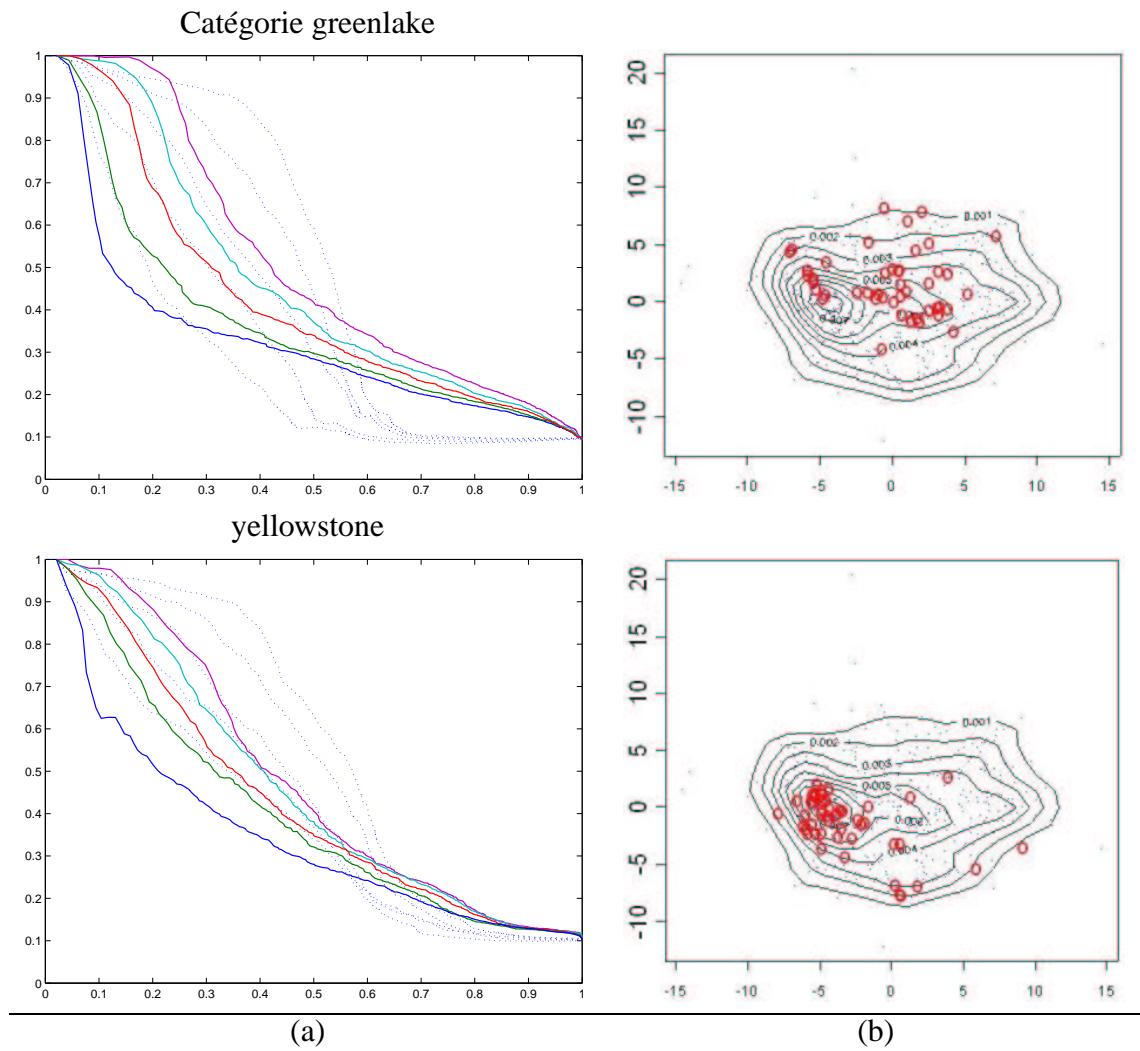


FIG. 5.7 – A gauche et de haut en bas, comparaison des courbes PR pour 1, 2, 3, 4, 5 bouclages de pertinence sur les images des deux catégories greenlake et yellowstone. Les courbes obtenues par EMiner sont en trait plein et les courbes obtenues par SVM en pointillées. A droite et de haut en bas, répartition des images de la catégorie (en "o" rouges) parmi l'ensemble des images de la base ANN ("." bleus) au sein de l'espace 2-D obtenu par la méthode de Sammon (Stress = 0.14). EMiner est plus performant pour N_r petit, ensuite SVM est plus performant pour un nombre N_r un peu plus élevé. EMiner retrouvent les points isolés à la fin pour N_r grand.

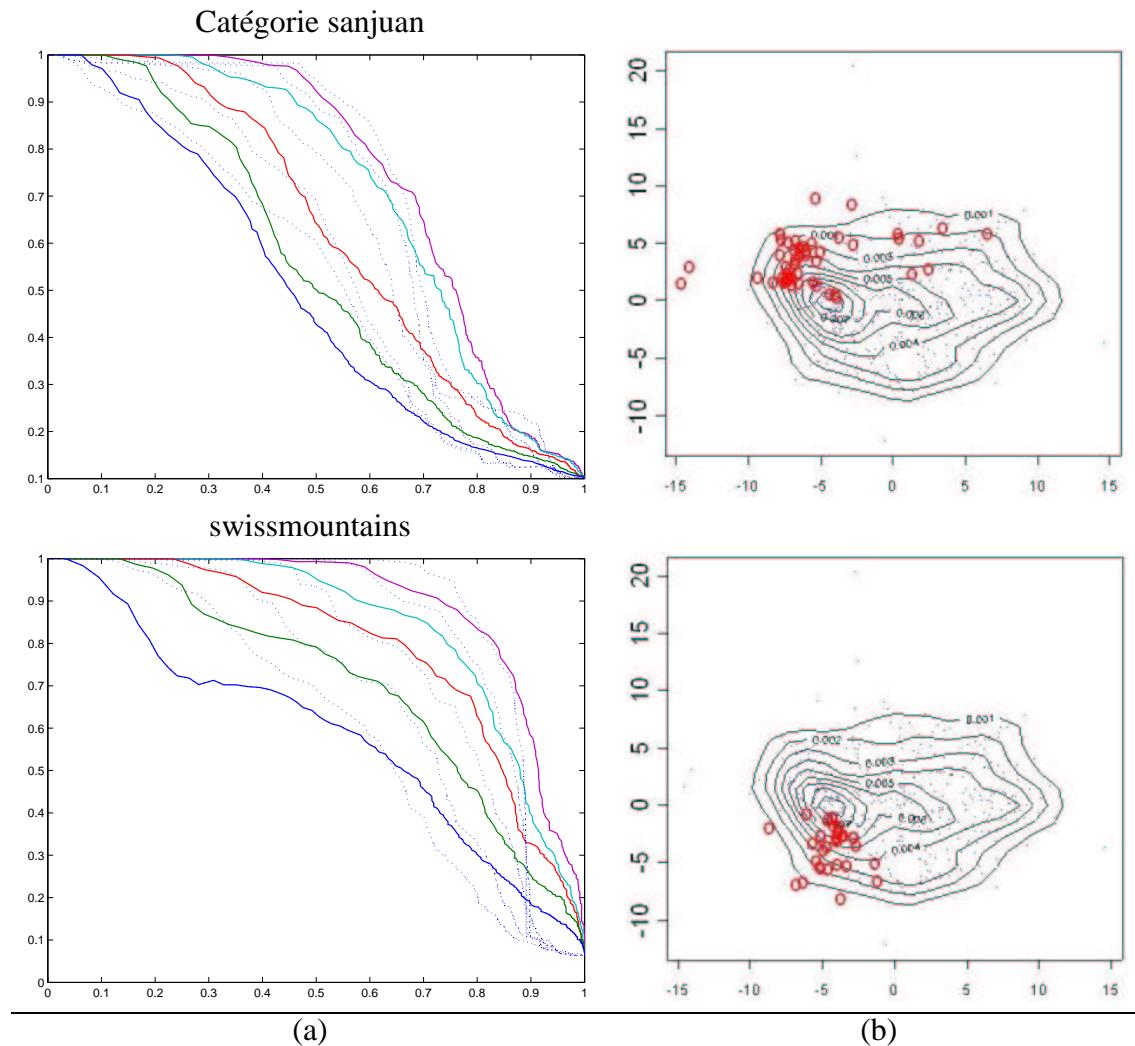


FIG. 5.8 – A gauche et de haut en bas, comparaison des courbes PR pour 1, 2, 3, 4, 5 bouclages de pertinence sur les images des deux catégories sanjuan et swissmountains. Les courbes obtenues par EMiner sont en trait plein et les courbes obtenues par SVM en pointillées. A droite et de haut en bas, répartition des images de la catégorie (en "o" rouges) parmi l'ensemble des images de la base ANN ("." bleus) au sein de l'espace 2-D obtenu par la méthode de Sammon (Stress = 0.14). Les deux algorithmes donnent à peu près les mêmes résultats.

mières itérations.

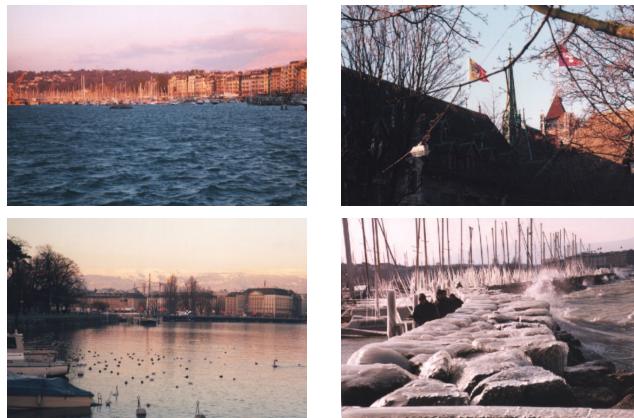


FIG. 5.9 – *Images de la catégorie "geneva".*

Pour la classe "cherries" (*cf.* figure 5.10), SVM est plus performant que EMiner. Cette classe est plutôt uni-modale et mélangée dans l'espace des attributs avec la classe "spring-flowers" (figure 5.11). Dans ce cas la précision apportée par la frontière des SVM peut expliquer ce résultat, puisqu'il sépare les deux classes sans difficulté. De plus, le caractère uni-modal de la classe ne contribue pas à la détermination d'une frontière inattendue. EMiner, par contre, semble moins précis, et sa capacité à mieux modéliser des classes multimodales lors des premières annotations ne peut être exploitée compte tenu de la forme de la classe.

Enfin, pour la classe "greenlake" (figure 5.12), nous sommes face à un cas situé entre les deux précédents, à la fois dispersée dans l'espace des attributs et proche d'une autre classe. Dans ce cas, il semble délicat d'affirmer la supériorité d'un algorithme par rapport à l'autre.

Une analyse simple du phénomène est montrée dans la table 5.4. La figure *a* de cette dernière représente la courbe *PR* obtenue par EMiner (ligne solide) et SVM (en pointillées) de la catégorie "greenlake" au cinquième bouclage de pertinence. Trois cas sont obtenus. Pour $N_r < 12$ EMiner est meilleur que SVM, pour $12 < N_r < 68$ SVM est plus performant et finalement pour $N_r > 68$ EMiner gagne SVM. Le nombre d'images retournées pertinentes N_{rp} (voir tableau 5.4) pour $N_r = 12$ ($N_r = 68$) est à peine près 11 (26).

Par ailleurs, les points de cette catégorie sont regroupés en trois modes (voir figure *b* de 5.4). Le plus grand mode contient 12 points et $45 - 29 = 16$ points sont dispersés dans l'espace 2-D.

Pour $N_r = \lceil 2.N_c \rceil = 90$ EMiner retrouve $N_{rp} = \lceil \text{précision} * N_r \rceil = 0.32 * 90 = 29$ images pertinentes et SVM retrouve $N_{rp} = 0.28 * 90 = 26$ images pertinentes. Cela montre peut-être que EMiner est efficace pour traiter des données dispersées.



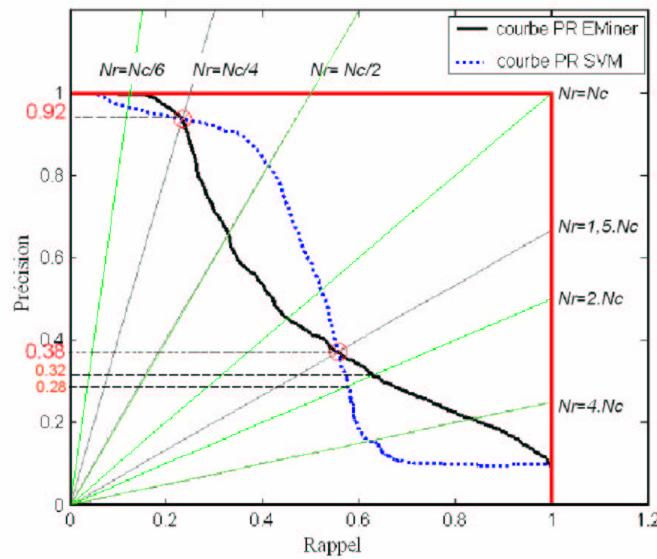
FIG. 5.10 – *Images de la catégorie "cherries".*



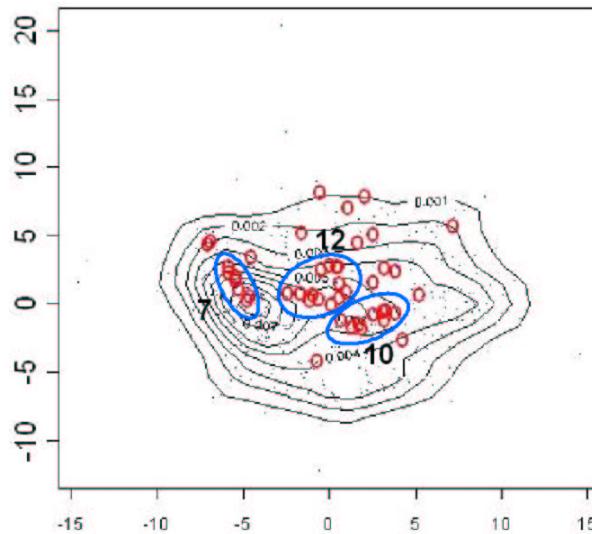
FIG. 5.11 – *Images de la catégorie "springflowers".*



FIG. 5.12 – *Images de la catégorie "greenlake".*



a- Courbe PR à la cinquième boucle de pertinence.



b- Répartition des images de la catégorie (Projection de Sammon).

| | | | |
|---------------------|---|------------------------------|------|
| N_c | 45 | | |
| # d'images annotées | 25 | | |
| N_r | $\lceil N_c/4 \rceil = 12$ | $\lceil 1.5.N_c \rceil = 68$ | 90 |
| précision | 0.92 | 0.38 | 0.32 |
| N_{rp} | $\lceil \text{précision} * N_r \rceil = 11$ | 26 | 29 |

c- Valeurs des points d'intersection des courbes PR de EMiner et SVM.

TAB. 5.4 – Exploration des résultats de recherche par EMiner et SVM sur la catégorie "greenlake".

Conclusion de la comparaison

L'intérêt croissant de la communauté CBIR pour la classification par SVM semble justifié. Cependant, face à un classifieur paramétrique comme EMiner, l'utilisation des SVM, ne permet pas d'obtenir des résultats aussi performants lorsque peu d'exemples sont classés. Au vu des résultats obtenus, les deux approches semblent plus se compléter que se concurrencer. D'une part, EMiner a davantage de recul sur les données puisqu'il apprend de toutes les images de la base (apprentissage semi-supervisé), ce qui lui permet de trouver des classes de formes plus complexes. D'autre part, les SVM ont une plus grande capacité à discriminer de manière précise une classe donnée. Un mélange des deux stratégies est une perspective à envisager. Par exemple, l'utilisation d'un EMiner lors des premières itérations afin de collecter un nombre suffisant de modes, suivi d'un SVM pour parfaire la discrimination, nous semble être une approche prometteuse.

5.3 Sélection de variables

La sélection de variables dans l'indexation et la recherche d'images demeure un problème non résolu à ce jour (Vasconcelos, 2003), du aux quantités massives de données utilisées dans la recherche et de la nécessité d'effectuer un apprentissage en ligne en réponse à l'interaction de l'utilisateur.

Pour pouvoir traiter différents types de problèmes de recherche d'images, différentes caractéristiques (liées à la texture, couleur, forme, ...) sont extraites de l'image pendant la phase d'indexation. D'autre part pendant la recherche en ligne, l'être humain est parfois intéressé par une seule caractéristique ou un sous-ensemble de caractéristiques. D'où le besoin d'une méthode flexible de sélection des variables (caractéristiques) qui réponde aux choix de l'utilisateur et qui rende l'apprentissage en ligne plus adapté aux besoins de l'être humain.

Par ailleurs, le problème de sélection de variables est un secteur de recherche très actif qui trouve ses applications dans les domaines de la reconnaissance de forme (classification, régression). De nombreuses méthodes de sélection de variables dédiés à la classification supervisée sont proposées dans la littérature. Le lecteur intéressé pourra consulter avec profit l'article de Kohavi et John (1997). Récemment, les propositions se sont multipliées dans le cadre de la sélection de variables dédiée à la classification non-supervisée (Dy et Brodley, 2000).

Dans ce qui suit, une méthode originale de sélection de variables dédiée à un apprentissage semi-supervisé est introduite. Elle est adaptée à notre méthode de recherche d'images basée sur une approche semi-supervisée et sur les lois de mélanges. Un critère d'évaluation basé sur les images annotées et non annotées ainsi qu'un algorithme rapide de sélection de variables, pour optimiser le temps de recherche en ligne, sont proposés.

Avant de détailler notre approche, nous dressons un état de l'art non exhaustif, des méthodes de sélection de variables existantes, consacrées à la classification supervisée ainsi que non supervisée.

5.3.1 Sélection de variables dédiée pour la classification supervisée

La sélection de variables est une technique de réduction de dimensionnalité, et est souvent employée comme prétraitement pour la classification (supervisée). La réduction de dimensionnalité des données a un certain nombre d'avantages. Les variables corrélées ou non pertinentes se comportent souvent comme un bruit dans les données et dégrade sa performance. L'élimination de telles variables a comme conséquence un sous-ensemble restreint de variables avec une performance de classification égale à (ou meilleure que) l'utilisation de tout l'ensemble. Un autre avantage est la nécessité de stocker moins de variables qui décrivent les données, et une classification plus rapide.

De plus, la réduction du nombre de variables (caractéristiques) aide l'expert humain à se focaliser sur un sous-ensemble de variables pertinentes, permettant une meilleure compréhension dans les processus décrits par les données.

Plusieurs études bibliographiques comparatives ont été réalisées sur les méthodes existantes en sélection de variables adressées à la classification supervisée. Le lecteur intéressé peut consulter (Molina et al., 2002; Kudo et Sklansky, 2000). Dans ce qui suit nous allons expliquer le schéma général regroupant ces méthodes.

Selon Molina et al. (2002), la relation entre l'algorithme de sélection de variables et l'algorithme de classification choisi pour évaluer l'utilité du procédé de sélection de variables, peut prendre trois formes principales : "incorporation" (embedded), "filtrage" (filter) et "bouclage" (wrapper).

- **Schéma "incorporation"** : L'algorithme de classification a son approche de sélection de variables (explicite ou implicite). Des outils traditionnels d'apprentissage statistique comme les arbres de décision ou certains réseaux de neurones artificiels sont inclus dans ce schéma (Mitchell, 1982).
- **Schéma "filtrage"** : Si le procédé de sélection de variables a lieu avant l'étape de classification (cf. figure 5.13), il peut être vu comme un filtre des variables non utiles avant la classification. Dans un sens général, ce schéma peut se voir comme cas particulier du schéma "incorporation" dans lequel la sélection de variables est employée comme un prétraitement. L'algorithme de sélection de variables dans ces schémas est indépendant de l'algorithme de classification (Liu et Setiono, 1996b; Kira et Rendell, 1992).
- **Schéma "bouclage"** : Dans ce schéma c'est l'algorithme de sélection de variables qui emploie l'algorithme d'apprentissage comme un sous-programme (Kohavi et John, 1997; Liu et Setiono, 1996a). L'évaluation d'un sous-ensemble de variables est réalisée par une méthode spécifique de classification, et employée soit par validation croisée sur l'ensemble d'apprentissage, soit sur deux ensembles séparés d'apprentissage et de test. De cette façon l'algorithme de sélection est bouclé "wrapped" autour de l'algorithme de classification. (voir figure 5.14). L'inconvénient principal de ce schéma est le temps de calculs énorme mis pour appeler l'algorithme de classification afin d'évaluer chaque sous-ensemble de variables.

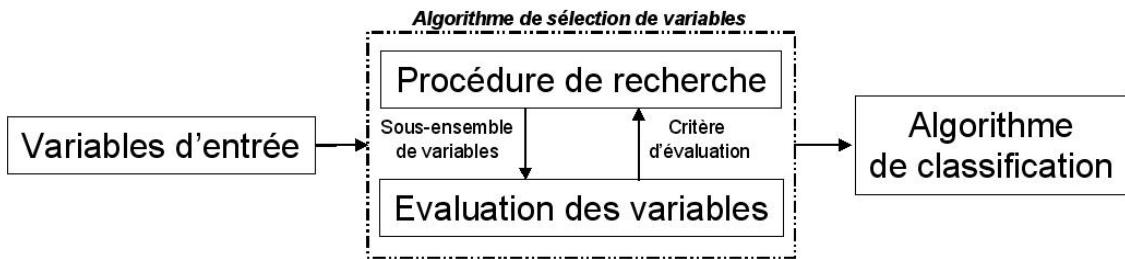


FIG. 5.13 – Schéma "filtrage". L'algorithme de sélection de variables est comme une phase de pré-traitement. Il est indépendant de la classification.

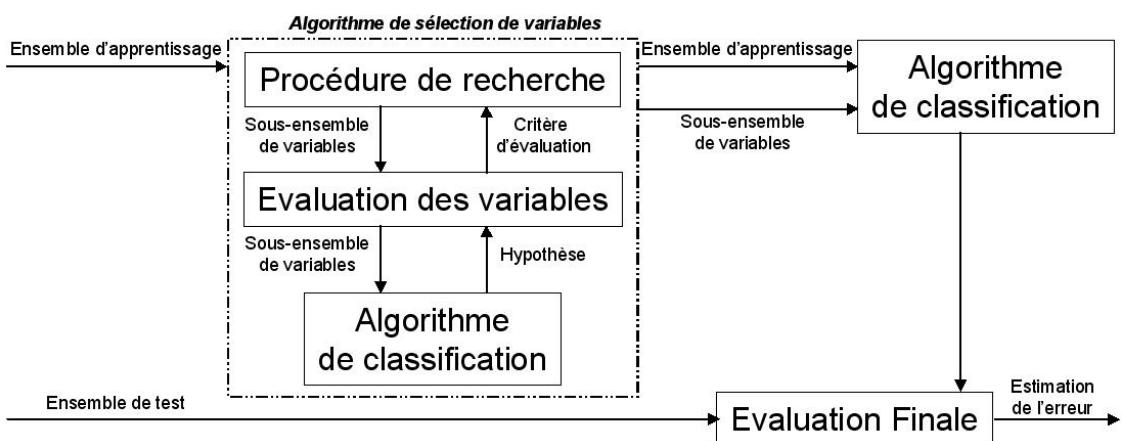


FIG. 5.14 – Schéma "bouclage". L'algorithme de classification est utilisé comme une boîte noire dans l'algorithme de sélection de variables.

Algorithmes de sélection de variables

Il est pratique de considérer les techniques de sélection de variables sous trois angles différents, trois composantes qu'il est possible de régler séparément :

1. un critère d'évaluation des variables, pour comparer différents sous-ensembles de variables et n'en retenir qu'un,
2. une procédure de recherche, pour explorer l'espace des différentes combinaisons de variables,
3. un critère d'arrêt, pour stopper la procédure de recherche ou déterminer l'ensemble de variables à sélectionner.

Critères d'évaluation Il est à rappeler que les critères d'évaluation que nous allons citer, sont utilisés dans les problèmes de sélection de variables dédiés aux classifications supervisées où les labels de l'ensemble d'apprentissage sont connus :

- Taux d'erreur en classification : ce critère est la base des méthodes de schéma "bouclage" dédiées pour une classification supervisée (Kohavi et John, 1997):

$$\hat{P}_e = 1 - \frac{|S_{TE}^+|}{|S_{TE}|}. \quad (5.9)$$

S_{TE} est l'ensemble test des données. S_{TE}^+ est le sous ensemble de S_{TE} que le classifieur a classé correctement.

- Divergence : elle mesure une distance probabiliste ou divergence entre les densités conditionnelles des classes. Si C_1 et C_2 sont deux classes, la formule générale de ce critère s'écrira :

$$\mathcal{J} = \int \mathcal{F}[f(\mathbf{x}|C_1), f(\mathbf{x}|C_2)]d\mathbf{x}. \quad (5.10)$$

\mathcal{F} peut être la fonction Chernoff, Bhattacharyya, Kullback-Leibler, Kolmogorov, Matusita ou Patrick-Fisher (Devijver et Kittler, 1982).

- Dépendance : ce critère mesure comment deux variables sont fortement associées entre elles, au sens où, sachant la valeur de l'une il est possible de prévoir la valeur de l'autre. Dans le contexte de la sélection de variables, une variable est mieux évaluée si elle prévoit mieux la classe. Le coefficient de corrélation est une mesure classique qui trouve toujours son application (Hall, 1999).
- Distances inter-classes : ces mesures sont fondées sur l'hypothèse que les exemples appartenant à des différentes classes sont éloignés dans l'espace des variables. Il est ainsi suffisant de définir une métrique entre les classes et de l'employer comme critère :

$$D(C_1, C_2) = \frac{1}{N_1 N_2} \sum_i^{N_1} \sum_j^{N_2} d(\mathbf{x}_{(i,C_1)}, \mathbf{x}_{(j,C_2)}). \quad (5.11)$$

N_1 (resp. N_2) étant le nombre d'exemples appartenant à C_1 (resp. C_2). $\mathbf{x}_{(i,C_1)}$ signifie un exemple i de C_1 .

- Information ou incertitude : De même que la dépendance probabiliste, nous pouvons observer \mathbf{x} et calculer la probabilité a posteriori $P(C_i|\mathbf{x})$ pour déterminer combien d'informations sur la classe de \mathbf{x} ont été obtenues, sachant sa probabilité *a priori*. Si toutes les classes deviennent également probables, alors le gain d'information est minimal et l'incertitude (entropie) est maximale.
- Consistence (Consistency) : plusieurs mesures de ce critère (voir Dash et al., 2000).

Procédures de recherche Les approches de sélection de variables calculent un critère fonction d'un ensemble (ou sous-ensemble) de variables. Cependant, une recherche approfondie nécessite le calcul du critère pour tous les sous-ensembles possibles et le sous-ensemble qui aura la valeur du critère désirée est retenu. C'est un travail exhaustif car ayant p variables, on aura $2^p - 1$ sous-ensembles. Des algorithmes sont utilisés pour réduire le temps de la recherche du sous-ensemble pertinent : "sélection séquentielle ascendante" (sequential forward selection), "élimination séquentielle descendante" (sequential backward elimination), "ascendante-descendante" (forward-backward) etc... (voir Kittler, 1978).

Critère d'arrêt Une fois que la méthode d'évaluation et celle de recherche ont été fixées, certaines méthodes de sélection de variables examinent tous les sous-ensembles fournis par la méthode de recherche.

Supposons que le critère d'évaluation (noté \mathcal{J}) est à maximiser et que p est le cardinal d'un sous-ensemble de variables. Il existe trois façons de choisir le sous-ensemble optimal et d'arrêter la sélection :

1. Fixer le cardinal du sous-ensemble à sélectionner, soit p_0 , et trouver le sous-ensemble de taille p_0 qui maximise le critère d'évaluation.
2. Choisir un seuil \mathcal{J}_0 du critère d'évaluation (la valeur minimale) et trouver le plus petit sous-ensemble de variables tel que le \mathcal{J} correspondant est supérieur ou égal à \mathcal{J}_0 .
3. Trouver un compromis entre minimiser p et maximiser \mathcal{J} .

Notons qu'un sous ensemble optimal n'est pas nécessairement unique.

5.3.2 Sélection de variables pour une classification non supervisée

Le travail dans la sélection de variables dédiée à la classification non supervisée est relativement nouveau. La plupart des approches sont adaptées à un algorithme de partitionnement (clustering) particulier. Devaney et Ram (1997) ont appliqué la recherche ascendante et descendante séquentielle pour évaluer chaque sous-ensemble de variables. Ils ont mesuré "la métrique de l'utilité de catégorie" des partitions (clusters) trouvées en appliquant le COBWEB (un algorithme hiérarchique de clustering) conjointement avec chaque sous-ensemble de variables. Talavera (1999) a appliqué l'approche "Blind" (semblable au schéma filtre) et l'approche "feedback" (analogique au bouclage) de COBWEB, et a employé une mesure de "dépendance" de variables pour choisir les variables. Vaithyanathan et Dom (1999) ont formulé une fonction cible pour choisir le sous-ensemble de variables et trouver

le nombre optimal de clusters pour un problème de clustering de documents en utilisant un cadre statistique bayésien d'évaluation. Agrawal et al. (1998) a présenté un algorithme de clustering (CLIQUE) qui procède étape-par-étape à partir d'une variable jusqu'à la dimension la plus élevée.

Inspirés par l'approche bouclage de Kohavi et John (1997), Dy et Brodley (2000) ont introduit une méthode de sélection de variables pour une classification non supervisée. Plutôt que de boucler la recherche du meilleur sous-ensemble de variables autour d'un algorithme de classification supervisée, elles ont bouclé la recherche autour d'un algorithme de classification non supervisée. Elles ont présenté deux méthodes : FSSEM (sélection de variables bouclée autour de l'algorithme EM) et FSSEM-k (FSSEM avec l'identification du nombre de clusters). Le terme EM représente l'algorithme Expectation-Maximisation (Dempster et al., 1977) appliqué pour estimer les paramètres de maximum de vraisemblance d'un mélange gaussien fini. Dy et Brodley proposent le critère de séparabilité $\text{trace}(\mathbf{S}_W^{-1}\mathbf{S}_B)$ comme critère de sélection de variables en apprentissage non-supervisé. Elles cherchent à trouver le sous-ensemble de variables qui sépare le mieux les classes (clusters en anglais). Ce critère a été choisi parmi plusieurs utilisés en analyse discriminante pour mesurer la séparabilité entre les classes (voir Fukunaga, 1990, page 446). Il est invariant par rapport aux transformations linéaires non singulières (Fukunaga, 1990).

Le terme \mathbf{S}_B désigne la matrice de dispersion inter-classe ("Between-class scatter matrix" en anglais) et montre la dispersion des centres des classes autour de la moyenne totale. \mathbf{S}_W est la matrice de dispersion intra-classe \mathbf{S}_W ("Within-class scatter matrix") et mesure la dispersion des membres d'une classe autour de son centre. Ayant K classes, ils sont définies comme suit :

$$\mathbf{S}_W = \sum_{k=1}^K p_k E^k \{(\mathbf{X} - \mathbf{M}_k)(\mathbf{X} - \mathbf{M}_k)^t\} = \sum_{k=1}^K p_k \boldsymbol{\Sigma}_k, \quad (5.12)$$

$$\mathbf{S}_B = \sum_{k=1}^K p_k (\mathbf{M}_k - \mathbf{M}_0)(\mathbf{M}_k - \mathbf{M}_0)^t, \quad (5.13)$$

$$\mathbf{M}_0 = E\{\mathbf{X}\} = \sum_{k=1}^K p_k \mathbf{M}_k, \quad (5.14)$$

$$\mathcal{J} = \text{trace}(\mathbf{S}_W^{-1}\mathbf{S}_B). \quad (5.15)$$

\mathbf{M}_k est le centre (moyenne) de la classe k et \mathbf{M}_0 est la moyenne de l'ensemble des données.

5.4 Notre approche

Dans ce chapitre, nous proposons une méthode originale de sélection de variables adaptée à l'apprentissage semi-supervisé basé sur les modèles de mélanges. L'approche proposée est similaire aux approches de type "bouclage" (voir figure 5.15) puisque l'algorithme de classification est intégré dans l'algorithme de sélection de variables. Le calcul du critère

d'évaluation est basé sur les paramètres estimés en apprentissage. En effet, les approches de type "bouclage", bien que souvent plus efficaces pour la recherche de variables pertinentes, sont coûteuses en temps de calcul. Désireux d'optimiser le temps de réponse de la procédure de recherche en ligne, nous avons opté pour une méthode rapide. Suite à l'estimation des paramètres du modèle, un critère de qualité est calculé pour chaque variable. Les variables ayant une valeur du critère inférieure (ou supérieure) à un seuil choisi à priori sont filtrées (retenues).

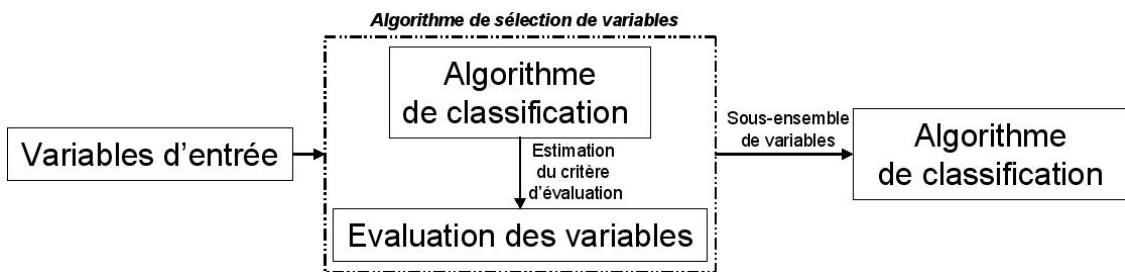


FIG. 5.15 – Schéma de notre approche de sélection de variables.

Cependant le schéma de notre procédure de recherche d'images avec la sélection des variables sera comme illustré dans la figure 5.16. Ainsi, à chaque boucle de pertinence :

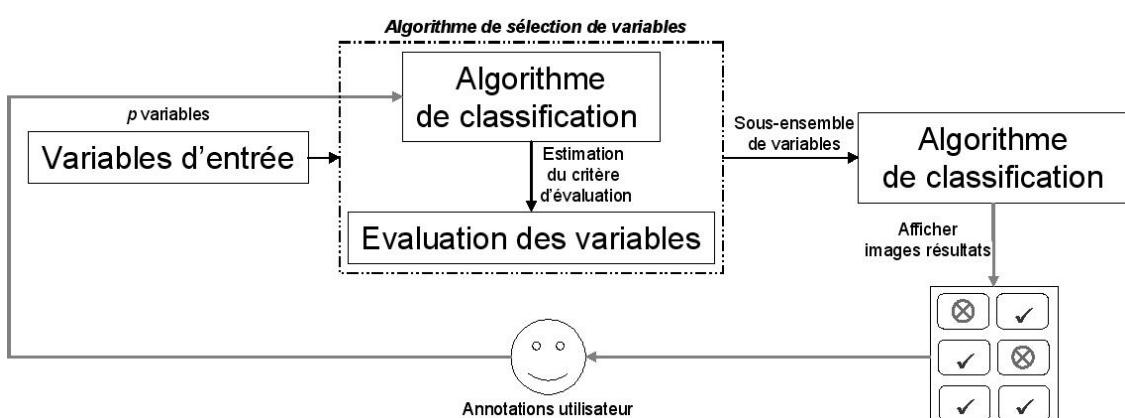


FIG. 5.16 – Schéma de la procédure de recherche d'images avec sélection des variables.

1. sélectionner les variables après estimation des paramètres du classifieur utilisant les nouvelles étiquettes,
2. classifier les images de la base de données en deux classes (images pertinentes et non pertinentes) en employant les images annotées et les variables sélectionnées,
3. afficher les images semblables,
4. récupérer les nouvelles annotations de l'utilisateur puis aller à l'étape 1 (s'il n'est pas satisfait).

5.4.1 Critère

Face à un problème d'apprentissage semi-supervisé où seulement certaines étiquettes sont disponibles, un critère fonction des données étiquetées et non étiquetées semble s'imposer. L'idée de base consiste à parier que le classifieur donne des résultats cohérents et à favoriser les variables qui permettent d'obtenir une séparation nette entre classe pertinente et classe non-pertinente. Pour chaque image, la classe d'appartenance la plus probable est déterminée. Le critère considérée est la moyenne des probabilités d'appartenance. Si cette moyenne est proche de la probabilité a priori de la classe la plus probable ($\max(\Pi_1, \Pi_2)$, voir équations 5.4) cela signifie que le classifieur ne fait que très peu de distinction entre les classes pertinente et non pertinente et a contrario si la moyenne observée est proche de 1, le classifieur donne des résultats très tranchés (mais éventuellement faux).

Lorsqu'une image est étiquetée, sa probabilité d'appartenance à sa classe d'origine est prise en compte et lorsque l'image ne possède pas d'étiquette la probabilité d'appartenance à la classe la plus probable est prise en compte.

$$Cr = \frac{1}{N} \sum_{i=1}^N \sum_{r=1}^R s_{ir} p(r|\mathbf{x}_i, \mathbf{z}_i), \quad (5.16)$$

avec

- si \mathbf{x}_i est étiqueté $s_{ir} = z_{ir}$,
- sinon (\mathbf{x}_i n'est pas étiqueté) :

$$s_{ir} = \begin{cases} 1 & \text{si } r \in C_k = \max(\sum_{l=1}^{R1} p(l|\mathbf{x}_i, \mathbf{z}_i), \sum_{l=R1+1}^{R1+R2} p(l|\mathbf{x}_i, \mathbf{z}_i)), \\ 0 & \text{sinon.} \end{cases}$$

Remarquons que le critère inventé pourra être utilisé en apprentissage partiellement supervisé.

5.4.2 Algorithme

La méthode proposée consiste à projeter le critère Cr sur les axes de l'espace \mathbb{R}^p . Ainsi chaque variable se caractérise par une valeur du critère. Lorsque le critère est élevé (proche de 1), les images pertinentes et non-pertinentes (au sens du mélange) sont très séparées.

Soit Cr^j , avec $j = \{1, \dots, p\}$, la projection de Cr sur l'axe j . De l'équation 5.16, Cr^j s'écrit de la façon suivante:

$$Cr^j = \frac{1}{N} \sum_{i=1}^N \sum_{r=1}^R s_{ir}^j p^j(r|\mathbf{x}_i, \mathbf{z}_i), \quad (5.17)$$

L'exposant j indice les grandeurs projetées sur l'axe j . Par exemple $p^j(r|\mathbf{x}_i, \mathbf{z}_i)$ est la projection de $p(r|\mathbf{x}_i, \mathbf{z}_i)$ sur l'axe j . s_{ir}^j sera :

- si \mathbf{x}_i est étiqueté $s_{ir}^j = z_{ir}$,

– sinon :

$$s_{ir}^j = \begin{cases} 1 & \text{si } r \in C_k = \max(\sum_{l=1}^{R1} p^j(l|\mathbf{x}_i, \mathbf{z}_i), \sum_{l=R_1+1}^{R1+R2} p^j(l|\mathbf{x}_i, \mathbf{z}_i)), \\ 0 & \text{sinon.} \end{cases}$$

Dans le cas présent, nous avons supposé les matrices de variance covariance diagonales :

$$\Sigma_r = \begin{pmatrix} \sigma_{r1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{r2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{rp}^2 \end{pmatrix}. \quad (5.18)$$

Cette dernière hypothèse nous permet de réduire le nombre de paramètres à estimer en gardant toujours la possibilité d'avoir des classes relativement complexes, et d'augmenter ainsi la robustesse de la méthode.

Lorsque l'on suppose l'indépendance entre les variables conditionnellement à l'appartenance à une composante (matrice de variance diagonale), la densité marginale liée à une variable d'un modèle de mélange est un modèle de mélange des densités marginales de ses composantes (voir 5.4.3). Ainsi on observe que :

$$p^j(r|\mathbf{x}_i) = \frac{\pi_r \cdot f_r^j(\mathbf{x}_i|\boldsymbol{\theta}_r)}{f^j(\mathbf{x}_i) = \sum_l \pi_l \cdot f_l^j(\mathbf{x}_i|\boldsymbol{\theta}_l)}. \quad (5.19)$$

Les composantes sont supposées être mélangées dans les mêmes proportions $\{\pi_r\}_{r=1}^R$ qu'avant projection. $f_r^j(\mathbf{x}_i|\boldsymbol{\theta}_r)$ est la projection (densité marginale) de $f_r(\mathbf{x}_i|\boldsymbol{\theta}_r)$ sur l'axe j .

Dans le cas d'un mélange gaussien, la densité $f_r(\mathbf{x}_i|\boldsymbol{\theta}_r)$ de la composante r sachant \mathbf{x}_i s'exprime comme

$$f_r(\mathbf{x}_i|\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) = \frac{(2\pi)^{-\frac{p}{2}}}{\det |\boldsymbol{\Sigma}_r|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_r)^t \boldsymbol{\Sigma}_r^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_r)}, \quad (5.20)$$

où $\boldsymbol{\mu}_r$ et $\boldsymbol{\Sigma}_r$ sont respectivement le vecteur moyenne et la matrice de variance covariance de la composante r . En conséquence (voir 5.4.4), les densités marginales se calculent sans estimer les paramètres du mélange dans chaque sous-espace

$$f_r^j(\mathbf{x}_i|\boldsymbol{\theta}_r) = \frac{1}{\sqrt{2\pi} \cdot \sigma_{rj}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x_{ij} - \mu_{rj}}{\sigma_{rj}} \right)^2}, \quad (5.21)$$

avec μ_{rj} est la composante j de la moyenne $\boldsymbol{\mu}_r = (\mu_{r1}, \dots, \mu_{rp})$.

Finalement, les équations 5.19 et 5.21 permettent de calculer rapidement le critère projeté de l'équation 5.17.

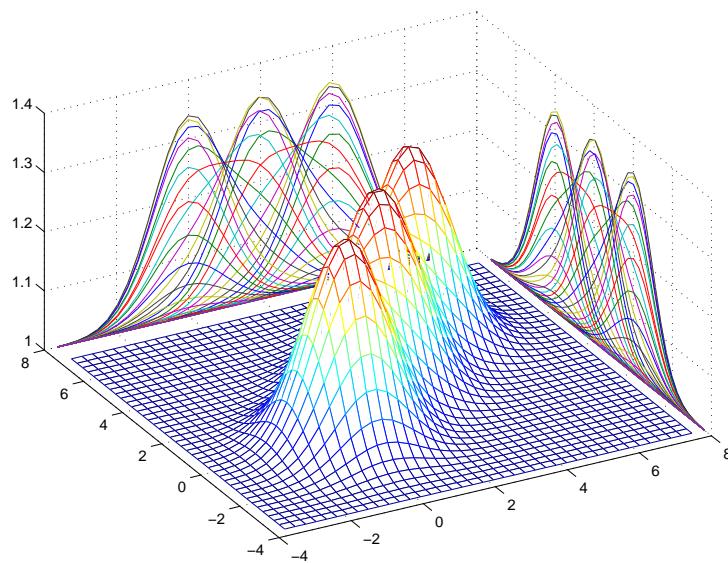


FIG. 5.17 – Projection d'un mélange de gaussiennes (cas de $p = 2$) sur les axes x et y .

5.4.3 Remarque 1

Soit un mélange de deux lois probabilistes de densités f_1 et f_2 fonctions de deux variables x et y . La densité de ce mélange est :

$$f(x,y) = p_1 f_1(x,y) + p_2 f_2(x,y). \quad (5.22)$$

Si $f(x)$ la densité marginale de $f(x,y)$, elle est égale, par sa définition, à :

$$\begin{aligned} f(x) &= \int_{V_y} (p_1 f_1(x,y) + p_2 f_2(x,y)) dy \\ &= \int_{V_y} p_1 f_1(x,y) dy + \int_{V_y} p_2 f_2(x,y) dy \\ &= p_1 \int_{V_y} f_1(x,y) dy + p_2 \int_{V_y} f_2(x,y) dy \\ &= p_1 f_1(x) + p_2 f_2(x). \end{aligned}$$

$f_1(x)$ et $f_2(x)$ étant les densités marginales de $f_1(x,y)$ et $f_2(x,y)$ respectivement.

Ainsi la densité marginale d'un mélange de lois probabilistes est un mélange (avec la même proportion) des densité marginales de chacun des lois.

5.4.4 Remarque 2

La densité d'une gaussienne de moyenne $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)$ et matrice de variance-covariance diagonale $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p^2 \end{pmatrix}$ est égale à :

$$\begin{aligned} f(x_1, \dots, x_p) &= \frac{(2\pi)^{-p/2}}{(\sigma_1 \dots \sigma_p)} \cdot e^{-\frac{1}{2} \cdot \sum_{j=1}^p \left(\frac{x_j - \mu_j}{\sigma_j} \right)^2} \\ &= \left\{ \frac{(2\pi)^{-1/2}}{\sigma_1} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2} \right\} \times \dots \times \left\{ \frac{(2\pi)^{-1/2}}{\sigma_p} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x_p - \mu_p}{\sigma_p} \right)^2} \right\} \\ &= f(x_1) \times \dots \times f(x_p). \end{aligned}$$

5.4.5 Expérimentations

Nos expérimentations sont conçues pour évaluer notre algorithme de sélection de variables selon 1) la capacité de choisir les variables pertinentes et 2) l'amélioration potentielle des résultats de recherche par EMiner. Nous présentons d'abord des expériences avec des données synthétiques et puis une analyse détaillée en utilisant des données réelles.

Illustration du critère avec des données synthétiques

Les données synthétiques sont constituées par des variables pertinentes et non-pertinentes. Les variables pertinentes sont les variables issues du modèle de mélange gaussien. Les variables non pertinentes sont produites uniformément distribuées.

En fait, nous avons simulé $N = 1000$ exemples $\{\mathbf{x}_i\}_{i=1,\dots,1000}$ décrits par 2 variables (pertinentes) et générés par un mélange de quatre gaussiennes, $\mathcal{N}_1(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $\mathcal{N}_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, $\mathcal{N}_3(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$ et $\mathcal{N}_4(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_4)$ selon l'équation :

$$f(\mathbf{x}_i) = p_1 \cdot f_1(\mathbf{x}_i) + p_2 \cdot f_2(\mathbf{x}_i) + p_3 \cdot f_3(\mathbf{x}_i) + p_4 \cdot f_4(\mathbf{x}_i);$$

Les quatre gaussiennes sont équiprobables ($p_1 = p_2 = p_3 = p_4 = 1/4$). Ensuite un autre ensemble, de 1000 exemples décrits par 7 variables (non pertinentes) suivant une loi normale uniforme $\{\mathbf{w}_i\}_{i=1,\dots,1000}$, est créé.

Le jeu de données utilisé dans les expérimentations est donc : $D = \{(\mathbf{x}_i, \mathbf{w}_i)\}_{i=1,\dots,1000}$.

Le problème est supposé un problème de deux classes C_1 et C_2 . La classe C_1 est formée par la gaussienne \mathcal{N}_1 et la deuxième classe C_2 est formée par les trois autres gaussiennes ($\mathcal{N}_2, \mathcal{N}_3, \mathcal{N}_4$). Les paramètres des gaussiennes sont choisis tels que les matrices de variance covariance sont diagonales et le taux d'erreur minimale (classifieur bayésien) est (a) 4% et (b) 24% (voir le tableau 5.5). Les figures (a) et (b) de 5.18 illustrent le mélange des gaussiennes choisi dans les deux cas.

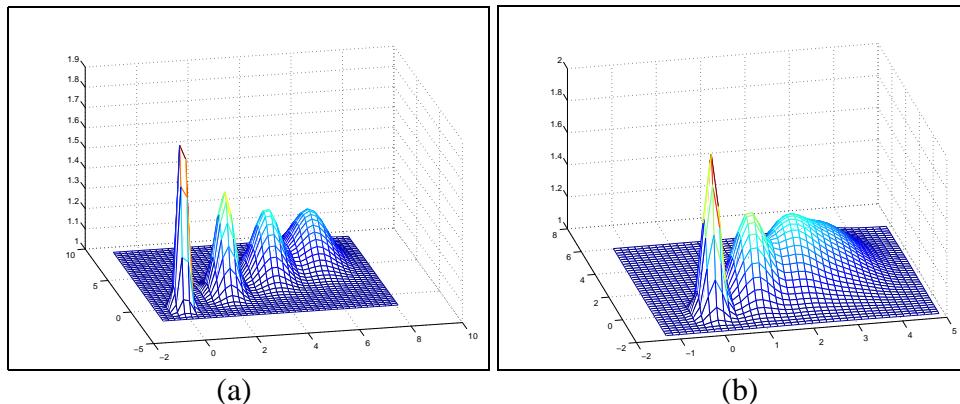


FIG. 5.18 – Données gaussiennes: (a) l’erreur minimale du classifieur bayésien est égale à 5% et dans (b) 24%.

| | \mathcal{N}_1 | \mathcal{N}_2 | \mathcal{N}_3 | \mathcal{N}_4 |
|-----------------|--|--|--|--|
| (a) μ | (0,0) | (2,2) | (4,4) | (6,6) |
| (b) μ | (0,0) | (1,1) | (2,2) | (3,3) |
| (a, b) Σ | $\begin{pmatrix} 0.163 & 0 \\ 0 & 0.381 \end{pmatrix}$ | $\begin{pmatrix} 0.327 & 0 \\ 0 & 0.763 \end{pmatrix}$ | $\begin{pmatrix} 0.491 & 0 \\ 0 & 1.145 \end{pmatrix}$ | $\begin{pmatrix} 0.654 & 0 \\ 0 & 1.527 \end{pmatrix}$ |

TAB. 5.5 – Paramètres des données gaussiennes : (a) l’erreur minimale est égale à 5%, (b) elle est égale à 24%.

Pour différentes proportions d’exemples classés allant de 0 (cas d’une classification non supervisée) jusqu’à 1 (cas d’une classification supervisée), le critère d’évaluation Cr^j de l’équation 5.17 est calculé pour les 9 variables.

En regardant la figure (a) de 5.19, on voit clairement que les deux variables produisant les gaussiennes ont un critère plus grand que $\Pi_2 = \max(\Pi_1, \Pi_2) = 3 \times 1/4$ et proche de 1. Plus la proportion des exemples classés augmente, plus le critère $\{Cr^j\}_{j=1,2}$ des variables pertinentes s’approche de 1 et $\{Cr^j\}_{j=3,\dots,9}$ des variables non pertinentes diminue et par conséquence le problème de séparations des variables pertinentes et non pertinentes devient plus facile. Cela confirme notre choix de critère comme expliqué dans le paragraphe 5.4.1. Les mêmes résultats à quelques dégradations près, sont obtenus dans le cas de la figure (b) de 5.19 où les gaussiennes sont conçues telles que l’erreur minimale de classification est à peu près 24%.

Recherche avec sélection de variables dans la base ANN

Dans ce paragraphe, nous évaluons notre algorithme de recherche EMiner sur les données (index) de la base ANN, accompagné de la méthode de sélection de variables que nous avons proposée. Le protocole d’évaluation adopté est celui de la section précédente (table 5.2) où l’étape 4 de classification est remplacée par le schéma de la figure 5.15. A l’issue de cette évaluation réalisée pour $m = 5$ images annotées à chaque itération, nous avons

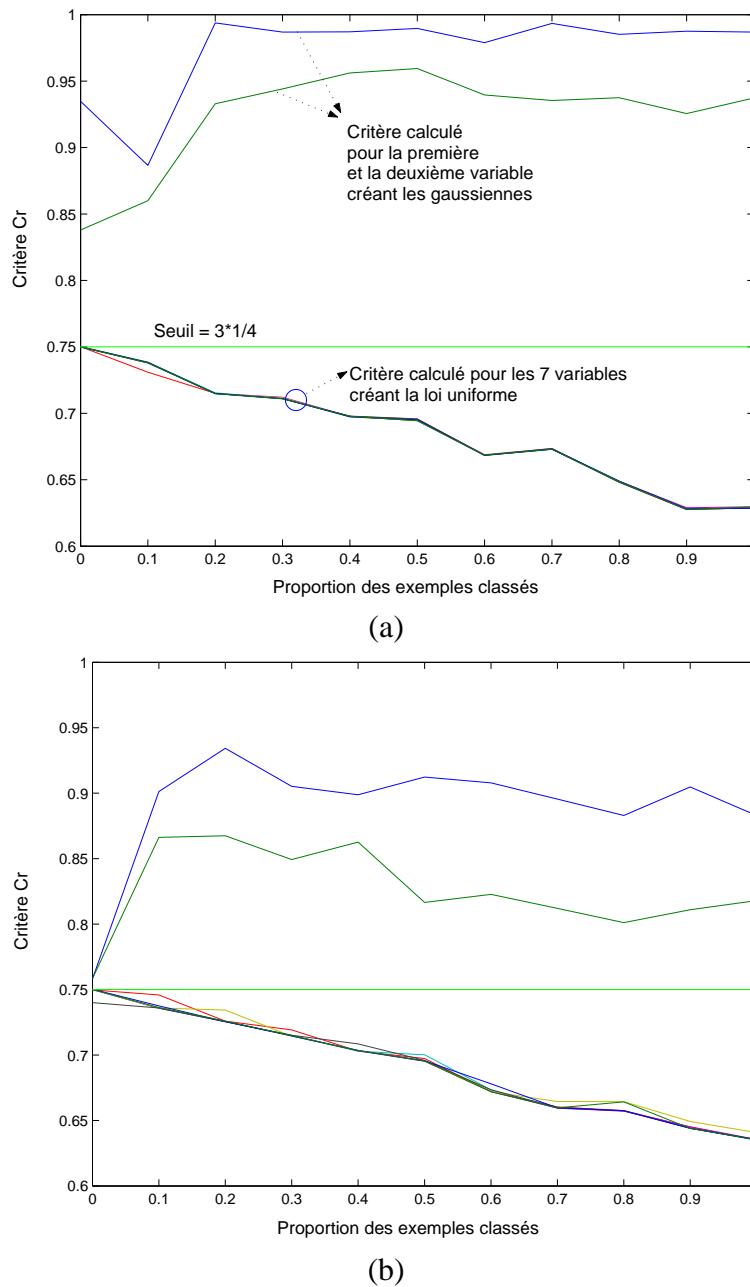


FIG. 5.19 – Critère d'évaluation Cr^j calculé pour toutes les variables en fonction de la proportion des exemples classés. Les variables générant les gaussiennes ont une valeur du critère au dessus du seuil qui est égal à $\max(\Pi_1, \Pi_2) = 3/4$. La valeur du critère des variables non pertinentes créant le bruit est inférieur au seuil. Plus qu'on augmente la proportion des exemples classés plus la distinction entre variable pertinente et non pertinente devient claire. Les gaussiennes sont générées telles que l'erreur minimale de classification est de 5% dans (a) et 24% (b).

tracé (figure 5.20) les courbes de précision/rappel PR (lignes solides) obtenues après 5 bouclages de pertinence et pour toutes les catégories de la base. Sur le même graphe (figure 5.20), nous illustrons (en pointillés) les courbes d'évaluation PR de EMiner avant sélection de variables ainsi que le nombre de variables retenues après sélection.

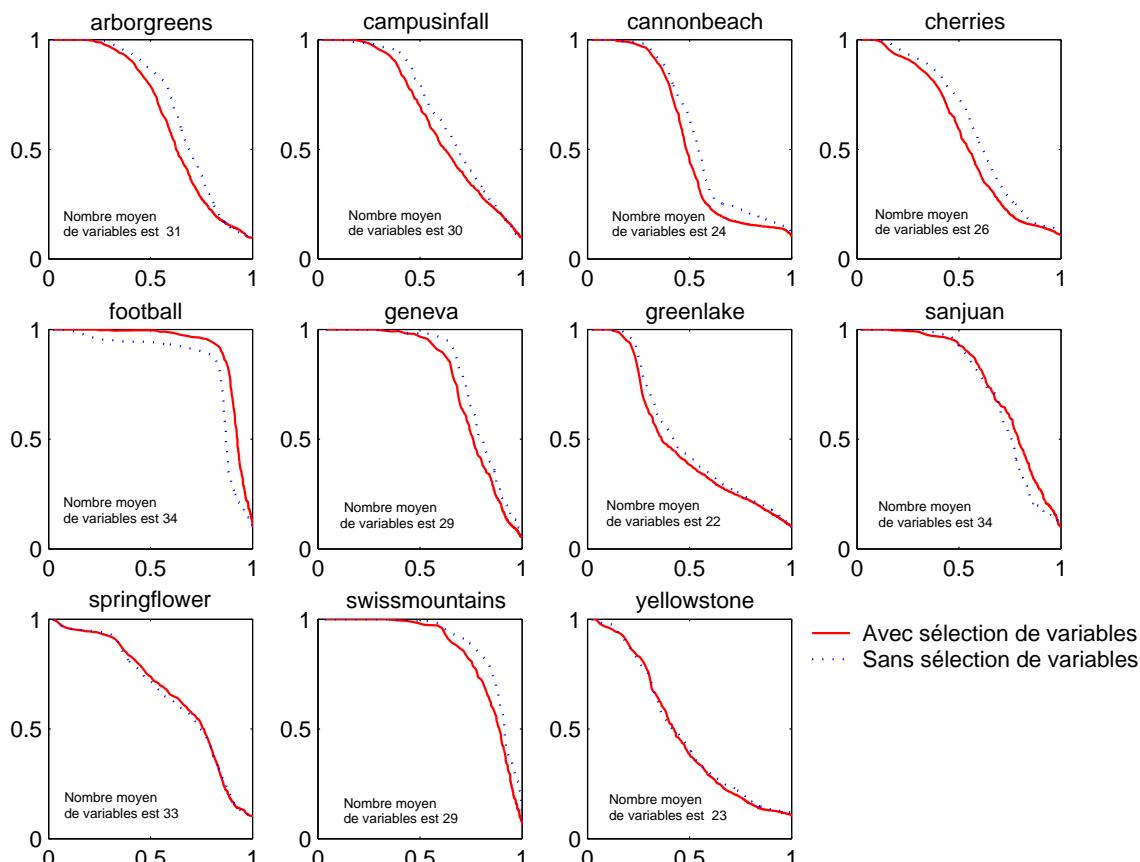


FIG. 5.20 – Les courbes PR obtenus après évaluation du système sans (ligne en pointillés) et avec sélection de variables (ligne solide). Ces courbes sont tracées pour $m = 5$ (images annotées à chaque bouclage) et à la 5ème boucle de pertinence. Après sélection de variables, ces courbes PR sont parfois mieux qu'avant sélection (cas de la catégorie "Football"), parfois moins bonnes à quelques dégradations près (cas de la catégorie "arborgreen") et finalement pareilles comme dans le cas de la catégorie "yellowstone". Le nombre moyen de variables retenues après sélection (affiché sur chaque catégorie) varie entre 22 et 34 partant de 50 variables au départ (avant sélection).

Au vu des résultats, nous distinguons trois cas :

1. une amélioration des performances après sélection de variables dans le cas des catégories "football", "sanjuan" et "springflower",
2. une conservation des performances, cas des catégories "greenlake" et "yellowstone",
3. une dégradation très faible des performances pour le reste.

La dégradation peut être expliquée par le fait que la catégorie en question est dispersée dans l'espace des attributs. En outre la sélection des variables est réalisée avec un faible nombre d'images annotées comme dans le cas du graphe b de la figure 5.19 pour une proportion d'exemples classés inférieur à 0,01 environ.

Rappelons que ces résultats sont obtenus après diminution du nombre de variables de 50 jusqu'à 25 (en moyenne) avec un algorithme de sélection de variables "rapide" (voir paragraphe 5.3.1) contrainte de l'optimisation du temps de recherche en ligne.

5.5 Conclusions

Nous avons introduit une nouvelle méthode EMiner de recherche interactive d'images par boucle de pertinence basée sur les modèles de mélange et sur l'algorithme EM. Nous avons montré la flexibilité de cette méthode pour trouver des catégories d'images de structure complexe même multimodales en effectuant des expérimentations sur la base ANN généraliste à l'aide d'un protocole d'évaluation strict.

Une comparaison de notre approche EMiner à une autre basée sur les classificateurs SVM a été réalisée. Nous avons abouti à la conclusion que EMiner permet de trouver des classes complexes multimodales par contre SVM semble plus puissant pour discriminer d'une manière précise une classe uni-modale.

Nous avons proposé un nouveau schéma de sélection de variables conçu d'une part pour la classification *semi-supervisée* (contribution au niveau du choix du critère de sélection) et d'autre part *rapide* (au niveau de la procédure du choix du sous ensemble de variables pertinentes) pour la recherche d'images par boucle de pertinence. Les performances de cette méthode sont illustrées à l'aide de données synthétiques et de données réelles (index de la base ANN). Nous avons pu éliminer les variables redondantes en conservant ou en améliorant les performances.

Chapitre 6

Application aux pathologies ostéo-articulaires

Dans ce chapitre nous décrivons notre application médicale de recherche par le contenu. Elle est dédiée aux images IRM ostéo-articulaires des mains.

6.1 Introduction

Le diagnostic radiologique des pathologies ostéo-articulaires repose sur l'analyse sémiologique des images (Cotten, 2000), c'est à dire sur l'analyse de différents signes dont la présence, l'absence ou l'association sont fondamentales pour reconnaître l'affection. La détection informatisée de ces signes pourrait donc permettre une indexation des images les contenant, et par conséquent, dans un deuxième temps, une indexation des affections ostéo-articulaires.

Nous avons choisi comme premier sujet d'étude la polyarthrite rhumatoïde, rhumatisme inflammatoire susceptible d'entraîner des destructions articulaires très importantes mais dont le pronostic précoce est rendu fondamental par le développement de nouveaux traitements susceptibles de stopper l'évolution de la maladie.

Le diagnostic repose sur l'étude IRM des mains (articulations métacarpo-phalangiennes notamment) (Sugimoto et al., 2000; Boutry et al., 2001, 2003) et sur la présence ou non de deux symptômes : une synovite (membrane articulaire inflammatoire et épaisse) et des érosions osseuses. La détection informatisée permet alors d'individualiser les patients présentant les premiers symptômes de cette affection. Il en résulte une détection et une indexation objective de la pathologie.

Le but de ce travail de recherche est bien sûr d'appliquer par la suite ce type d'analyse informatisée de l'image à d'autres indices visuels et par conséquent à d'autres pathologies avec pour objectifs :

- Une meilleure gestion et utilisation des banques d'images à partir de l'indexation

des symptômes. Celle-ci permettrait ainsi aux étudiants en Médecine (formation initiale), mais également aux praticiens (formation continue) analysant un dossier radiologique d'avoir accès à d'autres dossiers radiologiques présentant le même type de signes radiologiques.

- Un gain de temps dans l'accès à l'information. En effet, le développement de l'imagerie, notamment par exemple du scanner spiralé, est à l'origine de la production de plusieurs centaines d'images pour un même patient. L'indexation à partir des symptômes permet ainsi une orientation directe et rapide vers la ou les images les plus représentatives.

6.2 Description de l'application

6.2.1 Contenu de la base

Notre base d'images est formée par des images IRM des mains fournies par le service de Radiologie Ostéo-Articulaire à l'hôpital Roger Salengro du Centre Hospitalier Régional Universitaire (CHRU) de Lille.

En effet, il s'agit des images des coupes axiales des squelettes des deux mains (voir figure 6.1). Durant l'acquisition des images IRM d'un patient, une série de ces coupes est enregistrée, contenant des images des deux carpes, métacarpes et des phalanges. Notre base est constituée des coupes pertinentes de différentes séries (*cf.* figure 6.3). Ce sont des coupes au niveau des articulations choisies par le médecin pour faire le diagnostic et qui portent d'habitude les symptômes de la maladie (voir figure 6.2).

6.2.2 Contenu des images : indices visuels

Une phase essentielle de l'indexation des images de la base est la compréhension des indices visuels ou encore les éléments sémiologiques d'une pathologie qui sont d'après Cotten (2000) les "clés du diagnostic d'une pathologie". Ainsi la signature (l'index) extraite devra caractériser l'image en codant ses informations pertinentes et devra combler en partie la lacune sémantique en combinant l'interprétation humaine de haut niveau et le calcul des attributs des images de bas niveau.

Comme nous l'avons déjà signalé, les deux symptômes sont :

1. la synovite qui est la membrane articulaire inflammatoire et épaisse comme montré dans la figure 6.4,
2. l'érosion osseuse où l'inflammation commence à éroder l'os. Ceci crée une déformation de la forme du contour de la coupe axiale de l'os (voir figure 6.4).

Ainsi, la détection et la caractérisation de régions inflammatoires qui se présentent comme des zones claires autour des os et la détection de contours des ces derniers (érodés ou non) devront permettre d'élaborer une signature. D'où le choix d'une segmentation automatique en régions étiquetées chacune par un vecteur d'attributs.

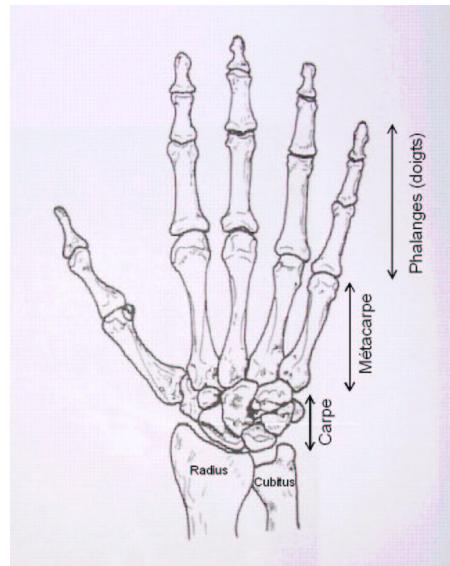


FIG. 6.1 – *Le squelette de la main : il est formé par le carpe, le métacarpe et les phalanges (doigts). Le carpe est l'ensemble des huit petits os du poignet, répartis en deux rangées, reliant l'avant-bras au métacarpe. Le grand os et l'os crochu font partie du carpe. Le métacarpe est la partie du squelette de la main située entre le carpe (poignet) et les doigts.*

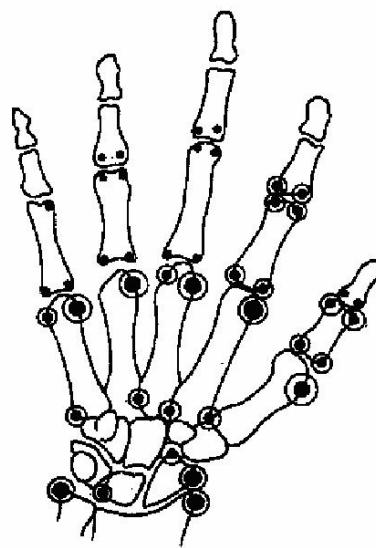


FIG. 6.2 – *Les érosions osseuses qui se présentent au niveau des articulations.*

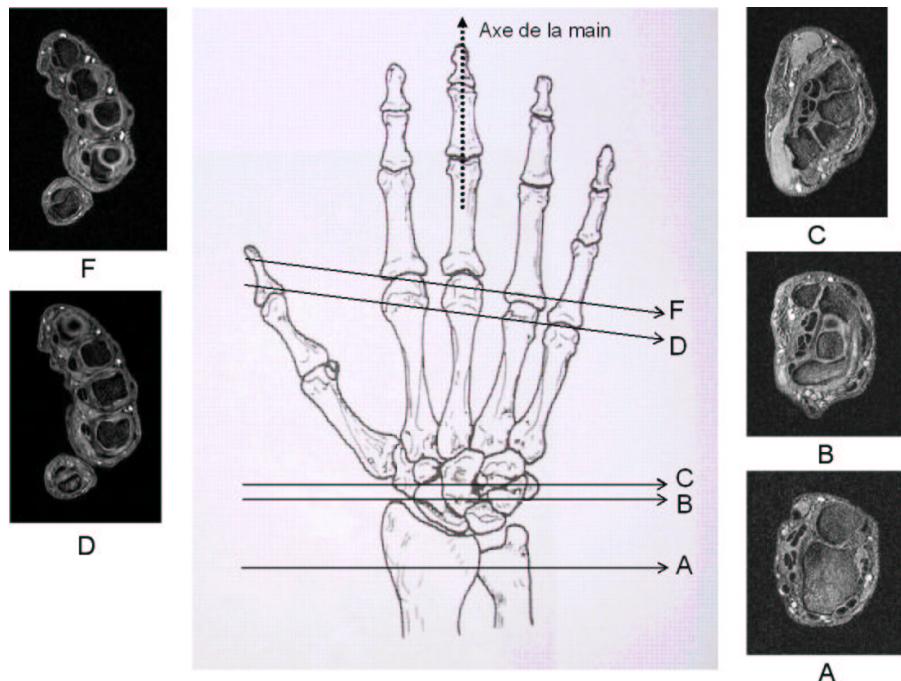


FIG. 6.3 – Les images IRM des coupes axiales du squelette de la main existantes dans la base. Les coupes D et F sont au niveau des articulations métacarpo-phalangiennes : têtes métacarpiques et bases phalangiennes.

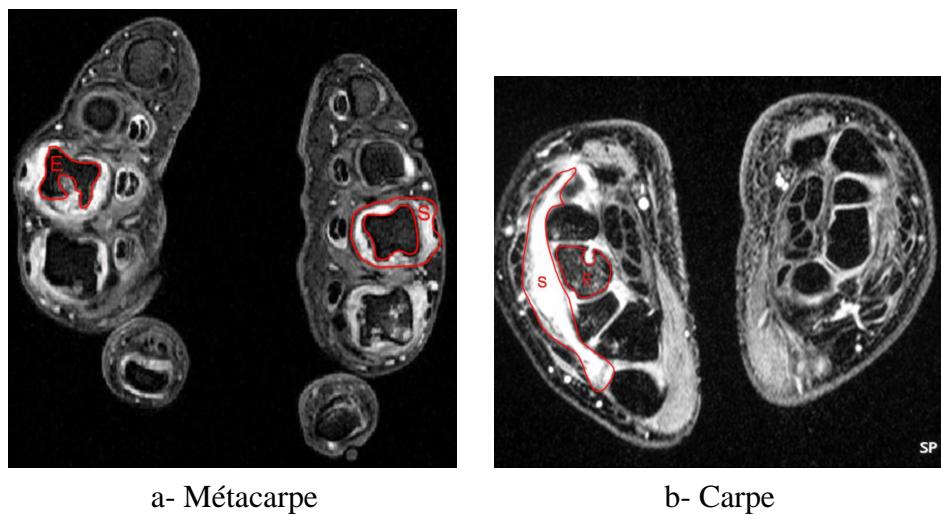


FIG. 6.4 – Les éléments sémiologiques de l'affection des mains: une synovite (S) qui est une membrane articulaire inflammatoire et épaisse et l'érosion osseuse (E).

Nous pouvons distinguer huit catégories ou classes d'images en séparant d'abord entre carpe et métacarpe (caractéristique de forme), et ensuite entre les symptômes (présence ou absence de synovite et érosion) :

1. métacarpe normal,
2. métacarpe, synovite,
3. métacarpe, érosion,
4. métacarpe, synovite et érosion,
5. carpe normal,
6. carpe, synovite,
7. carpe, érosion,
8. carpe, synovite et érosion.

Nous pouvons noter qu'il est très rare de trouver des images des mains présentant de l'érosion sans avoir de la synovite. Un échantillon des six différentes catégories est montré dans la figure 6.5.

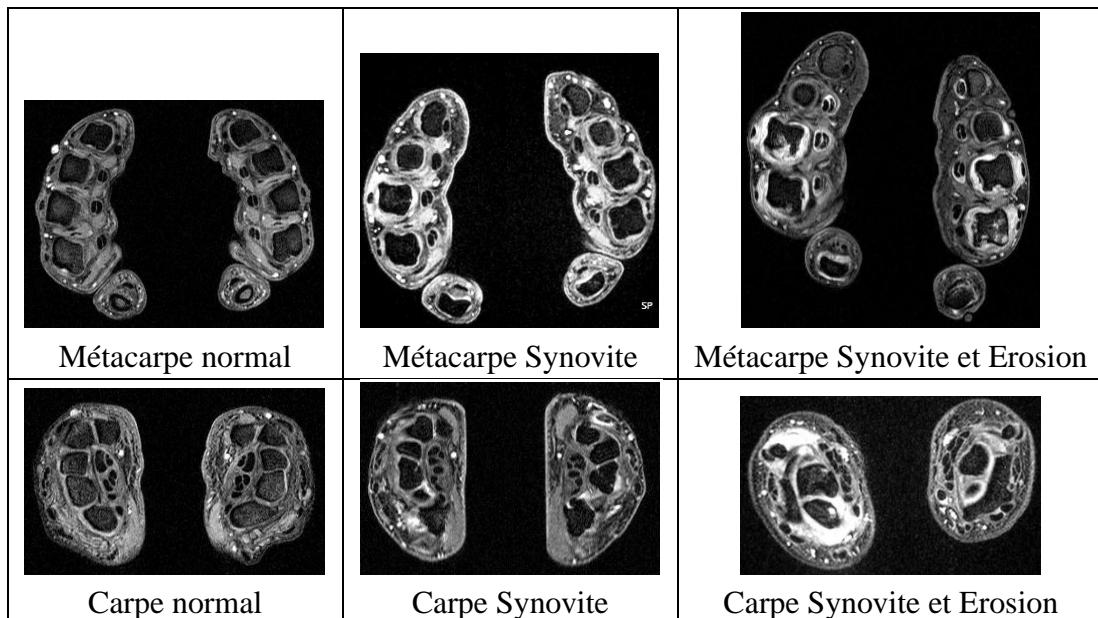


FIG. 6.5 – Un échantillon de six images exemples.

6.3 Présentation du système

Comme nous l'avons déjà indiqué dans le chapitre 2, le terme indexation recouvre un ensemble de techniques permettant de retrouver des images dans des banques d'images. Il existe aujourd'hui des techniques opérationnelles pour rechercher des informations à l'aide de mots-clés. L'identification de l'information est cependant plus difficile lorsqu'elle s'effectue à l'aide des seules données images. Le contenu d'une image, souvent très riche,

dissimule différentes informations non codées à l'avance. Ces informations qui se situent à différents niveaux d'abstraction doivent être extraites, dans le cas des images médicales, en accord avec l'analyse sémiologique effectuée par le praticien. Ces informations entreront dans la constitution d'un index qui sera utilisé pour classer les images et les retrouver efficacement.

Notre système (cf. figure 6.6) comprend classiquement une phase hors ligne d'indexation pour construire la base d'index. Le fonctionnement en ligne comprend le calcul d'index sur l'image requête, la phase de recherche et d'interaction avec l'utilisateur. La boucle de pertinence implique l'utilisateur dans une phase incrémentale d'introduction d'images annotées dans la base (cf. chapitre 3) et participe à la réduction du fossé sémantique (Smeulders et al., 2000).

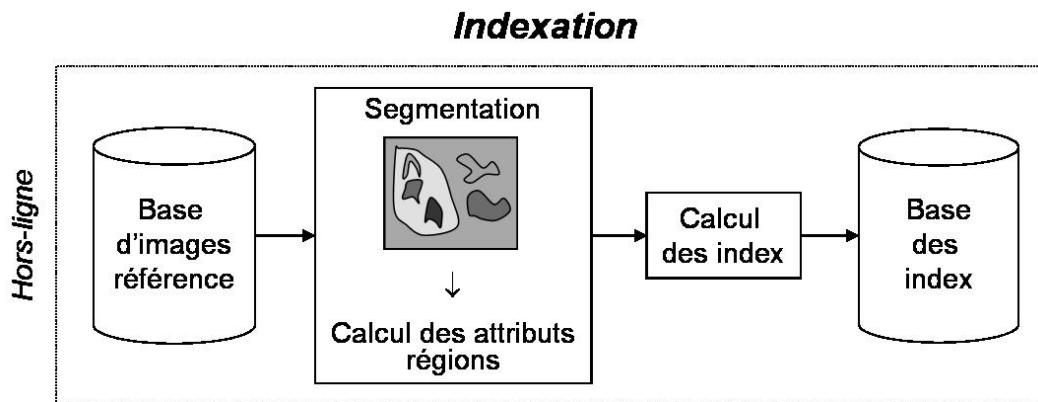


FIG. 6.6 – Schéma descriptif de la méthode d'indexation.

Comme nous l'avons déjà expliqué dans le paragraphe précédent, notre index doit être fortement lié aux régions pathologiques. Pour cela notre méthode d'indexation, comme illustré dans le schéma 6.6, doit comporter une étape de segmentation d'images. Signalons que cette segmentation doit être automatique vu le grand nombre d'images que la base doit contenir.

Ensuite et après segmentation, un ensemble de régions est extrait de chaque image. Le fait que le nombre des régions diffère d'une image à une autre, peut poser un problème de calcul d'index. Ces régions sont caractérisées par un ensemble d'attributs qui codent les informations de niveau de gris, de texture et de forme des régions obtenues.

Une phase de classification automatique (quantification) est proposée pour fabriquer des index de même dimension pour toutes les images de la base. Cette étape est délicate parce qu'il faut trouver un compromis entre le grand taux d'informations déjà extraites après segmentation et le taux de compression de ces informations pour avoir une méthode de recherche en ligne simple et rapide.

6.4 Segmentation

En regardant nos images, nous pouvons faire la distinction entre le fond et les mains qui comprennent les os et les régions claires (enflammées ou non). Le niveau de gris moyen du fond et des os est proche de zéro (noir) par contre celui des régions claires n'est pas loin de 255 (blanc). De cette façon, nous avons proposé les deux traitements consécutifs suivants :

- extraction des mains et des os du fond,
- extraction des zones claires.

Les contours des mains et des os ainsi que les frontières des régions claires sont extraits par une technique de contours déformables basée sur la théorie des ensembles de niveaux présentée ci-après. L'avantage de cette méthode est de fournir des contours fermés et qu'elle permet au contour en mouvement de se diviser en plusieurs formes ou de fusionner comme cela est illustré dans la figure C.3.

6.4.1 Prétraitement

Une phase de prétraitement des images qui doit être automatique et applicable sur toutes les images de la base est indispensable pour faciliter la tâche de la segmentation automatique et calculer des attributs comparables d'une image à une autre. Ainsi, quatre étapes consécutives ont été appliquées :

1. expansion de dynamique,
2. encadrement des mains dans un rectangle,
3. normalisation de la surface de l'image (ou redimensionnement),
4. spécification de l'histogramme.

La première phase consiste à utiliser au mieux l'échelle de niveaux de gris disponible sur le système d'acquisition et à rendre les images de la base dans la même dynamique de niveaux de gris.

Les étapes 2 et 3 ne sont que des étapes préparatoires pour la dernière étape : la "spécification d'histogramme" qui est indispensable pour faciliter la mise en oeuvre de la méthode de segmentation automatique proposée, c'est à dire le réglage des paramètres de cette dernière. Cette phase de prétraitement consiste à rendre la distribution d'intensité des images de la base voisine d'une distribution spécifiée à l'avance. Cette dernière est celle d'une image référence prise dans les conditions optimales d'acquisition.

A noter que les attributs extraits des images sont calculés sur les images ayant spécification d'histogramme.

Un schéma récapitulatif de la phase de prétraitement et de segmentation est présenté dans la figure 6.7.

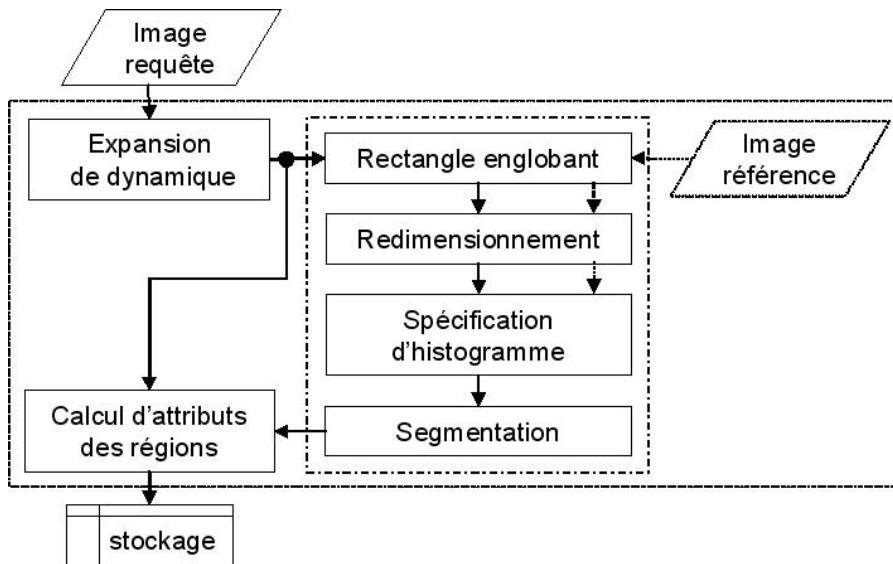


FIG. 6.7 – Un schéma récapitulatif de la phase de prétraitement et de segmentation.

Expansion de dynamique

Les images disponibles dans notre base sont de niveaux de gris codés sur 8 bits (la valeur de niveau de gris des pixels varie entre 0 et 255). En regardant l'histogramme de la plupart des images de la base (voir histogramme de l'image initiale de la figure 6.8)), nous observons que sa dynamique est resserrée entre $[a_0 > 0, a_1 = 255]$. Pour étendre les niveaux de gris entre $a_{min} = 0$ et $a_{max} = 255$, nous avons appliqué la transformation linéaire classique appelée "expansion de la dynamique" (Cocquerez et Philipp, 1995) :

Soient $A[i,j]$ l'image de départ et $A'[i,j]$ l'image après transformation. L'expansion de dynamique correspond à la transformation linéaire T suivante :

$$a' = \alpha + \beta a, \quad (6.1)$$

telle que :

$$\forall a \in [a_0, a_1], \quad a \xrightarrow{T} a' \in [a_{min}, a_{max}],$$

avec

$$\alpha = \frac{a_{min} \cdot a_1 - a_{max} \cdot a_0}{a_1 - a_0} \quad \text{et} \quad \beta = \frac{a_{max} - a_{min}}{a_1 - a_0}. \quad (6.2)$$

Cette transformation ne fait qu'améliorer l'aspect visuel de l'image (voir 6.9). Dans les conditions nominales d'acquisition, le fond d'une image est noir (valeur de niveau de gris est égale à 0) ce qui n'est pas le cas des images de la base qui ont un fond bruité. Cette transformation peut diminuer le bruit du fond.

Rectangle englobant les mains

Dans les conditions normales d'acquisition des images, le fond occupe 50% de la surface de l'image. L'extraction du rectangle englobant les mains est indispensable pour di-

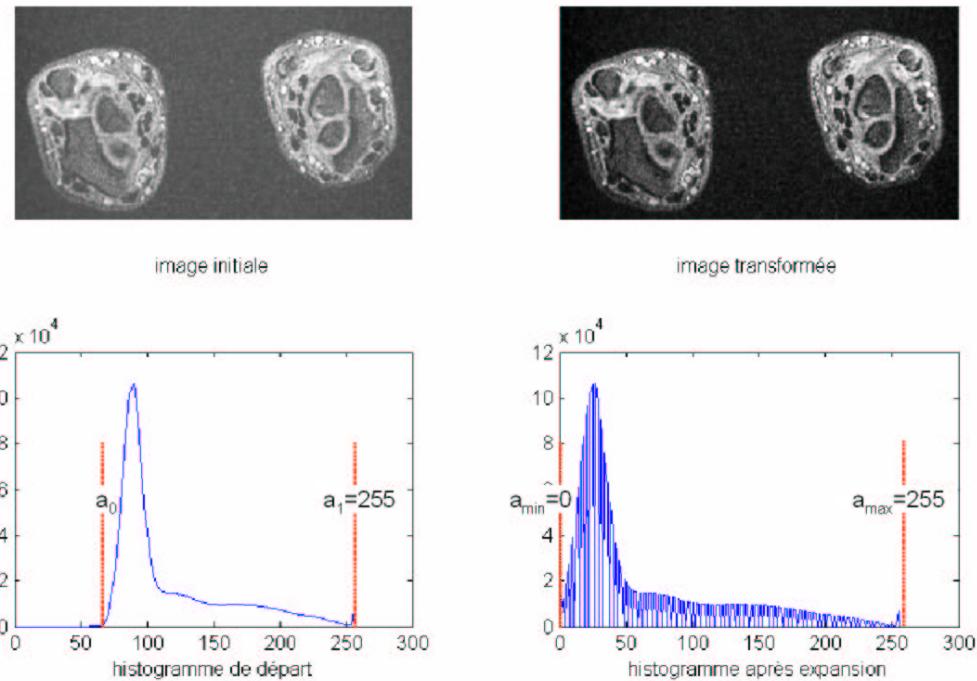


FIG. 6.8 – Expansion de dynamique des intensités.

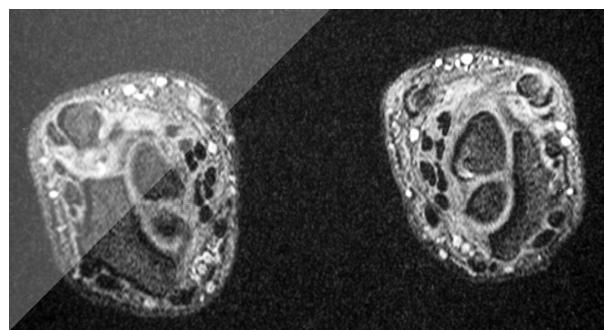


FIG. 6.9 – Expansion de dynamique: Image initiale / image transformée.

minuer la taille du fond et par conséquence réduire son influence pendant la spécification de l'histogramme d'une image par celui d'une image référence où le fond n'occupe pas la même surface.

Ainsi et après l'expansion de la dynamique, l'extraction du rectangle englobant les mains est réalisée par un simple seuillage après lissage de l'image.

normalisation de la surface des images

Les surfaces des images sont ramenées à une surface constante S_f . Si S_0 est la surface de l'image initiale, nous obtenons :

$$S_f = cte^2 \cdot S_0; \quad (6.3)$$

ce qui fait que la constante cte vaut :

$$cte = \sqrt{\frac{S_f}{S_0}}; \quad (6.4)$$

et la largeur et la hauteur de l'image finale sont celles de l'image initiale multipliées par la constante cte .

Un redimensionnement de l'image utilisant l'interpolation par le plus proche voisin (Parker et al., 1983) est appliqué connaissant les nouvelles dimensions. Cette méthode est choisie car elle est plus rapide que d'autres approches classiques existantes (interpolation bilinéaire, bicubique) .

Spécification de l'histogramme

Pour mettre en oeuvre la segmentation avec le même réglage de paramètre pour toutes les images de la base, ces dernières font l'objet d'une spécification d'histogramme (voir Cocquerez et Philipp, 1995, pg. 69). L'histogramme de référence est celui d'une image pathologique typique. Les figures 6.10 et 6.11 illustrent des résultats de cette transformation.

Nous rappelons que les attributs sont extraits des images avant spécification d'histogramme.

6.4.2 Détection du contour par l'utilisation de la théorie des ensembles de niveaux

La méthode est détaillée en annexe C. Nous indiquons ici la manière dont nous élaborons le contour initial et les paramètres de l'algorithme.

Initialisation des contours

L'approche "ensemble de niveau" est caractérisée, entre autres, par le fait que le contour déformable peut se diviser ou fusionner en cours d'évolution. L'initialisation peut donc

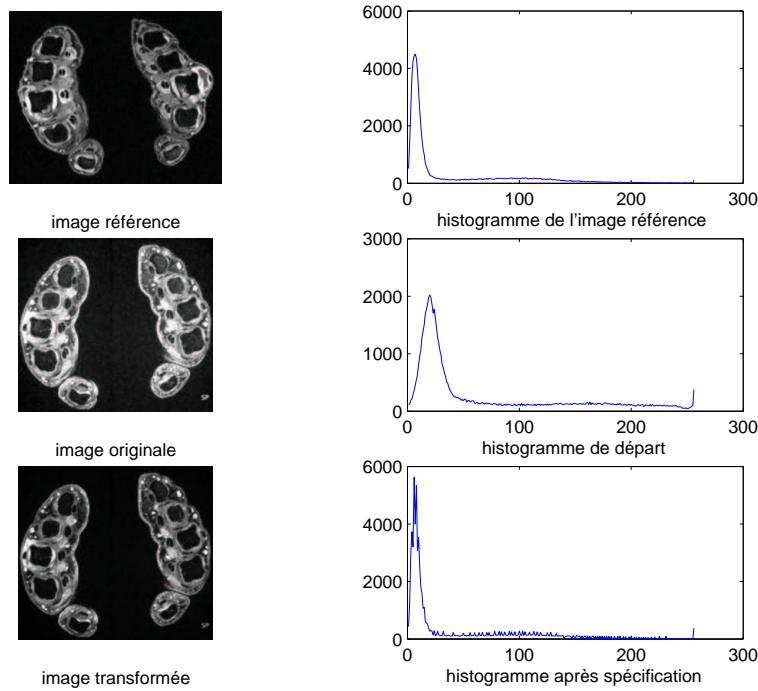


FIG. 6.10 – Spécification d'histogramme: exemple d'une image du métacarpe.

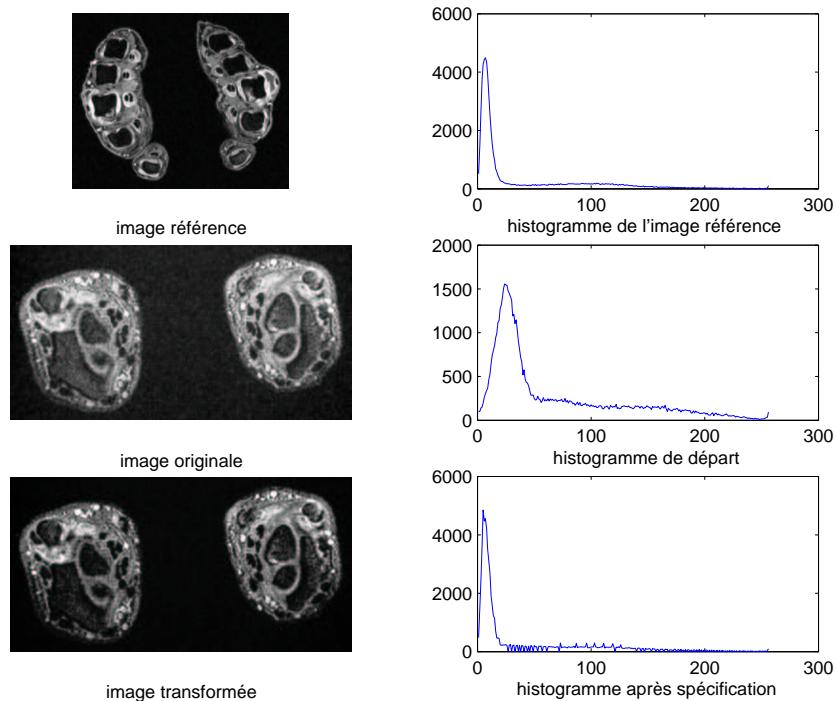


FIG. 6.11 – Spécification d'histogramme: exemple d'une image du carpe.

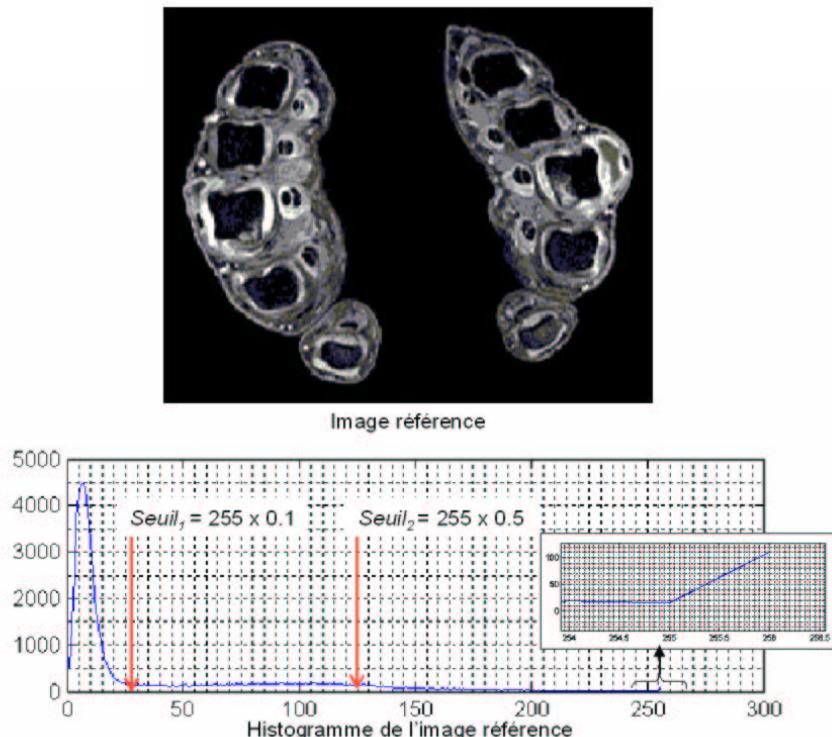


FIG. 6.12 – Histogramme de l'image référence. $Seuil_1 = 0.1 \times 255$ et $Seuil_2 = 0.9 \times 255$ sont les valeurs des seuils utilisés pour initialiser les contours déformables autour des mains et des régions claires respectivement.

se faire en bordure d'image. Cependant, comme cette méthode est coûteuse en temps de calcul, les contours initiaux sont choisis proches de la solution en prenant les frontières de régions "germes" fournies par les opérateurs élémentaires de morphologie mathématique (Serra, 1982) : dilatation et érosion sur l'image binaire obtenue par seuillage (*cf.* figure 6.12).

Résultats de la segmentation

Grâce à la spécification d'histogramme des images de la base par rapport à l'image référence, nous avons pu régler les paramètres F_A et F_G de la vitesse de propagation des contours (voir l'équation C.31) d'une façon qui convient à l'image référence :

$$F_{\text{mains}}(\kappa) = k_i * (500 + 100\kappa), \quad (6.5)$$

$$F_{\text{regions claires}}(\kappa) = k_i * (2000 + 1500\kappa). \quad (6.6)$$

Le principe du sens de propagation des contours est illustré dans la figure 6.13.

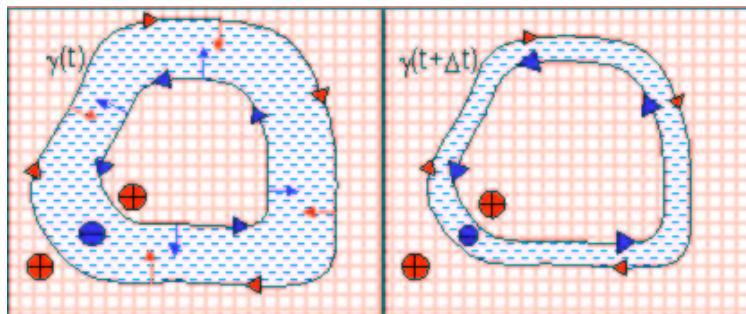


FIG. 6.13 – Illustration d'un exemple de propagation des ensembles de niveaux.

La figure 6.14 montre les résultats de détection des contours et des résultats de segmentation (image des étiquettes) d'une image du carpe présentant des symptômes d'érosion. L'étiquetage réalisé est en 4-connexité (voir Cocquerez et Philipp, 1995, pg.61).

Avec la théorie des ensembles de niveaux, nous avons pu détecter les contours des érosions qui présentent en général des angles aigus comme montré dans la figure 6.15. C'est une autre avantage de cette théorie par rapport à d'autres techniques de contours actifs (voir l'annexe C). Un autre résultat de segmentation d'une image du métacarpe pathologique est montré dans la figure 6.16.

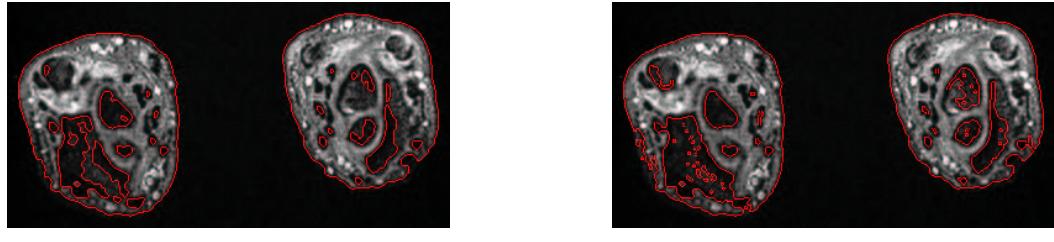
Les résultats de segmentation de six images exemples choisies parmi les six catégories pathologiques de la base sont exposés dans la figure 6.17.

6.5 Extraction des caractéristiques

Généralement les attributs extraits des images, sont classés en trois catégories d'attributs :

1. niveau de gris,

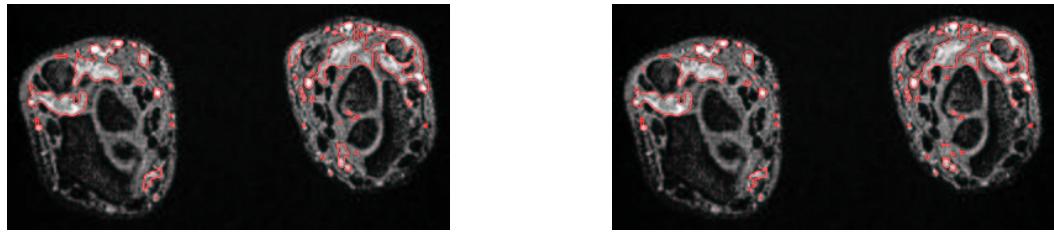
Détection des mains et des articulations



Contour initial

Contour final

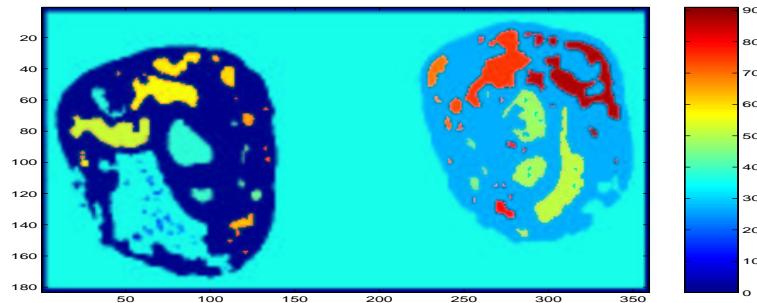
Détection des régions claires



Contour initial

Contour final

Propagation des contours



Résultats segmentation : image des étiquettes des régions

FIG. 6.14 – Détection des contours par la théorie des ensembles de niveaux.

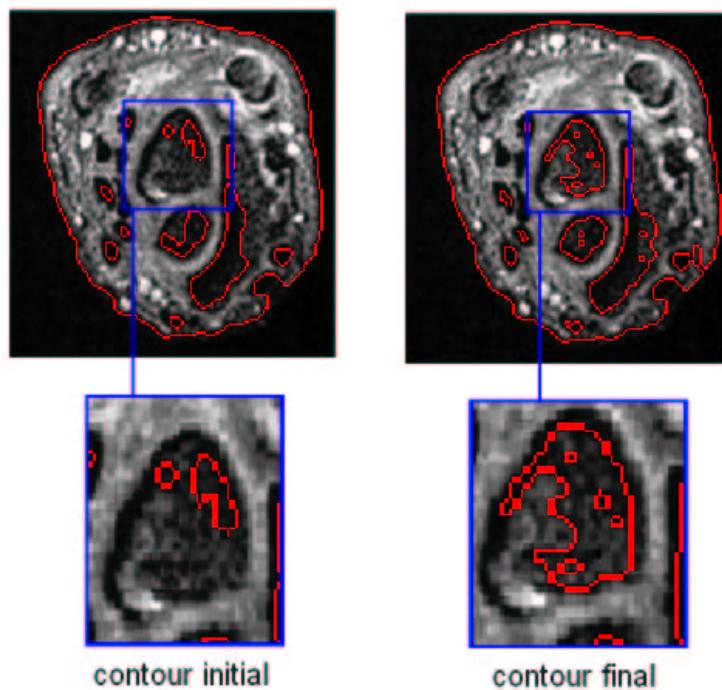


FIG. 6.15 – Détection de l'érosion.

2. forme,
3. texture.

Les indices visuels caractérisant les images de notre base sont :

- les zones pathologiques sont claires et peu texturées,
- les contours des érosions présentent une courbure plus élevée que les contours des régions saines,
- les os sont texturés.

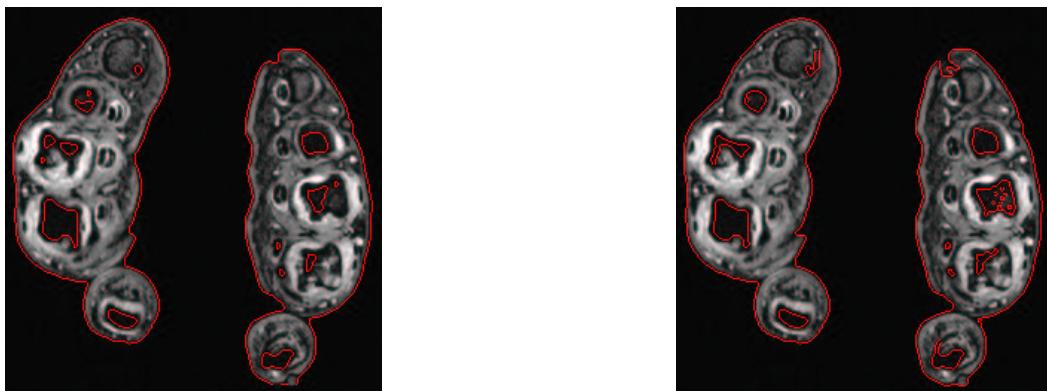
D'où le choix de nos attributs calculés pour chaque région (notée R) présentés ci-après.

Les attributs de niveau de gris calculés sur R sont la moyenne et l'écart type de niveau de gris des pixels appartenant à R . La forme de R a été caractérisée par sa surface, sa compacité et par les moments invariants de Hu. Les masques de Laws ont été employés pour décrire sa texture.

6.5.1 Attributs de forme: les moments invariants de Hu

Plusieurs attributs de forme sont rencontrés dans la littérature. Une comparaison des attributs de forme courants se trouve dans (Mehtre et al., 1997) où les moments invariants de Hu ont montré leur efficacité pour la recherche d'images basée sur les mesures de forme. Ces moments ont été employés avec succès dans d'autres domaines d'application par exemple en recalage d'images (Flusser et Suk, 1994; Dai et Khorram, 1999).

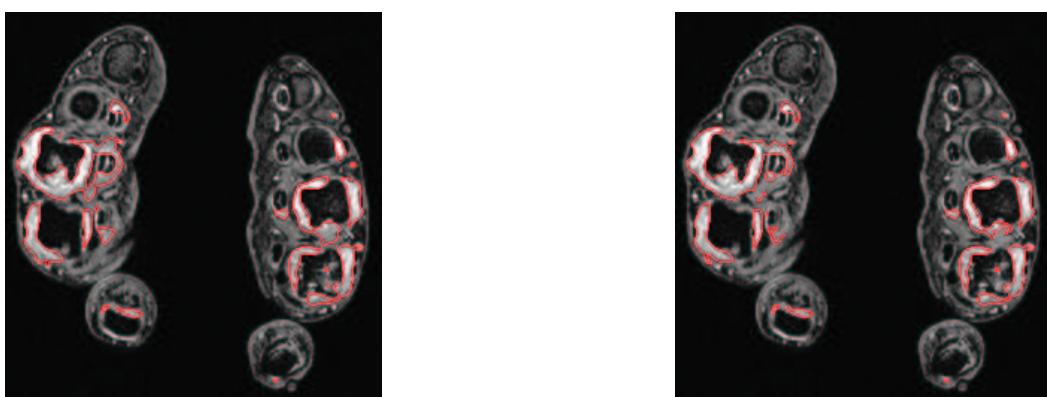
Détection des contours des mains et des articulations



Contour initial

Contour final

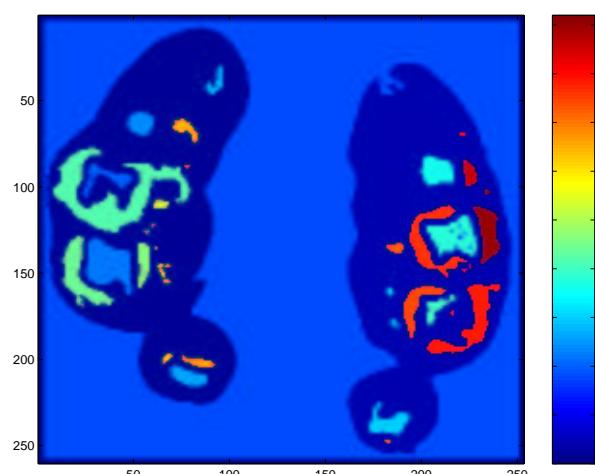
Détection des régions claires



Contour initial

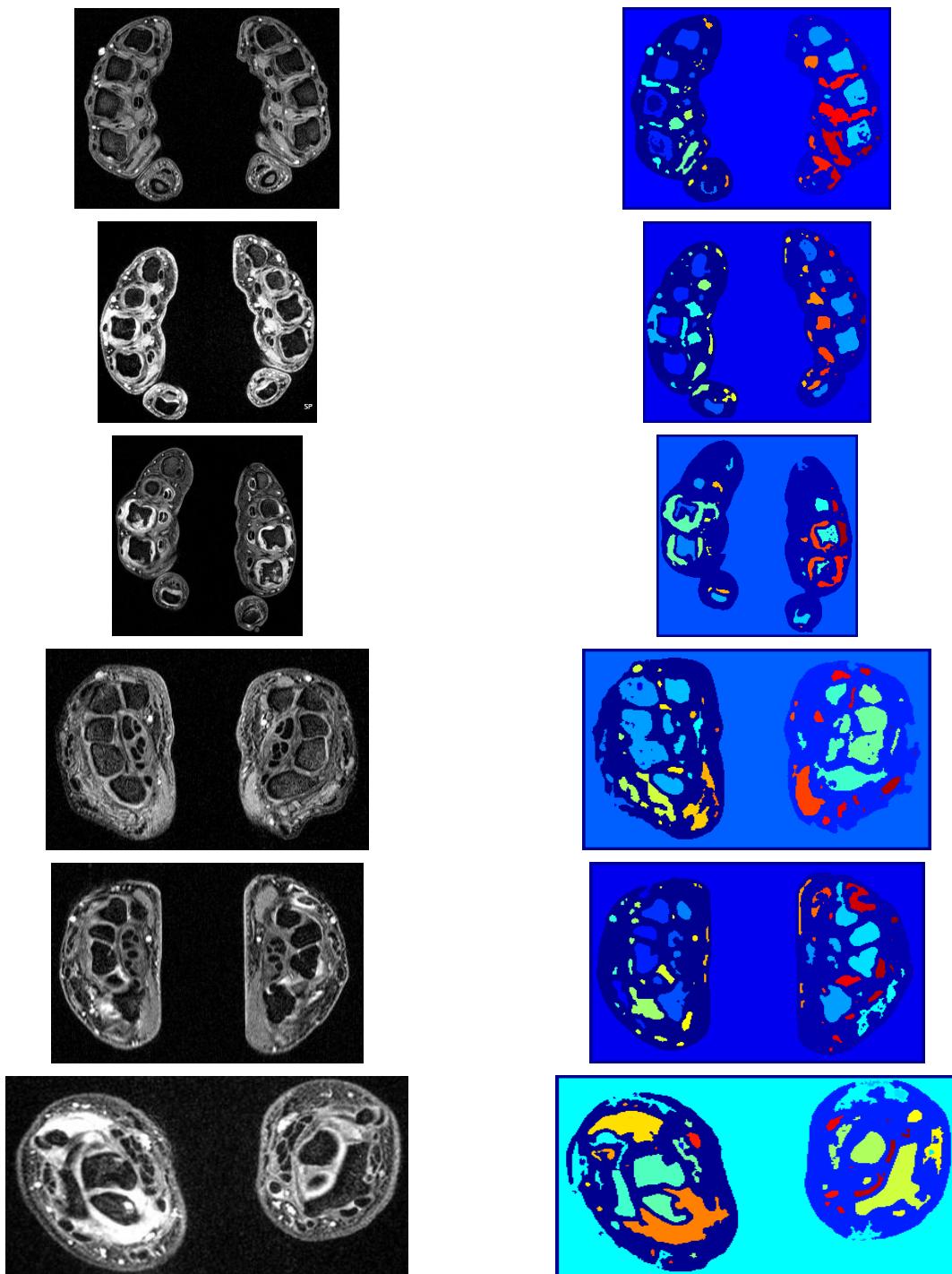
Contour final

Propagation des contours



Résultats segmentation : étiquetage des régions

FIG. 6.16 – Détection des contours par la théorie des ensembles de niveaux.



(a) Image à segmenter

(b) Image des étiquettes des régions après segmentation

FIG. 6.17 – Résultats de segmentation des six images exemples choisies de la base.

1. Surface normalisée (surface de la région divisée par la surface des mains);
2. Moyenne de niveau de gris;
3. Ecart type de niveau de gris;
4. Compacité;
5. Les 7 moments invariants de Hu : I_1, \dots, I_7 ;
6. Les neufs mesures d'énergie de texture de Laws : T_1, \dots, T_9 ;

TAB. 6.1 – *Liste des attributs par région*

Les moments invariants de Hu (Hu, 1962) sont des mesures de formes très populaires (Pratt, 1991). Ces moments sont invariants par rapport à la translation, rotation, et le changement d'échelle. Ils sont définis comme suit :

$$I_1 = \frac{1}{\mu_{00}^2}(\mu_{20} + \mu_{02}); \quad (6.7)$$

$$I_2 = \frac{1}{\mu_{00}^4} \{ (\mu_{20} - \mu_{02})^2 + 4(\mu_{11})^2 \}; \quad (6.8)$$

$$I_3 = \frac{1}{\mu_{00}^5} \{ (\mu_{30} - 3\mu_{12})^2 + (\mu_{03} - 3\mu_{21})^2 \}; \quad (6.9)$$

$$I_4 = \frac{1}{\mu_{00}^5} \{ (\mu_{30} + \mu_{12})^2 + (\mu_{03} + \mu_{21})^2 \}; \quad (6.10)$$

$$I_5 = \frac{1}{\mu_{00}^{10}} \{ (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12}).[(\mu_{30} + \mu_{12})^2 - 3(\mu_{03} + \mu_{21})^2] + (3\mu_{21} - \mu_{03})(\mu_{03} + \mu_{21}).[3(\mu_{30} + \mu_{12})^2 - (\mu_{03} + \mu_{21})^2] \}; \quad (6.11)$$

$$I_6 = \frac{1}{\mu_{00}^7} \{ (\mu_{20} - \mu_{02}).[(\mu_{30} + \mu_{12})^2 - (\mu_{03} + \mu_{21})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \}; \quad (6.12)$$

$$I_7 = \frac{1}{\mu_{00}^{10}} \{ (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12}).[(\mu_{30} + \mu_{12})^2 - 3(\mu_{03} + \mu_{21})^2] + (3\mu_{12} - \mu_{30})(\mu_{03} + \mu_{21}).[3(\mu_{30} + \mu_{12})^2 - (\mu_{03} + \mu_{21})^2] \}. \quad (6.13)$$

Les μ_{pq} sont les moments centraux où p et q définissent les ordres des moments. Dans le cas d'une image binaire $F(i,j)$ (voir l'image exemple de la figure 6.18) :

$$\mu_{pq} = \sum_i \sum_j (i - \bar{x})^p (j - \bar{y})^q F(i,j). \quad (6.14)$$

\bar{x} et \bar{y} sont les coordonnées du centre de gravité de l'objet :

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}. \quad (6.15)$$

$$(6.16)$$

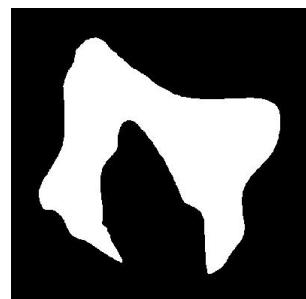
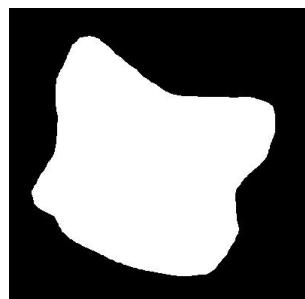


FIG. 6.18 – Exemple d'une image binaire contenant un objet dont il faut analyser la forme. Les pixels appartenant à l'objet ont une valeur de 1 et le reste est mis à zéro.

avec m_{pq} sont les moments :

$$m_{pq} = \sum_i \sum_j i^p j^q F(i,j). \quad (6.17)$$

Le tableau 6.2 nous donne un exemple de calcul des 7 moments de Hu dans les deux cas de doigt normal et anormal.



a- Forme d'un doigt normal b- Forme d'un doigt après une érosion

| | I_1 | I_2 | I_3 | I_4 | I_5 | I_6 | I_7 |
|---|--------|--------|-------------|-------------|--------------|--------------|--------------|
| a | 0.2840 | 0.0114 | 1.7403e-004 | 3.1464e-006 | -3.4731e-011 | -1.8656e-007 | 6.4922e-011 |
| b | 0.4572 | 0.0153 | 8.8934e-004 | 1.2595e-005 | 4.7900e-010 | -1.1979e-006 | -1.2440e-009 |

TAB. 6.2 – Les moments de Hu calculés sur deux formes différentes d'une coupe axiale du doigt.

6.5.2 Attributs de texture utilisant les masques de Laws

Laws (1979, 1980b) nous propose un ensemble de masques 3×3 (voir tableau 6.3) ou 5×5 trouvés de manière expérimentale en effectuant des tests sur un ensemble d'images de textures naturelles. L'idée consiste à représenter les textures par des mesures d'énergie réalisées à la sortie de ces masques. La convolution de l'image originale I par ces masques donne plusieurs images résultats.

Laws a utilisé les images de texture de l'album de Brodatz et autres images pour comparer ses mesures d'énergie à d'autres mesures de texture basées sur les matrices de co-occurrence et de corrélation. Il a obtenu des taux de succès de classification de pixels de 94%, de 72%, et de 65% respectivement.

L'intérêt de cette méthode est qu'elle est simple à mettre en oeuvre, et qu'elle est directement exploitable pour réaliser une segmentation en régions. Un certain nombre de chercheurs en plus de Laws lui-même (Laws, 1980a) ont trouvé ces mesures de texture meilleures et plus facilement implémentables que les approches alternatives.

$$\begin{array}{c}
 \frac{1}{36} \begin{bmatrix} +1 & +2 & +1 \\ +2 & +4 & +2 \\ +1 & +2 & +1 \end{bmatrix} \quad \frac{1}{12} \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} \quad \frac{1}{12} \begin{bmatrix} -1 & +2 & -1 \\ -2 & +4 & -2 \\ -1 & +2 & -1 \end{bmatrix} \\
 \frac{1}{12} \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \quad \frac{1}{4} \begin{bmatrix} +1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & +1 \end{bmatrix} \quad \frac{1}{4} \begin{bmatrix} -1 & +2 & -1 \\ 0 & 0 & 0 \\ +1 & -2 & +1 \end{bmatrix} \\
 \frac{1}{12} \begin{bmatrix} -1 & -2 & -1 \\ +2 & +4 & +2 \\ -1 & -2 & -1 \end{bmatrix} \quad \frac{1}{4} \begin{bmatrix} -1 & 0 & +1 \\ +2 & 0 & -2 \\ -1 & 0 & +1 \end{bmatrix} \quad \frac{1}{4} \begin{bmatrix} +1 & -2 & +1 \\ -2 & +4 & -2 \\ +1 & -2 & +1 \end{bmatrix}
 \end{array}$$

TAB. 6.3 – *Les Masques de Laws*
(voir Pratt, 1991, pg.588)

Notre méthode consiste à calculer l'énergie des textures des régions obtenues après segmentation. Ainsi après convolution de l'image I par les masques de Laws, 9 images $M_{m=1,\dots,9}$ de textures sont obtenues comme illustré dans la figure 6.19. Ensuite, les neuf mesures de l'énergie de texture des régions sont calculées suivant l'équation 6.18 suivante :

$$T_{m,R_r=1,\dots,n} = \frac{1}{S_{R_r}} \sum_{(i,j) \in R_r} M_m^2(i,j); \quad (6.18)$$

où R_r est une région r de l'image contenant au total n régions. S_{R_r} est la surface (nombre de pixels) de R_r .

6.6 Indexation

Le problème est de trouver des images qui se ressemblent. Cela se fait en général par mesure de similarité entre les index des images fabriqués à partir des attributs calculés sur les images (voir chapitre 2).

En effet, puisque les indices visuels de nos images sont liés fortement aux régions et à leur topologie dans l'image, (ex. une région enflammée entoure le doigt), nous avons absolument besoin d'un index codant les caractéristiques de l'image entière calculée à

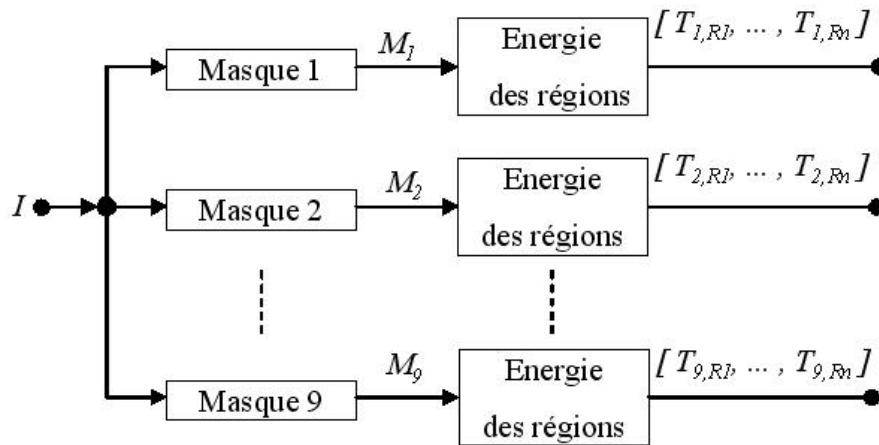


FIG. 6.19 – Méthode d'extraction des attributs de texture des régions utilisant les masques de Laws.

partir de celles des régions obtenues après segmentation (voir le Tableau 6.1). En outre, nous voulons trouver un index de même longueur pour toutes les images pour :

- trouver une méthode flexible de mesure de similarité et de recherche par boucle de pertinence,
- permettre une économie sur le temps de recherche.

D'où le choix de notre méthode présentée ci-après.

6.6.1 Calcul de l'index

Une classification non-supervisée des régions par la carte auto-organisatrice de Kohonen (Kohonen, 2001) est réalisée sur un nombre d'images tirées au hasard dans la base. Ainsi l'index de l'image i noté $x_i = \{x_{ij}; 1 \leq j \leq p\}$ est défini comme suit :

$$x_{ij} = \frac{1}{S_{\text{objet}}} \sum_{R_r \in C_j} S_{R_r}, \quad (6.19)$$

avec S_{objet} est calculé suivant l'équation suivante :

$$S_{\text{objet}} = S_{\text{image}} - S_{\text{fond}},$$

et C_1, C_2, \dots, C_p sont les classes des régions qui sont au nombre de p .

A noter que x_{ij} représente la proportion de pixels de l'image appartenant à la classe j , l'un des x_{ij} peut par exemple représenter la proportion de pixels appartenant à des zones enflammées dans les deux coupes présentes dans l'image. Dans la figure 6.21 nous présentons les résultats de la classification non supervisée des régions à travers l'image "dscn2991" (voir figure 6.5 b) où nous avons colorié les régions appartenant à la même classe avec la même couleur. Ce résultat est obtenu pour 30 classes de régions. Prenons par exemple la

classe numéro 26, elle correspond bien à l'index et au majeur des deux mains droites et gauches qui ont normalement la même forme d'os. La classe numéro 23 correspond aux régions claires enflammées de l'images. La figure 6.20 illustre d'autres résultats.

Signalons que cette façon de calculer l'index est généralisable à toutes les images où il est possible de distinguer un objet du fond.

6.6.2 Classification non supervisée des régions

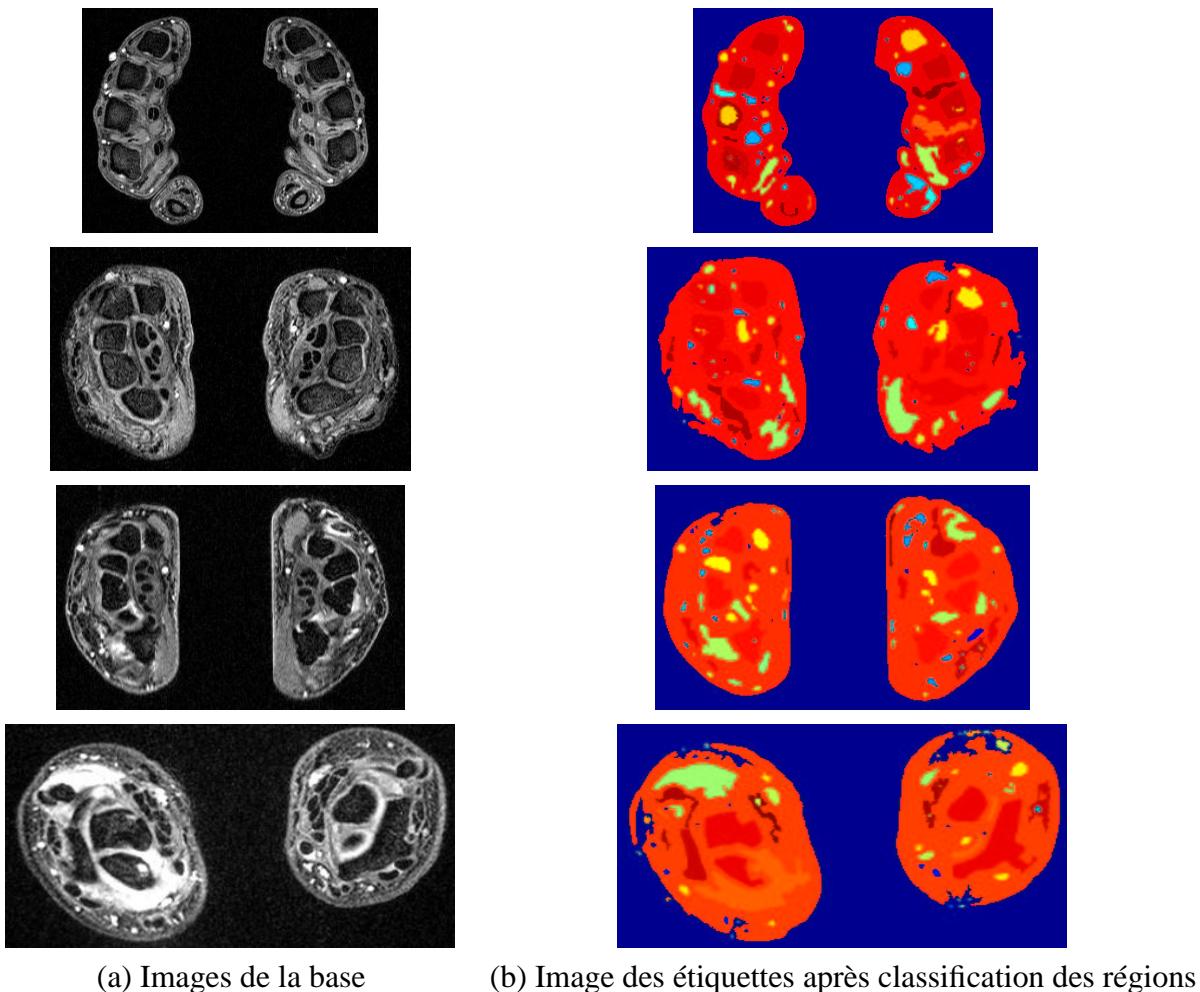


FIG. 6.20 – Résultats de classification des régions : les régions qui appartiennent à la même classe ont la même couleur.

Trois méthodes de classification non supervisée des régions à partir de leurs attributs ont été essayées : classification par les modèles de mélange et l'algorithme EM, Fuzzy c -means et la carte auto-organisatrice de Kohonen.

La carte de Kohonen a pour objectifs d'obtenir une représentation résumée des données (classification), et de faire en sorte que deux classes voisines dans l'"espace des neurones"

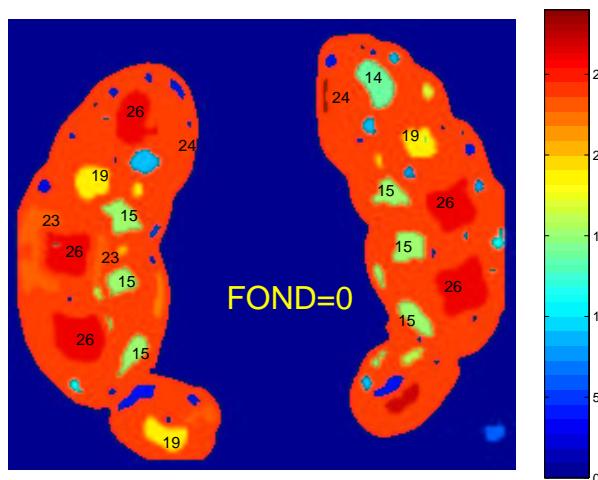


FIG. 6.21 – Résultats de classification des régions dans le cas de 30 classes : Les régions qui appartiennent à la même classe ont la même couleur.

le soient aussi dans l'espace des paramètres (préservation de la topologie). De plus, elle a prouvé sa capacité à trouver un partitionnement de l'espace des régions avec un nombre de classes non-vides élevé et par conséquent à bien représenter l'ensemble des régions.

En effet, vue la variété de l'ensemble des régions (caractérisées par les 20 attributs du tableau 6.1) de la base d'images, nous avions besoin d'un grand nombre de classes pour pouvoir les ordonner. Avec les modèles de mélange et l'algorithme EM, nous avons pu utiliser un nombre maximum de 15 classes. Par contre l'approche *Fuzzy c-means* nous a permis d'augmenter ce nombre jusqu'à 20. Avec la carte auto-organisatrice de Kohonen, 30 classes ont été facilement atteintes. Pour plus de détails sur ces trois algorithmes classiques utilisés, le lecteur intéressé peut consulter avec profit (Dang, 1998, chapitre 2).

Dans (Ambroise, 1996, chapitre 2) une étude comparative est réalisée entre les cartes de Kohonen et les modèles de mélange et l'algorithme EM : il a abouti à la conclusion que la carte de Kohonen est plus robuste c'est-à-dire stable, indépendante de l'initialisation et permettant de ne pas laisser de classes vides.

6.7 Expérimentations et validation réelle

Cette section vise à donner quelques éléments de validation de notre système de recherche d'images ostéo-articulaires.

Nous avons vu dans le chapitre 5 que l'évaluation d'un système de recherche d'images par le contenu demande une base d'images avec une vérité-terrain.

Notre base d'images est traitée comme un ensemble de huit classes (voir la section 6.2.2). La construction d'une telle base médicale étiquetée est un sacré problème qui demande beaucoup de temps et de travail pour archiver et étiqueter les dossiers patients au fur et à mesure de leur disponibilité. Elle demande ainsi la participation de personnels spécialisés : au moins un médecin spécialiste pour étiqueter ou diagnostiquer les images archivées

plus un technicien en informatique pour gérer les sauvegardes.

La base d'images étiquetées obtenue¹ dans ces conditions pour accomplir notre travail dans les délais prévus, est illustrée dans le tableau 6.4. Elle est de taille $N = 395$. La classe numéro 3 (métacarpe, érosion) est presque vide, ce qui s'explique par le fait qu'il est très rare d'avoir de l'érosion sans synovite. De même pour la classe numéro 7 (carpe, érosion).

| Numéro | Nom | Effectif |
|--------|--------------------------------|----------|
| 1 | Métacarpe, normal | 27 |
| 2 | Métacarpe, synovite | 90 |
| 3 | Métacarpe, érosion | 9 |
| 4 | Métacarpe, synovite et érosion | 63 |
| 5 | Carpe, normal | 34 |
| 6 | Carpe, synovite | 61 |
| 7 | Carpe, érosion | 23 |
| 8 | Carpe, synovite et érosion | 88 |
| | total | 395 |

TAB. 6.4 – *Contenu de la base expérimentale.*

Dans ce qui suit, nous proposons un protocole d'évaluation strict, qui tient compte de la faible taille de la base. Il suppose que l'image requête n'est pas contenue dans la base : c'est le principe du "leave one out" utilisé pour estimer l'erreur d'un classifieur.

6.7.1 Protocole d'évaluation

La technique de "validation croisée" consiste à diviser l'ensemble de données disponibles en V sous parties, et à définir le classifieur à l'aide de toutes les parties sauf une qui servira à l'estimation d'erreur. La procédure est répétée en utilisant successivement toutes les parties pour l'estimation, ce qui produit V différents taux d'erreur estimés. La version extrême de cette procédure, alors baptisée "leave one out", consiste à partager un ensemble de taille N en N sous parties (c'est à dire chaque exemple de l'ensemble de données forme une partie).

Notre protocole d'évaluation du système proposé peut aussi être appelé "leave one out" parce qu'il consiste à calculer le rappel et la précision (à la place de l'erreur) après recherche d'une image requête retirée de la base d'images utilisées ainsi comme ensemble d'apprentissage pour la classification non supervisée des régions et le calcul des index de la base. Le tableau 6.5 illustre les étapes du protocole d'évaluation. C'est le protocole utilisé dans le chapitre 5 sur lequel nous avons ajouté l'étape 2.

1. Nous remercions Pr. Anne Cotten, Dr. Nathalie Boutry et M. Patrick Potteau pour leur coopération.

Soient :

- N la taille de la base (le nombre total des images);
- b le nombre de boucles de pertinence effectuées;
- m le nombre des images annotées à chaque boucle de pertinence;
- lot l'ensemble des images parmi lesquelles sont tirées les meilleures, $N - b \times m$ est le nombre des images du lot.

pour i de 1 à N , **faire** :

1. image requête $\leftarrow i$ ème image de la base;
2. calculer les index de la base en se servant des images de la base sauf l'image requête pour l'apprentissage (classification non supervisée des régions);
3. pour m donné (égal par exemple à 5), **faire** :
 - o classifier par **EMiner** les images de la base ayant seulement l'image requête comme image annotée;
 - o **pour** b de 1 à b_{max} , **faire** :
 - ø sélection des m meilleures images parmi le *lot*;
 - ø annotation des m images non précédemment annotées;
 - ø classification par **EMiner** des images de la base en utilisant seulement les $b \times m$ images annotées;
 - ø calcul des critères précision/rappel pour N_r images retrouvées allant de 1 à N ;
 - o **fin** de la boucle de b .

fin de la boucle de i .

La précision et le rappel calculés sont **moyennés** sur toutes les images requêtes de la catégorie.

TAB. 6.5 – Protocole d'évaluation de notre système d'indexation et de recherche dans la base ostéo-articulaire basé sur le principe de "leave one out".

6.7.2 Résultats de recherche

Nous avons développé une interface utilisateur² pour la présentation de l'image requête et des images résultats (cf. figure 6.22). En dessous de chaque image retrouvée, il y a un bouton que l'utilisateur active quand l'image est pertinente. Quand l'utilisateur a annoté les images retrouvées, il active le bouton "AFFINER". Un bouclage de pertinence est alors exécuté (cf. algorithme EMiner du chapitre 5).

Des résultats d'images retrouvées ($N_r = 6$) pour une requête de chaque type sont présentés (figures 6.22 jusqu'à 6.29). Sur ces exemples, nous observons que les images retrouvées au bout de 6 itérations au maximum présentent des symptômes similaires à l'image requête :

1. La figure 6.22 résulte de la recherche d'une image de métacarpe normal. Les 6 images retrouvées sont pertinentes ($N_{rp} = N_r = 6$, ou précision = 1).
2. La figure 6.23 présente les 6 images similaires de la classe "métacarpe, synovite", résultats de recherche d'une image de métacarpe qui présente une synovite.
3. La figure 6.24 concerne une image de métacarpe qui présente une érosion. Deux images retrouvées sont similaires à l'image requête. Ceci est probablement dû au fait que si nous regardons les 9 images de la classe "métacarpe, érosion", (cf. figure 6.30), nous trouvons que seulement trois images a, f et g se ressemblent. Les autres images qui restent présentent les mêmes symptômes de pathologie mais pas la même coupe : par exemple, l'image e est une coupe au niveau des phalanges (changement de la forme de l'os des doigts). L'image b est tirée d'une séquence IRM sous des conditions d'acquisition d'image différentes de celles du reste (changement de texture).
4. La figure 6.25 affiche les 6 images similaires à une image de métacarpe qui présente de la synovite et de l'érosion. Les érosions détectées sont entourées par des cercles rouges.
5. La figure 6.26 est le résultat de la recherche d'une image de carpe normal.
6. La figure 6.27 montre les résultats de recherche de deux images de carpates qui présentent de la synovite : l'une est une coupe au niveau des deux os radius et cubitus (voir coupe A de la figure 6.3) et une autre au niveau du carpe (coupe B de la figure 6.3).
7. La figure 6.28 est le résultat de la recherche d'une image de carpe qui présente d'érosion. Deux images sont pertinentes (issues de la classe "carpe, érosion") sur les 6 retrouvées. Cette classe (numéro 7) présente le même problème que la classe 3 de métacarpe érosion. Cela s'explique par le fait que les symptômes d'érosion se manifestent rarement avant la synovite.
8. La figure 6.29 illustre les images similaires trouvées par le système de deux images requêtes de carpe présentant de la synovite et de l'érosion.

2. Les programmes ont été écrits en langage C et en MATLAB pour la thèse. Une version écrite en langage C et Visual Basic est transférée sur le système de fouille d'images développé par la société Archimed.

Les résultats sont obtenus en supposant que la recherche par bouclage de pertinence est réalisée par un spécialiste qui annote les images selon leur vérité-terrain. Notre système interactif permet au praticien de manifester son doute en pondérant ses avis. La boucle de pertinence du système prend en compte cet indice de confiance dans le calcul des résultats.

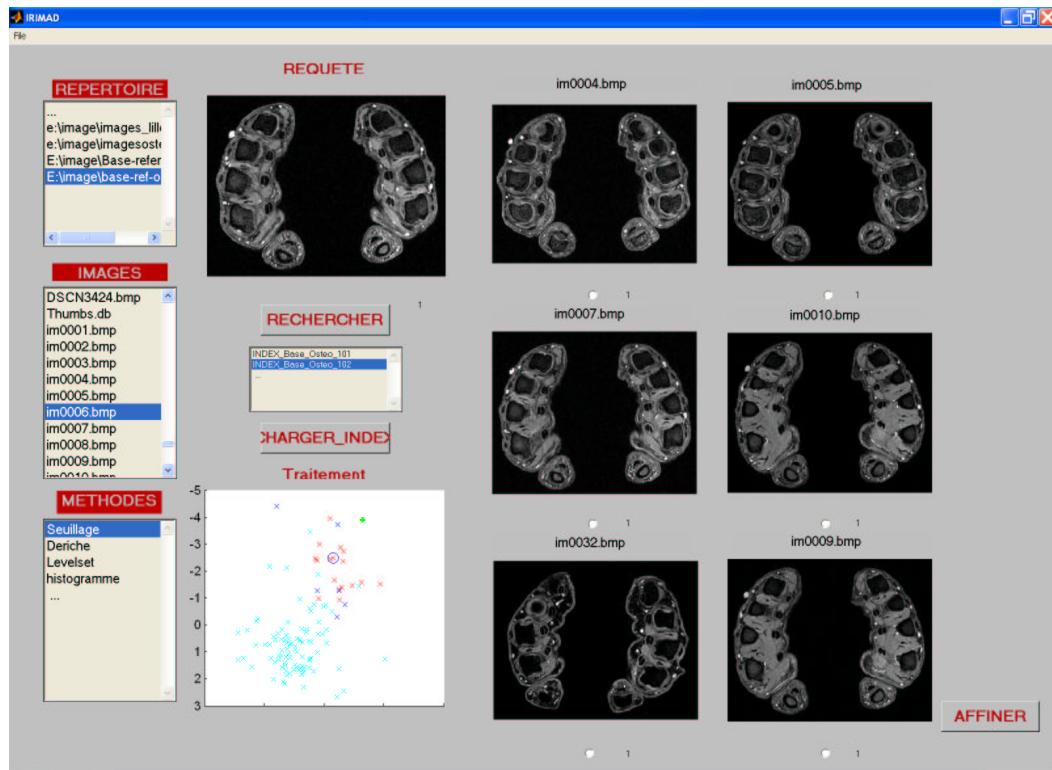


FIG. 6.22 – Résultats de recherche d'une image d'un métacarpe normal.

6.7.3 Résultats d'évaluation

La figure 6.31 montre les courbes de précision/rappel obtenues lors des recherches sur chacune des huit classes de la base pour 1, 2, 3, 4 et 5 bouclages de pertinence. Cinq images sont annotées ($m = 5$) à chaque itération. Dans la figure 6.32 nous illustrons les résultats pour $m = 10$.

Nous avons déjà présenté dans le chapitre 5 une méthode permettant de décoder facilement les courbes de précision/rappel (PR), c'est à dire que trouver le nombre d'images pertinentes retrouvées N_{rp} pour un nombre d'images retrouvées N_r donné. Prenons l'exemple de $N_r = 6, 10, 40$ et 80 : le tableau 6.6 donne le nombre des images pertinentes retrouvées N_{rp} après 5 itérations pour les huit classes de la base. Cinq images sont annotées ($m = 5$) à chaque itération. De même pour le tableau 6.7 mais pour $m = 10$.

Les 6 images retrouvées ($N_r = 6$: deuxième ligne du tableau 6.6) sont toutes pertinentes ($N_{rp} = 6$) pour les classes numéro 2, 4, 6 et 8. Pour les classes numéro 1, 3, 5 et 7,

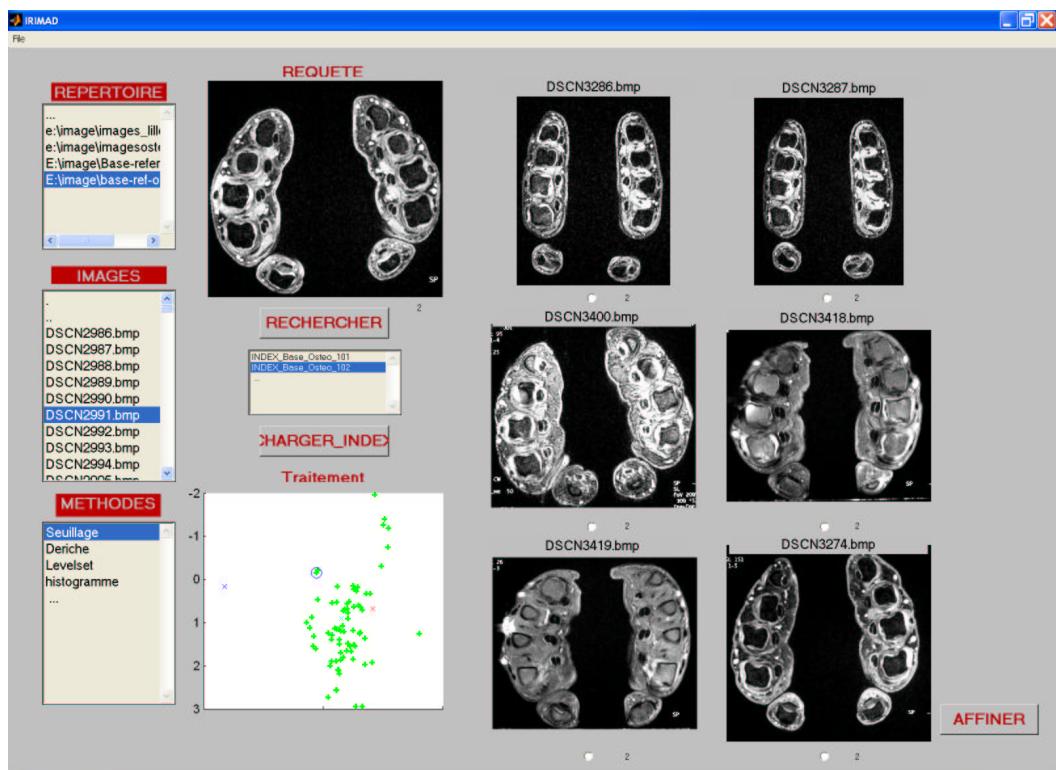


FIG. 6.23 – Résultats de recherche d'un métacarpe présentant de la synovite.

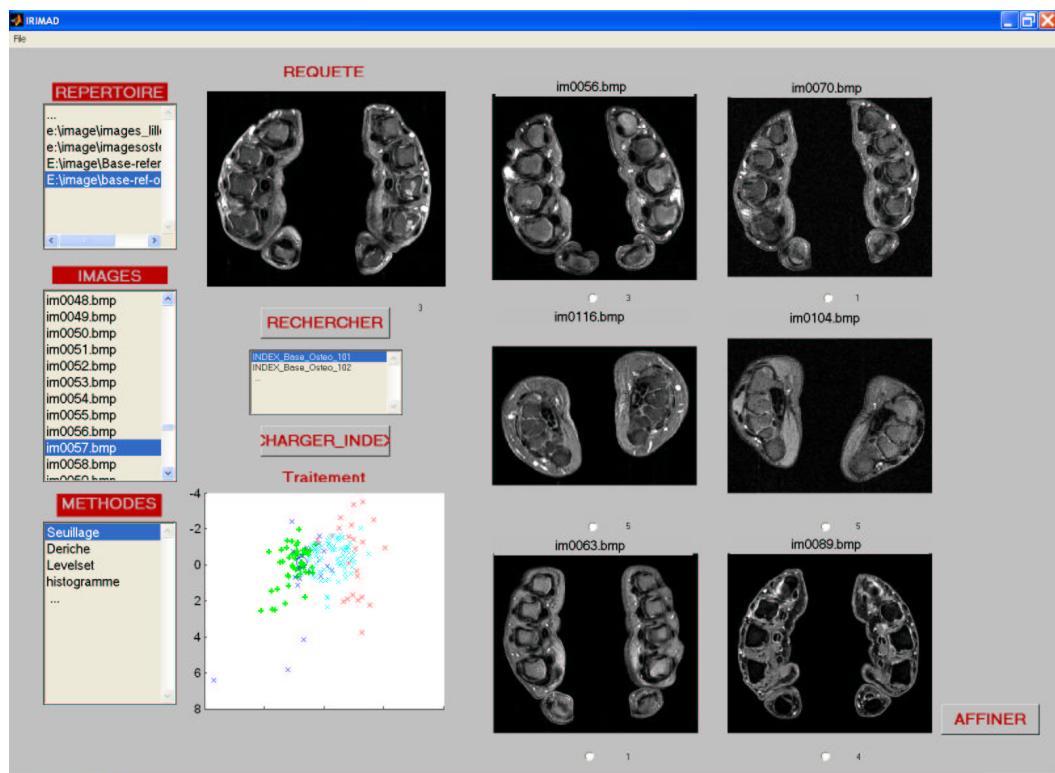


FIG. 6.24 – Résultats de recherche d'un métacarpe présentant de l'érosion.

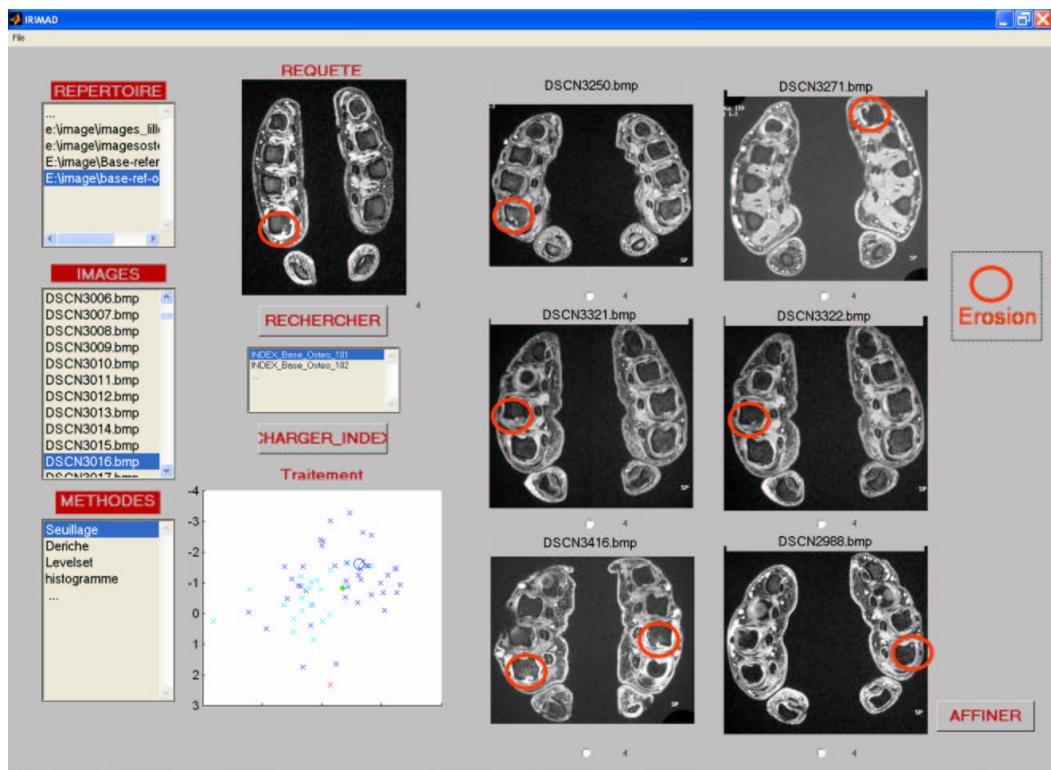


FIG. 6.25 – Résultats de recherche d'un métacarpe présentant de la synovite et de l'érosion.

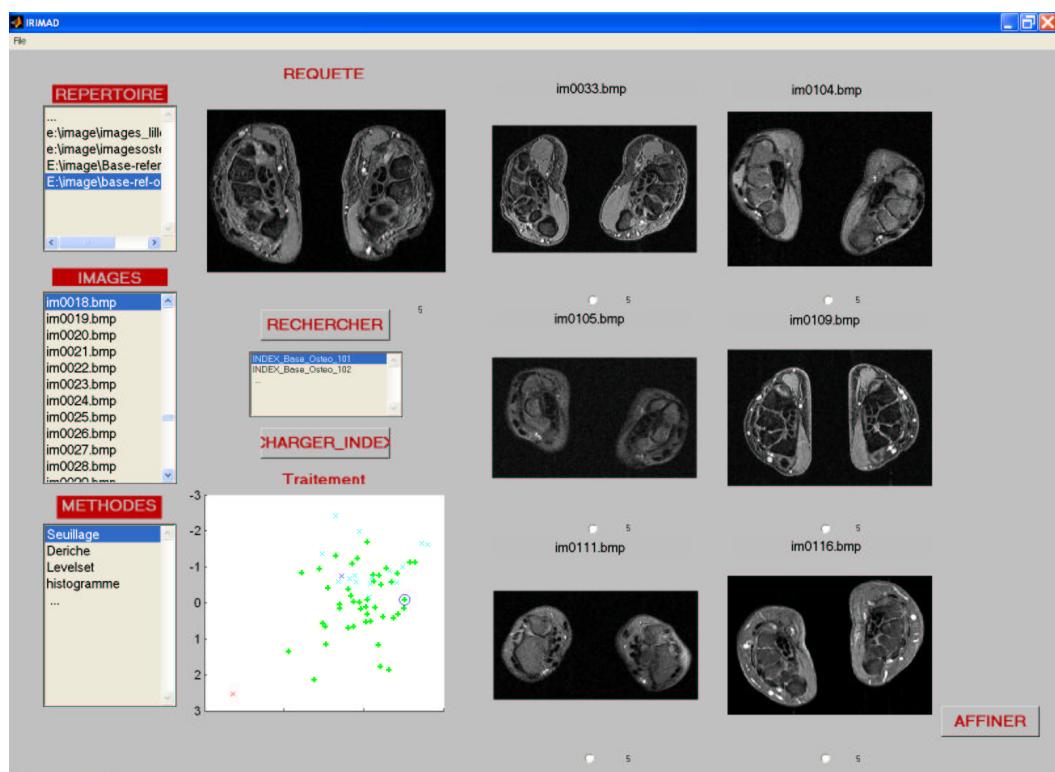


FIG. 6.26 – Résultats de recherche d'un carpe normal.

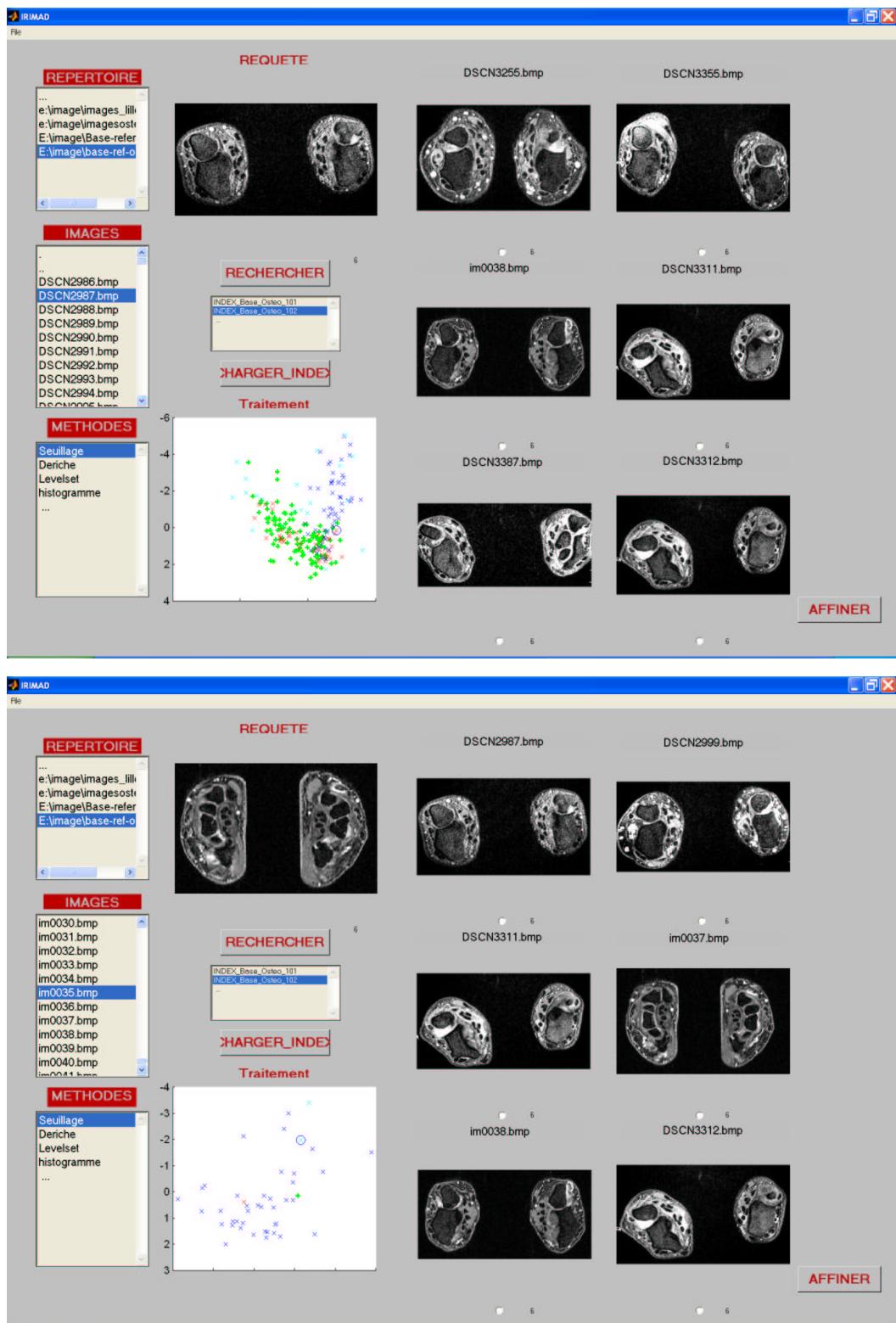


FIG. 6.27 – Résultats de recherche d'un carpe présentant de la synovite.

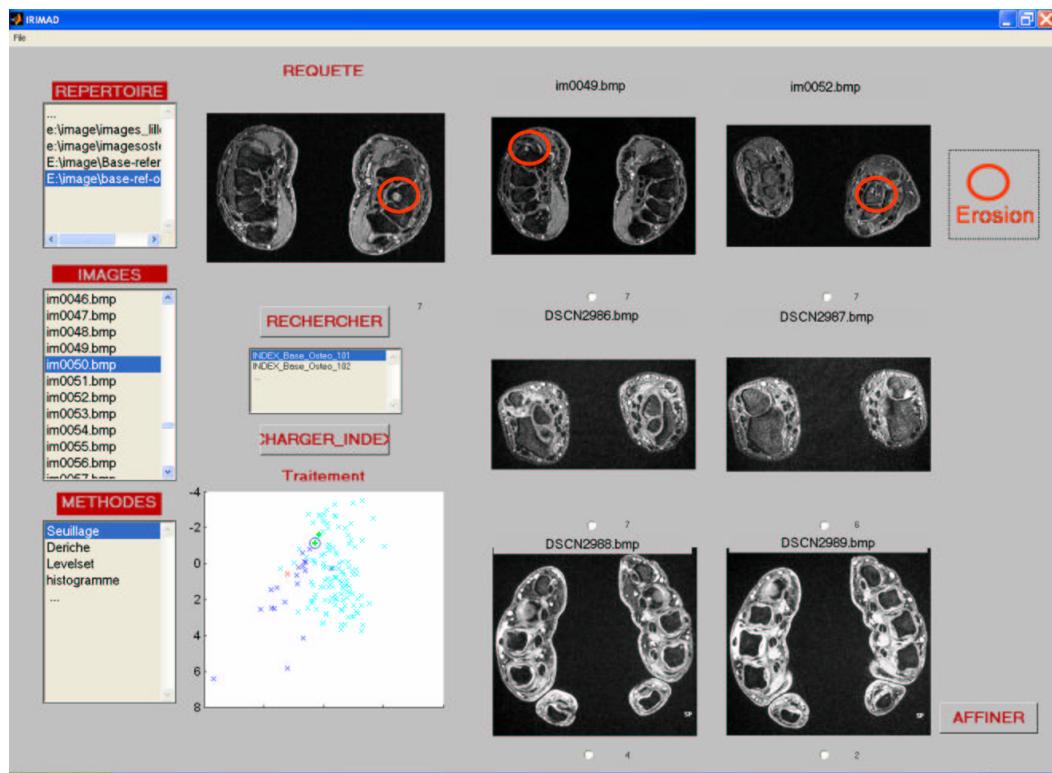


FIG. 6.28 – Résultats de recherche d'un carpe présentant de l'érosion.

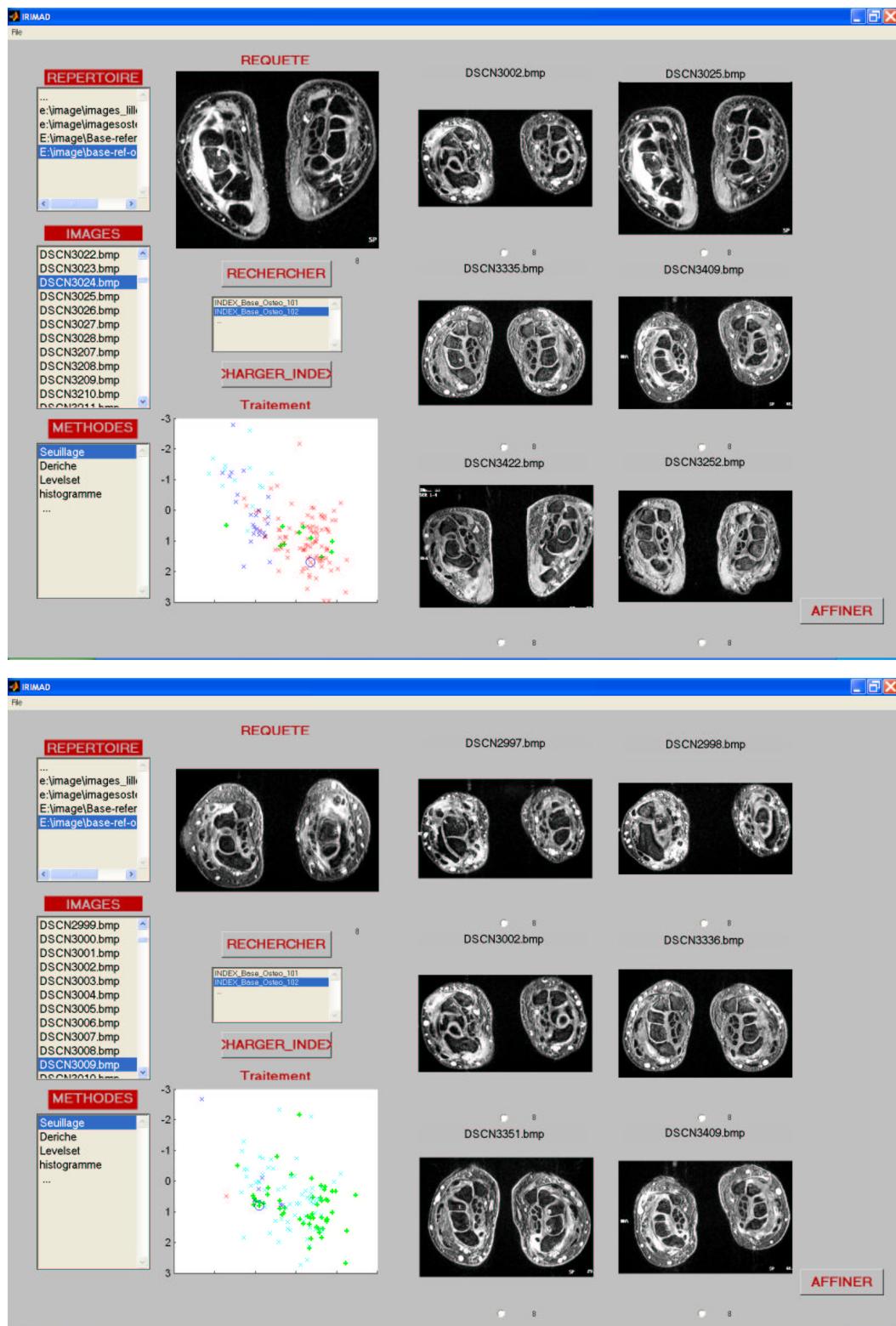


FIG. 6.29 – Résultats de recherche d'un carpe présentant de la synovite et de l'érosion.

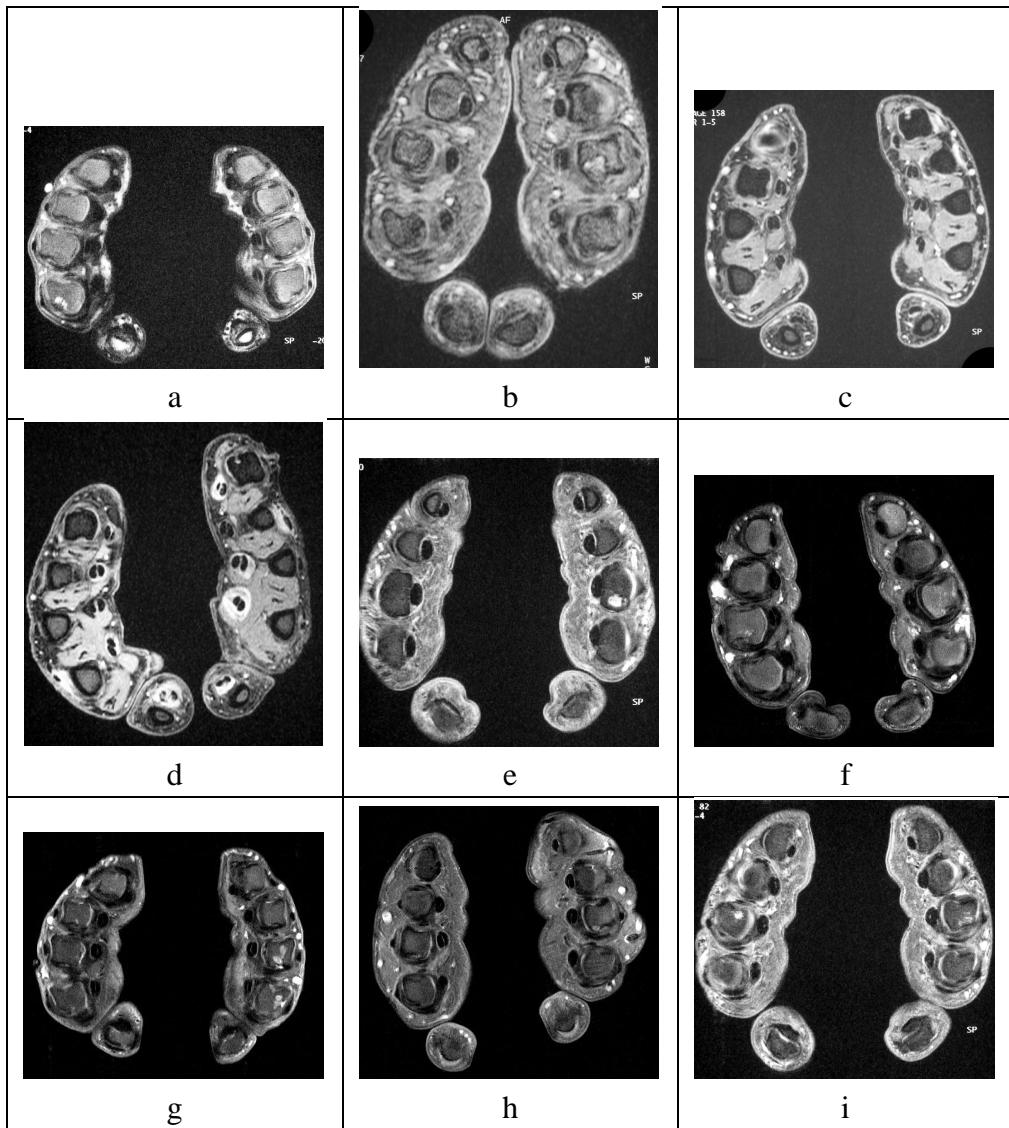


FIG. 6.30 – les neuf images de la classe "méta, érosion" dans la base.

très peu représentées (un faible effectif N_c , voir dernière ligne du même tableau 6.6), Notre système échoue à retrouver toutes les images semblables à l'image requête ($N_{rp} < 6$). Par contre, pour les classes de mains normales (1 et 5), nous ne sommes pas loin d'une précision égale à 1, elle est de 4/6 et 5/6 respectivement. Cela peut être expliqué par le fait que les mains normales (carpe et métacarpe) sont plus faciles à distinguer que les autres classes.

| $N_r \downarrow$; Catégorie → | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------------------------|----|----|---|----|----|----|----|----|
| 6 | 4 | 6 | 1 | 6 | 5 | 6 | 3 | 6 |
| 10 | 5 | 10 | 2 | 9 | 7 | 8 | 3 | 8 |
| 40 | 8 | 19 | 2 | 15 | 10 | 15 | 4 | 17 |
| 80 | 10 | 29 | 3 | 21 | 14 | 20 | 6 | 27 |
| $N_c \rightarrow$ | 27 | 90 | 9 | 63 | 34 | 61 | 23 | 88 |

TAB. 6.6 – Nombre d'images retrouvées pertinentes N_{rp} après 5 itérations. Cinq images sont annotées ($m = 5$) à chaque itération.

| $N_r \downarrow$; Catégorie → | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------------------------|----|----|---|----|----|----|----|----|
| 6 | 5 | 6 | 2 | 6 | 5 | 6 | 4 | 6 |
| 10 | 8 | 10 | 2 | 10 | 8 | 10 | 5 | 10 |
| 40 | 12 | 24 | 2 | 19 | 13 | 19 | 5 | 22 |
| 80 | 14 | 34 | 3 | 26 | 17 | 25 | 8 | 32 |
| $N_c \rightarrow$ | 27 | 90 | 9 | 63 | 34 | 61 | 23 | 88 |

TAB. 6.7 – Nombre d'images retrouvées pertinentes N_{rp} après 5 itérations. 10 images sont annotées ($m = 10$) à chaque itération.

6.7.4 Recherche par sélection de variables

Nous avons déjà conçu une méthode de sélection de variables dans la section 5.4 du chapitre 5 pour la recherche interactive d'images (voir schéma de la figure 5.16) dans le but de choisir l'ensemble de variables pertinentes pour chaque problème posé et d'améliorer, si c'est possible, les résultats de recherche.

Les résultats d'évaluation de la recherche par sélection de variables sont montrés dans la figure 6.33. Le protocole utilisé est le même que celui du tableau 6.5, mais nous avons remplacé la procédure de classification par **EMiner** par les procédures du schéma de la figure 5.15 (chapitre 5).

La figure 6.33 montre clairement les performances de la recherche par sélection de variables. Une amélioration des courbes de précision/rappel PR est constatée pour les huit classes. Le tableau 6.8 permet une lecture plus simple des performances. Une comparaison de ce tableau avec celui du paragraphe précédent (tableau 6.6) prouve l'amélioration des

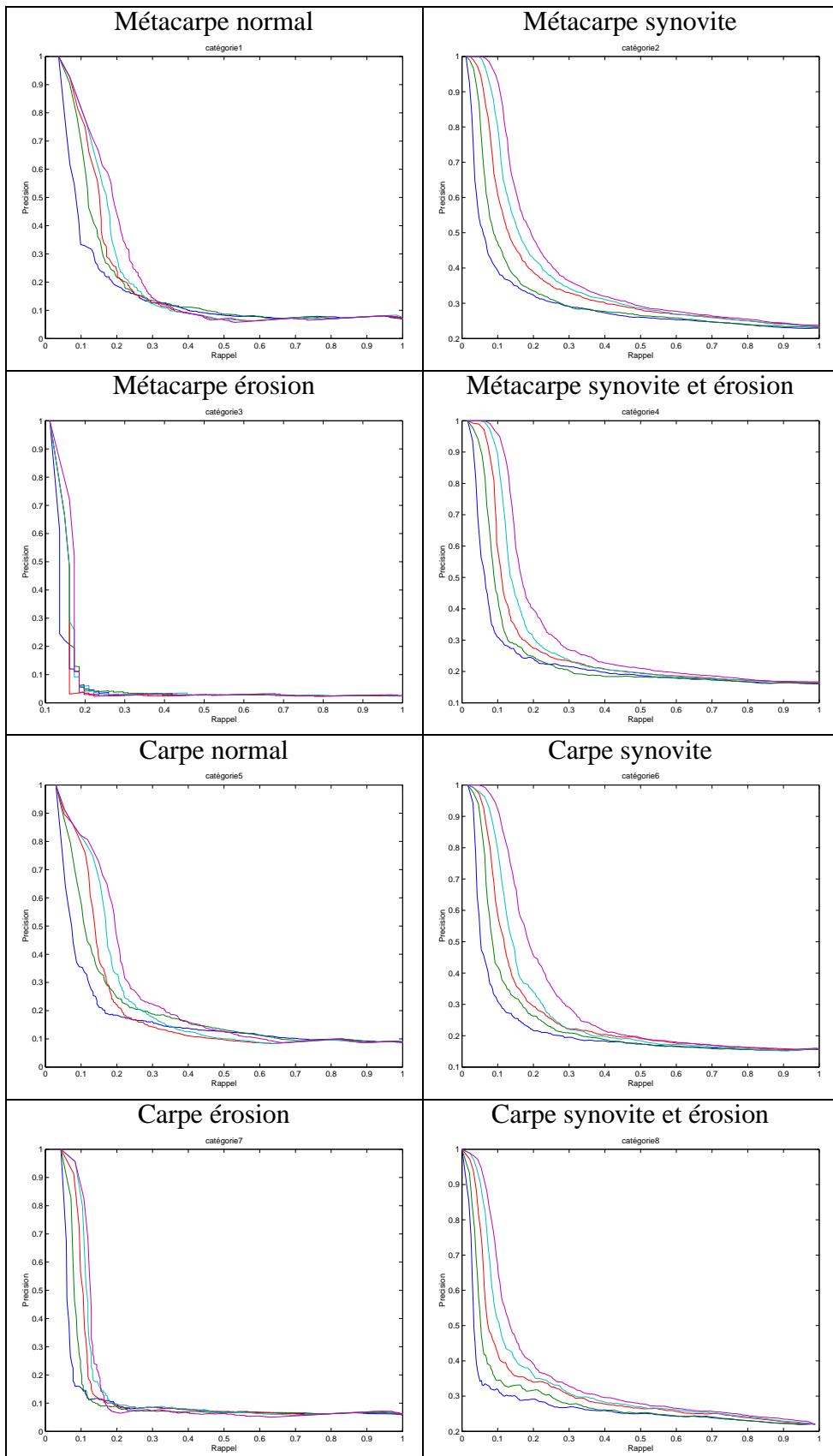


FIG. 6.31 – Evaluation : courbes de Précision versus Rappel des huit catégories après 1, 2, 3, 4 et 5 itérations. Index fabriqués par la carte de Kohonen avec 30 classes de régions. Cinq images sont annotées à chaque boucle de pertinence.

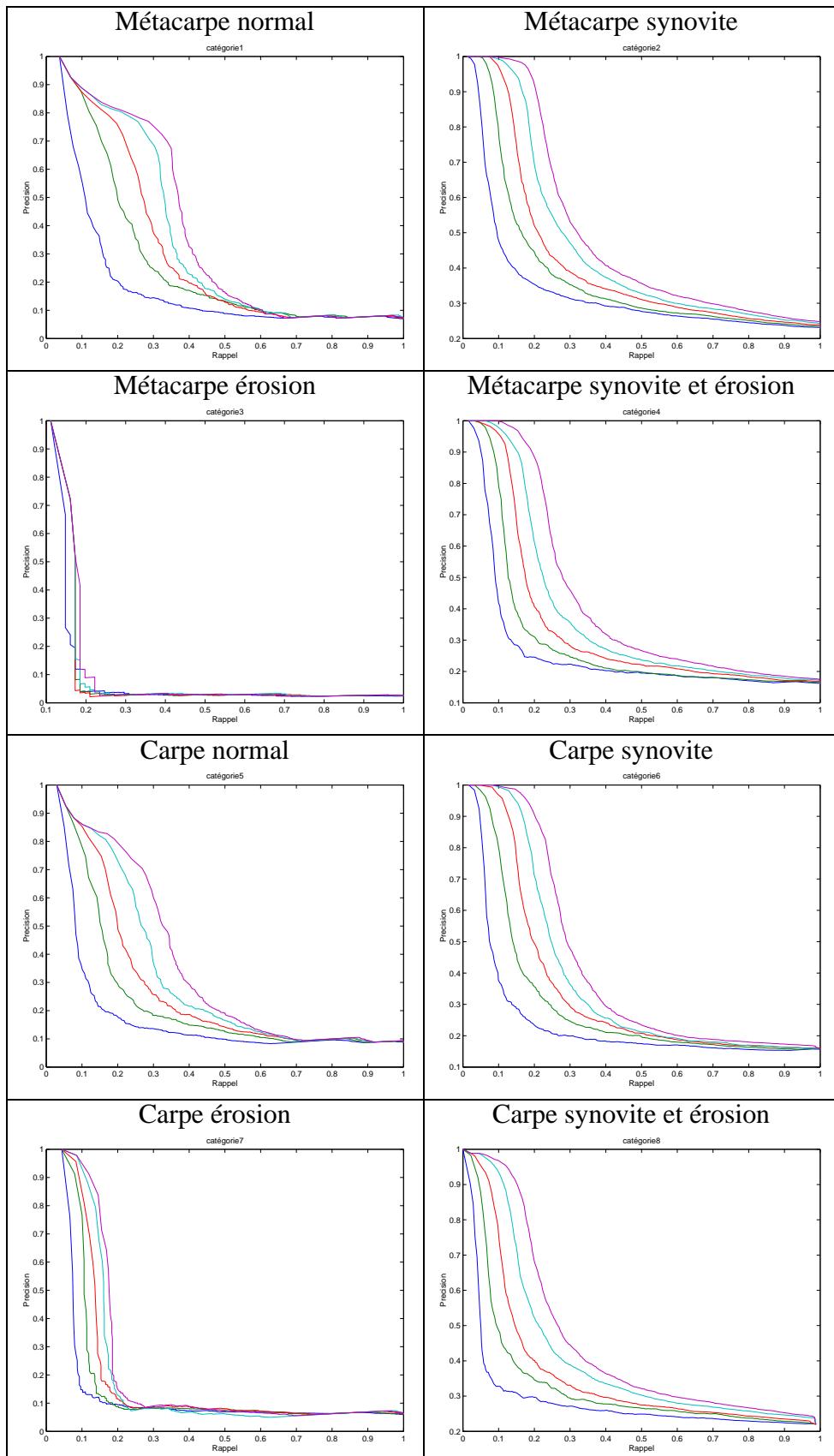


FIG. 6.32 – Evaluation : courbes Précision versus Rappel des huit catégories après 1, 2, 3, 4 et 5 itérations. Index fabriqués par la carte de Kohonen avec 30 classes de régions. 10 images sont annotées à chaque boucle de pertinence.

performances de la recherche avec sélection de variables et par conséquent celles de la méthode que nous proposons.

Partant de 30 variables au départ, le nombre de variables sélectionnées diminue au minimum d'une dizaine (voir dernière ligne du tableau 6.8). Il est à noter que l'ensemble des variables sélectionnées varie d'une recherche à une autre autrement dit d'une requête à une autre.

| $N_r \downarrow$; Catégorie → | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------------------------------------|----|----|----|----|----|----|----|----|
| 6 | 5 | 6 | 2 | 6 | 5 | 6 | 3 | 6 |
| 10 | 6 | 10 | 2 | 9 | 7 | 8 | 4 | 8 |
| 40 | 12 | 19 | 3 | 16 | 12 | 14 | 6 | 18 |
| 80 | 15 | 29 | 3 | 24 | 17 | 20 | 8 | 28 |
| $N_c \rightarrow$ | 27 | 90 | 9 | 63 | 34 | 61 | 23 | 88 |
| Nombre de variables sélectionnées → | 17 | 18 | 21 | 16 | 14 | 17 | 11 | 13 |

TAB. 6.8 – *Recherche avec sélection de variables : nombre d'images retrouvées pertinentes N_{rp} après 5 itérations. Cinq images sont annotées ($m = 5$) à chaque itération. Le nombre de variables au départ est de 30.*

6.8 Bilan

Nous avons proposé une méthode d'indexation des images IRM ostéo-articulaires des mains. Cette méthode repose sur l'analyse sémiologique effectuée par le médecin sur ces images. En plus des informations sur l'objet contenu dans l'image (carpe, métacarpe), les index fabriqués codent les indices visuels pour reconnaître les affections (synovite et érosion).

En effet, après une phase de prétraitement et suite à une segmentation automatique, chaque image de la base est représentée par un ensemble de régions chacune caractérisée par un ensemble d'attributs de forme, de niveau de gris et de texture. Cependant, pour obtenir un vecteur index qui a la même taille pour toutes les images, l'ensemble des régions de toute la base ont subi une classification non supervisée. L'index calculé pour chaque image est la distribution statistique des classes de ses régions. Cette distribution code, quelque part, la proportion des zones pathologiques dans l'image.

Ayant une base d'images de taille modeste, créée et étiquetée suivant huit classes par les médecins et l'informaticien du service ostéo-articulaire, nous avons proposé un protocole d'évaluation qui tient compte de sa faible taille. C'est un protocole basé sur la technique "leave one out" qui se sert des deux critères de qualité "rappel" et "précision". Les résultats de l'évaluation de notre système sont satisfaisants. Il retrouve par exemple six images qui sont toutes pertinentes (présentent les mêmes symptômes de pathologie) après 5 itérations au maximum avec 5 images annotées par le médecin à chaque itération. Nous avons amélioré ces résultats par l'approche de sélection de variables proposée dans le chapitre 5.

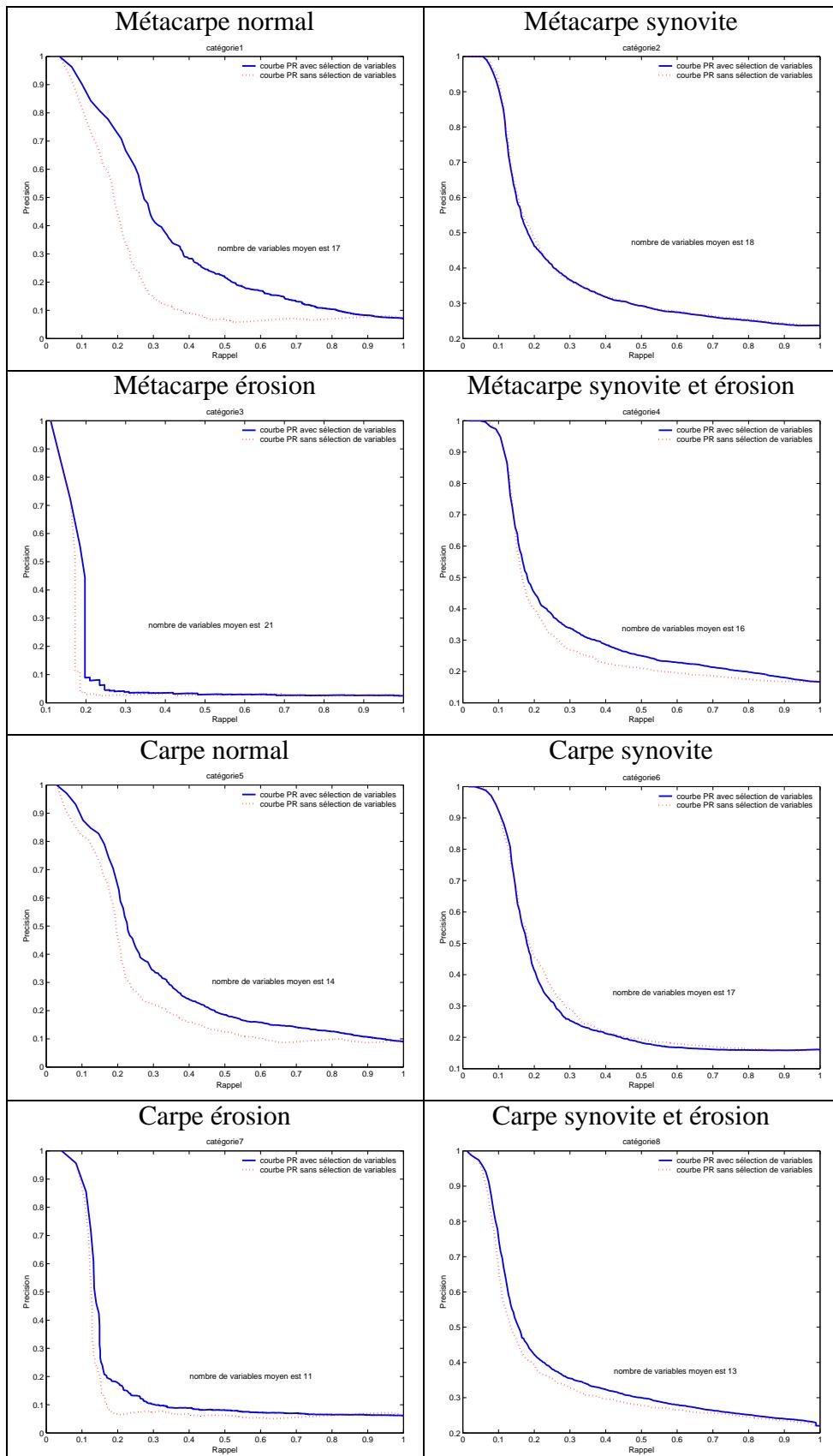


FIG. 6.33 – Courbes de précision/rappel PR avec et sans sélection de variables, après 5 itérations. Les courbes en trait plein (en pointillés) sont celles obtenues avec (sans) sélection de variables. Cinq images sont annotées à chaque boucle de pertinence ($m = 5$).

Chapitre 7

Conclusion générale

7.1 Récapitulatif du travail

Dans ce travail, nous nous sommes intéressés à la conception d'un système de recherche par le contenu pour une application aux images IRM ostéo-articulaires. Ce travail s'intègre dans le projet "Système de communication pour la formation en médecine basé sur l'indexation d'images". Il s'adresse au service de Radiologie Ostéo-Articulaire à l'hôpital Roger Salengro du CHRU de Lille.

L'indexation et la recherche d'images par le contenu (CBIR) demeurent des problèmes importants qui trouvent des applications potentielles en médecine, étant donné la place que l'image numérique médicale occupe à présent dans le quotidien des hôpitaux.

L'architecture de notre système est classique. Elle est formée de deux parties : l'indexation qui s'effectue "hors ligne" et la recherche qui s'effectue "en ligne". L'indexation ne peut être "universelle", elle est spécifique à chaque application.

Le *fossé sémantique*, qui sépare les attributs bas-niveaux calculés hors ligne dans les images et le sens que l'utilisateur place dans sa requête, conduit à des résultats de recherche erronés. La recherche par boucle de pertinence intègre l'expertise de l'utilisateur dans la recherche pour réduire le fossé sémantique et améliorer les résultats de la recherche grâce à ses annotations. Notre système utilise le principe de recherche par boucle de pertinence pour intégrer les connaissances des médecins dans la procédure de recherche.

La nouvelle approche que nous avons introduite est statistique. Elle a abordé le problème du bouclage de pertinence sous l'angle de la séparation des images pertinentes et non-pertinentes dans l'espace des index. Il s'agissait de construire le classifieur permettant de grouper les images pertinentes et de les discriminer des images non-pertinentes. Notre classifieur est basé sur un apprentissage semi-supervisé utilisant les modèles de mélange et l'algorithme EM. La technique proposée est appelée EMiner. Son originalité consistait à exploiter à la fois les images annotées et non annotées. Nous avons mis en évidence l'intérêt de l'apprentissage semi-supervisé à l'aide des mesures classiques en CBIR (rappel et précision). Un autre avantage de notre approche est qu'elle est capable de chercher

des catégories d'images multimodales, ce qui permet de traiter de structures de données complexes.

La sélection de variables demeure un problème important en CBIR. Les caractéristiques extraites d'une image pendant la phase d'indexation hors ligne codent plusieurs informations de l'image. Cependant, à chaque image requête correspond un sous ensemble de variables pertinentes qui permet de la distinguer au mieux de certaines images. Dans ce but, nous avons introduit une méthode originale de sélection de variables opérant pendant la recherche interactive.

Pour leur indexation, chaque image de la base est segmentée automatiquement par la détection des contours fermés utilisant la théorie de courbes de niveaux. L'initialisation des contours est basée sur les indices visuels décrits préalablement par des médecins. Après segmentation, chaque image est décrite par l'ensemble de ses régions chacune caractérisée par vingt attributs caractérisant le niveau de gris, la forme et la texture. Le nombre de régions obtenu est un peu élevé et diffère d'une image à une autre. Dans le but d'obtenir un index de même taille pour toutes les images afin d'accélérer la procédure de recherche en ligne, nous avons classé (classification non supervisée) les régions de la base par la carte auto-organisatrice de Kohonen. Ensuite l'index d'une image est calculé en fonction de la distribution statistique de ses régions par rapport aux classes des régions de la base.

La plupart des systèmes de recherche d'images médicales par le contenu développés, opèrent d'une façon semi automatique en indexation. Par exemple, pour le système Assert, l'utilisateur doit délimiter manuellement la région d'intérêt dans l'image, ce qui permettra par la suite le calcul d'un vecteur index traduisant au mieux les informations pertinentes de l'image. Dans notre système, l'utilisateur n'intervient pas dans l'étape de calcul d'index qui est ainsi effectué à partir de l'image entière. L'index d'une requête donnée peut contenir ainsi des informations pertinentes et non pertinentes pour l'utilisateur. Grâce à la méthode de sélection de variables intégrée dans la boucle de pertinence, nous choisissons les variables les plus intéressantes pour un problème donné. Ceci constitue un point d'originalité dans notre travail par rapport à d'autres systèmes.

L'évaluation a été réalisée sur une base d'images de taille modeste, étiquetée par les médecins et l'informaticien du service ostéo-articulaire. Nous pouvons distinguer huit classes dans la base qui sont des combinaisons des deux organes (carpe, métacarpe) avec les pathologies possibles (normal, synovite et érosion). La construction de cette base a demandé beaucoup de temps pour trouver dans les archives un nombre "suffisant" de dossiers représentatifs des cas pathologiques. Nous avons proposé un protocole d'évaluation qui tient compte de la faible taille de la base. Selon l'avis de l'équipe médicale, les résultats de l'évaluation de notre système sont satisfaisants.

Mise en place d'une plate forme

Une version écrite en language C et Visual Basic de notre système est transférée sur le système de fouille d'images développé par la société Archimed par l'intermédiaire d'Aurélien Flipo (2004) (étudiant UTC). Une démonstration est disponible sur le serveur de la Compagnie Archimed (<http://www.archimed.fr/>).

Bientôt, une plate forme sera mise en place dans le service de Radiologie Ostéo-Articulaire de l'hôpital Roger Salengro du CHRU de Lille.

7.2 Perspectives d'applications et de recherche

Applications

Deux applications potentielles de notre système ont été identifiées par les médecins du service ostéo-articulaire. Elles nécessitent la création d'une base d'images expertisée :

1. Evaluation thérapeutique :

A partir d'une image d'un patient, rechercher automatiquement les images semblables dans la base et retrouver, en particulier, les images d'exams antérieurs.

2. Formation :

Notre système interactif peut être utilisé comme un outil d'examen qui teste la capacité des étudiants à retrouver les images qui présentent les mêmes symptômes.

Perspectives de recherche

Vers un apprentissage à long-terme Le problème essentiel qui bloque les applications déjà citées, est la création de la base expertisée, annotée par les médecins et représentative de tous les cas rencontrés : elle demande du temps et la connaissance du médecin qui est en formation et évolution continues. D'où une autre application potentielle des systèmes de recherche par le contenu qui peut se mettre en amont de toutes les applications précédentes : l'apprentissage à long-terme (voir par exemple Vasconcelos et Lippman, 2001; Fournier et Cord, 2002). L'apprentissage à long-terme signifie la propagation de l'expertise que l'utilisateur introduit grâce au bouclage de pertinence, d'une session de recherche à l'autre.

Nous croyons qu'une conception d'une méthode d'apprentissage à long-terme permet la création d'une base structurée où les images sont étiquetées par catégories. Notre méthode d'apprentissage peut être facilement modifiée pour offrir un apprentissage à long-terme. Il suffit de sauvegarder les annotations z_i des utilisateurs d'une session de recherche à une autre et de mettre à jour la matrice de classification Z en faisant plus confiance à l'avis d'un même médecin à la session en cours qu'à celles précédentes.

EMiner Dans notre algorithme EMiner de recherche par boucle de pertinence, nous avons choisi de fixer *a priori* les nombres de gaussiennes des classes des images pertinentes et non pertinentes (R_1 et R_2) dans le but d'accélérer la recherche en ligne. Le calcul automatique de ces deux paramètres peut être possible pour améliorer notre approche. Govaert et al. (2000) proposent une méthode rapide de calcul de nombre de gaussiennes appliquée dans le cas d'un apprentissage non-supervisé. Nous pourrions adapter cette méthode dans le cas de l'apprentissage semi-supervisé afin de pouvoir l'intégrer dans EMiner.

Les travaux de Dy et Brodley (2000) en apprentissage non supervisé nous apparaissent importants parce qu'ils traitent le problème de sélection de variables en même temps que le

calcul du nombre de gaussiennes. Il serait donc intéressant d'envisager ce problème dans le futur.

La comparaison de EMiner avec une approche basée SVM offre plusieurs sujets de réflexions. Il serait intéressant de poursuivre cette étude avec l'"Equipe de Traitement des Images et du Signal" (Université de Cergy-Pontoise / ENSEA).

Protocole d'évaluation objective Nous avons déjà proposé dans ce mémoire un protocole strict d'évaluation , basé sur une vérité terrain, qui nous a permis de comparer notre méthode de recherche interactive avec une autre méthode basée SVM.

Un autre protocole d'évaluation du système nous paraît important : une évaluation "objective" basée sur l'avis de tous les utilisateurs potentiels. Il consiste à réaliser plusieurs essais avec les personnels du service hospitalier ayant différents profils : informaticiens, médecins spécialistes, étudiants en médecine, etc. C'est une évaluation objective qui montre la capacité du système à répondre au besoin de chacun.

Segmentation automatique Notre méthode d'indexation est basée sur une segmentation automatique réalisée par la théorie des courbes de niveaux. Il serait intéressant d'appliquer d'autres méthodes de segmentation automatique plus robustes au bruit.

Extraction d'attributs et calcul d'index L'extraction des attributs est un domaine de recherche très actif en reconnaissance des formes ainsi qu'en indexation. L'extraction et la sélection des informations pertinentes de l'image est une clé du problème de recherche d'image par le contenu. Notre stratégie d'indexation s'applique sur des attributs de niveau de gris, de texture et de forme calculés sur les régions de l'image. Nous pourrions envisager de compléter et d'affiner notre description de l'image par exemple par le recours à l'intégration de la structure spatiale des régions dans l'image.

Annexes

Annexe A

Algorithme EM et modèles de mélange

Nous illustrons dans cet annexe, les étapes de l'algorithme EM pour l'estimation des paramètres d'un modèle de mélange. Toutes sortes de supervision (voir définition dans chapitre 4) sont envisagées.

A.1 Hypothèses

Soit $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ un échantillon i.i.d de N observations générées par un modèle de mélange de R composantes. Ainsi, chaque composante r est modélisée par une distribution f_r paramétrique de paramètres représentés par le vecteur $\boldsymbol{\theta}_r$, π_r étant la proportion de la composante r , ($0 < \pi_r < 1$ et $\sum_{r=1}^R \pi_r = 1$). Ainsi nous avons :

$$f(\mathbf{x}_i | \Phi) = \sum_{r=1}^R \pi_r f_r(\mathbf{x}_i | \boldsymbol{\theta}_r), \quad (\text{A.1})$$

$$f(\mathcal{X} | \Phi) = \prod_{i=1}^N \sum_{r=1}^R \pi_r f_r(\mathbf{x}_i | \boldsymbol{\theta}_r), \quad (\text{A.2})$$

avec $\Phi = (\pi_1, \dots, \pi_{R-1}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R)$ le vecteur de paramètres à estimer.

Soit $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. $\mathbf{s}_i = (s_{ik}, r = 1, \dots, R)$ un vecteur binaire indiquant de quelle composante est issue \mathbf{x}_i : $s_{ir} \in \{0, 1\}$ et $\sum_{r=1}^R s_{ir} = 1$. $s_{ir} = 1$ signifie que \mathbf{x}_i provient de la $r^{i\text{ème}}$ composante.

Dans ce qui suit le vecteur \mathbf{s}_i est appelé *label* ou étiquette du vecteur \mathbf{x}_i .

En apprentissage non-supervisé \mathcal{S} est *inconnu*. Il est estimé en même temps que Φ . Dans le cas où chaque classe est formée par une seule composante, $\mathbf{s}_{i,(\forall i=1, \dots, N)}$ est *complètement connu* en apprentissage supervisé. Il est *partiellement connu* en apprentissage partiellement supervisé (voir section 4.1 du chapitre 4).

Prenons l'exemple de 5 classes, alors si \mathbf{x}_i est étiqueté par ne pas appartenir aux classes "1" et "2", alors \mathbf{s}_i est partiellement connu c'est à dire $s_{i1} = s_{i2} = 0$ (connus) et s_{i3}, s_{i4} et s_{i5} sont inconnus.

Par contre, si une classe est formée par plusieurs composantes, le label \mathbf{s}_i est partiellement connu en apprentissage semi et partiellement supervisé.

A.2 Algorithme EM

L'algorithme *Expectation Maximization* (EM) (Dempster et al., 1977) est conçu pour les problèmes d'estimation des paramètres des distributions où les données observées sont incomplètes (ayant des informations manquantes).

Désignant par \mathcal{O} les données observées incomplètes et par \mathcal{D} les données complétées.

La maximisation de la log-vraisemblance $L(\Phi; \mathcal{X}) = \log f(\mathcal{X}|\Phi)$ étant un problème difficile, le principe de l'algorithme EM consiste à maximiser itérativement la quantité Q définie comme suit :

$$L(\Phi; \mathcal{X}) = Q(\Phi|\Phi^{(q)}) - H(\Phi|\Phi^{(q)}), \quad (\text{A.3})$$

où

$$\begin{aligned} Q(\Phi|\Phi^{(q)}) &= \mathbb{E}[L(\Phi; \mathcal{D})|\mathcal{O}, \Phi^{(q)}]; \\ H(\Phi|\Phi^{(q)}) &= \mathbb{E}[L(\Phi; \mathcal{D}|\mathcal{O})|\mathcal{O}, \Phi^{(q)}]. \end{aligned}$$

avec $\Phi^{(q)}$ une valeur donnée du vecteur Φ .

Q s'interprète comme l'espérance de la log-vraisemblance complète, espérance prise par rapport aux données manquantes connaissant les données observées et les paramètres $\Phi^{(q)}$.

Les deux étapes de l'algorithme EM s'écrivent :

- **Etape E :** Calculer $Q(\Phi|\Phi^{(q)})$,
- **Etape M :** Trouver $\Phi^{(q+1)}$ maximisant $Q(\Phi|\Phi^{(q)})$:

$$\Phi^{(q+1)} = \arg \max_{\Phi} Q(\Phi|\Phi^{(q)}).$$

A.3 Apprentissage non-supervisé

Dans ce cas les informations manquantes sont l'ensemble \mathcal{S} des vecteurs \mathbf{s}_i . Les données complètes sont alors $\mathcal{D} = (\mathcal{X}, \mathcal{S})$, avec $\mathcal{O} = \mathcal{X}$.

Indiquons les paramètres à estimer par \mathbf{s}_i à la place de \mathbf{r} lorsque $s_{ir} = 1$ et écrivons la densité de l'échantillon $(\mathcal{X}, \mathcal{S})$:

$$g(\mathcal{X}, \mathcal{S}|\Phi) = \prod_{i=1}^N g(\mathbf{x}_i, \mathbf{s}_i) = \prod_{i=1}^N \pi_{s_i} f_{s_i}(\mathbf{x}_i | \boldsymbol{\theta}_{s_i}). \quad (\text{A.4})$$

Q s'écrit :

$$\begin{aligned}
Q(\Phi|\Phi^{(q)}) &= \mathbb{E}^r[L(\Phi; \mathcal{D}|\mathcal{O}, \Phi^{(q)})] = \mathbb{E}^r[L(\Phi; \mathcal{X}, \mathcal{S})|\mathcal{X}, \Phi^{(q)}], \\
&= \mathbb{E}^r[\log \prod_{i=1}^N \pi_{s_i} f_{s_i}(\mathbf{x}_i | \boldsymbol{\theta}_{s_i}) |\mathcal{X}, \Phi^{(q)}], \\
&= \sum_{i=1}^N \mathbb{E}^r[\log \pi_{s_i} f_{s_i}(\mathbf{x}_i | \boldsymbol{\theta}_{s_i}) |\mathcal{X}, \Phi^{(q)}].
\end{aligned}$$

D'où

$$Q(\Phi|\Phi^{(q)}) = \sum_{i=1}^N \sum_{r=1}^R t_r^{(q)}(\mathbf{x}_i) \log \pi_r f_r(\mathbf{x}_i | \boldsymbol{\theta}_r), \quad (\text{A.5})$$

avec

$$t_r^{(q)}(\mathbf{x}_i) = \frac{\pi_r^{(q)} f_r(\mathbf{x}_i | \boldsymbol{\theta}_r^{(q)})}{f_r(\mathbf{x}_i)}. \quad (\text{A.6})$$

Ainsi les deux étapes de l'algorithme EM s'écrivent :

- **Etape E** : calcul des parties de $Q(\Phi|\Phi^{(q)})$ ne dépendant pas de Φ , soit ici, pour $1 \leq i \leq N, 1 \leq r \leq R$

$$t_r^{(q)}(\mathbf{x}_i) = p(s_{ir} = 1 | \mathbf{x}_i, \Phi^{(q)}) = \frac{\pi_r^{(q)} f_r(\mathbf{x}_i | \boldsymbol{\theta}_r^{(q)})}{\sum_{r=1}^R \pi_r^{(q)} f_r(\mathbf{x}_i | \boldsymbol{\theta}_r^{(q)})}. \quad (\text{A.7})$$

- **Etape M** : Calcul de $\Phi^{(q+1)}$ qui maximise

$$Q(\Phi|\Phi^{(q)}) = \sum_{i=1}^N \sum_{r=1}^R t_r^{(q)}(\mathbf{x}_i) \log \pi_r f_r(\mathbf{x}_i | \boldsymbol{\theta}_r).$$

A.3.1 Cas d'un mélange gaussien

Les estimateurs du maximum de vraisemblance s'écrivent alors :

$$\boldsymbol{\mu}_r^{(q+1)} = \frac{\sum_{i=1}^N t_r^{(q)}(\mathbf{x}_i) \cdot \mathbf{x}_i}{n_r^{(q)}}; \quad (\text{A.8})$$

$$\boldsymbol{\Sigma}_r^{(q+1)} = \frac{1}{n_r} \sum_{i=1}^N t_r^{(q)}(\mathbf{x}_i) (\mathbf{x}_i - \boldsymbol{\mu}_r^{(q+1)}) (\mathbf{x}_i - \boldsymbol{\mu}_r^{(q+1)})^T; \quad (\text{A.9})$$

$$p_r^{(q+1)} = \frac{n_r^{(q)}}{N}; \quad (\text{A.10})$$

où

$$n_r^{(q)} = \sum_{i=1}^N t_r^{(q)}(\mathbf{x}_i). \quad (\text{A.11})$$

A.4 Prise en compte de labels complètement connus

Considérons maintenant le cas où m individus parmi les N observations ont des labels complètement connus (issus d'une composante de mélange connue). Les données observées sont notées :

$$\mathcal{O} = \{(\mathbf{x}_1, s_1), \dots, (\mathbf{x}_m, s_m), \mathbf{x}_{m+1}, \dots, \mathbf{x}_N\}.$$

La log-vraisemblance complète peut se décomposer en deux termes :

$$L(\Phi; \mathcal{D}) = L_m(\Phi; \mathcal{D}) + L_{N-m}(\Phi; \mathcal{D}), \quad (\text{A.12})$$

où

- $L_m(\Phi; \mathcal{D})$ est la vraisemblance de Φ par rapport aux observations de composante connue,
- $L_{N-m}(\Phi; \mathcal{D})$ est la vraisemblance de Φ par rapport aux observations de composante inconnue.

Lorsque nous calculons le critère Q , nous obtenons :

$$\begin{aligned} Q(\Phi | \Phi^{(q)}) &= \mathbb{E}^r [L(\Phi; \mathcal{D}) | \mathcal{O} = ((\mathbf{x}_1, s_1), (\mathbf{x}_2, s_2), \dots, (\mathbf{x}_m, s_m), \mathbf{x}_{m+1}, \dots, \mathbf{x}_N), \Phi^{(q)}], \\ &= \mathbb{E}^r [L_m(\Phi; \mathcal{D}) + L_{N-m}(\Phi; \mathcal{D}) | \mathcal{X}, \Phi^{(q)}], \\ &= L_m(\Phi; \mathcal{D}) + \mathbb{E}^r [L_{N-m}(\Phi; \mathcal{D}) | (\mathbf{x}_{m+1}, \dots, \mathbf{x}_N), \Phi^{(q)}]. \end{aligned}$$

D'où

$$Q(\Phi | \Phi^{(q)}) = \sum_{i=1}^m \log \pi_{s_i} f_{s_i}(\mathbf{x}_i | \boldsymbol{\theta}_{s_i}) + \sum_{j=m+1}^N \sum_{r=1}^R t_r^{(q)}(\mathbf{x}_j) \log \pi_r f_r(\mathbf{x}_j | \boldsymbol{\theta}_r). \quad (\text{A.13})$$

A.5 Prise en compte de labels partiellement connus

Considérons maintenant un échantillon où m individus parmi les N sont issus de composante connue et dont l individus sont connus *pour ne pas provenir de certaines composantes*. Les $N - l - m$ observations sont issues de composante inconnue. Nous avons donc un échantillon \mathcal{O} de la forme :

$$\begin{aligned} \mathcal{O} = & \{(\mathbf{x}_1, s_1), \dots, (\mathbf{x}_m, s_m), \\ & (\mathbf{x}_{m+1}, \{s_{m+1, r_{m+1}}\}_{r_{m+1} \in R_{m+1}}), \dots, (\mathbf{x}_{m+l}, \{s_{m+l, r_{m+l}}\}_{r_{m+l} \in R_{m+l}}), \\ & \mathbf{x}_{m+l+1}, \dots, \mathbf{x}_N\}, \end{aligned}$$

où R_j est l'ensemble des indices des composantes auxquelles x_j n'appartient pas, c'est à dire pour lesquelles les $s_{jr} = 0$ si $r \in R_j$.

La log-vraisemblance complète peut se décomposer en trois termes :

$$L(\Phi; \mathcal{D}) = L_m(\Phi; \mathcal{D}) + L_l(\Phi; \mathcal{D}) + L_{N-m-l}(\Phi; \mathcal{D}).$$

où

- $L_m(\Phi; \mathcal{D})$ est la vraisemblance de Φ par rapport aux observations de composante connue,
- $L_{N-m-l}(\Phi; \mathcal{D})$ est la vraisemblance de Φ par rapport aux observations de composante inconnue,
- $L_l(\Phi; \mathcal{D})$ est la vraisemblance de Φ par rapport aux observations qui n'appartiennent pas à certaines composantes.

L'espérance Q s'exprime comme :

$$\begin{aligned} Q(\Phi | \Phi^{(q)}) &= \mathbb{E}^r[L(\Phi; \mathcal{D}) | \mathcal{X}, \Phi^{(q)}], \\ &= \mathbb{E}^r[L_m(\Phi; \mathcal{D}) + L_l(\Phi; \mathcal{D}) + L_{N-m-l}(\Phi; \mathcal{D}) | \mathcal{X}, \Phi^{(q)}], \\ &= L_m(\Phi; \mathcal{D}) + \mathbb{E}^r[L_l(\Phi; \mathcal{D}) | \mathcal{X}, \Phi^{(q)}] + \mathbb{E}^r[L_{N-m-l}(\Phi; \mathcal{D}) | \mathcal{X}, \Phi^{(q)}]. \end{aligned}$$

Si nous détaillons le calcul de $\mathbb{E}^r[L_l(\Phi; \mathcal{D}) | \mathcal{X}, \Phi^{(q)}]$ nous obtenons :

$$\begin{aligned} \mathbb{E}^r[L_l(\Phi; \mathcal{D}) | \mathcal{X}, \Phi^{(q)}] &= \mathbb{E}^r[L_l(\Phi; \mathcal{D}) | (\mathbf{x}_{m+1}, \{s_{m+1, r_{m+1}}\}_{r_{m+1} \in R_{m+1}}), \\ &\quad \dots, (\mathbf{x}_{m+l}, \{s_{m+l, r_{m+l}}\}_{r_{m+l} \in R_{m+l}}), \Phi^{(q)}], \\ &= \sum_{j=m+1}^{m+l} \mathbb{E}^r[L_l(\Phi; \mathbf{x}_j, \mathbf{s}_j) | (\mathbf{x}_j, \{s_{j, r_j}\}_{r_j \in R_j}), \Phi^{(q)}], \\ &= \sum_{j=m+1}^{m+l} \sum_{r=1}^R p^{(q)}(s_{jr} = 1 | (\mathbf{x}_j, \{s_{j, r_j}\}_{r_j \in R_j}), \boldsymbol{\theta}^{(q)}) \log \pi_r f_r(\mathbf{x}_j | \boldsymbol{\theta}_r). \end{aligned}$$

avec

$$p^{(q)}(s_{jr} = 1 | (\mathbf{x}_j, \{s_{j, r_j}\}_{r_j \in R_j}), \boldsymbol{\theta}^{(q)}) = u_r^{(q)}(\mathbf{x}_j) = \begin{cases} 0 & \text{si } r \in R_j, \\ \frac{\pi_r^{(q)} f_r(\mathbf{x}_j | \boldsymbol{\theta}_r^{(q)})}{\sum_{r' \notin R_j} \pi_r^{(q)} f_{r'}(\mathbf{x}_j | \boldsymbol{\theta}_{r'}^{(q)})} & \text{sinon.} \end{cases} \quad (\text{A.14})$$

Ainsi $Q(\Phi | \Phi^{(q)})$ peut être développé comme suit :

$$\begin{aligned} Q(\Phi | \Phi^{(q)}) &= \sum_{i=1}^m \log \pi_{s_i} f_{s_i}(\mathbf{x}_i | \boldsymbol{\theta}_{s_i}) \\ &+ \sum_{j=m+1}^{m+l} \sum_{r=1}^R u_r^{(q)}(\mathbf{x}_j) \log \pi_r f_r(\mathbf{x}_j | \boldsymbol{\theta}_r) \\ &+ \sum_{j'=m+l+1}^N \sum_{r=1}^R t_r^{(q)}(\mathbf{x}_{j'}) \log \pi_r f_r(\mathbf{x}_{j'} | \boldsymbol{\theta}_r). \end{aligned} \quad (\text{A.15})$$

Le critère $Q(\Phi|\Phi^{(q)})$ peut s'écrire d'une manière condensée comme :

$$Q(\Phi|\Phi^{(q)}) = \sum_{i=1}^N \sum_{r=1}^R c_{ir}^{(q)} \log \pi_r f_r(\mathbf{x}_i|\boldsymbol{\theta}_r), \quad (\text{A.16})$$

avec

$$c_{ir}^{(q)} = \begin{cases} 1 & \text{si } s_{ir} = 1 \text{ et } \sum_{r=1}^R s_{ir} = 1, \\ 0 & \text{si } s_{ir} = 0, \\ \frac{\pi_r^{(q)} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r^{(q)})}{\sum_{r' \notin R_i} \pi_r^{(q)} f_{r'}(\mathbf{x}_i|\boldsymbol{\theta}_{r'}^{(q)})} & \text{si } (r \notin R_i) \text{ autrement } (s_{ir} = 1 \text{ et } 1 < \sum_{r=1}^R s_{ir} < R), \\ \frac{\pi_r^{(q)} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r^{(q)})}{f(\mathbf{x}_i) = \sum_{r=1}^R \pi_r^{(q)} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r^{(q)})} & \text{sinon.} \end{cases} \quad (\text{A.17})$$

A.6 Formules finales

Pour trouver une formule simplifiée de $c_{ir}^{(q)}$ (équation A.17) et pour une implémentation simple de l'algorithme, il suffit de coder les vecteurs s_i par \mathbf{z}_i comme suit :

- ($z_{ir} = 1$ et $\sum_{r=1}^R z_{ir} = 1$) pour s_i connu (\mathbf{x}_i issu de la composante r),
- ($z_{ir} = 1 \forall r \in R_i$ et $z_{ir} = 0$ sinon), pour s_i partiellement connu (R_i représente l'ensemble des composantes auxquelles l'exemple \mathbf{x}_i n'appartient pas),
- ($z_{ir} = 1 \forall r = 1, \dots, R$) pour s_i inconnu (l'exemple i est de composante inconnue).

Le vecteur \mathbf{z}_i est appelé vecteur *annotation* ou *label*.

Ainsi $c_{ir}^{(q)}$ est calculée comme suit :

$$c_{ir}^{(q)} = \frac{z_{ir} \pi_r^{(q)} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r^{(q)})}{f(\mathbf{x}_i) = \sum_{r=1}^R z_{ir} \pi_r^{(q)} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r^{(q)})}, \quad (\text{A.18})$$

et finalement d'après la définition de $c_{ir}^{(q)}$ nous obtenons :

$$p(r|\mathbf{x}_i, \Phi_r^{(q)}) = c_{ir}^{(q)} = \frac{z_{ir} \pi_r^{(q)} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r^{(q)})}{\sum_{r=1}^R z_{ir} \pi_r^{(q)} f_r(\mathbf{x}_i|\boldsymbol{\theta}_r^{(q)})}. \quad (\text{A.19})$$

A.6.1 Modèle de mélange gaussien

Les estimateurs du maximum de vraisemblance s'écrivent alors :

$$\boldsymbol{\mu}_r^{(q+1)} = \frac{\sum_{i=1}^N c_{ir}^{(q)} (\mathbf{x}_i) \cdot \mathbf{x}_i}{n_r^{(q)}}; \quad (\text{A.20})$$

$$\boldsymbol{\Sigma}_r^{(q+1)} = \frac{1}{n_r} \sum_{i=1}^N c_{ir}^{(q)}(\boldsymbol{x}_i) \cdot (\boldsymbol{x}_i - \boldsymbol{\mu}_r^{(q+1)}) (\boldsymbol{x}_i - \boldsymbol{\mu}_r^{(q+1)})^T; \quad (\text{A.21})$$

$$p_r^{(q+1)} = \frac{n_r^{(q)}}{N}; \quad (\text{A.22})$$

où

$$n_r^{(q)} = \sum_{i=1}^N c_{ir}^{(q)}(\boldsymbol{x}_i). \quad (\text{A.23})$$

Annexe B

Les Machines à vecteurs de support

Les Machines à vecteurs de support (SVM) sont principalement issues des travaux de Vapnik (voir Vapnik, 1995) à la fin des années 70. Contrairement aux méthodes traditionnelles qui cherchent à minimiser l'erreur sur l'ensemble d'apprentissage, Vapnik propose de la remplacer par la minimisation du risque structurel. Ce principe repose grossièrement sur la minimisation conjointe du risque empirique et de la complexité du modèle apprenant.

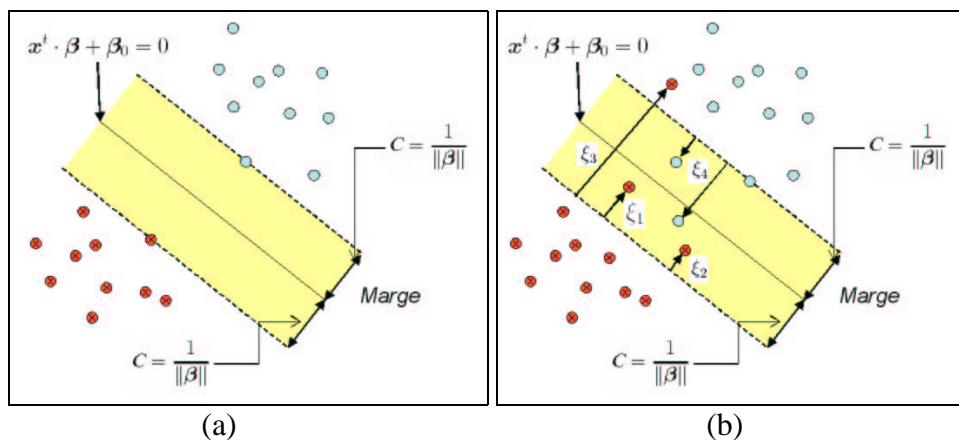


FIG. B.1 – *Classifieur à vecteurs de support. (a) cas de deux classes séparables. La frontière de décision est présentée par la ligne en trait plein, par contre les lignes en traits tirets limitent la marge maximale de largeur $2C$. (b)cas des deux classes non-séparables linéairement. La marge est maximisée sous la contrainte $\sum \xi_i \leq \text{constante}$: somme totale des erreurs commis.*

B.1 Détermination d'une frontière linéaire

B.1.1 Cas de deux classes séparables par un hyperplan

Considérant le problème de recherche de frontière linéaire dans le cas de deux classes. Nous disposons ainsi d'un ensemble de couple points-classes $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$. y_i est le label de l'exemple \mathbf{x}_i , il est égal à +1 pour les exemples positifs et -1 pour les exemples négatifs. Dans le cas où les classes sont séparables par un hyperplan, l'équation de celui-ci peut s'écrire :

$$\{\mathbf{x} : h(\mathbf{x}) = \mathbf{x}^t \cdot \boldsymbol{\beta} + \beta_0 = 0\}. \quad (\text{B.1})$$

La règle de classification est :

$$J(\mathbf{x}) = \text{signe}(\mathbf{x}^t \cdot \boldsymbol{\beta} + \beta_0).$$

Le but de SVM est de trouver $\boldsymbol{\beta}$ et β_0 de l'hyperplan optimal qui maximise la distance C entre l'hyperplan et le point \mathbf{x}_i le plus proche (cf. figure B.1.a) :

$$\begin{aligned} & \max_{\boldsymbol{\beta}, \beta_0, \|\boldsymbol{\beta}\|=1} C, \\ \text{sous la contrainte : } & y_i(\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0) \geq C, \quad \text{pour tout } i = 1, \dots, l. \end{aligned} \quad (\text{B.2})$$

Cela permet aussi de maximiser la marge entre les deux classes sur l'ensemble d'apprentissage.

Ce problème d'optimisation peut se transformer en une minimisation de $\frac{1}{2}\|\boldsymbol{\beta}\|^2$ comme suit :

$$\begin{aligned} & \min_{\boldsymbol{\beta}, \beta_0} \frac{1}{2}\|\boldsymbol{\beta}\|^2, \\ \text{sous la contrainte : } & y_i(\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0) \geq 1, \quad \forall i. \end{aligned} \quad (\text{B.3})$$

C'est un problème d'optimisation convexe (critère quadratique avec comme contrainte une inégalité linéaire). La fonction de Lagrange est donc :

$$L_P = \min_{\boldsymbol{\beta}, \beta_0} \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \sum_{i=1}^l \alpha_i [y_i(\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0) - 1]. \quad (\text{B.4})$$

La mise à zéro des dérivées par rapport à $\boldsymbol{\beta}$ et β_0 nous donne :

$$\boldsymbol{\beta} = \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i, \quad (\text{B.5})$$

$$0 = \sum_{i=1}^l \alpha_i y_i. \quad (\text{B.6})$$

Finalement et après la substitution de ces dernières équations dans B.4, le problème d'optimisation devient simple à résoudre :

$$\max_{\alpha} \sum_{i=1}^l \alpha_i + \frac{1}{2} \sum_{i,j}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i^t \mathbf{x}_j, \quad (\text{B.7})$$

sous la contrainte :

$$\alpha_i \geq 0 \quad \forall i.$$

L'équation de l'hyperplan s'écrit alors :

$$\{\mathbf{x} : h(\mathbf{x}) = \mathbf{x}^t (\sum_i \alpha_i y_i \mathbf{x}_i) + \beta_0 = \sum_i \alpha_i y_i \mathbf{x}^t \mathbf{x}_i + \beta_0 = 0\}. \quad (\text{B.8})$$

La règle de classification est :

$$J(\mathbf{x}) = \text{signe}(\sum_{i=1}^l \alpha_i y_i \mathbf{x}^t \mathbf{x}_i + \beta_0). \quad (\text{B.9})$$

B.1.2 Cas de deux classes non séparables linéairement

Supposons maintenant que les données ne sont pas linéairement séparables. L'idée est de relâcher les contraintes sur les données en introduisant un terme d'erreur $\xi \geq 0$ qui permet à quelques points d'être dans un mauvais endroit (ex. à l'intérieur de la marge) (voir figure B.1.b). Le problème d'optimisation devient :

$$\begin{aligned} & \max_{\boldsymbol{\beta}, \beta_0, \|\boldsymbol{\beta}\|=1} C, \\ & \text{sous les contraintes :} \\ & \begin{cases} y_i (\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0) \geq C - \xi_i \text{ ou } C(1 - \xi_i), \quad \forall i, \\ \xi_i \geq 0, \sum \xi_i \leq \text{constante}. \end{cases} \end{aligned} \quad (\text{B.10})$$

Afin de trouver la solution de ce problème d'optimisation utilisant les multiplicateurs de Lagrange, il est commode de le reformuler comme suit :

$$\min_{\boldsymbol{\beta}, \beta_0} \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \gamma \sum_{i=1}^l \xi_i, \quad (\text{B.11})$$

sous les contraintes :

$$\xi_i \geq 0, \quad y_i (\mathbf{x}_i^t \cdot \boldsymbol{\beta} + \beta_0) \geq 1 - \xi_i, \quad \forall i.$$

γ remplace la constante dans B.10.

On montre que l'équation de l'hyperplan B.1 peut aussi se mettre sous la même forme :

$$\{\mathbf{x} : h(\mathbf{x}) = \sum_i \alpha_i y_i \mathbf{x}^t \mathbf{x}_i + \beta_0 = 0\}. \quad (\text{B.12})$$

avec α_i une contrainte positive à valeur calculée fonction de γ . Le seul paramètre de réglage de cette procédure est γ .

B.2 Détermination d'une frontière non linéaire

La procédure décrite précédemment trouve des frontières linéaires dans l'espace des données \mathbf{x}_i . Pour trouver des frontières non linéaires, l'idée est de transformer les données de l'espace d'origine à un espace de dimension plus grande en utilisant une transformation non linéaire ϕ :

$$\mathbf{x} \rightarrow \phi(\mathbf{x}).$$

De B.12, l'équation de l'hyperplan peut être écrite :

$$\begin{aligned} h(\mathbf{x}) &= \phi(\mathbf{x})^t \boldsymbol{\beta} + \beta_0 \\ &= \sum_i \alpha_i y_i \langle \phi(\mathbf{x}), \phi(\mathbf{x}_i) \rangle + \beta_0. \end{aligned} \quad (\text{B.13})$$

Le problème est donc de ne pas trouver la transformation $\phi(\mathbf{x})$ mais de connaître la fonction Kernel \mathcal{K} appelée noyau :

$$\mathcal{K} = \langle \phi(\mathbf{x}), \phi(\mathbf{x}_i) \rangle = \phi(\mathbf{x})^t \phi(\mathbf{x}_i), \quad (\text{B.14})$$

qui calcule le produit scalaire dans l'espace transformé. Le noyau \mathcal{K} doit être une fonction symétrique (semi) définie positive. Le choix populaire de \mathcal{K} est :

1. *gaussien* (en anglais Radial Basis function RBF) : $\mathcal{K}(\mathbf{x}, \mathbf{x}_i) = e^{-\|\mathbf{x} - \mathbf{x}_i\|^2/2\sigma^2}$;
2. *polynomial* de degré d : $\mathcal{K}(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}^t \cdot \mathbf{x}_i + 1)^d$;
3. *sigmoïde* $\mathcal{K}(\mathbf{x}, \mathbf{x}_i) = \tanh(\mathbf{x}^t \cdot \mathbf{x}_i + \theta)$;

Finalement l'équation de la frontière non linéaire obtenue est :

$$h(\mathbf{x}) = \sum_{i=1}^l \alpha_i y_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i) + \beta_0, \quad (\text{B.15})$$

et la règle de classification est :

$$J(\mathbf{x}) = \text{signe} \left(\sum_{i=1}^l \alpha_i y_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i) + \beta_0 \right). \quad (\text{B.16})$$

Notons que si l'on omet l'opérateur *signe* de la fonction de décision, nous obtenons une mesure d'appartenance à la classe recherchée.

Annexe C

Théorie des courbes de niveaux pour la détection des contours

Nous présentons dans cet annexe la théorie des courbes de niveaux appliquée dans le chapitre 6 afin de segmenter des images ostéo-articulaires.

C.1 Evolution d'une courbe dans le plan

Soient $\gamma(0)$ une courbe lisse, fermée dans le plan Euclidien \mathbb{R}^2 et $\gamma(t)$ la famille de courbes avec un seul paramètre : le temps $t \in [0, +\infty[$. $\gamma(t)$ est obtenue par la propagation de $\gamma(0)$ suivant le champ des vecteurs normaux avec une vitesse F fonction scalaire de la courbure κ .

C.1.1 Formulation Lagrangienne

La formulation est dite "lagrangienne" quand le système de coordonnées suit le déplacement du front. Une approche naturelle consiste à fournir une expression paramétrique de l'évolution de la courbe.

Soit $\vec{X}(s,t) = [x(s,t), y(s,t)]$ le vecteur correspondant à un point M de la courbe paramétrée par s tel que : $0 \leq s \leq S$ et $\vec{X}(0,t) = \vec{X}(S,t)$. Soient \vec{N} la normale, dirigée vers l'extérieur, et $\kappa(s,t)$ la courbure de $\gamma(t)$ en $\vec{X}(s,t)$. La courbe est paramétrée de telle sorte que son intérieur soit à gauche quand nous nous déplaçons suivant les s croissants (voir figure C.1).

L'équation d'évolution s'écrit :

$$\frac{\partial \vec{X}}{\partial t} = \vec{N} \cdot F(\kappa), \quad (\text{C.1})$$

avec \vec{N} donné par :

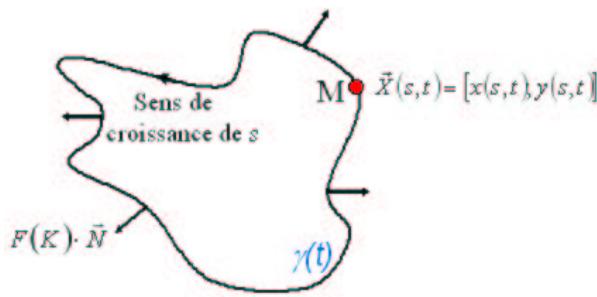


FIG. C.1 – Propagation d'une courbe avec une vitesse F de direction suivant la normale.

$$\begin{aligned}\vec{N} &= \left(\frac{\partial y}{\partial s}, -\frac{\partial x}{\partial s} \right) \times \frac{1}{\left(\left(\frac{\partial y}{\partial s} \right)^2 + \left(\frac{\partial x}{\partial s} \right)^2 \right)^{1/2}} \\ &= (y_s, -x_s) \times \frac{1}{(y_s^2 + x_s^2)^{1/2}}.\end{aligned}\quad (\text{C.2})$$

Par définition, la courbure κ vaut :

$$\kappa = \frac{y_{ss}x_s - x_{ss}y_s}{(x_s^2 + y_s^2)^{3/2}}. \quad (\text{C.3})$$

L'équation d'évolution devient alors :

$$\begin{cases} x_t = \frac{y_s}{(x_s^2 + y_s^2)^{1/2}} \cdot F(\kappa), \\ y_t = -\frac{x_s}{(x_s^2 + y_s^2)^{1/2}} \cdot F(\kappa). \end{cases} \quad (\text{C.4})$$

Ou encore :

$$\begin{cases} \frac{\partial \vec{X}}{\partial t} = \vec{N} \cdot F(\kappa), & \text{avec } t \in [0, +\infty[\text{ et } s \in [0, S[.} \\ \vec{X}(s, 0) = \gamma(0), \end{cases} \quad (\text{C.5})$$

Toute résolution numérique pose le problème de la précision d'approximation de la solution et la stabilité c'est à dire la sensibilité de l'approximation aux petits déplacements.

Sethian (1990) a montré que la résolution numérique de la formulation Lagrangienne exigeait de très petites tailles (prohibitives) de pas d'incrémentation temporelle. En particulier, ce défaut apparaît lors de l'approximation des dérivées par des différences finies centrées. Ce problème est résolu par une formulation "eulérienne" de l'évolution (car exprimée dans un repère fixe) introduite par Osher et Sethian (1988) et présentée dans le paragraphe suivant.

C.2 Théorie des courbes de niveaux "Level set"

C.2.1 Définitions

Cette approche a été formalisée par Osher et Sethian (1988). Elle consiste à suivre l'évolution de la courbe de niveau fermée $\gamma(0)$ située dans le plan \mathbb{R}^2 et correspondant à la courbe de niveau $\{\psi = 0\}$ de la surface $z = \psi(x,y,t)$ à l'instant initial :

$$\gamma(t=0) \equiv (\vec{X} | \psi(\vec{X}, t=0) = 0),$$

A l'instant initial la fonction $\psi(\vec{X}, t=0)$ est construite de la manière suivante :

$$\psi(\vec{X}, t=0) = \pm \text{distance de } \vec{X} \text{ à } \gamma(0), \quad (\text{C.6})$$

où le signe moins (plus) est choisi si \vec{X} est à l'intérieur (à l'extérieur) de $\gamma(t)$.

L'ensemble de niveau $\{\psi = 0\}$ de la surface $\psi(x,y,t)$ en évolution correspond au front $\gamma(t)$ en mouvement comme illustré dans la figure C.2. En effet, nous faisons correspondre la famille de courbes en mouvement $\gamma(t)$ avec la famille de surfaces en mouvement $\psi(x,y,t)$ de telle sorte que $\{\psi = 0\}$ fournit toujours le front en mouvement dans le plan \mathbb{R}^2 .

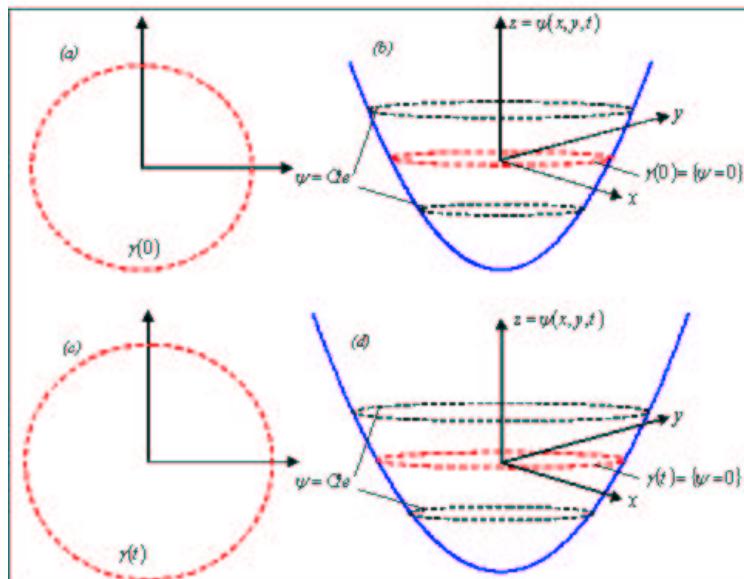


FIG. C.2 – (a) et (b) présentent le front γ et la surface $\psi(x,y)$ à l'instant $t = 0$. (c) et (d) présentent le front γ et la surface $\psi(x,y)$ correspondante à l'instant t .

C.2.2 Équation d'évolution

Quand $\psi(x,y,t)$ coïncide avec $\gamma(t)$ nous avons :

$$\psi(\vec{X},t) = 0 \Rightarrow \frac{\partial\psi}{\partial t} + \frac{\partial\psi}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial\psi}{\partial y} \cdot \frac{\partial y}{\partial t} = 0. \quad (\text{C.7})$$

Or, nous avons d'après l'équation C.1 :

$$\left(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t} \right) = \vec{N} \cdot F(\kappa). \quad (\text{C.8})$$

En incorporant l'équation C.8 dans C.7 nous obtenons :

$$\frac{\partial\psi}{\partial t} + \langle \vec{\nabla}\psi \cdot \vec{N} \cdot F(\kappa) \rangle = 0, \quad (\text{C.9})$$

soit

$$\psi_t + \langle \vec{\nabla}\psi \cdot \vec{N} \rangle \cdot F(\kappa) = 0. \quad (\text{C.10})$$

Le vecteur **gradient** $\vec{\nabla}\psi = \left(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t} \right)$ en un point M est **normal à la courbe de niveau** donnée par $\psi(x,y) = \text{constante}$ qui passe par M . En supposant que le vecteur normal est dirigé vers l'extérieur et en tenant compte de l'initialisation de $\psi(x,y,t)$ nous obtenons :

$$\vec{N} = \frac{\vec{\nabla}\psi}{|\vec{\nabla}\psi|}. \quad (\text{C.11})$$

L'équation C.10 devient alors :

$$\psi_t + |\vec{\nabla}\psi| \cdot F(\kappa) = 0. \quad (\text{C.12})$$

L'équation d'évolution est donc la suivante :

$$\begin{cases} \psi_t + |\vec{\nabla}\psi| \cdot F(\kappa) = 0, \\ \psi(\vec{X},0) \text{ donnée.} \end{cases} \quad (\text{C.13})$$

Par ailleurs, la courbure κ peut être exprimée en fonction des dérivées de ψ :

$$\kappa = \frac{\psi_{xx}\psi_y^2 - 2\psi_{xy}\psi_x\psi_y + \psi_{yy}\psi_x^2}{(\psi_x^2 + \psi_y^2)^{3/2}}. \quad (\text{C.14})$$

En outre, nous pouvons démontrer que :

$$\kappa = \text{div} \left(\frac{\vec{\nabla}\psi}{|\vec{\nabla}\psi|} \right). \quad (\text{C.15})$$

C.2.3 Avantages de la méthode

1. $z = \psi(x,y,t)$ reste toujours une fonction qui représente une surface pendant son évolution, la courbe de niveau $\{\psi = 0\}$ peut ne pas être simplement connexe et le front $\gamma(t)$ peut changer de topologie. La figure C.3 illustre cette caractéristique.
2. L'évolution du système est exprimée dans un repère fixe, en conséquence la grille de discrétisation est constante et l'instabilité de la formulation Lagrangienne est éliminée.
3. L'équation d'évolution C.13 obtenue appartient à la classe des équations d'Hamilton-Jacobi. Il est alors possible de trouver des approximations du gradient qui respectent une loi de conservation hyperbolique garantissant une solution calculée "douce" au voisinage des points anguleux (non différentiables) qui pourrait présenter la solution de l'équation d'évolution.
4. Les propriétés géométriques intrinsèques du front sont facilement calculées en fonction de $\psi(x,y,t)$. Par exemple, le vecteur normal et la courbure sont calculés à partir des équations C.15 et C.11.
5. La formulation Eulérienne est généralisable à N dimensions.

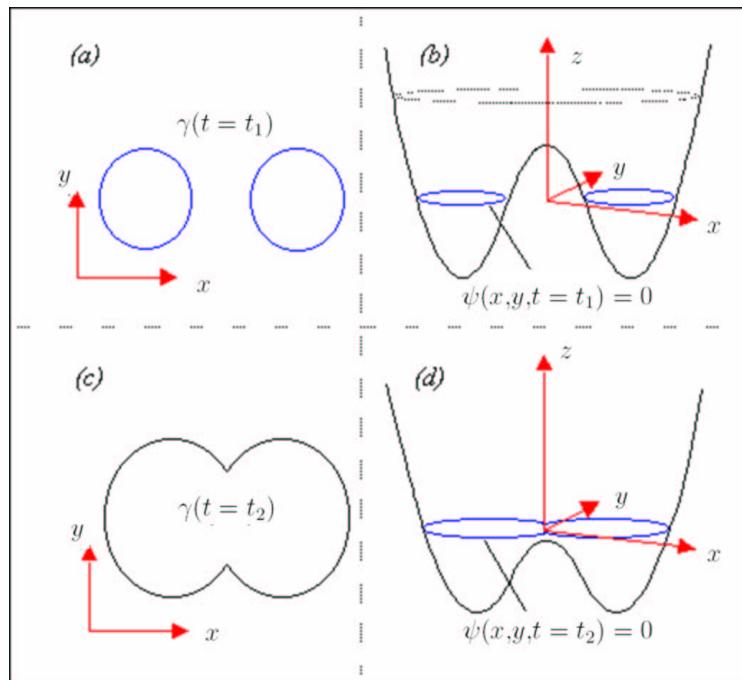


FIG. C.3 – *Changement de topologie du front $\gamma(t)$ en mouvement. Le contour (a) correspond à la courbe $\{\psi = 0\}$ de la surface $\{\psi(t = t_1)\}$ en (b). A ($t = t_2$) la surface en (d) a évolué et les contours en (c) ont changé de topologie.*

C.3 Schémas numériques

Le problème essentiel dans la détermination des schémas numériques réside donc au niveau de l'estimation des dérivées spatiales aux points anguleux où la fonction est continue mais non dérivable.

Osher et Sethian (1988) propose donc un schéma numérique qui estime les dérivées spatiales de telle sorte que la solution calculée :

1. ne fasse pas apparaître des points aigus artificiels,
2. respecte une loi de conservation hyperbolique.

C.3.1 Equation d'Hamilton-Jacobi

Proposons $F(\kappa) = 1$ dans l'équation C.13 qui revêt la forme standard d'une équation d'Hamilton-Jacobi :

$$\begin{cases} \psi_t + H(D\psi) = 0, \\ \psi(x,y,0) = \psi_0(x,y), \end{cases} \quad (\text{C.16})$$

avec :

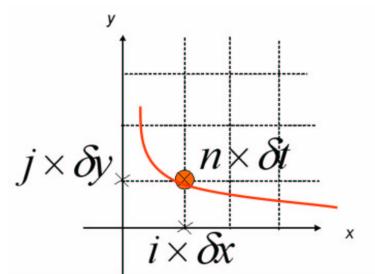
$$D\psi = (\psi_x, \psi_y) \text{ et } H(D\psi) = -(\psi_x^2 + \psi_y^2)^{1/2}.$$

C.3.2 Schéma monodimensionnel

En posant $u = \psi_x$, exprimant l'équation précédente C.16 en monodimensionnel et en dérivant par rapport à x , nous obtenons :

$$\begin{cases} u_t + [H(u)]_x = 0, \quad \text{où } H(u) = -(u^2)^{1/2}. \\ u(x,0) \text{ donnée.} \end{cases} \quad (\text{C.17})$$

Pour numériser l'équation précédente une définition est prise pour satisfaire la loi de conservation hyperbolique.



Définition : Soit u_i^n la valeur de u au point $i\delta x$ et au temps $n\delta t$. Les 3 points u_{i-1}^n , u_i^n et u_{i+1}^n respectent la loi de conservation hyperbolique s'il existe une fonction $g(u_1, u_2)$ telle que :

$$\begin{cases} \frac{u_i^{n+1}\delta x - u_i^n\delta x}{\delta t} = g(u_{i-1}^n, u_i^n) - g(u_i^n, u_{i+1}^n), \\ \text{où } g(u, u) = H(u). \end{cases} \quad (\text{C.18})$$

L'équation C.16 peut être écrite après discréétisation du temps :

$$\psi_i^{n+1} = \psi_i^n - \delta t \cdot H(u), \quad (\text{C.19})$$

ou encore :

$$\psi_i^{n+1} = \psi_i^n - \delta t \cdot g(u_{i-1/2}, u_{i+1/2}) = \psi_i^n - \delta t \cdot g(D_x^- \psi_i, D_x^+ \psi_i). \quad (\text{C.20})$$

Avec :

$$\begin{cases} D_x^- \psi_i = \frac{\psi_i^n - \psi_{i-1}^n}{\delta x}, \\ D_x^+ \psi_i = \frac{\psi_{i+1}^n - \psi_i^n}{\delta x}. \end{cases} \quad (\text{C.21})$$

Finalement, puisque $H(u) = f(u^2)$ et qui vaut dans notre cas $-(u^2)^{1/2}$, nous pouvons utiliser la fonction d'Hamilton-Jacobien donnée dans (Osher et Sethian, 1988) :

$$g(u_{i-1/2}, u_{i+1/2}) = g(D_x^- \psi_i, D_x^+ \psi_i) = f((\max(D_x^- \psi_i, 0))^2 + (\min(D_x^+ \psi_i, 0))^2), \quad (\text{C.22})$$

et l'équation C.19 devient :

$$\psi_i^{n+1} = \psi_i^n - \delta t \cdot ((\max(D_x^- \psi_i, 0))^2 + (\min(D_x^+ \psi_i, 0))^2)^{1/2}. \quad (\text{C.23})$$

C.3.3 Schéma bidimensionnel

La vitesse F peut être décomposée en deux composantes l'une constante et l'autre liée aux propriétés géométriques du front, par exemple la courbure ou la normale, soit alors :

$$F = F_A + F_G(\kappa). \quad (\text{C.24})$$

L'équation C.12 devient :

$$\psi_t = -(F_A |\nabla \psi| + F_G(\kappa) |\nabla \psi|). \quad (\text{C.25})$$

Dans le cas bidimensionnel l'équation C.23 peut être étendue en différenciant dans les deux directions pour obtenir :

$$\psi_{ij}^{n+1} = \psi_{ij}^n - \delta t F_A \left(\begin{array}{c} (\max(D_x^- \psi_{ij}, 0))^2 + (\min(D_x^+ \psi_{ij}, 0))^2 \\ + (\max(D_y^- \psi_{ij}, 0))^2 + (\min(D_y^+ \psi_{ij}, 0))^2 \end{array} \right)^{1/2} - \delta t F_G(\kappa) |\nabla \psi|. \quad (\text{C.26})$$

$$\left\{ \begin{array}{l} D_x^- \psi_{ij} = \frac{\psi_{ij}^n - \psi_{i-1,j}^n}{\delta x}, \\ D_x^+ \psi_{ij} = \frac{\psi_{i+1,j}^n - \psi_{ij}^n}{\delta x}, \\ D_y^- \psi_{ij} = \frac{\psi_{ij}^n - \psi_{i,j-1}^n}{\delta y}, \\ D_y^+ \psi_{ij} = \frac{\psi_{i,j+1}^n - \psi_{ij}^n}{\delta y}. \end{array} \right. \quad (\text{C.27})$$

Remarque : Les dérivées du terme $\delta t F_G(\kappa) |\nabla \psi|$ sont approchées par différences centrales. Les approximations des dérivées partielles de ψ doivent être les mêmes dans les calculs de κ et de $|\nabla \psi|$, sinon il y a des risques d'erreurs importantes, comme cela est justifié dans (Osher et Sethian, 1988).

C.4 Front d'onde pour obtenir les contours d'une image

Puisque la propagation du front doit s'arrêter sur les contours des objets dans une image, Malladi et al. (1995) ont proposé de multiplier la vitesse $(F_A + F_G)$ par un coefficient k_i lié à l'image :

$$k_i(x,y) = \frac{1}{1 + |\nabla \hat{I}|}, \quad (\text{C.28})$$

où \hat{I} est l'image originale I régularisée par un opérateur Gaussien G_σ et $|\nabla \hat{I}|$ est la norme du gradient sur \hat{I} :

$$|\nabla \hat{I}| = |\nabla(G_\sigma * I)|. \quad (\text{C.29})$$

Le coefficient k_i doit avoir les propriétés suivantes :

1. faible dans les zones de fort gradient pour fixer une vitesse nulle,
2. proche de 1 ($k_i \approx 1$) dans les zones de faible gradient, pour que les caractéristiques géométriques l'emportent et par conséquence faire évoluer le front.

Si nous souhaitons avoir un coefficient k_i de "décroissance" plus rapide que la fonction réciproque C.30, nous pouvons employer :

$$k_i(x,y) = e^{-|\nabla \hat{I}|}. \quad (\text{C.30})$$

C.4.1 Extension de k_i

D'après sa définition, k_i n'a pas de sens que sur le front de propagation $\gamma(t)$ donc sur l'ensemble $\{\psi = 0\}$.

En remarquant que :

- l'équation d'évolution s'applique sur la fonction ψ pour le domaine entier,
- $\gamma(t)$ et l'ensemble $\{\psi = 0\}$ sont dans le plan de l'image,

– et la fonction k_i n'a de sens que dans ce plan, il convient donc d'étendre k_i à tous les ensembles de niveaux $\{\psi = \text{constante}\}$ en le nommant \hat{k}_i tel que $\hat{k}_i = k_i$ en $\{\psi = 0\}$ comme illustré dans la figure C.4. L'équation d'évolution devient :

$$\psi_t + \hat{k}_i(F_A|\nabla\psi| + F_G(\kappa)|\nabla\psi|) = 0. \quad (\text{C.31})$$

C.4.2 Elaboration de l'extension de k_i

Il est nécessaire que l'équation d'évolution garde son sens physique pour tous les ensembles de niveaux. La fonction vitesse ne tient pas son sens physique de la géométrie de ψ , mais de la configuration de l'ensemble de niveau $\{\psi = 0\}$ dans le plan de l'image. Le problème est donc de définir une fonction vitesse qui soit définie sur tout le domaine. Ceci intervient donc dans la définition du coefficient k_i . Malladi et al. (1995) ont défini \hat{k}_i (extension de k_i) qui étend la signification physique de l'ensemble de niveau $\{\psi = 0\}$ aux autres ensembles de niveaux.

Pour cela, ils ont considéré que \hat{k}_i en un point $P \in \{\psi = C\}$ prend la valeur du point $Q \in \{\psi = 0\}$ le plus proche de P (voir la figure C.4).

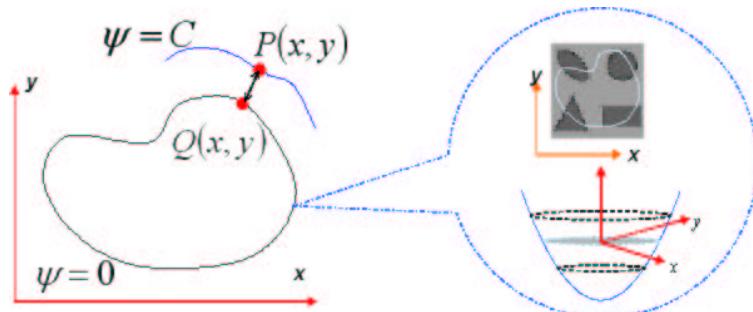


FIG. C.4 – Extension de vitesse de la courbe de niveau $\{\psi = 0\}$ dans le plan de l'image à une autre courbe de niveau $\{\psi = C\}$.

Cette extension se fait au fur et à mesure de l'évolution de la courbe. Il est inutile de construire explicitement \hat{k}_i .

C.5 Algorithmes de résolution de l'équation d'évolution

Une implémentation numérique directe de l'équation C.26 implique l'estimation de la fonction des courbes de niveaux ψ sur tous les pixels de l'image. Cette implémentation est coûteuse vu les opérations de calcul inutiles faites aux pixels loins du front de propagation. Trois méthodes ont été développées dans ce sens :

1. La méthode de bande étroite "narrow band" (Adalsteinsson et Sethian, 1995) où le domaine d'évolution de la fonction des courbes de niveaux est une bande étroite autour du front de propagation.

2. La méthode "Fast Marching" (Sethian, 1996) qui construit la fonction ψ ou en d'autres termes la bande étroite pas à pas.
3. La méthode "Hermes" (Paragios et Deriche, 1998) dont le principe est à la base de la combinaison des deux méthodes précédentes.

C.5.1 La méthode Narrow Band

Le front est mis à jour en actualisant la fonction des courbes de niveaux ψ pour un ensemble de points situés dans le voisinage réduit de l'ensemble de niveau $\{\psi = 0\}$. Ce voisinage est limité par deux courbes qui sont les ensembles de niveaux $\{\psi = \pm \frac{\delta}{2}\}$ (voir figure C.5). Nous effectuons donc une mise à jour des courbes des niveaux dans une bande étroite de largeur δ . Ceci est possible du fait que la résolution de la formulation Eulérienne de l'équation d'évolution en un point d'une courbe de niveau n'a besoin que des pixels voisins.

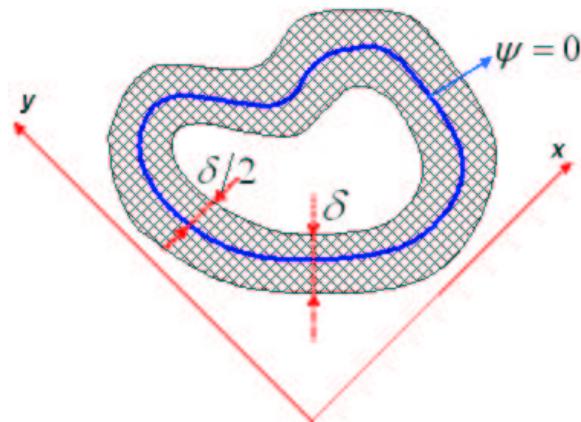


FIG. C.5 – Bande étroite de largeur δ autour de la courbe de niveau $\{\psi = 0\}$.

Le problème de cette méthode est que la position du front change d'une itération à une autre. Par suite la bande étroite doit être donc mise à jour à chaque itération. Ce qui est coûteux et complexe en temps de calcul. Pour palier cet inconvénient, la mise à jour de la bande n'est faite que dans le cas où le front s'approche de l'un des bords intérieur ou extérieur de la bande. D'où la nécessité d'un algorithme pour détecter la distance du front aux bords de la bande qui doit être inférieur à une valeur fixée à priori.

Par ailleurs, le calcul des dérivées aux frontières de la bande nécessite la connaissance des valeurs de la fonction des courbes de niveaux hors de la bande. Deux solutions sont proposées dans (Adalsteinsson et Sethian, 1995).

Il est important aussi de noter que la fonction ψ a besoin d'être ré-initialisée de temps à autre pour éviter tout cas d'instabilité développée dans la bande.

En résumé, cette méthode réduit le coût de calcul comparé à une autre implémentation directe, cependant son coût et sa complexité restent importants.

C.5.2 La méthode Fast Marching

Cette méthode a été introduite par Sethian (1996) pour réduire le coût de calcul de la méthode précédente. Elle n'est applicable que dans le cas où la vitesse du front ne change pas de signe pendant sa propagation. Elle ne peut être donc utilisée quand la vitesse est une fonction de la courbure, cas de larges applications en vision par ordinateur.

C.5.3 La méthode Hermes

Cette méthode est introduite par Paragios et Deriche (1998) pour la détection d'un objet en mouvement et la segmentation d'une image par la méthode des régions actives géodésiques. Elle combine les deux principes des méthodes précédentes en utilisant l'idée de la propagation sélective (Fast Marching) sur une bande relativement petite (Narrow Band).

L'équation aux dérivées partielles (EDP) de propagation du front peut être réécrite comme suit :

$$\psi^{t+1}(x,y) = \psi^t(x,y) + v(x,y,\psi^t). \delta t. \quad (\text{C.32})$$

Cette EDP montre que le front se propage suivant les caractéristiques de l'image et les propriétés géométriques de chaque pixel. La méthode Narrow Band met à jour la fonction des courbes de niveaux dans une bande étroite contenant des pixels qui ne bougent pas pendant une itération. Pour éviter tout calcul inutile, l'idée était d'évoluer le front localement suivant les vitesses de propagation associées. En effet, à chaque pas d'itération, le pixel qui a la plus grande vitesse absolue est choisi pour évoluer localement la fonction dans une fenêtre centrée sur ce pixel. La méthode est décrite par l'algorithme qui suit :

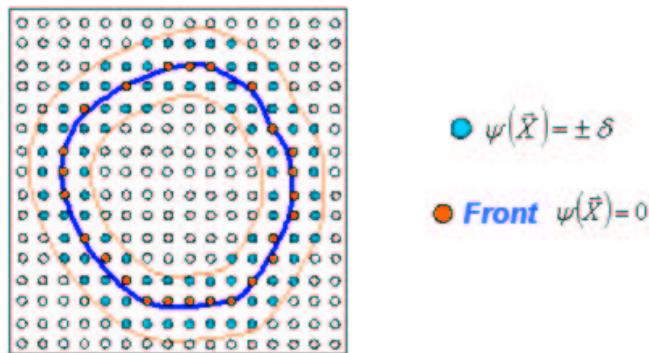


FIG. C.6 – Initialisation de l'algorithme Hermes : la largeur de la bande étroite est égale à celle de la fenêtre. les pixels formant le front constituent la liste des pixels Actifs.

A- Initialisation

La fonction des courbes de niveaux ψ est initialisée, suivant l'équation C.6, comme étant la distance algébrique au front initial. La valeur absolue maximale de la distance

est le rayon de la fenêtre (choisi *a priori* pour l'évolution locale). Ainsi les pixels formant le front initial constituent la liste Active. Les vitesses initiales de ces pixels sont alors calculées et sauvegardées.

B- Propagation du front

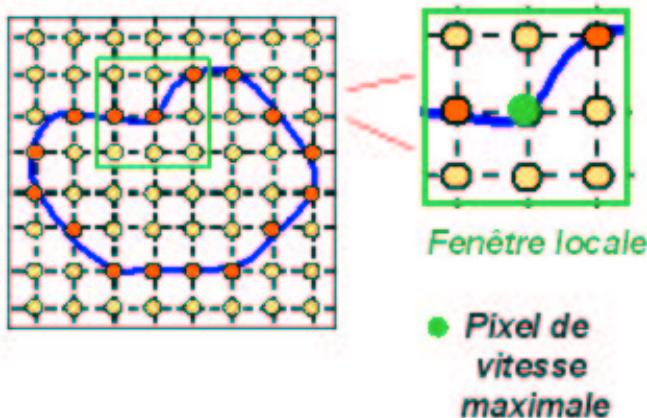


FIG. C.7 – L'étape "Propagation" de l'algorithme Hermes : évolution du front dans une fenêtre centrée sur le pixel qui a la plus grande vitesse absolue.

Pour un nombre d'itérations donné :

1. Sélectionner le pixel $[c]$ de la liste Active ayant une vitesse maximale en valeur absolue.
2. Mettre à jour la fonction ψ dans une fenêtre de rayon donné et de centre le pixel $[c]$. A chaque pixel $[s]$ de la fenêtre, est associé un pas de temps qui lui correspond :

$$\delta t(s) = \frac{\delta t}{1 + \|s - c\|}. \quad (\text{C.33})$$

Cette opération ne perturbe pas l'évolution de l'ensemble du front car si le front bouge, cela arrivera au pixel de plus grande vitesse $[c]$.

Les itérations locales sont suspendues si le front bouge.

- Si un pixel appartient au nouveau front et non pas à l'ancien, il est ajouté à la liste Active.
- Si le pixel n'appartient qu'à l'ancien front, il est éliminé de la liste Active.
- Si le pixel appartient à la fois à l'ancien et au nouveau front, il est donc affecté en terme de vitesse de propagation par le changement local de la fonction des courbes de niveaux. Une mise à jour de sa vitesse est nécessaire.

Les vitesses de propagation des pixels dans la fenêtre sont calculées au début de l'évolution locale et ne sont pas affectées par le changement de la fonction ψ .

3. Lorsque l'évolution locale est achevée, extraire le front en trouvant le passage par zéro de la fonction ψ dans la fenêtre.
4. Ré-initialiser localement la fonction des courbes de niveaux.
5. Le nouveau front local est comparé à l'ancien avant l'évolution. Trois cas sont à considérer pour les pixels formant la fenêtre :
6. Si l'évolution est achevée sortir de la boucle, sinon retourner à l'étape numéro 1.

C.6 Résultats

Nous illustrons deux résultats de détection de contour par la théorie des courbes des niveaux présentée dans cet annexe. L'implémentation de l'équation d'évolution du front est celle de l'algorithme Hermes. Les résultats de la figure C.8 montre le changement de topologie du contour en propagation. La figure C.9 affirme que la forme du contour final est indépendante de celle du contour de départ.

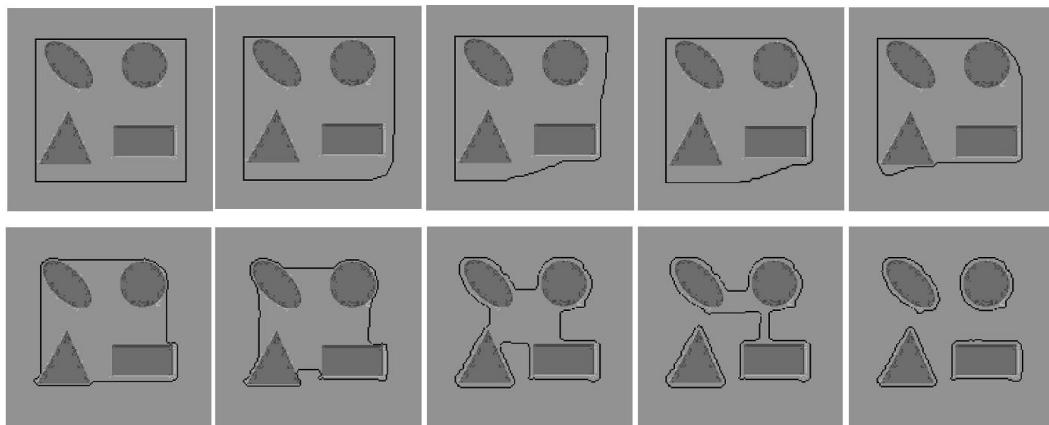


FIG. C.8 – *Propagation du Front pour la détection des contours (de gauche à droite, de haut vers bas) en appliquant la théorie Level Set et l'algorithme Hermes (le changement de topologie est illustré clairement).*

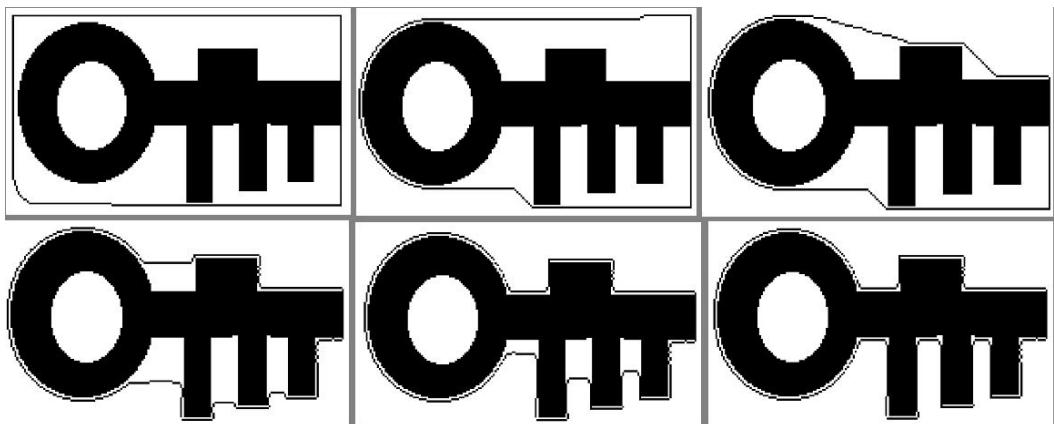


FIG. C.9 – Propagation du Front pour la détection des contours (de gauche à droite, de haut vers bas) en appliquant la théorie Level Set et l'algorithme Hermes (la forme de la courbe finale est indépendante de celle de l'initiale).

Bibliographie

- Adalsteinsson, D. et J. Sethian (1995). “A fast level set method for propagating interfaces.” *J. Comput. Phys.* 118, 269–277.
- Agrawal, R., J. Gehrke, D. Gunopoulos, et P. Raghavan (1998). “Automatic subspace clustering of high dimensional data for data mining applications.” In *ACM SIGMOD International Conference on Management of Data*, pp. 94–105.
- Aksoy, S. et R. Haralick (1998, June). “Textural features for image database retrieval.” In *IEEE Workshop on Content-Based Access of Image and Video Libraries in conjunction with CVPR’98*, Santa Barbara, CA, pp. 45–49.
- Amar, A., N.T. Labzour, et A. Bensaid (1997, August). “Semi-supervised hierarchical clustering algorithms.” In *Sixth Scandinavian Conference On Artificial Intelligence SCAI’97*, Helsinki, Finland, pp. 232–239.
- Ambroise, C. (1996). *Approche probabiliste en classification automatique et contraintes de voisinage*. Ph.D. thesis, Université de Technologie de Compiègne.
- Ambroise, C. et G. Govaert (2000). “EM algorithm for partially known labels.” In *IFCS2000, 7th Conference of the International Federation of Classification Societies*, Namur, Belgique, pp. 11–14.
- Bas, P. (2000, 5 octobre). *Méthodes de tatouages d’images fondées sur le contenu*. Ph.D. thesis, Institut National Polytechnique de Grenoble, France.
- Bennett, K.P. et A. Demiriz (1998). “Semi-supervised support vector machines.” In *Advances in Neural Information Processing Systems (NIPS)*, pp. 368–374.
- Bensaid, A. et J.C. Bezdek (1998). “Semi-supervised point prototype clustering.” *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)* 12(4), 625–643.
- Bensaid, A. M., L. O. Hall, J.C. Bezdek, et L.P. Clarke (1996). “Partially supervised clustering for image segmentation.” *Pattern Recognition* 29(5), 859–871,.
- Bloch, I. et H. Maître (1994). “Fusion de données en traitement d’images: modèles d’information et décisions.” *Traitemet du Signal* 11(6), 435–446.
- Boutry, N., A. Lardé, E. Solau, R. Flipo, et A. Cotten (2001, November). “The usefulness of MR imaging in early rheumatoid arthritis.” In *RSNA*, Chicago, pp. 25–30.
- Boutry, N., A. Larde, F. Lapegue, E. Solau-Gervais, RM. Flipo, et A. Cotten (2003, April). “Magnetic resonance imaging appearance of the hands and feet in patients with early rheumatoid arthritis.” *Journal of Rheumatologie* 30(4), 671–679.

- Carson, C., M. Thomas, S. Belongie, J. Hellerstein, et J. Malik (1999, June 2-4). “Blobworld : a system for region-based image indexing and retrieval.” In D. P. Huijsmans et A. W. M. Smeulders, eds., *Third Int. Conf. on visual Information Systems*, The Netherlands. Springer, Amesterdam.
- Chakrabarti, K., K. Porkaew, et S. Mehrotra (2000, February). “Efficient query refinement in multimedia databases.” In *IEEE International conference on data engineering ICDE*, San Diego, California, pp. 196.
- Chen, Y., X.S. Zhou, et T.S Huang (2001, October). “One-class SVM for learning in image retrieval.” In *International Conference in image processing (ICIP'01)*, Thessaloniki, Greece, pp. 34–37.
- Chu, W. W., C. C. Hsu, A. F. Cardenas, et R. K. Taira (1998). “A knowledge-based image retrieval with spatial and temporal constructs.” *IEEE Transactions on Knowledge and Data Engineering* 10(6), 872–888.
- Cocquerez, J. P. et S. Philipp (1995). *Analyse d'images: filtre et segmentation*. Masson.
- Coggins, J. M. et A. K. Jain (1985). “A spatial filtering approach to texture analysis.” *Pattern Recognition Letters* 3, 195–203.
- Comaniciu, D., P.M. David, et J. Foran (1999). “Image guided decision support system for pathology.” *Machine Vision and Applications* 11(4), 213–224.
- Cordella, L.P., P. Foggia, C. Sansone, et M. Vento (1998). “Subgraph transformations for the inexact matching of attributed relational graphs.” *Computing* (12), 43–52.
- Coster, M. et J. M. Chermant (1985). *Précis d'analyse d'images* (CNRS ed.).
- Cotten, A. (2000). *Imagerie de la pathologie ostéo-articulaire: Sémiologie pratique*. Masson.
- Cox, I.J., M.L. Miller, T.P. Minka, T.V. Papathomas, et P.N. Yianilos (2000). “The bayesian image retrieval system, PicHunter: theory, implementation and psychophysical experiments.” *IEEE transactions on image processing* 9(1), 20–37.
- Cozman, F.G., I. Cohen, et M.C. Cirelo (2003, August). “Semi-supervised learning of mixture models and bayesian networks.” In *20th International Conference on Machine Learning ICML-2003*, Washington, USA.
- Cozman, F. G. et I. Cohen (2002). “Unlabeled data can degrade classification performance of generative classifiers.” In *Fifteenth International Florida Artificial Intelligence Society Conference*, Pensacola, Florida, pp. 327–331.
- Dai, X. et S. Khorram (1999, September). “A feature-based image registration algorithm using improved chain-code representation combined with invariant moments.” *IEEE transactions on geoscience and remote sensing* 37(5), 2351–2362.
- Dang, M. (1998). *Classification de données spatiales : modèles probabilistes et critères de partitionnement*. Ph.D. thesis, Université de Technologie de Compiègne.
- Dash, M., H. Liu, et H. Motoda (2000). “Consistency based feature selection.” In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 98–109.
- Day, N. E. (1969). “Estimating the components of a mixture of normal distributions.” *Biometrika* 53(3), 463–474.

- Demiriz, A. et K.P. Bennett (2001). *Optimization Approaches to Semi-supervised Learning*, Chapter 1, pp. 121–141. Boston: Kluwer Academic Publishers.
- Demiriz, A., K.P. Bennett, et M.J. Embrechts (1999, November). “Semi-supervised clustering using genetic algorithms.” In *Artificial Neural Networks in Engineering ANNIE’99*.
- Dempster, A., N. Laird, et D. Rubin (1977). “Maximum likelihood from incomplete data via the EM algorithm.” *Journal of the Royal Statistical Society* 39, 1–38.
- Derrode, S. (1999, 15 Decembre). *Représentation de formes planes à niveaux de gris par différentes approximations de Fourier-Mellin analytique en vue d’indexation de bases d’images*. Ph.D. thesis, Université de Rennes I, Rennes.
- Derrode, S., R. Mezhoud, et F. Ghorbel (1999, 13-17 Septembre). “Reconnaissance de formes par invariants complets et convergents - application à l’indexation de bases d’objets à niveaux de gris.” In *17ème colloque GRETSI’99*, Vannes, France.
- Devaney, M. et A. Ram (1997). “Efficient feature selection in conceptual clustering.” In *International Conference on Machine Learning*, pp. 92–97.
- Devijver, P. A. et J. Kittler (1982). *Pattern Recognition - A Statistical Approach*. London, GB: Prentice Hall.
- Dy, J. et C. E. Brodley (2000). “Feature subset selection and order identification for unsupervised learning.” In *Proceedings of the Seventeenth International Conference on Machine Learning*, San Francisco.
- Eshera, M.A. et K.S. Fu (1984). “A graph distance measure for image analysis.” *IEEE Trans. on Systems, Man and Cybernetics* 14(5), 398–407.
- Flipo, A. (2004, Février). “Implémentation C d’une approche statistique de la fouille d’images.” Rapport de stage, Université de Technologie de Compiègne.
- Flusser, J. et T. Suk (1994). “A moment-based approach to registration of images with affine geometric distortion.” *IEEE Transactions on Geoscience and Remote Sensing* 32(2), 382–387.
- Fournier, J. (2002). *Indexation d’images par le contenu et recherche interactive dans les bases généralistes*. Ph.D. thesis, Université de Cergy-Pointoise.
- Fournier, J. et M. Cord (2002, September). “Long-term similarity learning in content-based image retrieval.” In *ICIP*, Volume 1, pp. 441–444.
- Fournier, J., M. Cord, et S. Philipp-Foliguet (2001). “RETIN: A content-based image indexing and retrieval system.” *Pattern Analysis and Applications Journal (Special Issue on Image Indexation)* 4(2/3), 153–173.
- Fournier, J., M. Cord, et S. Philipp-Foliguet (2002, Janvier). “Stratégie interactive d’exploration pour la recherche d’images par le contenu.” In *RFIA’2002*, Angers, France, pp. 211–220.
- Freund, Y. et R.E. Schapire (1999, September). “A short introduction to boosting.” *Journal of Japanese society for artificial intelligence* 14(5), 771–780.
- Friedman, J. (1989). “Regularized discriminant analysis.” *Journal of American Statistical Association* 84(405), 165–175.

- Fukunaga, K. (1990). *Introduction to statistical pattern recognition* (second ed.). Academic press.
- Fung, G. et O. L. Mangasarian (2001). “Semi-supervised support vector machines for unlabeled data classification.” *Optimization Methods and Software* 15, 29–44,.
- Ganesalingam, S. (1989). “Classification and mixture approaches to clustering via maximum likelihood.” *Applied Statistics* 38(3), 455–466.
- Ganesalingam, S. et G. J. McLachlan (1978). “The efficiency of a linear discriminant function based on unclassified initial samples.” *Biometrika* 65, 658–662.
- Ganesalingam, S. et G. J. McLachlan (1979). “Small sample results for a linear discriminant function estimated from mixture of normal populations.” *J. Statist. Comput. Simul.* 9, 151–158.
- Güld, M., M. Kohnen, D. Keysers, H. Schubert, B. Wein, J. Bredno, et T. Lehmann (2002). “Quality of dicom header information for image categorization.” In *SPIE*, Volume 4684, pp. 1215–1223.
- Gosselin, P.H., M. Najjar, M. Cord, C. Ambroise, et S. Philipp-Foliguet (2004, Janvier). “Méthodes d’apprentissage pour la recherche d’images par le contenu.” In *Workshop RFIA04, Analyse de données, Statistique et Apprentissage pour la Fouille d’images*, Toulouse, France.
- Govaert, G., C. Ambroise, et T. Denoeux (2000, Mars). “Développement d’un logiciel temps réel pour le contrôle par émission acoustique des appareils à pression.” Technical Report 1.8.1333, CETIM.
- Grewe, L. et A.C. Kak (1995, May). “Interactive learning of a multiple-attribute hash table classifier for fast object recognition.” *Computer Vision and Image Understanding* 61(3), 387–416.
- Gros, P., G. Mclean, R. Delon, R. Mohr, C. Schmid, et G. Mistler (1997). “Utilisation de la couleur pour l’appariement et l’indexation d’images.” Rapport de recherche RR-3269, INRIA, Rhone-Alpes.
- Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. Ph.D. thesis, University of Waikato.
- Haralick, R., B. Shanmugam, et I. Dinstein (1973). “Texture features for image classification.” *IEEE Transactions on Systems, Man and Cybernetics*, 610–622.
- Hartley, H. O. et J. N. K. Rao (1968). “Classification and estimation in analysis of variance problems.” *Review of international statistical institute* 36, 141–147.
- Hastie, T. et R. Tibshirani (1996). “Discriminant analysis by gaussian mixtures.” *Journal of the Royal Statistical Society B* 58, 155–176.
- Hastie, T., R. Tibshirani, et J. Friedman (2001). *The elements of statistical learning: Data Mining, Inference, and Prediction*.
- Heinrichs, A., D. Kourbaroulis, B. Levienaise-Obadia, P. Rovida, et J.M. Jolion (2000, april). “Image indexation and content based search using pre-attentive similarities.” In *conference RIAO2000 (Content-based Multimedia Access)*, paris, France, pp. 1616–1631.

- Hong, P., Q. Tian, et T.S. Huang (2000, September). "Incorporate support vector machines to content based image retrieval with relevance feedback." In *IEEE international conference on image processing (ICIP'2000)*, Vancouver, Canada.
- Hu, M. (1962). "Visual pattern recognition by moment invariants." *IRE Trans. on Information Theory* IT-8, 179–187.
- Huang, H.K. et R.K. Taira (1992, April). "Infrastructure design of a picture archiving and communication system." *American Journal of Roetgenology* 158(4), 743–749.
- Huet, F. et S. Philipp (1998). "fusion of images interpreted by a multi-scale fuzzy classification." *pattern analysis and applications journal* 1, 231–247.
- Ishikawa, Y., R. Subramanya, et C. Faloutsos (1998). "MindReader: Querying databases through multiple examples." In *24th Conference on very large data bases (VLDB)*, New-York, pp. 218–227.
- Jiang, J., A. Armstrong, et G.C. Feng (2002). "Direct content access and extraction from JPEG compressed images." *Pattern Recognition* (35), 2511–2519.
- Keysers, D., J. Dahmen, H. Ney, B.B. Wein, et T.M. Lehmann (2003). "A statistical framework for model-based image retrieval in medical applications." *Journal of Electronic Imaging* 12(1), 59–68.
- Kira, K. et L. Rendell (1992). "A practical approach to feature selection." In *the 9th Int. Conf. on Machine Learning*, Aberdeen, Scotland, pp. 249–256. Morgan Kaufmann.
- Kirby, M. et L. Sirovich (1990, January). "Application of the karhunenloeve procedure for the characterization of human faces." *IEEE Trans. on Pattern Analysis and Machine Intelligence* 12(1), 103–107.
- Kittler, J. (1978). "Feature set search algorithms." in *Pattern Recognition and Signal Processing* (1), 41–60.
- Kohavi, R. et G.H. John (1997). "Wrappers for feature subset selection." *Artificial Intelligence*, 273–324.
- Kohonen, T. (2001). *Self-Organizing Maps* (third extended ed.), Volume 30. Springer Series in Information Sciences.
- Kudo, M. et J. Sklansky (2000, January). "Comparison of algorithms that select features for pattern classifiers." *Pattern Recognition* 33(1), 25–41.
- Laaksonen, J., M. Koskela, et E. Oja (1999, July). "PicSOM: Self-organizing maps for content-based image retrieval." In *IJCNN'99*, Washington, DC.
- Larsen, J., A. Szymkowiak, et L.K. Hansen (2002, January). "Probabilistic hierarchical clustering with labeled and unlabeled data." *International Journal of Knowledge-Based Intelligent Engineering Systems* 6(1), 56–62.
- Laws, K.I. (1980a). "Rapid texture identification." In *SPIE Image Processing for Missile Guidance*, pp. 376–380.
- Laws, K.I. (1980b, January). *Textured image segmentation*. Ph.D. thesis, Dept. Electrical Engineering, University of Southern California.
- Laws, K. I. (1979). "Texture energy measures." In *DARPA Image Understanding Workshop*, DARPA, Los Altos, CA, pp. 47–51.

- Lehmann, TM., BB. Wein BB, J. Dahmen, J. Bredno, F. Vogelsang, et M. Kohnen (2000). “Content-based image retrieval in medical applications: A novel multi-step approach.” In *SPIE*, Volume 3972, pp. 312–320.
- Little, R. J. A. (1977). “Discussion on the paper by professor dempster, professor laird and dr. rubin.” *Journal of the Royal Statistical Society* 39(1), 25.
- Liu, H. et R. Setiono (1996a). “Feature selection and classification: a probabilistic wrapper approach.” In *the 9th Int. Conf. on Industrial and Engineering Applications of AI and ES*, pp. 129–135. Morgan Kaufmann.
- Liu, H. et R. Setiono (1996b). “A probabilistic approach to feature selection: a filter solution.” In *the 13th Int. Conf. on Machine Learning*, pp. 319–327. Morgan Kaufmann.
- Malladi, R., J. A. Sethian, et B. C. Vemuri (1995, February). “Shape modeling with front propagation: A level set approach.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(2), 158–175.
- Mao, J. et A. K. Jain (1992). “Texture classification and segmentation using multiresolution simultaneous autoregressive models.” *Pattern Recognition* 25(2), 173–188.
- Maître, H. (2003). *Le traitement des images*. Paris: Hermès.
- McLachlan, G. et K. Basford (1988). *Mixture models*. New York: Marcel Dekker.
- McLachlan, G. et D. Peel (2000). *Finite mixture models*. New York: John Wiley and Sons.
- McLachlan, G. J. (1975). “Iterative reclassification procedure for constructing an asymptotically optimal rule of allocation in discriminant analysis.” *Journal of the American Statistical Association* 70(350), 365–369.
- McLachlan, G. J. et S. Ganesalingam (1982). “Updating a discriminant function on the basis of unclassified data.” *Communications in Statistics: Simulation and Computation* 11(6), 753–767.
- McLachlan, G. J. et T. Krishnan (1997). *The EM algorithm and extensions*. New York: John Wiley and Sons.
- Mehtre, B., M. Kankanhalli, et W. Lee (1997). “Shape measures for content based image retrieval: A comparaison.” *Information Processing and Management* 33(3), 319–337.
- Meilhac, C. et C. Nastar (1999, June). “Relevance feedback and category search in image database.” In *IEEE international conference on multimedia Computing and systems*, Florence, Italy.
- Miller, D. J. et H. Uyar (1997). “A mixture of experts classifier with learning based on both labelled and unlabelled data.” In *Neural Information Processing Systems NIPS’9*, pp. 571–577.
- Mitchell, T. M. (1982). “Generalization as search.” *Artificial Intelligence* 18(2), 203–226.
- Müller, N., N. Michoux, D. Bandon, et A. Geissbuhler (2004). “A review of content-based image retrieval systems in medical applications—clinical benefits and future directions.” *International Journal of Medical Informatics*.
- Molina, L. C., L. Belanche, et A. Nebot (2002, December). “Feature selection algorithms: A survey and experimental evaluation.” In *IEEE International Conference on Data Mining (ICDM’02)*, Maebashi City, Japan, pp. 306–313.

- Murase, H. et S. K. Nayar (1995). “Visual learning and recognition of 3d objects from appearance.” *International Journal of Computer Vision* 14(5).
- Murray, G. D. et D. M. Titterington (1978). “Estimation problems with data from a mixture.” *Applied Statistics* 27(3), 325–334.
- Nadler, M. et E. P. Smith (1992). *Pattern recognition engineering*. New york: John Wiley & sons.
- Nastar, C., M. Mitschke, et C. Meilhac (1998). “Efficient query refinement for image retrieval.” In *IEEE Conf. on Comp. Vis. and Pattern Recognition*, Santa Barbara, California, pp. 547–552.
- Niblack, W., R. Barber, W. Equitz, M. Flickner, E.H. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, et G. Taubin (1993). “The QBIC project: Query by image by content using color, texture, and shape. Volume 1908, Bellingham, WA, pp. 173–187.
- Nigam, K., A. McCallum, S. Thrun, et T. Mitchell (2000). “Classification from labeled and unlabeled documents using EM.” *Machine learning* 39(2/3), 135–167.
- Olabarriaga, S.D. et A.W.M. Smeulders (2001). “Interaction in the segmentation of medical images: A survey.” *Medical image analysis* 5, 127–142.
- O’Neill, T. (1978, December). “Normal discrimination with unclassified observations.” *Journal of the american statistical association* 73(364), 821–826.
- Osher, S. et J. Sethian (1988). “Fronts propagating with curvature-dependent speed : Algorithms based on hamilton-jacobi formulations.” *Journal of Computational physics* 79, 12–49.
- Paragios, N. et R. Deriche (1998). “A PDE-based level-set approach for detection and tracking of moving objects.” In *ICCV*, pp. 1139–1145.
- Parker, J. A., R. V. Kenyon, et D. E. Troxel (1983, March). “Comparison of interpolating methods for image resampling.” *IEEE Transactions on Medical Imaging* 2(1), 31–39.
- Pauwels, E.J. et G. Frederix. (1999). “Finding salient regions in images: Nonparametric clustering for image segmentation and grouping.” *Computer Vision and Image Understanding* 75(1,2), 73–85.
- Pedrycz, W. (1985, January). “Algorithms of fuzzy clustering with partial supervision.” *Pattern Recognition Letters* 3(1), 13–20.
- Peng, J., B. Bhanu, et S. Qing (1999, July-August). “Probabilistic feature relevance learning for content-based image retrieval.” *Computer Vision and Image Understanding* 75(1-2), 150–164.
- Persoon, E. et K. S. Fu (1977). “Shape discrimination using fourier descriptors.” *IEEE Transactions on System Management and Cybernetics* 7(3), 629–639.
- Pham, D. L., C. Xu, et J. Prince (1998). “A survey of current methods in medical image segmentation.” technical report JHU/ECE 99-01, The Johns Hopkins University, Baltimore.
- Philipp, S. et M. Lekkat (2003, oct). “Object retrieval in image databases using image composition.” In *SIBGRAPI03*, Sao Carlos, Brésil.
- Pratt, W. K. (1991). *Digital image processing* (second ed.). New york: Wiley.

- Randen, T. et J. H. Husoy (1999). "Filtering for texture classification: A comparative study." *IEEE Trans. PAMI* 21(4), 291–310.
- Rao, A. R. (1990). *A Taxonomy for Texture Description and Identification*. New York: Springer-Verlag.
- Rauber, T. W. et A. Steiger-Garçao (1992, 30.Aug.-3.Sept). "Shape description by unl fourier features - an application to handwritten character recognition." In *11th IAPR International Conference on Pattern Recognition*, The Hague, The Netherlands.
- Reed, T. R. et J. M. H. Du Bur (1993). "A review of recent texture segmentation and feature extraction techniques." *Computer Vision, Graphics, and Image Processing* 57(3), 359–372.
- Rui, Y. et T. Huang (2000, June). "Optimizing learning in image retrieval." In *Conference on Computer vision and Pattern Recognition (CVPR)*, Hilton Head Island, SC., pp. 236–243.
- Rui, Y., T.S. Huang, et S. Mehrotra (1997). "Content-based image retrieval with relevance feedback in Mars." In *IEEE International conference in image processing*, pp. 815–818.
- Rui, Y., T. S. Huang, S. Mehrotra, et M. Ortega (1997, june). "A relevance feedback architecture in content-based multimedia information retrieval systems." In *IEEE Workshop on Content-based Access of Image and Video Libraries*, Puerto Rico, pp. 82–89.
- Rui, Y., A. She, et T. Huang (1996). "Modified fourier descriptors for shape representation - a practical approach." In *Proc of First International Workshop on Image Databases and Multi Media Search*.
- Saint-Jean, C. et C. Frélicot (2001). "Une méthode paramétrique et robuste de classification semi-supervisée avec rejet." In *Plateforme AFIA, CAP 2001*, Grenoble, pp. 85–100.
- Sammon, J (1969). "A nonlinear mapping for data structure analysis." *IEEE Trans on Computers* C-18(5).
- Schmid, C. et R. Mohr (1997, May). "Local grayvalue invariants for image retrieval." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(5), 530–534.
- Scholkopf, B., J. C. Platt, J. Shawe-Taylor, A. Smola, et R. Williamson (1999). "Estimating the support of a high-dimensional distribution." Technical Report MSR-TR-99-87, Microsoft Research.
- Serra, J. P. (1982). *Image analysis and mathematical morphology*. Academic Press.
- Sethian, J. (1990). "Numerical algorithm for propagating interfaces: Hamilton-jacobi equations and conservation laws." *J. Differential. Geom.* 31, 131–161.
- Sethian, J. (1996). "A fast marching level set method for monotonically advancing fronts proc." In *Nat. Acad. Sci.*, Volume 93, pp. 1591–1595.
- Shahshahani, B. et D. Landgrebe (1994). "The effect of unlabeled samples in reducing the small sample size problem and mitigating the hughes phenomenon." *IEEE Transactions on Geoscience and Remote Sensing* 32(5), 1087–1095.
- Shapiro, L. et R. Haralick (1981, September). "Structural descriptions and inexact matching." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 3(5), 504–519.

- Shyu, C., C. Brodley, A. Kak, A. Kosaka, A. Aisen, et L. Broderick (1999, July/August). “ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases.” *Computer Vision and Image Understanding* 75(1/2), 111–132.
- Smeulders, A.W.M., M. Worring, S. Santini, A. Gupta, et R. Jain (2000, December). “Content-based image retrieval at the end of the early years.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12), 1349–1380.
- Sugimoto, H., A. Takeda, et K. Hyodoh (2000, August). “Early-stage rheumatoid arthritis: prospective study of the effectiveness of MR imaging for diagnosis.” *Radiology* 216(2), 569–575.
- Tagare, H., C. Jaffe, et J. Duncan (1997). “Medical image databases: a content-based retrieval approach.” *J. Am. Med. Inform Assoc* 4(3), 184–189.
- Talavera, L. (1999). “Feature selection as a preprocessing step for hierarchical clustering.” In *the 16th International Conf. on Machine Learning*, pp. 389–397. Morgan Kaufmann, San Francisco, CA.
- Tamura, H., S. Mori, et T. Yamawaki (1978, June). “Texture features corresponding to visual perception.” *IEEE Trans. Systems, Man and Cybernetics* 8(6), 460–473.
- Thies, C., A. Malik, D. Keysers, M. Kohnen, B. Fischer, et T. M. Lehmann (2003). “Hierarchical feature clustering for content-based retrieval in medical image databases.” In *SPIES*, Volume 5032, pp. 598–608.
- Tieu, K. et P. Viola (2000, June). “Boosting image retrieval.” In *IEEE Conference on computer vision and pattern recognition*, Hilton Head Island, pp. 228–235.
- Titterington, D. M. (1976). “Updating a diagnostic system using unconfirmed cases.” *Applied Statistics* 25(3), 238–247.
- Tréneau, A., C. Fernandez-Maloigne, et P. Bonton (2004). *Image numérique couleur : De l'acquisition au traitement*. Dunod.
- Turk, M. et A. Pentland (1991). “Eigenfaces for recognition.” *Journal of Cognitive Neuroscience* 3(1).
- Tuytelaars, T. et L. Van Gool (1999). “Content-based image retrieval based on local affinely invariant regions.” In *Visual Information and Information Systems*, pp. 493–500.
- Unser, M. (1986, July). “Local linear transforms for texture measurements.” *Signal Processing* 11, 61–79.
- Vaithyanathan, S. et B. Dom (1999). “Model selection in unsupervised learning with applications to document clustering.” In *the 16th International Conf. on Machine Learning*, pp. 433–443. Morgan Kaufmann, San Francisco, CA.
- van Gool, L., P. Dewaele, et A. Oosterlinck (1985). “Texture analysis: Anno 1983.” *Computer Vision, Graphics and Image Processing* 29, 336–357.
- Vapnik, V. (1995). *The nature of Statistical learning theory*. New York.
- Vasconcelos, N. (2003, 14-17 September). “A family of information-theoretic algorithms for low-complexity discriminant feature selection in image retrieval.” In *IEEE International conference on image processing*, Barcelone, Spain.

- Vasconcelos, N. et A. Lippman (2000a). "Learning from user feedback in image retrieval systems." *Advances in Neural Information Processing Systems 12*.
- Vasconcelos, N. et A. Lippman (2000b). "Learning over multiple temporal scales in image databases." In *European Conference on Computer Vision*, Dublin, Ireland, pp. 33–47.
- Vasconcelos, N. et A. Lippman (2000c, September 03-08). "A unifying view of image similarity." In *International Conference on Pattern Recognition (ICPR'00)*, Volume 1, Barcelona, Spain.
- Vasconcelos, N. et A. Lippman (2001, October). "Content-based retrieval from image databases: current solutions and future directions." In *International Conference in image processing (ICIP'01)*, Thessaloniki, Greece, pp. 6–9.
- VeltKamp, R.C et M. Tanase (2001). "Content based image retrieval systems: A survey." Technical report uu-cs-2000-34, Utrecht University.
- Wang, J. et G. Wiederhold (2001). "SIMPLICITY: Semantics-sensitive integrated matching for picture libraries." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.
- Winter, A. et C. Nastar (1999, 22 June). "Differential feature distribution maps for image segmentation and region queries in image databases." In *IEEE CVPR'99 workshop on Content Based Access of Images and Video Libraries (CBAVL)*, Ft Collins, Colorado.
- Wood, M.E.J, N.W. Campbell, et B.T. Thomas (1998, September). "Iterative refinement by relevance feedback in content-based digital image retrieval." In *ACM Multimedia 98*, Bristol, UK, pp. 13–20.
- Wu, L., C. Faloutsos, K.P. Sycara, et T.R. Payne (2000). "FALCON: Feedback adaptive loop for content-based retrieval." In *VLDB 2000, Proceedings of 26th International Conference on Very Large Data Bases, September 10-14, 2000, Cairo, Egypt*, pp. 297–306.
- Wu, Y., Q. Tian, et T. Huang (2000, June). "Discriminant-EM algorithm with application to image retrieval." In *Computer Vision and Pattern Recognition*, Hilton Head Island, SC., pp. 222–227.
- Zhang, L., F. Lin, et B. Zhang (2001, October). "Support vector machine learning for image retrieval." In *IEEE International conference on image processing (ICIP'01)*, Thessaloniki, Greece, pp. 721–725.
- Zhou, X.S. et T.S. Huang (2001). "Comparing discriminating transformations and svm for learning during multimedia retrieval." In *ACM Multimedia*, Ottawa, Canada.