

**MÉTHODES DE FOUILLE DE DONNÉES POUR
L'INDEXATION D'IMAGES ET RECHERCHE
INTERACTIVE**

THÈSE DE DOCTORAT

Discipline: Informatique et télécommunication

Université Ibn Tofail

Directeur de Thèse

Pr. Abderrahmane SBIHI

PAR

Hassan TABOUT

Septembre 2010

RESUMÉ

MÉTHODES DE FOUILLE DE DONNÉES POUR L'INDEXATION D'IMAGES ET RECHERCHE INTERACTIVE

Avec le développement des nouvelles technologies, le domaine d'indexation et recherche d'images par le contenu acquiert de plus en plus d'importance. On est donc souvent amené à vouloir déterminer quelles sont les images d'une base qui sont similaires à une image donnée appelée « requête ». Indexer une image consiste à en extraire une signature. Rechercher une image dans une base consiste à comparer plusieurs signatures entre elles. La recherche d'images par le contenu est généralement basée sur des caractéristiques primaires ou bas niveaux telles que la couleur, la forme et la texture pour construire la signature.

Dans cette thèse, nous proposons une étude de l'indexation et de la recherche d'images couleurs d'algues, ceci afin d'en développer une plate-forme complète. Nous cherchons à représenter les images de façon locale, cela grâce à l'utilisation d'une méthode de segmentation automatique permettant des requêtes partielles. Ensuite, nous étudions les modèles de métriques permettant d'évaluer la similarité entre images. Enfin, nous proposons un système de bouclage de pertinence visant à faciliter la navigation au sein de la base d'images en tenant compte des retours de l'utilisateur.

Mots clef: Indexation d'images, Recherche par le contenu, Image couleur, Segmentation, Marches aléatoires, Bouclage de pertinence, Similarité.

ABSTRACT

DATAMINING METHODS FOR IMAGE INDEXING AND INTERACTIVE RETRIEVAL

With the development of new technologies, the field of content based image indexing and retrieval becomes increasingly important. We need often to retrieve images which are similar to a query image. Indexing an image consists in extracting a signature from it. Then, retrieving a query image from a database is to compare several signatures together. The content based image retrieval systems are usually based on low levels features such as color, shape and texture to build signatures.

The contribution of this thesis relates a color Seaweed image indexing study to, finally, develop a complete framework. We build a local image representation. For this purpose, we propose a random walks based segmentation method allowing partial requests. Afterwards, we study similarity evaluation between images. Finally, we propose a new relevance feedback algorithm aiming to facilitate image database browsing while taking account of the user desire and visual similarity interpretation.

Key words: Image indexing, Content-based retrieval, Color image, segmentation, Random Walks, Relevance feedback, Similarity.

I dedicate this work to
and
to

ACKNOWLEDGEMENTS

I would like to thank

TABLE DES MATIÈRES

CHAPITRE

1	INTRODUCTION	2
1.1	La recherche d'images par le contenu	3
1.2	Contexte applicatif	5
1.3	Contributions	7
1.4	Plan du mémoire	7
2	PANORAMA SUR LA RECHERCHE D'IMAGES PAR LE CONTENU	9
2.1	Introduction	9
2.2	La recherche d'images par le contenu	9
2.3	Architecture du système de recherche d'images	10
2.4	Requête de l'utilisateur :	11
2.5	Le Fossé sémantique	13
2.6	Domaines d'application	14
2.7	Extraction des caractéristiques	17
2.7.1	La Couleur	18
2.7.2	La texture	24
2.7.3	La forme	30
2.8	Segmentation et points d'intérêt	30
2.8.1	Segmentation	31
2.8.2	Les points d'intérêt	32
2.9	Mesure de similarité	34
2.9.1	Combinaison des descripteurs	36
2.10	Indexation Multidimensionnelle	37
2.11	Retour de pertinence	38
2.12	Quelques systèmes de recherche d'images	38
2.13	Conclusion	40
3	APPROCHE DE SEGMENTATION POUR LA DESCRIPTION LOCALE DE L'IMAGE	41
3.1	Introduction	41
3.2	Description locale	41
3.2.1	Segmentation d'images	43
3.3	Méthodes de segmentation en CBIR	44
3.4	Graphes et marches aléatoires	45
3.4.1	Graphes	46
3.4.2	Marches aléatoires	47
3.5	Regroupement spectrale	48
3.5.1	Matrice Laplacienne	48

3.5.2	Segmentation spectrale normalisée	50
3.6	Nouvelle méthode de segmentation basée sur les marches aléatoires	53
3.7	Méthode de segmentation	54
3.8	Résultats et Discussions	58
3.8.1	Evaluation et Comparaison	59
3.8.2	Limitations et améliorations	60
3.8.3	Comparaison à d'autres méthodes de segmentation	61
3.8.4	Exemple de résultats sur des images naturelles	63
3.9	Conclusion :	64
4	INDEXATION ET RECHERCHE D'IMAGES	66
4.1	Introduction	66
4.2	Procédé d'extraction de connaissances	67
4.3	SWIR : Application à une base d'algues	69
4.4	Description du contenu	73
4.4.1	Description locale	74
4.5	Techniques de recherche	75
4.5.1	Requête par région	76
4.5.2	Requête par plusieurs régions	76
4.6	Implémentation	78
4.6.1	Phase d'indexation	79
4.6.2	Procédé de recherche	80
4.6.3	Interface utilisateur	80
4.7	Evaluation des performances	81
4.7.1	Rappel et Précision	82
4.7.2	Base d'images	83
4.7.3	Signature visuelle	84
4.7.4	Recherche globale	85
4.7.5	Recherche partielle	87
4.7.6	Extension de la recherche d'images à l'environnement mobile.	89
4.8	Conclusion	90
5	RECHERCHE INTERACTIVE D'IMAGES	91
5.1	Introduction	91
5.2	Etudes Bibliographique	93
5.2.1	Méthodes géométriques	94
5.3	Classification supervisée	95
5.3.1	Méthodes à noyaux	96
5.3.2	Algorithmes à base de SVM	98
5.4	Nouvelle approche au RF	100
5.4.1	Introduction	100
5.4.2	Apprentissage semi supervisé à multiple instances	101
5.4.3	Régularisation	103
5.4.4	Méthode de Consistance	104
5.4.5	Description de la méthode proposée : "MISSL_based RF" .	105
5.5	Bouclage de pertinence Automatique	109
5.5.1	Introduction	109
5.5.2	Description de la méthode proposée : "MISOM_based ARF"	110
5.5.3	Apprentissage de la similarité	113

	viii
5.5.4 Evaluation des performances	114
5.6 Conclusion	118
6 CONCLUSION ET PERSPECTIVES	120
6.1 Bilan	120
6.2 Perspectives	122
 REFERENCES	 124
 APPENDIX	
A π^0 FORMULAE	135
 PUBLICATIONS	 136
 CURRICULUM VITAE	 138

LISTE DES TABLEAUX

TAB.

4.1	Valeurs nutritive de quelques espèces d'algues	70
4.2	Quelques exemples d'utilisation de certaines espèces d'algues . . .	72
4.3	Performance de descripteurs	85

TABLE DES FIGURES

FIG.

1.1	Interaction de la recherche d'images avec les autres domaines de recherche	3
2.1	Schéma d'un système de recherche d'image par le contenu	11
2.2	Correspondance longueur d'onde/couleur pour le spectre visible	19
2.3	Sensibilité spectrale des cônes	19
2.4	Cube de l'espace couleur RGB	20
2.5	Représentation de l'espace couleur XYZ	21
2.6	Représentation de l'espace couleur HSV	21
2.7	Espace de couleur Munsell	22
2.8	Exemple d'histogramme d'une image couleur	23
2.9	Exemple de textures et de module de la transformée de Fourier associée	27
2.10	Illustration de l'extraction de caractéristiques de texture par filtrage de Gabor	28
2.11	Ondelettes de Haar	29
2.12	Décomposition AMR par ondelettes de Haar	30
2.13	Exemple de segmentation par approche région	32
2.14	Exemple des points d'intérêt en utilisant Harris Couleurs	34
3.1	Image de synthèse représentant un type d'algue	43
3.2	Influence du nombre de sommets dans la coupe	52
3.3	Résultats de la segmentation par coupe normalisée	53
3.4	Utilisation des valeurs propres pour la segmentation d'images.	57
3.5	Trace (P) pour 6 types d'algues.	57
3.6	Exemple de résultat de segmentation	59
3.7	Exemple de résultats de segmentation	60
3.8	Exemple de résultats impliquant la couleur	61
3.9	Comparaison entre l'approche proposée et Ncut	62
3.10	Comparaison entre l'approche proposée et Ncut	63
3.11	Comparaison entre méthodes (scènes naturelles).	64
4.1	Schéma classique illustrant le processus d'extraction de connaissances [37]	67
4.2	Processus d'ECD pour la caractérisation d'images d'algues	69
4.3	Répartition moyenne d'utilisation et Intérêt nutritionnel des algues	70
4.4	Segmentation d'une image en entités visuelles	74
4.5	Extraction des attributs de la couleur et de la texture	75

4.6	Comparaison des régions entre l'image requête et une image de la base	77
4.7	L'architecture du système de recherche proposé	79
4.8	Interface Web du moteur de recherche d'images SWIR	81
4.9	Exemple d'images utilisées dans cette étude	84
4.10	Exemple de requête globale	86
4.11	Comparaison des deux méthodes d'indexation	87
4.12	Exemple de requête par région	88
4.13	Mesure de performance par region	89
4.14	Visualisation de l'image à chercher et retour de l'information concernant le site Chellah.	90
5.1	Le bouclage de pertinence	92
5.2	Exemple illustrant la reformulation de la requête par l'algorithme de Rocchio	94
5.3	Exemple illustrant l'algorithme de repondération	95
5.4	Hyperplan optimal séparant des deux classes	98
5.5	Problème de Bouclage de Pertinence.	105
5.6	Architecture du bouclage de pertinence	109
5.7	Bouclage de pertinence automatique proposé	111
5.8	Réduction de l'espace de signature à l'aide de SOM	112
5.9	Process d'identification de pertinence en utilisant MISOM	113
5.10	Courbes Rappel/Précision par itération	115
5.11	Exemple de recherche d'images sans bouclage	116
5.12	Exemple de recherche d'images avec bouclage de pertinence automatique	117
5.13	Courbes rappel précision comparant des méthodes de bouclage de pertinence	117

LIST OF SYMBOLS

Chapitre 1

INTRODUCTION

Avec l'émergence rapide et la révolution des appareils photographiques numériques, le nombre d'images numériques disponibles a subi une augmentation phénoménale ces dernières années. La diminution du coût de stockage et la disponibilité des techniques de numérisation de haute qualité permettent aussi aujourd'hui de constituer de très grandes bases d'images dans des domaines variés : images médicales, images d'art, images de la biodiversité, images satellitaires, ... Ainsi, trouver la bonne image est devenue un problème coûteux. Il est nécessaire de disposer des données organisées de façon à permettre l'exploration, la recherche et la récupération de ces données. Des solutions ont donc été développées pour permettre de faire des recherches dans des bases de données d'images de manière automatique. Ces solutions se répartissent en deux catégories : Recherche d'images par mots clés (Text-Based Image Retrieval (TBIR)) et Recherche d'images par le contenu (Content-Based Image Retrieval (CBIR)). La technique TBIR consiste à annoter les images manuellement par des mots, et utiliser ensuite des systèmes de gestion de base de données pour effectuer la recherche d'images. Toutefois, il existe deux difficultés majeures, en particulier lorsque la taille de la base d'images est importante. La première est le laborieux travail requis pour effectuer une annotation manuelle de chaque image. L'autre difficulté est la subjectivité de la perception humaine, qui fait que pour le même contenu de l'image, des personnes différentes peuvent le percevoir différemment. La subjectivité et l'imprécise annotation peuvent causer une inadéquation entre les images retournées et le besoin de l'utilisateur. Cela conduit à fournir beaucoup d'efforts sur la recherche d'images par le contenu. Grâce à cette technique, les images sont indexées par leur contenu visuel, comme la couleur, la texture ou la forme. CBIR est devenu de plus en plus important grâce à l'avancée de la technologie informatique, et fournit la réponse à ces deux inconvénients du système TBIR mentionnés ci-dessus.

Aujourd'hui, le problème principal posé par l'expansion rapide des docu-

ments multimédia est qu'il existe peu de systèmes de recherche d'images suffisamment performants pour permettre de retrouver une certaine image dans une base d'images. C'est pour cette raison que de nombreux chercheurs à travers le monde s'intéressent de près à la recherche d'images. C'est un problème ouvert qui se trouve à la rencontre de plusieurs thématiques de recherche diverses : les bases de données, les statistiques, l'interaction homme machine, l'intelligence artificielle, le traitement d'images, les réseaux et autres domaines (figure 1.1)

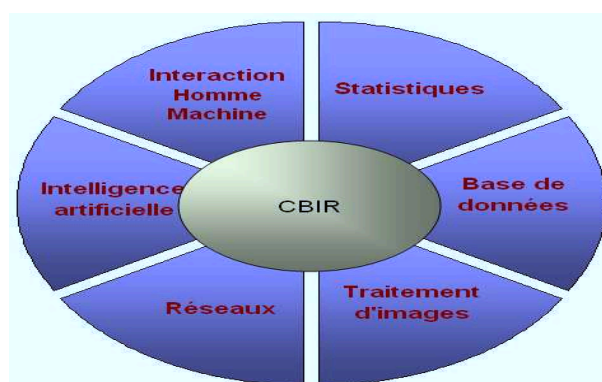


FIG. 1.1 – *Interaction de la recherche d'images avec les autres domaines de recherche.*

- Les bases de données pour le stockage et l'indexation des images,
- L'intelligence artificielle : pour l'apprentissage, mesure de similarité, classement des images les unes par rapport aux autres, . . .
- Le traitement d'images pour l'extraction des descripteurs discriminants d'images,
- Les statistiques, les distances métriques pour l'évaluation de distances entre images, le regroupement, la segmentation, . . .
- L'interaction Homme Machine, pour l'élaboration d'une interface graphique et conviviale permettant à l'utilisateur d'interroger et paramétrer le système de recherche.
- Les Réseaux, Communication pour la gestion de la base de données à distance avec tous les problèmes de sécurité, de la qualité de service et de transmission de l'information qui en découlent.

1.1 La recherche d'images par le contenu

"Une image vaut mille mots"

Récemment, ce proverbe chinois a été largement utilisé par les chercheurs de la communauté des technologies de l'information pour dire que les images ne peuvent pas être bien décrites par les annotations textuelles, et donc des systèmes de recherche d'information traditionnels ne sont pas appropriés pour la recherche d'images. De nombreuses techniques de recherche d'information ont vu le jour ces dernières années dont la fouille de données (Data Mining). La fouille de données consiste à analyser, par des méthodes statistiques et mathématiques, un large volume de données, afin d'en faire ressortir des tendances ou des règles. Il correspond à l'ensemble des techniques et des méthodes qui, à partir de données, permettent d'obtenir des connaissances exploitables. Les techniques de recherche de textes sont très efficaces car elles bénéficient de travaux de recherche éprouvés depuis une dizaine d'années. Ainsi, les moteurs de recherche basés sur le texte sont de plus en plus performants. Malheureusement, la recherche d'images pose encore de nombreux problèmes en raison de la difficulté à mettre en oeuvre des algorithmes efficaces d'analyse d'images. Il est important de souligner que la recherche par le contenu n'est pas un remplaçant, mais peut être un complément à la recherche d'images à base des mots textuels. L'intégration des deux peut aboutir à une recherche de performance satisfaisante.

Le but d'un système CBIR est de retrouver une ou plusieurs images parmi une base d'images pour répondre à une requête d'un utilisateur. La recherche sur les bases d'images nécessite de posséder une base d'images sur laquelle on puisse tester les méthodes mises en oeuvre. On peut distinguer deux types : les bases spécialisées, contenant un type réduit d'images (par exemple les bases médicales, les bases de la biodiversité, . . .), et les bases généralistes, contenant des images de natures très variées et de différentes provenances. Les méthodes de recherche dans les bases spécialisées se distinguent de celles dans les bases généralistes par le fait que l'on connaît par avance le type d'images rencontrées, ainsi que les recherches qui vont y être menées. Certaines bases d'images disponibles gratuitement existent sur Internet comme la base de Wang[151] ou la base Columbia[24], chacune contenant quelques milliers d'images. D'autres bases d'images contenant plusieurs millions d'images existent également mais sont commerciales, la base la plus populaire est la base de Corel[27]. Enfin, la plus grande source d'images au monde c'est l'Internet. Les moteurs de recherche contiennent les adresses de millions d'images, d'où l'intérêt commercial de la recherche d'images. Le choix de la base de données est très important pour l'étude d'un système de recherche d'images par le contenu. En effet, le contenu des images va influencer très fortement les choix de la méthode de recherche. Ainsi une méthode qui marche sur une base donnée peut donner des résultats très différents sur une autre base. Dans la suite de cette thèse, le champ d'application de notre méthode sera défini plus

précisément. En effet, nous nous intéressons dans cette thèse à une base spécialisée, plus précisément à une base d'images d'algues. On distingue généralement deux niveaux d'abstraction pour ce type d'images :

- Le bas niveau dit " numérique ", qui fait référence au contenu signal de l'image, autrement dit l'ensemble des pixels qui la constitue.
- Le haut niveau dit "sémantique ", qui fait référence au contenu interprétable de l'image, autrement dit l'ensemble des objets et significations que l'on peut lui associer.

Il existe souvent une forte différence entre la description bas-niveau et la sémantique d'une image, que l'on appelle le fossé sémantique. Il s'agit de la discordance entre l'information que l'on peut extraire des données visuelles en utilisant des caractéristiques bas niveaux de l'image et les interprétations des utilisateurs. Dans le but de combler le fossé sémantique, le premier point à traiter concerne la phase d'indexation. Celle-ci est en général déterminée en amont de l'utilisation du système de recherche. Concernant ce point, certaines techniques mobilisant des descripteurs de niveaux d'abstraction élevés sont utilisées. En effet, la segmentation de l'image en entités visuelles, la caractérisation de ces entités, la définition d'une mesure de similarité adéquate peuvent améliorer la recherche. La première difficulté réside dans la détermination d'un algorithme capable d'extraire des régions homogènes et respectant une certaine cohérence visuelle. La deuxième est dans le choix d'attributs puis dans la définition des descripteurs pour caractériser les composantes précédentes. Enfin, la mise en place d'une mesure de similarité entre requête et image de la base qui exploite les éléments cités précisément. L'automatisation du processus de recherche implique le traitement de plusieurs problèmes de natures différentes. Par suite, les performances d'un système de recherche est tributaire du traitement efficace des briques qui le constitue. Toutefois, pour atteindre une sémantique plus élevée, en adéquation avec le but poursuivi par l'utilisateur, il faut un traitement supplémentaire de haut niveau. Nous nous sommes inscrits dans cette problématique en essayant de proposer une recherche interactive. L'utilisateur fournit des informations au système dans le but de préciser la nature de sa recherche.

1.2 Contexte applicatif

Les algues jouent un rôle primordial dans le maintien de l'équilibre écologique du milieu aquatique. Elles sont les principales responsables de la production primaire. De point de vue économique, elles présentent actuellement une source nutritionnelle et un produit à valeur montante. Elles sont utilisées en agriculture comme engrais et fourrage, dans l'industrie alimentaire et pharma-

ceutique, dans le textile et dans d'autres domaines. Les algues marines sont l'une des plus importantes ressources marines vivantes. Elles pourraient être qualifiées comme la futuriste prometteuse plante [32]. Ces plantes sont la source de nourriture, d'alimentation et de la médecine. Agar-agar, Carrageenan et Alginate sont des exemples populaires de produits extraits des algues. Ceux-ci ont été utilisés comme nourriture pour les êtres humains, aliments pour animaux, engrais pour les plantes et la source de divers produits chimiques. Les produits à base d'algues sont utilisés dans notre vie quotidienne de différentes manières. Par exemple, certaines algues polysaccharides sont employées dans la fabrication de dentifrices, savons, shampooings, produits cosmétiques, lait, glaces, aliments transformés, désodorisants d'air et une foule d'autres produits [32, 2].

L'identification précise des espèces est primordiale pour de nombreuses disciplines, notamment pour l'élaboration de moyens de lutte performants contre les adventices, la prise de conscience de l'extinction d'espèces, lutte contre les espèces nuisible et les espèces invasives. L'objectif de cette thèse est de contribuer à gérer et d'utiliser les métadonnées multimédia pour faciliter l'accès à la diversité biologique du patrimoine algal. Ceci est fait en tenant compte, d'une part, des outils d'analyse et de traitement de l'image pour la description du contenu de ces images, et d'autre part, de l'installation d'un système de navigation et de recherche dans une telle base. L'indexation manuelle pour une telle tâche est difficile à cause du volume de données. Il y a donc un réel besoin d'automatiser pour une utilisation plus rapide et efficace. Parmi les applications les plus récentes du système de recherche d'images d'algues est la reconnaissance automatique des principales espèces d'algues sur des clichés sous marines. Un dispositif de prise en vue sous marines dont le but est de prendre des images des fonds marin où croît l'espèce. Cela, permettra d'éviter le risque que prennent les plongeurs par mauvais temps. D'où l'idée de remplacer ces prélèvements par des analyses automatisées d'images sous marines afin de diminuer le coût de réalisation de la cartographie et d'étendre les zones prospectées. Également, en botanique et surtout dans l'étude de l'expression des gènes, on aimerait souvent rapprocher les images pour étudier les effets des modifications des gènes et faire les conclusions adéquates. Aussi, la recherche d'images d'algues peut elle être utilisée en système éducatif où l'apprenant peut extraire des informations sur une espèce donnée et la rapprocher avec les espèces déjà existantes dans la base. La disponibilité des collections d'images pourra aussi aider le formateur à donner des exemples et des illustrations afin de réduire le temps de préparation et d'améliorer la qualité de l'enseignement.

1.3 Contributions

Plusieurs contributions sont apportées dans nos travaux de recherche d'images par le contenu visuel. Il s'agit :

- d'une proposition à l'utilisateur d'un moteur de recherches d'images d'algues (SWIR ou SeaWeed Image Retrieval), depuis la segmentation, jusqu'à la requête ;
- du développement d'une méthode de segmentation pour améliorer la pertinence et l'efficacité d'un tel moteur ;
- de la conception d'outils permettant de caractériser au mieux le contenu d'une image ;
- de la mise en place d'une métrique de similarité liée à des considérations visuelles ;
- de la mise au point d'un processus de recherche interactive d'image.
- de la mise en place des extensions à l'environnement mobile.

1.4 Plan du mémoire

Au cours de ce travail, nous nous sommes attachés à suivre la linéarité imposée par la description même d'un système de recherche d'images. Ainsi, le second chapitre, relate l'état de l'art de la recherche d'images. Nous y présentons d'abord, les caractéristiques décrivant le contenu visuel. Ensuite, nous y abordons les différentes méthodes pour caractériser ces attributs, ainsi que les principales distances existantes pour établir la similarité entre les attributs caractérisant les images et les techniques utilisées pour combiner les descripteurs d'une image. Ce chapitre se termine par une présentation des systèmes de recherche d'images les plus populaires.

Le troisième chapitre, présente une étude de la description locale d'une image. Sa première partie présente l'état de l'art sur les méthodes de segmentation reconnues, ainsi que les principales méthodes et leurs utilisations dans les systèmes CBIRs dont le but est de montrer les différences entre ces méthodes. La deuxième partie de ce chapitre, présente une nouvelle méthode de segmentation. Après justification de la nécessité de la décomposition de l'image en régions, et le rappel de quelques fondements théoriques, ce chapitre expose l'étude expérimentale de la méthode proposée, son évaluation et sa comparaison avec d'autres méthodes.

Le quatrième chapitre concerne la recherche d'images par le contenu. Les images sont déjà segmentées par la technique présentée dans le chapitre trois. Nous définissons les descripteurs de l'image utilisés. Nous présentons la méthode

de mesure de similarité utilisée. Nous abordons l'application aux images d'algues et nous présentons le système de recherche d'images élaboré. Enfin nous exposons l'évaluation du système.

Le cinquième chapitre est dédié au bouclage de pertinence. Nous proposons une vue générale des principales techniques de retour de pertinence. Nous exposons une nouvelle méthode de bouclage de pertinence. Aussi, et afin de minimiser le nombre d'itérations, voire éviter l'intervention de l'être humain, nous proposons une méthode de bouclage de pertinence automatique.

Enfin, nous concluons ce mémoire de thèse par le chapitre six, dans lequel nous revenons sur les résultats obtenus et nous présentons l'ensemble des perspectives de recherche ouvertes par nos travaux.

Chapitre 2

PANORAMA SUR LA RECHERCHE D'IMAGES PAR LE CONTENU

2.1 Introduction

Les systèmes de recherche d'images par le contenu possèdent une phase d'indexation et une phase de recherche. L'indexation a pour but d'extraire et de représenter le sens d'un document de manière à ce qu'il puisse être retrouvé par l'utilisateur. L'étape de recherche consiste en l'exploitation des caractéristiques extraites lors de l'étape précédente pour mesurer la similarité entre l'image requête et les images de la base de données.

L'objectif de ce chapitre est de dresser un état de l'art sur les méthodes d'indexation et de recherche d'image par le contenu. Nous détaillons comment les informations visuelles sont extraites des images et comment ces systèmes les utilisent pour rechercher l'information. Nous allons mettre en évidence les difficultés rencontrées pour construire un tel système. Nous décrivons quelques domaines d'applications. Puis nous parlons de quelques systèmes de recherche d'images par le contenu.

2.2 La recherche d'images par le contenu

La recherche d'images par le contenu est un problème très difficile en raison de la grande taille de la base de données, la difficulté d'interpréter l'image, aussi bien par l'être humain que par la machine, la difficulté de formuler une requête, et le problème de l'évaluation des résultats. Ces défis et les travaux antérieurs sont examinés ci-dessous.

Taille des données

Pour construire un système de recherche d'images pour les domaines tels que l'imagerie biomédicale, la biodiversité ou l'internet, il est nécessaire de stocker et traiter des milliers voire des millions d'images de manière efficace. Cela requiert un espace mémoire très important. Il n'est évidemment pas possible de traiter toutes ces données en ligne pour chaque requête. L'indexation hors ligne est alors nécessaire afin de réduire le temps de traitement de la requête.

Vision par ordinateur

On dit qu'une image vaut mille mots. En effet, on peut imaginer un millier de mots pour décrire le contenu d'une image. En outre, le "sens" d'une image dépend du point de vue et de l'expérience de celui qui l'interprète. En d'autres termes, la description fournie par un utilisateur concernant une image peut être différente de celle fournie par un autre utilisateur. La vision par ordinateur reste parmi les plus difficiles problèmes en informatique. Il est extrêmement difficile de reconnaître les objets en images et traduire leurs contenus et leurs structures en termes linguistiques.

2.3 Architecture du système de recherche d'images

Un système de recherche d'images par le contenu (CBIR) ne traite pas seulement les diverses sources d'information en différents formats (par exemple, texte, image, vidéo), mais aussi les besoins de l'utilisateur. En effet, il analyse le contenu de la source de l'information ainsi que la requête de l'utilisateur, puis après une mesure de similarité, il permet de retourner les images pertinentes. Les principales fonctions d'un tel système sont les suivantes :

1. Analyser le contenu de la base de données, et représenter le contenu des sources analysées d'une manière qui convient à la mesure de similarité avec la requête de l'utilisateur (espace de sources d'information se transforme en espace d'attributs pour le but de mesurer la similarité). Cette étape est gourmande en termes de temps de traitement, car l'extraction des caractéristiques se fait d'une façon séquentielle pour toutes les images de la base de données. Toutefois, il faut le faire une seule fois et peut se faire hors ligne.

2. Analyser la requête utilisateur et la représenter sous une forme qui correspond à la base de données source. Une partie de cette étape est similaire à l'étape précédente, mais elle est appliquée seulement sur l'image requête.

3. Définir une stratégie pour mesurer la similarité visant à répondre aux requêtes. Récupérer efficacement les informations qui sont pertinentes. Cette étape se fait en ligne et il est nécessaire qu'elle soit rapide. Des techniques d'indexations récentes tentent de réorganiser l'espace d'attributs afin d'augmenter la vitesse de traitement de similarité.

4. Faire les ajustements nécessaires dans le système, le plus souvent par un réglage des paramètres dans la mesure de similarité ou reformulation de la requête dans le cadre d'un système interactif.

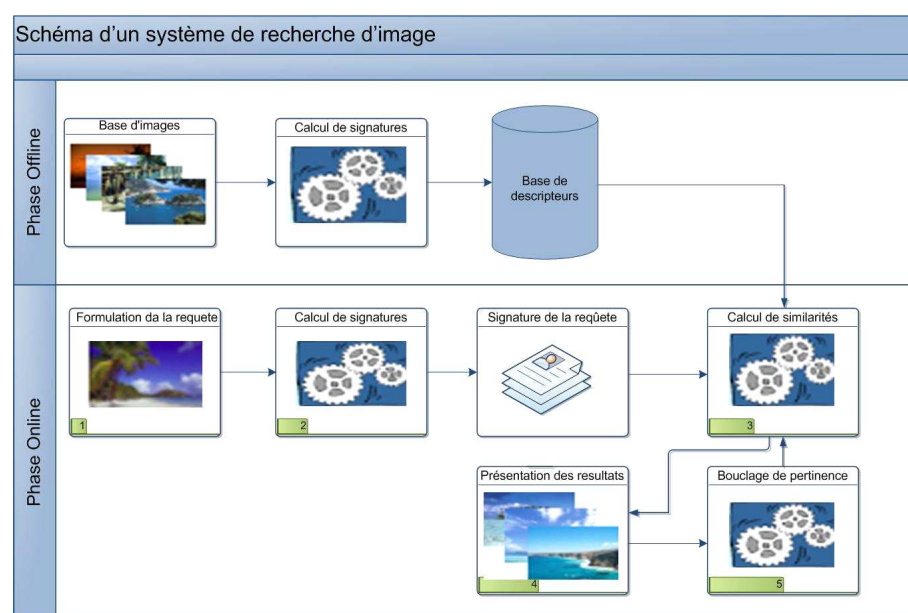


FIG. 2.1 – Schéma d'un système de recherche d'image par le contenu.

Il ressort clairement de ce qui précède que, d'une part, un système CBIR, traite les sources d'information visuelle dans différents formats et d'autre part il y a les requêtes de l'utilisateur. Ces deux parties sont liées par une série de tâches comme l'illustre la figure 2.1. Ces tâches seront explicitées en détails dans les sections qui suivent.

2.4 Requête de l'utilisateur :

Il existe de nombreuses façons par lesquelles on peut lancer une requête visuelle. Ci dessous les différentes méthodes qui sont couramment utilisées dans les systèmes de recherche d'image par le contenu :

Requête par l'exemple (QBE, Query by Example) :

Dans ce type de requête, l'utilisateur du système spécifie une requête de l'image cible avec laquelle l'image de la base de données doit être comparée. L'image requête peut être une image normale, une image de faible qualité (image de fax, images sous marines,...), ou un dessin qui est tracé en utilisant une interface graphique avec des outils de dessin appropriés. Un premier avantage de ce type de système est qu'il constitue un moyen naturel aussi bien pour les experts que pour les utilisateurs en général qui sont à la recherche d'une image de la base de données.

Requête par Caractéristiques (QBF, Query by Feature) :

Dans ce type de système, l'utilisateur précise explicitement les caractéristiques qui l'intéressent. Par exemple trouver toutes les images dont la gauche contient 25% de jaune et 30% de vert. Cette requête est indiquée en utilisant une interface graphique. Ce type de requête est destiné pour des utilisateurs spécialisée. QBIC [40] est un exemple de système qui utilise ce type de requêtes.

Requête en combinant le texte et l'image :

Dans ce type de système [15], l'utilisateur emploie le texte comme une clé primaire de recherche pour renforcer la recherche par l'exemple. Ce type de représentation implique un haut degré d'abstraction qui est difficile à atteindre parce que l'image contient une grande quantité d'informations qu'il est difficile de résumer en quelques mots-clés. Cette méthode est généralement facile à implémenter, mais elle nécessite un laborieux travail pour annoter manuellement la base d'images et présente un degré élevé de subjectivité comme mentionné précédemment.

Requête par image mentale :

La recherche par croquis est une première approche de ce que l'on appelle la recherche par image mentale [117]. Cette recherche correspond à la situation où on a simplement une mémorisation d'une situation ou un contexte. En d'autres termes, l'utilisateur a une image précise en tête et il utilise les fonctionnalités du système pour réussir à la retrouver dans la base. Une version plus évoluée de cette approche consiste à utiliser un thésaurus de patches visuels. C'est un dictionnaire regroupant des parties d'image caractéristiques de la base. Par composition de ces patches, l'utilisateur exprime une requête [55]. Il s'agit d'un mi-chemin entre la requête par l'exemple et la requête par croquis.

2.5 Le Fossé sémantique

La principale difficulté rencontrée dans la recherche d'image par le contenu est que les utilisateurs font généralement leurs requêtes à un niveau sémantique, tandis que la requête est caractérisée par des descripteurs de bas niveau. Ces descripteurs ne couvrent qu'une gamme limitée des caractéristiques de l'image. Le problème est multiple et il porte sur l'ensemble du domaine de la vision par ordinateur. Il est très difficile de trouver des descripteurs qui permettent de prendre en compte d'une part toutes les représentations possibles d'une même scène et d'autre part la subjectivité des requêtes formulées par l'utilisateur. Plusieurs aspects doivent être pris en considération quand on aborde la recherche d'image par le contenu : le fossé sémantique, le fossé sensoriel et le fossé numérique.

Le fossé numérique.

Il s'agit principalement de l'utilisation incomplète ou de la confusion à décrire le contenu visuel. En d'autres termes, c'est l'incapacité de la signature (caractéristiques bas niveau de l'image) de décrire le contenu de l'image. Pour réduire ce fossé numérique, il faut développer des signatures qui soient à la fois riches et fidèles au contenu. Pour rendre la recherche plus efficace, ces signatures doivent également être compactes. Sans signatures de bonne qualité, nous ne pouvons pas avoir de bons résultats. En effet, des précautions particulières doivent être prises au moment de choix des descripteurs car la qualité de recherche dépend directement de la fidélité de la signature à l'aspect visuel qu'elle décrit.

Le fossé sensoriel.

Les bases de données hétérogènes ont un contenu visuel qui couvre des types d'images de différentes provenances. Il est un facteur de confusion qui ne vient pas de l'interprétation réelle de l'image, mais plutôt de la façon avec laquelle l'image a été enregistrée : Qualité médiocre de l'image, angles de visions contraignants, manque d'éclairage, bruit. Ces facteurs qui s'ajoutent à la difficulté d'interprétation d'images sont généralement perçus comme contribuant à l'écart sémantique, mais leur cause est plutôt physique : différence entre l'objet lui-même et sa photo. Cet écart provient de la différence entre l'objet et l'information qu'on peut tirer d'une image de cet objet. Le fossé sensoriel peut être contrôlé en quelque sorte pour des projets spécialisés, où les images sont enregistrées de manière contrôlée, ce qui permet d'assurer la qualité du résultat. Mais pour les bases d'images généralistes, il est inévitable, car la source des images

est généralement inconnue et la qualité des images, dans la plupart du temps, est non contrôlée.

Le fossé sémantique.

Le fossé sémantique est le manque de concordance entre l'information que l'on peut extraire des données visuelles en utilisant des caractéristiques bas niveau de l'image et les interprétations des utilisateurs du système [118]. C'est la principale source de problèmes dans un CBIR, puisque les résultats retournés par un système sont évalués en fonction de leur concordance avec les attentes de l'utilisateur (qui sont formulés à un niveau sémantique). Les caractéristiques bas niveau sont souvent incapables d'offrir une description détaillée du contenu sémantique d'une image. Cela est d'autant plus vrai que le contenu sémantique diffère d'un utilisateur à l'autre. La description sémantique du contenu visuel est généralement préférable de le faire à l'aide des mots clés et des annotations textuelles [105], mais la présence de mots clés dans les bases de données image reste rare en raison des coûts impliqués dans les annotations et aussi en raison de la subjectivité implicite, en utilisant les annotations provenant de plusieurs sources. Une approche pour résoudre ce problème a été tentée en [8] où les auteurs utilisent un système web pour naviguer sur Internet à la recherche d'images et les annoter en utilisant le texte qui les entourent. Toutefois, il existe parfois des aspects du contenu visuel qu'on ne peut décrire même en utilisant un nombre raisonnable des mots-clés. D'autre part, l'étiquetage seul est rarement suffisant pour capturer tout le contenu sémantique. Il est généralement incomplet, sensible au contexte et souffre des erreurs venant de l'annotateur [1, 52]. Dans cette perspective, des travaux récents, pour pallier au problème du fossé sémantique utilisent les descripteurs visuels avec la technique du bouclage de pertinence. En outre, le fossé sémantique est responsable de certaines difficultés rencontrées lors de la définition des méthodes d'évaluation de la qualité et l'efficacité des algorithmes de recherche d'image. Le problème le plus important dans l'évaluation du système CBIR est l'absence de la vérité terrain.

2.6 Domaines d'application

Les domaines d'application des systèmes de recherche d'image par le contenu sont immenses. En fait là où il y'a des archives multimédia, les systèmes CBIR trouvent pleinement leurs domaines d'application. Ci dessous quelques exemples :

La criminalité

En général, les services de police maintiennent de grandes archives d'images concernant les photos de visages, les empreintes digitales des anciens suspects, criminels ou terroristes, des images de semelles de chaussure, etc. Chaque fois qu'un crime est commis, ils peuvent comparer la preuve sur les lieux du crime par rapport aux images préstockées dans leurs archives.

Militaire

La reconnaissance des avions de l'ennemi via les écrans de radar, l'identification des cibles via les images satellites et le guidage des missiles de croisière sont des exemples parmi d'autres qui illustrent le rôle des systèmes CBIR dans le domaine militaire.

La propriété intellectuelle

Lors de dépôt de la marque de fabrication pour enregistrement, la nouvelle marque est comparée avec les marques déjà existantes afin de s'assurer qu'il n'existe aucun risque de confusion.

Diagnostic médical

Les systèmes CBIR attirent de plus en plus d'intérêt dans le domaine médical. Ces systèmes sont utilisés pour faciliter le diagnostic par l'identification des cas similaires déjà traités.

Les systèmes d'information géographiques SIG et la télédétection

Les agronomes et les géographes utilisent largement les images satellites, aussi bien dans la recherche que dans des domaines pratiques, telles que l'identification des champs qui ont été affectés de certaines maladies ou qui ont un manque d'éléments nutritifs, ou détecter certaines transgressions.

Education et formation

Il est souvent difficile d'identifier le bon outil didactique pour illustrer les points clés dans une conférence ou dans une séance d'auto-apprentissage d'un module. La disponibilité des collections de clips vidéo et d'images pourra aider à donner des exemples et des illustrations afin de réduire le temps de préparation et d'améliorer la qualité de l'enseignement.

La mode et Le design

Dans les domaines de la mode et du design, la recherche d'une collection de tissus pour trouver une combinaison particulière de la couleur ou de la texture est de plus en plus reconnue comme une aide utile au processus de conception. Par exemple, rechercher une texture spécifique pour l'industrie textile.

Presse et publicité

Les journaux et agences de presse maintiennent des archives d'images, des interviews, des clips vidéo pour illustrer des articles ou des textes publicitaires. Ces archives peuvent souvent être de très grandes tailles, ce qui devient coûteux pour une indexation textuelle détaillée. En effet, les sociétés de radiodiffusion sont confrontées à un problème encore plus grand, devoir traiter des millions d'heures de séquences vidéo et d'images, presque impossible à annoter d'une façon manuelle.

L'héritage culturel

Les musées et galeries d'art traitent souvent les objets visuels. La capacité à identifier des objets, partageant certaines similarités peut être utile aussi bien aux chercheurs qui veulent déduire les influences historiques, qu'aux amateurs d'art qui sont à la recherche d'autres exemples de tableaux ou de sculptures qui satisferont leur goût. Par exemple recherche d'un tableau ou d'une oeuvre d'art spécifique.

La Biodiversité

La biodiversité désigne la diversité des organismes vivants, qui s'apprécie en considérant la diversité des espèces, celle des gènes au sein de chaque espèce, ainsi que l'organisation et la répartition des écosystèmes. Le maintien de la biodiversité est une composante essentielle du développement durable [155]. La diversité biologique est la diversité de toutes les formes du vivant. Elle est habituellement subdivisée en trois niveaux :

- (1) La diversité génétique, qui se définit par la variabilité des gènes au sein d'une même espèce ou d'une population. Elle est donc caractérisée par la différence de deux individus d'une même espèce ou sous-espèce (diversité intraspécifique).
- (2) La diversité spécifique, qui correspond à la diversité des espèces (diversité interspécifique). Ainsi, chaque groupe défini peut alors être caractérisé par le nombre d'espèces qui le compose.

- (3) La diversité écosystémique, qui correspond à la diversité des écosystèmes présents sur Terre, des interactions des populations naturelles et de leurs environnements physiques.

L'identification précise des espèces est primordiale pour de nombreuses disciplines, notamment pour l'élaboration de moyens de lutte performants contre les adventices des cultures, la prise de conscience de l'extinction d'espèces, lutte contre les espèces nuisibles et les espèces invasives,... Or cette identification est généralement difficile pour des non-spécialistes, tels que des agronomes, des techniciens ou des étudiants. Les difficultés rencontrées avec les outils classiques, comme les flores ou les manuels, sont généralement liées aux méthodes d'identifications structurées sous forme de clés dichotomiques, et à la terminologie employée qui reste réservée surtout aux spécialistes. De plus, ces méthodes sont généralement basées sur l'emploi de caractères des organes reproducteurs (fleurs, fruits), qui ne s'expriment qu'une courte partie de l'année chez les plantes, alors qu'il est parfois nécessaire de les identifier avant leur stade de développement adulte, notamment pour toutes les espèces invasives. Dans l'étude de l'expression des gènes, par exemple, on aimerait souvent rapprocher les images pour étudier les effets des modifications des gènes et faire les conclusions adéquates. En effet, être capable de ramener les expressions des gènes comparables et leurs permettre donc de tirer les conclusions et de faire les études comparatives d'une manière efficace et plausible. Afin de remédier à ces contraintes, des systèmes CBIR ont commencé à voir le jour. Ikona [59], par exemple, est un système de recherche d'image par le contenu est utilisé pour la reconnaissance d'espèces végétales à partir d'images de leurs organes constitutifs (feuille, organe sexuel...). Notre travail s'inscrit dans cette optique. Notre objectif est de gérer et d'utiliser les métadonnées multimédia pour faciliter l'accès à la diversité biologique de notre patrimoine et plus précisément nous sommes intéressés aux images d'algues.

2.7 Extraction des caractéristiques

L'extraction des caractéristiques est à la base des systèmes de recherche d'images par le contenu. Dans un sens large, les caractéristiques peuvent être basées sur des attributs visuelles bas niveau (couleur, texture, forme) et/ou sur des attributs haut niveau comme par exemple les annotations textuelles. Pour les attributs visuels auxquels nous nous intéressons, il n'existe pas une seule meilleure représentation pour une fonction donnée. Vu que les attributs représentent l'image, leur choix est déterminant pour la suite de la méthode. Si les attributs sont mal choisis, la méthode de recherche donnera de mauvais résultats. Comment choisir de bons attributs ? Il n'y a pas de réponse générale à cette

question car le choix des attributs va dépendre des images en question.

Les attributs sont en général choisis par un expert du domaine des images de la base. L'expert justifie le choix des attributs par son expérience, sur les caractéristiques qui lui semblent importantes et sur le champ applicatif de la méthode de recherche. Suivant l'application développée, l'expert pourra choisir différents attributs.

Le choix des attributs est fortement dépendant des images de la base. Ainsi, les attributs qui donnent d'excellents résultats sur une base d'images peuvent donner des résultats médiocres sur une autre base. Il n'y a pas d'attributs universels donnant de bons résultats sur n'importe quelle base d'images : Les attributs dépendent du contexte (context-sensitive). Dans notre travail, nous proposons un choix des attributs utilisés afin de valider et d'illustrer notre méthode.

Il existe deux familles d'attributs, les attributs globaux qui sont calculés à partir de l'image entière et les attributs locaux qui sont calculés sur une région de l'image considérée. Des attributs de différents types sont utilisés pour représenter le contenu de l'image. Les attributs sont classés en trois familles principales : la couleur, la texture et la forme. Pour l'étude de la base d'images d'algues, la forme n'est pas utilisée, vu que nous ne sommes pas intéressés à l'étude d'un type d'algue microscopique où la forme peut jouer un rôle primordial. Par contre, la recherche des macroalgues sera effectuée à l'aide d'attributs issus de la couleur et de la texture.

2.7.1 La Couleur

La couleur est un des attributs les plus employés en indexation d'images. La couleur est une notion purement perceptuelle nous permettant d'exprimer la distinction que nous faisons entre deux sensations colorées. Elle peut donc être définie comme étant la réaction d'une chaîne de traitements comprenant l'oeil et le cerveau face à un stimulus visuel.

Historiquement, depuis l'expérience de décomposition de la lumière blanche par un prisme réalisée par Isaac Newton vers 1665 et les travaux de James Clerk Maxwell vers 1865, on sait que la lumière est composée d'un ensemble d'ondes électromagnétiques de longueurs d'ondes différentes.

La lumière est une énergie électromagnétique comportant une partie ondulatoire et une partie corpusculaire (les photons). L'énergie lumineuse possède une longueur d'onde caractérisée par sa couleur. La figure 2.2 donne la correspondance longueur d'onde/couleur pour le spectre visible par l'oeil humain : une valeur de l'ordre de 380 nm à 780 nm environ.

Le physicien James Clerk Maxwell a prouvé que toute couleur peut être



FIG. 2.2 – Correspondance longueur d'onde/couleur pour le spectre visible.

obtenue par synthèse de trois couleurs seulement : le rouge, le vert et le bleu. À partir de cette découverte, la colorimétrie, la science qui étudie la couleur, a vu le jour. Le principe des trois couleurs primaires est employé dans la télévision ou le moniteur de l'ordinateur. La biologie nous indique que l'oeil dispose de deux types de cellules réceptrices : les bâtonnets sensibles à la luminance (quantité de lumière) et les cônes sensibles à la chrominance (couleur). Il existe trois types de cônes : rouge, vert et bleu. La figure 2.3 présente la courbe de sensibilité de l'oeil aux différents stimulus colorés.

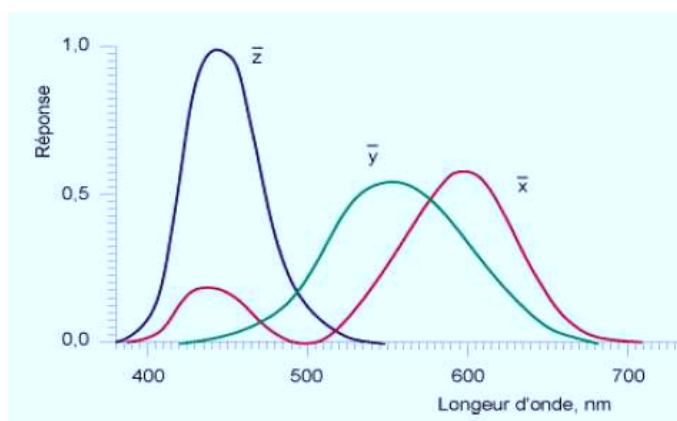


FIG. 2.3 – Sensibilité spectrale des cônes.

L'espace RGB (Red-Green-Blue) défini par la compagnie internationale de l'éclairage (C.I.E.) en 1931 représente les couleurs par synthèse additive. Chaque couleur est représentée par trois composantes : le rouge (R), le vert (G) et le bleu (B) dont les longueurs d'onde normalisées sont respectivement $\lambda_R = 700\text{nm}$, $\lambda_G = 546,1\text{nm}$ et $\lambda_B = 435,8\text{nm}$. En général, les trois composantes sont représentées sur l'ordinateur par trois octets (24 bits). Avec ce choix pour le stockage des couleurs, on obtient une palette de 256^3 soit 16 777 216 couleurs possibles. Le système RGB est représenté par un cube (figure 2.4) dont les sommets sont les couleurs de base, noir, blanc, rouge, vert, bleu, cyan, magenta, jaune ; le noir et

le blanc étant diamétralement opposés.

$$C = [R, V, B] \text{ avec } R \in [0, 255], G \in [0, 255] \text{ et } B \in [0, 255]$$

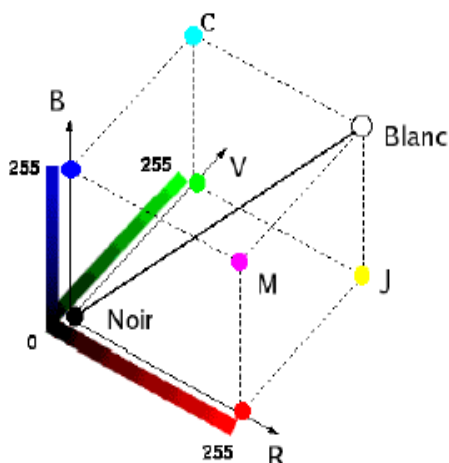


FIG. 2.4 – Cube de l'espace couleur RGB.

Le principal inconvénient de cet espace de représentation vient du fait qu'il ne tient pas compte de la variation de la sensibilité de l'oeil.

La C.I.E. a défini un espace de représentation qui prend en compte la sensibilité de l'oeil. Il s'agit de l'espace XYZ. La figure 2.5 donne ce diagramme qui permet de tenir compte de la sensibilité de l'oeil aux différents stimuli.

L'espace XYZ est rarement utilisé dans les recherches d'images car il n'est pas perceptuellement uniforme. C'est-à-dire que les différentes nuances d'une même teinte (par exemple rouge) peuvent être très éloignées dans l'espace. On lui préfère d'autres espaces qui ont des propriétés plus favorables au traitement informatique de la couleur. L'espace HSV (Hue-Saturation-Value ou teinte, saturation et valeur (luminance)) possède une composante de luminance, une composante de teinte qui correspond à la couleur et une saturation de cette teinte (qui correspond à la pureté de la couleur). Il est représenté par un cône (figure 2.6) dont les sommets sont le noir et le blanc. Cet espace présente l'avantage de simuler le comportement visuel humain dans lequel la couleur est décomposée en une image de luminance (la valeur) et une image de chrominance (teinte et saturation). La teinte fournit directement l'information de la couleur dominante dans les différentes régions de l'image.

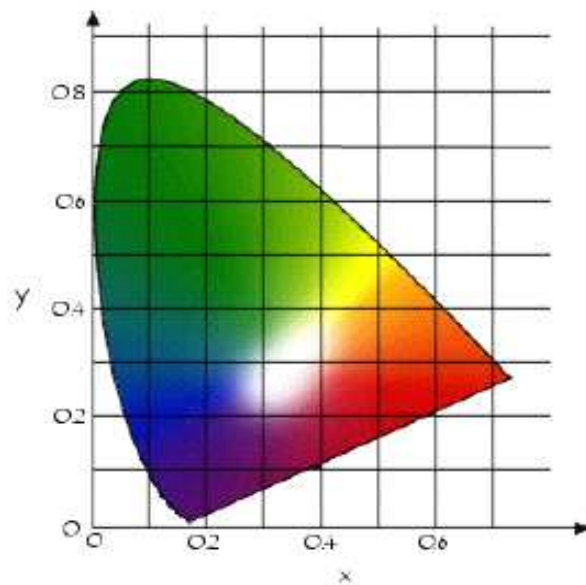


FIG. 2.5 – Représentation de l'espace couleur XYZ.

L'espace HSV n'est malheureusement pas uniforme, c'est-à-dire qu'une distance entre couleurs visuellement proches calculée dans cet espace peut être très grande. Un autre espace a donc été créé par la C.I.E. pour résoudre cette difficulté.

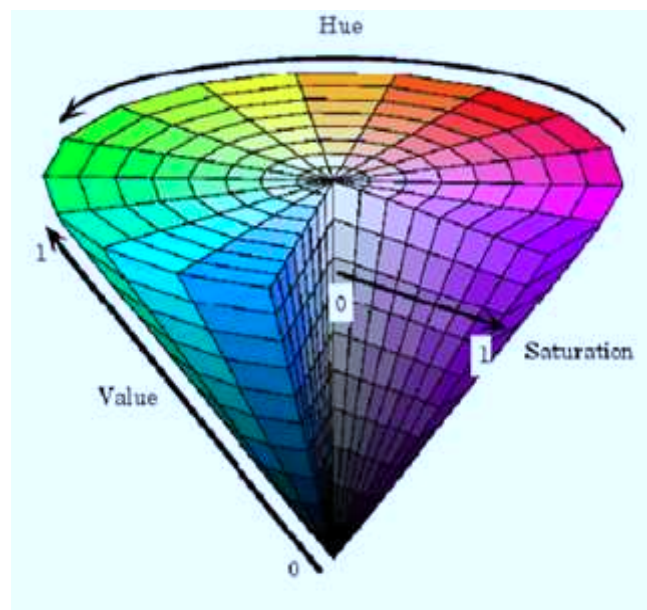


FIG. 2.6 – Représentation de l'espace couleur HSV.

L'espace L^*a^*b est un espace couleur perceptuellement uniforme, c'est-à-dire qu'il possède la bonne propriété de respecter les distances entre couleurs visuellement proches. Il est défini à partir de l'espace XYZ par des relations non-linéaires et par le choix d'un blanc de référence.

L'espace de couleur L^*a^*b a été créé afin de fournir une expression quantitative de classification des couleurs dans le système de Munsell qui fait partie des espaces de représentation de couleurs en teinte, en luminance, et en saturation (TLS). Le modèle de Munsell est composé de trois axes : hue (teinte), chroma (saturation) et value (luminosité). La principale différence avec les systèmes actuels de type TLS est que les axes sont faiblement échantillonnés. En d'autres termes, chaque axe est représenté par un faible nombre de valeurs (ex. : 10 pour la teinte). Ainsi, la teinte est représentée par un disque divisé en 10 secteurs comme le montre la figure 2.7. Ces valeurs correspondent aux teintes suivantes : rouge R, jaune-rouge YR, jaune Y, jaune-vert GY, vert G, bleu vert BG, bleu B, bleu-pourpre PB, pourpre P et rouge-pourpre RP. Un secteur peut être subdivisé en sous-sections pour une meilleure précision. Dans ce cas, un numéro précèdera le label donné à la section (3YR par exemple).

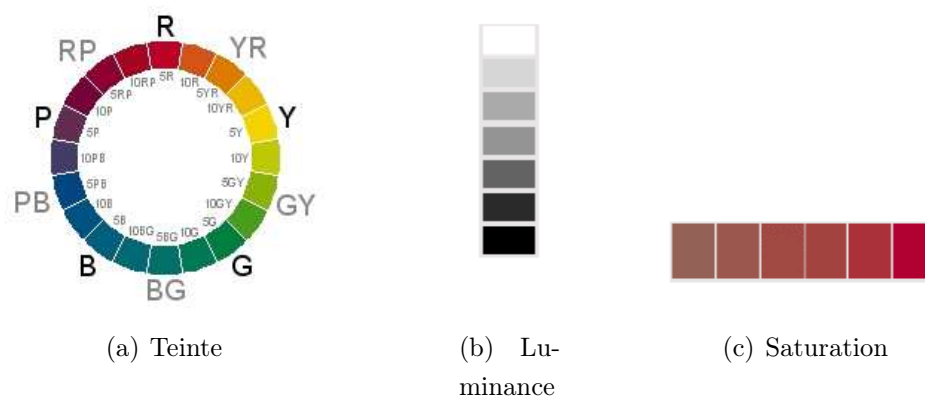


FIG. 2.7 – *Espace de couleur Munsell.*

Histogramme de couleur

Les histogrammes de couleur (RVB, Lab, HSV, etc.) sont souvent utilisés pour rechercher des images similaires en se basant sur la distribution générale des couleurs de l'image requête. L'histogramme contient les fréquences d'occurrence (c.-à-d. le nombre de pixels) des couleurs contenues dans l'image. Plus une couleur de l'image sera prédominante, plus sa fréquence sera forte dans l'histogramme (cf. Figure 2.8).

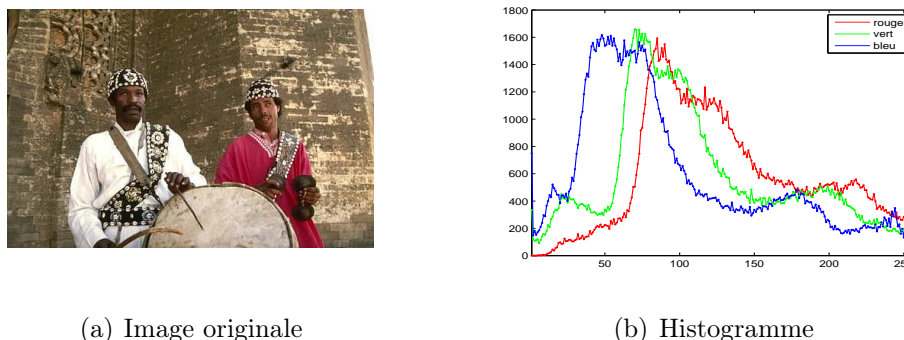


FIG. 2.8 – Exemple d’histogramme d’une image couleur

Notons que l’utilisation d’un histogramme entraîne un compromis entre la vitesse et la précision : plus l’histogramme contient d’intervalles, plus les comparaisons d’histogrammes prendront de temps, mais plus aussi les résultats deviendront discriminants.

Les histogrammes sont normalisés avant d’effectuer les comparaisons, tels que leur aire soit unitaire (division par le nombre de pixels de l’image). Ainsi, la taille de l’image requête n’affecte pas le processus de comparaison. La méthode est bonne lorsque la requête est une image connue et que l’on souhaite retrouver les images ayant une distribution de couleur similaire. Elle n’est pas adaptée aux requêtes basées sur des extraits d’images. On note par ailleurs l’influence de la couleur du fond sur les histogrammes dans le cas particulier d’une requête où l’image représente un objet 3D. Aussi, l’histogramme ne contient-il aucune information spatiale concernant la localisation d’un objet, une extension consiste à utiliser un corrélogramme [58]. Celui-ci se définit comme un histogramme à trois dimensions, dans lequel les deux premières dimensions représentent tous les couples de couleurs possibles et la troisième traduit la distance spatiale entre les couleurs concernées.

Toutefois, ces approches restent assez limitées à des requêtes bas niveau, et sont assez sensibles aux conditions d’éclairage ou aux ombres portées sur les objets

Les moments statistiques

La méthode d’histogramme utilise la distribution complète de la couleur. Les données stockées sont d’une taille assez grande, cela nous prend beaucoup de temps et aussi de mémoire. Donc pour résoudre ce problème, au lieu de calculer la distribution complète dans les systèmes de recherche d’images, on calcule seulement des dominantes caractéristiques de couleur tel que l’espérance, la variance

et d'autres moments. Dans [130] les auteurs ont prouvé que les méthodes utilisant des moments statistiques sont plus rapides et donnent des résultats meilleurs que les méthodes d'histogrammes.

On peut calculer l'espérance, la variance, les moments sur chaque composante couleur par la formule suivante :

$$E = \frac{1}{N} \sum_{j=1}^N p_j \quad (2.1)$$

$$\delta = \left(\frac{1}{N} \sum_{j=1}^N (p_j - E)^2 \right)^{1/2} \quad (2.2)$$

$$S = \left(\frac{1}{N} \sum_{j=1}^N (p_j - E)^3 \right)^{1/3} \quad (2.3)$$

Où E est l'espérance, δ est la variance, S est le troisième moment.

2.7.2 La texture

La notion de texture est utilisée pour traduire un aspect homogène de la surface d'un objet sur une image. La texture se manifeste donc par une information visuelle qui permet de la décrire qualitativement à l'aide des adjectifs suivants : grossière, fine, lisse, tachetée, granuleuse, marbrée, régulière ou irrégulière. Même s'il n'existe pas de définition universelle de la texture, elle est généralement conçue comme une structure spatiale constituée de l'organisation de primitives ayant chacune un aspect aléatoire [43]. De nombreuses méthodes sont référencées dans la littérature pour la décomposition de l'image et le calcul de caractéristiques dites de texture. Parmi les plus connues, on trouve la décomposition paramétrique de Wold utilisée par Liu et Picard [73] dans le système Photobook [98]. De façons similaires, les caractéristiques de texture utilisées dans le système QBIC [95] sont le grossièreté (coarseness), le contraste et l'orientation. Les travaux de Haralick sur la texture, avec en particulier les matrices de cooccurrences et les différents indices qui en découlent, sont largement utilisés en indexation [49]. Les modèles autorégressifs multi échelle MRSAR sont aussi couramment employés, au même titre que les filtres de Gabor [82] [53] et les ondelettes. Elles permettent, de plus, d'envisager l'indexation directement sur les données compressées [92].

Il existe deux types de signatures pour caractériser la texture [6] :

- La première est déterministe et fait référence à une répétition spatiale d'un motif de base dans différentes directions. Cette approche structurale cor-

respond à une vision macroscopique des textures. La texture de la peau de zèbre est un exemple macroscopique dans lequel la primitive est la zébrure.

– La seconde est probabiliste et cherche à caractériser l’aspect chaotique qui ne comprend ni motif localisable, ni fréquence de répétition principale. Cette approche correspond à une vision microscopique de la texture.

Les méthodes d’extraction des attributs dépendent du modèle de texture recherché. Pour les textures macroscopiques, l’accent est mis davantage sur les répétitions d’un motif dans une certaine direction à l’aide des matrices de co-occurrence. Pour les textures microscopiques, les techniques de recherche sont basées sur des calculs statistiques sur les niveaux de gris des images. L’analyse en ondelettes et filtres de Gabor permettent de calculer des attributs de texture mixtes [6].

La matrice de cooccurrence

La texture d’une image peut être interprétée comme la régularité d’apparition de couples de niveaux de gris selon une distance donnée dans l’image. La matrice de cooccurrence des niveaux de gris (GLCM) [48, 102] contient les fréquences spatiales relatives d’apparition des niveaux de gris selon quatre directions : $\theta = 0$, $\theta = \pi/2$, $\theta = \pi/4$ et $\theta = 3\pi/4$. La matrice de cooccurrence est une matrice carrée $n \times n$ où n est le nombre de niveaux de gris de l’image. On définit la matrice des fréquences relatives F par :

$$F(d, \theta) = (f(i, j/d, \theta)) \quad (2.4)$$

$f(i, j/d, \theta)$ représente le nombre de fois où un couple de points séparés par la distance d dans la direction θ a présenté les niveaux de gris i et j . Pour obtenir de véritables fréquences relatives, il faut normaliser les éléments de la matrice en les divisant par le nombre total de paires de points élémentaires séparés par la distance d dans la direction θ dans toute l’image. À partir de la matrice de cooccurrence, on extrait différents attributs de texture. 14 mesures statistiques ont été initialement proposées dans [48]. Des études comparatives ont montrés que seulement une partie de ces mesures ont donnés de bonnes performances et qui sont communément utilisées [89, 26, 154] :

L’énergie :

$$E = \sum_{i=1}^n \sum_{j=1}^n (f(i, j/d, \theta))^2 \quad (2.5)$$

L’entropie :

$$ENT = \sum_{i=1}^n \sum_{j=1}^n f(i, j/d, \theta) \log(f(i, j/d, \theta)) \quad (2.6)$$

Le contraste :

$$C = \sum_{i=1}^n \sum_{j=1}^n (i-j)^2 f(i, j/d, \theta) \quad (2.7)$$

L'homogénéité :

$$HOM = \sum_{i=1}^n \sum_{j=1}^n \frac{f(i, j/d, \theta)}{1 + |i-j|} \quad (2.8)$$

La transformée de Fourier discrète

De nombreuses méthodes d'extraction d'attributs de texture sont basées sur la transformée de Fourier. Cette dernière permet de passer du domaine spatial de l'image au domaine fréquentiel.

Puisque les textures sont la répétition d'un motif, elles seront caractérisées par des pics dans la représentation fréquentielle de l'image (les textures fines dans les fréquences hautes et les textures grossières dans les basses fréquences). Etant donnée une image $i(x, y)$, la transformée de Fourier $F(f_x, f_y)$ s'exprime de la façon suivante :

$$F(f_x, f_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} i(x, y) e^{-2i\pi(f_x x + f_y y)} dx dy \quad (2.9)$$

f_x Et f_y sont appelées les fréquences spatiales de l'image $i(x, y)$. On a, en outre, la relation inverse, dite de décomposition de la fonction image $i(x, y)$ sur une base de fonctions de Fourier :

$$i(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(f_x, f_y) e^{2i\pi(f_x x + f_y y)} dx dy \quad (2.10)$$

L'idée consiste donc à décrire l'image $i(x, y)$ comme étant un signal à deux variables spéciales x et y . Ce signal est décomposé comme une somme pondérée de fonctions sinusoïdales (équation 2.10). Le poids de chacune de ces fonctions dans la décomposition est rendu par le coefficient de Fourier $F(f_x, f_y)$. Comme on travaille dans le plan complexe, pour éviter d'interpréter la phase, on utilise souvent le module de la transformée de Fourier ou spectre de Fourier. Le spectre est très riche en information, il permet de déterminer des attributs dans le plan fréquentiel de l'image.

La figure 2.9 présente des exemples de textures, fines et grossières ainsi que les modules $|F(f_x, f_y)|$ des transformées de Fourier associées.

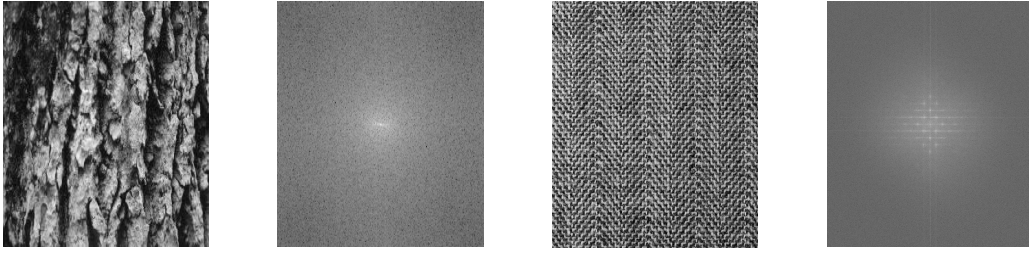


FIG. 2.9 – Exemple de textures et de module de la transformée de Fourier associée

Les descriptions de type Fourier décomposent l'image sur les fonctions sinusoidales, qui ont la propriété de s'étendre indéfiniment dans l'espace. Une telle description est donc adaptée à des signaux périodiques ou quasi-périodiques, mais trouve sa limitation pour des signaux plus complexes, présentant de fortes discontinuités. Or, les images présentent justement ce type de caractéristiques. C'est pourquoi d'autres décompositions ont été utilisées, comme celle dite de Gabor.

Filtres de Gabor

Outre son efficacité établie en reconnaissance de formes texturées et pour la modélisation des champs récepteurs des cellules du cortex visuel, différentes comparaisons réalisées dans le cadre de la recherche d'images, placent cette approche parmi les techniques les plus efficaces [157].

L'idée est de décomposer l'image sur les fonctions $G_{n,m}(x, y)$ limitées dans l'espace, afin d'en étudier indépendamment les fragments. On qualifie les $G_{n,m}(x, y)$ de fonctions analysantes. Dans le cas de la décomposition de Gabor, ces fonctions analysantes sont obtenues à partir d'une fonction mère $G(x, y)$, définie comme une fonction sinusoidale orientée sur l'axe des x , modulée par une enveloppe gaussienne dans les directions x et y .

$$G(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right) \cdot \exp(2i\pi f_x x) \quad (2.11)$$

Les fonctions $G_{n,m}(x, y)$ sont ensuite obtenue par dilatation et rotation de $G(x, y)$:

$$G_{n,m}(x, y) = a^{-m} G(x', y') \quad (2.12)$$

$$x' = a^{-m} \left(x \cos\left(\frac{n\pi}{K}\right) + y \sin\left(\frac{n\pi}{K}\right) \right) \quad (2.13)$$

$$y' = a^{-m} \left(-x \sin\left(\frac{n\pi}{K}\right) + y \cos\left(\frac{n\pi}{K}\right) \right) \quad (2.14)$$

Où a et K sont des paramètres à fixer.

L'image i est alors analysée par cette série de fonctions par convolution successives, conduisant à une série de coefficients $w_{n,m}(x, y)$:

$$w_{n,m}(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G_{n,m}^*(x - a, y - a) i(a, b) da db \quad (2.15)$$

On parle de décomposition espace-fréquence, car les fonctions analysantes $G_{n,m}(x, y)$ sont appliquées localement à l'image (espace) et à différentes fréquences du fait des dilatations et rotations. Cette décomposition a été utilisée par Manjunath et al. [82] pour des indexations par les textures. Torralba et Oliva [147] utilisent également cette décomposition sur filtres de Gabor, afin de pouvoir classifier des images en familles sémantiques (paysages naturelles, bâtiments, etc...), d'après les valeurs de coefficients $w_{n,m}(x, y)$.

Les filtres de Gabor sont relativement utilisés aujourd'hui, notamment du fait de leur pertinence en regard du système visuel humain. En effet, Marcelja [83] a montré que les cellules du cortex humain pouvaient être modélisées par les fonctions de Gabor à une dimension. Daugman [30] a élargi ce modèle à deux dimensions.

Les principaux inconvénients de cette méthode résident dans le choix des paramètres pour déformer la fonction mère analysante. Aussi, la description obtenue via les coefficients $w_{n,m}(x, y)$ pourrait être redondante.

Nous utilisons principalement un banc de 12 filtres (CF. figure 2.10), avec quatre orientations (horizontales, verticales et diagonales), et trois fréquences (basses et moyennes fréquences privilégiées).

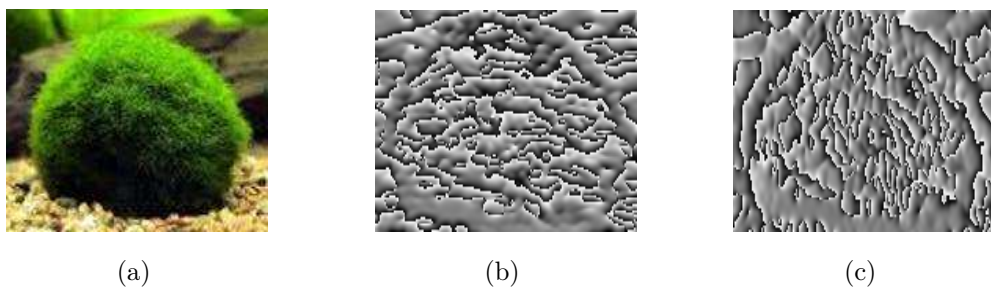


FIG. 2.10 – *Illustration de l'extraction de caractéristiques de texture par filtrage de Gabor : Présentation des images filtrées par deux des filtres avec même fréquence (0.8) et deux orientations : horizontale(b) et verticale(c).*

Les ondelettes

L'idée fondatrice [81] est la même que pour les filtres de Gabor : décomposer l'image sur des fonctions $\Psi_{a,b}(x)$ limitées dans l'espace, afin d'en étudier

les fragments indépendamment. Pour simplifier les notations, mais sans limitations aucune, plaçons nous dans le cas d'un signal à une dimension. Une analyse par ondelettes d'une fonction $i(x)$ consiste à projeter celle ci sur une famille de fonctions $\Psi_{a,b}(x)$, appelées fonctions analysantes :

$$W(a,b) = \int_{-\infty}^{+\infty} i(x)\Psi_{a,b}(x)dx \quad (2.16)$$

Ces fonctions analysantes $\Psi_{a,b}(x)$ sont obtenues à partir de la déformation d'une fonction mère $\Psi(x)$, telle que :

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}}\Psi\left(\frac{x-a}{b}\right) \quad (2.17)$$

Le paramètre a permis d'appliquer la fonction analysante en différents points de la fonction i (translation) et le paramètre b permet d'appliquer la fonction analysante à différentes échelles (dilatation de la fonction mère). La figure 2.11 présente deux déformées des ondelettes de Haar (en rouge). La fonction mère est affichée en bleue.

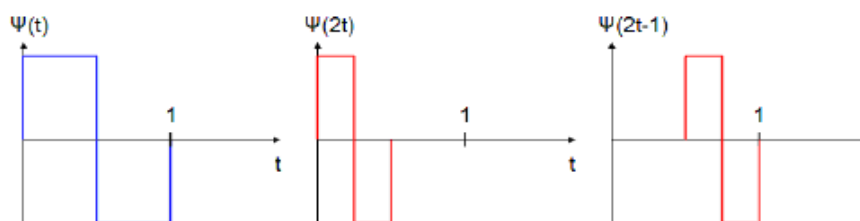


FIG. 2.11 – *Ondelettes de Haar*

En pratique, la description par ondelettes d'une image, c'est-à-dire d'un signal à deux dimensions, traite séparément chacune des composantes horizontales (x), verticales (y) et diagonales (x et y), sous échantillonnées d'un facteur 2 par rapport à l'image originale. Il en résulte un schéma de décomposition à cadrans, comme présenté dans la figure 2.12. Par exemple, lorsque la composante y est traitée, on obtient une image sous-échantillonnée d'un facteur 2, qui contient les détails verticaux (hautes fréquences, coin inférieur gauche). De même, le traitement de la composante x donne une image sous échantillonnée qui contient les détails horizontaux. L'image originale peut alors être reformée, à partir des composantes basses fréquences sur x et y (approximation grossière de l'image, coin supérieur gauche) auxquelles on ajoute les détails hautes fréquences des trois

autres cadrans. L'opération ainsi répétée est appelée analyse multi résolution (AMR) par ondelettes.

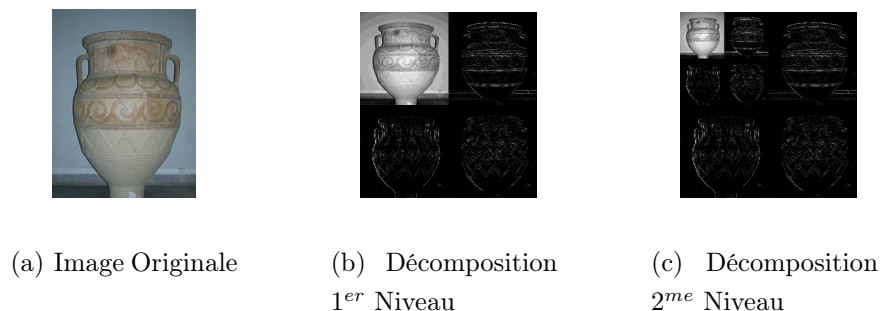


FIG. 2.12 – *Décomposition par ondelettes de Haar.*

À partir de la transformée en ondelettes on peut extraire des attributs de différents types et à différents niveaux de résolution. L'image d'approximation donne des informations sur les régions qui composent l'image, d'une résolution fine à une résolution grossière. Les images de détails donnent des informations horizontales, verticales et diagonales sur l'image

Les moments des coefficients d'ondelettes sont directement disponibles. On a ainsi des mesures à différents niveaux de résolution pour décrire la texture.

2.7.3 La forme

La forme est un descripteur important dans les bases d'images. L'objectif de la description par la forme est d'extraire une caractéristique géométrique pour coder un profil, estimer la ressemblance et reconnaître un objet [47]. Pour la recherche d'images par le contenu, la forme d'un objet devra être invariante aux transformations affines : translation, rotation et facteur d'échelle [87]. Généralement les descriptions du profil d'un objet se divisent en deux catégories : les contours et les régions. Les méthodes basées sur les contours comprennent des descripteurs comme les chaînes de Freeman [38] et la méthode CSS [88]. Pour la caractérisation des régions, les moments invariants [57] et méthode ART [65] sont souvent utilisés comme descripteurs. Dans cette thèse, la forme n'est pas utilisée, vu que nous sommes intéressés à l'étude des macros algues où la recherche est effectuée à l'aide d'attributs issus de la couleur et de la texture.

2.8 Segmentation et points d'intérêt

L'extraction des descripteurs globaux sur l'image entière permet de réduire le nombre de calcul nécessaire, la taille de la base des signatures ainsi que le

coût de recherche des images les plus similaires. Cependant, l'approche globale ne permet pas une recherche efficace d'objets dans l'image. En revanche, les descripteurs extraits localement sont efficaces, mais coûteux. Les descripteurs locaux peuvent être :

- (1) Des régions de l'image obtenues soit par segmentation de l'image entière, soit par recherche de régions d'intérêt
- (2) Des points d'intérêt

2.8.1 Segmentation

La meilleure approche pour interpréter une image sémantiquement reste l'utilisation d'une forte segmentation de la scène. La segmentation forte automatique est cependant difficile à atteindre [118].

L'humain sait naturellement séparer des objets dans une image. Pour cela il se base notamment sur des connaissances de haut niveau qui lui permettent de détecter ce qui l'intéresse dans l'image. En traitement d'image, on caractérise une région comme étant un ensemble de pixels ayant des propriétés commune d'intensité, de texture, de couleur,..., qui la différencie des régions voisines. Il y a plusieurs méthodes de segmentation, on distingue cependant deux grandes familles d'algorithmes :

- (1) **La segmentation par contour** (*edge-based segmentation*). Un contour est une frontière entre deux milieux différents. On utilise souvent l'image dérivée pour mettre en évidence les variations de niveau. Quand il y a une forte variation, cela est souvent dû à un changement de teinte ; d'éclaircissement ou de texture, et de façon générale à un changement de régions. Un algorithme classique d'extraction des contours est l'algorithme de Canny [14].
- (2) **La segmentation par région** (*region based segmentation*). Le principe est de trouver les régions en regroupant les pixels ayant des caractéristiques similaires et en séparant ceux qui sont différents (techniques de clustering, méthodes spectrales,...). Figure 4.4 un exemple d'image segmentée en utilisant NCUT [116].



FIG. 2.13 – *Exemple de segmentation par approche région [116]*

Une des difficultés de la segmentation est de savoir en combien doit être découpée une image, car certaines images peuvent être segmentée en moins de régions e.g. un portrait sur un fond uniforme, alors que d'autre nécessite plus de régions.

Des études [111] ont été faites pour comparer les méthodes classiques de segmentation : Mean shift, Clustering, approche morphologique. Leurs résultats montrent qu'il n'existe pas de segmentation générique ou meilleure que les autres, mais qu'en fonction de la tâche à accomplir certaines méthodes sont plus efficaces que d'autres.

Dans [5], les auteurs proposent une étude comparative des techniques de segmentation pour la reconnaissance d'objets dans les images. Pour cette tâche, il est nécessaire d'obtenir des segmentations qui ne découpent pas un objet en plusieurs régions. Une partie du ciel ou d'un champ peut être séparée en plusieurs régions. Par exemple dans la figure 2.13, il est difficile de mélanger les parties blanches et rouges du bus. Cependant, les régions obtenues à partir de segmentation sur des images décrites par des traits visuels de bas niveau ne correspondent pas souvent à des objets sémantiques. Les quatre algorithmes de segmentation comparés sont : Expectation Maximisation (EM) [15], l'algorithme normalized cuts [116] et l'algorithme mean shift [25]. Les résultats montrent que l'algorithme normalized cuts donne de meilleurs résultats que mean shift et que la méthode EM de Blobworld.

2.8.2 Les points d'intérêt

Les points d'intérêt d'une image sont les points où le signal change. Ces points peuvent être les coins, les jonctions ou les points de forte variation de

texture. Les points d'intérêt seront similaires dans les images similaires. Pour extraire ces points, il existe 3 types d'approches :

- (1) Approches contours : les contours d'une image sont d'abord détectés, puis les points d'intérêt sont extraits le long des contours en considérant les points de courbures maximales ainsi que les intersections des contours.
- (2) Approches intensité : la fonction d'intensité est utilisée pour extraire directement les points de discontinuités des images
- (3) Approches à base de modèle : les points d'intérêts sont identifiés dans l'image par mise en correspondance de la fonction d'intensité avec un modèle théorique

Les approches intensité sont les plus utilisées, car elles sont indépendantes des contours et du type de points d'intérêt. On peut citer au moins deux avantages des points d'intérêts par rapport aux régions : ils ne nécessitent pas d'enchaînement pour détecter les contours des régions, et on peut les extraire efficacement de la plupart des images. Le détecteur de points d'intérêt le plus populaire est le détecteur de Harris [50]. Ce détecteur étant seulement applicable aux images en niveau de gris, Montesinos et al. [90] l'ont généralisé pour les images couleur. Les points d'intérêt produits par ce détecteur sont définis comme étant les extrema locaux positifs de l'image intermédiaire en niveau de gris :

$$P(x, y) = Det(M(x, y)) - k.trace^2(M(x, y)) \quad (2.18)$$

Où k vaut typiquement 0.04 et $M(x, y)$ est une matrice définie comme suit :

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12} & M_{22} \end{pmatrix} \quad (2.19)$$

construite en chaque point de l'image en fonction des trois images intermédiaires en niveau de gris $M_{11}(x, y)$, $M_{12}(x, y)$ et $M_{22}(x, y)$ définies par :

$$\begin{aligned} M_{11}(x, y) &= G_{\tilde{\sigma}} \otimes (r_x^2 + g_x^2 + b_x^2) \\ M_{12}(x, y) &= G_{\tilde{\sigma}} \otimes (r_x r_y + g_x g_y + b_x b_y) \\ M_{22}(x, y) &= G_{\tilde{\sigma}} \otimes (r_y^2 + g_y^2 + b_y^2) \end{aligned}$$



FIG. 2.14 – *Exemple des points d'intérêt en utilisant Harris Couleurs*

Où $G_{\tilde{\sigma}}$ est une gaussienne 2D isotropique de variance $\tilde{\sigma}^2$, C_x et C_y représentent les dérivées gaussiennes du premier ordre des canaux c de l'image originale $I(x, y)$ avec $c \in \{r, g, b\}$. Ces dérivées sont implémentées en utilisant une gaussienne de variance σ^2 . Selon les comparaisons faites par Gouet et Boujemaa [46], le détecteur ci-dessus apparaît comme étant le plus sûr parmi les détecteurs de points d'intérêt couleur connus, et ce par rapport aux changements d'illumination, au bruit, aux rotations et aux changements de point de vue. La figure 2.14 donne un exemple de points d'intérêt détectés dans une image d'algues.

2.9 Mesure de similarité

Dans cette section, nous présentons les principales distances existantes pour établir la similarité entre les attributs caractérisant les images. La notion de similarité est un des aspects les plus difficiles de la recherche d'images par le contenu. L'une des raisons est la perception subjective de l'homme. Considérons l'exemple suivant : Quelqu'un veut trouver toutes les images similaires à l'image, qui montre des bâtiments dans une scène urbaine. La seconde personne veut par exemple trouver les bâtiments de même couleur. Une autre personne pourrait vouloir trouver la même architecture, abstraction faite de la couleur. Une quatrième personne peut être n'est pas intéressée par les bâtiments, mais par l'arrière plan, etc. En effet une image peut conduire à un ensemble d'images jugées similaires, dépendamment de l'intention de l'utilisateur. Cette perception subjective peut être traitée en utilisant des méthodes interactives pour comprendre le profil de l'utilisateur.

La mesure de similarité est souvent métrique, les images sont considérées ressemblantes si la distance est faible. La complexité de calcul d'une distance doit

être raisonnable parce que dans un système de recherche d'images cette tâche s'exécute en temps réel. D'autres paramètres entrent en jeu telle la dimension de l'espace caractéristique, la taille de la base, ... La méthode basique de recherche calcule la distance entre la requête et toutes les images de la base puis les ordonne selon leurs scores. Ceci par conséquent rend le temps de réponse proportionnel au nombre d'images. Les méthodes d'indexation multidimensionnelles permettent par ailleurs de réduire cette complexité comparée à la recherche séquentielle.

Nous n'avons pas l'intention de décrire ici toutes les mesures en détail. En général, chaque image dans une base de données est représentée par un n-dimensionnel vecteur $v = \{v_1, v_2, \dots, v_n\}$. Ensuite, la similarité entre un vecteur caractéristique de l'image requête ($q = q_1, q_2, \dots, q_n$), et les vecteurs caractéristiques des images dans la base de données v est donnée par une métrique. Nous listons ici les métriques les plus utilisées :

Distances de Minkowski

Une famille de métriques très utilisée est la famille des distances de Minkowski. Dans un espace de dimension finie, la forme générale de la famille de distances de Minkowski s'écrit :

$$d_r(q, v) = \left(\sum_{i=1}^n |q_i - v_i|^r \right)^{1/r} \quad (2.20)$$

Avec $r \geq 1$

Où n est la dimension du vecteur et r l'ordre de la distance.

Parmi cette famille de distances, trois d'entre elles possèdent des propriétés intéressantes :

- (1) La distance de Manhattan (ou city block), $r = 1$,

$$d_c(q, v) = \left(\sum_{i=1}^n |q_i - v_i| \right) \quad (2.21)$$

- (2) La distance Euclidienne, $r=2$;

$$d_e(q, v) = \left(\sum_{i=1}^n |q_i - v_i|^2 \right)^{1/2} \quad (2.22)$$

La distance quadratique

La distance de Minkowski traite les éléments du vecteur de caractéristiques d'une manière équitable. La distance quadratique en revanche favorise les éléments les plus ressemblants. Sa forme générale est donnée par :

$$d_Q(q, v) = \sqrt{(q_i - v_i)^t A (q_i - v_i)} \quad (2.23)$$

Où $A = [a_{ij}]$ est la matrice de similarité, a_{ij} représente la distance entre deux éléments des vecteurs q et v . Pour construire la matrice A , Hafner et al. [56] propose la formule suivante :

$$a_{ij} = 1 - \frac{d_{ij}}{\max(d_{ij})} \quad (2.24)$$

Où d_{ij} est la distance euclidienne entre les classes de couleur dans un espace de Munsell.

Les propriétés de cette distance la rendraient proche de la perception humaine de la couleur, ce qui est en fait une métrique attractive dans les systèmes CBIR [40].

Intersection d'histogrammes

L'intersection d'histogrammes a été proposée par Swain et Ballard [132] : Elle mesure la partie commune entre deux histogrammes.

$$d(h_1, h_2) = \frac{\sum_i \min(h_1(i), h_2(i))}{\sum_i h_2(i)} \quad (2.25)$$

Deux images présentant une intersection normalisée d'histogrammes proche de 1 sont considérées comme similaires.

2.9.1 Combinaison des descripteurs

Vu la richesse du contenu de l'image, un seul descripteur n'est pas en mesure de saisir toutes les informations nécessaires pour une représentation adéquate de l'image. En effet, des complémentarités entre les descripteurs sont utilisées pour maximiser la quantité d'informations du contenu de l'image. Ainsi, les résultats de recherche peuvent être améliorés si plus d'un descripteur est employé dans un système CBIR [64]. Plusieurs possibilités existent dans la littérature pour combiner les descripteurs d'une image. Le reste de cette section passe en revue quelques méthodes existantes pour les calculs de similarité de plusieurs descripteurs.

Un seul vecteur

Une manière de combiner les différents descripteurs est de les empiler comme un seul vecteur et de calculer la distance entre ces deux vecteurs dans un espace de caractéristiques.

Combinaisons linéaires

Au lieu d'empiler tous les descripteurs dans un seul vecteur, la similarité entre les images peut être calculée directement à partir de la combinaison de caractéristiques. Dans ce cas, on calcule les distances entre les descripteurs de même nature. Ces distances sont combinées pour obtenir la similarité globale entre images. PicHunter [29], qui emploie la distance de Manhattan pour déterminer la similarité entre les vecteurs caractéristiques, obtient une distance totale entre deux images en combinant les distances entre caractéristiques en utilisant une moyenne pondérée. Dans cette approche, la même mesure est employée pour obtenir la distance entre les descripteurs de la même nature.

Approches booléennes

Les systèmes tels que MARS [96] ont essayé de surmonter les limitations structurelles de la moyenne pondérée par adoption de la notion booléenne utilisée dans la recherche textuelle. MARS [96] représente la requête de l'utilisateur comme une expression booléenne floue des attributs de l'image et la similarité entre les images devient une évaluation de cette expression. Garlic [36] utilise min et max comme opérateurs de la logique floue AND et OR, respectivement, pour s'adresser aux requêtes conjonctives et disjonctives. Bien que les approches basées sur les expressions booléennes peuvent modéliser un large spectre de requêtes. Pourtant, la prise de décision basée sur l'agrégation des critères n'est généralement pas conjonctive ou disjonctive [101].

2.10 Indexation Multidimensionnelle

L'indexation multidimensionnelle débuta originellement avec les quad-tree développés pour les bases de données traditionnelles. QBIC [95, 40] fut le premier système à utiliser ce genre d'indexation pour diminuer les temps de recherche, où les paramètres de texture et de couleur sont indexés par un R-tree [7] qui partitionne hiérarchiquement un espace multidimensionnel. Ainsi les objets sont représentés par l'indice du groupe auquel ils appartiennent. Ainsi, lors d'une recherche, les similarités des groupes avec la requête sont évaluées. Comme le nombre de groupes est moins important que le nombre d'objets, la recherche devient moins complexe. D'autre part l'utilisation d'arbres, tels que R-tree, permet de diminuer encore plus la complexité de recherche en diminuant le nombre de comparaisons hiérarchiquement.

2.11 Retour de pertinence

Initialement utilisé sur les documents textuels [114], Le bouclage de pertinence (Relevance Feedback) est actuellement très utilisé dans le domaine de la recherche d'images par le contenu, car il permet à l'utilisateur de raffiner sa requête en fournissant au système plusieurs exemples de ce qu'il souhaite obtenir. Le bouclage de pertinence suppose que le jugement de l'utilisateur est plus pertinent que le jugement du système. Cette technique est donc un moyen d'apporter de la sémantique à la recherche d'images. Cependant, elle nécessite que l'utilisateur indique au système quelles images l'intéressent. Un bon système de bouclage de pertinence est donc un système qui propose à l'utilisateur les images qu'il recherche en peu d'itérations.

Un système classique de bouclage de pertinence comporte deux composants : un système de sélection et un système d'apprentissage. Pour rechercher les images, un utilisateur réalise d'abord une première requête. Le système lui renvoie les premiers résultats. Puis parmi les résultats retournés, l'utilisateur sélectionne certains exemples comme pertinents et d'autres comme non pertinents. Le système prend alors en compte la sélection de l'utilisateur pour réaliser un apprentissage et retourner des résultats plus pertinents. L'utilisateur peut alors sélectionner d'autres exemples et réitérer le processus jusqu'à ce qu'il obtienne des résultats satisfaisants ou bien jusqu'à ce que les résultats fournis se stabilisent.

Un état de l'art et contributions concernant les techniques de bouclage de pertinence pour la recherche d'images sont explicités dans le chapitre 5.

2.12 Quelques systèmes de recherche d'images

La recherche dans le domaine de traitement d'image a sensiblement progressé au cours des dernières années et a abouti à un certain nombre de systèmes CBIR commerciaux et de recherche. Cette section décrit quelques-uns des principaux systèmes CBIR.

QBIC

IBM a développé un système appelé QBIC (Query By Image Content) [95, 40]. Il procède à une extraction des caractéristiques simples à partir d'images : La couleur, la texture et la forme. Les attributs de la couleur sont calculés en utilisant la moyenne de la couleur dans les espaces RGB, YIQ, L^*a^*b et Munsell et les histogrammes de couleur dans l'espace RGB. La texture est caractérisée par des versions modifiées des attributs de Tamura [144]. La forme est décrite par la

circularité, l'excentricité, l'orientation des principaux axes et les moments invariants. Le QBIC a également mis en oeuvre une méthode de recherche d'images basée sur croquis de l'utilisateur. L'interface utilisateur de QBIC permet aux utilisateurs d'effectuer une variété de recherches fondées sur les principes de propriétés visuelles. Les images peuvent être recherchées par les pourcentages de propriétés telles que la couleur, la forme et la texture qui se produisent dans les images, par exemple 40% de vert, 20% de blanc. QBIC permet également des recherches en utilisant des mots-clés combinés avec les caractéristiques visuelles dans une requête unique.

MARS

MARS (Multimedia Analysis and Retrieval System) [113] élaboré à l'université d'Illinois permet une combinaison des caractéristiques. En outre, il propose une combinaison entre les caractéristiques globales ou locales de l'image avec des mots-clés textuels. La couleur est représentée au moyen d'un histogramme 2D en utilisant les coordonnées H et S de l'espace HSV. La texture est représentée par deux histogrammes, une mesure de la grossièreté et une autre l'orientation de l'image, et un scalaire définissant le contraste. Afin d'extraire les caractéristiques, l'image est divisée en 5x5 zones. La forme est représentée par le biais des descripteurs de Fourier. MARS utilise des techniques de la recherche d'information afin de permettre un système CBIR interactif.

VIRAGE

VIRAGE [4] est un produit commercial pour la recherche d'images et de vidéo par le contenu. Il emploie la couleur, la texture et la forme comme attributs. VIRAGE offre à l'utilisateur une interface graphique conviviale pour formuler sa requête d'une façon interactive. Cette interface permet une insertion d'image, réglage des poids, outils de dessin,...

BLOBWORLD

Blobworld [15] est un système de recherche d'image développé à l'université de Berkeley. Il utilise l'algorithme EM (Expectation Maximization) pour segmenter les images en régions uniformes de couleur et de texture (blobs). La couleur est décrite par un histogramme de 218 bin dans l'espace de couleur L^*a^*b . La texture est représentée au moyen du contraste et de l'anisotropie sur la région d'intérêt. La surface, l'excentricité, et l'orientation représentent la forme. La Requête est effectuée sur la base d'une région de l'une des images présentées à l'utilisateur. Blobworld permet à l'utilisateur de visualiser les différentes régions

de la requête. L'utilisateur sélectionne une région est caractérisée son importance selon deux valeurs : 'somewhat' et 'very'. Aussi, l'utilisateur est invité à indiquer l'importance de chaque attribut selon trois critères : 'not', 'somewhat' et 'very'. La mesure de similarité est effectuée en combinant la distance quadratique (pour la couleur) avec la distance Euclidienne (pour la texture et la forme).

Tous ces systèmes sont malheureusement difficilement comparables car ils travaillent tous avec des bases d'images différentes. Ils offrent tous une interface conviviale à l'utilisateur qui doit néanmoins entrer quelques paramètres avant de lancer la recherche.

2.13 Conclusion

Dans ce chapitre nous avons cerné l'état de l'art des techniques de traitement d'image ou de vision utiles dans le cadre de la recherche d'images par le contenu. Dans la première partie nous avons parlé de la description de l'image par la couleur, la texture et la forme. Ces attributs constituent les descripteurs fondamentaux des systèmes CBIR. Le résumé de l'image contient donc toute information nécessaire au calcul de la similarité. La mesure de cette dernière repose souvent sur un calcul de distance dans un espace de caractéristiques. Le choix de la distance est lié également à l'attribut utilisé, par exemple pour l'histogramme, on utilise généralement l'intersection d'histogrammes ou la distance quadratique. Cependant la mesure de ressemblance devient plus complexe pour l'approche par région : pour faire correspondre deux images, cela nécessite un traitement particulier qui se ramène à l'appariement sur graphe bipartite. La recherche d'images basée sur les régions requiert donc une détection préalable des objets ou régions significatives, c'est le but de la segmentation. Le prochain chapitre aborde les principales approches de segmentation. Nous présentons les techniques de partitionnement adaptées aux systèmes CBIR. Nous allons ensuite exposer notre méthode de segmentation appliquée aux images d'algues.

Chapitre 3

APPROCHE DE SEGMENTATION POUR LA DESCRIPTION LOCALE DE L'IMAGE

3.1 Introduction

La segmentation s'avère une étape cruciale en vision par ordinateurs. Elle est au centre des travaux intenses de recherche. Le contenu d'une image peut être décrit par un ensemble de composantes visuelles. Leur détection est une tâche ardue et un problème fondamental du traitement d'images. La segmentation est peut être considérée comme un problème de clustering où l'on cherche à regrouper les pixels de la même région dans le même ensemble. Le clustering consiste à organiser des objets dans des groupes dont les membres partagent des propriétés communes et vérifient donc un critère de similarité. En référence à la capacité que possède le système visuel humain de distinguer les objets d'une image. La segmentation vise à extraire des régions saillantes de l'image, qui pourront ensuite être manipulées par des tâches interprétatives de plus haut niveau, en vue de reconnaître des objets. Les avantages de manipuler de telles régions au lieu des simples pixels sont nombreux : tout d'abord, la complexité des traitements futurs se voit considérablement réduite. Ensuite, le fait de manipuler des primitives de type région ou contour permet de disposer d'outils plus proches du niveau objet. Or, nous avons déjà vu dans le chapitre 2 que la segmentation forte peut s'avérer particulièrement difficile à atteindre.

3.2 Description locale

Une région de l'image correspond à un ensemble de pixels qui présentent une certaine homogénéité visuelle par exemple de couleur, de texture, de forme, etc. Une région peut contenir un objet ou une partie de celui-ci. La notion d'ob-

jet est étroitement liée à la perception humaine. C'est ainsi que les régions et les relations entre elles accèdent à une certaine sémantique basique et limitée du contenu image. Les techniques de recherche d'images fondées sur les approches locales annoncent des résultats prometteurs par rapport aux méthodes optant pour un traitement global sur l'image [15, 78]. En effet la décomposition de l'image en plusieurs entités permet un plus grand degré de libertés dans le processus de recherche justement favorisé par cette sémantique de bas niveau. Le but des systèmes CBIR basés sur les régions d'intérêt est d'obtenir une description la plus complète possible d'un ou plusieurs objets dans l'image. Une fois ces derniers détectés, il est plus facile au système d'identifier les images cibles. D'où un gain en performances [152].

Les méthodes locales sont donc conditionnées par la détection automatique d'objets qui requiert une segmentation précise. La segmentation consiste à grouper les pixels en régions, selon des critères bas niveau comme la couleur et la texture. On aboutit ainsi à une partition de l'image en régions, chacune d'entre elles étant homogène en termes de couleurs ou de texture, selon le critère utilisé. Toutefois, ces approches sont très limitées car elles ne permettent pas d'extraire des régions composées de plusieurs parties avec pour chacune des descripteurs bas niveaux différents. Cette description demeure très difficile à obtenir avec les méthodes actuelles et à ce niveau de segmentation, en particulier pour les objets constitués de plusieurs entités visuellement différentes (par construction, occlusion, ombrage etc.). Shi et Malik [116] supposent que la segmentation ne pourra pas et ne devra pas fournir un résultat final objet complet, mais qu'elle produise une séquence de partitionnement hiérarchique basée sur les attributs de bas niveau et assurant une cohérence de couleur, de texture ou de mouvement. Le rôle de la segmentation serait alors limité à regrouper en régions homogènes des points qui partagent les mêmes propriétés. Ce d'autant plus que les techniques de CBIR montrent une certaine tolérance à la segmentation imparfaite parce que le processus de recherche de régions similaire s'accommodent d'une mise en correspondance partielle entre composantes de l'image. Une détection de région sera donc considérée pertinente si la région détectée est incluse dans l'objet cible.

Notre but est de capter des zones d'intérêt qui seront signifiantes pour l'utilisateur, afin d'apparier les images d'une manière plus sûre. L'image de la figure 3.1 représente de l'eau, des algues et des roches. Si l'utilisateur choisit un objet bien déterminé cela permet de guider la recherche dans une direction plus précise. Nous abordons cette notion dans le prochain chapitre.



FIG. 3.1 – Image de synthèse représentant un type d'algue.

3.2.1 Segmentation d'images

La segmentation est essentielle parce qu'elle constitue l'étape préliminaire à l'analyse et la reconnaissance de formes. Aussi en recherche d'images par le contenu, la segmentation s'avère une étape critique parce que le résultat final de recherche en dépend.

La segmentation suscite beaucoup d'intérêt ces trois dernières décennies mais elle reste un problème difficile dans son aspect universel. Comme décrit ci-dessus, on admet donc que le rôle de la segmentation consiste à partitionner l'image en régions respectant une certaine cohérence visuelle et une continuité spatiale. Les éventuelles imperfections de segmentation pourront être surmontées par le processus de mise en correspondance fondant la recherche en base de données. Nous retenons donc la définition donnée par Horowitz et Pavlidis [54] :

Soit $I(x, y)$ la valeur du pixel en (x, y) et P un prédicat qui mesure l'uniformité sur un groupe de pixels connexes : la segmentation est définie comme une partition de I en n sous ensembles où régions $\{R_1, R_2, \dots, R_n\}$ tels que :

$$\cup_{i=1}^n R_i = I$$

$$R_i \cap R_j = \Phi \quad (i \neq j)$$

$$P(R_i) = \text{vrai} \quad \forall i \in \{1, \dots, n\}$$

$$P(R_i \cup R_j) = \text{faux} \quad \text{si } i \neq j$$

Une grande variété de méthodes de segmentation a été proposée dans la littérature. Les techniques de segmentation sont généralement catégorisées en 4 groupes : fondés sur les contours, sur les régions, par regroupement et les hybrides. Des études de synthèse ont été introduites dans [97]. Ces techniques ont le plus souvent été mises au point sur des images mono canal à cause des contraintes de calcul et de la capacité mémoire. Cela n'a pas été trop gênant parce que selon plusieurs modèles de la vision humaine, nous serions plus sensibles aux variations de luminance qu'à la sensation couleur. Cette hypothèse trouve une application par exemple en compression d'images couleurs (JPEG) où on sépare la chrominance de la luminance pour coder la couleur sur moins de bits. Les avancées technologiques des moyens de calculs et de stockage ont permis la banalisation des images en couleur, d'où l'utilisation plus récente de cette dernière comme primitive dans le processus de segmentation. Pendant cette dernière décennie on note une augmentation remarquable des techniques de segmentation traitant la couleur et la texture. La section suivante trace un état de l'art des méthodes de segmentation utilisées en systèmes CBIR.

3.3 Méthodes de segmentation en CBIR

Dans le contexte du CBIR, la segmentation est confrontée à des problèmes plus larges. Comme déjà dit en chapitre 2, la recherche d'images par le contenu est en partie subjective et la segmentation vise à réduire le fossé sémantique en créant des régions significatives. Nous listons ici les principaux systèmes de CBIR qui utilisent la segmentation pour fonder une recherche dans les bases d'images. Cette liste n'est évidemment pas exhaustive : Le système « BlobWorld » [15] repose sur l'algorithme Expectation Maximization (EM) [31] pour extraire des régions. L'algorithme employé pour le regroupement des pixels en exploitant la couleur, la texture et la position comme primitives avant un post traitement supposé éliminer le bruit de la segmentation. Pour le choix du nombre de segments, les Valeurs 2 à 5 sont testées et le nombre retenu est celui qui vérifie le critère de Minimum Description Length [107]. Comme pour les algorithmes précédents, le choix automatique du nombre de classes ne peut être fait qu'en exécutant plusieurs fois l'algorithme. "Netra " [78] est un autre système segmentant les images en régions homogènes grâce à un algorithme basé sur le modèle prédictif nommé «Edge flow» [77] avec les mêmes primitives. « Simplicity » adopte une stratégie de recherche différente : la segmentation exploite les K-moyennes [79]. L'image est divisé en bloc 4x4, chacun représenté par 6 attributs de couleur et de texture. Citons encore un couple de travaux qui dressent la même problématique et différent par les méthodes de segmentation [128]. La plupart des méthodes

citées ci-dessus emploient des algorithmes fondés sur le regroupement partant du principe que la perception humaine tend à regrouper les points homogènes. Notre travail est motivé par ce même principe. Nous développons un algorithme optimisé pour notre type d'application. La section suivante introduit les graphes et marches aléatoires comme outils de regroupement très utilisés récemment.

3.4 Graphes et marches aléatoires

Depuis quelques années, l'utilisation de graphes dans de nombreux domaines révèle un engouement de plus en plus fort. Les graphes sont utilisés comme outil en automatisme pour les réseaux de petri ou les grafcet qui représentent un déroulement séquentiel d'une application. Les graphes sont également utilisés en base de données pour la représentation conceptuelle des données et dans les systèmes d'information géographiques. Dans le domaine de la bioinformatique, les graphes permettent de manipuler des représentations moléculaires. Dans le cadre de l'analyse de textes, les graphes permettent de représenter les documents afin de mettre en évidence l'agencement spatial d'un document.

En traitement d'images, le graphe le plus couramment utilisé est le graphe des pixels, où chaque pixel représente un noeud du graphe et chaque arc est défini par la connexion entre les noeuds. Une représentation plus compacte de l'image est donnée par exemple par le graphe d'adjacence construit à partir d'une sur-segmentation. Dans ce cas, on associe souvent aux noeuds représentant les régions des attributs comme le niveau de gris moyen, la surface, des indices texturaux, etc. Quant aux arcs entre deux régions adjacentes, on leur associe la longueur du contour ou une mesure de contraste. Des graphes également couramment utilisés sont les graphes de Voronoi construits à partir d'un nuage de points ou, de façon duale, la triangulation de Delaunay [67].

Comme nous pouvons le constater, le but principal des graphes est de permettre une représentation structurée des données. Il est ainsi possible de représenter et de manipuler plus aisément des objets complexes. Nous proposons ici d'utiliser les graphes comme outil de représentation d'images. Nous allons donc dans un premier temps rappeler la notion de graphe. Nous introduirons ensuite la problématique de segmentation d'images en utilisant les graphes. Un état de l'art des travaux sur le sujet sera proposé et mis en parallèle avec les travaux effectués dans cette thèse.

3.4.1 Graphes

Un graphe non-orienté $G = (V, E)$ est composé d'un ensemble V de sommets (ou noeuds) et d'un ensemble E de paires (non ordonnées) de sommets nommées arêtes (ou liens). Nous noterons n le nombre de sommets ($n = |V|$) et m le nombre d'arêtes ($m = |E|$). Les arêtes du graphe peuvent être pondérées grâce à une fonction de poids $w : E \rightarrow \mathbf{R}^+$ permettant de modéliser plus finement les interactions entre sommets, nous obtenons ainsi un graphe pondéré $G = (V, E, w)$. Le poids d'une arête $\{i, j\}$ entre deux sommets i et j sera noté $w_{ij} > 0$. Par convention, un poids nul est attribué dans le cas où l'arête n'existe pas ($w_{ij} = 0$ si $\{i, j\} \notin E$). Dans le cas d'un graphe non pondéré les poids des arêtes de E sont fixés à 1, ainsi dans ce cas particulier $\forall i, j \in V, w_{ij} \in \{0, 1\}$.

Des graphes orientés peuvent aussi être définis, leur ensemble E est alors composé de couples (ordonnés) de sommets nommés arcs. Un graphe non-orienté peut être vu comme un graphe orienté dans lequel tous les liens sont réciproques. Les propriétés des marches aléatoires diffèrent grandement entre les graphes orientés et non-orientés. Dans la suite de cette thèse, nous ne considérerons que des graphes non-orientés.

Le degré $d(v)$ d'un sommet $v \in V$ est le nombre d'arêtes incidentes au sommet v ; il s'agit du nombre de sommets voisins de v . Nous définissons aussi le poids $w(i)$ d'un sommet i comme la somme des poids de ses arêtes incidentes :

$$w(i) = \sum_{j \in V} w_{ij} \quad (3.1)$$

Notons que le poids d'un sommet coïncide avec la définition du degré d'un sommet dans le cas des graphes non-pondérés. La matrice d'adjacence W représentant les arêtes d'un graphe pondéré non-orienté est définie par :

$$W_{ij} = W_{ji} = \begin{cases} 0 & \text{Si } \{i, j\} \notin E; \\ w_{ij} & \text{Si } \{i, j\} \in E. \end{cases} \quad (3.2)$$

Dans le cas d'un graphe non-orienté, la matrice d'adjacence W est symétrique ($W^T = W$). Par ailleurs, dans le cas d'un graphe non pondéré $W_{ij} \in \{0, 1\}$ (car nous fixons un poids w_{ij} égal à 1 pour toutes les arêtes de E). Dans de nombreux cas, le poids d'un sommet d'un graphe pondéré est une bonne généralisation du degré d'un sommet d'un graphe non-pondéré. Cependant nous pouvons toujours définir la notion de degré d'un sommet dans un graphe pondéré comme le nombre de ses voisins, cette notion sera alors différente de la notion de poids. Nous définissons aussi la matrice diagonale D des poids des sommets de la façon suivante :

$$D_{ij} = \begin{cases} w(i) & \text{Si } i = j; \\ 0 & \text{Si } i \neq j. \end{cases} \quad (3.3)$$

Cette matrice D peut aussi être vue, dans le cas particulier des graphes non-pondérés, comme la matrice des degrés. Une mesure de la « taille » d'un sous-ensemble $A \subset V$ est le poids de ses arêtes :

$$vol(A) = \sum_{i \in A} d_i \quad (3.4)$$

3.4.2 Marches aléatoires

Nous définissons dans ce paragraphe un processus de marche aléatoire à temps discret dans G . Le temps est discrétisé ($t = 0, 1, 2, \dots$) et un marcheur est localisé à chaque instant t sur un sommet du graphe G . Le marcheur se déplace à chaque instant aléatoirement et uniformément vers l'un de ses sommets voisins. La suite des sommets visités constitue une marche aléatoire. Lorsque le marcheur se trouve à un instant donné sur le sommet i , nous notons P_{ij} la probabilité d'aller du sommet i au sommet j à l'instant suivant. Dans le cas d'un graphe non pondéré, le marcheur a une probabilité de $\frac{1}{d(i)}$ d'aller vers l'un des $d(i)$ voisins du sommet i . Lorsque nous considérons un graphe pondéré, les différentes arêtes empruntées pour relier les différents voisins n'ont pas le même poids et donc n'ont pas la même importance. Nous considérons alors que la probabilité d'emprunter chaque arête est proportionnelle au poids de l'arête. La probabilité de transition P_{ij} d'un sommet i à un sommet j est donc donnée par l'expression suivante :

$$P_{ij} = \frac{W_{ij}}{w(i)} = \frac{w_{ij}}{\sum_{k=1}^n w_{ik}} \quad (3.5)$$

Ces probabilités définissent la matrice de transition P de la chaîne de Markov associée au processus de marche aléatoire. Notons qu'il n'y a pas de problème de définition car les contraintes que nous avons fixées sur le graphe imposent $w(i) > 0$ pour tous les sommets i . Nous pouvons définir cette matrice de transition à partir de la matrice d'adjacence W du graphe et de la matrice diagonale D des poids des sommets selon la formule suivante :

$$P = D^{-1} W \quad (3.6)$$

3.5 Regroupement spectrale

Le regroupement spectral (Spectral Clustering) a été introduit par Donath et Hoffman [34]. Ils ont proposé de construire les partitions des graphes à base des vecteurs propres de la matrice d'adjacence. Dans la même année, Fiedler [39] a montré que les bi-partitions d'un graphe sont liées avec le second vecteur propre de la matrice de Laplace, et il a suggéré d'utiliser ce vecteur propre à la partition d'un graphe. Depuis lors, le regroupement spectral a connu un essor important dans différents domaines de recherche. Dans le domaine de segmentation d'image, le regroupement spectral a été devenu populaire grâce aux travaux de Shi et Malik [116], Meila et Shi [86]. Par la suite, il a été étendu à de nombreux domaines, par exemple, l'apprentissage de la fonction de similarité [3], ou l'annotation automatique d'image [104].

Les données de grande taille telle que les caractéristiques ou signatures d'images sont difficiles à interpréter ou regrouper. Toutefois, ces données complexes peuvent être les résultats de quelques changements de paramètres intrinsèques. Une série d'images dans un système CBIR par exemple, pourraient être régie par trois paramètres : la rotation horizontale, la rotation verticale et les changements d'éclairage. Ces caractéristiques peuvent être récupérées par des méthodes non-linéaires de réduction de dimensionnalité. Contrairement aux méthodes linéaires telles que PCA ou LDA [51], les méthodes d'analyse spectrale n'ignorent pas la convexité ou la concavité des données et sont donc en mesure de traiter un plus grand nombre d'ensembles de données [75]. Dans cette thèse, nous nous concentrerons sur le regroupement spectral pour une méthode de segmentation en utilisant les marches aléatoires.

3.5.1 Matrice Laplacienne

La matrice Laplacienne est la partie principale d'un regroupement spectral d'un graphe. Un domaine consacré à ces matrices, est décrit en [21].

Laplacienne non normalisée : La matrice Laplacienne **non normalisée** est donnée par :

$$L = D - W \quad (3.7)$$

Où W est la matrice d'adjacence et D est la matrice diagonale avec :

$$D_{ii} = \sum_{j=1}^n W_{ij} \quad (3.8)$$

Les propriétés les plus importantes sont :

– pour chaque vecteur $f \in \mathbf{R}^n$,

$$\begin{aligned} f'Lf &= f'Df - f'Wf \\ &= \sum_{i=1}^n d_i f_i^2 - \sum_{i,j=1}^n f_i f_j w_{ij} \\ &= \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2 \end{aligned}$$

- La matrice L est symétrique et semidéfinie-positve. La symétrie de L vient de la symétrie de W et D . Le fait qu'elle soit semi définie positive est une conséquence de $f'Lf \geq 0, \forall f \in \mathbf{R}^n$,
- la plus petite valeur propre de L est 0 et correspond au vecteur propre constant $\vec{1}$
- L possède des valeurs propres non négatives $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$

Laplacienne normalisée : La matrice Laplacienne **normalisée** est définie principalement dans la littérature[21] par deux matrices liées entre elles comme suit :

$$L_s = D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2}. \quad (3.9)$$

$$L_{ma} = D^{-1} L = I - D^{-1} W = I - P. \quad (3.10)$$

L_s est la matrice Laplacienne normalisée symétrique et L_{ma} est la matrice Laplacienne normalisée qui est proche de la matrice de transition P associée au processus de marche aléatoire. Les propriétés les plus importantes sont :

– pour chaque vecteur $f \in \mathbf{R}^n$,

$$f' L_s f = \frac{1}{2} \sum_{i,j=1}^n w_{ij} \left(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right)^2. \quad (3.11)$$

- 0 est une valeur propre de L_s correspondant au vecteur propre $D^{1/2} \vec{1}$.
- 0 est une valeur propre de L_{ma} correspondant au vecteur propre $\vec{1}$.
- L_s et L_{ma} sont des matrices semidéfinies positives et possèdent n valeurs propres non négatives $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.
- λ est une valeur propre de L_{ma} avec le vecteur propre u si et seulement si λ résoud $Lu = \lambda Du$.

Ces deux types de matrices Laplaciennes jouent un rôle central dans de nombreux algorithmes d'apprentissage semi supervisé à base des graphes. Cela est dû à certaines propriétés importantes de Laplacien. La matrice Laplacienne est toujours semidéfinie positive et par conséquent un valide noyau. Ensuite,

l'utilisation de la matrice de Laplace normalisée (ou, non normalisée selon le contexte) est motivée par sa relation au noyau de la chaleur (heat kernel), ou le noyau de diffusion (diffusion kernel) [106] qui sont bien connus en physique.

Dans le cadre du regroupement spectral, une comparaison a été faite dans [75] entre les trois matrices Laplaciennes. Il existe plusieurs arguments qui favorisent l'utilisation du regroupement spectral normalisé plutôt que non normalisé. Le premier argument est le partitionnement du graphe. Une opération de regroupement vise à :

- (1) Trouver une partition de telle sorte que les points appartenant à des clusters différents sont non similaires, c'est-à-dire minimiser la similitude entre les deux partitions.
- (2) Trouver une partition de telle façon que les points appartenant au même cluster sont similaires, c'est-à-dire maximiser la similarité intra-cluster.

Les auteurs dans [75] ont montré que le regroupement spectral non normalisé implémente les deux objectifs ci-dessus. Tant que le regroupement spectral non normalisé implémente seulement le premier objectif. Le deuxième argument en faveur du regroupement spectral est la convergence. En effet L_{ma} et L_s convergent vers une partition utile dans un espace de donnée. Contrairement au regroupement spectral normalisé, il peut être prouvé que le regroupement spectral non normalisé peut ne pas converger, ou qu'il peut converger vers des solutions triviales [75].

3.5.2 Segmentation spectrale normalisée

De nombreuses méthodes de segmentation s'appuyant sur la théorie des graphes ont été proposées. L'idée de base est la suivante : A partir d'un graphe construit sur les pixels de l'image, on cherche à partitionner le graphe de telle sorte que les parties obtenues correspondent à des zones "homogènes" de l'image (au sens de l'attribut des noeuds).

Arbre couvrant de poids minimal

Une première approche proposée par Morris [91] se fonde sur la recherche de l'arbre couvrant le poids minimal pour réaliser la segmentation. Le graphe est construit en pondérant les arcs par la valeur absolue de la différence de niveaux de gris, ou tout autre attribut significatif pour la segmentation considérée, des sommets qu'ils relient. Sur ce graphe, on calcule alors l'arbre couvrant de poids minimal. On supprime les arêtes les plus coûteuses du graphe pour obtenir la segmentation. En effet, on obtient alors une forêt, dans laquelle chaque arbre

représente une zone homogène (puisque toutes les arêtes ont des coûts faibles, et donc lient des sommets dont les attributs sont proches). Le critère de suppression des arêtes peut porter sur la valeur de l'arête ou sur le nombre total d'arêtes à supprimer. La recherche de l'arbre couvrant de poids minimal se fait par l'algorithme de Kruskal ou par l'algorithme de Prim [18]. Dans les deux cas l'idée est de trier les arêtes du graphe par ordre de poids croissants et de les ajouter pour constituer l'arbre. C'est une méthode de segmentation bien adaptée aux images sans bruit, qui n'introduit pas d'information sur la forme des régions cherchées, et qui suppose que les régions à segmenter sont bien délimitées les unes par rapport aux autres. Une mise en échec sera par exemple due à l'existence d'une connexion entre deux régions similaires mais appartenant à des objets différents.

Coupe de capacité minimale

Dans le même ordre d'idée, des méthodes de segmentation recherchant la coupe de capacité minimale ont été proposées [156]. A nouveau, le principe est de rechercher un partitionnement du graphe en sous-graphes. Une coupe partitionne un graphe en deux parties A et B telles que $A \cap B = \phi$ et $A \cup B = V$ et la capacité de la coupe est définie par :

$$cut(A, B) = \sum_{x \in A, y \in B} w(x, y) \quad (3.12)$$

Avec $w(x, y)$ capacité de l'arc (x, y). Le graphe considéré est à nouveau le graphe des pixels en 4-connexité, mais cette fois-ci le poids des arcs est une fonction de similarité (et non l'inverse comme précédemment). La coupe de poids minimale va donc déconnecter les pixels très dissemblables.

Coupe Normalisée(Ncut)

Le problème de la coupe de capacité minimale est que la capacité est toujours fonction du nombre d'arcs de la coupe, et donc pour des poids positifs, on aura toujours intérêt à partitionner le graphe de telle sorte que le moins d'arcs possibles interviennent dans la coupe. Ce phénomène est illustré sur la figure 3.2.

Il est donc toujours moins coûteux d'isoler un noeud, même si structurellement la partition (A, B) serait optimale. Pour remédier à ce problème Shi et Malik ont proposé une nouvelle définition du coût d'une coupe [116], qui supprime l'influence du nombre d'arcs dans la coupe :

$$Ncut(A, \bar{A}) = \left(\frac{1}{volA} + \frac{1}{vol\bar{A}} \right) \sum_{i \in A, j \in \bar{A}} W_{ij} \quad (3.13)$$

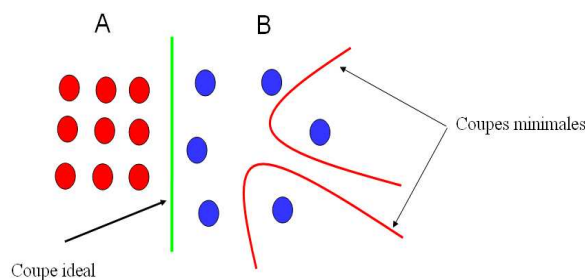


FIG. 3.2 – Influence du nombre de sommets dans la coupe.

\bar{A} est le complément du sous-ensemble A .

Minimiser N_{cut} revient à trouver une coupe de faible poids entre deux classes avec de fortes connexions internes. On retrouve une propriété similaire à celle de la classification pour laquelle maximiser la variance interclasses revient à minimiser la variance intra-classes. Dans [116], les auteurs ont montré que la recherche de la coupe normalisée minimale est un problème NPcomplet. Néanmoins une solution discrète approximée peut être trouvée de façon efficace en utilisant les valeurs et vecteurs propres [116]. En effet, La solution est donnée par le vecteur propre associé à la seconde plus petite valeur propre de l'équation :

$$Lx = \lambda Dx \quad (3.14)$$

Avec L est la matrice Laplacienne et D est la matrice diagonale formée avec les degrés des noeuds.

L'algorithme N_{cut} est résumé comme suit :

Algorithme 3.5.1 Coupe Normalisée(N_{cut}).

- (1) Étant donné une image, Construire un graphe pondéré $G = (V, E)$ et la matrice d'affinités W correspondante.
 - (2) Construire la matrice Laplacienne non normalisée $L = D - W$.
 - (3) Résoudre $Lx = \lambda Dx$ pour les vecteurs propres avec les plus petits valeurs propres.
 - (4) Utiliser le vecteur propre correspondant au seconde plus petite valeur propre pour bi-partitionner le graphe.
 - (5) Décider s'il faut subdiviser la partition courante par une répartition récursive.
-

L'algorithme N_{cut} se base sur la résolution de l'équation $Lx = \lambda Dx$ pour

généraliser des vecteurs propres qui correspondent, selon les propositions de 3.5.1, aux vecteurs propres de la matrice L_{ma} . En effet, l'algorithme utilise les vecteurs propres de la matrice Laplacienne normalisée L_{ma} . D'où l'appellation de la segmentation spectrale normalisée.

Un exemple de segmentation via l'algorithme Ncut est montré sur la figure 3.3.

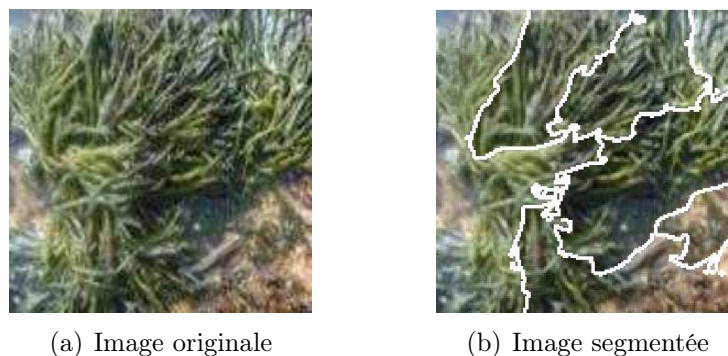


FIG. 3.3 – Résultats de la segmentation par coupe normalisée.

3.6 Nouvelle méthode de segmentation basée sur les marches aléatoires

Choix de l'attribut

Le processus de segmentation suppose l'image transformée pour être codée selon des primitives caractéristiques à partir desquelles on saura définir une homogénéité sur un ensemble de points. Nous avons vu que traditionnellement la couleur et la texture constituent les principaux attributs pour segmenter l'image en région. Elles sont utilisées indépendamment ou fusionnées. La couleur dépend initialement des trois valeurs de l'espace de couleurs adopté, en revanche la texture peut concerner l'intensité lumineuse seule. S'il est facile d'obtenir dès l'acquisition l'information de couleur ou de l'intensité en un pixel, la texture a besoin d'une certaine extension spatiale autour du pixel pour capter suffisamment d'informations. Les applications que nous visons, concernent des images où la texture est une caractéristique dominante. Nous avons donc opté pour l'utilisation de la texture afin de segmenter l'image. Nous allons voir plus loin qu'avec les contraintes spatiales nous pourrions obtenir une certaine homogénéité sur les textures.

Contribution du voisinage

Il s'agit de définir une caractéristique locale pour capter de l'information spatiale. Nous exploitons particulièrement cette propriété afin d'introduire une contrainte supplémentaire au niveau des primitives.

Targhi et al. ont présenté dans [143] un nouveau descripteur de texture «Eigentransform». La transformation consiste en une mesure de la rugosité en considérant les valeurs propres d'une matrice qui est formée par les niveaux de gris d'un carré autour directement d'un pixel. La valeur propre la plus grande donne une version lissée de l'image originale, mais les valeurs propres de plus petite grandeur encodent la fréquence élevée de la texture. Un avantage majeur de la transformation en valeurs propres est qu'elle est moins sensible à la luminosité locale, au lieu de réagir presque entièrement à la texture. Ceci est en contraste avec de nombreux autres descripteurs de texture [143]. Ces propriétés le rendent remarquablement bien adaptées aux applications pratiques en traitement d'image et en particulier la segmentation d'images.

L'algorithme consiste en une décomposition en valeur propres d'une fenêtre de taille $w * w$ autour d'un pixel d'une image donnée. Les valeurs propres sont triées d'ordre décroissant, $\{\|\lambda_1\|, \|\lambda_2\|, \dots, \|\lambda_n\|\}$ décrivant chaque pixel dans l'image. Il faut signaler que les vecteurs propres ne sont pas utilisés dans cet algorithme. Pour mieux segmenter l'image et réduire le bruit qui en découle, les auteurs ont défini la transformation Γ pour chaque pixel de l'image comme suit :

$$\Gamma(l, w) = \frac{1}{w - l + 1} \sum_{k=l}^w \|\lambda_k\|, \quad 1 \leq l \leq w. \quad (3.15)$$

Pour bien caractériser la texture, les auteurs ont choisi l dans l'intervalle $[2, w/3]$.

3.7 Méthode de segmentation

Nous considérons un Graphe $G = (V, E, W)$ construit sur les pixels de l'image à segmenter. La matrice de similarité W , qui contient les coefficients w_{ij} doivent être soigneusement définis. Deux pixels i et j appartenant à la même classe doivent avoir une valeur de similarité élevée et dans le cas inverse, la similarité devrait être nulle. La façon classique pour construire la matrice de poids est basée sur un noyau gaussien de 0-moyenne et la variance σ . Soit $d(i, j)$ la distance entre les noeuds i et j . Alors le poids ω_{ij} sera calculé comme suit :

$$\omega_{ij} = \exp\left(-\frac{d(i, j)}{\sigma^2}\right) \quad (3.16)$$

Où σ est un paramètre d'échelle et $d(i, j)$ est une distance entre certaines caractéristiques locales de l'image. La fonction de la distance peut être :

– Locales, par exemple :

$$d(i, j) = \|i - j\|_2 \quad (3.17)$$

La fonction $d(i, j)$ exprime la proximité spatiale des pixels i et j .

– Non local, par exemple :

$$d(i, j) = \|F(i) - F(j)\|_2 \quad (3.18)$$

où F est un vecteur caractéristique. La fonction $d(i, j)$ exprime la proximité entre les descripteurs locaux.

– Combinaison de services locaux et non locaux, par exemple :

$$\omega_{ij} = \exp \left(-\frac{\|i - j\|_2}{\sigma_1^2} - \frac{\|F(i) - F(j)\|_2}{\sigma_2^2} \right) \quad (3.19)$$

Dans ce cas, les fonctions locales et non-locales sont combinées.

Pour obtenir un résultat de segmentation optimale le choix des caractéristiques est une tâche critique. Pour ce faire, on peut considérer un vecteur caractéristique de chaque pixel provenant d'un banc de filtres comme suggéré par exemple dans [80]. Un moyen efficace pour représenter la texture sans bancs de filtres est d'utiliser un patch autour de pixel comme vecteur caractéristique. Cette idée a été introduite d'abord pour la synthèse de texture [35, 72] puis, dans le débruitage d'images. Buades et al. [12] ont proposé d'utiliser la différence entre patches pour calculer une matrice de poids et ensuite effectuer une moyenne non locale sur une image bruitée. Enfin, dans [10], Bresson et al. ont proposé une méthode de segmentation non supervisée basée sur cette représentation non locale de l'image. Nous allons maintenant introduire la matrice de similarité dans l'opération de segmentation. Nous sommes intéressé ici à extraire les informations locales et non locales. Le graphe est construit en utilisant la formule suivante :

$$W(i, j) = \begin{cases} \exp \left(-\frac{\|i - j\|_2}{\sigma_1^2} - \frac{\|F(i) - F(j)\|_2}{\sigma_2^2} \right) & \text{Si } i, j \in V_{p \times p}(i); \\ 0 & \text{Sinon.} \end{cases} \quad (3.20)$$

$V_{p \times p}(i)$ est un voisinage carré de taille p autour de i .

Nous allons nous intéresser à un processus de marche aléatoire (ou de diffusion) dans le graphe G . Le temps est discrétisé $t = (0, 1, 2, \dots)$. A chaque instant, un marcheur est localisé sur un sommet et se déplace à l'instant suivant vers un sommet choisi aléatoirement et uniformément parmi les sommets voisins. Nous proposons un algorithme qui utilise les marches aléatoires basées sur la similarité des noeuds du graphe. Les marches aléatoires sur graphe est un processus stochastique qui permet de se déplacer d'un noeud à l'autre aléatoirement. Il peut

être interprété comme une tentative de trouver une partition du graphe tels que la marche aléatoire reste longuement dans le même groupe et que le déplacement à un autre groupe arrive rarement. La suite des sommets visités est alors une marche aléatoire, et la probabilité de transition du sommet i au sommet j est à chaque étape : $P(i, j) = \frac{\omega(i, j)}{d(i)}$ et par suite $P = D^{-1}W$ avec D matrice diagonale des degrés des sommets. Pour un graphe connexe et non bipartite, la marche aléatoire possède toujours une distribution unique et stationnaire [74] : $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ où $\pi_i = d_i / \text{vol}(V)$. Comme expliqué plus haut, il y'a une relation entre la matrice de transition P et L_{ma} . Comme $L_{ma} = I - P$, en conséquence, λ est une valeur propre de L_{ma} avec u vecteur propre si et seulement si $1 - \lambda$ est une valeur propre de P avec u vecteur propre. Il est bien connu que de nombreuses propriétés d'un graphe peuvent être exprimées en termes de la matrice de transition P correspondante aux marches aléatoires sur graphe. De ce point de vue, il n'est pas surprenant que les valeurs propres de P peuvent être utilisés pour décrire le regroupement du graphe.

Dans le cadre de l'analyse et de la caractérisation de la texture, nous allons nous intéresser aux valeurs propres pour trouver un descripteur de texture unique et simple à calculer. Les valeurs propres de la matrice de transition P de faibles valeurs correspondent aux vecteurs propres de hautes fréquences, et celles de fortes valeurs correspondent aux vecteurs de basses fréquences. Si on considère alors que la texture est caractérisée par les hautes fréquences et si on ordonne les valeurs propres λ_i par module croissant, on peut espérer que les valeurs propres à partir de l'ordre 2 ou 3, qui sont une sorte de projection locale de l'image sur les vecteurs de haute fréquence, captent une grande partie de l'information texture. La figure (3.4) montre une illustration d'utilisation des valeurs propres pour la segmentation d'une image d'algues (*Cladophora aegagropila*). A chaque pixel de l'image, est associé un vecteur constitué par un motif composé des valeurs de l'image autour d'un voisinage 16x16 autour du pixel. Ces différents profils illustrent la capacité de discriminer différentes textures à partir des valeurs propres. Dans la figure (3.4b) on considère la plus grande valeur propre, les hautes intensités donnent une version lissée de l'image originale. La petite valeur propre (figure 3.4c) correspond aux hautes fréquences et donne une version bruitée de l'image cible. Pour caractériser la texture Nous avons considéré la moyenne des valeurs propres, ce qui revient à calculer la trace de P (figure 3.4e). Les résultats de segmentation en utilisant la trace de P restent comparables au calcul de la moyenne des importantes valeurs propres (les 3 premières valeurs propres à partir de l'ordre 2)(figure 3.4d). La texture est donc caractérisée par le descripteur suivant :

$$\Gamma(G) = \sum_{i=1}^w \lambda_i = \text{tr}(p) \quad (3.21)$$

Avec $\text{tr}(p)$ est la trace de la matrice P.



(a) Image originale



(b) Plus grande VP



(c) Plus faible VP



(d) Moyen des VP important

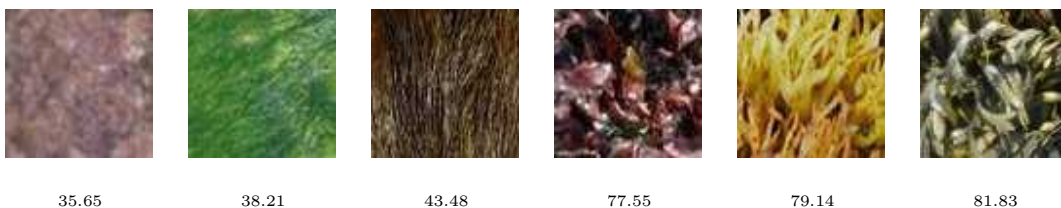


(e) Trace (P)

FIG. 3.4 – *Utilisation des valeurs propres pour la segmentation d'images.*

Pour segmenter l'image, nous avons proposé une méthode basée sur la trace de P. Ainsi, nous pouvons calculer le descripteur sans passer par la décomposition spectrale de la matrice de transition P.

Pour illustrer davantage le rôle du trace de P dans la description de la texture, la figure 3.5 montre la trace de P pour 6 type d'algues. On constate que les valeurs élevées décrivent les structures grossières, tandis que les valeurs inférieures correspondent aux textures lisses et fines.



35.65

38.21

43.48

77.55

79.14

81.83

FIG. 3.5 – *Trace (P) pour 6 types d'algues.*

L'algorithme suivant résume les principales étapes de notre approche de segmentation :

Algorithme 3.7.1 Segmentation basée sur P.

- (1) Étant donné une image I , construire une fenêtre F de taille $w * w$ autour de chaque pixel de I .
 - (2) Construire un graphe pondéré $G = (V, E)$ correspondant à F et sa matrice d'affinités W .
 - (3) Construire la matrice de transition $P = D^{-1}W$
 - (4) Calculer la trace de P .
 - (5) Assigner sa valeur au pixel concerné.
-

3.8 Résultats et Discussions

Il est très difficile d'évaluer les résultats d'une segmentation. En effet, ces traitements ne constituent pas une fin en soi, mais une étape indispensable à des traitements postérieurs. Martin et al. [85] proposent un cadre d'évaluation, qui repose sur la comparaison des résultats avec des segmentations effectuées par des opérateurs humains.

Toutefois, ces segmentations manuelles demeurent assez subjectives et en tout cas totalement déconnectées de traitements postérieurs. Ainsi, une image donnée peut être vue comme sous-segmentée ou au contraire sur-segmentée, suivant le but recherché. Par exemple, si nous sommes intéressés par l'extraction des objets par rapport au fond des images, le résultat sur l'oiseau dans la figure 3.6 est excellent. Si maintenant nous cherchons à comparer l'image avec des modèles stockés dans une base, en prenant en compte une dimension structurelle, ce résultat devient sous-segmenté. Encore une fois, nous ne pourrions évaluer quantitativement nos résultats qu'une fois revenus à des cas d'utilisations d'indexation. Cette évaluation globale du système sera proposée dans la section 4.5 du chapitre suivant.

Il faut rappeler que notre objectif est de déployer une chaîne complète de traitements pour l'indexation et la requête aux bases d'images d'algues. Nous verrons par la suite que certains traitements permettent de corriger certaines erreurs de segmentation, lors de la comparaison entre requête et une image de la base. Ce n'est donc qu'une fois toute la chaîne considérée, que nous pourrions pointer les éventuels problèmes liés à la segmentation, dans notre perspective d'indexation.

Nous nous concentrons donc pour l'instant sur une évaluation qualitative de la segmentation. Nous allons maintenant comparer qualitativement notre système à d'autres systèmes existants.



FIG. 3.6 – *Exemple de résultat de segmentation.*

3.8.1 Evaluation et Comparaison

Nous avons testé notre processus de segmentation sur un échantillon d'images d'algues. Des exemples de résultats sont montrés à la figure 3.7. La première colonne exhibe des images originales d'algues. La deuxième colonne montre l'image pré-segmentée en utilisant des patches de taille 16x16. La troisième colonne présente l'image segmentée après un post-traitement basé sur des opérations morphologiques. La dernière colonne présente les régions de l'image avec la moyenne de couleur de chaque région.

On peut constater que la méthode de segmentation proposée arrive à extraire efficacement les régions inhérentes aux espèces d'algues de leurs fonds et permet ainsi une bonne localisation des régions. Aussi la segmentation est capable de réduire de manière significative le bruit induit par l'étape de pré-segmentation : les régions de très petites tailles sont fusionnées de manière à créer des régions de taille significative. En outre, elle a tendance à faire émerger de l'image des zones qui correspondent aux différents objets sémantiques. On peut également remarquer que les autres artefacts d'éclairage induits par l'acquisition de ces images sont traités de manière assez efficace vu que notre méthode est moins sensible à la luminosité.

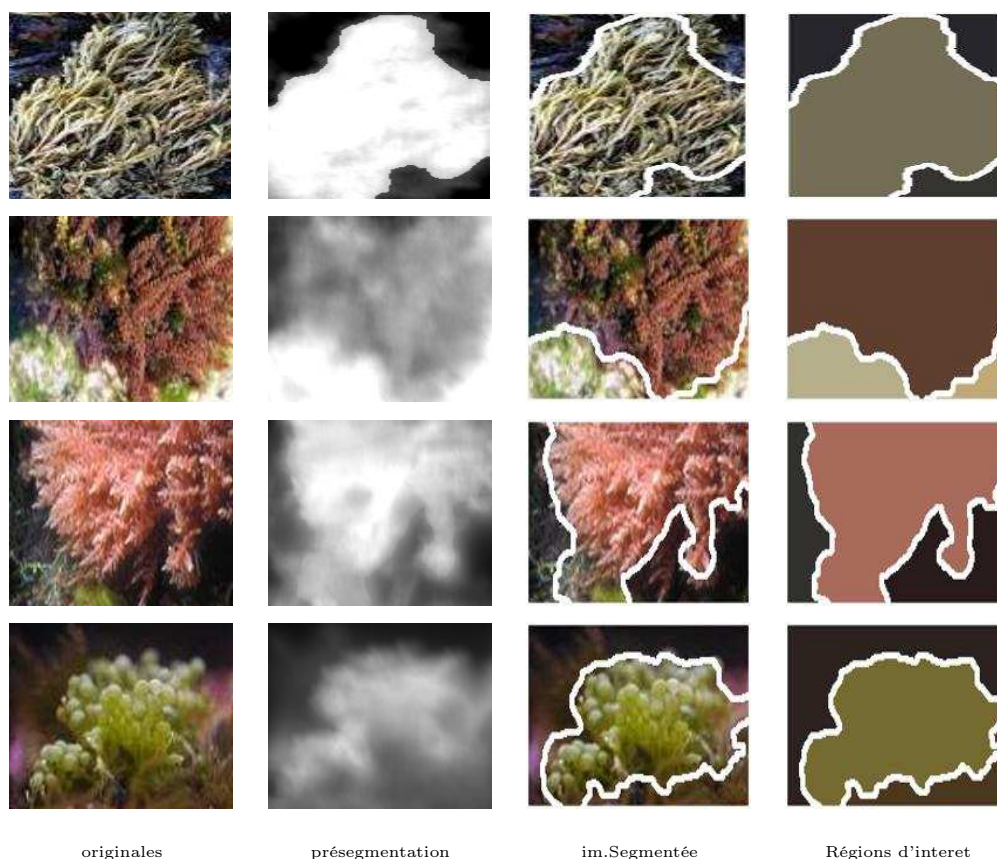


FIG. 3.7 – *Exemple de résultats de segmentation.*

3.8.2 Limitations et améliorations

Au fil de nos expérimentations nous avons pu détecter quelques limitations. Cela arrive lorsque deux régions voisines ont des textures proches (figure 3.8). La contrainte nécessaire au bon fonctionnement de cette technique est que les objets présents dans les images soient relativement bien séparées du fond et entre eux. En effet, l'algorithme est basé seulement sur la texture et ne prend pas en considération une caractéristique aussi importante qui est la couleur. Certes, la texture joue un rôle important dans la description des images d'algues et permet une bonne localisation des objets figurant dans l'image. Mais, il serait incorrect de ne pas considérer la couleur dans l'algorithme de segmentation.

Pour améliorer les résultats de notre processus de segmentation, nous avons ajouté les composantes de la couleur a et b de l'espace L^*a^*b à la composante de texture.

Des comparaisons entre les deux approches ont été effectuées. Les résultats sont présentés dans la figure 3.8. La première colonne exhibe des images originales dont la segmentation présente quelques limitations. La deuxième colonne représente les images segmentées via la méthode impliquant la texture. Bien que

des résultats de qualité aient été obtenus avec ce système, l'importance donnée à la similarité texture seule a conduit à des erreurs de segmentation : comme dans la première image par exemple où deux régions de types d'algues ont été fusionnées malgré qu'elles ont des couleurs différentes. Au contraire, la méthode combinant la couleur et la texture (troisième colonne) est capable d'effectuer des groupements homogènes et fidèles au contenu.

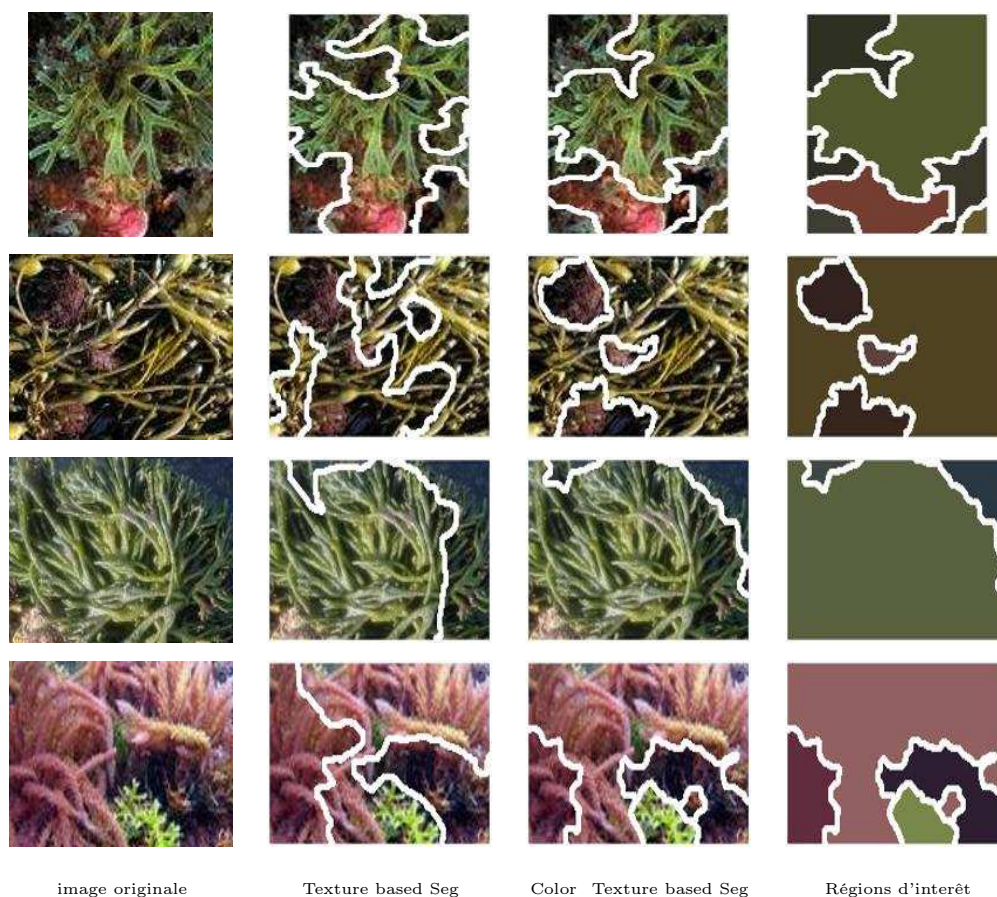


FIG. 3.8 – *Exemple de résultats impliquant la couleur.*

3.8.3 Comparaison à d'autres méthodes de segmentation

Pour comparer la méthode proposée à d'autres systèmes de segmentation, nous avons choisi la technique de Shi and Malik (Ncut) dont nous avons déjà parlé. Rappelons que ce dernier opère un regroupement de pixels sur la base des coupes normalisés sur les graphes. Nous avons utilisé les paramètres optimaux donnés par les auteurs [116]. Les résultats de segmentation (figure 3.9 et figure 3.10) obtenus sur ces images montrent que notre méthode localise mieux les régions. Aussi nous avons remarqué que les résultats obtenus avec le sys-

tème Ncut sont souvent sur-segmentés par rapport aux nôtres. Ceci vient du fait que le processus de groupement suivi est basé sur les coupes. Vu que Les images d'algues sont des images texturées, NCUT génère des partitions inutiles qui peuvent fausser l'opération de recherche. Au contraire, comme notre système impose à plusieurs propriétés locales et globales d'interagir pour activer un groupement, il est capable d'effectuer plus de groupements avec moins d'erreurs. Ceci permet un gain certain en robustesse de recherche.

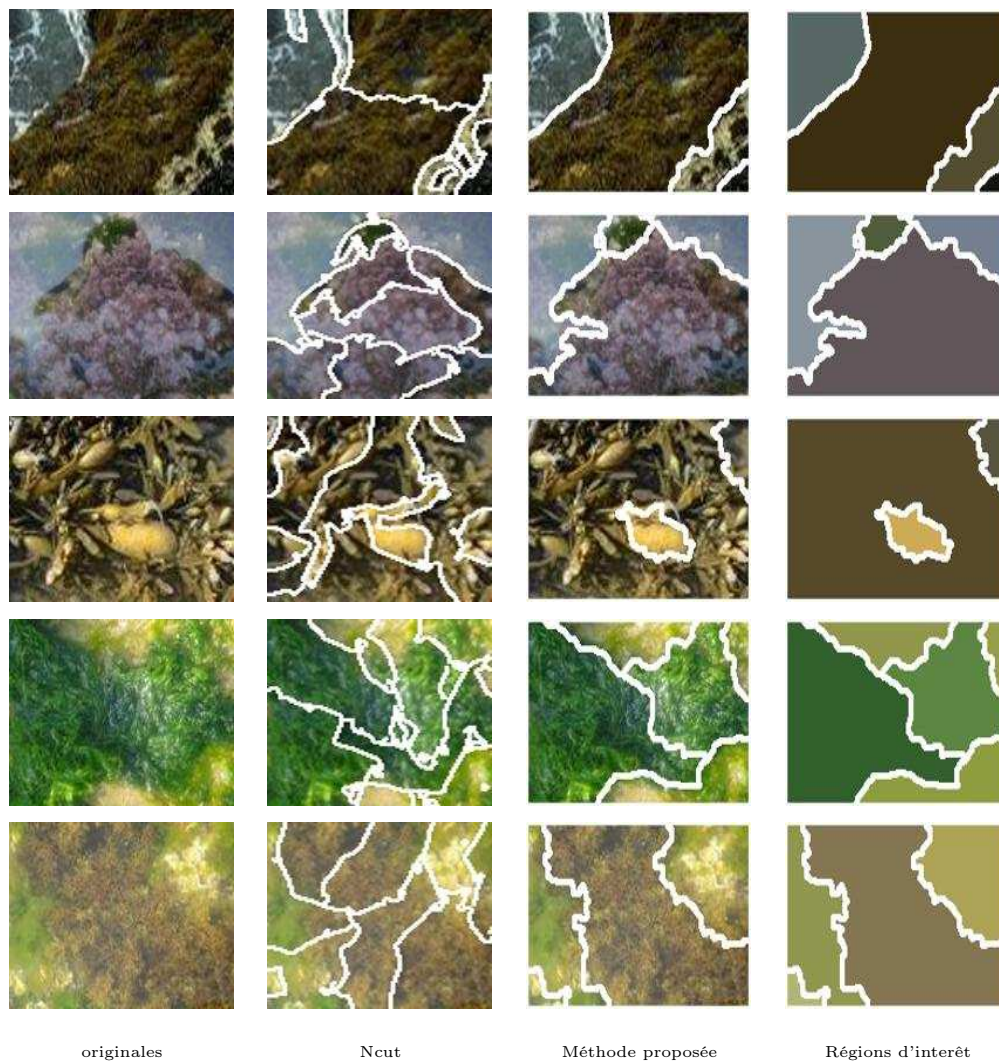


FIG. 3.9 – *Comparaison entre l'approche proposée et Ncut.*

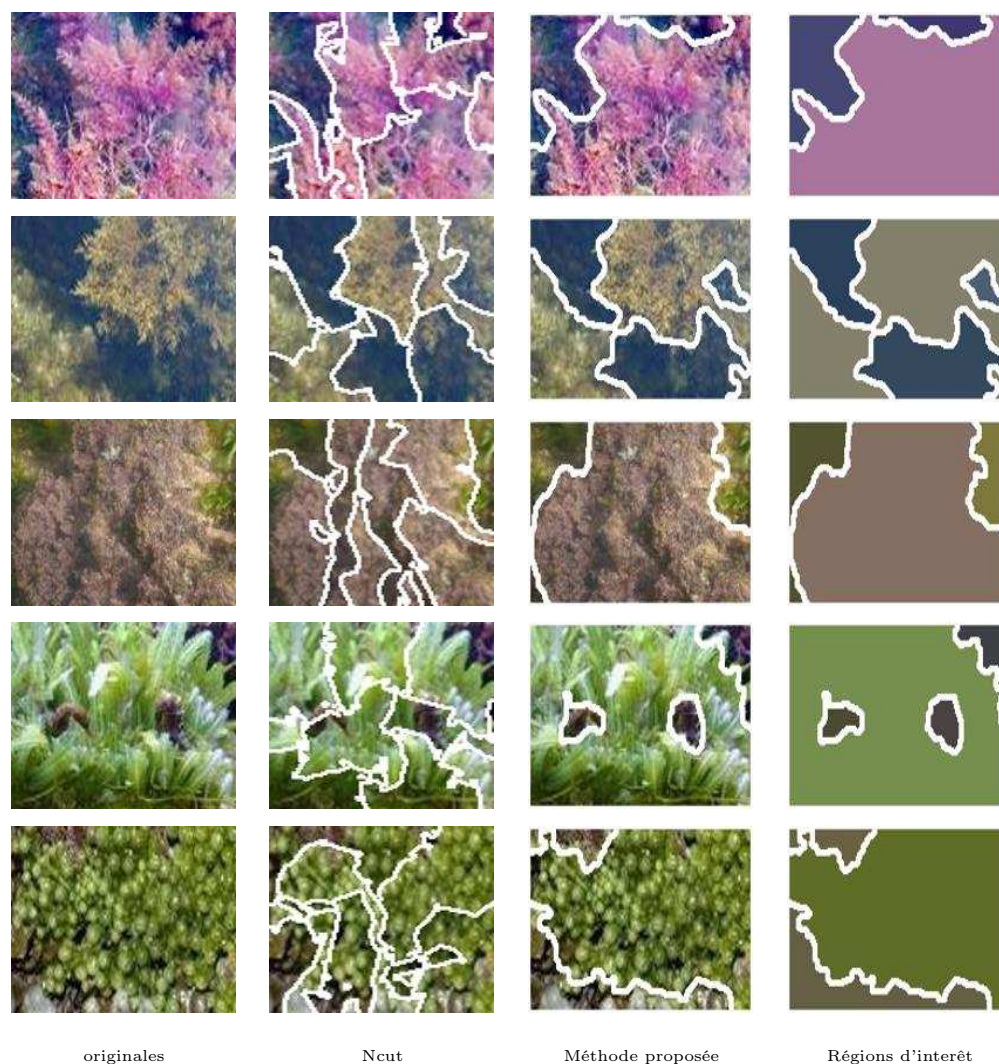


FIG. 3.10 – Comparaison entre l’approche proposée et Ncut

3.8.4 Exemple de résultats sur des images naturelles

Nous avons testé notre méthode de segmentation sur un échantillon d’images issues de la base bien connue Corel ©. Des exemples de résultats sont montrés dans la figure 3.11. La figure présente aussi les résultats obtenus par les systèmes de segmentation Ncut [116] et Blobworld [15], dont nous avons déjà parlé. Rappelons que ce dernier opère un regroupement de pixels sur la base de trois critères : Couleur, texture et proximité spatiale. Les zones affichées en gris correspondent à des régions qui ont été écartées des traitements. On constate que notre méthode permet de fusionner des régions qui ont des descripteurs de couleurs ou de texture légèrement différents, mais qui présentent tout de même une certaine unité. Ainsi, les animaux de la figure 3.11 sont correctement extraits du fond par notre méthode, alors que la segmentation Blobworld ou Ncut les laissent

partitionnés en plusieurs régions. On remarque aussi que, du fait des limitations inhérentes à la segmentation, les résultats de Blobworld ou Ncut présentent des régions dues aux conditions d'éclairages (image 4). Au contraire notre méthode est moins sensible aux conditions d'éclairage grâce aux propriétés spectrales ainsi que la non considération de la valeur de luminance dans la caractérisation de la couleur.



FIG. 3.11 – *Comparaison entre méthodes(scènes naturelles).*

3.9 Conclusion :

La recherche d'images basée sur les régions nécessite une détection préalable des objets ou zones d'intérêt, c'est le but de la segmentation. Nous avons donc commencé par l'énumération de quelques méthodes de segmentation d'images. Les techniques de partitionnement sont les mieux adaptées aux systèmes CBIR.

Dans ce chapitre nous avons présenté aussi les méthodes de partitionnement de graphes. Nous voulions, d'une part, souligner le lien entre le partitionnement et la diffusion. D'autre part, nous avons remarqué à quel point la diffusion sur graphe peut conduire à des techniques de segmentations robustes. A cet effet, nous avons proposé une méthode fondée sur la diffusion où les deux informations locales et non locales ont été utilisées. Les chapitres suivants présentent l'application de notre méthode dans l'indexation d'images et recherche interactive.

Chapitre 4

INDEXATION ET RECHERCHE D'IMAGES

4.1 Introduction

Dans le chapitre 2, nous avons évoqué que les systèmes d'indexation permettent généralement de rechercher des images d'après leurs caractéristiques bas niveaux comme la couleur ou la texture. Ces caractéristiques peuvent être extraites de l'image d'une façon globale sans distinguer entre les différentes régions existantes dans l'image. Néanmoins, si cet aspect est sans nul doute pertinent, l'utilisateur cherche le plus souvent une image d'après ce qu'elles représentent, en particulier les objets qu'elle représente. Ainsi, un utilisateur aimerait formuler une requête du type : je cherche dans les images d'algues celles qui contiennent le *gelidium sesquipedale* ou dans les bases généralistes celles qui contiennent des fleurs par exemple.

Dans ce chapitre nous nous focalisons sur la recherche d'images basée sur une description locale. Ceci nous a ramené à deux étapes. D'une part, les outils d'indexation doivent intégrer des caractéristiques images qui permettront de se rapprocher du niveau sémantique des images. En effet, nous l'avons déjà vu, le fossé sémantique entrave une description directe et non ambiguë. Toutefois, l'utilisation des descripteurs structurels permettent par exemple une première extension du pouvoir descriptif des données images.

D'autre part, un système de recherche d'images doit offrir aux utilisateurs une interface capable de les aider à formuler au mieux ce qu'ils recherchent. Le paradigme désormais classique de la requête par l'exemple, largement diffusé et reconnu, permet de retrouver des images similaires à une image requête. Durant les travaux de cette thèse, nous avons proposé un nouveau système d'indexation qui prend en compte ces constatations. La validité de nos méthodes est testée dans une application qui relève du domaine de la biodiversité : il s'agit des

images d'algues. Les résultats de la recherche et les performances supportent la comparaison de différentes méthodes.

4.2 Procédé d'extraction de connaissances

La fouille de données consiste à analyser, par des méthodes statistiques et mathématiques, un large volume de données, afin d'en faire ressortir des tendances ou des règles. Il correspond à l'ensemble des techniques et des méthodes qui, à partir de données, permettent d'obtenir des connaissances exploitables. En effet, les avancées théoriques ont permis de s'attaquer à des problèmes réels, qui en général impliquent de grandes bases de données. C'est ainsi que la fouille de données est née. Formellement donc, la fouille de données (ou data mining) a pour objectif l'extraction des informations utiles, auparavant inconnues et présentes de manière implicite à partir de grands ensembles de données.

La naissance de ce nouveau point de vue a donné un nouvel élan à la communauté et rapidement on s'est aperçu qu'il y avait de nouvelles problématiques. En particulier on s'est rendu compte que la fouille de données n'est qu'un maillon dans le processus de découverte de connaissance.

Ceci fut bien illustré par Fayyad et al [37] dans son schéma de l'extraction de connaissances à partir des bases de données (ECD) (figure 4.1).

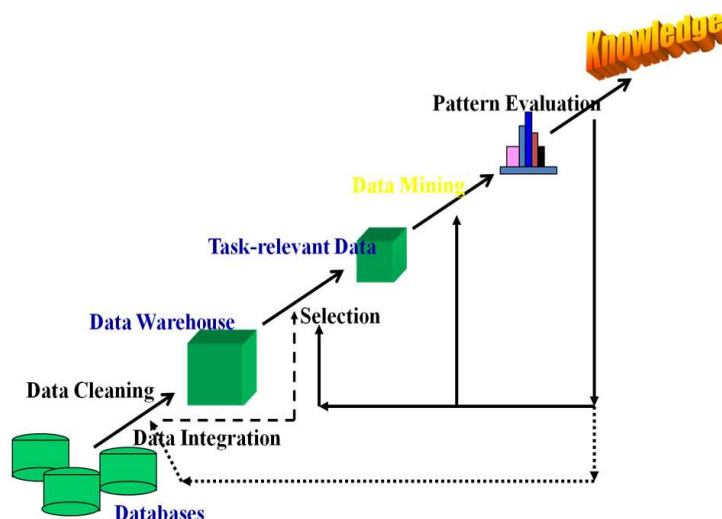


FIG. 4.1 – Schéma classique illustrant le processus d'extraction de connaissances [37]

Sur ce schéma, on voulait insister sur le fait qu'on part de données brutes qu'il faut stocker et gérer (typiquement dans des bases de données ou Data

warehouse). Puis on essaye de ne faire remonter que ce qui est utile pour appliquer les techniques de la fouille de données. Les résultats étant ensuite interprétés pour donner de nouvelles connaissances. Dans le cas des bases données d'images, on se rend compte que le problème est bien plus compliqué à cause de la nature même des données à la quelle s'ajoute le problème du fossé sémantique. Tout de même, ce schéma illustre ce que nous cherchons à faire : partir des données brutes pour arriver à des connaissances, et plus précisément dans notre cas, donner du sens au signal (la sémantique).

Dans le but de recherche d'images et de navigation dans une base d'images d'algues, nous présentons dans la figure 4.2, le procédé adopté dans notre travail en se basant sur le modèle d'extraction de connaissances à partir des données.

- (1) Sélection d'images d'essai : Vu la diversité des images d'algues provenant de différents sites et par différents moyens d'acquisition et de numérisation, il est fortement recommandé de construire une base prototype de cette population d'images qui soit la plus représentative.
- (2) Preprocessing : Les applications visant l'utilisation des images pour la segmentation, la classification, la recherche d'images ou autres, nécessitent une phase d'analyse du problème et la nature de ces images. Il s'agit dans cette étape de transformer les images de la base en paramètres numériques ayant une signification physique. L'extraction des descripteurs permet de projeter les images dans un espace ayant une dimension importante faisant appel aux techniques de traitement d'images. Le nombre de paramètres des caractéristiques obtenu est important et complexe à exploiter. Par voie de conséquence, il peut s'avérer judicieux de procéder à une réduction de l'espace de paramètres.
- (3) Construction de l'espace de signatures : La construction de l'espace de signatures consiste à réduire l'espace de représentation des images en un espace de dimension inférieure et à ne garder que les descripteurs les plus pertinents caractérisant au mieux l'image sans perdre leur signification physique du départ. La réduction de dimension peut aussi s'effectuer par l'extraction des descripteurs de la couleur et de la texture qui consistent à créer de nouveaux attributs par des combinaisons linéaires ou non-linéaires des paramètres initiaux.
- (4) Interprétation : Lorsque l'ensemble pertinent de descripteurs est identifié, différentes applications peuvent être envisagées. Parmi ces applications, nous citons la classification supervisée et non-supervisée, la recherche d'images, la reconnaissance d'objets, la navigation et le bouclage de pertinence.

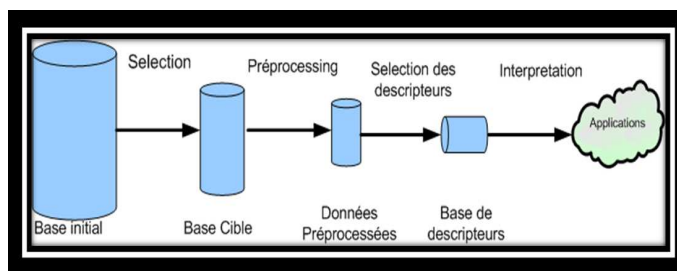


FIG. 4.2 – *Processus d'ECD pour la caractérisation d'images d'algues.*

4.3 SWIR : Application à une base d'algues

Les algues jouent un rôle primordial dans le maintien de l'équilibre écologique du milieu aquatique. Elles sont les principales responsables de la production primaire. De point de vue économique, elles présentent actuellement une source nutritionnelle et un produit à valeur montante (Figure 4.3 et table 4.1). Elles sont utilisées en agriculture comme engrais et fourrage, dans l'industrie alimentaire et pharmaceutique, dans le textile et dans d'autres domaines. Les algues marines sont l'une des plus importantes ressources marines vivantes. Elles pourraient être qualifiées comme la future prometteuse plante [32, 2]. Ces plantes sont la source de nourriture, d'alimentation et de médecine. Agar-agar, Carrageenan et Alginate sont des exemples populaires de produits extraits des algues. Celles-ci ont été utilisées comme nourriture pour les êtres humains, aliments pour animaux, engrais pour les plantes et la source de divers produits chimiques. Les produits à base d'algues sont utilisés dans notre vie quotidienne en l'un ou l'autre manière. Par exemple, certaines algues polysaccharides sont employées dans la fabrication de dentifrices, savons, shampooings, produits cosmétiques, lait, glaces, aliments transformés, désodorisants d'air et une foule d'autres produits [32, 2]. La table 4.2 présente quelque exemple d'utilisation de certaines espèces d'algues. Toutefois, L'apparition croissante d'algues nuisible ou HAB («Harmful Algal Blooms») en Méditerranée a conduit à de graves problèmes de santé publique causés par la consommation de fruits de mer contaminés par des algues toxiques.

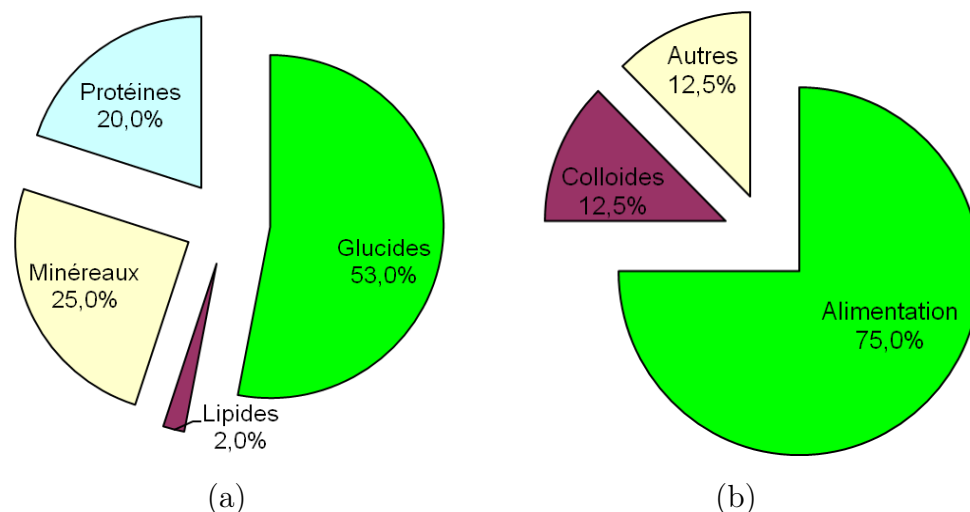


FIG. 4.3 – Répartition moyenne. (a) Intérêt nutritionnel des algues, (b) Utilisation des algues dans le monde [84].

TAB. 4.1 – Valeurs nutritive de quelques espèces d’algues [32].

Espèce	Proteine	Lipide	Carbohydrates	Ash
Ulva	26.1	2.1	42.0	7.8
Enteromorpha	19.5	0.3	64.9	15.2
Monostroma	20.0	1.2	63.9	14.9
Laminaria	16.1	2.4	39.3	19.6
Alaria	17.1	3.6	39.8	14.9
Sargassum	19.0	2.9	33.0	16.2
Padina	18.81	1.7	31.6	10.3
Porphyra	28.4	4.5	45.1	6.9
Rhodymenia	21.5	1.7	44.6	5.3
Gracilaria	24.37	1.8	61.75	11.3

L’objectif de cette étude est de gérer et d’utiliser les métadonnées multi-média pour faciliter l’accès à la diversité biologique de notre patrimoine algal. Ceci est fait en tenant compte, d’une part, des outils d’analyse et de traitement de l’image pour la description du contenu de ces images, et, d’autre part, de l’installation d’un système de navigation et de recherche dans une telle base. On aimerait souvent savoir les informations liées à une image d’algue et la comparer avec les images proches. Il y a donc un réel besoin d’automatiser pour une utilisation plus rapide et efficace. Parmi les applications de ce système de recherche d’images d’algues est la reconnaissance automatique des principales espèces d’algues sur des clichés sous marines. Un dispositif de prise en vue sous marines a pour but de prendre des images des fonds marins où croît l’espèce. Cela, permettrait d’éviter le risque que prennent les plongeurs par mauvais temps. D’où

l'idée de remplacer ces prélèvements par des analyses automatisées d'images sous marines afin de diminuer le coût de réalisation de la cartographie et d'étendre les zones prospectées.

Aussi peut il être utilisé en système éducatif où l'apprenant peut extraire des informations sur une espèce d'algue donnée et la rapprocher avec les espèces déjà existantes dans la base. La disponibilité des collections d'images pourrait aussi aider le formateur à donner des exemples et des illustrations afin de réduire le temps de préparation et d'améliorer la qualité de l'enseignement.

Également, en botanique et en particulier dans l'étude de l'expression des gènes, on aimerait souvent rapprocher les images pour étudier les effets des modifications des gènes et faire les conclusions adéquates d'une façon plausible et efficace.

SWIR (SeaWeed Image Retrieval) est un moteur de recherche d'images d'algues par le contenu. Son objectif est de pouvoir retrouver des images d'algues à partir de signatures visuelles calculées par des descripteurs d'images. L'application se présente selon une architecture client-serveur. Le client est une interface graphique mettant en évidence les similitudes entre les images. Le serveur analyse les images pour en extraire des descripteurs visuels, et effectue des comparaisons avec les requêtes formulées par le client. Il gère également les bases de données d'images ainsi que les annotations, ce qui permet aux utilisateurs d'avoir accès à la base et à des informations supplémentaires sur les images : l'espèce, groupe taxonomique, caractéristiques, habitat, distribution, Usage,...

Des travaux de recherche, actuellement en cours de finalisation, utilisent SWIR pour la reconnaissance d'espèces d'algues à partir d'images de leurs organes constitutifs ou via une description textuelle. Il est ainsi possible de retrouver en très peu de manipulations les images associées à une description donnée par l'utilisateur (couleur, taille et forme du stipe, habitat,...).

TAB. 4.2 – *Quelques exemples d'utilisation de certaines espèces d'algues [32].*

Espèce	Alimentation humaine	Alimentation animale	utilisation industrielle	Medecine	Fertilisant
Ulva fasciata	+	+	-	+	-
Enteromorpha compressa	+	+	-	+	-
Monostroma oxyspermum	+	+	-	-	-
Cladophora fascicularis	+	+	-	-	-
Chaetomorpha media	+	+	-	-	+
Codium fragile	+	+	-	+	-
Caulerpa sertularioides	+	+	-	-	-
Dictyota dichotoma	+	+	+	-	-
Spatoglossum asperum	-	-	+	-	+
Hydroclathrus clathratus	-	-	+	-	+
Stoechospermum marginatum	-	-	+	-	+
Colpomenia sinuosa	-	-	+	-	+
Dictyopteris australis	-	-	+	-	+
Padina tetraströmatica	-	-	+	-	+
Sargassum cinereum	-	-	+	+	+
Sargassum ilicifolium	-	+	+	+	+
Laminaria digita	-	-	+	+	
Macrocystis pyrifera	-	-	+	+	+
Porphyra vietnamensis	+	+	-	-	-
Amphiroa fragilissima	+	-	-	-	-
Jania adhaerens	-	-	-	+	-
Gracillaria corticata	+	+	+	-	-
Hypnea musciformis	+	+	+	-	-
Centroceros clavulatum	+	-	+	-	-
Laurencia papillosa	+	+	+	-	+
Chondrus crispus	+	-	+	-	-
Eucheuma uncinatum	+	+	+	-	-
Gelidiella acerosa	-	-	+	-	-

SWIR est utilisé pour la reconnaissance des algues marines. Certaines des algues rouges menacées sont exploitées industriellement pour l'extraction de l'agar-agar. *Gelidium sesquipedale*, l'espèce la plus recherchée pour son rendement élevé en agar, accuse une régression drastique due à sa surexploitation et dans une certaine mesure à la pollution, et ce malgré sa grande capacité de régénération végétative; sa récolte, autrefois par simple ramassage à marée basse, nécessite aujourd'hui la plongée à des profondeurs de plus en plus importantes. D'où la nécessité du développement d'outils d'identification rapides et accessibles par des non-spécialistes afin d'assurer une exploitation durable de ces ressources. Or, cette famille d'algues rouge est complexe et l'identification est reconnue comme étant difficile. Par ailleurs, il n'existe que peu de ressources multimédia fiables sur ces plantes.

4.4 Description du contenu

La description du contenu visuel se fait généralement par des caractéristiques bas niveaux comme la couleur, la texture ou la forme de zones définies mais encore des caractéristiques qui peuvent être plus élaborées telles que les points d'intérêts, géométrie ou mesure des contours,...La sémantique image, quant à elle, va beaucoup plus loin que ces attributs visuels. Mais il est difficile de la définir exactement et surtout de produire des règles pour son expression. Plusieurs méthodes automatiques sont proposées introduisant la sémantique ou son exploitation accrue dans la reconnaissance d'image via des relations d'interférence entre entités visuelles au sein de l'image. Les méthodes manuelles se contentent d'associer à l'image du texte exploitable lors d'une requête. En l'absence d'une caractérisation rigoureuse de la sémantique image on parlera d'une sémantique limitée, codée via des attributs associés aux entités extraites de l'image. Ces attributs réfèrent de manière plus ou moins exacte et directe aux intentions ou à l'action portées par la requête. Le contenu visuel sera donc pour nous une caractéristique importante dans la mesure de ressemblance entre images car il offre un pouvoir discriminant élevé.

Les performances d'un système de recherche d'images par le contenu dépendent principalement de la base d'images utilisée, de la nature des descripteurs et attributs visuels ainsi que leurs méthodes de calcul qui peuvent être de type globale ou local. L'approche globale consiste à extraire des caractéristiques représentant l'image entière par un traitement impliquant tous les pixels de l'image : l'histogramme ou correlogramme de l'image sont par exemple souvent utilisés pour cette fin. Une description globale s'adapte mal aux requêtes partielles où l'on fait correspondre des parties d'images. L'approche locale au contraire per-

met une description partielle offrant ainsi une représentation plus fine et plus naturelle. L'approche locale opère sur des partitions de l'image. La manière la plus simple pour obtenir ces partitions est la division de l'image en blocs égaux dont on extrait des descripteurs. On peut alors calculer ces descripteurs par blocs simplement avec les mêmes procédures que celles utilisées globalement (e.g. histogrammes). Dans ce type de partitionnement les régions ne sont pas significantes. Il peut donc s'avérer plus raisonnable d'utiliser un partitionnement en régions homogènes, qui relève de la segmentation.

4.4.1 Description locale

Notre procédure se base essentiellement sur la décomposition de l'image en entités visuelles à travers une segmentation en régions (C.F. chapitre 3). Les différentes parties de l'image issues de cette décomposition sont chacune représentées par des descripteurs visuels. La figure 4.4 montre un exemple de segmentation de l'image en plusieurs objets. On remarque que chaque région présente une certaine homogénéité visuelle. Les applications que nous visons, images d'algues, concernent des images où la texture est une caractéristique importante. Pour représenter cette dernière nous utilisons des descripteurs de type filtres de Gabor introduits dans le chapitre 2.



(a) Image originale

(b) image segmentée

FIG. 4.4 – *Segmentation d'une image en entités visuelles*

La couleur joue aussi un rôle très important dans la discrimination des algues. Ces dernières se sont scindées en trois principales couleurs : algues rouges, algues vertes et algues brunes. La couleur est représentée par les moments de couleurs. Notre choix de ces descripteurs est motivé par leurs puissances de discrimination des textures et de la couleur.

L'extraction de descripteurs locaux est définie comme suit : Soit une image décomposée en M régions, chaque région R est représentée par son vecteur ca-

ractérisant F , l'ensemble des descripteurs est défini par :

$$Ed = \{d_1, d_2, d_3, \dots, d_m\}$$

Avec :

$$d_i = (R_i, F_i)$$

$$F_i = S(R_i)$$

S est un opérateur qui calcule des attributs bas niveaux ou métriques. Un descripteur est représenté par le couple : région et vecteur caractéristiques associé. La procédure de l'extraction des composantes visuelles est présentée dans la figure 4.5

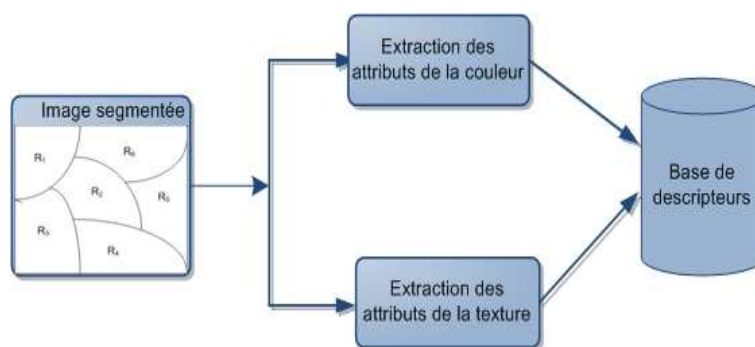


FIG. 4.5 – *Extraction des attributs de la couleur et de la texture*

4.5 Techniques de recherche

Au cours de cette section nous détaillons les différentes stratégies de recherches employées pour traiter une requête. Quand un utilisateur présente une image requête au système, on ne sait pas s'il cherche une région particulière ou des régions multiples. Un système fiable devra prendre en compte les différents critères. Par analogie avec les systèmes de recherches textuelles on peut autoriser plusieurs options pour composer sa requête : par exemple un mot seul, plusieurs mots dans le texte ou plusieurs mots associés. Une telle stratégie offre une marge de manoeuvre supplémentaire à l'utilisateur pour raffiner le résultat. Dans ce contexte notre système supporte ces modes de recherches : recherche précise et recherche approximative. Sur le plan technique la recherche par région est facile à implanter et les performances ne dépendent que des descripteurs visuels. En revanche dans le cas de plusieurs régions, la similarité est contrainte par les caractéristiques visuelles mais aussi par la méthode de mise en correspondance entre

composantes visuelles respectives de la requête et de la cible. Nous aborderons cet aspect dans la suite de ce chapitre. Soulignons que ce n'est pas l'utilisateur qui segmente l'image requête, mais il utilise l'image segmentée fournie par la machine.

4.5.1 Requête par région

La méthode de recherche suivie à ce niveau consiste à extraire toutes les images qui comportent des régions similaires à celles distinguées dans l'image requête. La région requête est donc comparée à toutes les régions de l'image cible de la base, celle qui affiche une distance faible entre attributs sera retenue. Les attributs utilisés dans ce cas sont de type visuel.

La mesure de similarité entre deux régions requête R_i et cible R'_j est déterminée par une distance Euclidienne appliquée aux vecteurs de caractéristiques.

$$dist(R_i, R'_j) = \left(\sum_{k=1}^n (f_k - f'_k)^2 \right)^{1/2} \quad (4.1)$$

Où f_k et f'_k représentent respectivement les composantes des vecteurs de caractéristiques de R_i et R'_j . n est donc la dimension du vecteur caractéristique.

4.5.2 Requête par plusieurs régions

Dans ce type de recherche tous les segments de l'image sont impliqués. L'objectif ici est d'obtenir le nombre maximum de régions requêtes correspondant à celle de la cible. Quand il s'agit d'une recherche par région singulière ou d'une approche globale, la mise en correspondance des images ne pose pas de problème pour estimer la similarité. Par contre, le problème devient plus complexe avec une requête à plusieurs régions car il faut satisfaire un critère global impliquant toutes les composantes, ce critère réalisant un compromis entre par exemple le nombre de régions impliquées dans l'appariement global et le degré de ressemblance dans chaque couple de régions. Le mécanisme de la mise en correspondance est illustré par la figure 4.6 où chaque noeud du graphe correspond à une région dans l'image.

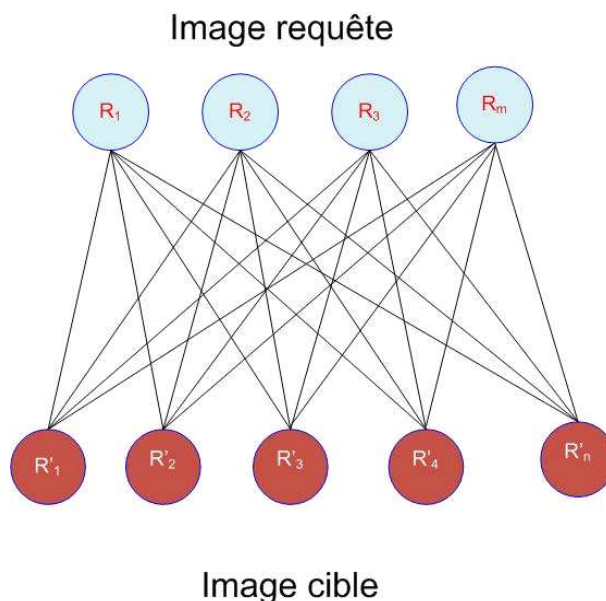


FIG. 4.6 – *Comparaison des régions entre l'image requête et une image de la base*

L'appariement de deux images avec régions en nombres différents pose les difficultés supplémentaires du choix des intervenants dans le calcul de distance. Une solution est de garder les meilleurs couples, un sous graphe est alors considéré. D'autres solutions consistent à rajouter des noeuds fictifs pour compenser la différence en nombre de régions [68]. De plus la segmentation ne fournit pas toujours un résultat fiable. Les éventuelles erreurs de segmentation conduisent également à une mise en correspondance de sous graphes qui favorise les recoupements. D'autres méthodes reposent sur des principes différents visent à compenser les erreurs de segmentation en testant plusieurs appariements simultanés ou en fusionnant des noeuds d'après les premiers résultats d'appariement. La plus populaire est la méthode "Integrated Region Matching" [76] qui consiste à donner une grande priorité aux paires de régions présentant une distance minimale i.e. ressemblance maximale. Elle s'appuie principalement sur les distances entre attributs. Dans notre méthode, on tient compte de ces notions principales. Nous la décrivons dans la section suivante.

Distance de Hausdorff moyenne

L'appariement bidirectionnel est une méthode qui repose sur un appariement à double sens, autrement dit : mesurer la similarité des zones d'intérêt requêtes vers les zones de la cible et vice versa. Cette technique permet d'estimer la préférence de la requête entre les cibles d'une part, et d'autre part à la cible d'exprimer son taux d'attraction selon une formule identique. Pour palier au

problème de l'incertitude de la segmentation automatique, on fait correspondre un objet à plusieurs objets homologues. Un plus grand poids est attribué aux couples qui se ressemblent. L'introduction de cette méthode peut s'avérer pertinente dans le calcul de similarité afin de pouvoir réduire les erreurs que peut produire une requête avec régions incomplètes. La similarité est assurée par l'aspect bidirectionnel de la méthode. Si une région, par exemple, appartient à un objet dans l'image requête correspond à plusieurs régions dans l'image qui appartiennent à des objets différents. L'objet le plus similaire dans l'image cible est celui qui correspond au mieux à la région requête. Empiriquement, on a vérifié que l'appariement bidirectionnel est plus pertinent que chacun des deux après mises en correspondance monodirectionnel.

Soit $Q = \{R_1, R_2, \dots, R_m\}$ l'image requête et $C = \{R'_1, R'_2, \dots, R'_m\}$ une image cible de la base. La distance entre Q et C est basée sur l'appariement des régions, en comparant chaque segment de Q avec tous les segments de C. La distance de Hausdorff moyenne (Average Hausdorff Distance AHD) peut s'écrire sous la forme suivante :

$$h_{avg}(Q, C) = \frac{1}{m} \sum_{a \in Q} \min_{b \in C} dist(a, b) \quad (4.2)$$

$$h_{avg}(C, Q) = \frac{1}{n} \sum_{b \in C} \min_{a \in Q} dist(b, a) \quad (4.3)$$

Le calcul de la distance finale permet de déterminer le degré de similarité entre deux images. La mesure doit inclure le coût global de mise en correspondance, pour cela nous utilisons la formule suivante :

$$D(Q, C) = (h_{avg}(Q, C) + h_{avg}(C, Q)) / 2 \quad (4.4)$$

Nous constatons ici que la distance entre Q et C est le résultat de la double mise en correspondance qui exprime leur interaction.

4.6 Implémentation

Dans les sections précédentes, nous avons décrit le mode d'implémentation d'une mesure de similarité entre images. Nous allons maintenant récapituler le processus complet de recherche d'image. Le schéma général de système de recherche est constitué de plusieurs procédés dont chacun s'occupe d'une tâche spécifique. Le processus de traitement, de l'indexation à la recherche, est entièrement automatisé et l'intervention humaine, à ce stade, est presque absente et ne porte que sur quelques paramètres. On distingue deux phases de traitement,

hors ligne et en ligne, qui favorisent la recherche et la navigation de l'utilisateur. La figure 4.7 présente l'architecture du système de recherche proposé

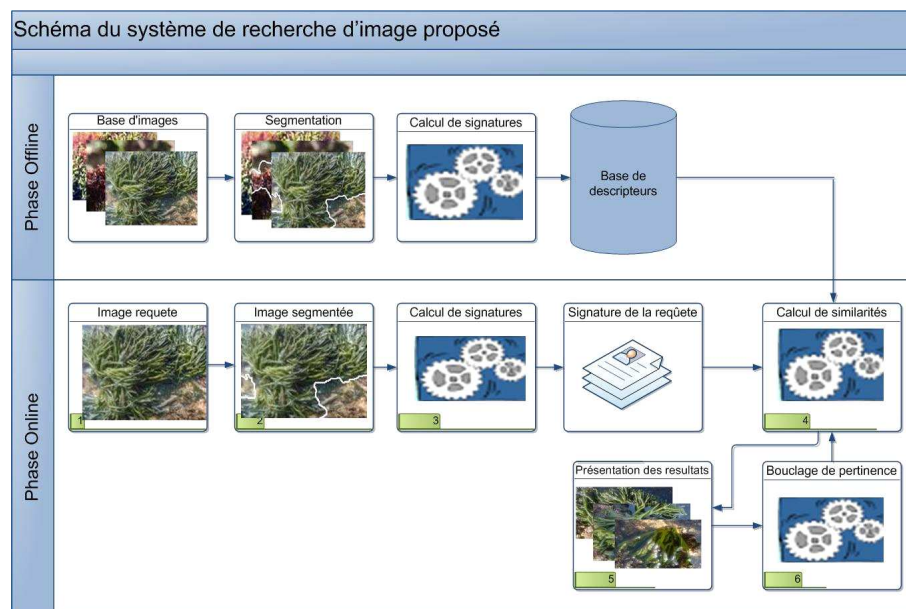


FIG. 4.7 – L'architecture du système de recherche proposé

4.6.1 Phase d'indexation

La phase d'indexation est une tâche qui s'effectue sans véritable contrainte temporelle puisqu'elle s'effectue hors ligne et elle est indépendante du processus de recherche. Elle permet d'extraire une signature de chaque image sur une base d'attributs. La phase d'indexation de notre système se compose de trois étapes principales : la segmentation pour extraire les composantes visuelles, l'extraction des signatures caractérisant chaque image puis leur stockage dans un fichier plat (figure 4.7). Ce fichier est utilisé par la suite comme base de caractéristiques images. En effet, toute modification de la base nécessite une mise à jour de ce fichier de descripteurs.

Nous allons maintenant récapituler le processus complet d'indexation. Soit I l'image à indexer :

- (1) L'image I est segmentée en régions par l'algorithme basé sur la diffusion locale détaillé dans le chapitre précédent.
- (2) On crée, à partir de I une liste d'objets, possédant des attributs de couleur et de texture. On calcule, pour chaque objet, un ensemble de signatures sous forme de vecteurs, possédant chacun un nombre variable de dimen-

sions et rattachés à une métrique spécifique. Les statistiques (moyenne et écart-type) de ces descripteurs sont remises à jour et sauvegardées à chaque phase d'indexation permettant ainsi une future normalisation de la distance dans la phase de recherche.

- (3) La signature est ensuite sauvegardée dans un fichier ;

La phase d'indexation représente la phase la plus importante en termes de temps de calcul. Ce dernier est très largement lié au nombre de descripteurs utilisés ainsi qu'à leurs complexités. Nous avons implémenté des méthodes pour l'extraction de la couleur et de la texture déjà étudiées dans le chapitre 2.

4.6.2 Procédé de recherche

Dans le processus de recherche, l'utilisateur présente l'image ou partie de l'image clé, le système explore la base pour trouver les images les plus similaires. Le traitement en ligne de notre système comporte deux parties : d'une part, il compare l'image ou partie de l'image clé avec les images de la base et affiche les résultats. La procédure de comparaison exploite l'une des techniques de recherche décrite dans la section 4.5 en se basant sur la base de caractéristiques. L'affichage des images retournées se fait suivant le score engendré par la procédure de comparaison. D'autre part, si l'utilisateur n'est pas satisfait des résultats de recherche, il a la possibilité de fournir des précisions quant à la catégorie qu'il recherche, par exemple sous la forme d'annotations. A l'aide de ces précisions, le système peut calculer un nouveau classement et présenter de nouvelles images. Cette opération est baptisée bouclage de pertinence et va être détaillée en chapitre 5.

4.6.3 Interface utilisateur

En raison de sa facilité de développement et de ses possibilités de traitement d'images, les premiers essais de la méthode ont été réalisés sous Matlab. Les temps de calculs étaient assez bons pour de petites bases d'images. Toutefois les performances commencent à se dégrader avec l'augmentation de la taille de la base. Matlab a démontré la faisabilité de la méthode mais une solution plus rapide et pratique a été préférée.

Notre système de recherche d'images par le contenu est orienté client serveur pour permettre d'effectuer une recherche sur une base d'images à partir de n'importe quel navigateur Internet. La volonté a été de créer des pages web dynamiques qui réagissent aux actions de l'utilisateur.

Le système se compose donc d'un serveur web qui gère des pages web dynamiques capables d'accéder à des données d'une base d'images sous l'envi-

ronnement UNIX. Les clients se connectent au système par le protocole HTTP et interagissent avec lui à l'aide de formulaires HTML.

Un utilisateur peut rechercher une image ou catégorie d'images à l'aide de l'interface graphique dont une capture d'écran est présentée en figure 4.8. L'interface se décompose en deux parties principales. La première, qui prend la majeure partie de l'écran, présente une partie du classement de la base par le système. Par exemple, sur la figure 4.8 nous pouvons voir les 8 images les plus pertinentes selon le système. Le classement est fait de gauche à droite puis de haut en bas. L'image en haut à gauche est l'image la plus pertinente, puis celle à sa droite est la deuxième plus pertinente, etc. La deuxième partie se trouve en haut du classement de la base et présente l'image requête et la région sélectionnée.

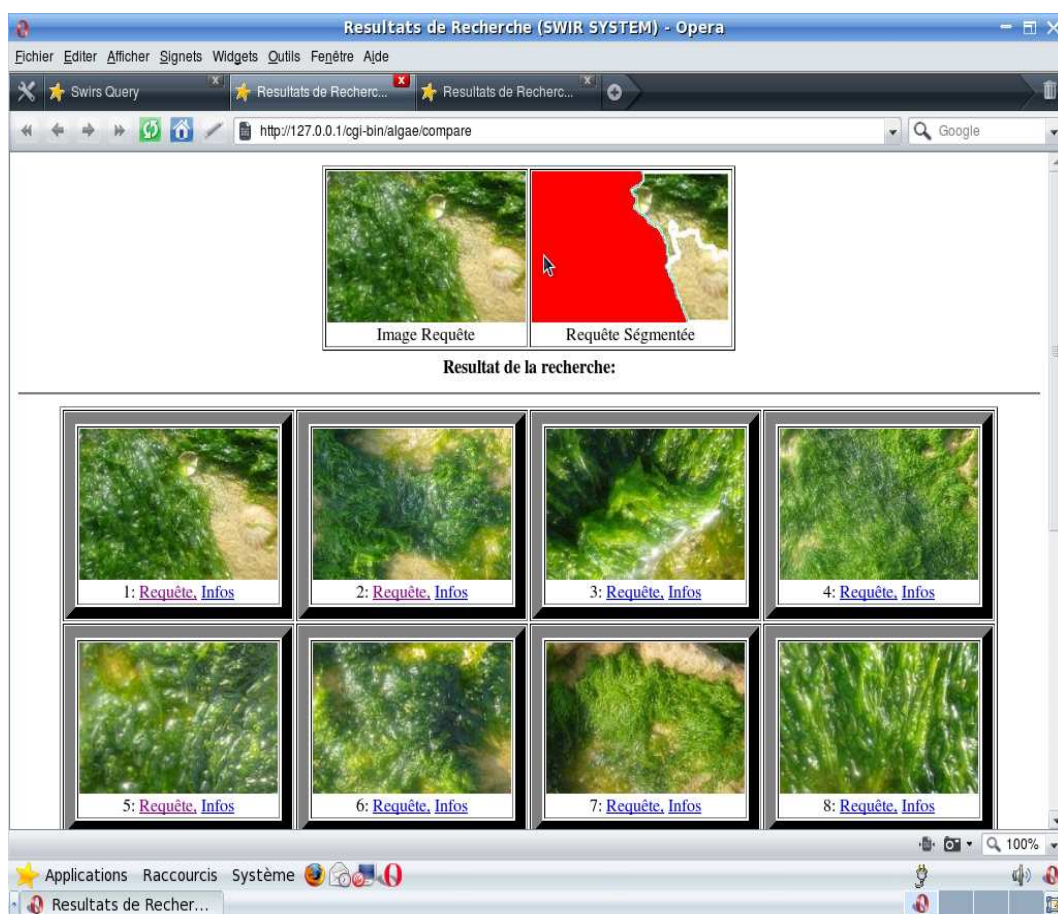


FIG. 4.8 – Interface Web du moteur de recherche d'images SWIR

4.7 Evaluation des performances

La mesure de pertinence d'un système reste l'une des grandes problématiques dans la recherche d'images par le contenu. Cela provient du fait que les

utilisateurs habilités à juger les performances du système n'ont pas toujours un sens commun. Par suite, L'élaboration d'une plateforme d'évaluation standard en CBIR se heurte au fait que les humains ne sont pas d'accord sur la pertinence des résultats retournés par un système ou l'autre. La solution basique consiste à établir une base d'images communes avec des classes où le jugement de pertinence serait commun. La difficulté d'adoption de cette solution vient de la diversité des applications visées qui impliquent des images spécialisées tel que notre cas, mais encore de la subjectivité liée à l'action envisagée ou au but poursuivi qui motive l'opération de recherche.

Malgré cela, plusieurs efforts ont été déployés pour créer des collections diversifiées afin de répondre aux différents domaines de spécialisations. La base la plus répandue dans le domaine de recherche d'image est la base généraliste Corel [27] où l'on peut trouver des groupes contenant des images jugées comme similaires [15]. Ces groupes favorisent la mesure de pertinence et permettent de distinguer entre différentes classes qui contiennent des sujets différents (roses, tigres, bâtiments, . . .). Pour juger leurs résultats, plusieurs recherches ont utilisé cette base. Toutefois certains travaux basés sur des applications spécifiques ont vu le jour. Pour ce genre d'applications l'expertise est demandée afin de juger la pertinence de recherche. En effet, ces applications génèrent plus de difficulté que les images généralistes. Notre travail vise justement de tels domaines spécifiques comme celui de la biodiversité et plus précisément les images d'algues. Il n'existe pas des bases communes organisées (Benchmark) pour tester la recherche de ces images vue que nous sommes les premiers qui avons abordé ce type d'application d'après nos recherches bibliographiques.

4.7.1 Rappel et Précision

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. La construction de cette vérité terrain est liée à notre propre expertise. Nous nous conformons à l'utilisation des critères d'évaluation classiques et de référence [121, 115]. Appelons A l'ensemble des images pertinentes (au sens d'une requête donnée) de la base et B l'ensemble des images retournées par le système. Nous définissons les critères suivants :

$$Précision = \frac{|A \cap B|}{|B|} = \frac{\text{nombre d'images pertinentes retrouvées}}{\text{nombre d'images retournées}} \quad (4.5)$$

$$Rappel = \frac{|A \cap B|}{|A|} = \frac{\text{nombre d'images pertinentes retrouvées}}{\text{nombre total d'images pertinentes}} \quad (4.6)$$

Le système est utilisé en recherche par exemple : une unique image exemple est proposée au système et ce dernier retourne les k images les plus proches de cette requête au sens de la fonction de similarité employée. La précision et le rappel sont estimés pour un nombre fixé d'images retournées par le système. En faisant varier ce paramètre, il devient possible de décliner en nombre de mesures de qualité différentes [121]. Pour notre part, nous employons principalement la courbe de rappel-précision présentant les variations de la précision en fonction de celles du rappel. Par ailleurs, l'évaluation est caractérisée par les deux propriétés suivantes :

- la précision et le rappel sont systématiquement moyennés sur plusieurs requêtes (le nombre de recherches est précisé pour chaque test),
- les performances sont établies et classées par catégorie dont le nombre d'images par catégorie est fixe, car les variations d'effectifs d'une classe à l'autre ainsi que la variabilité du contenu des images au sein d'une catégorie donnée se traduisent par d'importants changements de précision et de rappel.

4.7.2 Base d'images

Pour effectuer notre évaluation, nous avons utilisé une base d'images contenant environ 1000 images d'algues (voir figure 4.9). La base que nous avons construite contient les différentes espèces d'algues et qui sont divisées en trois principales taxonomies : algues rouges, algues vertes, et algues brunes.

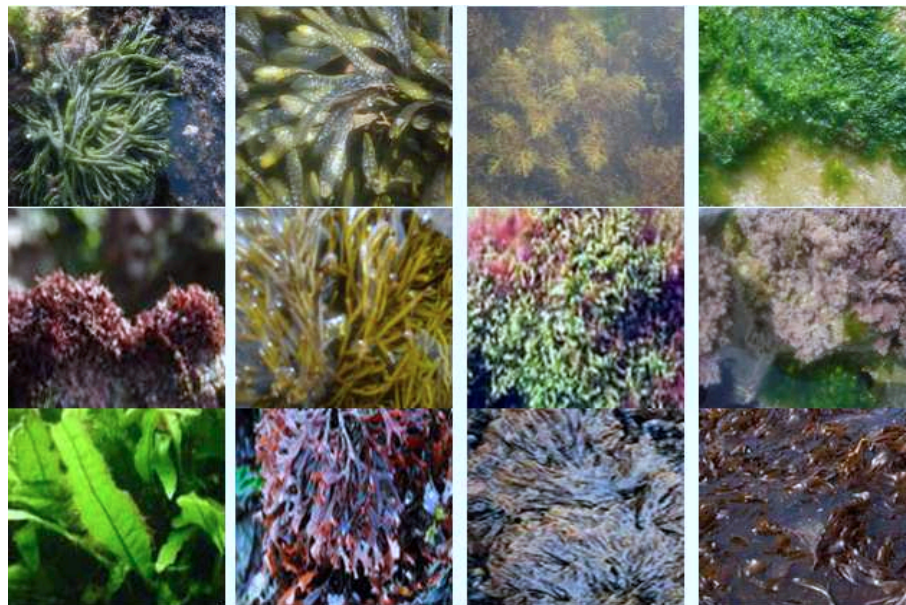


FIG. 4.9 – *Exemple d'images utilisées dans cette étude*

Pour favoriser une meilleure mesure de pertinence, les images similaires sont triées en groupes. Les images d'un groupe sont des images contenant la même espèce d'algue. Le choix des images avec la même espèce facilite le jugement de pertinence même avec un utilisateur non expert. Nous avons isolé 40 groupes, chacun d'entre eux contenant 25 images sémantiquement proches. Ces 40 groupes représentent notre vérité terrain et l'ensemble des images requêtes proviennent de ces mêmes groupes.

Les tests sont effectués dans les mêmes conditions pour toutes les caractéristiques afin de garantir une cohérence des mesures.

4.7.3 Signature visuelle

La mise au point de signatures visuelles performantes, implantées efficacement, est un préalable à la recherche d'image par le contenu. Chaque image est représentée comme un ensemble de vecteurs (des caractéristiques ou signatures) et une fonction distance est utilisée pour faire la recherche. SWIR contient plusieurs signatures de qualité qui couvrent les principaux aspects de la description des images d'algues : couleur et texture. Ces descripteurs sont régis par le paramétrage suivant :

- Matrice de Cooccurrence (GLCM) : distance=1, orientations= 0° , 45° , 90° , 135° avec symétrie. Dimension de l'espace de caractéristiques : $n=16$.
- Filtres de Gabor : 3 fréquences ($\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$) et 4 directions, $\theta = 0^\circ$, 45° , 90° , 135° , Dimension de l'espace de signatures : $n=12$.

- Moments d’ondelettes (WM) : Moyenne et déviation standard des coefficients de la transformée pour 3 décompositions en ondelettes, $n=20$.
- Moments de couleurs, Dimension de l’espace de caractéristiques : $n=9$.

Une implantation efficace de ces signatures a été réalisée et évaluée pour choisir la meilleure combinaison. Les signatures ont été choisies pour avoir de bonnes performances sans trop augmenter la dimension de l’espace de représentation résultant. On peut voir dans la table 4.3 une comparaison des performances de ces signatures.

TAB. 4.3 – *Précision Moyenne pour choix de descripteurs.*

Caractéristiques	Texture seule	Combinée avec la couleur
Gabor	0.24	0.36
WM	0.22	0.31
GLCM	0.20	0.28

4.7.4 Recherche globale

Dans ce type de recherche, l’utilisateur est intéressé à l’image toute entière. Nous étudions dans cette section le comportement du système vis-à-vis de la méthode d’indexation locale de différentes régions existantes dans l’image et la comparer avec l’indexation globale sans passer par la segmentation automatique de l’image. La figure 4.10 présente un exemple de résultat de recherche globale, avec 5 images pertinentes. L’image requête (en milieu haut) est composée de quatre régions. Nous remarquons sur le résultat que la plupart des images retournées comprennent ces régions. Les autres images que nous considérons non pertinentes contiennent des régions d’aspect visuel similaire à la requête.

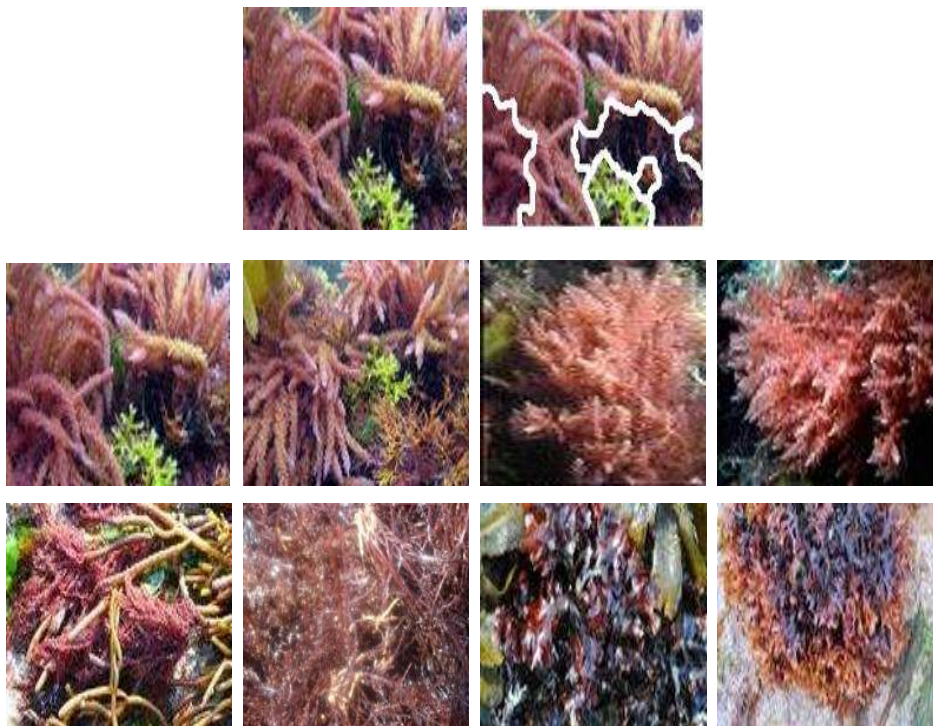


FIG. 4.10 – *Exemple de requête globale.*

Indexation locale vs indexation globale

Dans cette section nous allons comparer deux méthodes d'indexation : indexation globale, et indexation locale. Dans la première approche, nous considérons l'aspect visuel de l'image toute entière sans tenir en compte des objets ou des régions qui la forment. D'une autre manière ; dans l'indexation globale on considère l'image entière comme une seule région et on fait extraire les descripteurs qui lui sont afférents. Pour l'indexation locale, après segmentation de l'image, les descripteurs de chaque région sont stockés dans la base d'index et sont considérés dans la mesure de similarité.

La figure 4.11 montre la réponse que notre système fournit pour ces deux méthodes. La figure représente les courbe Rappel / Précision pour 4 espèce d'algues : *Gelidium Sesquipedale*, *Codium Fragile*, *Fucus Spiralis* et *Blidinga minima*. Pour les mêmes descripteurs utilisés, les courbes montrent une amélioration signifiante apportée lors de l'utilisation de l'approche locale qui reste relativement lente par rapport à l'approche globale.

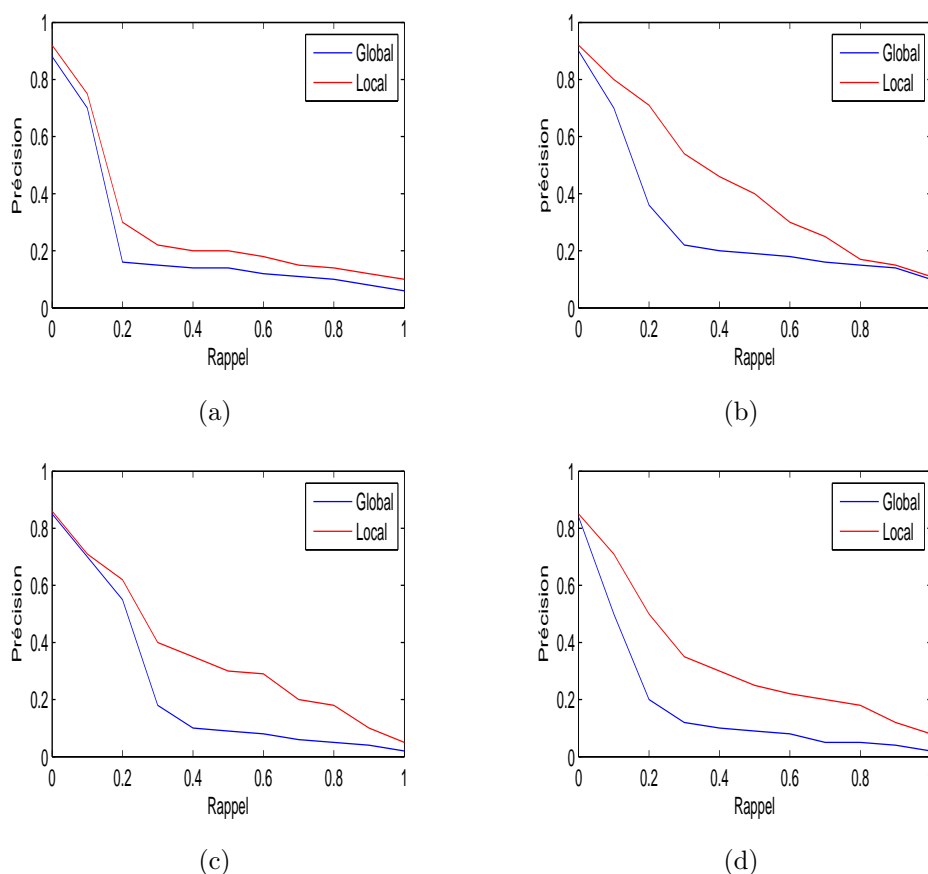


FIG. 4.11 – Comparaison des deux méthodes d'indexation : globale et locale pour 4 types d'algues (a) *Gelidium Sesquipedale* (b) *Codium Fragile* (c) *Fucus Spiralis* (d) *Blidinga minima*

4.7.5 Recherche partielle

La recherche partielle consiste à choisir une région de l'image en vue de trouver les images qui contiennent cette partie. Ce type de recherche donne des bons résultats sachant que l'utilisateur précise la région de l'image recherchée ; par conséquent le système ne rencontre pas le problème d'appariements de régions. L'image segmentée est présentée à l'utilisateur, qui sélectionne une région de son choix via l'interface graphique. La figure 4.12 illustre un exemple de recherche partielle où l'utilisateur a choisi une région qui correspond à l'espèce *Codium Fragile*. Les images résultats sont triées selon le score de gauche à droite et de haut en bas.



FIG. 4.12 – *Exemple de requête par région.*

Recherche partielle vs recherche intégrale

Dans une recherche partielle, l'utilisateur est intéressé à une partie de l'image. Pour évaluer l'apport de la recherche partielle par rapport à la recherche intégrale, nous étudions dans cette expérience le comportement du système vis-à-vis de la méthode de recherche choisie. La figure 4.13 présente les mesures effectuées pour un ensemble de 40 images requêtes. Dans cette figure on compare trois modes de recherche : recherche partielle, recherche intégrale et recherche globale. Dans la première méthode, l'utilisateur choisit, d'une manière interactive, la région qui lui intéresse. La deuxième méthode consiste à une recherche intégrale en tenant compte de toutes les régions de l'image. La recherche globale tient compte de l'aspect visuel général de l'image et considère l'image entière comme une seule région. Ces mesures indiquent que la recherche partielle présente les meilleures performances avec une précision moyenne supérieure d'environ 10% par rapport à la recherche intégrale. Il en ressort de ce qui précède que plus qu'on est précis dans le choix de la requête, plus que la réponse du système est meilleure.

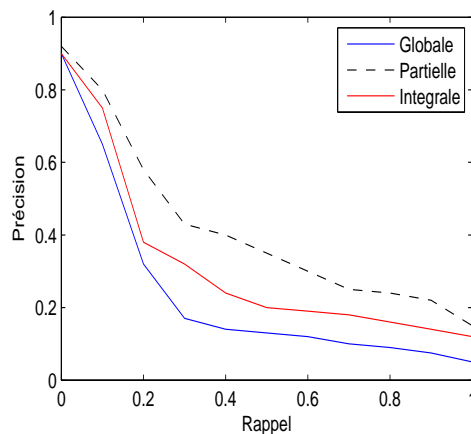


FIG. 4.13 – *Mesure de performance par mode de recherche*

4.7.6 Extension de la recherche d'images à l'environnement mobile.

Avec l'émergence rapide et la démocratisation des téléphones portables équipés d'appareils photos, beaucoup de nouvelles applications et services ont été émergés. L'image d'une scène prise par une caméra peut devenir une requête à un serveur distant pour identifier la scène et retourner les informations pertinentes. Cependant, les images des scènes prises dans différentes conditions, et avec différents points de vues est un problème ouvert dans le domaine de recherche d'images par le contenu. Ainsi, dans un autre projet de recherche, nous avons développé un système de recherche d'informations via l'image numérique dans l'environnement mobile [126, 125, 124]. Le système permet de retourner l'information à partir des requêtes d'images prises à partir d'un téléphone portable équipé d'un appareil photo. Pour rechercher la description d'un monument, un touriste peut utiliser directement l'image pour accéder efficacement à l'information voulue. Ainsi, en prenant la photo du monument visité avec un téléphone portable et en l'envoyant à un fournisseur de service par l'intermédiaire du service de messages multimédias (MMS). Peu de temps après, l'utilisateur reçoit un message texte (SMS) qui décrit le monument en question et fournit des informations supplémentaires. La figure 4.14 montre un exemple de recherche d'information via poste mobile [126].



FIG. 4.14 – Visualisation de l'image à chercher et retour de l'information concernant le site Chellah.

4.8 Conclusion

Dans ce chapitre, nous avons présenté notre système de recherche d'images par le contenu et la base d'image d'algues utilisée. Du point de vue de l'indexation, l'évaluation a, en premier lieu, mis en évidence l'apport de la segmentation automatique. Ceci permet d'augmenter les performances de la recherche.

Du point de vue de la recherche, l'interrogation du système par requête partielle s'est montrée intéressante. Ce mode de recherche, non seulement donne des résultats pertinents par rapport à la recherche approximative, mais aussi répond aux besoins de l'utilisateur.

Pour améliorer nos résultats et vaincre le problème lié au gap sémantique, nous allons présenter au chapitre suivant un processus de bouclage de pertinence. Le système propose à l'utilisateur d'agir sur la qualité des résultats proposés. Nous présentons pour cela une nouvelle méthode de boucle de rétroaction, permettant une navigation plus aisée et plus pertinente au sein de la base d'images.

Chapitre 5

RECHERCHE INTERACTIVE D'IMAGES

Les méthodes de recherche d'images les plus rencontrées font souvent appel au bouclage de pertinence ou retour de pertinence RF (Relevance Feedback), qui vise à combler le fossé sémantique ou numérique à l'aide des annotations de l'utilisateur. Ce chapitre vient clore l'ensemble de la mise en place d'une plateforme complète d'indexation et de recherche d'images d'algues, en proposant à l'utilisateur une navigation intelligente dans la base et d'agir sur la qualité des résultats proposés par le système. Nous présentons pour cela une nouvelle méthode de bouclage de pertinence, permettant une navigation plus aisée et plus pertinente au sein de la base d'images.

5.1 Introduction

Le but de la recherche est de retrouver les images similaires ou appartenant à la catégorie recherchée par l'utilisateur. Le processus est démarré par une requête, en général une image que présente l'utilisateur. Cette requête permet un premier classement des images en fonction de leur pertinence, i.e. leur appartenance à la catégorie recherchée. L'utilisateur a la possibilité de fournir des précisions quant au type d'image qu'il recherche, par exemple sous la forme d'annotations. A l'aide de ces précisions, le système peut calculer un nouveau classement et présenter de nouvelles images. L'utilisateur peut fournir de nouvelles annotations autant de fois qu'il le souhaite, et à chaque nouvelle mise à jour le système recalcule la pertinence des images.

Nous présentons sur le schéma 5.1 une architecture de bouclage de pertinence pour la recherche d'images [146, 16].

Deux étapes clés sont à considérer au sein du schéma du bouclage de pertinence. L'étape du calcul estime la pertinence de chaque image. Il en résulte

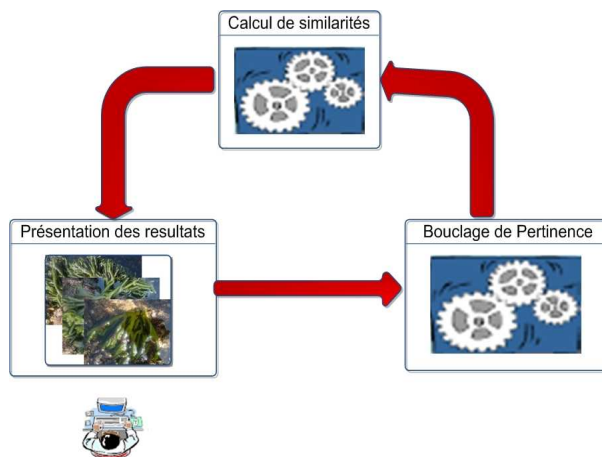


FIG. 5.1 – *Le bouclage de pertinence*

un classement qui est présenté à l'utilisateur, qui peut choisir de terminer la session s'il est satisfait, ou bien de poursuivre. Dans le cas où l'utilisateur souhaite poursuivre sa session, le système propose, lors de l'étape de sélection, des images que l'utilisateur peut annoter. Une fois les images annotées, le système utilise ces nouvelles informations pour calculer de nouvelles pertinences, et ainsi de suite jusqu'à la satisfaction de l'utilisateur.

Il existe plusieurs manières de sélectionner les exemples à annoter par l'utilisateur. L'approche la plus simple consiste à tirer aléatoirement des images dans la base. Cette technique est particulièrement inefficace puisque les concepts recherchés ont souvent très peu d'images au sein d'une même base, et la probabilité de tirer une image dans un concept donné est très faible. Il devient alors difficile de déterminer les images similaires avec peu d'annotations.

Annotations

Une technique très courante est de demander à l'utilisateur d'annoter les images les plus pertinentes. Cependant, cette solution n'est pas nécessairement la plus efficace. C'est le cadre d'étude de l'apprentissage actif, où d'autres critères que la pertinence peuvent être choisis pour sélectionner les images [22, 145].

La manière la plus simple d'annoter est de spécifier si une image ressemble ou non à l'image ou à la catégorie d'images recherchée (annotation binaire). Nous dirons qu'une image est pertinente lorsqu'elle ressemble à l'image recherchée, et non pertinente dans le cas contraire. L'annotation binaire peut sembler assez limitée. En effet, certains chercheurs proposent une annotation plus fine, par exemple une valeur entre 0 et 1 [112].

D'autres chercheurs proposent une interface graphique où les images sont disposées dans le plan selon leurs similarités [110]. Les images les plus proches de la requête sont présentées à l'utilisateur sous la forme d'une mosaïque bidimensionnelle. L'objectif est de traduire d'une manière efficace la notion de similarité existante au sein de l'espace de recherche. L'utilisateur peut, à l'aide de la souris, entourer l'ensemble des images qu'il juge pertinentes. D'autres chercheurs proposent plusieurs formes de raffinement au sein de la même interface [13]. Trois fenêtres distinctes permettent à l'utilisateur d'interagir avec le système. Une première fenêtre affiche un échantillon de la base tiré aléatoirement, une deuxième affiche les images annotées dans la première, et une troisième affiche les images annotées dans un plan 2D. Dans cette dernière fenêtre, l'utilisateur peut déplacer les images de manière à rapprocher celles qui sont dans la même catégorie. Une annotation plus fine que l'annotation binaire est discutable. En effet, il est difficile pour un utilisateur non expert du système d'évaluer précisément la pertinence d'une image. Comment dire si une image est similaire à 60%, ou 90% dans une catégorie particulière ?

Le problème se pose également lorsqu'il faut déplacer des images sur un plan 2D afin de révéler au système la similarité sémantique avec ses voisines. En outre, un argument fort en faveur des annotations binaires est leur succès dans le domaine de la recherche textuelle. Il existe aujourd'hui de nombreuses approches très efficaces avec de solides fondements théoriques. En plus des techniques de classification, l'approche active de l'apprentissage permet aussi de proposer un affichage intéressant pour l'utilisateur.

5.2 Etudes Bibliographique

La plupart des travaux actuels s'articulent autour de deux grandes approches : méthodes géométriques et méthodes basées sur la classification [149, 13, 29]. L'approche géométrique vise à modifier la requête et la distance utilisée dans le calcul de similarité en minimisant la distance à l'ensemble des exemples fournis par l'utilisateur [112]. Nous présentons tout d'abord le formalisme de ce type d'approche du bouclage de pertinence, pour présenter ensuite le système que nous avons développé dans ce cadre de recherche. Ensuite, les méthodes statistiques pour l'apprentissage en ligne sont présentées. Nos travaux basés sur une classification binaire et une optimisation de l'apprentissage en ligne sont introduits. Il ne s'agit pas ici de dresser un récapitulatif exhaustif des toutes les méthodes de bouclage de pertinence, mais d'évoquer uniquement les grands courants algorithmiques. Le bouclage de pertinence doit donc permettre à l'application de se rapprocher de l'image « idéale » si tant est qu'elle existe, censée

représenter ce que souhaite l'utilisateur. Nous noterons Q_0 l'image requête initiale et \vec{Q}_0 sa signature dans l'espace des descripteurs. De la même manière, on notera I_P^i et I_{NP}^j les images respectivement pertinentes et non pertinentes spécifiées par l'utilisateur, et d'indice i et j .

5.2.1 Méthodes géométriques

L'algorithme de Rocchio

Il s'agit ici du déplacement du point modélisant l'image requête dans l'espace des descripteurs vers une seconde image requête « idéale » (voir figure 5.2), cette dernière n'existe pas obligatoirement dans la base.

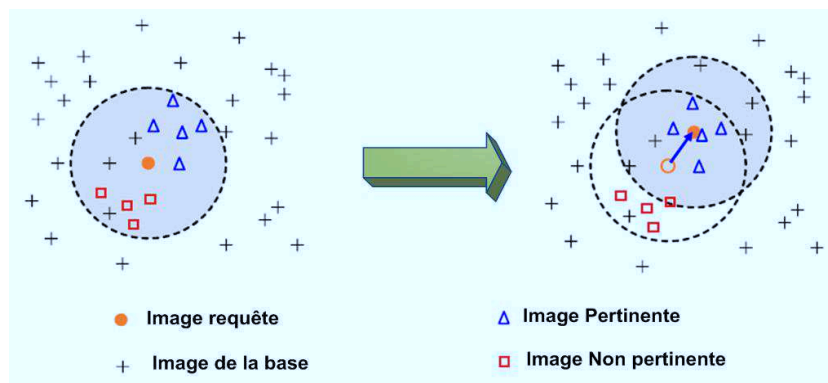


FIG. 5.2 – Exemple illustrant la Reformulation de la requête par l'algorithme de Rocchio

La formule standard donnant la position de cette image requête idéale Q_1 est donnée par la formule de Rocchio [108] suivante :

$$\vec{Q}_1 = \alpha \vec{Q}_0 + \frac{\beta}{n_p} \sum_{i=1}^{n_p} \vec{I}_p^i + \frac{\gamma}{n_{np}} \sum_{j=1}^{n_{np}} \vec{I}_{np}^j \quad (5.1)$$

Où α , β et γ sont les poids relatifs associés respectivement à la requête initiale, à la pertinence et à la non pertinence des images. De plus, n_p représente le cardinal de l'ensemble des images pertinentes désignées par l'utilisateur et n_{np} le cardinal de l'ensemble des images non pertinentes.

L'algorithme de repondération

Cet algorithme porte aussi le nom de méthode de l'écart-type. Il s'agit de prendre en compte la forme de la distribution statistique des images retournées par l'utilisateur. Si, par exemple, l'écart type σ^p de la distribution des réponses

qu'il juge pertinentes est important pour le descripteur, cela signifie certainement que ce dernier ne joue pas un rôle discriminant important. Il va donc falloir affecter à ce dernier une faible pondération. Ainsi, la pondération de ce descripteur est inversement proportionnelle à son écart type σ^p [113]. La nouvelle pondération sera alors de la forme $\omega = \frac{1}{\sigma^p}$. Cette fonction est empirique. Nous trouvons d'ailleurs dans la littérature d'autres fonctions telles que [61] : $\frac{1}{\sqrt{\sigma^p}}$ ou $\frac{1}{\log(\sigma^p)}$

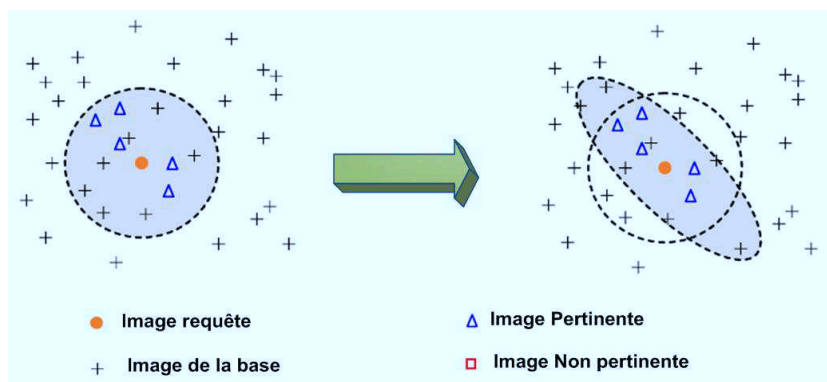


FIG. 5.3 – Exemple illustrant l'algorithme de repondération

D'une façon générale, toutes ces approches consistent à effectuer des déformations géométriques de l'espace des descripteurs afin de se rapprocher le plus pertinemment possible de l'espace perceptuel subjectif de l'utilisateur [118, 71, 113]. Ces déformations se caractérisent par une modification de la métrique associée. À ce niveau de l'analyse, il convient de remarquer que toutes ces méthodes géométriques restent unimodales, ce qui est une considérable limitation du modèle perceptuel [41].

5.3 Classification supervisée

La recherche actuelle sur le bouclage de pertinence est centrée sur la classification supervisée de la base de données. Un problème de classification supervisée est défini comme suit :

Définition. Étant donné un ensemble fini des données d'apprentissage $\{(x_i, y_i)\}_{i=1}^l \subseteq \mathcal{X} \times \mathcal{Y}$, où \mathcal{X} est l'espace d'entrée et $\mathcal{Y} = \{y_1, y_2, \dots, y_c\}$ est l'ensemble des classes possibles, trouver une fonction $f : \mathcal{X} \rightarrow \mathcal{Y}$ avec une faible erreur d'approximation aussi bien sur les données d'apprentissage que sur les données de test.

Les méthodes de classification sont divisées en deux catégories : les méthodes génératives et les méthodes discriminatives. L'idée de l'approche générative est de modéliser le système complexe qui a généré les données observées et, à partir de cette modélisation, construire une règle de décision. L'approche discriminative construit, quant à elle, directement la règle de décision à partir des données observées et ne s'intéresse pas aux caractéristiques du système qui a donné naissance aux données. L'avantage de l'approche générative est de fournir une règle de décision facilement compréhensible par un opérateur humain. Toutefois, ces dernières années, les méthodes génératives ont été dépassées par les méthodes discriminatives en terme de performance de classification. Cela est principalement dû aux caractéristiques des données modernes : grande dimension et nombre d'exemples limités. En effet, les méthodes génératives les plus utilisées, basées sur le modèle de mélange gaussien, sont particulièrement sensibles à ces deux facteurs alors que les méthodes discriminatives ne sont généralement pas affectées dans les espaces de grande dimension. Actuellement, les méthodes discriminatives sont donc particulièrement utilisées pour résoudre des problèmes où les données sont de grande dimension car elles fournissent des résultats quantitativement très bons, toutefois, on peut déplorer que ce soit au dépend de la qualité d'interprétation des décisions prises.

5.3.1 Méthodes à noyaux

Probablement, les classificateurs discriminatifs les plus populaires sont les méthodes à noyaux. Les fonctions noyaux définissent une projection (normalement non-linéaire) des données d'un espace vectoriel dit d'attributs vers un autre espace appelé de caractéristiques. En pratique, elles calculent les produits scalaires des arguments dans ce nouvel espace. Les arguments sont passés à la fonction en format de l'espace original et elle calcule implicitement leur produit scalaire dans l'espace de caractéristiques. Dans le cadre de l'apprentissage automatique et de la reconnaissance de formes, une observation potentielle appartient à l'espace vectoriel $\mathcal{X} \in R^p$ et l'analyse repose sur un ensemble fini d'échantillons $X = \{x_1, x_2, \dots, x_n\} \subseteq \mathcal{X}$. Pour les tâches supervisées, à chaque échantillon x_i est associée une étiquette $l_i \in \mathcal{R}$ pour concrétiser l'appartenance de cette échantillon à une classe.

En analyse de données, il est habituel de procéder à un changement d'espace afin d'accéder à une représentation des données plus informative. Ce changement d'espace implique l'utilisation d'une fonction ϕ et d'un espace \mathcal{F} :

$$\phi : \mathcal{X} \rightarrow \mathcal{F}$$

Les données de l'espace de départ sont appelées "attributs" et les données transformées "caractéristiques". Les méthodes à noyau sont les méthodes d'apprentissage qui nécessitent uniquement des produits scalaires entre échantillons afin d'effectuer la prédiction et l'apprentissage. Ces produits scalaires ne doivent pas nécessairement être définis sur l'espace d'entrée \mathcal{X} , mais sur n'importe quel espace de Hilbert \mathcal{H} pour lequel il existe une application $h : \mathcal{X} \rightarrow \mathcal{H}$. Cela donne un produit scalaire dans \mathcal{H} , $\langle h(x), h(x') \rangle$. Bien que l'utilisation de toute application dans un espace de Hilbert donne aux méthodes à noyau une bonne flexibilité pour le choix de la fonction de classification f , le choix de h est limité dans la pratique par le calcul effectué afin d'évaluer le produit scalaire. Considérons par exemple le cas où l'espace de caractéristiques est défini à partir des monômes d'ordre d d'entrées x_j de x , c.-à-d. x_{j_1}, \dots, x_{j_d} , où $j_1, \dots, j_d \in \{1, \dots, m\}$. Malheureusement, la dimension de l'espace des caractéristiques est : $(m+d-1)!/d!(m-1)!$. Ainsi, un mapping explicite dans un tel espace de caractéristique est impossible dans la pratique.

Ce problème peut être résolu en utilisant "l'astuce de noyau". Comme la méthode noyau exige que l'évaluation des produits scalaires $\langle h(x), h(x') \rangle$, et non les données transformées $h(x)$ explicitement, on peut espérer de trouver une fonction $K : X \times X \rightarrow \mathbf{R}$ avec $K(x, x') = \langle h(x), h(x') \rangle$, qui requiert des calculs plus simples. K est appelé une fonction du noyau ou "kernel". Pour l'exemple ci-dessus la fonction noyau est le noyau polynomial $K(x, x') = (x \cdot x')^d$.

En général toute fonction $K : X \times X \rightarrow \mathbf{R}$ peut servir de noyau s'elle répond à deux exigences : elle doit être symétrique $K(x, y) = K(y, x)$ et semi-définie positive : $\int_{\mathcal{X} \times \mathcal{X}} K(x, y) f(x) f(y) dx dy \geq 0$ pour tout $f \in L_2(\mathcal{X})$.

Comme dans la pratique, seules les données des ensembles finis X sont considérées, toutes les informations disponibles au classificateur peuvent être représentés dans la matrice de Gram $K = K(x_i, x_j)_{ij}$, $1 \leq i, j \leq n$ correspondant à la fonction noyau donnée. Cette matrice est connue comme la matrice noyau. Les noyaux valides peuvent donc être encore définis comme suit :

Proposition. Une fonction $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbf{R}$ est un noyau valide ssi pour tout ensemble fini $X \subseteq \mathcal{X}$ elle produit des matrices de Gram symétriques et semi-définies positives.

Une autre famille de fonctions noyaux les plus utilisées est celle des noyaux "radiaux", dérivant d'un calcul de distance entre vecteurs, représentées par le noyau gaussien (RBF).

$$K(x, y) = e^{-\|x-y\|^2/2\sigma^2} \quad (5.2)$$

Le noyau gaussien correspond à des fonctions $K(x, \cdot)$ centrées en x avec un maximum local en ce centre. La zone d'influence de chaque fonction est paramé-

trée par le facteur σ .

5.3.2 Algorithmes à base de SVM

Les Supports à Vaste Marge (SVM)[148] sont une technique de classification qui utilise une fonction noyau pour trouver l'hyperplan optimale. On cherche à séparer la base en deux groupes : les images pertinentes et les images non pertinentes. L'utilisation d'un réseau de neurones de type perceptron permet d'effectuer cette classification en évaluant la position des points par rapport à l'hyperplan séparateur. L'inconvénient de ce type de méthode est qu'il nous renvoie un résultat binaire : le point se situe dans la classe pertinente ou non pertinente. L'utilisation des Machines à Vecteurs de Support (SVM) permet de pallier à cet inconvénient en proposant de fournir la distance à l'hyperplan en supplément d'information. Cette méthode cherche à construire un hyperplan optimal, c'est à dire maximisant la distance entre celui-ci et les points d'apprentissage. L'équation de cet hyperplan est donné par $\vec{w} \cdot \vec{x} + b = 0$, où \vec{x} est l'ensemble des images de la base d'apprentissage (voir figure 5.4). On considère donc l'ensemble d'apprentissage (\vec{x}_i, y_i) avec $i \in [1.. N_l]$, et $y_i \in [-1.. 1]$ et selon \vec{x}_i que soit désigné comme étant pertinent ou non.

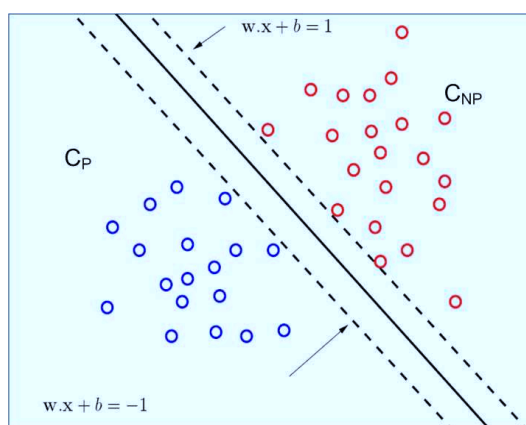


FIG. 5.4 – Hyperplan optimal séparant les deux classes Pertinente (C_P) et Non Pertinente (C_{NP}).

$$f_y(x_i) = \langle w, x_i \rangle + b \quad (5.3)$$

L'objectif est de trouver une fonction de discrimination \mathbf{f} (5.3) telle que tous les éléments de même classe soient du même côté de l'hyperplan. Autrement dit, il faut trouver un vecteur w et un réel b tel que :

$$\forall i \in I (\langle w, x_i \rangle + b) y_i > 0 \quad (5.4)$$

Dans un cas séparable, il existe de nombreux hyperplans vérifiant cette équation. Les SVM se distinguent des autres techniques par un choix particulier d'hyperplan séparateur. L'approche choisit parmi tous les hyperplans séparateurs celui qui maximise la marge, i.e. tel que la plus petite distance entre les points et l'hyperplan soit maximale. Par exemple, la zone entre les deux lignes pointillées sur la figure 5.4. Un tel hyperplan, en plus d'être unique, a de nombreux avantages sur le plan théorique [148]. La fonction de décision étant invariante par changement d'échelle, on choisit de trouver l'hyperplan tel que $w \cdot x + b = \pm 1$ pour les éléments les plus proches de la marge (cf. figure 5.4), ce qui revient à minimiser $\|w\|^2$ tel que :

$$\forall i \in I y_i (\langle w, x_i \rangle + b) \geq 1 \quad (5.5)$$

En utilisant les Lagrangiens, le problème devient :

$$\arg \max_{\alpha} W(\alpha) = \sum_{i \in I} \alpha_i - \frac{1}{2} \sum_{i,j \in I^2} \alpha_i \alpha_j y_i y_j \langle X_i, X_j \rangle \quad (5.6)$$

Avec :

$$\sum_{i \in I} \alpha_i y_i = 0 \text{ et } \alpha_i \geq 0, \forall i \in I$$

La fonction de pertinence s'écrit alors :

$$f_y(x) = \sum_{i \in I} y_i \alpha_i \langle X, X_i \rangle + b \quad (5.7)$$

Cette première méthode suppose que les données sont linéairement séparables. Afin d'assouplir la discrimination, une marge souple peut être introduite, en acceptant la mauvaise classification de certains éléments. Ceci revient à majorer chacun des α_i par une constante C [150]. De plus, on peut facilement utiliser une fonction noyau avec cette méthode, qui est à l'origine de l'essor des ces fonctions.

Puisque la méthode s'écrit avec un produit scalaire $\langle x; x' \rangle$, on peut utiliser le truc du noyau :

$$\arg \max_{\alpha} W(\alpha) = \sum_{i \in I} \alpha_i - \frac{1}{2} \sum_{i,j \in I^2} \alpha_i \alpha_j y_i y_j K(X_i, X_j) \quad (5.8)$$

Avec :

$$\sum_{i \in I} \alpha_i y_i = 0 \text{ et } 0 \leq \alpha_i \leq C \forall i \in I$$

La fonction de pertinence est alors :

$$f_y(x) = \sum_{i \in I} \alpha_i y_i K(x, x_i) \quad (5.9)$$

5.4 Nouvelle approche au RF

5.4.1 Introduction

L'étude précédente des différentes méthodes de bouclage de pertinence proposées dans la littérature nous amène à nous poser quelques questions essentielles pour le développement d'un algorithme efficace et pertinent. A cet effet, avant de commencer le développement d'un algorithme de bouclage de pertinence, il convient de bien définir le but que nous cherchons à atteindre. Nous avons fixé trois objectifs : prendre en compte aussi bien les valeurs positives que les valeurs négatives, souplesse de la recherche, et réactivité.

Comme nous l'avons vu précédemment, au cours du bouclage de pertinence, l'utilisateur se voit présenter un certain nombre d'images auxquelles il doit affecter une valeur correspondante à son jugement quant à la pertinence des réponses qui lui sont présentées. Cette vision nous oblige à nous attarder sur la validité d'une valeur négative. En effet, la plupart des moteurs ne prennent pas en considération les valeurs négatives en retour du bouclage de pertinence, ils considèrent juste les valeurs positives et cherchent à s'approcher de ces derniers. Il serait non judicieux d'omettre les images négatives. Cette proposition semble tout à fait acceptable pour une simple raison : les images négatives véhiculent eux aussi une information et doivent être incluses dans le calcul. Cela nous ramène à un problème de classification cherchant à séparer la base en deux catégories : les images pertinentes et les images non pertinentes.

Les auteurs dans [29], ont défini trois grandes catégories de recherche :

- (1) la recherche *parfaitement ciblée*. Dans ce cas, l'utilisateur recherche une image spécifique et précise, et ne se satisfera pas d'une approximation. Il s'agit d'une position prise par la plupart des moteurs de recherches ;
- (2) la recherche par *catégorie*. Il s'agit ici d'effectuer une classification de la base d'images entre les images pertinentes (celle de la catégorie recherchée par l'utilisateur) et les non-pertinentes ;

- (3) la recherche *ouverte*. Dans ce cas, on considère que l'utilisateur n'a qu'une vague idée de ce qu'il recherche et qu'il veut pouvoir naviguer d'une façon cohérente, au sein de la base [69].

Nous devons donc nous positionner au sein de ces catégories. Cela implique que notre algorithme doit bénéficier d'une souplesse le rendant réactif à tout type de recherche effectué par l'utilisateur.

Le bouclage de pertinence fait directement appel à une action de l'utilisateur qui en attend un retour instantané. Cela situe le processus à un point critique et l'oblige à respecter un faible temps de réactivité afin de fonctionner en temps réel. Il s'agit d'un point crucial que nous devons prendre en compte afin que notre application soit réaliste. Compte tenu de l'importance des dimensions de l'espace des descripteurs dans lesquelles nous travaillons, ainsi que le nombre important d'images que peut contenir une base, ce point est loin d'être trivial et peut rapidement mener à des impossibilités pratiques. Pour cela nous avons amélioré le temps de réponse et réduit le délai de convergence.

5.4.2 Apprentissage semi supervisé à multiple instances

Nous avons parlé de l'apprentissage supervisé où, le système apprend à classer des données, selon un modèle de classement, à partir d'une base d'apprentissage des exemples x_i et leur étiquettes y_i $D_L = \{(x_i, y_i) | i = 1, \dots, l\}$. Toutefois, dans de nombreuses applications du monde réel, l'apprentissage de la fonction de classification ($f : \mathcal{X} \rightarrow \mathcal{Y}$) en ne considérant que les données étiquetées pourrait s'avérer impraticable. Ceci est principalement dû au fait que, souvent, les données étiquetées sont difficiles à obtenir.

Considérons par exemple la tâche d'étiqueter les images d'une base de données dans le cadre de la recherche d'images. Il s'agit d'un travail manuel qui est généralement très coûteux. Ainsi, il semble naturel d'utiliser les informations supplémentaires fournis par les données non étiquetées afin d'améliorer le processus d'apprentissage. Ce problème est connu sous le nom « apprentissage semi-supervisé ». Nous pouvons donc formuler la tâche de classification semi-supervisée comme suit :

Définition. Considérons un ensemble fini d'exemples étiquetés $D_L = \{(x_i, y_i)\}_{i=1}^l \subseteq \mathcal{X} \times \mathcal{Y}$ et un ensemble fini des exemples non étiquetés $D_U = \{x_{l+1}, \dots, x_{(l+u)=n}\}$, l'apprentissage semi-supervisé consiste à trouver une fonction $f : \mathcal{X} \rightarrow \mathcal{Y}$ à faible erreur d'approximation sur toutes les données.

Alors, comment les données non étiquetées peuvent servir dans le proces-

sus d'apprentissage? L'hypothèse générale est que les exemples non étiquetés peuvent nous aider à dévoiler une sorte de structure ou de régularité dans les données qui ne peuvent pas certainement être découvert lors de l'utilisation seulement des données étiquetés.

L'hypothèse commune dans la plupart des algorithmes d'apprentissage semi-supervisé est appelé « hypothèse de clusters » [162, 17] et qui peut être énoncé comme suit :

- les points voisins sont susceptibles d'avoir la même étiquette;
- les points sur la même structure (cluster ou variété) sont susceptibles d'avoir la même étiquette.

Si cette hypothèse est satisfaite il serait possible d'effectuer une classification avec seulement un exemple étiqueté de chaque classe.

Il y a eu beaucoup de travaux sur l'application de l'apprentissage semi supervisé (Semi Supervised Learning : SSL) pour résoudre des problèmes pratiques par la propagation d'étiquette. Zhu a présenté une étude détaillée sur l'apprentissage semi-supervisé dans [164]. Toutefois, la plupart entre elles s'intéressent à l'image toute entière (Single Instance Learning), alors que la recherche d'images est souvent présentée comme un apprentissage d'instances multiples.

Des travaux sur l'apprentissage d'instances multiples (Multi instance Learning : MIL) ont été développés pour pallier au problème de recherche d'images par régions [153]. L'apprentissage MIL a été introduit par Dietterich et al. [33] dans les activités de classification des molécules, puis il a été largement utilisé dans les systèmes de recherche d'images par le contenu et l'annotation automatique d'images [153, 159, 158]. Le Multiple Instance Learning est une technique visant à apprendre à partir de données dont les classes ne sont pas connues pour chaque élément individuellement, mais plutôt pour des groupes d'éléments (bags), pour lesquels il est connu si ce groupe possède ou non des éléments (instances) d'une classe donnée. L'objectif d'un algorithme MIL est de générer un classificateur qui pourra classer correctement les bags inconnues. Un bag est positive si au moins un élément appartenant à ce bag est positif, et négatif si tous les éléments sont négatifs. Dans le cas de la reconnaissance d'objets, les éléments sont les régions d'une image, et un groupe d'éléments est l'image elle-même. Cet apprentissage est utilisable dans le cas où on a l'information de la classe de l'image sans avoir l'information de la classe de chacune des régions séparément, ce qui est bien souvent le cas.

Récemment des algorithmes combinant les deux approches ont vu le jour. Ainsi , Rahmani et Goldman [104] ont présenté une méthode d'apprentissage semi supervisée à multiple instances (Multi Instances Semi Supervised Learning MISSL) basée sur les graphes pour traiter le problème de recherche d'images

par régions, qui transforme le problème de multiples instances en un problème d'instance unique.

Zhou et al. [163] ont proposé un algorithme basé sur un apprentissage semi supervisé à multiples instances en utilisant les machines à vecteurs de support (MissSVM). Dans ce contexte, nous proposons un algorithme MISSL où nous avons considéré à la fois, les propriétés de l'apprentissage semi supervisé et l'apprentissage multi-instances.

5.4.3 Régularisation

L'apprentissage à partir d'exemples revient à créer une fonction par approximation à partir des échantillons. C'est un problème mal posé résolu avec la théorie de la régularisation, qui trouve une fonction \mathbf{f} minimisant une certaine fonctionnelle sur un espace de Hilbert \mathcal{H} : Supposons que nous avons une tâche de classification binaire, donc nous devons trouver une fonction $f_c : \mathcal{X} \rightarrow \{-1, 1\}$. Nous définissons une fonction $f : \mathcal{X} \rightarrow [-1, 1]$ tel que : $f_c(x) = \text{sgn}(f(x))$. La fonction \mathbf{f} est définie sur le graphe $G = (X, E)$. Afin d'exploiter les hypothèses énoncées ci-dessus, nous exigeons de notre fonction \mathbf{f} de varier sur le graphe, de telle façon que les noeuds voisins (tels qu'ils sont définis par les poids des arêtes) devraient avoir les mêmes valeurs. En outre, nous exigeons une faible erreur d'apprentissage $\sum_{i=1}^l V(y_i, f(x_i))$, où $V(., .)$ est une fonction de perte. Comme \mathcal{X} est fini, nous pouvons utiliser un vecteur $\mathbf{f} \in \mathbf{R}^n$ pour décrire \mathbf{f} sur G : $f = [f(x_1), \dots, f(x_n)]^T$. Ainsi, nous posons notre problème comme une fonction de régularisation définit comme suit :

$$\min_{f \in \mathcal{H}_k} \left\{ \sum_{i=1}^l V(y_i, f(x_i)) + \lambda J(f) \right\}. \quad (5.10)$$

Où λ est le paramètre de régularisation qui détermine le compromis entre l'erreur d'apprentissage et le lissage de la fonction. $\mathbf{J}(\mathbf{f})$ est un terme de pénalisation qui devrait pénaliser \mathbf{f} de ne pas violer certaines contraintes de lissage. \mathcal{H}_K est l'espace des fonctions $f : \mathcal{X} \rightarrow \mathcal{Y} = [-1, 1]$. Considérons maintenant le produit scalaire de \mathcal{H}_K , $\langle ., . \rangle_{\mathcal{H}_k}$ qui est défini comme suit :

$$\langle f, f' \rangle_{\mathcal{H}_k} = f^T L f', \quad f, f' \in \mathcal{H}_k. \quad (5.11)$$

La semi-norme induite est $\|f\|_{\mathcal{H}_k} = \sqrt{\langle f, f \rangle_{\mathcal{H}_k}}$. Il s'agit d'une semi-norme parce que $\|f\|_{\mathcal{H}_k} = 0$ si \mathbf{f} est un vecteur constant. Cela peut être vu en notant que (voir 3.5.1) :

$$\|f\|_{\mathcal{H}_k}^2 = \frac{1}{2} \sum_{i,j=1}^m (f_i - f_j)^2 W_{ij}. \quad (5.12)$$

Ce qui est plus important dans (5.12), c'est qu'elle montre pourquoi la matrice Laplacienne \mathbf{L} est utilisable comme une matrice de régularisation : il favorise un lissage sur le graphe sous-jacent représenté par \mathbf{W} . Ainsi, en utilisant (5.12), comme terme de pénalisation (5.10), conduit à un problème d'optimisation.

5.4.4 Méthode de Consistance

Les auteurs dans [162] ont proposé un algorithme qui l'ont appelé méthode de consistance. Il s'agit d'un algorithme itératif qui réalise les hypothèses de regroupement (Section 5.4.2). En effet, les points doivent être étiquetés non seulement d'une manière localement consistante, c'est à dire que les points voisins devraient avoir le même label, mais aussi d'une manière globalement consistante, de telle sorte que les points dans le même groupe doivent avoir mêmes étiquettes.

Soit W une matrice d'affinité de noyau gaussien RBF tel qu'elle est définie par l'équation (5.2), avec éléments de la diagonale $W_{ii} = 0$. Nous construisons la matrice normalisée symétriquement $S = D^{-1/2}WD^{-1/2}$ où D est la matrice diagonale avec $D_{ii} = \sum_{j=1}^n W_{ij}$. Soit F et Y des non-négative $n \times c$ matrices. F correspond à une classification sur \mathcal{X} , par étiquetage de chaque point x_i comme label $y_i = \operatorname{argmax}_{j \leq c} F_{ij}$. $Y_{ij} = 1$ si x_i est étiqueté comme $y_i = j$ et -1 sinon. La procédure d'itération est définie comme suit :

$$F(t+1) = \alpha SF + (1 - \alpha)Y \quad (5.13)$$

α est un paramètre dans $(0, 1)$ et détermine le taux d'information qui se propage entre les données voisins de chaque point et son étiquette initiale au cours de chaque étape d'itération. Le déroulement des itérations conduit à :

$$F(t) = (\alpha S)^{t-1}Y + (1 - \alpha) \sum_{i=0}^{t-1} (\alpha S)^i Y \quad (5.14)$$

Comme $\lim_{t \rightarrow \infty} (\alpha S)^{t-1} = 0$ et $\lim_{t \rightarrow \infty} \sum_{i=0}^{t-1} (\alpha S)^i = (I - \alpha S)^{-1}$, F converge vers

$$F^* = (1 - \alpha)(I - \alpha S)^{-1}Y. \quad (5.15)$$

Cela montre que l'itération décrit un processus de diffusion [162].

Des variantes de cet algorithme, ont utilisés \mathbf{P} et \mathbf{P}^T , respectivement, au lieu de S . $P = D^{-1}W$ où P est la matrice de transition. Par ailleurs [162] a montré que F^* est aussi la solution pour minimiser la fonction de coût :

$$Q(F) = \frac{1}{2} \left(\sum_{i,j=1}^n W_{ij} \left\| \frac{1}{\sqrt{D_{ii}}} F_i - \frac{1}{\sqrt{D_{jj}}} F_j \right\|^2 + \mu \sum_{i=1}^n \|F_i - Y_i\|^2 \right) \quad (5.16)$$

Cette formule décrit une fonction de régularisation (voir section 5.4.3) avec le paramètre de régularisation μ où $\alpha = \frac{\mu}{1+\mu}$. Le premier terme correspond à une contrainte de lissage, pénalisant ainsi la fonction de classification F de ne pas trop changer entre les points voisins. Le second terme est la contrainte d'ajustement ou fonction de perte, ce qui signifie que F ne doit pas trop changer par rapport à l'étiquette initiale. Dans ce cas, la fonction de perte quadratique a été choisie.

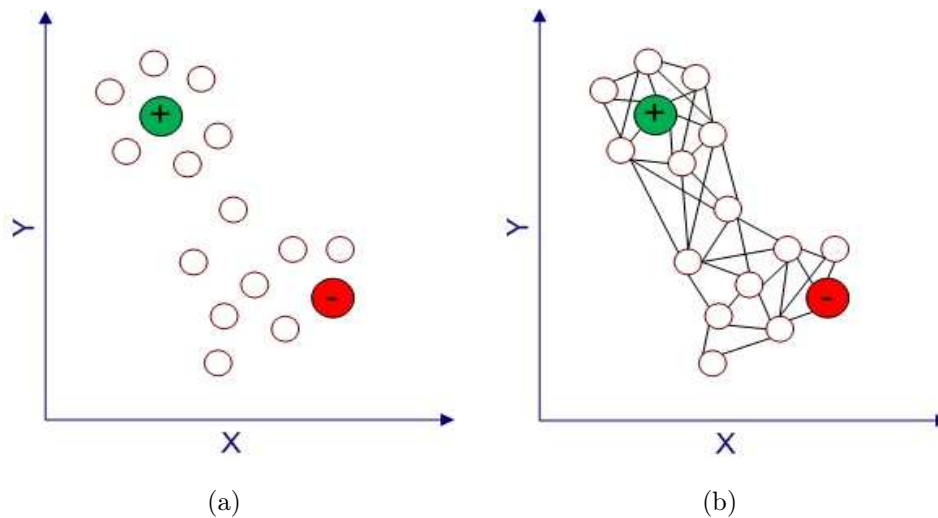


FIG. 5.5 – *Problème de Bouclage de Pertinence.* (a) montre l'espace de caractéristiques avec deux images marquées : une image est positive (en vert) et l'autre est négative (en rouge). Les autres images sont non marquées. (b) est une représentation graphique de (a).

5.4.5 Description de la méthode proposée : "MISSL_based RF"

Le but de ce travail est de prédire les étiquettes des images non reconnues. Comme évoqué ci dessus, l'apprentissage semi supervisé est très utile pour ce genre de problèmes. Nous allons présenter un algorithme d'apprentissage basé sur les graphes.

Pour commencer, le bouclage de pertinence peut se ramener facilement à un problème de classification binaire avec sortie $y \in \{-1, 1\}$. Pour se faire, nous considérons un graphe connecté $G(V, E)$ défini sur la base X (comme montré dans la figure 5.5) avec les noeuds V correspondant aux images de la base et les arêtes E (connexions entre les noeuds) pondérés par $n \times n$ matrice d'affinité symétrique W . Soit X la base des descripteurs d'images $X = \{x_1, \dots, x_l, x_{l+1}, \dots, x_n\} \subset \mathfrak{R}^m$ et L l'ensemble des labels $L = \{+1, -1\}$; les

premiers ℓ points $x_i (i \leq \ell)$ ont des labels $y_i \in L$. Les autres points restants $x_u = (l + 1 \leq u \leq n)$ sont non étiquetés. y_i est égale à $+1$ si l'image x_i appartient à la classe d'intérêt de l'utilisateur et $y_i = -1$, Autrement. Soit N_p le total d'images désignées comme pertinentes par l'utilisateur et N_{np} le nombre total de retours négatifs désignés par ce même utilisateur.

L'idée principale est de trouver une fonction à valeur réelle $f : V \rightarrow \mathbf{R}$ qui ne varie pas beaucoup entre les points voisins non marqués (lissage) et la valeur de la fonction f sur le point marqué f_i est égale à y_i son étiquette d'origine. C'est à dire, $f_i = y_i$, pour $i = 1, \dots, l$. Nous avons utilisé la fonction de classification F (5.19) décrite dans la section (5.4.3).

La méthode de consistance est basée sur les hypothèses de regroupement (Section 5.4.2) :

- les points voisins sont susceptibles d'avoir la même étiquette ;
- les points sur la même structure (cluster ou variété) sont susceptibles d'avoir la même étiquette.

Toutefois, dans les applications réelles, il est difficile de satisfaire ces hypothèses . En effet, Les images peuvent être proches dans la base de caractéristiques, mais ne le sont pas réellement. Pour rester fidèles aux hypothèses de regroupement, nous avons non seulement considéré une représentation globale de l'image, mais aussi une représentation locale.

Dans cette section, nous allons présenter une extension de la méthode de consistance [162], adaptée au bouclage de pertinence. Tout d'abord, nous allons parler de la représentation d'images adoptée. Après, nous proposons un apprentissage semi-supervisé à multiple instances par combinaison de l'information locale et globale.

Représentation locale

La représentation d'images par région est un moyen efficace pour améliorer les performances de la recherche d'images par le contenu. Les descripteurs globaux ne représentent pas toujours les objets saillants dans une image. Ces derniers sont moins gourmands au niveau de la capacité mémoire et du temps de calcul, mais ils donnent une représentation grossière du contenu de l'image. Ainsi, la performance de recherche d'images est fortement liée à la prise en compte des informations locales. Notre méthode de segmentation utilise une approche basée sur les marches aléatoires. Cette approche est décrite dans le chapitre 3.

En MIL, l'apprentissage est composé de plusieurs bags ; chacun comprend de nombreuses instances. Un bag est positivement marqué s'il contient au moins un élément positif, sinon il est étiqueté comme négatif. Le but est d'apprendre

quelques concepts de l'apprentissage pour l'étiquetage correct des bags non marqués. Dans la recherche d'images par objet, chaque région est considérée comme instance, et l'ensemble des régions qui proviennent de la même image peuvent être traitées comme un bag. Nous annotons une image d'algue comme étant positive si au moins une région de l'image contient l'espèce demandé.

Compte tenu d'une image contenant plusieurs régions, nous pouvons nous attendre à ce qu'au moins une région corresponde à l'intérêt de l'utilisateur. Par conséquent, le problème de la recherche d'images par région est essentiellement identique aux concepts de l'apprentissage à instances multiples. Chaque image de la base d'images est segmentée. Les attributs de la couleur et de la texture sont ensuite calculés pour chaque région. L'image est alors considérée comme un bag contenant de nombreuses instances (vecteurs caractéristiques).

Sur la base des résultats de la segmentation, les caractéristiques de la couleur et de la texture sont calculées pour chaque région. Ces descripteurs locaux contribuent dans la construction de la matrice de similarité globale comme illustré dans les sections suivantes.

Représentation globale

Une segmentation imparfaite peut affecter la représentation des bags dans l'apprentissage de multiples instances et par conséquent diminuer le rendement du système de recherche. En effet, nous avons ajouté des fonctionnalités globales pour faire face à ce problème. Afin de compenser les limitations inhérentes aux espaces de couleur et la représentation de la texture, nous construisons les descripteurs globaux d'une façon différente à celle utilisée dans la création des descripteurs locaux. Après avoir obtenu toutes les caractéristiques globales, elles sont intégrées dans un graphe de représentation globale.

Mise en place du procédé

Le système interactif consiste à demander à l'utilisateur des questions telles que ses réponses permettent de réduire le fossé sémantique selon les étapes suivantes :

- (1) Le système compare l'image requête avec les images de la base en utilisant les vecteurs caractéristiques. La mesure de similarité de la première extraction est réalisée par la distance Euclidienne. Des images les plus similaires, point de vue signatures, sont retournées et affichées. Ainsi, la première phase de recherche est achevée.
- (2) L'utilisateur annoté les images affichées comme positives ou négatives en fonction de son besoin.

- (3) Le système prévoit les labels des images non étiquetées selon l'algorithme suivant :

Les vecteurs caractéristiques sont arrangés dans un graphe indirecte et pondérés. Le graphe est caractérisé par la matrice de poids W , dont les éléments $W_{ij} \geq 0$ sont les mesures de similarité entre les vertices i et j , et par le vecteur initiale des labels. La mesure de similarité utilisée est définie par une fonction gaussienne :

$$W_{ij} = \exp(-d(x_i, x_j)^2 / (2\sigma)^2). \quad (5.17)$$

Avec σ est la déviation standard (paramètre de largeur gaussien), $d(x_i, x_j)$ est la distance entre les vecteurs caractéristiques x_i et y_i .

En tenant compte des deux aspects : global et local, La matrice de similarité finale du graphe est définie comme suit :

$$W_{ij} = W_{ij}^L * W_{ij}^g \quad (5.18)$$

W^L et W^g matrice de similarité locale et globale respectivement.

L'algorithme proposé est récapitulé comme suit :

Algorithme 5.4.1 Apprentissage Semi Supervisé à Multiple instances.

- (1) Chaque image de la base est segmentée en régions homogènes. pour chaque région d'image, on procède à extraire les vecteurs caractéristiques.
- (2) On construit la matrice de similarité en utilisant l'algorithme de k plus proche voisin (K-Nearest Neighbors).

$$W_{ij} = W_{ij}^L * W_{ij}^g$$

W^L et W^g matrice de similarité locale et globale respectivement.

- (3) On construit la matrice Laplacienne $S = D^{-1/2} * W * D^{1/2}$ dont D est une matrice diagonale et $D_{ii} = \sum_{j=1}^n w_{ij}$
- (4) Calculer la fonction de classification F :

$$F = A^{-1}Y. \quad (5.19)$$

Avec : $A = (I + \mu(I - S))$ et $\mu = \frac{\alpha}{1-\alpha}$

- (5) Décider la pertinence de l'image
-

La figure 5.6 présente un schéma récapitulatif de l'ensemble des opérations rentrant en jeu dans le processus de bouclage de pertinence.

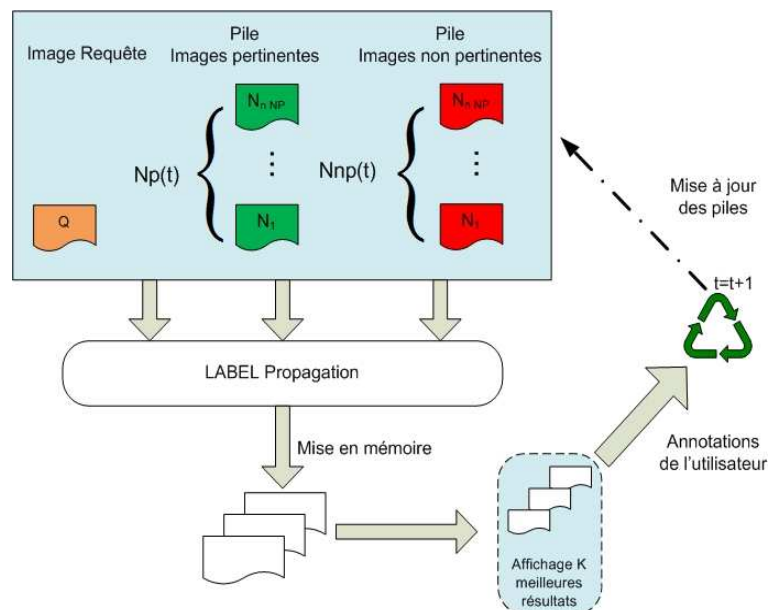


FIG. 5.6 – Architecture du bouclage de pertinence

5.5 Bouclage de pertinence Automatique

5.5.1 Introduction

Les méthodes de bouclage de pertinence contribuent à l'amélioration des résultats de recherche et diminuent le fossé sémantique. Toutefois, l'utilisateur est obligé d'annoter à chaque session les images pertinentes ou non pertinentes, et afin de satisfaire son besoin, l'utilisateur doit faire un laborieux travail et perdre plus de temps.

Pour remédier à ce problème, une approche de bouclage de pertinence automatique est proposée. L'idée de base de cette méthode vient de l'hypothèse que le jugement de pertinence d'un système de recherche d'images est fortement lié à un problème de classification. En effet, nous avons proposé une approche où l'identification de pertinence est assurée en utilisant un réseau de neurones de kohonen modifié qu'on l'a appelé MISOM (Multi Instance Self Organizing Maps). Ce réseau permet un traitement à multiple instances où les informations globale et locale sont intégrées. En effet, à chaque itération, les descripteurs globaux et locaux sont combinées pour former le réseau de neurones. Cela permet au système de faire la distinction entre les images pertinentes et non pertinentes automatiquement. Ainsi, le système, par l'intermédiaire du processus d'apprentissage, augmente son rendement de recherche, sans intervention humaine. Si l'utilisateur n'est pas satisfait des résultats retournés, il pourra continuer à rechercher en an-

notant les images manuellement. Cela permettra de réduire l'effort humain, de minimiser les itérations et de gagner en matière du temps de recherche.

Dans des travaux précédents, les auteurs dans [93, 62], ont proposé une méthode de bouclage de pertinence basée sur un apprentissage non supervisé. Cet apprentissage implique un nouveau réseau de neurones : SOTM (Self Organizing Tree Maps). Toutefois, ces approches ne prennent pas en compte les spécificités locales de l'image. Nous avons proposé dans [136] une approche de recherche d'images basée sur les régions d'intérêt en utilisant les cartes auto-organisées de Kohonen (SOM) [66]. Ce travail vient pour étendre les travaux déjà élaborés par l'introduction d'un processus de bouclage de pertinence automatique.

5.5.2 Description de la méthode proposée : "MISOM_based ARF"

Le processus de recherche d'images avec retour de pertinence automatique consiste à extraire les descripteurs globaux et locaux pour chaque image. Le système compare ensuite l'image requête à chaque image de base de données en utilisant ces caractéristiques bas niveaux. La mesure de similarité, dans la première extraction est réalisée par la distance euclidienne. Ainsi, les premières images sont retournées et affichées. Cette opération vient pour clore la première phase de recherche. Les images retournées sont ensuite classées en utilisant MISOM. Une nouvelle requête est ensuite reformulée afin de s'approcher des images pertinentes. Ce processus va se poursuivre pour deux cycles. Une vue du système est présentée dans la figure 5.7.

Pour améliorer la performance du système de recherche, les descripteurs de recherches et de classification devraient être différents.

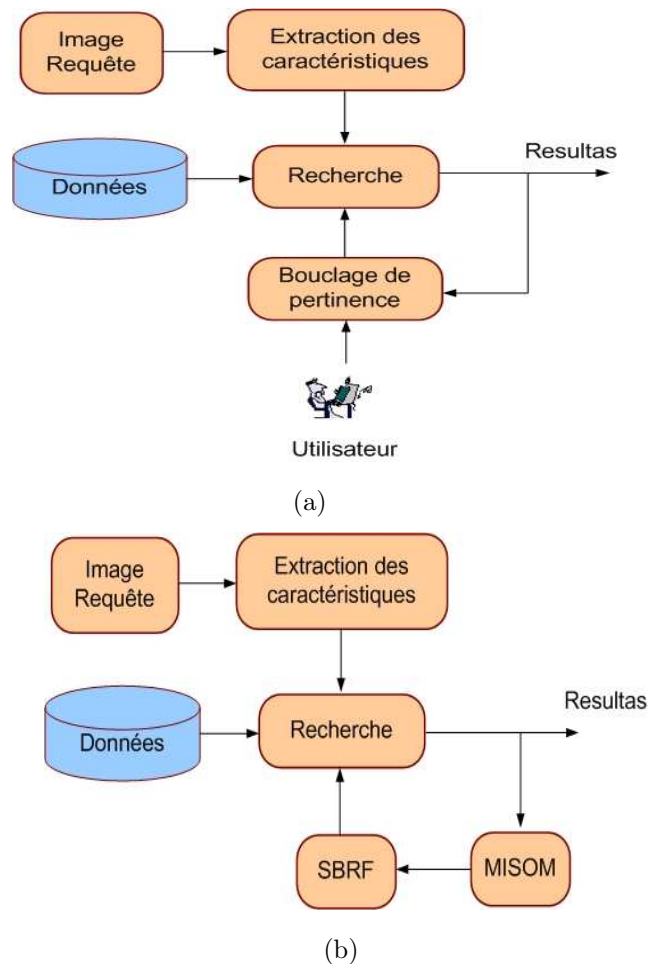


FIG. 5.7 – Architecture d'un Système de recherche d'images. (a) Bouclage de pertinence classique. (b) bouclage de pertinence automatique proposé

Regroupement d'images utilisant MISOM

Dans cette phase, une méthode basée sur le réseau de neurones de type SOM a été introduite pour parvenir au retour de pertinence d'images sans l'interaction de l'utilisateur. En d'autres termes, nous mettons en place un réseau d'apprentissage non supervisé pour la classification de la pertinence afin d'assister le système. Ainsi, ses performances peuvent être améliorées automatiquement.

La technique de Kohonen permet de créer un réseau qui stocke les informations de telle manière que les relations topologiques en phase d'apprentissage soient maintenues. Le réseau SOM est un moyen de représentation de données multidimensionnelles en espaces de dimensions réduites (figure 5.8). Ce processus de réduction de la dimension des vecteurs est essentiellement une technique de compression de données connue sous le nom de «quantification vectorielle». L'algorithme d'apprentissage SOM [66] est résumé comme suit :

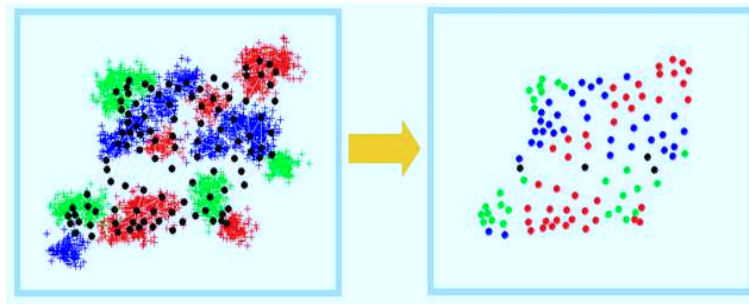


FIG. 5.8 – Réduction de l'espace de signatures à l'aide de SOM

Algorithme 5.5.1 Apprentissage SOM.

Soit $X = \{x_i / 1 \leq i \leq n\}$ les vecteurs caractéristiques de chaque image.

Etape 1 : Choisir arbitrairement un vecteur de x_i .

Etape 2 : Trouver le neurone gagnant j de poids synaptique w_j qui est proche de x_i :

$$\|x_i - w_j\| = \min_{1 \leq k \leq J} \|x_i - w_k\|$$

Etape 3 : Pour tout neurone l en voisinage du neurone gagnant, Mettre à jour son poids :

$$w_l(t+1) = w_l(t) + \alpha(t)(x_i(t) - w_l(t))$$

Avec $\alpha(t)$ est le taux d'apprentissage qui décroît avec le temps, $0 < \alpha(t) < 1$.

Etape 4 : Incrémenter t .

Si $t \leq T$, Avec T est le nombre max d'itérations,

Rétrécir le rayon de voisinage. Aller à l'étape 1.

Si non Apprentissage terminé.

Notre but est d'organiser les images retournées en un ensemble de clusters de telle façon que les images similaires se situent dans le même groupe. L'algorithme SOM organise des vecteurs de grandes dimensions dans une carte à deux dimensions en fonction de la similarité entre les vecteurs. L'application de l'algorithme SOM sur les descripteurs des images retournées, nous permettra d'effectuer le processus de regroupement suivant : les images similaires appartiennent au même neurone ou aux neurones voisins afin de construire des clusters. L'identification de pertinence est effectuée en deux étapes. Après extraction de caractéristiques locales, l'image est représentée par un certain nombre de régions d'intérêt. Les caractéristiques globales et locales alimentent un réseau MISOM. Il s'agit d'un réseau de neurones à deux couches. Tout d'abord, nous utilisons la propriété SOM de réduction de dimension pour traiter les descripteurs locaux.

Ces descripteurs alimentent la couche SOM inférieure en vue de réduire la dimension du vecteur d'entrée. Pour démarrer le processus d'apprentissage, la couche SOM inférieure est formée par les apports de toutes les régions d'images. Lorsque l'apprentissage est réalisé, chaque région est associée à son neurone gagnant sur la carte SOM.

Les positions des noeuds gagnants sont combinées avec les caractéristiques globales. Ainsi, les signatures combinées, sont ensuite traitées avec la couche supérieure SOM. Par conséquent, la couche inférieure est utilisée pour coder les régions, alors que le niveau haut est utilisé pour l'identification de pertinence. D'un autre point de vue, chaque neurone dans la carte SOM est marqué par une liste d'images qui sont similaires et qui sont dans le même groupe. En outre, les images appartenant au même groupe que la requête sont jugées pertinentes, et vice versa. La figure 5.9 montre le traitement automatique de la pertinence.

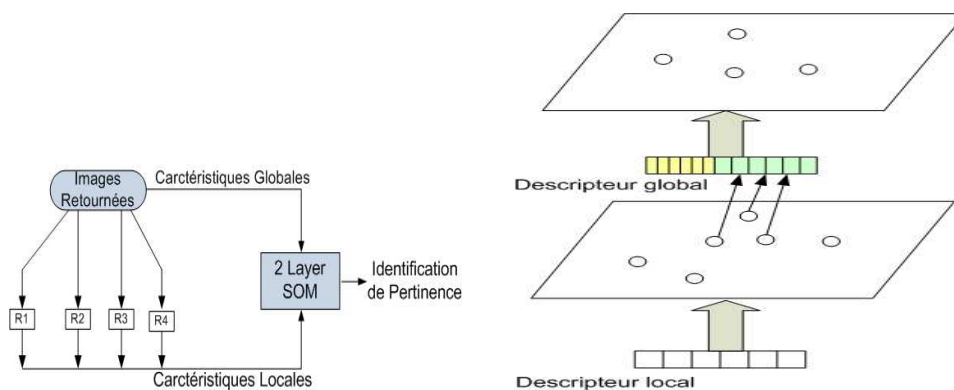


FIG. 5.9 – *Process d'identification de pertinence en utilisant MISOM*

5.5.3 Apprentissage de la similarité

Les images positives et négatives retournées sont ensuite traitées par le processus de mesure de similarité. Afin d'améliorer les performances de recherche dans les itérations suivantes, une méthode non linéaire de retour de pertinences utilisant une fonction radiale (Radial Basis Function (RBF)) est utilisée [93]. La méthode baptisée "Single Radial Basis Function based RF (SRBF_RF)" consiste en une fonction d'une seule variable avec deux paramètres de contrôle : centres C et déviation standard σ . La fonction RBF de forme gaussienne est définie comme :

$$f(x, c) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - c)^2}{2\sigma^2}\right) \quad (5.20)$$

Le processus RBF utilise les descripteurs des images pertinentes et non

pertinentes, à savoir X_r et X_n respectivement, et mis à jour les centres et les déviations à travers les retours de pertinence à chaque itération suivant les formules ci-dessous :

$$c(t+1) = \bar{x}_r - \alpha_n(\bar{x}_n - c(t)) \quad (5.21)$$

$$\sigma = \exp(-\beta.Std(x_r)) \quad (5.22)$$

Avec :

α_n est un paramètre de poids négatif, β est un paramètre de réglage, Std est la déviation standard et \bar{x}_r , \bar{x}_n sont respectivement la moyenne des descripteurs d'images pertinentes et non pertinentes et elles sont définies comme suit :

$$Std = \left(\frac{1}{N_r - 1} \sum_{i=1}^{N_r} (x_{ri} - \bar{x}_r)^2 \right)^{1/2} \quad (5.23)$$

$$\bar{x}_r = \frac{1}{N_r} \sum_{i=1}^{N_r} x_{ri} \quad (5.24)$$

$$\bar{x}_n = \frac{1}{N_n} \sum_{i=1}^{N_n} x_{ni} \quad (5.25)$$

Les centres définis dans l'équation (5.21) dénotent la location des caractéristiques similaires et la largeur σ dans l'équation (5.22) déterminent la déviation des caractéristiques autour du centre. La mesure de similarité entre le vecteur x et c est définie comme suit :

$$s(x, c) = \sum_{i=1}^p f(x_i, c_i) \quad (5.26)$$

5.5.4 Evaluation des performances

Méthode : MISSL_based RF

Pour effectuer notre évaluation, nous avons utilisé notre base d'images d'algues. Un nombre total de 40 images, une image de chaque espèce, sont choisies aléatoirement pour former des requêtes. Les premières images sont ensuite affichées pour commencer le processus de bouclage de pertinence. Dans le cas idéal, toutes les images affichées contiennent la même espèce. La performance est mesurée pour 40 images requêtes.

Dans cette expérience, on mesure la performance du système pour chaque espèce. L'efficacité de la recherche est définie en terme des taux rappel et précision. La figure 5.10 montre les courbes rappel/précision pour chaque itération.

Les observations suivantes sont déduites des résultats : D'abord, les performances du système après deux itérations se sont considérablement améliorées par rapport au schéma non interactif ($t=0$) et les améliorations sont importantes. Ensuite, après les sessions de l'apprentissage interactif, le système répond favorablement et les résultats s'améliorent davantage au fil des itérations.

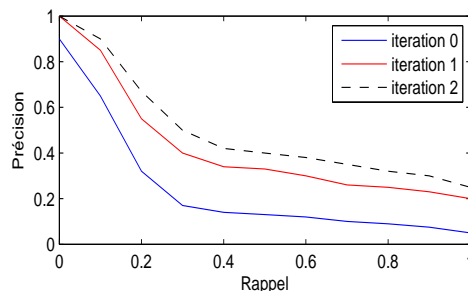


FIG. 5.10 – Courbes Rappel/Précision par itération.

Méthode :MISOM_based ARF

Pour valider la méthode que nous avons introduite, nous étudions, tout d'abord, le comportement du système vis-à-vis de la méthode du bouclage de pertinence automatique proposée et la comparer avec une méthode sans bouclage. Ensuite nous comparons les performances du système d'une part avec une technique de bouclage de pertinence automatique où les signatures sont globales et d'autre part avec le bouclage de pertinence classique où l'utilisateur pourra affiner la recherche en jugeant manuellement les résultats. Comme mentionné en haut et afin d'améliorer la performance du système de recherche en bouclage automatique, les descripteurs de recherche et de classification devraient être différents.

Les figures 5.11 et 5.12 illustrent un exemple de recherche sans et avec retour de pertinence automatique respectivement. Elles montrent que la méthode sans bouclage a des difficultés à retourner des images pertinentes et que le bouclage automatique l'emporte sur cette dernière.

Pour montrer les performances de notre approche, nous avons comparé la méthode proposée avec la technique introduite dans [62] où le retour de pertinence automatique est conduit par des descripteurs globaux. La figure 5.13 présente le résultat de ces deux méthodes. Les courbes rappels / précision montrent une amélioration significative apportée par notre approche. Toutefois, cette dernière présente un temps de recherche légèrement supérieur par rapport à celle proposée dans [62].



(a) Image requête

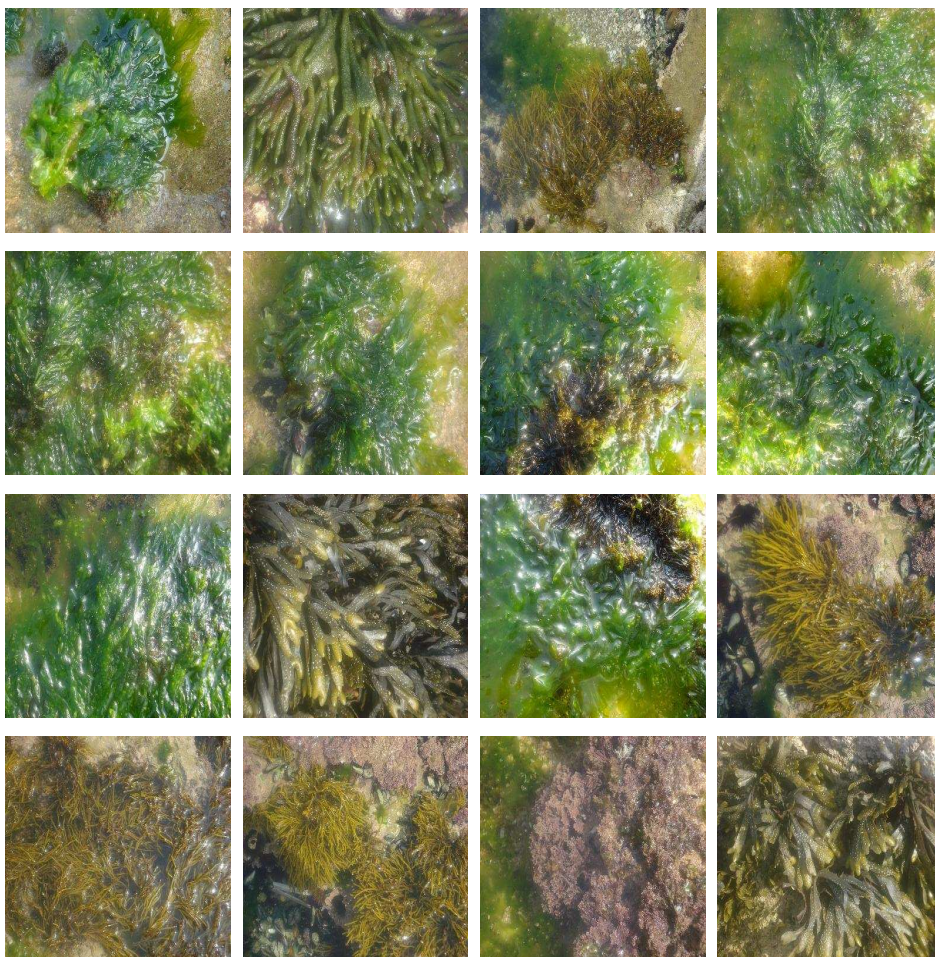


FIG. 5.11 – *Exemple de recherche d'images sans bouclage.*

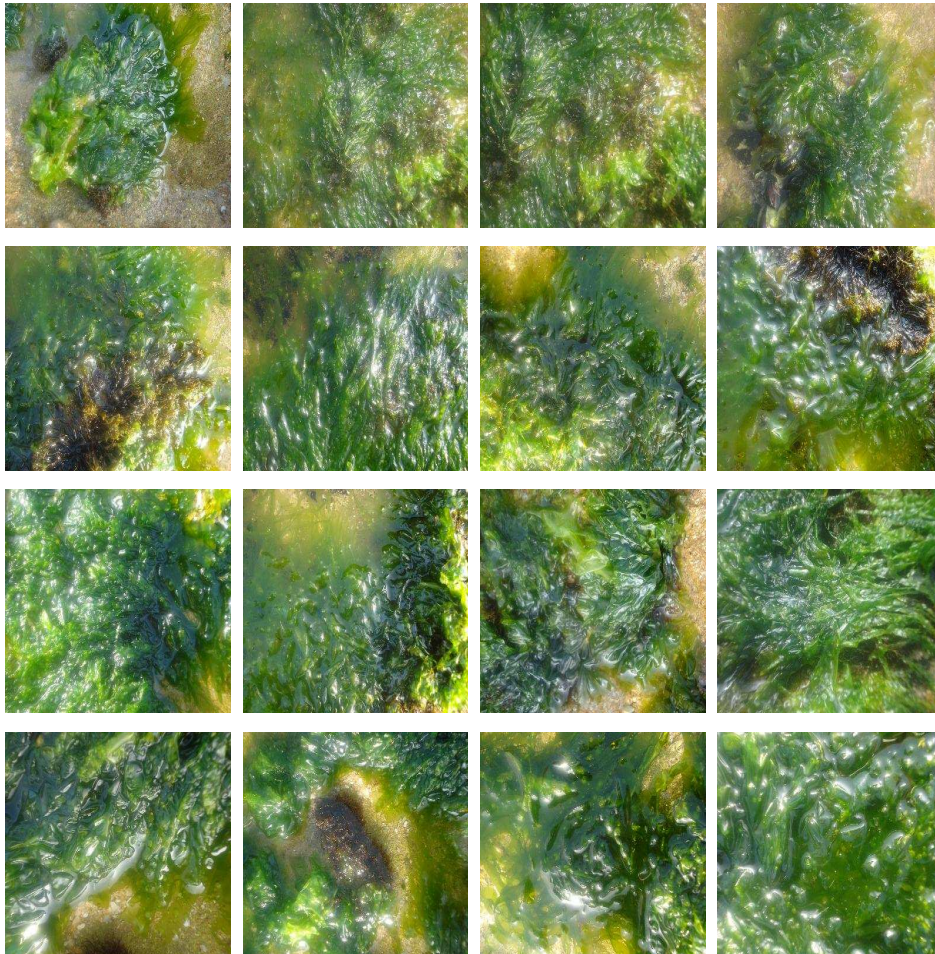


FIG. 5.12 – *Exemple de recherche d'images avec bouclage de pertinence automatique.*

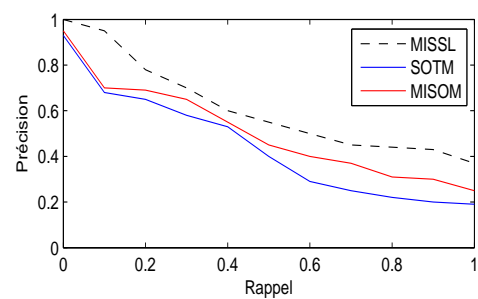


FIG. 5.13 – *Courbes rappel précision comparant des méthodes de bouclage de pertinence après trois itérations*

Aussi, Dans la figure 5.13 nous avons comparé les résultats des méthodes

de recherche avec retour de pertinence automatique et ceux de la recherche interactive utilisant la méthodes "MISSL_based RF".

La comparaison de ces résultats montre que la recherche interactive produit une amélioration par rapport aux méthodes de bouclage de pertinence automatique. En effet, le bouclage de pertinence automatique ne remplace pas le bouclage interactive géré par l'utilisateur, mais il contribue à l'amélioration des résultats d'une recherche sans bouclage. En outre, Après avoir reçu le résultat de recherche en utilisant le retour de pertinence automatique, l'utilisateur pourra affiner manuellement la recherche suivant son besoin. On appelle ce type de recherche : bouclage de pertinence combiné ou semi automatique. Ce type de recherche permet d'améliorer les résultats obtenus par le bouclage automatique et de minimiser le nombre d'itérations de la recherche par bouclage classique.

5.6 Conclusion

Au cours de ce chapitre, nous avons développé un algorithme de bouclage de pertinence. Pour cela, nous avons pris en compte la distribution statistique des images de la base au sein de l'espace des descripteurs pour trouver les images similaires aux images annotées positivement. Notre système prend en compte aussi bien les retours positifs que négatifs. A chaque retour utilisateur, une prédiction de label des images non encore étiquetées est effectuée dans l'espace de descripteurs. Cet algorithme est basé sur une méthode de propagation des labels des images étiquetées positives ou négatives (+1/-1). À partir de cela, on extrait alors les nouvelles images résultantes qui sont présentées à l'utilisateur. Nous avons formulé le processus de la recherche interactive comme étant un apprentissage semi supervisé et présenté une nouvelle solution en utilisant les caractéristiques de l'apprentissage de multiples instances. Cette méthodologie nous permet d'obtenir une propriété de distribution des images dans l'espace des caractéristiques en accord avec la notion de similarité visuelle humaine. De plus, ce modèle nous a permis de pratiquer une liaison entre la notion d'objets et d'images globales.

Nous avons aussi présenté une méthode de bouclage de pertinence automatique. Nous avons proposé un réseau de neurones de type SOM afin de classifier les images pertinentes et non pertinentes automatiquement. Cela permettra d'éviter l'intervention de l'être humain pour juger à chaque itération les résultats de recherche. En outre, l'utilisateur pourra affiner le résultat de recherche s'il n'est pas satisfait. Par conséquent, la méthode proposée ne remplace pas le bouclage de pertinence classique où l'utilisateur annote manuellement les résultats, mais elle contribue à l'amélioration des résultats d'une recherche sans bouclage, et aide à minimiser les itérations du processus de bouclage de pertinence.

Ces travaux ouvrent la voie à des améliorations. En effet, nous avons considéré deux partitions bien distinctes de l'espace des descripteurs : les images pertinentes et les images non-pertinentes. Cette segmentation de l'espace reste unimodale [11]. Il pourrait être intéressant d'utiliser un bouclage de pertinence permettant de réaliser un partitionnement multimodal de l'espace.

Chapitre 6

CONCLUSION ET PERSPECTIVES

La recherche d'images par le contenu est une problématique majeure pour la gestion des grandes bases d'images, et attire beaucoup plus d'importance avec l'augmentation du nombre d'images numériques. Dans cette thèse, nous nous sommes intéressés à l'application de techniques de fouille de données et d'apprentissage dans le cadre de la recherche d'images dans une base d'images d'algues. Afin de pouvoir traiter le caractère spécifique du problème, nous nous sommes tournés vers trois aspects fondamentaux de la recherche d'images par le contenu : indexation, recherche, et retour de pertinence. Notre étude s'est concrétisée en la réalisation d'une chaîne de traitement automatisé. Nous avons ainsi développé de nouvelles méthodes de segmentation et de reconnaissance d'images en veillant à respecter les spécificités des systèmes CBIR. Pour conclure ce manuscrit, nous allons passer en revue l'ensemble de nos apports au domaine de la recherche d'images. Ensuite, nous relaterons l'ensemble des perspectives de recherche nous semblant intéressantes.

6.1 Bilan

Cette thèse a été consacrée au développement d'un système de recherche d'images fixes couleurs. La nature même de ce sujet nous a amené à traiter un très large panel de domaines ne se limitant pas seulement à celui du traitement de l'image. De plus, outre l'ensemble des sujets abordés, nous avons dû garder à l'esprit, tout au long de cette thèse, que la finalité résidait dans la conception d'une plate-forme utilisable dans des conditions réelles et actuelles. Cela nous a amené à développer des outils facilement appréhendables pour un utilisateur non expert. Nous avons donc embrassé l'ensemble des processus impliqués dans l'indexation d'images. Nous allons passer rapidement en revue l'ensemble des points traités et en préciser nos apports. Dans une première partie, nous avons effectué un retour sur la notion de perception visuelle, ainsi qu'une remise en cause du

concept de similarité tel qu'il est abordé dans la littérature, en nous appuyant sur des considérations relevant des domaines du traitement de l'image. Une bonne compréhension du système visuel humain et la nature des images utilisées nous semblait nécessaire à l'élaboration de descripteurs pertinents, et nous a permis de mieux cibler les contraintes associées. La conduite d'expérimentations en situations réelles a mis en relief les subtilités de la notion de similarité visuelle. De l'ensemble de ces études nous avons tiré plusieurs leçons :

- Traiter une image dans sa globalité et en lui associant un ensemble de descripteurs est contraire aux processus mis en jeu dans l'évaluation de la similarité. Afin de respecter la logique humaine, nous avons donc développé une nouvelle méthode de segmentation pour la détection grossière de régions des principaux objets de l'image.
- Une simple métrique géométrique va à l'encontre des résultats expérimentaux portant sur la similarité visuelle. C'est afin de modéliser plus fidèlement celle-ci que nous avons appliqué des concepts cognitifs à l'élaboration d'une métrique contextuelle.
- Dans des conditions réelles de recherche par l'exemple, un utilisateur a besoin d'un outil lui permettant d'initialiser le plus précisément possible sa requête, celle-ci servant de base à toute procédure de recherche de similarité. C'est pour cela que nous avons distingué deux principes de recherche laissés au choix de l'utilisateur : Recherche par Région et recherche par image entière.
- Enfin, le fossé sémantique, de par sa nature même, restera dans une certaine proportion toujours infranchissable. C'est pour cela qu'il nous a semblé nécessaire de disposer d'un processus de bouclage de pertinence, évoluant et faisant la liaison entre les objets et l'image dans leur globalité.

Dans cette thèse, nous proposons deux stratégies de recherche interactive. La première travaille dans le cadre de l'apprentissage semi supervisé et présente une nouvelle solution en utilisant les caractéristiques de l'apprentissage de multiples instances. Cet algorithme est basé sur une méthode de propagation des labels des images annotées positives ou négatives. La méthode présente un temps de calcul contrôlé et l'évaluation montre son apport vis-à-vis des performances de recherche.

La seconde stratégie de recherche par bouclage de pertinence repose sur le retour de pertinence automatique. L'algorithme utilise un réseau de neurones de type SOM à deux couches afin de classifier les images retournées automatiquement et procédera ensuite à un raffinement de la requête. Cela permettra d'éviter l'intervention de l'être humain pour juger les résultats de recherche ou minimiser

le nombre d'itérations de recherche lancées par l'utilisateur s'il n'est pas satisfait. Cette technique fonctionne correctement lorsque l'utilisateur recherche des images proches de la requête du point de vue des caractéristiques primaires.

L'application visée concerne les images d'algues. Un système de traitement a été proposé (SWIR ou SeaWeed Image Retrieval), depuis la segmentation, jusqu'à la requête. Nous avons testé les performances de notre système en étudiant l'influence de chaque élément de la chaîne sur le processus de recherche. Les résultats obtenus sont satisfaisant au sens des mesures classiques dans le domaine reposant sur la dépendance Rappel / Précision. Les résultats obtenus pour l'ensemble des techniques élaborées ont non seulement permis d'obtenir un moteur cohérent et efficace dans sa globalité, mais sont aussi simple d'utilisation. En effet, nous avons, tout au long de ce travail, cherché à rester le plus généraliste possible tout en évitant à l'utilisateur d'avoir des réglages à effectuer de quelque nature qu'ils soient. De plus, comme nous l'avons précisé en début de chapitre, le but était de fournir une plate-forme d'indexation et de recherche dédiée à une application aux images d'algues. Ce travail a abouti et sert actuellement de base à des recherches plus poussées au sein du laboratoire.

6.2 Perspectives

Les évolutions possibles de ce travail sont nombreuses en raison du nombre de paramètres impliqués dans les différentes étapes. Dans le cadre de l'extraction des zones d'intérêt, la description de l'image par régions permet de modéliser de façon pertinente la notion de similarité. Mais, comme nous l'avons vu dans les chapitres précédents, la segmentation n'est pas toujours optimale car elle ne permet pas de regrouper des régions sémantiques spatialement disjointes. Il serait donc intéressant de remédier à cela en améliorant la méthode de segmentation, en la rendant capable de prendre en compte ce type de problème. Il semble que cela puisse conduire à l'obtention d'une meilleure estimation de la similarité visuelle. À ce niveau, l'utilisation des lois de Gestalt peut servir de base pratique à un agencement cohérent avec la notion de similarité psycho-visuelle. Il serait ensuite intéressant de classifier ces régions selon un graphe ou une chaîne, et d'effectuer une mesure de similarité basée sur la comparaison de graphes. Les résultats apportés par l'utilisation d'une telle méthode structurale ne pourront qu'améliorer le moteur.

Dans l'introduction de cette thèse, nous avons commencé par évoquer l'annotation d'images par mots-clés comme première approche pour la recherche d'images. Plus précise que l'annotation d'images, l'annotation des régions détectées permettrait de décrire les images par leur composition en termes d'ob-

jets sémantiques. Des requêtes sophistiquées pourraient alors être formulées sur le contenu aussi bien en termes de composition visuelle que sémantique. Par ailleurs, toutes les régions annotées par des mots-clés pourraient être regroupées en catégories sémantiques. Il serait alors pertinent de formuler des requêtes par composition logique à la fois de composantes visuelles et de composantes sémantiques. L'inconvénient majeur de l'annotation demeure la pénibilité de la tâche d'annotation sur une base entière. Pour remédier à cet inconvénient, l'annotation des régions pourrait être semi-automatisée en propageant les mots-clés de régions annotées vers les régions d'aspect visuel similaire.

Concernant l'apprentissage interactif, nous avons considéré deux partitions bien distinctes de l'espace des descripteurs : les images pertinentes et les images non-pertinentes. Cette segmentation de l'espace reste unimodale. Il pourrait être intéressant d'utiliser un bouclage de pertinence permettant de réaliser un partitionnement multimodal de l'espace.

Enfin, il est envisageable d'étudier le comportement de la méthode avec des bases plus grandes pour voir l'influence sur le temps de calcul des descripteurs, des signatures et surtout du bouclage de pertinence. Bien que les calculs soient rapides, le fait d'utiliser une base de grande taille va sans doute poser des problèmes de temps de calcul. Nous envisageons aussi d'élargir le champ d'application. En effet, des travaux de recherche, actuellement en cours de finalisation, utilisent SWIR pour la reconnaissance d'espèces d'algues à partir d'images de leurs organes constitutifs ou via une description textuelle. Il est ainsi possible de retrouver en très peu de manipulations les images associées à une description donnée par l'utilisateur (couleur, taille et forme du stipe, habitat, ...).

BIBLIOGRAPHIE

- [1] Armitage, L. & Enser, P. 'Analysis of user need in image archives', *Journal of Information Science* ,23(4), 287–299.1997.
- [2] S. Ayyappan, "Seaweed Cultivation and Utilization" National Academy of Agricultural Sciences, pp 22, 2003.
- [3] Bach, F. and Jordan, M. Learning spectral clustering. In S. Thrun, L. Saul, and B. Scholkopf (Eds.), *Advances in Neural Information Processing Systems 16 (NIPS)* (pp. 305 - 312). Cambridge, MA : MIT Press. 2004.
- [4] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jam, and C. F. Shu, "The Virage Search Engine : An Open Framework for Image Management",In *Proc. of the SPIE Storage and Retrieval for Still Image and Video Databases IV*,volume 2670, pages 77-87, February 1996.
- [5] K.Barnard, P. Duygulu, R.Guru, P.Gabbur et D. Forsyth. "the effects of segmentation and feature choice in a translation model of object recognition".*Computer Vision and Pattern Recognition*, 675-682.2003
- [6] P. Bolon, J.-M. Chassery, J.-P. Cocquerez, D. Demigny, C. Graffigne, A. Montanvert,S. Philipp, R. Zédoudj, and J. Zérubia. *Analyse d'images, filtrage et segmentation*. Masson, 1995.
- [7] Beckmann, N., Kriegel, H.-P., Schneider, R., and Seeger, B. The R*Tree : An efficient and robust access method for points and rectangles. In *Proc. of ACM SIGMOD*.1990.
- [8] Beigi, M., Chang, S., Smith, J. R. & Benitez, A. 'Visual information retrieval from large distributed online repositories', *Comm. ACM* ,40(12), 63 – 71.1997.
- [9] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Functions*. Plenum, New York NY, 1981.
- [10] X. Bresson and T.F. Chan. Non-local Unsupervised Variational Image Segmentation Models, UCLA CAM Report 08-67, 2008.
- [11] R. Brunelli and O. Mich. Image Retrieval by Examples. *IEEE Transactions on Multimedia*, 2(3) :164-171, 2000.
- [12] A. Buades, B. Coll, and J.M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2) :490-530, 2005.
- [13] G. Caenen, G. Frederix, A. Kuijk, E. Pauwels, and B. Schouten. Show me what you mean! PARISS : A CBIR-interface that learns by example. In *International Conference on Visual Information Systems (Visual'2000)*, volume 1929, pages 257-258, 2000.

- [14] J. Canny. "A computational approach to edge detection ". IEEE transactions on Pattern Analysis and Machine Intelligence, 8(6). 679-698. 1968
- [15] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld : Image Segmentation Using Expectation-Maximization and Its Application to Image Querying", IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(8) : pp. 1026-1038, August 2002.
- [16] Chang, E., Li, B. T., Wu, G., & Goh, K. (2003). Statistical learning for effective visual information retrieval. IEEE International Conference on Image Processing. Barcelona, Spain.
- [17] O. Chapelle, J. Weston, and B. Scholkopf. Cluster kernels for semi-supervised learning. volume 15 of NIPS. 2003.
- [18] I. Charon, A. Germa, and O. Hudry. Méthodes d'optimisation combinatoire, Collection Pédagogique de Télécommunication. Masson, 1986.
- [19] Y. Chen, X. Zhou, and T. S. Huang. One-Class SVM for Learning in Image Retrieval. In IEEE International Conference on Image Processing, Thessaloniki, Grèce, 2001.
- [20] V. Chitkara "Color-Based Image Retrieval Using Compact Binary Signatures" Technical Report, University of Alberta, 2001
- [21] Chung, F. Spectral graph theory. Conference Board of the Mathematical Sciences, Washington. 1997.
- [22] Cohn, D. Active learning with statistical models. Journal of Artificial Intelligence Research, 4, 129-145. 1996
- [23] R.R Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, S. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data Diffusion map. Proceeding of the National Academy of Science, 102 : 7426-7431, May 2005.
- [24] Base Columbia. <http://www.cs.columbia.edu/CAVE/research/softlib/coil-100.html>.
- [25] D. Comaniciu, P. Meer. " Mean Shift : A Robust Approach Toward Feature Space Analysis. IEEE Trans. Pattern Anal. Mach. Intell., 24(5), 603-619. 2002
- [26] R.W. Connors and C.A. Harlow. A theoretical comparison of texture algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2(3) :204-222. 1980.
- [27] Base de Corel. <http://www.corel.com>.
- [28] T. M. Cover and J. A. Thomas. Elements of information theory. John Wiley & Sons, 1991.
- [29] I. J. Cox, M. L. Miller, T. P. Minka, T. Pappas, and P. N. Yianilos. The Bayesian Image Retrieval System, PicHunter : Theory, Implementation and Psychophysical Experiments. IEEE Transactions on Image Processing, 9(1) :20-37, Janvier 2000
- [30] J. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression", IEEE Trans., 1988, ASSP-36(7).
- [31] A.P. Dempster, N.M. Laird and D.B. Rubin. Maximum Likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 39(1) :1-38, 1976.

- [32] V.K. Dhargalkar and N. Pereira, "Seaweed : Promising plant of the millennium" *Science and Culture*, vol.71 (3-4), pp 60-66, 2005.
- [33] Dietterich, T. G., Lathrop, R. H., and Lozano-Perez, T. "Solving the multiple instance problem with axis parallel rectangles". *Artif. Intell.* Vol 98, pp 31-71. 1997
- [34] Donath, W. E. and Hoffman, A. J. (1973). Lower bounds for the partitioning of graphs. *IBM J. Res. Develop.*, 17, 420 - 425. 1973.
- [35] A.A. Efros and W.T. Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of the Conference on Computer graphics and interactive techniques, SIGGRAPH*, pages 341-346, New York, NY, USA, 2001. ACM.
- [36] R. Fagin, "Combining fuzzy information from multiple systems," *Journal of Computer and System Science*, vol. 58, no. 1, pp. 83–99, 1999.
- [37] U.Fayyad, G.Piatestsky-Shapiro and P.Smyth. "From Data Mining to Knowledge Discovery in Databases". *AI Magazine* 17(3), pp. 37-54, 1996.
- [38] H Freeman and L.S. Davis. "Corner finding algorithm for chain-coded curves". *IEEE trans. Computers*, 26(3) : 297 303.1977.
- [39] Fiedler, M. Algebraic connectivity of graphs. *Czechoslovak Math. J.*, 23, 298 - 305. 1973.
- [40] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkhani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by Image and Video Content : The QBIC System", *IEEE Computer Magazine*, 28(9), pp. 23-32, September 1995.
- [41] J. Fournier. Indexation d'images par le contenu et recherche interactive dans les bases généralistes. PhD thesis, Université de Cergy-Pontoise, octobre 2002.
- [42] J. M. Francos, A. Z. Meiri, and B. Porat. "A unified texture model based on a 2-d wold like decomposition". *IEEE Trans. Signal Process.* 41 :2665–2678, 1993.
- [43] A.Gagalowicz, "Vers un modèle de textures ", Thèse de doctorat univ. Pierre et Marie Curie, Paris V, mai 1983
- [44] [7] B. Gaveau, A. Lesne, L.S Schulman, Spectral signatures of hierarchical relaxation, *Physical Letters A* : 258, pp 222-228
- [45] M.Gondran and M.Minoux. Grpahes et algorithmes. Collection de la Direction des Etudes et Recherches d'Electricité de France, Eyrolles, 1979.
- [46] V. Gouet and N. Boujemaa. About optimal use of color points of interest for content-based image retrieval. Internal Report, INRIA Rocquencourt, 2002.
- [47] R. C Gonzales and R.E. Woods. *Digital image processing*. Prentice-Hall. New Jersey, 2002.
- [48] R.M.Haralick, K.Shanmugam, and I.Dinstein. Textural features for image classification. *IEEE Trans. Sys. Man. Cybern.* 3(6) :610 321,1973
- [49] S. Aksoy and R. Haralick. Textural features for image database retrieval. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, in conjunction with CVPR'98, pages 45–49, Santa Barbara, CA, June 1998.

- [50] J. Harris and M. Stephens. A combined corner and edge detector. 4th ALVEY Vision Conference, pages 147–151, 1988.
- [51] Hastie, T, Tibshirani, R and Friedman, J. H. The elements of statistical learning, Springer, August 2001.
- [52] Hastings, S. Query categories in a study of intellectual access to digitized art images, *in* ‘Proc. of the 58th Annual Meeting of the American Society for Information Science’. 1995.
- [53] A. Heinrichs, D. Koubaroulis, B. Levienaise-Obadia, P. Rovidia, and J. Jolion. Image indexing and content based search using pre-attentive similarities. In RIAO2000, volume 2, pages 1616–1631, Paris, France, April 2000.
- [54] S. L. Horowitz and T. Pavlidis. Picture segmentation by a tree traversal algorithm. *J. ACM*, 23(2) :368-388. 1976.
- [55] H. Houissa and N. Boujemaa. A new angle-based spatial modeling for query by visual thesaurus composition. In IEEE International Conference on Image Processing, 2007.
- [56] J. Hafner, H.S.Sawhney and W.Equitiz. “Efficient colour histogram indexing for quadratic form distance functions. *IEEE Transactions on pattern Analysis and Machine Intelligence*,17(7) :729-736, 1995.
- [57] M. K. Hu. Visual pattern recognition by moment invariant, computer methods in image analysis. *Transactions of information theory*, 8(2) :179-187, 1962.
- [58] J. Huang, S.Kumar, M. Mitra, W.J. Zhu and R. Zabih. “Image Indexing Using Colour Correlograms”. *Proc. of Conf. on Comp. Vision and Patt. Recog.(CVPR)*, San Juan (Puerto Rico), pp. 762-768, 1997.
- [59] <http://www-rocq.inria.fr/imedia/cbir-demo.html>
- [60] [6] S. Ingve, A. Eriksen Kasper, M. Sergei and Sneppen Kim, Diffusion on complex networks : a way to probe their large-scale topological structures. *Physica A :Statistical Mechanics and its Applications*, 336(1-2) : 163-173, May 2004.
- [61] Y. Ishikawa, R. Subramanya, and C. Faloutsos. Mindreader : Query databases through multiple examples. In *International Conference on Very Large Data Bases*, pages 218-227, New York, USA, Août 1998. ACM.
- [62] K .Jarrah , P.Muneesawang, I.Lee and L.Guan "Minimizing Human-Machine Interactions in Automatic Image Retrieval" *CCGEI* may 2004.
- [63] S. Jeannin “MPEG-7 Visual part of experimentation model, Version 9.0”. 55th Mpeg Meeting. Pisa, Italy.2001.
- [64] S. Jeong, K. Kim, B. Chun, J. Lee, and Y.J. Bae, “An effective method for combining multiple features of image retrieval,” in *Region 10 Conference*. IEEE, pp. 982–985. 1999.
- [65] W.-Y. Kim et Y.-S. Kim “A new region based shape descriptor”. ISO/IEC MPEG99/ M5472, Maui, Déc.1999.
- [66] T. Kohonen, "Self-Organizing Maps", 3rd ed. New York : Springer-Verlag, vol. 30, Springer Series in Information Sciences. 2001
- [67] W. Kropatsch and H. Bishof. *Digital Image Analysis*. Springer-Verlag, 2001.

- [68] H.W.Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2 : 83—97, 1955
- [69] C.C. Kuhlthau. Principle for Uncertainty for Information Seeking. *Journal of Documentation*, 49(4) :339-355, 1993.
- [70] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22 :79–86, 1951.
- [71] T. Kurita and T. Kato. Learning of Personal Visual Impression for Image Database Systems. In *Second International Conference on Document Analysis and Recognition*, pages 547-552,1993.
- [72] L. Liang, C. Liu, Y.Q. Xu, B. Guo, and H.Y. Shum. Real-time texture synthesis by patch-based sampling. *ACM Trans. Graph.*, 20(3) :127-150, 2001.
- [73] F. Liu and R. W. Picard. Periodicity, directionality, and randomness : Wold features for image modeling nad retrieval. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(7) :722–733, July 1996.
- [74] Lovasz, L . Random walks on graphs : A survey. *Combinatorics, Paul Erdos is Eighty 2* , pp. 1-46.1993.
- [75] von Luxburg, V. A Tutorial on Spectral Clustering. *Statistics and Computing*, 17(4), 395-416. 2007.
- [76] J.Li, J. Z Wang and G.Wiederhold. IRM : integrated region matching for image retrieval. *ACM Multimedia*, pp 147-156, 2000.
- [77] W. Y. Ma and B. S. Manjunath. Edge flow : A framework for boundary detection and image segmentation. In *Proc. IEEE International conference on Computer Vision and Pattern Recognition (CVPR)*, June 1997.
- [78] W. Ma and B.S Manjunath. Netra : A toolbox for navigating large image databases. *Multimedia Syst.* 7(3) :184-198.1999
- [79] J. B. MacQueen (1967) : "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1 :281-297
- [80] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 43(1) :7-27, 2001.
- [81] S. Mallat "A Wavelet tour of signal Processing". Academic Press, 2nd edition, 637 p.1999.
- [82] B. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence (Special Issue on Digital Libraries)*, 18(8) :837–842, August 1996.
- [83] MARCELJA, S. "Mathemathical description of the response of simple cortical cells » *Journal of Optical Society of America*, 70(11) :1297-1300. 1980.
- [84] H.Marfaing. L'alguie alimentaire : une réglementation stricte pour un aliment d'avenir. *Option Qualité 298* : 15-20.Novembre 2010.
- [85] Martin et al. [2001] D. Martin, C.Fowlkes, D.Tal, J. Malik "A Database of Human Segmented Natural Images and its Application to Evaluating

- Segmentation Algorithms and Measuring Ecological Statistics". IEEE International Conference of Computer Vision (ICCV), Volume 2, pages 416-423.2001.
- [86] Meila, M. and Shi, J. A random walks view of spectral segmentation. In 8th International Workshop on Artificial Intelligence and Statistics (AISTATS). 2001.
- [87] B.M Mehtre, M.S. Kankanhalli, and W.F Lee. "Shape measures for content based image retrieval : a comparison. *Inf. Process. Manage.* 33(3) : 319-337, 1997.
- [88] F. Mokhtarian et A. Mackworth "A theory of multiscale, curvative-based shape representation for planar curve ". *IEEE Transactions on pattern Analysis and machine Intelligence*, 14 :789-805.1992.
- [89] A. Monadjemi. Towards Efficient Texture Classification and Abnormality Detection. PhD thesis, University of BRISTOL. 2004.
- [90] P. Montesinos, V. Gouet, and R. Deriche. Differential invariants for color images. *International Conference on Pattern Recognition*, pages 838–840, 1998.
- [91] O.J. Morris, J.Lee, and A.Constantinidès. Grpah theory for image analysis : an approach based on the shortest spanning tree. *IEEE Proceedings*, 133(2) :193-229. 1986.
- [92] P. Muneesawang and L. Guan. Multiresolution-histogram indexing and relevance feedback learning for image retrieval. In *International Conference on Image Processing (ICIP'2000)*, Vancouver, Canada, September 2000.
- [93] P. Muneesawang, and L. Guan, "Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture", *IEEE Transactions on Neural Networks*, vol. 13, no 4, pp. 821-834, 2002.
- [94] M. Nadenau. "*Integration of human color vision models into high quality image compression*". PhD thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), 2000.
- [95] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The QBIC project : Querying images by content, using color, texture, and shape. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 173–187, February 1993.
- [96] M. Ortega, Y. Rui, K. Chakrabarti, K. Porkaew, S. Mehrotra, and T.S. Huang, "Supporting ranked boolean similarity queries in mars," *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 6, pp. 905–925, 1998.
- [97] N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition* , 26(9) : 1277-1294, 1993.
- [98] A. Pentland, R. Picard, and S. Sclaroff. Photobook : Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3) :233–254, June 1995.
- [99] A. Perchant. Morphisme de graphes des attributs flous pour la reconnaissance structurelle de scènes, PhD thesis, Ecole Supérieure de Télécommunications, 2000.

- [100] E. Persoon and K.S.Fu. "Shape discrimination using Fourier descriptors. IEEE Trans. Sys, Man, Cyb, 7 :170 179, 1977.
- [101] K.N. Plataniotis and A.N. Venetsanopoulos, Color image processing and applications. Springer-Verlag, 2000.
- [102] J.-G. Postaire. *De l'image à la décision*. Dunod informatique, 1987.
- [103] T. Quack, U. Monich, L. Thiele, and B. Manjunath. Cortina : A system for large-scale, Content-based web image retrieval. In *ACM Multimedia 2004*, <http://vision.ece.ucsb.edu>, Oct 2004.
- [104] Rahmani, R. and Goldman, S. A. "MISSL : multiple instance semi-supervised learning". In Proc. of ICML, 2006.
- [105] Rilo, E. and Hollaar, L., 'Text databases and information retrieval', *ACM Computing Surveys* ,28(1), 133–135.1996
- [106] Risi Imre Kondor and John D. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In ICML '02 : Proceedings of the Nineteenth International Conference on Machine Learning, pages 315-322, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.2002
- [107] J. Rissanen. Modeling by shortest data description. *Automatica*, 1978.
- [108] J. Rocchio. Relevance Feedback in Information Retrieval, pages 313-323. *The Smart Retrieval System-Experiments in Automatic Document Processing*. Geraxd Slaton, prentice-hall edition, 1971.
- [109] Y. Rubner, C. Tomasi, et L. Guibas. The earth mover distance as a metric for image retrieval. Stanford University Technical Report, 1998.
- [110] Rubner, Y. Perceptual metrics for image database navigation. Doctoral dissertation, Stanford University.1999.
- [111] J. Da Rugna, et H.Konik."Etude comparative de methode de segmentation dans une approche orientée indexation ». Actes du Congrès Reconnaissance des Formes et Intelligences Artificielle, 2004.
- [112] Rui, Y., and Huang, T. (2000). Optimizing learning in image retrieval. Conf on Computer Vision and Pattern Recognition (CVPR) pp. 236-243. Hilton Head, SC.
- [113] Y. Rui, T.S. Huang, and S. Mehrotra, "Content-based Image Retrieval with Relevance Feedback in MARS", in Proc. of the IEEE International Conference on Image Processing , ICIP 97, pages 815-818, Santa Barbara, California, USA , October 1997.
- [114] G.Salton. *Automatic Information Organization and Retrieval*". McGraw Hill. 1968
- [115] G.Salton et C.Buckley . Improving retrieval performance by relevance feedback. *Journal of the American society for information science*, vol. 41(4), pp. 288–297. 1990
- [116] J. Shi et J. Malik."Normalized Cuts and Image Segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*,22(8), 888-905.2000
- [117] Simon P.Wilson, Julien Fauqueur, and Nozha Boujemaa. *Machine Learning Techniques for Multimedia*, chapter Mental Search in Image Databases : Implicit Versus Explicit Content Query, pages 189-204. Springer, 2008.

- [118] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12) :1349-1380, Décembre 2000.
- [119] J. Smith, and S.F. Chang, "VisualSEEK : a fully automated content-based image query system", in *Proceedings of the 4th ACM Multimedia Conference*, Boston, MA, pp. 87-98,1996.
- [120] J. R. Smith and S.F. Chang, "An image and video search engine for the world-wide web", In *Proc. of the SPIE Storage and Retrieval for Image and Video Databases*, volume 3022, pages 85-95, San Jose, CA, USA, February 1997.
- [121] J.R Smith. Image retrieval evaluation. *IEEE Workshop on Content based Access of Image and Video Libraries (CBAIVL'98)*, pp. 112-113. June 1998.
- [122] sfgj
- [123] A. Souissi, H.TABOUT, A. Sbihi "Système de recherche et de navigation dans une base d'arabesques par descripteurs locaux". *VEILLE STRATEGIQUE SCIENTIFIQUE & TECHNOLOGIQUE (VSST'2007)*. octobre 2007. MARRAKECH, Morocco.
- [124] A. Souissi, H.TABOUT, A. Sbihi "GSMGUIDE application GSM destiné à la recherche de l'information" *Workshop sur les technologies de l'information et de la communication (WOTIC'07)*. 5-7 Juillet 2007. RABAT, Morocco.
- [125] A.Souissi, H.Tabout and A.Sbihi, "MIR system for mobile information retrieval by image querying" , *International Journal of Computer Science and Network* VOL.8 No.4, pp. 65-73, April 2008.
- [126] A.Souissi, H.Tabout and A.Sbihi, "Système MMSINFO destiné pour la recherche d'information par des requêtes image dans le réseau mobile 2G et 3G",*9th African Conference on Research in Computer Science - CARI*, Rabat, Morocco, October 27-30, pp161-168. October 2008.
- [127] V. Sridhar M. A. Nascimento and X. Li. Effective and efficient region-based image retrieval. *Journal of visual languages and computing*, 14(2) : 151-179, 2003
- [128] V. Sridhar M. A. Nascimento and X. Li. Effective and efficient region-based image retrieval. *Journal of visual languages and computing*, 14(2) : 151-179, 2003
- [129] R. Stoica, J. Zerubia, and J.M. Francos. Image retrieval and indexing : A hierarchical approach in computing the distance between textured images. In *International Conference in Image Processing (ICIP'01)*, Chicago, USA, October 1998.
- [130] M. A. Stricker and M. Orengo. "Similarity of color images". In *Proc. of the SPIE conference on the Storage and Retrieval for Image and Video Databases III*, pages 381–392, 1995.
- [131] Sutton (R.) et Barto (A.). *Reinforcement learning : An introduction*. MIT Press, Cambridge, MA., 1998.
- [132] M. Swain et D. Ballard. Color indexing. *International Journal of Computer Vision (IJCV)*,7(1) :11–32, 1991.

- [133] H. Tabout, A. Sbihi "Image Indexing Using Color and Texture". Workshop sur les technologies de l'information et de la communication (WOTIC'05). Juin 2005, Kenitra, Morocco.
- [134] H. Tabout, A. Souissi and A. Sbihi "A neural network based image retrieval using color and texture features". Maghrebian Conference on Software Engineering and Artificial Intelligence (MCSEAI) pp. 274-277 December 07-09, 2006, AGADIR, Morocco
- [135] H.Tabout and A.Sbihi "Algae image indexing approach using color and texture features" SITA, November, 2006
- [136] H. Tabout, A. Sbihi "Algae image indexing approach using color and texture features" Conference en Systemes Intelligent : Theories ET Applications (SITA) pp. 88-92, Mohammadia, Morocco.2006.
- [137] H. Tabout, A. Souissi, A. Sbihi "Region-Based Automated Relevance Feedback in Algae Image Retrieval", International Journal of Computer Science and Network Security, VOL.7 No.12 pp. 124-131, December 2007.
- [138] H.Tabout, A. sbihi "A method for algae image indexing and retrieval based on neural network" Information and communication technologies international symposium (ICTIS) ; 3-5 April 2007. FES, Morocco.
- [139] H. Tabout, A. Souissi, A. Sbihi "A TREE STRUCTURED BASED ALGAE IMAGE RETRIEVAL" Systèmes Electroniques-Informatiques & traitement de l'information (SEITI) pp 15-18; Janvier 2007. Mohammedia; Morocco.
- [140] H.Tabout, Y.Chahir, Abdelmoghith Souissi and Abderrahmane Sbihi, "Random walks based segmentation approach for image retrieval", IEEE International Conference on Pervasive Computing and Applications - ICPCA , Alexandria, Egypt, October 06-08, pp. 593-597. IEEE Press, October 2008. ISBN : 978-1-4244-2020-9.
- [141] H.Tabout, Y.Chahir, A.Souissi and A.Sbihi, "An interactive machine for algae image retrieval", IEEE international conference on Computer Science and Information Technology ICCSIT 2009, Vol.4, pp 176-180, Beijing, China, ISBN 9781424432608.
- [142] H.Tabout, Y.Chahir, A.Souissi and A.Sbihi, "A diffusion approach for interactive image retrieval", Accepted, Studia Informatica universalis, Hermann Editions.
- [143] Alireza Tavakoli Targhi, Mårten Björkman, Eric Hayman and Jan-Olof Eklundh : Real-Time Texture Detection Using the LU-Transform. In the Workshop on Computation Intensive Methods for Computer Vision, in conjunction with ECCV 2006.
- [144] H.Tamura, S.Mori and T.Yamawaki. "Textural features corresponding to visual perception". IEEE Transactions on systems, Man and Cybernetics,8(6) : 460-473, 1978.
- [145] Tong, S., and Koller, D. (2000). Support vector machine active learning with applications to text classification. International Conference on Machine Learning, 999-1006.

- [146] Tong, S., and Koller, D. (2001). Support vector machine active learning with application to text classification. *Journal of Machine Learning Research*, 2 :45-66.
- [147] TORRALBA, A. et OLIVA, A "Semantic organization of scenes using discriminant structural templates». *IEEE international conference on Computer Vision (ICCV)*, Volume 2, pages 1253-1258.1999.
- [148] Vapnik, V. N. *Statistical learning theory*. Wiley-Interscience, New York.1999.
- [149] N. Vasconcelos and M. Kunt. Content-based retrieval from image databases : current solutions and future directions. In *International Conference in Image Processing (ICIP'01)*, volume 3, pages 6-9, Thessaloniki, Greece, October 2001
- [150] Veropoulos, K. Controlling the sensivity of support vector machines. *International Joint Conference on Artificial Intelligence (IJCAI)*. Stockholm, Sweden, 1999.
- [151] Base de Wang . [http ://wang.ist.psu.edu/](http://wang.ist.psu.edu/).
- [152] J.Z. Wang, J.Li and G. Weiderhold. Simplicity : Semantics-sensitive integrated matching for picture libraries. *IEEE Trans. Pattern Anal. Mach. Intell.*,23(9) :947-963.2001
- [153] Changhu Wang et al. "Graph-based Multiple-Instance Learning for Object-based Image Retrieval". *International Multimedia Conference*, pp 156-163. 2008.
- [154] J.S.Weszka, C.R. Dyer and A.Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transaction on systems, Man and Cybernetics*, 6(4) :269-285,1976.
- [155] [http ://www.wikipedia.org/](http://www.wikipedia.org/)
- [156] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering : theory and its application to image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15 (11) : 1101-1113, 1993.
- [157] C.C Yang Content-based image retrieval : a comparison between query by example and image browsing map approaches. February 2004.
- [158] C. Yang, M. Dong, F. Fotouhi, "Region based image annotation through multiple instance learning", *ACM Conference on Multimedia (MM'05)*, Singapore, pp. : 435-438, Nov., 2005
- [159] Changbo Yang and Ming Dong, "Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning". *Computer Vision and Pattern Recognition (CVPR'06)*. Pp 2057-2063. 2006
- [160] T. Zaharia ET F. Preteux "Comparative study for 3D and 2D/3D Shape descriptors“. *Research Report Munchen, Germany*.2004.
- [161] L. Zhang, F. Lin, and B. Zhang. Support Vector Machine for Learning Image Retrieval. In *IEEE International Conference on Image Processing*, Thessaloniki, Grèce, 2001.
- [162] Dengyong Zhou et al. "Learning with Local and Global Consistency". *Advances in Neural Information Processing Systems 16*, pp 321-328. 2004

- [163] Zhi-Hua Zhou and Jun-Ming Xu, "On the Relation between Multi-Instance Learning and Semi-Supervised Learning". IMC, pp. 1167-1174. 2007.
- [164] Zhu, X. "Semi-Supervised Learning Literature Survey". Computer Sciences Technical Report 1530, University of Wisconsin-Madison, 2005.

Annexe A

π^0 FORMULAE

PUBLICATIONS

- [1] H.Tabout, Y.Chahir, A.Souissi and A.Sbihi, "A diffusion approach for interactive image retrieval", *STUDIA INFORMATICA UNIVERSALIS*, Hermann Editions, VOL.8, pp. 111-127, 2010. ISBN 978 2 7056 80299.
- [2] H.Tabout, Y.Chahir, A.Souissi and A.Sbihi, "An interactive machine for algae image retrieval", *IEEE international conference on Computer Science and Information Technology ICCSIT 2009*, Vol.4, pp 176-180, Beijing, China, ISBN 9781424432608.
- [3] A.Souissi, H.Tabout and A.Sbihi, "MIR system for mobile information retrieval by image querying" , *International Journal of Computer Science and Network* VOL.8 No.4, pp. 65-73, April 2008.ISSN : 1738-7906.
- [4] H.Tabout, Y.Chahir, Abdelmoghith Souissi and Abderrahmane Sbihi, "Random walks based segmentation approach for image retrieval", *IEEE International Conference on Pervasive Computing and Applications - ICPCA* , Alexandria, Egypt, October 06-08, pp. 593-597. IEEE Press, October 2008. ISBN : 978-1-4244-2020-9.
- [5] A.Souissi, H.Tabout and A.Sbihi, "Système MMSINFO destiné pour la recherche d'information par des requêtes image dans le réseau mobile 2G et 3G", *9th African Conference on Research in Computer Science - CARI*, Rabat, Morocco, October 27-30, pp161-168. October 2008. ISBN : 2-7261-1299-4.
- [6] H. Tabout, A. Souissi, A. Sbihi "Region-Based Automated Relevance Feedback in Algae Image Retrieval", *International Journal of Computer Science and Network Security*, VOL.7 No.12 pp. 124-131, December 2007.ISSN : 1738-7906.
- [7] A. Souissi, H.TABOUT, A. Sbihi "Système de recherche et de navigation dans une base d'arabesques par descripteurs locaux". *VEILLE STRATEGIQUE SCIENTIFIQUE & TECHNOLOGIQUE (VSST'2007)*. octobre 2007. MARRAKECH, Morocco.
- [8] A. Souissi, H.TABOUT, A. Sbihi "GSMGUIDE application GSM destiné à la recherche de l'information" *Workshop sur les technologies de l'information et de la communication (WOTIC'07)*. 5-7 Juillet 2007. RABAT, Morocco.
- [9] H.Tabout, A. sbihi "A method for algae image indexing and retrieval based on neural network" *Information and communication technologies international symposium (ICTIS)* ; 3-5 April 2007. FES, Morocco.
- [10] H. Tabout, A. Souissi, A. Sbihi "A TREE STRUCTURED BASED ALGAE IMAGE RETRIEVAL" *Systèmes Electroniques-Informatiques & traitement de l'information (SEITI)* pp 15-18 ; Janvier 2007. Mohammedia ; Morocco.

- [11] H. Tabout, A. Souissi and A. Sbihi "A neural network based image retrieval using color and texture features". Maghrebian Conference on Software Engineering and Artificial Intelligence (MCSEAI) pp. 274-277 December 07-09, 2006, AGADIR, Morocco
- [12] H. Tabout, A. Sbihi "Algae image indexing approach using color and texture features" Conference en Systemes Intelligent : Theories ET Applications (SITA) pp. 88-92, Mohammadia, Morocco.
- [13] H. Tabout, A. Sbihi "Image Indexing Using Color and Texture". Workshop sur les technologies de l'information et de la communication (WOTIC'05). Juin 2005, Kenitra, Morocco.

CURRICULUM VITAE

Your Curriculum Vitae goes here