



**HAL**  
open science

# TRAITEMENT D'IMAGES PAR MODÈLES DISCRETS SUR GRAPHS ET CLASSIFICATION DE DONNÉES D'IMAGES PAR APPRENTISSAGE

Olivier Lezoray

► **To cite this version:**

Olivier Lezoray. TRAITEMENT D'IMAGES PAR MODÈLES DISCRETS SUR GRAPHS ET CLASSIFICATION DE DONNÉES D'IMAGES PAR APPRENTISSAGE. Traitement des images [eess.IV]. Université de Caen Basse-Normandie, 2007. tel-01100436

**HAL Id: tel-01100436**

**<https://hal.science/tel-01100436v1>**

Submitted on 6 Jan 2015

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

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

# UNIVERSITÉ de CAEN BASSE-NORMANDIE

U.F.R. Sciences

ÉCOLE DOCTORALE SIMEM

Mémoire présenté par

**Mr Olivier LÉZORAY**

et soutenu le

**22 Novembre 2007**

en vue de l'obtention de l'

**HABILITATION À DIRIGER DES RECHERCHES**

## **TRAITEMENT D'IMAGES PAR MODÈLES DISCRETS SUR GRAPHE ET CLASSIFICATION DE DONNÉES D'IMAGES PAR APPRENTISSAGE**

### MEMBRES du JURY

**Abderrahim Elmoataz, Professeur, Université de Caen Basse-Normandie ..... Directeur de HDR**

**Jean-Marc Chassery, Directeur de recherches CNRS, GIPSA-Lab Grenoble ..... Rapporteur**

**Jean-Michel Jolion, Professeur, Université de Lyon ..... Rapporteur**

**Louis Wehenkel, Professeur, Université de Liège, Belgique ..... Rapporteur**

**Marinette Revenu, Professeur, ENSICAEN ..... Examineur**

**Étienne Grandjean, Professeur, Université de Caen Basse Normandie ..... Examineur**



## Remerciements



Les travaux de recherches que je présente dans ce mémoire d'Habilitation à Diriger des Recherches ne sont pas le résultat de mon seul travail mais bien celui de collaborations avec différents doctorants et chercheurs.

Je tiens tout d'abord à remercier mes trois rapporteurs, Jean-Marc Chassery, Jean-Michel Jolion et Louis Wehenkel pour l'honneur qu'ils m'ont fait en acceptant la charge de rapporter ce mémoire. Depuis que j'ai débuté mes activités de recherches, la lecture de leurs travaux a fortement influencé les miens et c'est un plaisir de les voir à présent les évaluer.

J'adresse également mes remerciements à Marinette Revenu, pour avoir accepté de présider ce jury. Elle m'a soufflé l'idée de m'atteler à cette habilitation alors que cela semblait effectivement le bon moment. Je la remercie pour ses encouragements (depuis ma thèse).

Je remercie également Etienne Grandjean, futur directeur du GREYC, pour avoir accepté d'être membre de ce jury et m'avoir manifesté ses encouragements.

Je remercie enfin Abderrahim Elmoataz, mon directeur de recherche. Nos échanges vont bien au delà des travaux de recherches et du co-encadrement de doctorants. Sa culture scientifique et son dynamisme de tous les instants nous permettent d'avancer à grands pas.

Je tiens à remercier de tout coeur tous les étudiants avec qui j'ai eu l'honneur de travailler et en particulier les doctorants passés ou présents. Merci à Cyril Meurie, Gilles Lebrun, Sébastien Bougleux et Vinh Thong Ta. Ils sont la cheville ouvrière des laboratoires et ce mémoire n'existerait pas sans eux et sans la motivation qu'ils m'ont donnée ou qu'ils me donnent tout au long de leur thèse.

Je remercie également tous les membres des équipes de recherches normandes (le léopard) auxquelles j'ai pu être attaché : mon ancienne équipe Vision et Analyse d'Images et mon nouveau laboratoire, le GREYC. Je ne saurais les citer tous, mais ils ont tous su faciliter mon intégration. Une mention spéciale à l'attention de Régis Clouard pour la bibliothèque Pandore.

Je remercie tous mes collègues sur l'antenne universitaire de Saint-Lô (la licorne), qu'ils soient du département SRC ou GTE. Parmi tous ceux et celles-ci, on retrouve Christophe Charrier (ours de Gaspésie Nantaise), Marie-Pierre Besnard (la com, c'est un métier), Philippe Hervieu (VW man), Jérôme Clouet (Thanks for your proof reading), Athony Paul (notre sysadmin), etc.

Enfin, je dédie ce mémoire à ma famille. Merci à mes parents et frères (les racines, le triskell) pour leurs encouragements et spécialement à mon père pour le goût de la culture scientifique qu'il a su donner à ses enfants. Merci à ma femme Sophie pour sa patience envers son mari qui travaille beaucoup et qui a passé ses vacances à rédiger ce manuscrit. Merci également à mes deux enfants Manon et Quentin, la première pour avoir prêté son image à colorer et le second pour avoir laissé son papa travailler alors qu'il aurait préféré jouer avec. La présence de tous trois à mes côtés m'importe plus que tout.





---

---

# TABLE DES MATIÈRES

---

<b>1</b>	<b>Introduction générale</b>	<b>1</b>
<b>2</b>	<b>Traitement d'images par modèles discrets sur graphes</b>	<b>3</b>
2.1	Graphes et traitement d'images . . . . .	4
2.1.1	Préliminaires sur les graphes . . . . .	4
2.1.2	Représentations par graphes . . . . .	5
2.2	Ordres de données vectorielles pour le traitement d'images . . . . .	7
2.2.1	Ordres de données vectorielles . . . . .	8
2.2.2	Ordres et filtrage . . . . .	9
2.2.3	Chemin Hamiltonien dans un graphe . . . . .	12
2.2.4	Marches aléatoires . . . . .	18
2.2.5	Conclusion et perspectives . . . . .	26
2.3	Régularisation discrète sur graphes pondérés . . . . .	28
2.3.1	Introduction et problématique . . . . .	28
2.3.2	Géométrie différentielle sur graphes . . . . .	29
2.3.3	Régularisation discrète . . . . .	31
2.3.4	Liens vers d'autres approches . . . . .	34
2.3.5	Applications . . . . .	35
2.3.6	Conclusion et perspectives . . . . .	39
2.4	Hierarchies de partitions . . . . .	42
2.4.1	Introduction et Définitions . . . . .	42
2.4.2	Un critère connectif : les zones homogènes . . . . .	43
2.4.3	Diffusion-Fusion . . . . .	46
2.4.4	Coalescence-Fusion . . . . .	51
2.4.5	Conclusion et perspectives . . . . .	55
<b>3</b>	<b>Classification de données d'images par apprentissage</b>	<b>57</b>
3.1	Principes de la classification . . . . .	58

3.1.1	Généralités . . . . .	58
3.1.2	Apprentissage supervisé : concepts . . . . .	59
3.2	Contexte des travaux . . . . .	61
3.3	Fusion de décisions de classificateurs . . . . .	62
3.3.1	Évaluation de classifications de pixels . . . . .	63
3.3.2	Schémas de classification . . . . .	64
3.3.3	Combinaison de classifications . . . . .	66
3.4	Construction de fonctions de décision parcimonieuses . . . . .	70
3.4.1	Sélection de modèle et parcimonie . . . . .	70
3.4.2	Construction de fonctions de décision simplifiées . . . . .	71
3.4.3	Application à la classification de données d'images . . . . .	77
3.5	Schémas multi-classes de combinaison de classificateurs binaires . . . . .	84
3.5.1	Schémas de décomposition . . . . .	85
3.5.2	Schémas de décodage . . . . .	87
3.5.3	Propositions de schémas de décodage . . . . .	90
3.6	Conclusions et perspectives . . . . .	95
<b>4</b>	<b>Conclusion générale et perspectives</b>	<b>97</b>
4.1	Synthèse des contributions . . . . .	97
4.2	Perspectives et Projets . . . . .	99
<b>5</b>	<b>Bibliographie</b>	<b>103</b>
<b>A</b>	<b>Curriculum Vitae</b>	<b>117</b>
A.1	Diplômes . . . . .	117
A.2	Carrière . . . . .	117
A.3	Responsabilités pédagogiques et administratives . . . . .	118
A.3.1	Enseignements et responsabilités pédagogiques . . . . .	118
A.3.2	Responsabilités administratives . . . . .	120
A.4	Activités liées à la recherche . . . . .	121
A.4.1	Thèmes de recherche . . . . .	121
A.4.2	Activités de recherche . . . . .	121
A.4.3	Co-encadrement de thèses et de DEA . . . . .	122
A.5	Animation de la recherche . . . . .	123
A.5.1	Participation à des réseaux . . . . .	123

---

A.5.2	Activités éditoriales . . . . .	123
A.5.3	Membre de comités de lecture, de programme . . . . .	123
A.5.4	Organisation de Journées, de Conférences . . . . .	124
A.6	Projets de recherche et valorisation . . . . .	124
A.6.1	Projets nationaux ANR . . . . .	124
A.6.2	Projets régionaux du Comité Régional Imagerie et Technologies de l'In- formation et de la Communication (CRITIC) . . . . .	124
A.6.3	Collaborations Industrielles . . . . .	125
A.6.4	Consultances . . . . .	125
<b>B</b>	<b>Publications</b>	<b>127</b>
	<b>Résumé</b>	<b>135</b>



## Introduction générale

---



If we knew what we were doing, it wouldn't be called research, would it ?

ALBERT EINSTEIN

**A**PRÈS des études en Informatique à l'Université de Caen Basse-Normandie, j'ai débuté mes activités de recherches en 1995 à Caen au GREYC (Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen) dans le cadre du DEA d'Intelligence Artificielle et Algorithmique, sous la direction de Hubert Cardot. Après une année de service militaire, j'ai continué en thèse de 1996 à 2000 au sein du groupe VAI (Vision et Analyse d'Images) du LUSAC (Laboratoire Universitaire des Sciences Appliquées de Cherbourg) sous la direction de Marinette Revenu du GREYC. Cette thèse était financée par la ligue contre le cancer et s'est déroulée à Cherbourg en étroite collaboration avec le laboratoire d'anatomie et de cytologie pathologiques de l'hôpital Pasteur de Cherbourg qui s'appelle désormais le Centre Hospitalier Public du Cotentin. L'objectif visé de cette thèse était la conception d'un système de vision par ordinateur pour la reconnaissance de cellules dans des images microscopiques couleur de cytologie des séreuses. Cet objectif fût atteint et nous avons conçu un prototype nommé ARCTIC (Aide à la Recherche en Cytologie par le Tri Informatique Cellulaire). J'ai pu concevoir un schéma de traitement d'image approprié à la segmentation d'images microscopiques, définir des attributs de couleur et de texture cellulaires et enfin catégoriser les objets cellulaires à l'aide d'une architecture de réseaux de neurones optimisée.

Après une année d'ATER (1999 – 2000) au sein du département d'Informatique de l'Université de Caen Basse-Normandie, j'ai été nommé Maître de Conférences en Informatique à l'IUT Cherbourg Manche, Antenne de Saint-Lô, en Septembre 2000. Je suis resté rattaché à mon laboratoire de thèse, le LUSAC. Ma nomination coïncidait avec l'ouverture d'un nouveau département de l'IUT Cherbourg Manche, le département SRC (Services et Réseaux de Communication) délocalisé sur le site universitaire de Saint-Lô. Dès mon arrivée, nous avons eu à mettre en place la structure de ce nouveau département qui ouvrait avec seulement quatre titulaires. Avec mes collègues de l'époque, nous avons donc mis en place ce nouveau département. Ces premières années furent donc très prenantes car j'ai dû assurer la tâche de directeur des études et monter de toutes pièces tous les cours en Informatique pour les premières années et également m'occuper avec mes collègues d'énormément de tâches annexes afférentes à l'ouverture d'un nouveau département d'IUT. Ce contexte de site délocalisé éloigné de mon laboratoire de recherche fût un peu particulier, mais la qualité des infrastructures que j'ai pu trouver à Saint-Lô m'a permis de continuer mes activités de recherche, même à distance.

A mon arrivée à Saint-Lô, je restais rattaché au LUSAC dans l'équipe Vision et Analyse d'Images. Avec l'arrivée de nouveaux collègues Maîtres de Conférences sur le site de Saint-Lô, nous avons pu mettre en place un groupe de recherche et l'obtention d'une salle dédiée à nos

activités en 2004. Ceci a facilité grandement nos activités notamment pour l'accueil de deux des doctorants que j'ai encadrés (Cyril Meurie et Gilles Lebrun). En 2007, le LUSAC s'étant donné des orientations en termes de recherches qui ne correspondaient pas à notre groupe VAI, nous avons demandé notre rattachement au laboratoire GREYC UMR CNRS 6072 et il sera effectif à partir de 2008.

Les travaux que je présente dans ce mémoire d'habilitation à diriger des recherches correspondent à des recherches effectuées depuis mon arrivée à Saint-Lô, en septembre 2000. Globalement, ceci est le fruit de travaux que j'ai réalisés avec des étudiants de DEA, des doctorants, ainsi qu'avec des collègues chercheurs. La recherche est avant tout un travail d'équipe qui nécessite énormément de communication, surtout lorsque la localisation géographique demande de la mobilité.

Ce mémoire est composé de deux parties principales qui peuvent se lire indépendamment l'une de l'autre. La première partie concerne la description de mes activités de recherche pour le traitement d'images à partir de modèles discrets sur graphes. La seconde partie concerne la description de mes activités de recherche pour la classification de données d'images par apprentissage. Dès ma thèse, j'ai mené de front deux activités distinctes de recherche que sont le traitement d'images et la fouille de données avec comme point commun les images pour support.

Le premier chapitre présente plusieurs voies que nous avons explorées pour un traitement d'images à partir de modèles discrets sur graphes. Travaillant depuis ma thèse sur la segmentation d'images couleur, j'ai été rapidement amené à considérer des graphes d'adjacence de régions. Les graphes étant naturellement adaptés aux problèmes de vision par ordinateur, je me suis alors intéressé à diverses problématiques du traitement d'images (filtrage, simplification, segmentation, etc.) selon ce point de vue. Ce chapitre présente les voies que nous avons explorées.

Le second chapitre présente les travaux que nous avons menés pour la conception d'apprenants artificiels pour la classification de données issues d'images. Dès ma thèse, je me suis attaché à ne pas valider uniquement les algorithmes d'apprentissage que nous avons proposés sur des données issues d'images, mais bien sur des bases de données de référence en apprentissage. En conséquence, la portée de nos contributions concerne plus largement la thématique de l'apprentissage artificiel. Ce chapitre présente les voies que nous avons explorées.

Le mémoire se conclut par une synthèse des principaux résultats obtenus et montre les perspectives de recherche ouvertes par ces travaux.

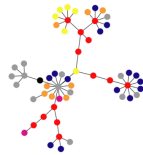
Une première annexe est consacrée à la présentation de mon parcours en détaillant mes responsabilités pédagogiques, administratives, de recherche ainsi que mes projets en cours. Une seconde annexe présente la liste de mes publications<sup>1</sup>. Une troisième annexe sous la forme d'un document séparé contient quelques publications choisies<sup>2</sup>.

---

<sup>1</sup>Dans ce mémoire, les références à mes publications seront faites à l'aide de chiffres et les références à d'autres travaux seront faites à l'aide de la notation de l'American Mathematical Society (nom de l'auteur et année).

<sup>2</sup>Toutes mes publications sont également disponibles à <http://www.info.unicaen.fr/~lezoray/Publications.php>

## Traitement d'images par modèles discrets sur graphes



Graph Theory Uncovers the Roots of Perfection.

DANA MACKENZIE

### Sommaire

<b>2.1 Graphes et traitement d'images</b> . . . . .	<b>4</b>
<b>2.2 Ordres de données vectorielles pour le traitement d'images</b> . . . . .	<b>7</b>
<b>2.3 Régularisation discrète sur graphes pondérés</b> . . . . .	<b>28</b>
<b>2.4 Hiérarchies de partitions</b> . . . . .	<b>42</b>

**C**E chapitre présente les recherches que j'ai menées pour proposer des formulations de traitement d'images par modèles discrets sur graphes. Ces travaux ont été réalisés essentiellement avec Abderrahim Elmoataz soit dans le cadre de thèses que nous avons co-encadrée (Cyril Meurie [MEURIE05B], pour les ordres de données vectorielles et les hiérarchies de partitions) où que nous co-encadrons actuellement (Vinh Thong Ta [TAON], pour la régularisation discrète). Dans le cadre de la régularisation discrète, j'ai également collaboré avec Sébastien Bougleux [BOUGLEON], doctorant d'Abderrahim Elmoataz.

Pour clarifier l'exposé de ce chapitre, en voici un plan rapide :

- Dans la Section 2.1, sont rappelées des notions essentielles de la théorie des graphes. Nous nous attardons ensuite sur l'intérêt des représentations à base de graphes pour le traitement d'images. Dans ce cadre, un graphe peut être associé à n'importe quelle représentation de données d'images pour peu que l'on se munisse d'une mesure de distance ou de similarité.
- Dans la Section 2.2, nous nous attachons à la définition d'ordres de données vectorielles. Nous proposons alors deux méthodes ; l'une basée sur la recherche d'un chemin Hamiltonien dans un graphe et l'autre basée sur l'exploitation de la réduction de dimension via les marches aléatoires sur graphes. Certains éléments sont détaillés car ils n'ont pas été encore publiés ou sont en cours de publication [18, 55].
- Dans la Section 2.3, un cadre général pour la régularisation basée sur le p-Laplacien discret pour le traitement d'images et de données représentées par des graphes pondérés de topologies arbitraires est présenté. Ce cadre unifie la régularisation locale ou non locale sur les images et l'étend naturellement au traitement de données discrètes sur graphes. Certains éléments sont détaillés car ils sont en cours de publication [19, 58, 72–74].
- Dans la Section 2.4, nous proposons plusieurs méthodes pour la construction de hiérarchies de partitions dans le cadre d'une approche morphologique, d'une approche combinant approches espace-échelle par diffusion sur graphe et fusion de régions ainsi que d'une



approche hybride par coalescence colorimétrique et fusion de régions avec détermination automatique du paramètre d'échelle.

## 2.1 Graphes et traitement d'images

### 2.1.1 Préliminaires sur les graphes

Nous rappelons ici des notions de base de la théorie des graphes (voir dans [DIESTE05] pour plus de détails). Un graphe  $\mathcal{G}$  est un couple  $\mathcal{G} = (V, E)$  où  $V$  est un ensemble fini de noeuds (de taille  $|V|$ ) et  $E$  un ensemble d'arêtes (de taille  $|E|$ ) :  $E \subseteq V \times V$ . Deux noeuds  $u$  et  $v$  dans un graphe sont adjacents si l'arête  $(u, v)$  appartient à  $E$ , les deux noeuds sont alors appelés des noeuds voisins.

Le degré (simple)  $\delta(v)$  d'un noeud  $v$  est le nombre d'arêtes incidentes en ce noeud.  $\delta : V \rightarrow \mathbb{N}^+$  est défini par  $\delta(v) = |u \sim v|$  où  $|\cdot|$  désigne le cardinal d'un ensemble. La relation  $u \sim v$  désigne l'ensemble des noeuds  $u$  adjacents au noeud  $v$  via les arêtes  $(u, v) \in E$  :  $u$  est un voisin de  $v$ . Si un noeud a un degré simple de un, il est appelé une feuille.

Un chemin  $C$  est un ensemble de noeuds  $C = (v_1, v_2, \dots, v_k)$  tel qu'il existe une arête pour tout couple de noeuds successifs du chemin :  $\forall i \in [1, k[$ , l'arête  $(v_i, v_{i+1}) \in E$ . La longueur d'un chemin correspond à son nombre d'arêtes. Un chemin est simple s'il ne passe qu'une et une seule fois par toute arête. Un chemin est Hamiltonien s'il passe par tous les noeuds du graphe et ceci une et une seule fois (ce problème est NP-complet). Un graphe complet est un graphe où chaque paire de noeuds est reliée par une arête. Un graphe complet contenant  $n$  noeuds a donc exactement  $n(n-1)/2$  arêtes et le degré simple de chaque noeud est  $(n-1)$ .

Un graphe est connexe si, pour chaque paire de noeuds  $u$  et  $v$ , il existe un chemin tel que  $v_1 = u$  et  $v_k = v$ . Un graphe est non-orienté quand l'ensemble des arêtes est symétrique, c'est à dire que pour chaque arête  $(u, v) \in E$ , nous avons également  $(v, u) \in E$ . Nous considérerons toujours des graphes simples où au maximum une arête relie deux noeuds, ces graphes seront également connexes et non orientés [DIESTE05].

Nous pouvons à présent définir l'espace des fonctions sur graphes. Soit  $\mathcal{H}(V)$  l'espace de Hilbert des fonctions sur les noeuds, dans lequel chaque fonction  $f : V \rightarrow \mathbb{R}^p$  associe un vecteur  $f(v)$  à un noeud  $v$ . L'espace des fonctions  $\mathcal{H}(V)$  peut être muni du produit scalaire usuel  $\langle f, g \rangle_{\mathcal{H}(V)} = \sum_{v \in V} f(v)g(v)$  où  $f$  et  $g$  sont deux fonctions de  $\mathcal{H}(V)$ . La norme d'une fonction  $f$  obtenue à partir du produit scalaire est  $\|f\| = \sqrt{\langle f, f \rangle}$ . De la même manière, nous pouvons définir  $\mathcal{H}(E)$  comme l'espace de Hilbert des fonctions sur les arêtes, dans lequel chaque fonction  $g : E \rightarrow \mathbb{R}^+$  associe une valeur réelle à une arête  $e$ . Cet espace de fonctions peut être muni du produit scalaire usuel  $\langle h, l \rangle_{\mathcal{H}(E)} = \sum_{(u,v) \in E} h(u, v)l(u, v)$  où  $h$  et  $l$  sont deux fonctions de  $\mathcal{H}(E)$ .

Un graphe sera dit pondéré si une fonction de poids  $w \in \mathcal{H}(E)$  lui est associé. Cette fonction  $w$  satisfait en outre  $w(u, v) > 0$  si  $(u, v) \in E$ ,  $w(u, v) = 0$  si  $(u, v) \notin E$  et  $w(u, v) = w(v, u)$  pour toutes les arêtes de  $E$  puisque nous ne considérerons que des graphes non orientés. Un graphe peut être pondéré par une distance ou une similarité entre noeuds. Si l'on considère la

matrice d'adjacence associé à un graphe, on notera  $W$  cette matrice si les poids sont des distances et  $S$  si les poids sont des similarités. Une fonction de distance classique est la norme  $L_2$  :  $w_{uv} = w(u, v) = \|f(u) - f(v)\|_2$ . Deux exemples classiques de fonctions de similarité sont :  $w(u, v) = \exp\left(-\frac{\|f(u)-f(v)\|_2}{2\sigma^2}\right)$  et  $w(u, v) = \frac{1}{\|f(u)-f(v)\|_2}$ .

Le degré d'un noeud  $v \in V$  peut être défini par  $d_i = d(v_i) = \sum_{u \sim v_i} w(u, v_i)$ . La matrice des degrés  $D$  est définie comme la matrice diagonale avec les degrés  $d_i$  des noeuds  $v_i \in V$  sur la diagonale. A partir d'un sous ensemble  $A \in V$ , nous notons  $\bar{A}$  son complément. La taille de cet ensemble peut être mesurée en considérant soit  $|A|$ , son nombre de noeuds, ou bien  $vol(A) = \sum_{v_i \in A} d(v_i)$ , en terme de poids de ses arêtes.

Un arbre est un graphe simple connexe acyclique. Un arbre de recouvrement d'un graphe connexe non orienté  $\mathcal{G}$  est un arbre composé de tous les noeuds du graphe et de quelques arêtes de  $\mathcal{G}$ . Un arbre de recouvrement minimum (un Minimum Spanning Tree (MST)) est un arbre de recouvrement dont la somme des poids des arêtes est inférieure à celle de n'importe quel autre arbre de recouvrement. Ainsi, un arbre de recouvrement minimum  $\mathcal{T}(\mathcal{G})$  d'un graphe  $\mathcal{G}$  est un graphe connexe pondéré  $\mathcal{T}(\mathcal{G}) = (V', E')$  où la somme des poids des arêtes  $\sum_{(u,v) \in E'} w(u, v)$  est minimale. Pour un graphe  $\mathcal{G}$  de  $n$  noeuds, son arbre de recouvrement minimum (ou MST)  $\mathcal{T}(\mathcal{G})$  a exactement  $(n - 1)$  arêtes. Le MST peut être calculé efficacement en  $O(|E| \log |V|)$  par l'algorithme de Prim avec des structures de données appropriées.

## 2.1.2 Représentations par graphes

Les graphes sont la structure de données par excellence en informatique et elles sont tout naturellement adaptées à la représentation de données quelconques en modélisant les propriétés de voisinage entre ces données. En traitement d'images, on utilise habituellement le support de l'image comme une grille où les relations de connexité spatiale entre les pixels déterminent le voisinage considéré. En conséquence, sur une grille 2D, le graphe de voisinage associé à une image correspond par exemple à l'utilisation d'une 4 ou d'une 8 connexité au niveau pixel (Figures 2.1(a) et 2.1(b)). Une grille est une tessellation régulière dont les éléments sont des carrés, mais il existe d'autres tessellations constituées de polygones dont la plus célèbre est la tessellation de Voronoï (Figure 2.1(c)). A cette tessellation est associée également un graphe de voisinage nommé le graphe de Delaunay (Figure 2.1(d)).

Plus spécifiquement, lorsque que l'on segmente une image en régions, on peut associer à la carte de segmentation un graphe de voisinage nommé le graphe d'adjacence de régions où chaque noeud représente une région et chaque arête représente une relation de voisinage entre ces régions (Figure 2.2).

De manière similaire, à un ensemble quelconque de données, nous pouvons associer un graphe de proximité afin de modéliser les relations spatiales de voisinage entre les données. On citera par exemple [VONLUX06] le graphe complet, le graphe des  $k$  plus proches voisins, le graphe de  $\varepsilon$ -voisinage, l'arbre de recouvrement minimum. La figure 2.3 présente quelques-uns de ces graphes de similarité sur une base de données d'images.

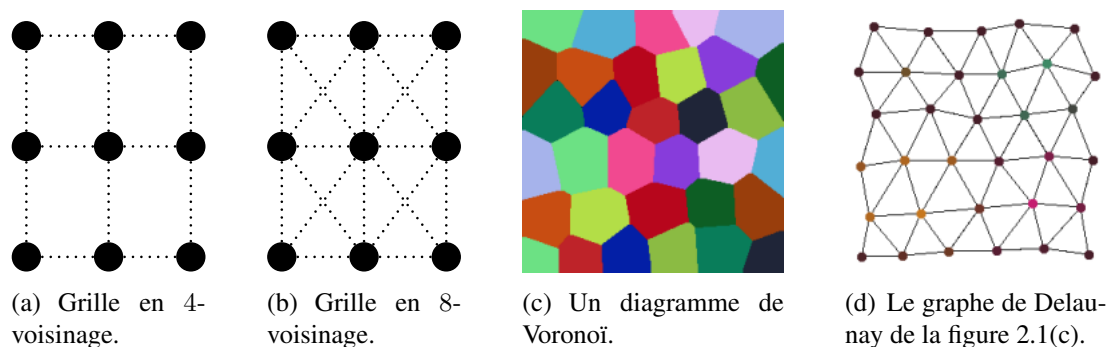


FIG. 2.1 – Exemples de graphes de type grille ou de Delaunay.

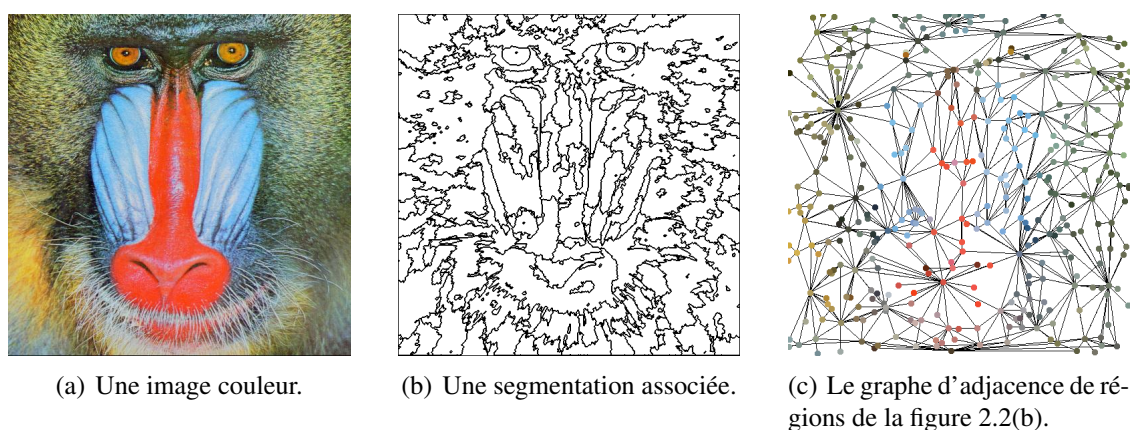


FIG. 2.2 – Exemple de graphe d'adjacence de régions.

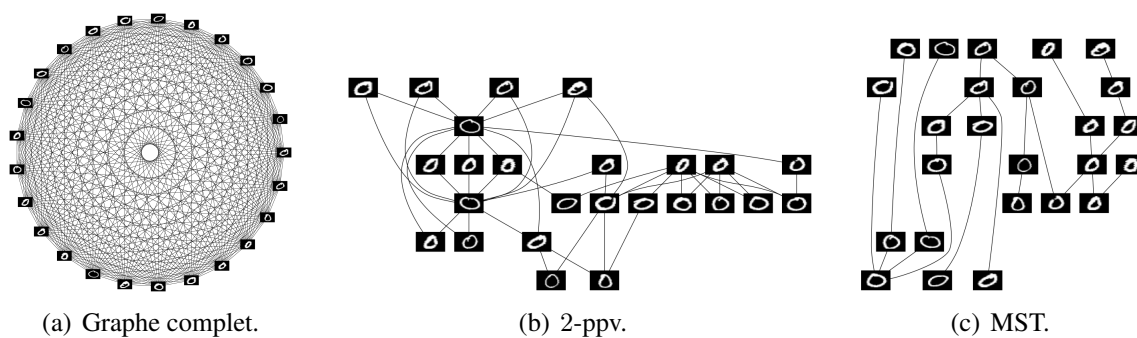


FIG. 2.3 – Exemples de graphes de proximité.

Comme nous venons de le voir, n'importe quel domaine discret peut être modélisé par un graphe. En conséquence, les graphes sont bien adaptés à la résolution de problèmes de vision par ordinateur. Les principaux graphes utilisés pour le traitement de données d'images sont donc les suivants :

- Les graphes grilles [CHAN01B] qui sont des structures qui correspondent naturellement aux images : les noeuds sont les pixels et les arêtes les relations de connexité entre pixels.
- Les graphes d'adjacence de régions [TRÉ00] qui permettent de décrire la structure d'une image en terme de régions : les noeuds représentent les régions et les arêtes les relations de voisinage spatial entre les régions.
- Les graphes de proximité (graphe complet, graphe des  $k$  plus proches voisins, arbre de recouvrement minimum, graphe de  $\varepsilon$ -voisinage) [VONLUX06] qui décrivent les relations spatiales entre des données de dimensions quelconques. Chaque noeud correspond à un vecteur dans un espace de description et les arêtes correspondent à la proximité spatiale entre les vecteurs associés aux noeuds. La topologie du graphe dépend d'un critère donné et certains de ces graphes peuvent être non connexes.

Finalement, un graphe peut être associé à n'importe quelle représentation de données d'images pour peu que l'on se munisse d'une mesure de distance ou de similarité. Dans ce cadre, traiter des données d'images revient à traiter des graphes dont les propriétés attachées aux noeuds et aux arêtes dépendent du problème considéré. La topologie du graphe dépendra donc du problème considéré : des grilles pour la restauration d'image, des graphes d'adjacence de régions pour la segmentation, des graphes de proximité pour la catégorisation de données. Nous montrerons dans la suite qu'à partir de telles représentations, nous pouvons résoudre des problèmes très différents à l'aide de formalismes exprimés directement sur graphes de topologies parfois arbitraires.

## 2.2 Ordres de données vectorielles pour le traitement d'images

☞ **Mes publications associées :** [18, 45, 55, 69]

Beaucoup de filtres non linéaires, tels que par exemple les filtres médians et morphologiques, nécessitent d'ordonner les pixels de la fenêtre d'analyse. Dès qu'un pixel est décrit par des données multivariées, se pose le problème de l'extension des filtres scalaires à des données vectorielles. Comme il n'existe pas d'ordre naturel pour les données vectorielles, cette extension n'est pas triviale. La définition d'un ordre de données vectorielles présente donc un fort intérêt pour le traitement d'images multivariées. Les images couleur étant devenues communes, un ordre de vecteurs couleur est essentiel pour le traitement d'images couleur [LUKAC06A, LUKAC05] car cette notion d'ordre peut être utilisée à différentes fins : filtrage, détection de contour, interpolation spatiale, comparaison d'histogrammes, création de palettes, etc. [LUKAC06A]. Nous nous sommes intéressés à cette problématique dans le cadre du filtrage (médian ou morphologique) d'images couleur, mais les ordres que nous avons proposés peuvent s'appliquer à n'importe laquelle des précédentes problématiques. Ce problème d'ordre de données vectorielles ne concerne pas uniquement la communauté du traitement d'images mais également celle d'apprentissage [AGARWA06, ZHOU03], le point commun étant le traitement des variétés (« manifolds ») [BELKIN05] notamment pour la réduction de dimension [BELKIN03].

Après un survol des techniques usuelles d'ordre de données vectorielles, nous présentons leurs avantages et défauts dans un but de filtrage d'image (médian ou morphologique). Nous proposons ensuite deux ordres de données vectorielles qui pallient les défauts des ordres de la littérature. Le premier ordre proposé consiste à construire un chemin Hamiltonien sur un graphe complet construit sur la fenêtre d'analyse du filtrage. Le second ordre est basé sur une réduction de dimension par marches aléatoires sur graphe.

## 2.2.1 Ordres de données vectorielles

Pour traiter des données multivariées telles que des vecteurs couleur (qui ne sont pas forcément de dimension 3 comme nous le verrons plus loin), les trois principaux ordres de données vectorielles sont le pré-ordre, l'ordre partiel et l'ordre total. Avant la description de ces ordres, nous rappelons les définitions essentielles des relations utiles pour caractériser un ordre de données vectorielles [HANBUR02, VERTAN96, 45].

**Definition 1.** Soit  $R$  une relation binaire définie sur un ensemble donné  $A$ .

- $R$  est réflexive si  $\forall \mathbf{x} \in A, \mathbf{x}R\mathbf{x}$ ,
- $R$  est transitive si  $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in A, \mathbf{x}R\mathbf{y}$  et  $\mathbf{y}R\mathbf{z} \Rightarrow \mathbf{x}R\mathbf{z}$ ,
- $R$  est anti-symétrique si  $\forall \mathbf{x}, \mathbf{y} \in A, \mathbf{x}R\mathbf{y}$  et  $\mathbf{y}R\mathbf{x} \Rightarrow \mathbf{x} = \mathbf{y}$ .

**Definition 2.** Une relation binaire  $R$  sur un ensemble  $A$  est un pré-ordre si  $R$  est réflexive et transitive.

**Definition 3.** Une relation binaire  $R$  sur un ensemble  $A$  est un ordre partiel si  $R$  est réflexive, transitive et anti-symétrique.

**Definition 4.** Une relation binaire  $R$  sur un ensemble  $A$  est un ordre total si  $R$  est un ordre partiel et si  $\forall \mathbf{x}, \mathbf{y} \in A, \mathbf{x}R\mathbf{y}$  ou  $\mathbf{y}R\mathbf{x}$ .

Un ordre complet sur un ensemble  $A$  est un ordre où n'importe quelle paire de vecteurs peut être ordonnée (par exemple la relation binaire  $\leq$  sur  $\mathbb{R}$ ).

Une image multivariée peut être représentée par une projection  $\mathbb{Z}^l \rightarrow \mathbb{R}^p$  où  $l$  est la dimension de l'image et  $p$  le nombre de composantes de l'image multivariée. Soit  $F = \{\mathbf{x}_k \in \mathbb{Z}^l; k = 1, 2, \dots, N\}$  une fenêtre de filtrage de taille  $N$  centrée sur  $\mathbf{x}_1$ . A partir de  $F$ , nous obtenons un ensemble  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  de  $N$  vecteurs de dimension  $p$  :  $\mathbf{x}_i = \{\mathbf{x}_i^1, \mathbf{x}_i^2, \dots, \mathbf{x}_i^p\}$ ,  $\mathbf{x}_i \in \mathbb{R}^p$ . Une façon classique de définir une relation d'ordre entre des vecteurs est d'utiliser une transformation [GOUTS195]  $h$  de  $\mathbb{R}^p$  dans  $\mathbb{R}^q$  suivie de l'ordre naturel sur chaque dimension de l'espace  $\mathbb{R}^q$ . Avec  $h : \mathbb{R}^p \rightarrow \mathbb{R}^q$ , et  $\mathbf{x} \rightarrow h(\mathbf{x})$  alors  $\forall (\mathbf{x}_i, \mathbf{x}_j) \in \mathbb{R}^p \times \mathbb{R}^p, \mathbf{x}_i \leq \mathbf{x}_j \Leftrightarrow h(\mathbf{x}_i) \leq h(\mathbf{x}_j)$ . Quand  $h$  est bijective, cela correspond à définir une courbe de remplissage de l'espace qui passe une et une seule fois par tout point de l'espace  $\mathbb{R}^p$  et qui induit donc un ordre complet. A partir de ces définitions, nous pouvons définir une taxonomie [BARNET76] des différents types d'ordres de données vectorielles (i.e. multivariées). Les différents ordres que nous présentons sont l'ordre marginal, l'ordre réduit et l'ordre conditionnel. Nous rappelons ici leurs principes.

- Dans l'ordre marginal, les vecteurs sont ordonnés indépendamment dans chaque dimension ( $q = p, h = \text{Identité}$ ). Cet ordre est un ordre partiel et il est admis que ce type d'ordre n'est pas satisfaisant car il peut produire des nouveaux vecteurs qui n'appartiennent pas à l'ensemble initial [LAMBER02, LUKAC05].

- Dans l'ordre réduit,  $q = 1$  et  $h$  désigne une métrique qui différencie le type d'ordre réduit. Chaque vecteur est réduit à une valeur scalaire et les vecteurs sont triés selon ces scalaires. Une manière classique d'utiliser un ordre réduit, est de prendre  $h$  comme étant une fonction de distance cumulative. Ce type d'approche est par exemple utilisé dans les filtres médians couleur [LUKAC06A, LUKAC06B]. Cet ordre est un pré-ordre.
- Dans l'ordre conditionnel (ou lexicographique), les vecteurs sont ordonnés selon un ordre hiérarchique des composantes du vecteur. Nous avons donc  $q = 1$  et  $h : \mathbb{R}^p \rightarrow \mathbb{R}, \mathbf{x}_i \rightarrow \mathbf{x}_i^k$  où  $k$  est déterminé de manière adaptative. Pour deux vecteurs  $\mathbf{x}_i$  et  $\mathbf{x}_j$ , on définit :

$$\mathbf{x}_i \leq \mathbf{x}_j \begin{cases} \mathbf{x}_i^1 < \mathbf{x}_j^1, \text{ ou} \\ \mathbf{x}_i^1 = \mathbf{x}_j^1, \text{ et, } \mathbf{x}_i^2 < \mathbf{x}_j^2 \text{ ou } \dots \\ \mathbf{x}_i^1 = \mathbf{x}_j^1, \text{ et, } \mathbf{x}_i^2 = \mathbf{x}_j^2 \dots \mathbf{x}_i^p < \mathbf{x}_j^p \end{cases}$$

Cet ordre est un ordre total mais il introduit une forte dissymétrie entre les composantes des vecteurs.

Comme le traitement de données multivariées est présent dans de nombreux domaines de recherches (du traitement d'images couleur aux images multi-spectrales), un nombre important de différents ordres de données vectorielles est présent dans la littérature [ANGULO05B, ASTOLA90, CHANUS98, COMER99, HANBUR01, IWANOW99, KOPPEN99, LOUVER02, ORTIZ01, ORTIZ00, PETERS97, PITAS91, TALBOT98, TSALID02, VARDAV01]. La plupart de ces travaux reposent sur des ordres lexicographiques modifiés ; ce qui est naturel étant donné que l'ordre lexicographique est l'ordre total le plus communément utilisé. Cependant, un autre ordre total de données vectorielles, nommé l'ordre par entrelacement de bits, a été proposé par CHANUSSOT [CHANUS98]. L'ordre par entrelacement de bits est basé sur une représentation binaire de chaque composante du vecteur considéré  $\mathbf{x}$ . Si les  $p$  composantes de  $\mathbf{x}$  sont codées sur  $b$  bits, les  $p \cdot b$  bits sont entrelacés afin de construire un entier  $h(\mathbf{x})$  codé sur  $p \cdot b$  bits [CHANUS98, LAMBER00A]. La transformation considérée peut être écrite comme suit :

$$h(\mathbf{x}) = \sum_{k=1}^b \left\{ 2^{p \cdot (b-k)} \sum_{i=1}^p 2^{(p-i)} \mathbf{x}^{i \triangleright k} \right\}$$

où  $\mathbf{x}^{i \triangleright k}$  désigne le  $k^{\text{ème}}$  bit de la  $i^{\text{ème}}$  composante de  $\mathbf{x}$ . Par exemple, pour le cas des images couleur codées en *RGB*,  $p = 3$  et  $b = 8$ .

Nous pouvons également mentionner les travaux de REGAZONNI [REGAZZ97, STRING99] qui définit un ordre total à l'aide de courbes de remplissage de l'espace. Nous verrons dans la suite que, sur des données organisées sur un voisinage, un ordre total peut être défini par construction d'une courbe de remplissage de l'espace et que ceci est équivalent à définir un chemin Hamiltonien.

## 2.2.2 Ordres et filtrage

Un important champ d'application du traitement d'images multivariées est le filtrage [PITAS91, CHANUS98, LUKAC06A, PLATAN00]. Nous commençons par nous intéresser au lien qui existe entre la définition d'un ordre de données vectorielles et le filtrage (qu'il soit par exemple morphologique [RONSE90, GOUTSI95] ou médian [LUKAC05, CHANUS98]). En effet, les filtres morphologiques ou médian sont basés sur des ordres de données vectorielles spécifiques qui ont des avantages et désavantages. Nous les détaillons dans la suite.

### 2.2.2.1 Filtrés morphologiques vectoriels

La morphologie mathématique est une approche non linéaire du traitement d'images qui repose sur une structure fondamentale, le treillis complet  $\mathcal{L}$ . Un treillis complet est défini tel que [RONSE90] :

- Une relation d'ordre  $\leq$  est définie sur  $\mathcal{L}$ ,
- Pour tout sous ensemble fini  $\mathcal{K}$  de  $\mathcal{L}$ , existent un supremum  $\vee \mathcal{K}$  et un infimum  $\wedge \mathcal{K}$ .

Une approche marginale peut être évidemment utilisée, mais une approche purement vectorielle est préférable. En effet, avec une approche marginale, le supremum et l'infimum n'appartiennent pas toujours au treillis et des fausses couleurs peuvent apparaître [CHANUS98]. Ce défaut est à présent bien établi dans la littérature. Une autre contrainte pour la définition de filtres morphologiques est donc d'imposer le fait que le supremum et l'infimum d'un ensemble appartiennent à celui-ci. En conséquence, des ordres totaux sont généralement considérés pour l'utilisation de filtres morphologiques. Ainsi, les principaux ordres de données vectorielles en morphologie mathématique sont l'ordre lexicographique [HANBUR01, ANGULO05B] et l'ordre par entrelacement de bits [CHANUS98] qui sont des ordres totaux et qui remplissent toutes les contraintes du treillis complet  $\mathcal{L}$ .

Cependant, l'ordre lexicographique est très asymétrique. En conséquence, la plupart des décisions d'ordre de données vectorielles sont prises au niveau de la première composante ce qui implique d'attribuer une priorité aux composantes. Cela définit alors des opérateurs dont le comportement n'est pas homogène dans un espace vectoriel donné tel qu'un espace couleur. Ce choix de priorité entre les composantes étant difficile [ANGULO03, ORTIZ01, ANGULO05A], il peut être contourné, pour le cas des images couleur, en changeant de représentation et en considérant des espaces couleur perceptuels basés sur la luminance, la teinte et la saturation [ANGULO05B, LOUVER02]. Dans ce type d'espace couleur, il devient plus naturel (d'un point de vue de l'interprétation humaine) de choisir la priorité entre les composantes. Même si l'ordre lexicographique a des désavantages, c'est le plus communément utilisé et il a été étudié par de nombreux auteurs pour le filtrage morphologique [IWANOW99, LOUVER02, ORTIZ00, PETERS97, TALBOT98, TSALID02, VARDAV01].

L'ordre par entrelacement de bits permet de limiter l'asymétrie entre les composantes. Cependant, puisqu'il est basé sur un entrelacement de bits, il ne peut opérer que sur des données vectorielles entières, ce qui n'est pas le cas de beaucoup d'espaces couleur ni, plus généralement, de données multivariées. L'ordre par entrelacement de bits a donc été principalement conçu pour traiter des images couleur dans l'espace couleur *RGB*.

Une fois qu'un ordre de données vectorielles a été choisi, nous pouvons appliquer les deux principales opérations morphologiques, à savoir l'érosion  $\epsilon$  et la dilatation  $\delta$ . Indépendamment de l'ordre vectoriel considéré, l'érosion et la dilatation définies sur une fenêtre de filtrage  $F$  centrée en  $\mathbf{x}$  sont définies par :

$$\epsilon(\mathbf{x}) = \{\mathbf{y} : \mathbf{y} = \wedge F\} \text{ et } \delta(\mathbf{x}) = \{\mathbf{y} : \mathbf{y} = \vee F\}$$

où  $\vee$  et  $\wedge$  sont respectivement le supremum et l'infimum d'un ensemble. D'autres opérations morphologiques élémentaires peuvent être obtenues par composition : l'ouverture ( $\gamma = \delta\epsilon$ ) et la fermeture ( $\varphi = \epsilon\delta$ ) par exemple.

### 2.2.2.2 Filtres médians vectoriels

Le plus populaire des filtres vectoriels est sans aucun doute le filtre médian vectoriel (VMF) [LUKAC06B] qui a été initialement proposé par ASTOLA [ASTOLA90]. La littérature sur les filtres médian vectoriels étant très importante, nous ne ferons pas un état de l'art sur ce type de filtre et le lecteur intéressé pourra consulter plusieurs articles et ouvrages de référence [LUKAC06A, LUKAC06B, LUKAC05, PLATAN00]. Les filtres médian vectoriels sont basés sur un ordre de données vectorielles dans une fenêtre de filtrage  $F$ . A chaque vecteur  $\mathbf{x}_k \in F$  est associé une mesure de distance cumulative :

$$R_k = \sum_{i=1}^N \|\mathbf{x}_k - \mathbf{x}_i\|_\gamma \text{ pour } k = 1, 2, \dots, N$$

où  $\|\mathbf{x}_k - \mathbf{x}_i\|_\gamma$  désigne la distance de Minkowski entre deux vecteurs

$$\|\mathbf{x}_k - \mathbf{x}_j\|_\gamma = \left( \sum_{i=1}^p |\mathbf{x}_k^i - \mathbf{x}_j^i|^\gamma \right)^{\frac{1}{\gamma}}$$

$\gamma$  définit le type de norme qui est habituellement la norme Euclidienne ( $\gamma = 2$ ). La sortie d'un filtre médian vectoriel est le vecteur  $\mathbf{x}_{(1)}$  auquel est associé la plus petite mesure de distance cumulative :

$$\mathbf{x}_{(1)} = \arg \min_{\mathbf{x}_i \in F} R_i$$

Suivant le type de distance utilisée entre deux vecteurs, il est possible de différencier les techniques selon le fait qu'elles opèrent sur des distances entre vecteurs [LUKAC06B], sur des angles entre vecteurs [TRAHAN96] ou bien les deux [KARAKO95].

### 2.2.2.3 Discussion

A partir de ce que nous venons d'exposer, nous pouvons faire plusieurs observations. La première est que les ordres réduits utilisés pour les filtres médian vectoriels ne sont pas adaptés au filtrage morphologique puisqu'ils ne définissent pas un treillis complet. Ils sont cependant adaptés à n'importe quel type de données multivariées et donc à n'importe quel espace couleur. La seconde observation est que les ordres vectoriels utilisés pour le filtrage morphologique sont principalement des ordres totaux (même si des ordres basés sur la combinaison d'un ordre réduit basé sur des distances et d'un ordre lexicographique ont été récemment proposés [ANGULO05A, ANGULO07]). Ces ordres totaux ont tout de même le désavantage majeur de n'être adaptés qu'à un type de représentation des données vectorielles, et notamment pour les données vectorielles couleur : l'espace  $RGB$  pour l'ordre par entrelacement de bits et l'espace  $IHSL$  [HANBUR01] pour l'ordre lexicographique. En conséquence, ces ordres de données vectorielles ne peuvent être utilisés pour effectuer des opérations morphologiques sur des données vectorielles quelconques provenant par exemple d'images multi-spectrales. Pour l'ordre lexicographique, le rôle prioritaire donné à la première composante est difficile à contourner, à part pour les images couleur en utilisant un espace couleur de type  $IHSL$ . Nous n'avons cependant pas décrit ici toutes les approches possibles pour établir un ordre de données vectorielles, mais uniquement les principales. On trouvera une revue très complète dans l'article récent [APTOUL07] qui classe notamment, dans le cadre de la morphologie mathématique multivariée, les approches suivant trois propriétés : génériques (indépendantes de la représentation



vectorielle considérée, c'est à dire sans connaissances *a priori*), préservant les vecteurs, et sans priorité entre les composantes. Nos contributions respectent toutes ces propriétés. Pour conclure, nous pouvons dire que tous les ordres usuels de la littérature ont des désavantages et il serait intéressant, à partir de ces observations, de définir un nouvel ordre vectoriel qui soit suffisamment général pour permettre d'effectuer un filtrage morphologique ou médian et ceci sur des données vectorielles de dimensions quelconques.

## 2.2.3 Chemin Hamiltonien dans un graphe

### 2.2.3.1 Motivations

Comme nous l'avons mentionné précédemment il y a une équivalence

$$(\text{ordre total sur } \mathbb{R}^p) \Leftrightarrow (\text{bijection } h : \mathbb{R}^p \rightarrow \mathbb{R})$$

On peut également relier la notion d'ordre vectoriel à la notion de courbe de remplissage de l'espace. Une telle courbe est une courbe qui parcourt un espace multidimensionnel en passant, de point en point, et de manière unique, par tous les points de cet espace. Cette courbe remplit donc l'espace. Nous ne considérons que ces courbes dans le cas discret car nous opérons sur des images représentées par une grille discrète. Nous pouvons alors établir la double équivalence suivante :

$$(\text{ordre total sur } \mathbb{R}^p) \Leftrightarrow (\text{bijection } h : \mathbb{R}^p \rightarrow \mathbb{R}) \Leftrightarrow (\text{courbe de remplissage de l'espace } \mathbb{R}^p)$$

La figure 2.4 présente une courbe de remplissage de l'espace (de type PEANO) d'un domaine discret qui est une image 2D. Sur cette figure sont également superposées quelques valeurs de  $h(\mathbf{x})$ .

A partir du moment où l'espace multidimensionnel est discret, nous pouvons le représenter par un graphe connexe. Ainsi, nous pouvons établir la triple équivalence suivante :

$$(\text{ordre total sur } \mathbb{R}^p) \Leftrightarrow (\text{bijection } h : \mathbb{R}^p \rightarrow \mathbb{R}) \Leftrightarrow (\text{courbe de remplissage de l'espace } \mathbb{R}^p) \Leftrightarrow (\text{chemin Hamiltonien sur } \mathbb{R}^p)$$

Les définitions habituelles des courbes de remplissage de l'espace dans  $\mathbb{R}^p$  pour le filtrage d'images [STRING99, REGAZZ97] utilisent des courbes (par exemple une courbe de Peano) qui sont indépendantes de l'organisation spatiale et elles ne respectent donc pas la topologie. En effet, trouver une courbe de remplissage de l'espace qui respecte la topologie revient à construire un chemin Hamiltonien. Trouver un chemin Hamiltonien est un problème trop difficile (NP-complet) pour être directement résolu sur  $\mathbb{R}^p$  : pour un graphe de  $n$  noeuds, il y a  $(n - 1)!$  chemins Hamiltoniens possibles. La recherche d'un chemin Hamiltonien a déjà été utilisé dans le cadre des algorithmes de réorganisation de palettes couleur (provenant par exemple d'images compressées) [PINHO04, BATTIA04]. Nous avons choisi de prendre ce point de vue pour chercher à ordonner des données vectorielles à savoir construire une courbe de remplissage de l'espace à travers la définition d'un chemin Hamiltonien. Ce chemin sera défini de manière dynamique, mais plutôt que de chercher à définir ce chemin directement sur l'espace  $\mathbb{R}^p$ , nous proposons de le faire sur une fenêtre d'analyse  $F$ .

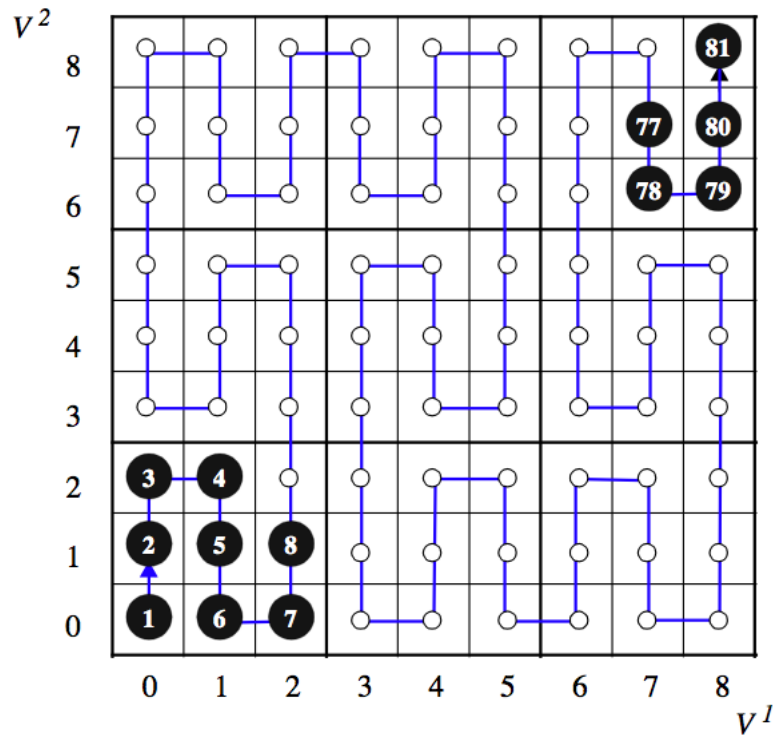


FIG. 2.4 – Courbe de PEANO sur un domaine discret en 2D (tiré de [LAMBER02]).

A une fenêtre d'analyse  $F$ , nous pouvons associer un graphe complet dont les noeuds correspondent aux pixels de  $F$ . A chaque pixel est associée une couleur et cela correspond à avoir une fonction  $f \in \mathcal{H}(V)$ ,  $f : V \rightarrow F$  qui associe un vecteur couleur  $\mathbf{x} \in F$  à chaque noeud. De la même manière, nous pouvons associer un poids à chaque arête du graphe,  $w \in \mathcal{H}(E)$ ,  $w : E \rightarrow \mathbb{R}^+$ . Classiquement, nous considérons la norme Euclidienne, et  $w(u, v) = \|f(u) - f(v)\|_2$ . Cela revient donc à considérer une matrice  $W$  représentant le graphe complet sur  $F$ . Cette matrice est symétrique et de trace nulle car nous ne considérons pas de boucles. La matrice est pondérée par les distances entre les noeuds.

Dans la suite, nous nous focaliserons sur le cas des images couleur, mais le principe reste le même pour des vecteurs de plus grandes dimensions. Pour illustrer notre approche, nous utiliserons une image de référence communément utilisée en morphologie mathématique (figure 2.5, une peinture de MIRO nommée « Le chanteur ») [ANGULO05B, HANBUR01, 45]. Nous étudierons le comportement de différents ordres ainsi que celui que nous proposons sur une fenêtre d'analyse pré-définie  $F$  (spécifiée par la zone extraite et agrandie dans le coin haut droit de la figure 2.5).

### 2.2.3.2 Extraction de l'infimum et du supremum

Pour définir un chemin Hamiltonien sur une fenêtre d'analyse  $F$ , nous considérons le graphe complet  $\mathcal{G}_0$  défini sur  $F$ . Puisqu'il est difficile de trouver un chemin Hamiltonien optimal sur  $\mathcal{G}_0$  parmi les  $(|F| - 1)!$  différentes possibilités, nous proposons d'en obtenir une approximation. Au lieu de directement chercher à définir un tel chemin Hamiltonien, nous commençons par en déterminer les bornes ( $\wedge$  et  $\vee$ ).



FIG. 2.5 – Image de référence (peinture de Joan Miro : « Le chanteur ») avec une zone mise en valeur.

Soit  $\mathcal{T}_0 = \mathcal{T}(\mathcal{G}_0)$  le MST de  $\mathcal{G}_0$ . Un MST étant une généralisation aux grandes dimensions d'une liste triée de scalaires [THEOHA05], nous pouvons utiliser sa structure afin de déterminer les noeuds qui sont des candidats possibles pour être ses bornes. Un noeud  $v$  d'un chemin Hamiltonien est l'une de ses bornes si  $\delta(v) = 1$ , c'est à dire que c'est une feuille. Nous utilisons ce principe pour les extraire. Soit  $\mathcal{N}_0 = \{u | \delta(u) = 1, u \in \mathcal{T}_0\}$  l'ensemble des feuilles de  $\mathcal{T}_0$ . Les noeuds de  $\mathcal{N}_0$  sont les seuls candidats possibles comme bornes du chemin Hamiltonien. Puisque la plupart du temps  $|\mathcal{N}_0| > 2$ ,  $\mathcal{N}_0$  doit être réduit à deux éléments. Pour ce faire, nous répétons le même procédé sur le graphe complet construit sur les noeuds de  $\mathcal{N}_0$  jusqu'à ce que  $|\mathcal{N}_i| = 2$  avec  $i$  le numéro de l'itération (évidemment  $|\mathcal{N}_{i+1}| < |\mathcal{N}_i|$ ).

En résumé, pour extraire les bornes d'un chemin Hamiltonien, le principe peut être décrit comme suit :

```

i ← 0
Construire le graphe complet  $\mathcal{G}_i$  sur la fenêtre d'analyse  $F$ 
Répéter
     $\mathcal{T}_i = \mathcal{T}(\mathcal{G}_i)$ 
     $\mathcal{N}_i = \{u | \delta(u) = 1, u \in \mathcal{T}_i\}$ 
    Construire le graphe complet  $\mathcal{G}_{i+1}$  sur les noeuds de  $\mathcal{N}_i$ 
    i ← i + 1
jusqu'à ce que ( $|\mathcal{N}_{i-1}| = 2$ )
 $\mathcal{T}_i = \mathcal{G}_i$ 
 $\mathcal{N}_i = \{u \in \mathcal{T}_i\}$ 
    
```

Algorithme 1: Algorithme d'extraction des bornes d'un chemin Hamiltonien.

A la fin du processus,  $\mathcal{N}_i = \{u, v\}$  contient exactement deux noeuds qui sont les bornes du chemin Hamiltonien. Cependant, il reste à déterminer lequel de ces deux noeuds est le  $\vee$

(respectivement le  $\wedge$ ). Le  $\wedge$  est identifié comme le noeud le plus proche à un vecteur de référence  $\mathbf{x}_{ref}$  [HANBUR01, SORIA-02] :

$$\vee = \arg \max_{v \in \mathcal{N}_i} \|f(v), \mathbf{x}_{ref}\|_2 \text{ et } \wedge = \arg \min_{v \in \mathcal{N}_i} \|f(v), \mathbf{x}_{ref}\|_2$$

Cela correspond à utiliser une couleur de référence pour le cas des images couleur. Cette couleur de référence est généralement le noir. Utiliser une couleur de référence pour la définition d'un ordre vectoriel de couleurs a été initialement proposé par SORIA-FRISCH dans [SORIA-02] et HANBURY pour l'espace couleur *IHSL* [HANBUR01].

Calculer le MST d'un graphe a une complexité en  $O(|E|\log|V|)$  et en  $O(N^2\log N)$  pour un graphe complet puisque  $|V| = |F| = N$ . On peut alors montrer que la complexité moyenne de l'algorithme est en  $O(N^3\log N)$ .

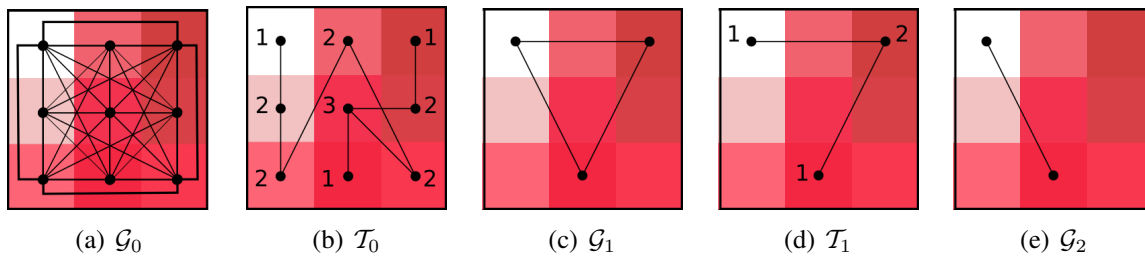


FIG. 2.6 – Les graphes  $\mathcal{G}_i$  et  $\mathcal{T}_i$  lors des différentes étapes de l'algorithme pour l'extraction des bornes du chemin Hamiltonien. Pour chaque  $\mathcal{T}_i$ , les degrés des noeuds sont précisés sur la figure.  $\mathcal{T}_2$  n'est pas montré ici.

La figure 2.6 présente les différentes étapes de l'algorithme. Le graphe complet  $\mathcal{G}_0$  est construit sur la fenêtre d'analyse  $F$  (Figure 2.6(a)) et son MST  $\mathcal{T}_0$  est calculé (Figure 2.6(b)). Un nouveau graphe complet est défini  $\mathcal{G}_1$  (Figure 2.6(c)) sur les feuilles de  $\mathcal{T}_0$  et son MST  $\mathcal{T}_1$  est calculé (Figure 2.6(d)). Finalement, un graphe complet  $\mathcal{G}_2$  contenant deux noeuds est obtenu (Figure 2.6(e)). Avec le noir comme couleur de référence  $\mathbf{x}_{ref}$ , le  $\vee$  est le pixel du coin haut gauche de  $F$  et le  $\wedge$  le pixel central de la dernière ligne.

### 2.2.3.3 Construction de l'ordre de données vectorielles

Une fois que les bornes du chemin Hamiltonien ont été déterminées, le chemin Hamiltonien peut être entièrement construit. Sur la fenêtre d'analyse  $F$ , un chemin Hamiltonien  $p = (v_1, v_2, \dots, v_k)$  doit respecter les propriétés suivantes :

- La longueur de  $p$  est  $|F| = N$ ,
- Ses bornes sont des feuilles :  $\delta(v_1) = \delta(v_k) = 1$ ,
- $v_1 = \wedge$  et  $v_k = \vee$ ,
- Les noeuds (hors feuilles) ont un degré (simple) égal à deux :  $\forall v \in \{v_2, \dots, v_{k-1}\}, \delta(v) = 2$ .

La construction du chemin Hamiltonien que nous proposons est basée sur le principe du plus proche voisin, le voisinage étant considéré sur le graphe complet initial  $\mathcal{G}_0$  afin de pouvoir couvrir n'importe quel chemin. Pour déterminer le plus proche voisin  $u$  d'un noeud  $v$ , nous considérons le poids de l'arête  $w(u, v)$  mais également la saillance du noeud voisin  $u$ . La saillance d'un noeud  $v$  nous permet de quantifier son importance globale dans la cascade des MST qui ont

été générés lors de l'extraction de bornes du chemin Hamiltonien. Nous définissons la saillance d'un noeud comme suit :

$$\Delta(v) = \sum_{i=0}^{i_{max}} (i + 1) \cdot \delta(\mathcal{T}_i, v) \text{ où } \delta(\mathcal{T}_i, v) = \delta(v), v \in \mathcal{T}_i$$

où  $i$  correspond à une itération de l'algorithme d'extraction des bornes du chemin Hamiltonien et  $i_{max}$  est le nombre total d'itérations. Par exemple, la saillance du noeud  $v_j$  correspondant au pixel du coin haut gauche de  $F$  dans la figure 2.6(a) est :

$$\Delta(v_j) = 1 \cdot \delta(\mathcal{T}_0, v_j) + 2 \cdot \delta(\mathcal{T}_1, v_j) + 3 \cdot \delta(\mathcal{T}_2, v_j) = 1 \cdot 1 + 2 \cdot 1 + 3 \cdot 1 = 6$$

La construction du chemin Hamiltonien est en  $O(N^2)$  et peut être résumée ainsi :

$$\left\{ \begin{array}{l} v_1 = \wedge \text{ et } v_N = \vee \\ v_{j+1} = \arg \min_{\substack{u \sim v_j \\ u \notin \{v_1, \dots, v_j\}}} (w_{uv_j} \Delta(u)) \quad \forall j = 1, \dots, (N - 2) \end{array} \right.$$

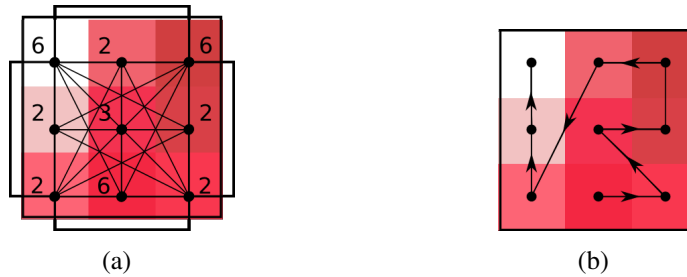


FIG. 2.7 – (a) Le graphe  $\mathcal{G}_0$  avec les saillances des noeuds  $\Delta(v)$ , (b) Le chemin Hamiltonien construit.

La figure 2.7 illustre la construction de l'ordre vectoriel sur  $F$  comme un chemin Hamiltonien : sur le graphe complet  $\mathcal{G}_0$  (Figure 2.7(a), avec les saillances des noeuds précisées), nous obtenons le chemin de la figure 2.7(b).

### 2.2.3.4 Résultats

Finalement, nous montrons quelques résultats expérimentaux et comparaisons de l'ordre proposé avec l'ordre réduit basé sur des distances, l'ordre lexicographique et l'ordre pare entre-lacement de bits. Pour avoir une idée de l'influence du choix de l'ordre de données vectorielles, nous appliquons plusieurs d'entre eux sur la fenêtre d'analyse  $F$  de la figure 2.5. Les résultats sont montrés par la figure 2.8. La première ligne présente la fenêtre d'analyse  $F$ , les autres lignes, les différents ordres vectoriels obtenus. Notre ordre est noté « Ordre basé graphe ».

Comme attendu, l'ordre marginal produit des fausses couleurs même si cela n'est pas visuellement évident. Pour l'ordre lexicographique, nous avons alterné l'ordre de comparaison des composantes pour illustrer le rôle dominant qui est attribué à la première composante. Ici, la composante Verte comme composante dominante produit les meilleurs résultats mais cela risque fort de ne pas être le cas pour une autre fenêtre d'analyse. Pour l'ordre réduit, la distance

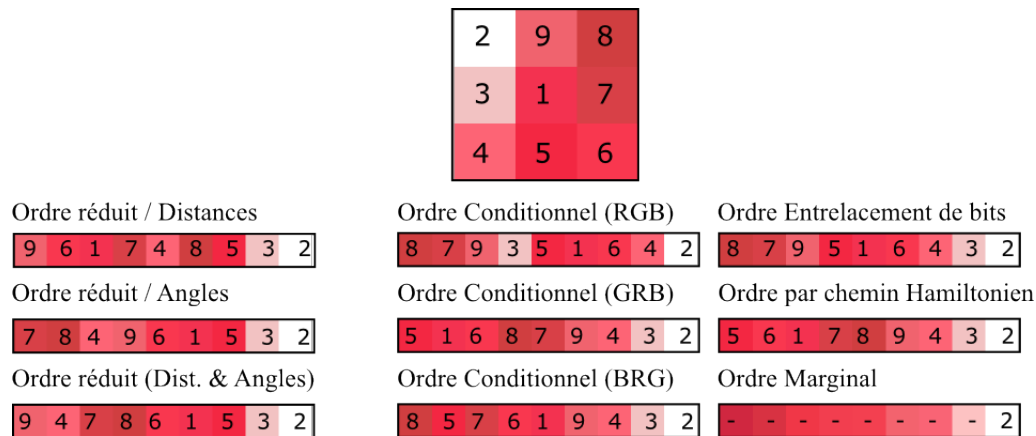


FIG. 2.8 – Comparaison entre différents ordres vectoriels sur une fenêtre d'analyse  $F$  de la figure 2.5. Chaque couleur de  $F$  est indexée par un numéro qui est précisé ensuite dans chacun des ordres obtenus (- indique que la couleur obtenue n'appartient pas à  $F$ ).

choisie a une grande influence sur l'ordre induit. L'ordre par entrelacement de bits produit des résultats proches de l'ordre lexicographique  $RGB$ , ce qui montre que la façon dont les bits sont entrelacés induit encore de l'asymétrie. Les résultats obtenus avec notre ordre sont visuellement satisfaisants et proches ici de ceux obtenus avec un ordre lexicographique  $GRB$ .

Dans le cadre du filtrage morphologique, la figure 2.9 montre quelques résultats d'érosion et de dilatation (espace couleur  $RGB$ , fenêtre d'analyse de taille  $3 \times 3$  et noir comme couleur de référence). On voit que les résultats sont similaires pour les trois ordres vectoriels sauf pour l'ordre lexicographique dans les zones rouges de l'image. L'ordre par entrelacement de bits semble moins bon que les deux autres ordres pour la dilatation. En changeant la couleur de référence, on obtient des comportements différents des opérations morphologiques qui privilégient ou non les vecteurs couleur proches de la référence (plus de détails dans [18]). Un intérêt de notre ordre est qu'il peut s'appliquer à des données de dimensions quelconques. Par exemple, nous pouvons effectuer n'importe quelle opération morphologique dans l'espace couleur  $CIECAM02$  [MORONE02], ce qui n'est envisageable avec aucun autre ordre vectoriel. La figure 2.10 présente des exemples de traitements morphologiques plus complexes (gradient, réhaussement de contraste, waterfall par watershed) dans l'espace couleur  $CIECAM02$  avec notre ordre basé sur les graphes. Nous rappelons ici que dans cet espace couleur, une couleur est décrite par un vecteur de dimension  $p = 6$ .

Dans le cadre du filtrage médian, la figure 2.11 présente quelques résultats avec différents ordres vectoriels pour le traitement de bruit impulsionnel. Les meilleurs résultats sont obtenus avec un filtre médian vectoriel basé sur un ordre réduit. Notre ordre ne permet pas de surpasser l'ordre réduit sauf lorsque l'on modifie la pondération des arêtes pour considérer des filtres de type DDF (Directional Distance Filter). Pour plus de détails, on consultera [18].

En terme de complexité, les ordres vectoriels classiques (lexicographiques, entrelacement de bits, réduit) ont une complexité en  $O(N^2)$  avec un algorithme classique de tri de scalaires. La complexité de notre approche est  $O(N^3 \log N)$ .



FIG. 2.9 –  $\epsilon$  et  $\delta$  avec différents ordres vectoriels sur la figure 2.5.

## 2.2.4 Marches aléatoires

Nous proposons ici une autre manière de définir un ordre de données vectorielles (issus d'images dans nos expérimentations). Nous cherchons cette fois à définir l'ordre par une réduction non linéaire de dimension grâce à des marches aléatoires sur graphes.

### 2.2.4.1 Principe

Une marche aléatoire sur un graphe est un processus stochastique qui parcourt le graphe en sautant aléatoirement de noeud en noeud [VONLUX06, COIFMA06]. La probabilité de transition d'un noeud  $v_i$  vers un noeud  $v_j$  est donnée par

$$p_{ij} = p_{v_i v_j} = \frac{w(v_i, v_j)}{d(v_i)} = \frac{w_{v_i v_j}}{d_i}$$

car la probabilité de transition en un saut d'un noeud  $v_i$  à un noeud  $v_j$  est proportionnelle à  $w(v_i, v_j)$ , la valuation de l'arête reliant ces deux noeuds. La matrice de transition  $P = (p_{ij})$  associée est alors définie par

$$P = D^{-1}S$$

où  $D$  est la matrice des degrés des noeuds et  $S$  une matrice de similarité associée au graphe.  $p_{ij}$  peut être interprété comme la probabilité de transition du noeud  $v_i$  au noeud  $v_j$  en un saut



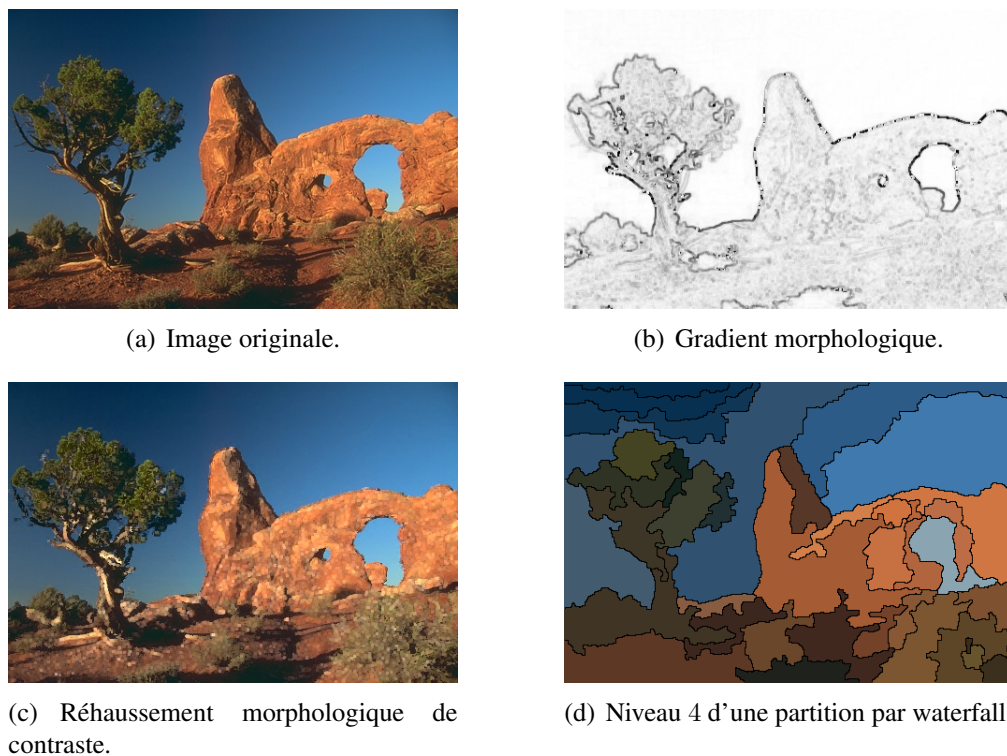


FIG. 2.10 – Exemples de traitements morphologiques dans l'espace *CIECAM02* avec un ordre basé sur les graphes.

(une marche).  $P$  est généralement symétrique et pour chaque colonne la somme des éléments est de 1. Cette matrice est intéressante car elle reflète la géométrie intrinsèque des données [ROBLES04]. Une marche aléatoire correspond à une chaîne de Markov homogène puisque les probabilités de transition restent les mêmes à chaque fois que l'on revient sur un noeud du graphe (les probabilités ne dépendant pas d'un facteur temps). Les chaînes de Markov sont définies en termes d'états et de transitions entre ces derniers. Les états sont dans notre cas les noeuds du graphe. Dans une chaîne de Markov, deux états  $i$  et  $j$  sont dits communicants si l'on peut atteindre l'un à partir de l'autre avec une probabilité finie ; ce qui signifie que le graphe est connexe. Si l'on veut décrire la probabilité de transition  $p_t(v_i, v_j)$  d'un noeud  $v_i$  à un noeud  $v_j$  en  $t$  sauts, il suffit de considérer des voisinages plus larges, ce qui correspond à élever la matrice

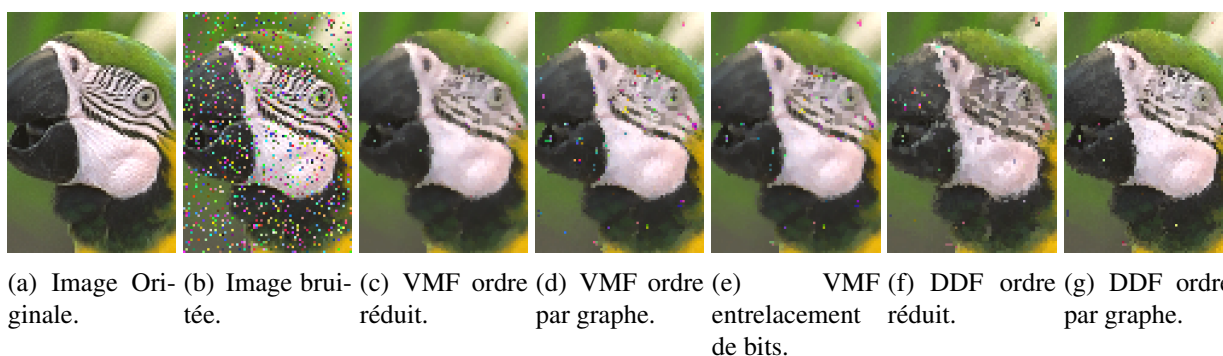


FIG. 2.11 – Image bruitée par 15% de bruit impulsionnel et débruitée par un VMF et un DDF avec différents ordres vectoriels.



$P$  à la puissance  $t$ . Si le graphe est connexe et non bipartite, alors la marche aléatoire converge vers une distribution stationnaire  $\pi = [\pi_1, \dots, \pi_n]$  satisfaisant  $\pi P^t = \pi$  avec  $\pi_i = \frac{d_i}{\text{vol}(V)}$  c'est à dire  $\lim_{t \rightarrow \infty} p_t(v_j, v_i) = \pi_i$ .

Il existe une connexion forte entre les marches aléatoires et le clustering spectral [HEIN05]. En effet, les vecteurs propres  $\mathbf{v}$  de  $P$  obtenus en résolvant  $P\mathbf{v} = \lambda\mathbf{v}$  sont exactement les mêmes que ceux obtenus sur le Laplacien normalisé. Le laplacien non normalisé est défini par :

$$\Delta = D - S$$

et le Laplacien normalisé par

$$L = D^{-1}\Delta = I - D^{-1}S = I - P$$

Les vecteurs propres de  $(I - D^{-1}S)\mathbf{v} = \lambda'\mathbf{v}$  sont donc les mêmes que ceux de la matrice de transition  $P$  et l'on a  $\lambda = 1 - \lambda'$ . Les valeurs propres de  $P$  sont  $\lambda_1 = 1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq -1$  et l'on a

$$P = \sum_{i=1}^{|V|} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

Un parallèle peut alors être fait entre la matrice de transition  $P$  et le critère de coupe normalisé de SHI et MALIK [SHI00]. Le critère de coupe normalisé  $Ncut$  est un critère pour trouver une coupe optimale des noeuds  $V$  d'un graphe en deux sous-ensembles  $A$  et  $\bar{A}$  et est défini comme suit :

$$Ncut(A, \bar{A}) = \left( \frac{1}{\text{Vol}(A)} + \frac{1}{\text{Vol}(\bar{A})} \right) \sum_{v_i \in A, v_j \in \bar{A}} S_{ij}$$

Ce problème étant NP-dur, SHI et MALIK ont proposé une méthode spectrale de coupe basée sur le deuxième plus petit vecteur propre du Laplacien dont le signe des éléments définit la coupe. Avec la relation précédente, cela permet de mettre en relation pourquoi le deuxième vecteur propre est utilisé pour trouver la meilleure coupe et non le premier qui correspond au plus grand vecteur propre  $\mathbf{v}_1 = I$  de  $P$  et qui ne contient donc aucune information pour trouver une coupe. Le vecteur propre  $\mathbf{v}_1$  de  $P$  associé à  $\lambda_1$  est connu comme le vecteur de FIEDLER de  $P$  et fournit de l'information concernant la structure en groupes du graphe. Cette mise en relation entre le problème de coupe normalisée et les marches aléatoires est due à MEILA et SHI [MEILA01].

### 2.2.4.2 Réduction de dimension

Une manière idéale de concevoir un ordre de données vectorielles est, comme l'a proposé GOUTSIAS [GOUTSI95], de concevoir une transformation  $h : \mathbb{R}^p \rightarrow \mathbb{R}^q$  avec  $q \ll p$ . Idéalement, si  $q = 1$  cela revient à notre approche précédente qui définit un chemin Hamiltonien sur les données à ordonner. Une autre manière de procéder repose dans la construction de la transformation  $h$  comme une opération de réduction de dimension. Le problème général de la réduction de dimension est le suivant. A partir d'un ensemble  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  de  $N$  vecteurs de dimension  $p$  avec  $\mathbf{x}_i \in \mathbb{R}^p$ , on désire trouver un nouvel ensemble  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$  de  $N$  vecteurs de dimension  $q$  avec  $\mathbf{y}_i \in \mathbb{R}^q$  et  $q \ll p$  de manière à ce que  $\mathbf{y}_i$  représente au mieux  $\mathbf{x}_i$ . On cherche donc à avoir  $\|\mathbf{y}_i - \mathbf{y}_j\|_2$  qui soit faible lorsque  $\mathbf{x}_i$  et  $\mathbf{x}_j$  sont proches. C'est le principe énoncé par BELKIN et NIYOGI [BELKIN03] appelé « Laplacian Eigenmaps » qui correspond à minimiser sous contraintes

$$\frac{1}{2} \sum_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|_2 S_{ij} = Tr(\mathbf{Y}^T \Delta \mathbf{Y})$$

avec  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$ . Ce critère peut être reformulé et les solutions du précédent problème d'optimisation correspondent à  $(D - S)\mathbf{y} = \lambda D\mathbf{y}$  et donc à  $P\mathbf{y} = \lambda\mathbf{y}$  pour le Laplacien normalisé, ce qui nous ramène donc directement aux marches aléatoires. Nous pouvons donc définir une transformation de réduction de dimension comme  $h : \mathbf{x}_i \rightarrow (y_2(i), \dots, y_q(i))$  où  $y_k(i)$  désigne le  $i^{\text{ème}}$  élément du vecteur propre  $\mathbf{y}_k$ . On remarque que l'on ne considère pas le vecteur propre  $\mathbf{y}_1$  correspondant à  $\lambda_1 = 1$  car il ne porte aucune information.  $q$  désigne le nombre de vecteurs propres retenus. Grâce à ce type de méthode, nous pouvons détecter des structures de faibles dimensions dans un espace initial de très grande dimension : ceci est désigné par le terme de manifold learning dans la communauté. Les Laplacian Eigenmaps permettent de détecter des structures non linéaires avec un graphe comme représentation discrète d'une variété (un manifold). LAFON et COIFMAN [LAFON06A, LAFON06B, COIFMA06] ont étendu ce principe aux puissances  $P^t$  de la matrice de transition  $P$  et l'ont nommé « diffusion maps ». La figure 2.12 montre, pour un ensemble de points représentant une structure non linéaire discrète (le maintenant classique « SwissRoll »), la réduction de dimension obtenue. On remarque que la projection effectuée respecte bien la géométrie initiale des points. Les vecteurs propres obtenus sont également présentés dans la figure 2.12.

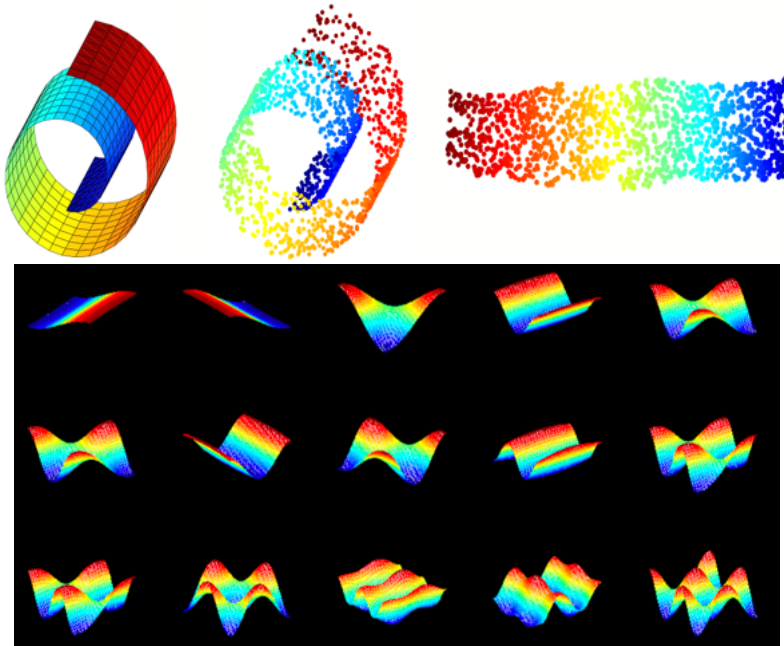


FIG. 2.12 – Réduction de dimension sur des points provenant d'une distribution de type « SwissRoll ». La première ligne présente le SwissRoll, quelques points de celui-ci et sa représentation après réduction de dimension. Les lignes suivantes présentent les vecteurs propres obtenus.

### 2.2.4.3 Ordre de données multivariées provenant d'images

Dans cette section, nous proposons de nous baser sur une réduction de dimension pour ordonner des données multivariées. Le principe est le suivant. A partir d'un ensemble  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  de  $N$  vecteurs de dimension  $p$  avec  $\mathbf{x}_i \in \mathbb{R}^p$ , nous définissons la réduction de dimension  $h : \mathbb{R}^p \rightarrow \mathbb{R}$ . La réduction de dimension est effectuée en calculant les vecteurs propres de la matrice de transition des marches aléatoires et en ne conservant que le deuxième plus grand vecteur propre. Nous pourrions également retenir les  $q$  plus grands vecteur propres (hors le premier)

et ordonner les données multivariées grâce à un ordre lexicographique. Cependant, la représentation la plus parcimonieuse basée sur un seul vecteur propre est généralement suffisante comme nous allons le voir par la suite. Effectuer une réduction de dimension vers une représentation scalaire ( $q = 1$ ) nécessite, si l'on veut bénéficier d'un ordre total, de connaître les bornes de l'ordre ainsi défini. En effet, la réduction de dimension proposée revient à définir directement un chemin Hamiltonien, mais il reste à en définir les bornes minimum et maximum. Comme précédemment, cela se réalise en convenant d'un vecteur de référence (dans l'espace initial), le minimum étant la borne du chemin qui est la plus proche de la référence.

L'algorithme complet est donné ci-dessous :

Construire un graphe connexe sur les  $N$  données multivariées à ordonner  
 Calculer la matrice de Similarité  $S \in \mathbb{R}^{N \times N}$  et les degrés  $D$   
 Calculer la matrice de transition  $P = D^{-1}S$   
 Calculer les vecteurs propres  $\mathbf{v}_1, \dots, \mathbf{v}_N$  de  $P\mathbf{v} = \lambda\mathbf{v}$   
 Retenir  $\mathbf{v}_2$  comme représentation des données.  
 Définir les bornes  $\vee$  et  $\wedge$ .

Algorithme 2: Algorithme pour ordonner des données vectorielles par réduction de dimension.

La complexité de l'algorithme est dominée par le calcul des vecteur propres qui est en  $O(N^3)$ .

### 2.2.4.4 Résultats

Reprenons tout d'abord l'exemple du filtrage que nous avons considéré pour la constitution d'un chemin Hamiltonien. La figure 2.13 présente l'ordre obtenu par marches aléatoires sur une fenêtre d'analyse. On considère un graphe complet et l'on effectue la réduction de dimension. On constate que l'ordre obtenu est très proche des précédents résultats que nous avons obtenu. Ceci illustre bien la validité de l'approche.

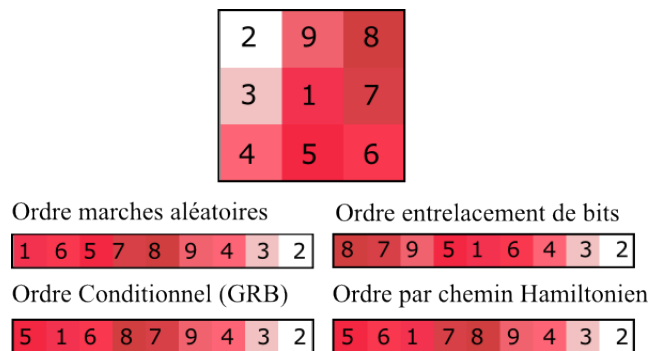


FIG. 2.13 – Comparaison entre différents ordres vectoriels sur une fenêtre d'analyse  $F$  de la figure 2.5. Chaque couleur de  $F$  est indexée par un numéro qui est précisé ensuite dans chacun des ordres obtenus.

Dans une optique de filtrage, nous pouvons, comme précédemment, considérer des opérations morphologiques opérant sur des images multivariées. Pour cela, il suffit d'appliquer la réduction de dimension sur la fenêtre d'analyse  $F$  du filtrage considéré. La réduction de dimension proposée permet d'obtenir un ensemble de scalaires sur  $F$ , si bien que l'on peut appliquer classiquement n'importe quel algorithme de filtrage d'image en niveaux de gris. Le graphe considéré pour la réduction de dimension sur une fenêtre d'analyse  $F$  de taille  $3 \times 3$  est un graphe complet et la couleur de référence le noir. La figure 2.14 présente quelques résultats d'opérations morphologiques simples (érosion et dilatation) sur une image. Comme la réduction de dimension est

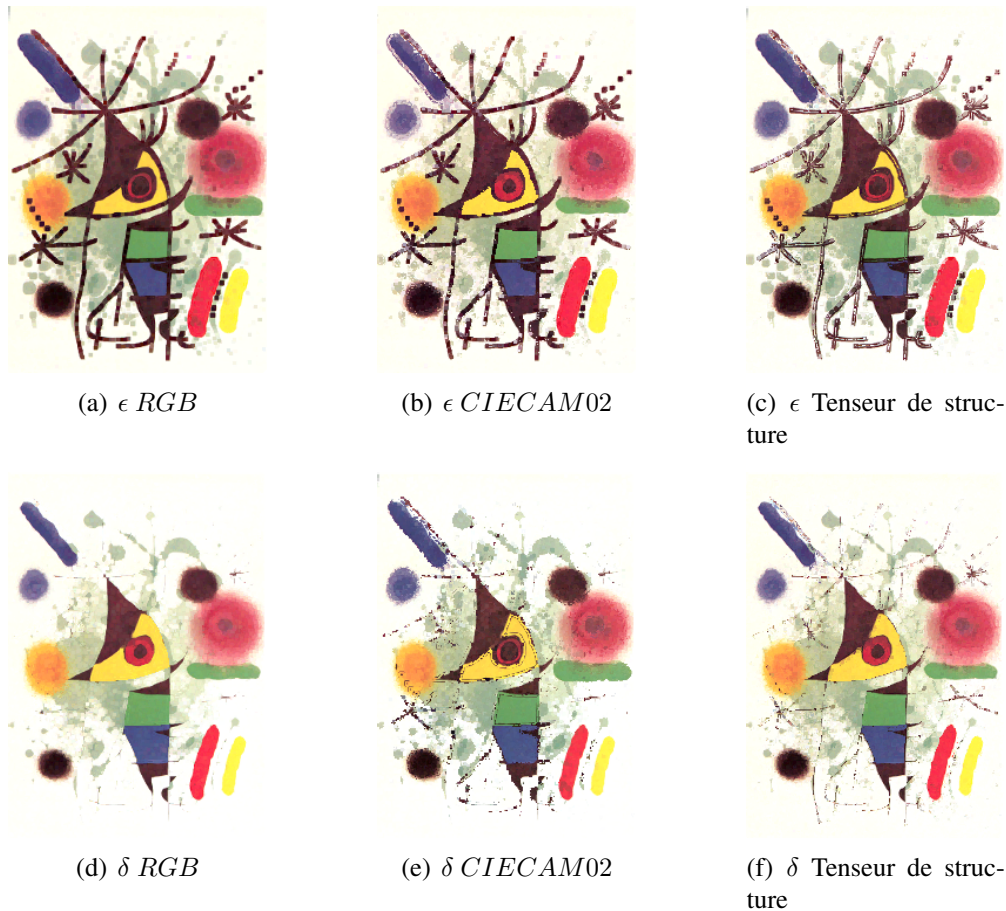


FIG. 2.14 –  $\epsilon$  (érosion) et  $\delta$  (dilatation) avec un ordre vectoriel basé sur les marches aléatoires sur la figure 2.5, ceci avec différentes représentations associées à un pixel.

basée sur l'utilisation d'une matrice de similarité entre pixels, les seuls éléments à déterminer pour l'appliquer sont le type de graphe considéré ainsi que la mesure de similarité employée. Nous considérons un graphe complet sur  $F$  et la mesure de similarité entre deux pixels est définie simplement par :  $s(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{d(\mathbf{x}_i, \mathbf{x}_j)}$  où  $d$  est une mesure de distance entre deux données multivariées. Nous avons considéré jusqu'ici que les données associées à un pixel sont des données vectorielles, et dans ce cas  $d$  est une norme  $L_2$  (distance Euclidienne classique). Cependant, nous pouvons associer à un pixel des données multivariées sous forme de matrice à travers, par exemple, le tenseur de structure. Pour un pixel d'une image, le tenseur de structure est défini comme suit pour une image  $I$  en niveaux de gris :  $J = (\nabla I \nabla I^T) = \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix}$  et par

$J = \sum_{i=1}^p (\nabla I_i \nabla I_i^T)$  pour une image multivariée. On peut définir la distance entre deux tenseurs à l'aide de la norme de FROBENIUS [BURGET07]. La norme de FROBENIUS sur une matrice

$2 \times 2$  représentant un tenseur de structure  $A$  est définie par :  $\|A\| = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \sqrt{\sum_{i,j=1}^2 a_{ij}^2}$

Ainsi, la distance entre deux tenseurs  $A$  et  $B$  est donnée par  $\|A - B\|$ . La matrice de référence considérée pour déterminer les bornes est la matrice nulle. On constate sur la figure 2.14 que suivant la représentation considérée, les résultats diffèrent pour l'érosion et la dilatation. L'effet global reste cependant le même. Tout l'intérêt réside ici dans le fait que pour effectuer une quelconque opération morphologique sur des données multivariées (un vecteur  $RGB$ , un vecteur  $CIECAM02$  ou bien une matrice représentant un tenseur de structure), nous avons juste besoin de définir une mesure de distance ; le reste de l'algorithme est identique. Nous pourrions donc à présent appliquer n'importe quelle opération morphologique à une image d'IRM de tenseur de diffusion comme cela est proposé par BURGETH [BURGET07]. Cependant, notre approche est plus simple et ne change pas quelle que soit la représentation multivariée. Un exemple de traitement sur une image 3D artificielle de tenseur de diffusion (chaque voxel est décrit par une matrice  $3 \times 3$ ) est donnée par la figure 2.15 où l'on applique une érosion, une dilatation et un gradient morphologique avec un élément structurant de taille  $5 \times 5 \times 5$ .

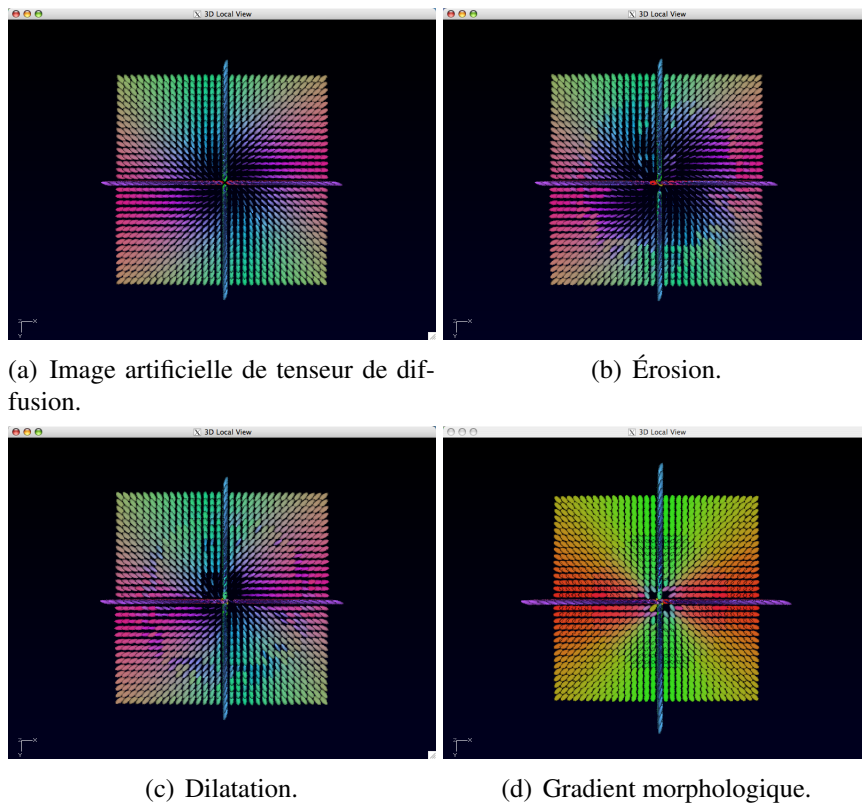


FIG. 2.15 – Opérations morphologiques sur une image de tenseur de diffusion.

Afin d'illustrer le comportement de l'algorithme, nous l'appliquons au problème de réorganisation de palettes couleur [PINHO04, BATTIA04]. On dispose d'une image couleur  $I$  codée par une image d'index  $I_{index} \in \mathbb{R}^2 \rightarrow \mathbb{N}$  et d'une palette  $P \in \mathbb{N} \rightarrow \mathbb{R}^3$  où à chaque élément de  $P$  est associé un vecteur  $RGB$   $\mathbf{x}_i$ . Le vecteur couleur d'un pixel  $\mathbf{p}(x, y)$  dans l'image couleur  $I$  est donc défini par  $P(I_{index}(\mathbf{p}(x, y)))$ . Nous appliquons l'algorithme de réduction de dimension



sur le graphe complet de  $P$  et la transformation  $h : \mathbb{R}^3 \rightarrow \mathbb{R}$  permet de définir un ordre sur les couleurs désigné par la transformation  $\psi : \mathbb{R}^3 \rightarrow \mathbb{N}$  c'est à dire  $\psi(\mathbf{x}) = rang(h(\mathbf{x}))$  où la fonction  $rang$  permet d'obtenir le rang de son argument parmi l'ensemble des valeurs possibles. Ceci redéfinit une nouvelle palette  $P'$ . L'image d'index peut ensuite être transformée à partir des valeurs de  $\psi$  par  $I'_{index}(\mathbf{p}(x, y)) = \psi(P(I_{index}(\mathbf{p}(x, y))))$  où  $\mathbf{p}(x, y)$  désigne un pixel. En conséquence, réorganiser la palette des couleurs revient également à réorganiser l'image d'index. Les figures 2.16 et 2.17 illustrent ce principe. On constate que la palette est ré-ordonnée et que l'image d'index associée a cette fois-ci une signification visuelle cohérente avec le contenu de l'image couleur. En effet, l'image d'index transformée se rapproche d'une image de luminance et l'on remarque que la structure spatiale des pixels est bien représentée alors que le traitement de réduction de dimension n'en tient absolument pas compte. On remarque de plus que l'image d'index obtenue peut être l'inverse d'une image de luminance (Figure 2.17(d)) si l'on ne prend pas de précautions afin de définir quelle est l'origine de l'espace initial (la référence est le noir pour la figure 2.16 et le blanc pour la figure 2.17).

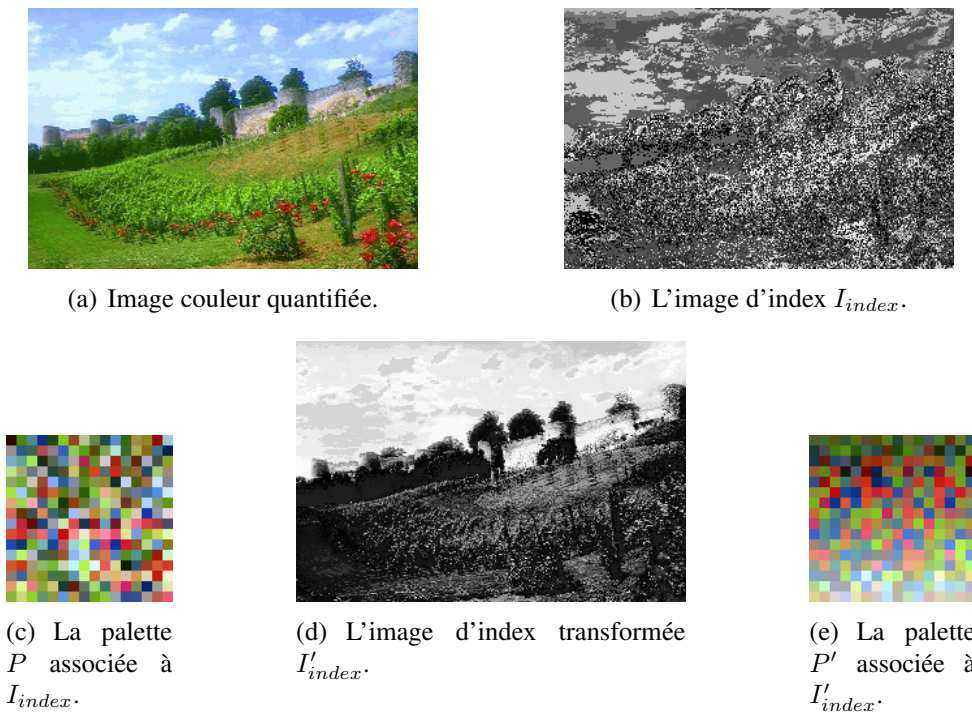


FIG. 2.16 – Réorganisation de la palette d'une image couleur quantifiée en 256 couleurs.

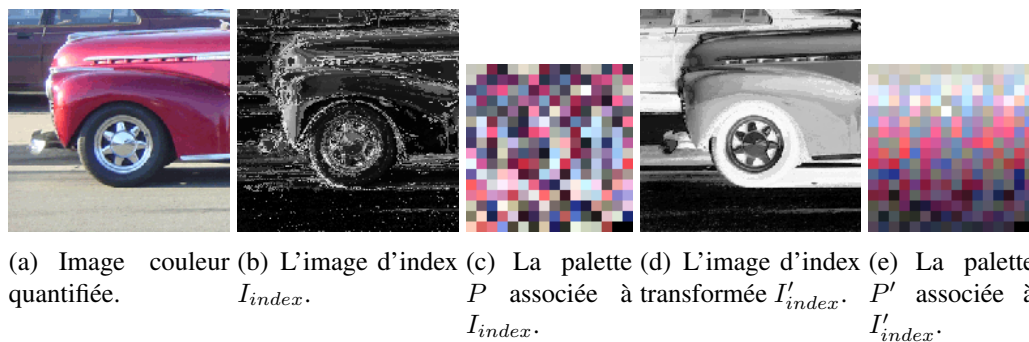
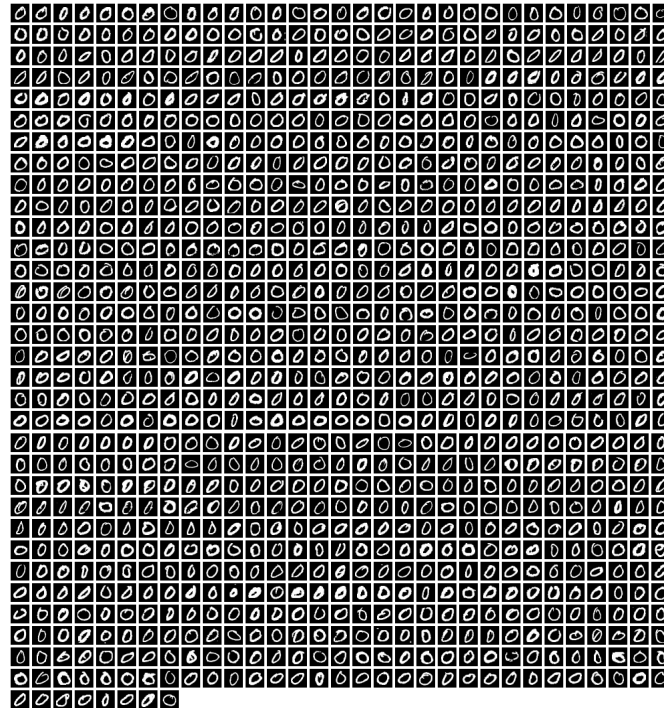


FIG. 2.17 – Réorganisation de la palette d'une image couleur quantifiée en 256 couleurs.

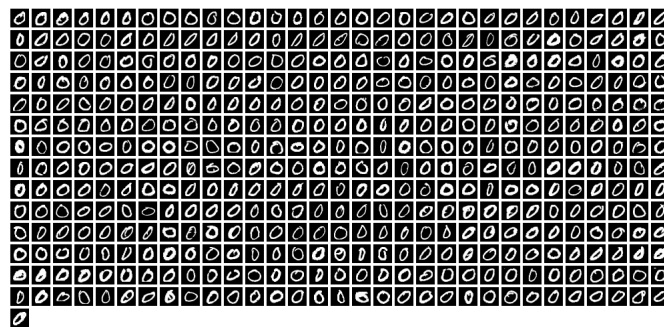
## 2.2.5 Conclusion et perspectives

Dans cette section, nous avons étudié la problématique des ordres de données vectorielles pour réaliser des opérations de traitement d'images. Nous nous sommes placés dans un cadre discret pour lequel un graphe de voisinage est associé aux données à ordonner. Nous avons pu alors mettre en avant que définir un ordre de données vectorielles revient à définir un chemin Hamiltonien sur un graphe. Deux approches ont alors été proposées. La première approche consiste à construire explicitement un chemin Hamiltonien sur un graphe. Contrairement aux algorithmes classiques, nous cherchons tout d'abord à déterminer les bornes de ce chemin et seulement ensuite nous réalisons sa construction. La deuxième approche consiste à construire directement ce chemin Hamiltonien par une réduction de dimension non linéaire des données initiales sur le graphe de voisinage associé. Ces deux approches diffèrent des approches classiques que l'on trouve dans la littérature du traitement d'images qui reposent principalement sur des ordres réduits basés sur des distances ou totaux basés sur un ordre lexicographique. Nos approches proposent donc des formalismes plus généraux qui peuvent s'appliquer à des images dont les données sont multivariées dès que l'on dispose d'un graphe de voisinage et d'une mesure de distance ou de similarité.

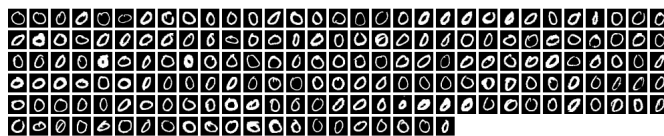
Une perspective directe des ces approches repose dans le traitement d'images multivariées quelconques par des approches morphologiques sans chercher à modifier les algorithmes tels qu'ils sont définis en niveaux de gris. De plus, étant donné que nos approches reposent sur deux données d'entrée associées aux données (un graphe de voisinage et une mesure de distance ou de similarité), nous pouvons également les utiliser pour traiter des variétés plus complexes à l'aide d'opérations morphologiques. A titre d'exemple, nous avons considéré l'ensemble des images représentant le chiffre zéro dans la base de test MNIST (1000 images). En leur associant un graphe de voisinage connexe (un 2-ppv fusionné avec un MST) et une mesure de distance Euclidienne entre les images, nous pouvons effectuer des opérations morphologiques sur cette variété. Par exemple, une érosion ou une dilatation reviennent à effectuer des opérations de contraction du graphe selon l'ordre de données vectorielles défini. La figure 2.18 présente de telles opérations pour la base d'images avant puis après contraction par érosion ou dilatation. L'application de méthodes morphologiques à des variétés n'a pas encore été exploré et les opérations réalisées restent à interpréter car cela est plus difficile qu'avec des images 2D pour lesquelles le résultat d'une opération morphologique est facilement interprétable visuellement.



(a) Chiffre 0.



(b) Érosion.



(c) Dilatation.

FIG. 2.18 – Érosion et dilatation sur les données de test MNIST pour le chiffre 0.



## 2.3 Régularisation discrète sur graphes pondérés

☞ Mes publications associées : [16, 19, 37, 51, 54, 57, 58, 67, 68, 72–74]

### 2.3.1 Introduction et problématique

En traitement d'image et par exemple en restauration d'image, on peut répartir les méthodes selon trois différents modèles de traitement [PEYR07] :

- Traitement local : ce modèle est le modèle traditionnel où seules les interactions locales autour d'un pixel sont prises en compte. Beaucoup de méthodes de traitement d'images sont conformes à ce modèle. On peut citer les filtres gaussien, médian [LUKAC06B], anisotropique [PERONA90], basés sur des ondelettes [MALLAT98] ou la minimisation de la variation totale par équations aux dérivées partielles [CHAN05B, RUDIN92].
- Traitement semi local : ce modèle correspond à prendre en compte des interactions de voisinage plus large qui peuvent être favorisées par la géométrie de l'image. On peut citer le filtre de YAROSLAVSKY [YAROSL85], le filtre bilatéral [TOMASI98], le filtre de SUSAN [SMITH97],
- Traitement non local : ce modèle a été récemment proposé par BUADES et al. [BUADES05A, BUADES05B] qui ont proposé de supprimer les contraintes spatiales locales et de les remplacer par des contraintes sur des blocs de pixels (des patches). Ce modèle est devenu très populaire et on le retrouve à présent dans de nombreux travaux [GILBOA07, KERVRA06, PEYR07, SZLAM06].

On trouve deux notions sous-jacentes dans tous ces modèles : le voisinage de traitement et le type de variété à traiter. Nous avons vu dans la section précédente que les graphes sont des structures de données tout à fait adaptées pour ce type de problématique. Nous proposons donc ici un formalisme général de traitement de données d'images de dimensions quelconques sur des graphes de topologies arbitraires. Ces travaux sont dans la continuité de nos premiers travaux de diffusion sur graphes d'adjacence de régions pour la création de hiérarchies de partitions [37, 51] (voir Section 2.4.3).

Parmi les modèles de traitement local, les méthodes variationnelles basées sur une régularisation fournissent un cadre formel qui a été extrêmement étudié pour une grande variété de problèmes de vision par ordinateur. Les solutions de modèles variationnels peuvent être obtenues en minimisant des fonctions d'énergies appropriées et cette minimisation est habituellement réalisée à l'aide d'équations aux dérivées partielles (EDP) [AUBERT02]. Les EDP sont définies en continu et lorsque l'existence et l'unicité d'une solution ont été prouvées, elles sont discrétisées afin d'obtenir une solution numérique [CHAN05B, TANG01, TSCHUM05, WEICKE99]. Nous proposons de nous inspirer de ces travaux définis en continu pour définir un cadre formel de régularisation discrète sur graphes de topologies arbitraires. Ainsi, contrairement aux EDP, notre méthode de régularisation pourra s'appliquer sur des données ayant une organisation spatiale différente d'une grille et elles ne nécessiteront aucun schéma de discrétisation puisqu'exprimées directement en discret. De plus, notre approche permet naturellement d'effectuer des traitements selon n'importe lequel des précédents modèles (local, semi local ou non local) en modifiant simplement la topologie du graphe et les poids des arêtes de celui-ci. Un certain nombre des approches de la littérature peuvent donc être vues comme un cas particulier de notre approche et elle présente également des liens avec la théorie spectrale sur graphe.

Nous formalisons la régularisation discrète d'une fonction  $f^0 \in \mathcal{H}(V)$  par le problème suivant de minimisation sur graphe :

$$\min_{f \in \mathcal{H}(V)} \left\{ E_p(f, f^0, \lambda) = \frac{1}{p} \sum_{v \in V} \|\nabla_v f\|^p + \frac{\lambda}{2} \sum_{v \in V} \|f - f^0\|^2 \right\}$$

où  $p \in [1, +\infty)$  est le degré de régularité,  $\lambda$  un terme d'attache aux données et  $\nabla f$  représente un gradient pondéré de la fonction  $f$  sur un graphe. La solution de ce problème de minimisation amène à la définition d'une famille de filtres non-linéaires. Nous commençons par définir des opérateurs de géométrie différentielle sur graphe, puis nous présentons le problème de régularisation discrète et nous montrons les liens avec d'autres approches. Enfin nous présentons des résultats sur des données de dimensions quelconques réparties sur des graphes de topologies arbitraires.

## 2.3.2 Géométrie différentielle sur graphes

Ce que nous présentons ici est une discrétisation de la géométrie différentielle classique en continu et des définitions similaires ont été proposées pour l'apprentissage semi-supervisé [ZHOU04] ainsi que pour le calcul différentiel sur graphes [BENSOU05].

### 2.3.2.1 Opérateurs gradient et divergence

Nous rappelons tout d'abord que nous considérons que les fonctions  $f \in \mathcal{H}(V)$  associent un vecteur de dimension  $q$  à chaque noeud :  $f : V \rightarrow \mathbb{R}^q$ . L'opérateur de différence  $df : \mathcal{H}(V) \rightarrow \mathcal{H}(E)$  sur un graphe  $\mathcal{G} = (V, E)$  d'une fonction  $f \in \mathcal{H}(V)$  sur une arête  $(u, v)$  reliant deux noeuds  $u$  et  $v$  est défini pour toute arête  $(u, v) \in E$  par

$$(df)(u, v) = (df)_{uv} = \sqrt{w(u, v)} (f(v) - f(u)) \quad (2.1)$$

Dans le cas d'un graphe représentant une grille discrète, la différence revient à  $(f(v) - f(u))$  qui est l'opérateur de différence classique en analyse numérique. La dérivée  $\frac{\partial f}{\partial e} \Big|_v : \mathcal{H}(V) \rightarrow \mathbb{R}^+$  d'une fonction  $f$  au noeud  $v$  selon une arête  $e = (u, v)$  est définie par

$$\frac{\partial f}{\partial e} \Big|_v = (df)(u, v)$$

Nous la noterons également  $\partial_v f(u)$ . Cette définition est cohérente avec la définition continue de la dérivée d'une fonction, par exemple si  $f(v) = f(u)$  alors  $\frac{\partial f}{\partial e} \Big|_v = 0$ . Nous avons donc  $\frac{\partial f}{\partial e} \Big|_v = -\frac{\partial f}{\partial e} \Big|_u$  ( $\partial_v f(u) = -\partial_u f(v)$ ) et  $\partial_u f(u) = 0$ .

Pour une fonction  $f \in \mathcal{H}(V)$  et un noeud  $v$ , le gradient pondéré de  $f$  au noeud  $v$  est l'opérateur vectoriel défini comme  $\nabla : V \rightarrow \mathbb{R}^N$  par

$$\nabla f(v) = \nabla_v f = (\partial_v f(u) : (u, v) \in E, u \sim v)^T \quad (2.2)$$

Ainsi, la norme du gradient pondéré sur graphe notée  $\|\nabla f\|$  en un noeud  $v$  est définie comme  $\|\nabla\| : \mathbb{R}^N \rightarrow \mathbb{R}^+$  par

$$\|\nabla_v f\| = \|\nabla f(v)\| = \sqrt{\sum_{u \sim v} (\partial_v f(u))^2} = \sqrt{\sum_{u \sim v} w(u, v) (f(v) - f(u))^2} \quad (2.3)$$

La norme du gradient mesure la régularité d'une fonction autour d'un noeud. Soit  $\mathcal{R}$  une fonctionnelle sur  $\mathcal{H}(V)$  définie par  $\mathcal{R}_p(f) = \sum_{v \in V} \|\nabla_v f\|^p$ , pour chaque  $p \in [1, +\infty)$ . Cette fonctionnelle  $\mathcal{R}_p$  peut être vue comme une mesure de régularité de  $f$  puisque c'est la somme des variations locales en chaque noeud.

La divergence sur graphe est l'opérateur  $div : \mathcal{H}(E) \rightarrow \mathcal{H}(V)$  qui satisfait

$$\langle df, h \rangle_{\mathcal{H}(E)} = \langle f, -div(h) \rangle_{\mathcal{H}(V)}$$

avec  $f \in \mathcal{H}(V)$  et  $h \in \mathcal{H}(E)$ . Cet opérateur  $-div$  est donc l'opérateur adjoint  $d^*$  de l'opérateur de différence  $d$ . A partir des définitions des produits scalaires dans  $\mathcal{H}(V)$  et  $\mathcal{H}(E)$  et de l'équation (2.1), on peut prouver que la divergence sur graphe d'une fonction  $h \in \mathcal{H}(E)$  en un noeud  $v$  peut s'exprimer comme

$$(d^*h)(v) = (-div(h))(v) = \sum_{u \sim v} \sqrt{w(v, u)} (h(u, v) - h(v, u)) \quad (2.4)$$

### 2.3.2.2 Opérateur p-Laplacien

Le Laplacien sur graphe est un opérateur  $\Delta : \mathcal{H}(V) \rightarrow \mathcal{H}(V)$  défini par  $\Delta f = -div(df) = d^*(df)$ . Le Laplacien sur graphe est un opérateur linéaire car le gradient et la divergence sont des opérateurs linéaires. De plus, le Laplacien est auto-adjoint :

$$\langle \Delta f, g \rangle_{\mathcal{H}(V)} = \langle df, dg \rangle_{\mathcal{H}(E)} = \langle f, \Delta g \rangle_{\mathcal{H}(V)}$$

et positif semi-défini :

$$\langle \Delta f, f \rangle_{\mathcal{H}(V)} = \langle df, df \rangle_{\mathcal{H}(E)} = \mathcal{R}_p(f) \geq 0$$

ce qui implique que

$$\Delta f = \frac{\partial \mathcal{R}_p(f)}{\partial f} \quad (2.5)$$

La courbure sur graphe est un opérateur non linéaire  $\kappa : \mathcal{H}(V) \rightarrow \mathcal{H}(V)$  défini par

$$\kappa f = -div \left( \frac{df}{\|\nabla f\|} \right) = d^* \left( \frac{df}{\|\nabla f\|} \right) \quad (2.6)$$

Nous pouvons généraliser le Laplacien et la courbure par un opérateur qui peut être considéré comme un analogue discret du  $p$ -Laplacien dans le cas continu [TROYAN00]. Le  $p$ -Laplacien sur graphe est un opérateur  $\Delta_p : \mathcal{H}(V) \rightarrow \mathcal{H}(V)$  avec  $p \in [1, +\infty)$  défini par

$$\Delta_p f = -\frac{1}{p} div (\|\nabla f\|^{p-2} df) = \frac{1}{p} d^* (\|\nabla f\|^{p-2} df) \quad (2.7)$$

En conséquence, nous avons  $\Delta_1 = \kappa$  et  $\Delta_2 = \Delta$ . En substituant (2.1) et (2.4) dans la définition (2.7) de  $\Delta_p f$ , nous obtenons alors :

$$(\Delta_p f)(v) = \frac{1}{p} \sum_{u \sim v} \gamma(u, v) (f(v) - f(u)) \quad (2.8)$$

où  $\gamma(u, v)$  est la fonction définie par

$$\gamma(u, v) = w(u, v) (\|\nabla f(v)\|^{p-2} + \|\nabla f(u)\|^{p-2}) \quad (2.9)$$

qui généralise les opérateurs classiques de Laplacien et courbure sur graphe. En général  $\Delta_p$  est non linéaire (sauf pour le cas de  $p = 2$ ) et est positif semi-défini.

Afin d'éviter le cas d'un dénominateur nul pour le calcul de la courbure ( $p = 1$ ), le gradient  $\|\nabla_v f\|$  est remplacé par sa version régularisée  $\|\nabla_v f\|_\beta = \sqrt{\beta^2 + \|\nabla_v f\|^2}$  où  $\beta$  est un paramètre ayant une faible valeur positive. Il est appelé le paramètre de régularisation par CHAN [CHAN01B].

### 2.3.3 Régularisation discrète

Nous proposons un cadre général pour la régularisation sur graphes, un graphe étant une représentation discrète de données dont la dimensionnalité et la topologie sont arbitraires.

#### 2.3.3.1 Principe général

Étant donné un graphe  $\mathcal{G} = (V, E)$  auquel est associé une fonction de pondération  $w : E \rightarrow \mathbb{R}^+$ , nous désirons effectuer la régularisation  $p$ -Laplacienne d'une fonction  $f^0 \in \mathcal{H}(V)$  en utilisant le  $p$ -Laplacien. Cela revient à considérer des problèmes variationnels généraux sur graphes. Le but est alors, étant donné une fonction  $f^0 \in \mathcal{H}(V)$ , de trouver une fonction  $f^* \in \mathcal{H}(V)$  qui est non seulement lisse sur  $\mathcal{G}$  mais suffisamment proche de  $f^0$ . Ce problème d'optimisation peut être formalisé comme la minimisation d'une somme pondérée de deux termes d'énergies :

$$f^* = \min_{f \in \mathcal{H}(V)} \left\{ E_p = \frac{1}{p} \mathcal{R}_p(f) + \frac{\lambda}{2} \|f - f^0\|^2 = \frac{1}{p} \sum_{v \in V} \|\nabla_v f\|^p + \frac{\lambda}{2} \sum_{v \in V} \|f - f^0\|^2 \right\} \quad (2.10)$$

Le premier terme est le terme de régularité qui impose que  $f$  ne change pas trop entre des noeuds proches. Le second terme est l'attache aux données qui impose que  $f$  ne soit pas trop éloignée de  $f^0$ .  $\lambda \geq 0$  est un paramètre de fidélité aux données initiales, appelé multiplicateur de Lagrange, qui spécifie le compromis entre ces deux termes d'énergie. Les deux termes de l'énergie  $E_p$  sont des fonctions strictement convexes de  $f$  [CHAN01B, CHAMBO97], et ce problème a une solution unique pour  $p = 1$  ou  $p = 2$  qui satisfait :

$$\left. \frac{\partial E_p}{\partial f} \right|_v = \frac{1}{p} \frac{\partial}{\partial f} \|\nabla_v f^*\|^p + \lambda (f^*(v) - f^0(v)) = 0, \forall v \in V \quad (2.11)$$

En utilisant la propriété (2.5) du  $p$ -Laplacien pour calculer la dérivée du premier terme de  $E_p$ , l'équation (2.11) peut être réécrite ainsi :

$$(\Delta_p f^*)(v) + \lambda (f^*(v) - f^0(v)) = 0, \forall v \in V \quad (2.12)$$

La solution  $f^*$  de (2.10) est également la solution de (2.12). En substituant l'expression du  $p$ -Laplacien dans (2.12), nous obtenons :

$$\left( \lambda + \frac{1}{p} \sum_{u \sim v} \gamma_{uv} \right) f^*(v) - \frac{1}{p} \sum_{u \sim v} \gamma(u, v) f^*(u) = \lambda f^0(v), \forall v \in V. \quad (2.13)$$

Parmi les méthodes existantes [BRONSH97, HAGEMAS81] qui peuvent être utilisées pour résoudre (2.13), nous utilisons un algorithme itératif de type Gauss-Jacobi [GREENB97, KINCAI02]. Soit  $t$  le numéro d'itération, et  $f^{(t)}$  la solution de (2.13) à l'itération  $t$ . La fonction initiale  $f^{(0)}$  peut être initialisée avec  $f^0$ . L'itération correspondante est donnée par :

$$\begin{cases} f^{(0)} = f^0 \\ \gamma^{(t+1)}(u, v) = w(u, v) \left( \|\nabla f^{(t+1)}(v)\|^{p-2} + \|\nabla f^{(t+1)}(u)\|^{p-2} \right), \forall (u, v) \in E, \\ f^{(t+1)}(v) = \frac{p\lambda}{p\lambda + \sum_{u \sim v} \gamma^t(u, v)} f^0(v) + \frac{\sum_{u \sim v} \gamma^t(u, v) f^t(u)}{p\lambda + \sum_{u \sim v} \gamma^t(u, v)}, \forall v \in V, \end{cases} \quad (2.14)$$

où  $\gamma^{(t)}$  est la fonction  $\gamma(u, v)$  à l'itération  $t$ . Nous définissons la fonction  $\varphi : V \rightarrow \mathbb{R}^+$  par

$$\varphi_{vu}^{(t)} = \frac{\gamma^t(u, v)}{p\lambda + \sum_{u \sim v} \gamma^t(u, v)} \text{ si } u \neq v \text{ et } \varphi_{vv}^{(t)} = \frac{p\lambda}{p\lambda + \sum_{u \sim v} \gamma^t(u, v)} \text{ sinon.}$$

Ainsi, une itération de notre algorithme de régularisation 2.14 peut être réécrite par :

$$f^{(t+1)}(v) = \varphi_{vv}^{(t)} f^0(v) + \sum_{u \sim v} \varphi_{vu}^{(t)} f^{(t)}(u) \quad (2.15)$$

On peut noter que la valeur de  $f(v)$  à une itération donnée ( $t + 1$ ) dépend de deux quantités : la valeur originale de  $f$  au noeud  $v$  (i.e.  $f^0(v)$ ) et une somme pondérée des valeurs à l'itération  $t$  dans le voisinage de  $v$ . Ainsi, itérer (2.15) permet d'obtenir un filtre passe bas dont les coefficients sont mis à jour au fur et à mesure des itérations. En modifiant la topologie du graphe, la représentation associée à chaque noeud du graphe et les poids des arêtes de celui-ci, nous obtenons une famille de filtres non linéaires simples et rapides.

### 2.3.3.2 Représentation associée aux noeuds du graphe

Comme nous l'avons précédemment mentionné, à chaque noeud  $v$  du graphe est associé un vecteur  $f(v)$ . Par exemple, dans le cas d'un graphe de type grille pour une image couleur,  $f(v)$  est le vecteur  $RGB$  associé au pixel. Dans le cas des images couleur  $RGB$  classiques, nous avons donc  $q = 3$ . On peut également considérer par exemple une représentation de type Luminance-Chromaticité [CHAN01A] et restaurer les deux séparément pour ensuite les re-combiner. Dans le cas où les données associées à un noeud sont multi-variées, nous effectuons  $q$  schémas de régularisation indépendants (un par composante). Pour le cas de  $p = 1$ , le  $p$ -Laplacien est alors différent pour chaque composante. Cependant, afin de prendre en compte le couplage existant entre les différentes composantes, les régularisations des composantes ne doivent pas être basées sur des géométries locales indépendantes mais sur une géométrie vectorielle commune. Ceci est nécessaire afin de prendre en compte la corrélation qui existe entre les différentes composantes du vecteur  $f(v)$ . En conséquence, nous considérons la norme suivante

$$\|\nabla f(v)\| = \sqrt{\sum_{i=1}^q \|\nabla f_i(v)\|^2} \text{ où } f_i(v) \text{ désigne la } i^{\text{ième}} \text{ composante de } f(v).$$

### 2.3.3.3 Poids des arêtes du graphe

Même si nous avons associé à chaque noeud  $v$  du graphe un vecteur  $f(v)$ , nous pouvons intégrer des informations supplémentaires (extraites du voisinage de  $v$  ou de  $f(v)$ ) dans le calcul des poids des arêtes. Nous noterons  $F_f(v)$  un vecteur de caractéristiques calculé à partir de  $f(v)$  et de  $f(u)$  tel que  $u \sim v$ . Dans le cas le plus simple, on peut avoir  $F_f(v) = f(v)$ . Plus généralement, à chaque noeud  $v \in V$  du graphe, nous pouvons associer un vecteur d'attributs  $F_f(v) \in \mathbb{R}^m$  avec  $m \in \mathbb{N}^+$ . Ce vecteur peut donc être de dimension quelconque et incorporer des attributs variés décrivant un noeud du graphe (couleur, texture, etc.) à partir de mesures locales, semi locales ou non locales selon la topologie du graphe considéré.

Nous pouvons ensuite associer une fonction de poids à chaque arête du graphe. Cette fonction de poids détermine le type de régularisation induite par  $\mathcal{R}_p(f)$ . Les poids sont positifs et symétriques ; ils permettent de quantifier la similarité entre les données multi-variées associées à deux noeuds reliés par une arête. La formulation générale d'une fonction de poids est alors la suivante  $w(u, v) = g(F_f(u), F_f(v)), \forall (u, v) \in E$ . A titre d'exemples, nous pouvons en considérer plusieurs :

1.  $g_1(F_f(u), F_f(v)) = \frac{1}{\varepsilon + \|F_f(u) - F_f(v)\|}$
2.  $g_2(F_f(u), F_f(v)) = \exp\left(-\frac{\|F_f(u) - F_f(v)\|^2}{\sigma^2}\right)$
3.  $g_3(F_f(u), F_f(v)) = \exp\left(-\frac{\|F_f(u) - F_f(v)\|^2}{\sigma^2}\right) \exp\left(-\frac{\|u - v\|^2}{2\sigma_d^2}\right)$

Pour  $g_1$ , le paramètre  $\varepsilon$  est nécessaire pour éviter une division par zéro et a une très petite valeur ( $10^{-4}$  dans notre cas). Pour  $g_2$  et  $g_3$ , le paramètre  $\sigma$  est généralement fixé *a priori*, mais il est préférable de l'estimer. Nous avons choisi de l'estimer localement pour chaque noeud comme l'écart type mesuré sur l'image originale  $f^0$  [54] autour de celui-ci. Dans  $g_3$ ,  $\sigma_d$  contrôle la décroissance spatiale entre les coordonnées des noeuds.

### 2.3.3.4 Constantes de régularisation

La constante de régularisation  $\lambda$  détermine un compromis entre régularisation et attache aux données. Elle joue un rôle important et plutôt que fixé *a priori*, il est préférable de l'estimer au fur et à mesure des itérations. Une façon d'estimer  $\lambda$  est de considérer le problème équivalent de minimisation sous contraintes [BROOK03] de (2.10) formulé par  $\min_f \sum_{v \in V} \|\nabla_v f\|^p$  sous la contrainte  $\frac{1}{|V|} \sum_{v \in V} \|f - f^0\|^2 = \sigma^2$ . En utilisant (2.5) et (2.8), on obtient, pour une itération donnée  $t$ , une estimation de  $\lambda$  [CHAN01B] :

$$\lambda_t = \frac{1}{\bar{\sigma}_t^2} \frac{1}{|V|} \sum_{v \in V} \left( (f^t(v) - f^0(v)) \sum_{u \sim v} \gamma_{uv} (f^t(v) - f^t(u)) \right). \quad (2.16)$$

Pour  $t = 0$ , nous fixons  $\lambda_0 = \frac{1}{\bar{\sigma}_t^2}$  où  $\bar{\sigma}_t^2$  désigne la variance du bruit estimée sur l'image  $f^t$ . Lorsque  $p = 1$ , le paramètre de régularisation  $\beta$  est nécessaire pour éviter le cas d'un dénominateur nul dans le calcul de la courbure (ce qui intervient donc lorsque  $\|\nabla_v f\| \approx 0$ ), mais cela amène à une approximation de la solution de (2.10). Selon [CHAN01B], la performance de la régularisation induite est indépendante de  $\beta$  tant qu'il conserve une faible valeur. Nous fixons donc  $\beta^2 = \frac{1}{\bar{\sigma}_0^2}$ .

### 2.3.3.5 Critère d'arrêt

Afin de déterminer automatiquement le nombre d'itérations, nous pouvons utiliser un critère de terminaison défini comme  $\frac{\|f^t(v) - f^{t-1}(v)\|}{\|f^t(v)\|} < \varepsilon$ . Ceci permet de stopper les itérations lorsque peu de modifications interviennent sur  $f^t$ .

## 2.3.4 Liens vers d'autres approches

Nous présentons ici les liens qui existent entre notre approche par régularisation et d'autres approches de la littérature. Notre approche définit une famille de filtres anisotropiques linéaires ( $p = 2$ ) et non linéaires ( $p = 1$ ). En modifiant la topologie du graphe et les poids des arêtes, nous obtenons naturellement les trois modèles de traitement que sont les traitements local, semi local et non local. Par exemple, pour le traitement d'image, si l'on considère un graphe de type grille en 8-voisinage avec une fonction de poids dont le vecteur de caractéristiques est  $F_f(v) = f(v)$ , on obtient un traitement local. Si l'on augmente la connectivité des noeuds en les reliant à des noeuds proches spatialement on obtient un traitement semi-local. Enfin, si l'on considère un vecteur caractéristique représentant un patch (un ensemble des valeurs autour du noeud courant), on obtient un traitement non local. Ce traitement non local peut, à l'extrême, être effectué sur le graphe complet pour considérer un traitement complètement non local. En modifiant ensuite la fonction de pondération associée aux arêtes du graphe, nous obtenons plusieurs connexions avec des approches de la littérature.

Prenons tout d'abord le cas de  $p = 2$  et  $\lambda = 0$ . Avec une fonction de poids de type  $g_3$  et  $F_f(v) = f(v)$ , nous obtenons un traitement équivalent au filtre bilatéral. De même, avec une fonction de poids de type  $g_2$ , nous obtenons un traitement équivalent au filtre de YAROSLAVSKY. Si l'on considère un traitement non local avec une fonction de type  $g_2$  et  $F_f(v) = [f(u), v \in B_{v,s}]^T$  un carré centré en  $v$  de taille  $s$ , on obtient un traitement équivalent au filtre non local de BUADES. On peut montrer également le lien qui existe entre le filtrage proposé et le filtrage spectral sur graphe pour  $p = 2$  et  $\lambda = 0$  [COIFMA05, SZLAM06]. Dans ce cas, l'expression (2.15) se réduit à  $f^{(t+1)}(v) = \sum_{u \sim v} \varphi_{vu}^{(t)} f^{(t)}(u)$  où  $\varphi_{vu}$  peut être interprété comme la probabilité de passer d'un noeud  $v$  à un noeud  $u$  en une marche. En prenant la matrice Markov  $P(v, u) = \varphi_{vu}$  et  $F$  la matrice de  $f$ , la précédente équation itérative peut être réécrite comme  $F^{t+1} = PF^{(t)} = P^t F^{(0)}$ . Une façon équivalente de considérer les puissances de  $P$  est de décomposer  $F$  selon les premiers vecteurs propres de  $P$ . Ceci permet de définir une base pour n'importe quelle fonction  $f \in \mathcal{H}(v)$  et celle-ci peut être décomposée selon les  $k$  premiers vecteurs propres de  $P$  par  $f = \sum_{i=1}^{i=k} \langle f, \mathbf{v}_i \rangle_{\mathcal{H}(V)} \mathbf{v}_i$ , ce qui peut être interprété comme un filtrage dans le domaine spectral. Quand  $\lambda \neq 0$  et que les poids sont constants, on obtient un traitement équivalent au EDP discrètes proposées par CHAN [CHAN01B] (pour  $p = 1$ , c'est la régularisation par variation totale et pour  $p = 2$  c'est la régularisation  $L_2$ ). Plus généralement, si les poids ne sont pas constants, cela correspond à une régularisation  $L_2$  ou par variation totale sur des graphes de topologies arbitraires. On remarque que cette régularisation est naturellement non locale en augmentant la connectivité des noeuds du graphe.

### 2.3.5 Applications

Nous présentons ici brièvement les potentialités de la méthode qui permet, comme nous allons le voir, d'effectuer une régularisation non locale sur des images ou des variétés. Nous considérons tout d'abord le cas du filtrage pour  $p = 1$  afin de bien préserver les contours. La figure 2.19 présente des résultats de débruitage sur une image considérée comme un graphe de type grille avec un voisinage en 8-connexité. Selon le type de bruit (impulsionnel ou Gaussien), des fonctions de poids sont préférables. Pour le bruit impulsionnel une fonction de type  $g_1$  est préférable et pour le bruit de type Gaussien on préférera une fonction de type  $g_2$  ou  $g_3$ . Nous avons pu étudier expérimentalement [16] les propriétés de la régularisation proposée dans le cadre du filtrage, et les capacités de notre approche sont équivalentes aux approches classiques tout en étant très simples et efficaces. La figure 2.19(f) présente une restauration basée sur un changement de représentation du vecteur couleur attaché à chaque noeud du graphe où la luminosité et la chromaticité sont restaurés séparément.

Le filtrage que nous avons effectué dans la figure 2.19 est un filtrage local car l'on considère un graphe de voisinage en 8-connexité. Nous allons voir ici les bénéfices d'une approche non locale sur une image réelle comportant du bruit 2.20(a). Nous considérons une fonction de pondération de type  $g_1$  et  $g_2$  opérant en local (figures 2.20(b) et 2.20(c)), une fonction de pondération de type  $g_2$  opérant en non-local en représentation *RGB* (figure 2.20(d) ou bien Luminance-Chromaticité (figure 2.20(e)). Dans le cas du non-local, on considère un voisinage de recherche de type  $7 \times 7$  (chaque pixel est relié à ses 48 voisins directs) avec un vecteur  $F_f$  défini sur une fenêtre de type  $5 \times 5$ . La norme utilisée dans  $g_2$  pour comparer les vecteurs  $F_f$  est  $\|F_f(u) - F_f(v)\|_{2,a}^2$ . Comme cela a été constaté par BUADES [BUADES05A], un filtrage selon un modèle non local permet de mieux préserver les informations textures. C'est donc évidemment le cas également de notre approche par régularisation non locale. La figure 2.21 en présente une illustration. Les paramètres sont les mêmes que précédemment.

Notre méthode peut également facilement s'adapter à d'autres problématiques telles que la retouche d'images pour supprimer des objets indésirables dans des images. Les figures 2.22(a)-2.22(d) présentent quelques résultats de retouche avec  $p = 2$  et  $\lambda = 0$  en non-local. Bien que la méthode ne soit pas pleinement apte à restaurer des grandes zones manquantes texturées, on peut vérifier la qualité des résultats obtenus. On notera surtout que ces résultats ont été obtenus en quelques secondes de traitement pour un traitement non-local, ce qui en fait une méthode de retouche d'images très compétitive. Dans le même ordre d'idée, nous avons considéré une application de coloration d'image en niveaux de gris. Pour cela, nous considérons un ensemble de couleurs germes disposés par l'utilisateur. Ces germes sont diffusés ( $p = 2$  et  $\lambda = 0$ ) dans l'image par une approche locale avec une représentation de type chromaticité-luminance où seule la chromaticité est régularisée. On constate la qualité des résultats obtenus en quelques secondes également. Ce critère de temps est important étant donné le caractère très interactif de la coloration qui nécessite de corriger les germes (en position et couleur) afin d'obtenir le résultat voulu.

De par sa formulation, notre approche permet d'effectuer naturellement une régularisation non locale sur des variétés. A titre d'exemple, nous considérons un ensemble d'images représentant le chiffre manuscrit zéro dans la base USPS. Nous pouvons effectuer une régularisation non locale sur ces variétés en considérant un graphe complet et une distance euclidienne simple entre les images. Chaque image est considérée comme un vecteur de taille 64 qui est régularisé soit 64 régularisations effectuées en parallèle. La figure 2.23(a) présente l'ensemble des images





(a) Image originale.



(b) Image bruitée par 15% de bruit impulsionnel.

(c) Image bruitée par un bruit Gaussien ( $\sigma = 15$ ).(d) Restauration de 2.19(b) avec  $g_1$ .(e) Restauration de 2.19(c) avec  $g_2$ .(f) Restauration de 2.19(c) avec  $g_2$  et une représentation de type Luminance-Chromaticité.

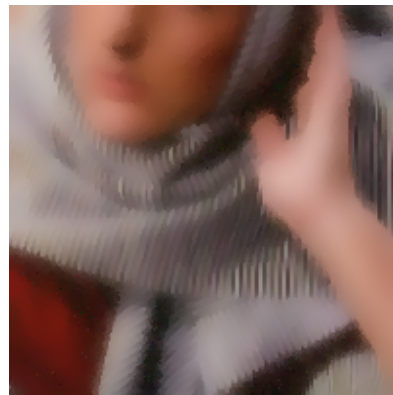
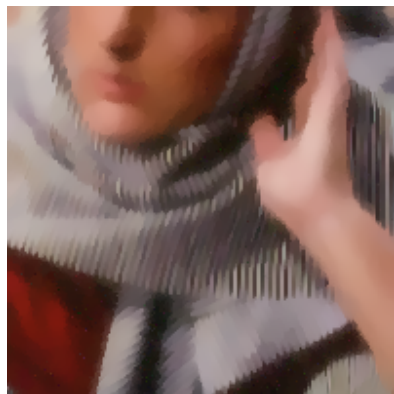
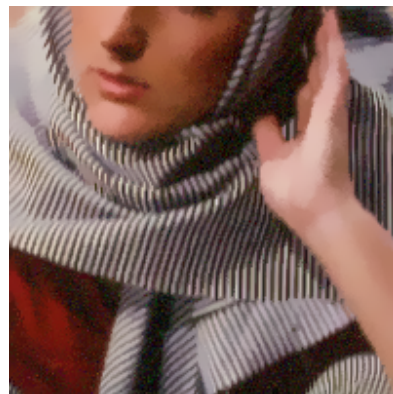
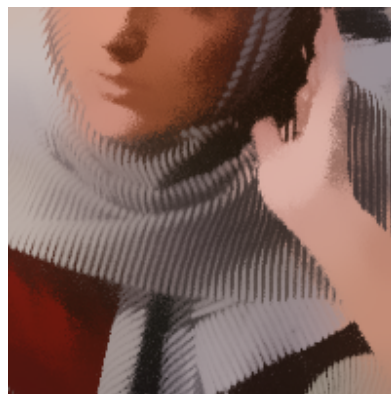
FIG. 2.19 – Exemple des capacités de filtrage pour deux types de bruits et deux fonctions de pondération différentes.



(a) Image originale bruitée.

(b) Image restaurée en local avec  $g_1$ .(c) Image restaurée en local avec  $g_2$ .(d) Image restaurée en non-local avec  $g_2$ .(e) Image restaurée en non-local avec  $g_2$  en Luminance-Chromaticité.

FIG. 2.20 – Exemples de capacités de filtrage sur une image réelle avec une approche locale ou non locale.

(a) Image bruitée ( $\sigma = 15$ ).(b)  $p = 2$ , Régularisation avec  $g_2$  en local.(c)  $p = 1$ , Régularisation avec  $g_2$  en local.(d)  $p = 1$ , Régularisation avec  $g_2$  en non-local (patch  $3 \times 3$ ).(e)  $p = 1$ , Régularisation avec  $g_2$  en non-local (patch  $7 \times 7$ ).FIG. 2.21 – Exemple des bénéfices du modèle de traitement non local pour la préservation de la texture lors d'un filtrage d'une image bruitée par un bruit Gaussien ( $\sigma = 15$ ).



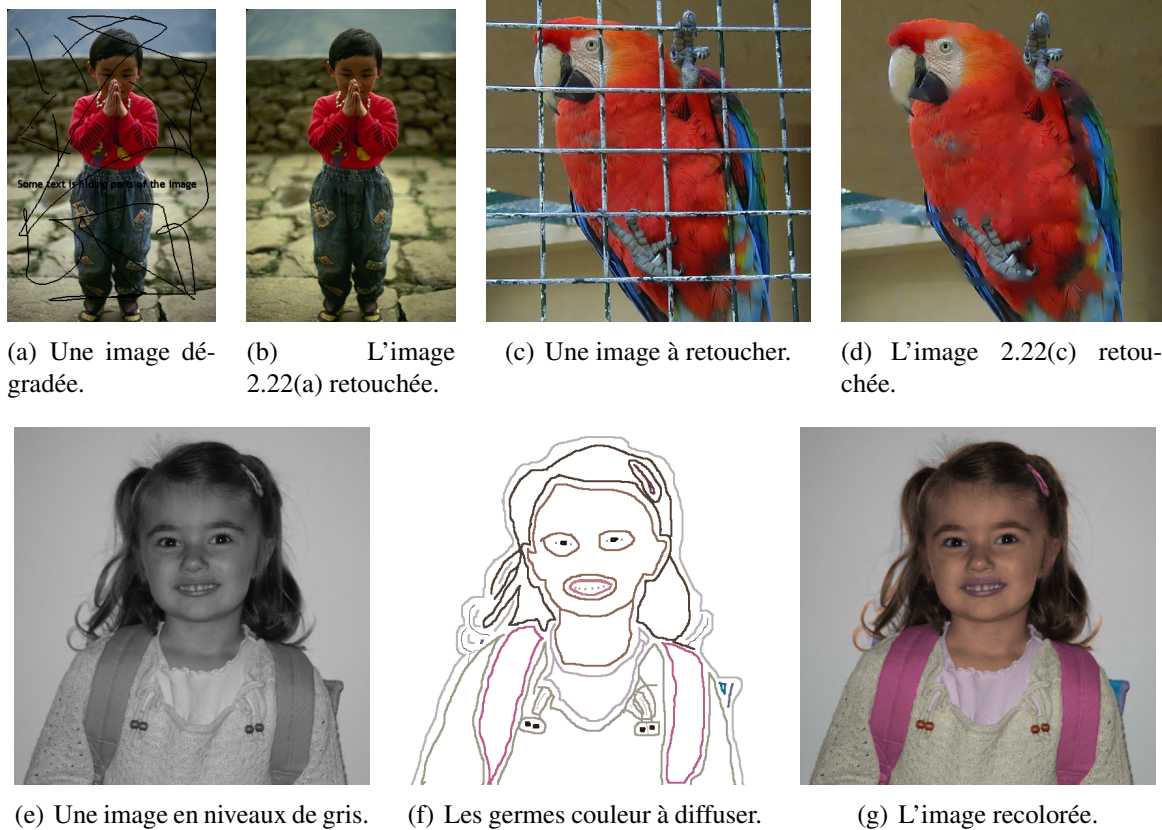


FIG. 2.22 – Exemples de retouche et de coloration d'images.

et sa régularisation non locale ( $p = 2$ ) pour différentes quantités d'attache aux données (figures 2.23(b)-2.23(d)). On remarque bien évidemment que le terme d'attache aux données permet de ne pas trop s'éloigner des données initiales. Notre méthode permet donc de réaliser une régularisation discrète non locale sur des variétés, ce qui n'est le cas d'aucune méthode de la littérature jusqu'à présent.

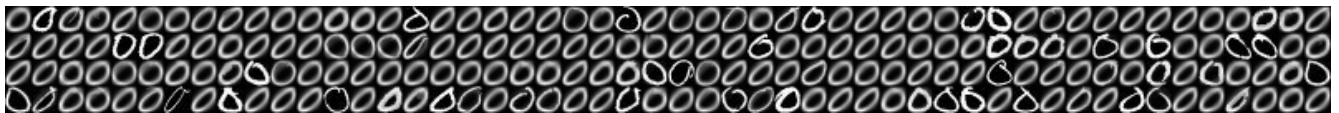
En se basant sur une diffusion simple ( $p = 2$  et  $\lambda = 0$ ), notre approche permet également d'effectuer un apprentissage semi-supervisé c'est à dire d'effectuer de la diffusion de labels sur graphe à partir de quelques éléments labélisés. Nous présentons deux applications de ce principe pour l'extraction d'objets dans les images (Figure 2.24). La première application concerne un graphe de type grille avec deux ensembles de labels initiaux et la seconde application un graphe complet construit sur une pré-segmentation avec trois ensembles de labels initiaux. On remarque qu'avec un graphe complet, notre approche permet de labéliser des éléments qui ne sont pas spatialement connexes dans l'image, ce qui permet de minimiser le nombre de labels initiaux à positionner.

### 2.3.6 Conclusion et perspectives

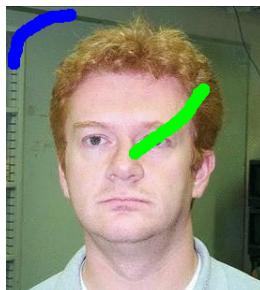
Dans cette section, nous avons proposé une formulation discrète de la régularisation qui repose sur des graphes pondérés de topologies arbitraires qui représentent des données de dimensions quelconques. Nous avons pu montrer les liens de cette approche avec différentes approches de la littérature. Notre approche dispose d'un certain nombre d'attraits par rapport aux approches



(a) Base d'image originale.

(b) Régularisation non locale avec  $\lambda = 1$ .(c) Régularisation non locale avec  $\lambda = 0.01$ .(d) Régularisation non locale avec  $\lambda = 0$ .FIG. 2.23 – Régularisation non locale de variété ( $p = 2$ ).

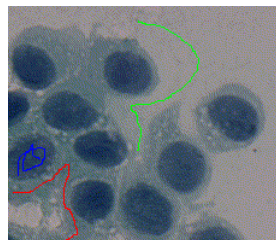
(a) Une image.



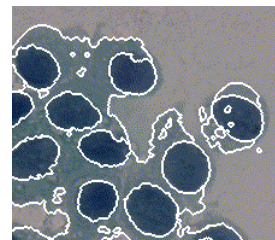
(b) Un ensemble de labels initiaux.



(c) La partition obtenue après diffusion des labels.



(d) Une image de microscopie et trois ensembles de labels.



(e) La segmentation obtenue après diffusion des labels.

FIG. 2.24 – Exemple d'apprentissage semi-supervisé pour la segmentation interactive d'image.

classiques. Premièrement, nous avons une formulation exprimée directement en discret, ce qui permet de considérer des graphes de topologies arbitraires. Comme la topologie peut être arbitraire, nous pouvons naturellement effectuer une régularisation locale, semi locale ou non locale en modifiant simplement la connexité du graphe. En conséquence, nous pouvons alors effectuer une régularisation discrète non locale sur des variétés, ce qu'aucune autre méthode ne permet d'effectuer actuellement. Deuxièmement, la résolution du problème de régularisation est simple à mettre en oeuvre. De plus, avec une formulation locale ou semi locale, la résolution est rapide, ce qui nous a permis d'appliquer notre formalisme à différentes problématiques (restauration, filtrage, retouche, coloration).

Les perspectives de ces travaux sont nombreuses. Tout d'abord, le filtrage de variétés telles que des bases de données (d'images ou non) présente un intérêt certain qui n'a pratiquement jamais été exploré en fouille de données où les données sont considérées brutes et jamais filtrées. Cependant, si l'on considère le traitement d'image, la simplification est un pré-traitement important facilitant la segmentation. On peut donc supposer que la simplification de bases de données devrait également faciliter leur découpage. D'autre part, nous nous intéressons également au traitement de données de grandes dimensions par des approches discrètes et non locales dans une optique de catégorisation de bases de données d'images ou bien extraites d'images. Il s'agira également de s'intéresser à déterminer l'influence de la topologie des graphes utilisés dans le cadre d'une approche non locale. En effet, le graphe complet, même s'il présente l'attrait d'exprimer directement un caractère non local, est d'une plus grande complexité que certains autres graphes locaux ou semi locaux (grille, k-ppv, MST ou autres). Ce problème est généralement totalement éludé dans la littérature et il s'agira de l'étudier.

## 2.4 Hiérarchies de partitions

### 2.4.1 Introduction et Définitions

La segmentation d'images consiste en la délimitation de zones plus ou moins régulières ou cohérentes suivant un critère donné. On cherche donc une partition de l'image en zones dans lesquelles des valeurs suivent un modèle donné d'organisation [ADAMS94]. Segmenter des images de façon automatique est un problème important en traitement et analyse d'images car une segmentation en régions permet d'extraire les composants élémentaires d'une image, ces derniers servant de base pour l'identification d'objets. Il y a cependant un fossé de représentation entre l'image elle-même et sa description sous forme d'une partition en régions. Une façon de contourner ce fossé représentationnel est de se concentrer sur, d'une part, la segmentation en régions, mais également sur les groupements perceptuels et sur une vision hiérarchique des images. L'union de régions forme un groupe qui est à nouveau une région avec des propriétés locales internes et globales de par son voisinage. La plupart du temps, les méthodes de segmentation bas-niveau ne peuvent pas construire directement une bonne partition finale d'une image. Il faut donc avoir recours à des méthodes hiérarchiques dont les hiérarchies de partitions font partie.

Le problème classique de la segmentation automatique est généralement envisagé comme une division de l'image en régions disjointes, le résultat d'un algorithme de segmentation est alors une partition du domaine de l'image [ZUCKER76]. Une image  $\mathcal{I}$  est formée par un ensemble de pixels :  $\mathcal{I} = \{p_1, p_2, \dots, p_n\}$ . Une région  $\mathcal{R}$  est un sous-ensemble de pixels de l'image constituée de  $|\mathcal{R}|$  pixels.

**Definition 5.** Une partition  $\mathcal{P}$  est un ensemble de régions  $\mathcal{P} = \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_k\}$  tel que :

- (1) l'union des régions de la partition donne l'ensemble de départ :  $\mathcal{I} = \bigcup_{i=1}^k \mathcal{R}_i$ ,
- (2) les régions ont une intersection nulle :  $\forall i, j, i \neq j, \mathcal{R}_i \cap \mathcal{R}_j = \emptyset$ .

Parmi les algorithmes permettant de définir une partition d'une image on trouve les algorithmes pour la segmentation hiérarchique qui ne produisent pas seulement une partition mais une hiérarchie ou une pyramide irrégulière de partitions avec des niveaux croissants de partition. Nous pouvons alors définir une relation d'ordre entre deux partitions : une partition  $\mathcal{P}$  est incluse dans une partition  $\mathcal{Q}$  si toute région  $\mathcal{R}_j^{\mathcal{P}}$  est complètement incluse dans une région  $\mathcal{R}_i^{\mathcal{Q}}$ . Ceci permet de définir une hiérarchie de partitions emboîtées d'une image [HOROWI76B]. Soit  $\mathcal{H}$  un ensemble de partitions associé à une image,  $\mathcal{H}$  forme une hiérarchie de partitions s'il est possible d'établir un ordre d'inclusion parmi toute paire d'éléments de l'ensemble  $\mathcal{H}$ .

**Definition 6.** Une hiérarchie de partitions emboîtées d'une image est un ensemble de partitions  $\mathcal{H} = \{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_l\}$  tel que les régions de la partition  $\mathcal{P}_i = \{\mathcal{R}_1^i, \mathcal{R}_2^i, \dots, \mathcal{R}_k^i\}$  sont incluses dans les régions de la partition  $\mathcal{P}_j = \{\mathcal{R}_1^j, \mathcal{R}_2^j, \dots, \mathcal{R}_{k'}^j\}$  avec  $j > i, k > k'$  et  $\mathcal{R}_m^i \subseteq \mathcal{R}_p^j$  ou  $\mathcal{R}_m^i \cap \mathcal{R}_p^j = \emptyset$ .

C'est à dire que deux régions quelconques appartenant à des partitions différentes de la hiérarchie sont soit disjointes soit incluses l'une dans l'autre. On appelle la partition  $\mathcal{P}_i$  le niveau  $i$  de la hiérarchie.  $\mathcal{P}_0$  est la partition la plus fine (que l'on nomme partition fine) et constitue le niveau inférieur de la hiérarchie,  $\mathcal{P}_l$  est la partition la plus grossière et constitue le niveau supérieur de la hiérarchie. Les régions des niveaux inférieurs étant toujours incluses dans les

régions des niveaux supérieurs, une partition de niveau  $i + 1$  peut être obtenue par fusion de plusieurs régions du niveau  $i$  : c'est ce que l'on appelle le principe d'inclusion dans un principe multi-résolution qui permet de définir des pyramides irrégulières.

Pour construire des pyramides irrégulières (i.e. des hiérarchies de partitions), plusieurs approches sont possibles :

- Nous pouvons considérer des approches morphologiques basés sur des critères connectifs qui respectent le principe d'inclusion [SERRA03, SERRA02].
- Nous pouvons nous placer dans un cadre espace-échelle qui consiste à analyser une image à différentes échelles générées par un schéma de diffusion anisotropique par exemple [PERONA90, SUMENG05, VANHAM03]. A chacune de ces échelles, nous pouvons associer une partition de l'image simplifiée. Cependant, appliqué tel quel, ceci ne respecte pas le principe d'inclusion entre les échelles et il est nécessaire d'avoir recours à un recalage d'une échelle à l'autre.
- Nous pouvons également partir d'une partition fine (qui peut être l'image originale) et fusionner progressivement les régions entre elles. Ceci peut se faire soit par des critères de décimation [JOLION92, MEER90], des noyaux de contraction [KROPAT95, BRUN03], soit par des critères de fusion de régions [FELZEN04, GARRID98, SALEMB97, SALEMB00, NOCK04, GUIGUE03C, GUIGUE06].

On notera que plusieurs représentations peuvent être utiles pour construire une hiérarchie de partitions. La plus commune est le graphe d'adjacence de régions [FELZEN04], mais l'on peut également se baser sur le graphe dual [KROPAT95] ou bien des cartes combinatoires [BRUN03]. Nous ne considérerons que la représentation par graphe d'adjacence.

Dans nos travaux, nous avons considéré les trois précédentes approches pour construire des hiérarchies de partitions. Nous présentons nos différents apports dans les sections qui suivent.

## 2.4.2 Un critère connectif : les zones homogènes

☞ **Mes publications associées :** [51, 56, 67, 70]

En morphologie mathématique, le fait d'avoir un ordre entre les partitions implique que la hiérarchie de partitions forme un treillis complet. Les principaux critères morphologiques permettant de définir une hiérarchie de partitions sont basés sur la notion de connexion. Cette notion réside dans la définition d'un critère : une image est segmentée en zones au regard d'un critère donné. La notion de connexion, d'un point de vue mathématique, revient à dire que si l'on considère une famille d'ensembles connexes possédant au moins un point en commun, alors leur union est elle aussi connexe. Ce qui peut être traduit sous la forme suivante :  $\{A_i \text{ connexes}\}$  et  $\{\cap A_i \neq \emptyset\} \Rightarrow \{\cup A_i \text{ connexes}\}$ . Partant de ceci, SERRA [SERRA88] a établi une nouvelle notion de connexion que nous avons reprise ci-dessous [SERRA03, SERRA02] :

**Definition 7.** Soit un espace arbitraire noté  $E$ . On appelle une classe connexe ou connexion  $C$  toute famille dans  $P(E)$  telle que :

- $\emptyset \in C$  et quel que soit un élément ponctuel de  $P(E)$  noté  $x \in E$ ,  $\{x\} \in C$ ,
- pour toute famille  $\{C_i\}$  dans  $C$ ,  $\cap C_i \neq \emptyset$  entraîne  $\cup C_i \in C$

Les principales connexions de segmentation sont les zones plates et la Ligne de Partage des Eaux. Les zones plates d'une image  $\mathcal{I}$  sont les composantes connexes ayant une valeur constante,



ce qui constitue évidemment un critère connectif à seuil de segmentation. MEYER a introduit ensuite la notion de zones quasi-plates [MEYER98].

La Ligne de Partage des Eaux (LPE) est quant à elle un opérateur de croissance de régions définissant une connexion par cheminement basée sur le gradient morphologique d'une image. Les germes de la LPE correspondant aux minima du gradient morphologique. L'algorithme des cascades de LPE [BEUCHE94] (ou « waterfall ») permet de construire une LPE hiérarchique non paramétrique qui procède à une fusion des bassins versants et qui correspond à une hiérarchie de partitions.

Nous proposons à présent un nouveau critère connectif qui est une variation des connexions par cheminements et des connexions par seuil, nous les appellerons les connexions homogènes.

**Definition 8.** Deux points  $p$  et  $q$  appartiennent à une même zone homogène d'une image  $I$  ssi  $\|I(p) - I(q)\| \leq k \times \Psi(\text{Germe}(p))$ , avec  $\text{Germe}(p)$  le pixel germe de la région de  $p$  et  $\Psi(p) = \frac{1}{n_v} \sum_{p_v \in V(p)} \|I(p) - I(p_v)\|$ .

$I(p)$  désigne les données associées au pixel  $p$ ,  $V(p)$  désigne l'ensemble des voisins du point  $p$  et  $n_v$  le cardinal de cet ensemble,  $\|\cdot\|$  est une norme  $\mathcal{L}_2$  et  $k$  un nombre réel correspondant à la finesse de la partition générée.  $\Psi(p)$  étant proche d'une mesure de gradient, les pixels se trouvant dans des zones homogènes (la variation dans leur voisinage est faible) seront traités en premier comme germes de région. Chaque région est composée initialement d'un pixel et croît en agrégrant progressivement les pixels adjacents à celle-ci selon la règle précédente. Ceci se traduit par le fait qu'un pixel  $q$  est agrégé à une région  $\mathcal{R}$  si la distance entre un pixel  $p$  de  $\mathcal{R}$  voisin de  $q$  est  $k$  fois inférieure à l'homogénéité du pixel germe de  $\mathcal{R}$ .  $k$  représente le saut d'homogénéité accepté pour que deux pixels appartiennent à une même région. Nous utilisons une implantation efficace à base de files hiérarchiques. Les zones homogènes produisent des segmentations dont la finesse décroît avec l'augmentation du paramètre  $k$  (la contrainte est relâchée). Une hiérarchie de partitions obtenues pour des valeurs croissantes de  $k$  est bien évidemment non stratifiée car cela diminue le nombre de germes initiaux tout en relâchant la contrainte d'homogénéité.

Si l'on désire construire une hiérarchie de partitions de zones homogènes qui soit stratifiée, il faut se contraindre à respecter le principe d'inclusion des régions entre deux niveaux successifs de la hiérarchie. Tout comme pour l'algorithme des cascades, une façon de réaliser ceci est d'appliquer le principe des zones homogènes sur un graphe d'adjacence de régions obtenu par une partition fine par zones homogènes. Ceci nous amène à reformuler le principe des zones homogènes sur graphe qui permet de définir une partition d'un graphe.

**Definition 9.** Deux noeuds  $v_i$  et  $v_j$  d'un graphe  $G$  appartiennent à une même zone homogène si  $\|f(v_i) - f(v_j)\| \leq k' \times \Psi(\text{Germe}(v_i))$  avec  $\text{Germe}(v_i)$  le noeud germe de la région de  $v_i$  et  $\Psi(v) = \frac{1}{\delta(v)} \sum_{u \sim v} \|f(v) - f(u)\|$

De la même manière que pour les zones homogènes définies sur une image, tous les noeuds du graphe sont considérés comme des germes et sont enfilés dans une file de priorité selon  $\Psi(v)$ .  $f(v)$  est un vecteur associé à chaque noeud du graphe ; cela sera un vecteur couleur moyen calculé à partir de la partition initiale. L'algorithme de construction d'une hiérarchie de partitions de zones homogènes est alors donné par l'algorithme 3.

```

 $\alpha$  : entier ;  $k$  : réel ;  $k'$  : réel ;
 $\alpha \leftarrow 1$  ; Définir  $\alpha_{end}$ 
 $P_\alpha \leftarrow$  Zones homogènes de finesse  $k$  de l'image initiale.
 $G_\alpha = (V_\alpha, E_\alpha)$  pour une partition initiale  $P_\alpha$ .
Tant que ( $\alpha \leq \alpha_{end}$ ) faire
     $G_{\alpha+1} \leftarrow$  Zones homogènes de finesse  $k'$  de  $G_\alpha$ 
     $\alpha \leftarrow \alpha + 1$ 
Fait

```

Algorithme 3: Hiérarchie de partitions par zones homogènes.

L'algorithme prend deux paramètres  $k$  et  $k'$ .  $k$  définit la finesse de la partition initiale et  $k'$  définit la finesse des partitions successives de la hiérarchie. Le réglage de  $k$  et  $k'$  est primordial pour la performance de l'algorithme. La figure 2.25 présente quelques niveaux d'une hiérarchie de partitions par zones homogènes stratifiées avec  $k = 0$  et  $k' = 0.5$ . Le premier niveau est très sur-segmenté et les suivants permettent d'extraire les principaux composants visuels de l'image en décimant le graphe.

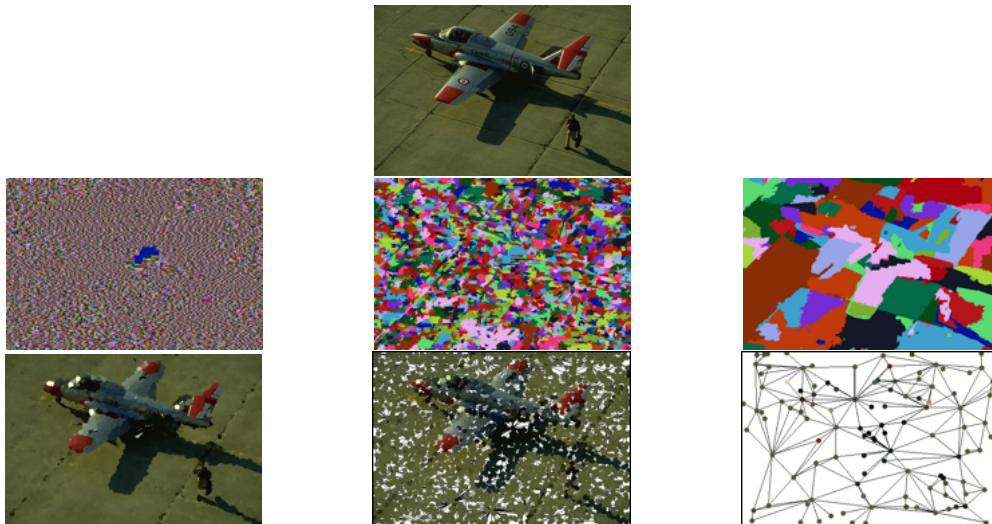


FIG. 2.25 – Une hiérarchie de partitions par zones homogènes stratifiées (niveaux 0, 1 et 10). La première ligne présente l'image originale. La deuxième ligne présente les cartes de régions obtenues. La dernière présente les graphes d'adjacence de régions où chaque région est représentée par un disque coloré par la couleur moyenne de la région qu'il représente.

Le comportement des zones homogènes se situe entre les connexions à seuils (zones quasi-plates) et les connexions par cheminement (ligne de partage des eaux). Cependant, elles sont beaucoup plus adaptées à une utilisation dans le cadre de segmentations automatiques et notamment celles induites par les hiérarchies de partitions. Ceci pour plusieurs raisons : elles ne nécessitent pas de marqueurs, elles fonctionnent sur l'image et non sur son gradient morphologique ce qui ne pose pas de problème pour l'utilisation de la couleur [HANBUR01] et ceci quel que soit l'espace couleur utilisé. De plus, contrairement aux zones quasi-plates, les zones homogènes s'adaptent localement au contenu de l'image car le seuil d'appartenance à une région dépend de l'homogénéité mesurée en ce point. Elles préservent également les structures

finies qui peuvent être détruites lors du calcul d'un gradient. La figure 2.26 présente plusieurs niveaux de hiérarchies de partitions produites par zones quasi-plates, zones homogènes stratifiées et par cascades. La carte de saillance (dernière colonne de la figure 2.26) illustre l'importance de chaque pixel tout au long des niveaux. La saillance d'un pixel est définie comme le plus haut niveau auquel ce pixel est à la frontière entre deux régions. On constate bien alors le caractère intermédiaire des hiérarchies par zones homogènes stratifiées par rapport aux critères connectifs classiques. Une carte de saillance extraite par l'une ces méthodes peut être interprétée comme un détecteur de contours basé régions [ARBELA04]. Nous avons comparé expérimentalement le comportement des zones homogènes comparativement aux deux approches classiques de hiérarchies de partitions que sont les zones quasi-plates et la ligne de partage des eaux. Afin d'avoir une réelle validation expérimentale, nous avons mené nos tests sur 200 images de la base d'images naturelles de Berkeley. On notera qu'habituellement, les approches morphologiques proposées dans la littérature le sont sans validation expérimentale à grande échelle [ANGULO07]. Nous avons alors pu constater que les zones homogènes sont de très bons candidats pour la segmentation automatique d'images [56, 70]. Elle sont d'ailleurs à présent utilisées comme composant de base de segmentation automatique en vidéo-surveillance dans le cadre du projet BOSS (on Board Wireless Secured Video Surveillance) à l'INRETS [56].

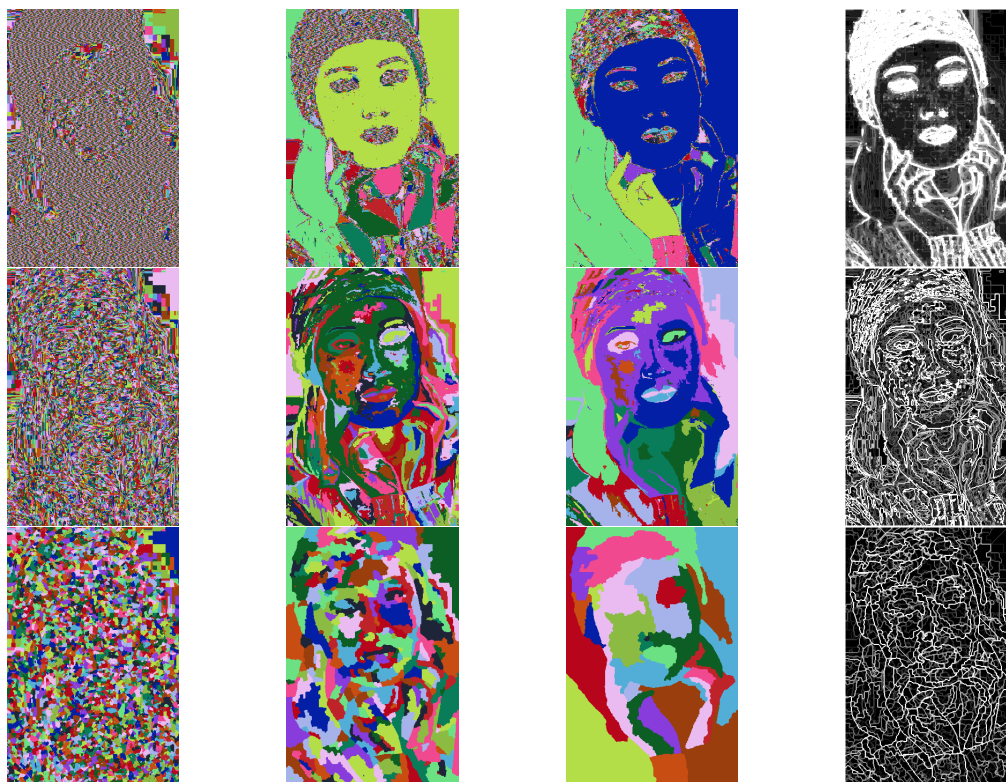


FIG. 2.26 – Hiérarchies de partitions par les zones quasi plates (ligne du haut), zones homogènes (ligne du milieu) et cascades (ligne du bas). Les partitions présentées correspondent aux niveaux 1, 5 et 10 de la hiérarchie (de gauche à droite) et la dernière colonne présente la carte de saillance correspondante sur les dix niveaux.

### 2.4.3 Diffusion-Fusion

☞ Mes publications associées : [16, 37, 51, 67]

Comme nous l'avons vu, une partition fine est une partition sur-segmentée d'une image qui est très proche du contenu original de l'image. Dans le cas où l'on considère les zones strictement plates, l'image mosaïque associée à la partition fine correspond exactement à l'image originale, mais elle permet d'en obtenir une représentation plus compacte (moins de régions que de pixels). De plus, l'image mosaïque (à chaque pixel est associé la couleur moyenne de la région à laquelle il appartient) est une représentation simplifiée d'une image. Il semble donc plus intéressant de travailler directement au niveau région plutôt qu'au niveau pixel. Partant d'une partition fine, nous pouvons associer un modèle à chaque noeud du graphe d'adjacence associé à une région de la partition fine. Nous considérons ici le modèle le plus simple possible à savoir la couleur moyenne de la région.

Habituellement, dans une approche espace-échelle, on travaille au niveau pixel et l'image est simplifiée par diffusion [VANHAM03], ce qui a généralement comme défaut de déplacer les frontières des régions à travers les échelles lorsqu'on leur associe une partition. Nous préférons donc considérer une représentation ensemble-échelle [GUIGUE06] qui consiste à travailler au niveau région et à effectuer un schéma de diffusion sur les régions. La diffusion opérera donc directement sur le graphe d'adjacence de régions d'une partition fine et non sur l'image originale. Ceci revient à simplifier les modèles associés aux régions et permet d'obtenir un ensemble de graphes simplifiés à différents niveaux d'échelle. Pour simplifier un graphe d'adjacence, nous effectuons une régularisation discrète sur ce graphe avec  $p = 2$  et  $\lambda = 0$ , ce qui revient à une diffusion linéaire sur le graphe. La figure 2.27 présente un exemple de graphes simplifiés



FIG. 2.27 – Un ensemble d'images simplifiées dans une approche ensemble-échelle sur un graphe d'adjacence de régions après 0, 5, 15, 50 et 200 itérations (de gauche à droite) sur une image (première ligne) et une version bruitée de celle-ci. La partition initiale est obtenue par les zones strictement plates.

à différents niveaux d'échelles. Nous présentons l'image mosaïque pour illustrer le résultat de la simplification. La partition fine initiale correspondait ici aux zones strictement plates c'est à dire que l'image mosaïque de la partition initiale correspond exactement à l'image originale. Les images mosaïques présentées correspondent à une simplification du graphe après 0, 5, 15, 50 et 200 itérations. Nous présentons également le même traitement sur la même image originale bruitée par un bruit gaussien ( $\sigma = 5$ ). On remarque tout d'abord qu'effectuer une diffusion sur le graphe permet effectivement d'obtenir une simplification à différents niveaux d'échelles. De plus, effectuer une simplification sur le graphe d'adjacence est plus rapide que sur le graphe

grille car il contient moins de noeuds ce qui en fait une alternative intéressante à la simplification d'image classique. Enfin, la simplification proposée ne déplace pas les frontières des régions à travers les échelles (puisque nous considérons le graphe d'adjacence) lorsqu'on lui associe une partition sur graphe contrairement à une approche espace-échelle classique qui ne respecte pas le principe d'inclusion.

Comme nous le constatons, simplifier le graphe d'adjacence permet d'obtenir facilement une représentation simplifiée des modèles associés à la partition fine initiale. Il semble alors naturel de coupler cette simplification avec une approche par fusion de régions [MAKROG01] pour simplifier également la structure du graphe à travers les échelles. En effet, puisque les modèles associés aux noeuds sont simplifiés à travers les échelles, des régions similaires tendent vers des modèles similaires et elles peuvent alors être fusionnées. Ceci permettra d'accélérer à nouveau le traitement puisque la simplification opérera sur un graphe décimé après fusion des régions similaires.

Nous considérons que deux noeuds  $v_i$  et  $v_j$  peuvent fusionner si un critère de fusion  $C(v_i, v_j)$  est vérifié pour ces deux noeuds. Pour générer une hiérarchie de partitions nous alternons donc simplification du graphe par diffusion et fusion des régions similaires. L'algorithme complet est décrit par l'algorithme 4 où  $\alpha$  est le niveau de la hiérarchie, on passe d'un niveau au suivant seulement si au moins deux régions ont fusionné. Lorsque deux régions fusionnent, le modèle de la nouvelle région correspond à l'union des deux modèles des régions qui fusionnent.

```

 $\alpha$  : entier ;  $\alpha_{end}$  : entier ;
 $\alpha \leftarrow 0$  ; Définir  $\alpha_{end}$ 
 $G_\alpha = (V_\alpha, E_\alpha)$  pour une partition fine initiale  $P_\alpha$ .
Tant que ( $\alpha \leq \alpha_{end}$ ) faire
    Pour les noeuds  $v_i \in V_\alpha$  faire
        | Simplifier les modèles de noeuds  $v_i$ .
    Fin Pour
    Pour les arêtes  $E_l = (v_i, v_j) \in V_\alpha \times V_\alpha$  faire
        | Si ( $C(v_i, v_j)$ ) Alors
            | | Ajouter  $E_l$  au noyau de contraction  $NC_{\alpha, \alpha+1}$ 
        Fin Si
    Fin Pour
    Si ( $|NC_{\alpha, \alpha+1}| > 0$ ) Alors
        | Contracter le graphe  $G_\alpha$  à l'aide du noyau de contraction
        |  $NC_{\alpha, \alpha+1} : G_{\alpha+1} = Contraction[G_\alpha, NC_{\alpha, \alpha+1}]$ 
        |  $\alpha \leftarrow \alpha + 1$ 
    Fin Si
Fait

```

Algorithme 4: Algorithme pour la création d'une hiérarchie de partition par diffusion-fusion sur un graphe d'adjacence.

Selon le critère de fusion  $C(v_i, v_j)$ , nous pouvons obtenir différentes hiérarchies de partitions. Nous avons considéré trois types de critères : un seuil fixe, un seuil évolutif et un seuil adaptatif. Pour un seuil fixe, le critère de fusion est  $\|f(v_i) - f(v_j)\| < T$  où  $f(v)$  associe un vecteur couleur à chaque noeud. Pour un seuil évolutif, le critère de fusion est le même, mais le

seuil augmente au fur et à mesure des itérations par  $T = T + \Delta T$ . Pour un seuil adaptatif, nous considérons la formulation statistique de Nock [NOCK04]. La figure 2.28 présente plusieurs hiérarchies de partitions obtenues avec l'approche proposée avec différents critères de fusion pour les niveaux 1, 4, 9, 15 et 20. Le seuil fixe  $T$  vaut 1 et  $\Delta T = 0.5$  pour le seuil évolutif. Pour chaque hiérarchie sont présentés la partition obtenue, l'image mosaïque et la carte de saillance (dernière ligne de la figure 2.28). La partition initiale a été générée par le critère connectif des zones homogènes avec  $k = 0.5$  présenté dans la première colonne de la septième ligne de la figure 2.28. Avec un seuil fixe, le nombre de niveaux de la hiérarchie de partitions est important alors que pour les autres critères, des segmentations plus grossières sont obtenues rapidement à travers les niveaux. Les cartes de saillance nous montrent que les principaux composants visuels de l'image sont bien extraits. Ainsi, contrairement à d'autres approches qui sont basées sur une approche espace-échelle et segmentation [SUMENG05], notre méthode permet de regrouper dans un même principe la simplification et la fusion directement sur le graphe d'une partition fine. A l'extrême, la partition fine initiale pourrait très bien être le graphe grille en 8-voisinage, mais considérer une partition fine est plus intéressant en terme de rapidité. L'approche que nous proposons peut être considérée comme une alternative à la détection de contours basée région proposée dans [ARBELA04, SUMENG05].



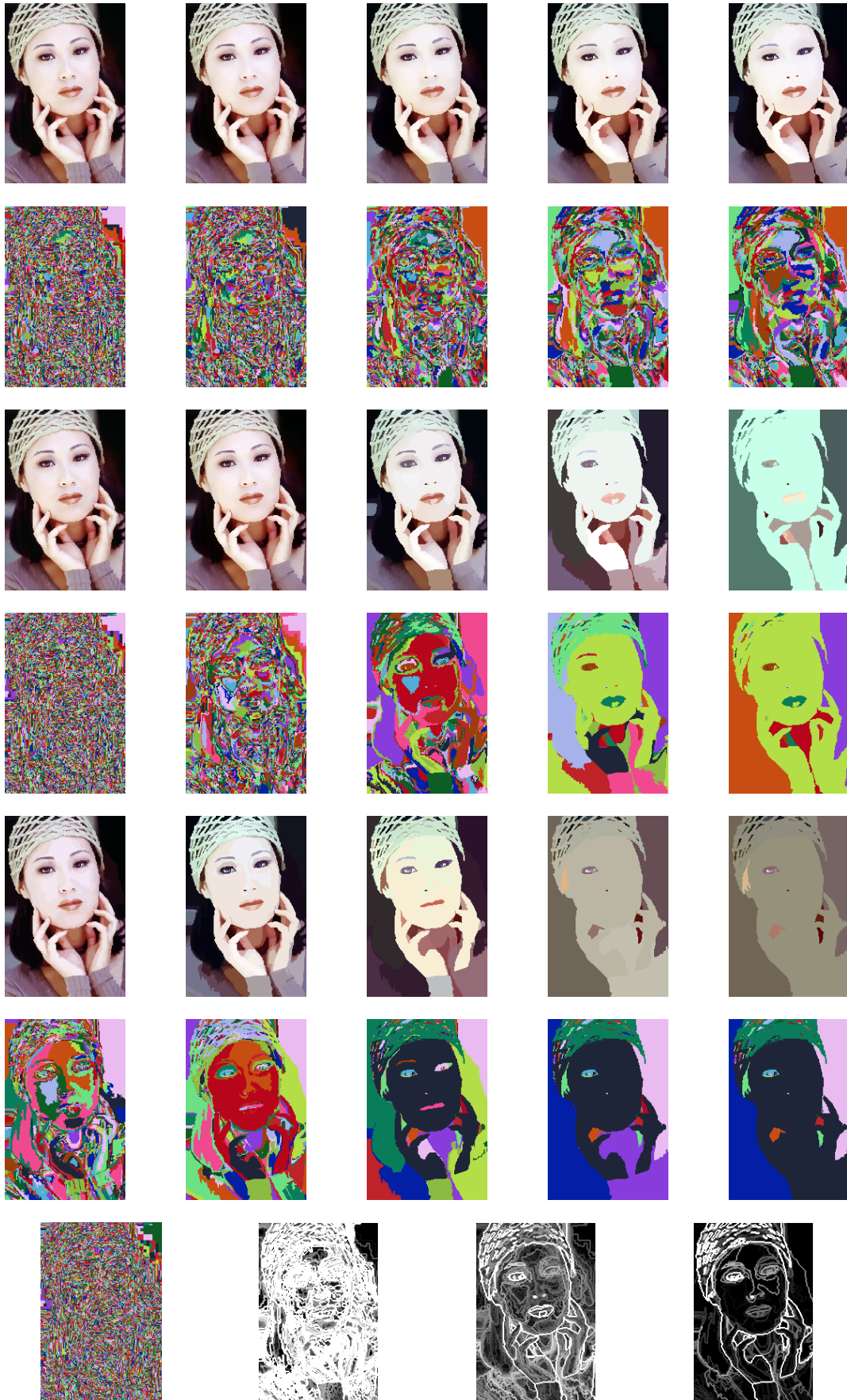


FIG. 2.28 – Hiérarchies de partitions produites par diffusion-fusion sur une partition fine initiale obtenue par les zones homogènes. Les deux premières lignes présentent l'image mosaïque et la partition avec un seuil fixe. Les lignes suivantes présentent les mêmes résultats obtenus avec des seuils évolutif et adaptatif. La dernière ligne présente la partition fine et les cartes de saillance pour les différents critères de fusion.

### 2.4.4 Coalescence-Fusion

☞ **Mes publications associées :** [15, 22, 23, 35, 36, 41, 66]

Nous décrivons ici une approche hybride de la segmentation d'images par génération d'une hiérarchie de partitions. Cette approche repose sur deux éléments : une coalescence morphologique non supervisée permettant de construire une partition fine et une fusion des régions de la partition fine. Nous introduisons également une mesure d'énergie permettant de s'affranchir des paramètres de la méthode. La mesure d'énergie proposée est alors vue comme un critère d'évaluation des partitions obtenues. L'obtention de la partition fine est réalisée par une approche espace-échelle dans l'espace colorimétrique avec détermination automatique du niveau d'échelle requis via la mesure d'énergie. Cette mesure d'énergie est réutilisée afin de déterminer un bon niveau d'échelle lors de la fusion des régions de la partition fine obtenue.

#### 2.4.4.1 Coalescence morphologique non supervisée

En traitement d'images, la majorité des méthodes de coalescence colorimétriques (i.e. clustering) travaillent soit sur des projections 1D des images (des histogrammes) [BUSIN04, CELENK90] soit directement dans l'espace colorimétrique 3D [POSTAI93B, GÉRA01, COMANI02]. On retrouve alors les oppositions habituelles entre approches marginale et vectorielle. Nous avons proposé une approche intermédiaire basée sur des projections bivariées (des histogrammes 2D) [MATAS95A, KURUGO01, XUE03, MACAIR06]. Le principal intérêt des histogrammes bivariées repose dans le fait que ce ne sont rien d'autre que des images en niveaux de gris de tailles réduites ( $256 \times 256$ ) et leur traitement peut se faire par des approches classiques. Un histogramme 2D est une projection d'un histogramme 3D sur chacune des paires de composantes. La figure Figure 2.29 présente une image couleur artificielle (voir les travaux de Macaire et al. [MACAIR06]) et ses projections 3D et 2D (colorées pour une meilleure visualisation).

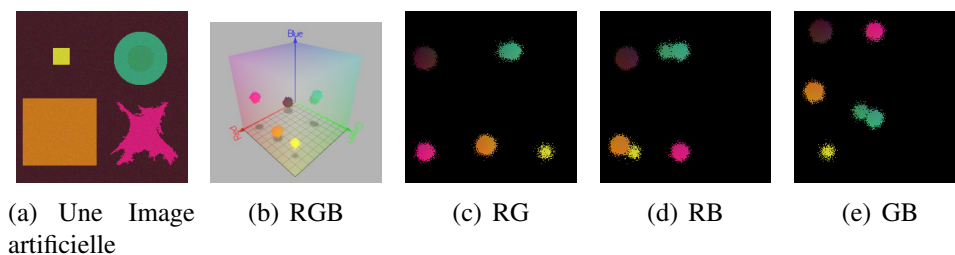


FIG. 2.29 – Une image couleur artificielle et ses projection 3D et 2D.

Nous avons choisi une approche morphologique pour effectuer la coalescence d'un histogramme 2D. Effectuer une coalescence par une approche morphologique a été proposé par Soille pour le cas des histogrammes 1D [SOILLE96, SOILLE04]. Comme beaucoup d'approches de coalescence, un paramètre *a priori* est nécessaire, à savoir le nombre de classes. Afin de nous affranchir de ce nombre, nous préférons rechercher les clusters dominants d'un histogramme 2D par une approche morphologique dérivée de celle proposée par Soille.

La coalescence d'un histogramme 2D  $\mathcal{H}$  se fait en plusieurs étapes. Il est tout d'abord simplifié à l'aide d'un filtre exponentiel symétrique : un nouvel histogramme lissé  $\varphi_\beta(\mathcal{H})$  est obtenu où  $\beta$  spécifie le degré de simplification. Celui-ci est reconstruit dans l'histogramme original afin



d'obtenir une version simplifiée mais proche de l'histogramme original :  $\mathcal{H}^{(2)} = \psi(\varphi_\beta(\mathcal{H}), \mathcal{H})$  où  $\psi$  désigne la reconstruction. Les clusters dominants sont déterminés par le calcul des érodés ultimes :  $\mathcal{H}^e = \zeta(\mathcal{H}^{(2)})$  où  $\zeta$  désigne l'extraction de l'ensemble des érodés ultimes. Ceci permet d'extraire de manière non supervisée les clusters dominants d'un histogramme 2D sans *a priori* sur leur nombre. Les érodés ultimes représentent les centres des classes et pour obtenir une coalescence finale de l'histogramme 2D, une ligne de partage des eaux est effectuée sur l'image de distance aux érodés ultimes avec comme marqueurs ces derniers :  $W_{\mathcal{H}} = \text{Watershed}(\mathcal{H}^e, \delta(\mathcal{H}^e))$  avec  $\delta$  la fonction distance. Toutes ces étapes sont décrites par la figure 2.30 pour l'histogramme  $RB$  de l'image de la figure 2.29(a). On remarque que la coalescence est très efficace, même dans des régions où les clusters sont relativement mélangés.

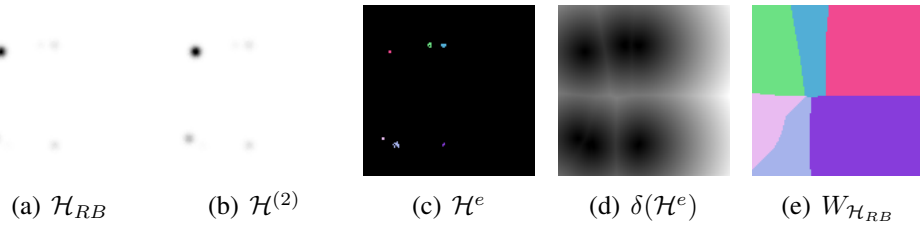


FIG. 2.30 – Illustration des étapes de la coalescence morphologique non supervisée sur l'histogramme 2D  $RB$  de la figure 2.30(a). Les figures 2.30(a) et 2.30(b) sont inversées.

A partir d'une partition d'un histogramme 2D, nous pouvons obtenir une coalescence de l'image originale en effectuant la transformation inverse de celle de la construction de l'histogramme. Comme pour une image couleur, nous disposons de trois histogrammes 2D, trois coalescences sont obtenues. Ceci est décrit par les figures 2.31(a)-2.31(c). On peut alors constater que dans la coalescence de l'image obtenue par la coalescence de l'histogramme 2D  $RG$ , les deux clusters des pixels de couleur verte n'ont pas été dissociés. On remarque également que les numéros des clusters ne correspondent évidemment pas d'une partition à l'autre car leur nombre diffère. Nous avons également utilisé cette coalescence pour la segmentation d'images de microscopie de lavages bronchio-alvéolaires [66].

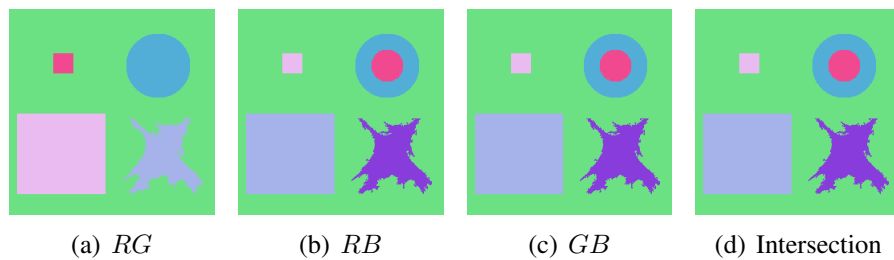


FIG. 2.31 – Coalescences obtenues ((a)-(c)) à partir de la figure 2.29(a) pour les trois projections bivariées. La figure (d) présente l'intersection des figures (a)-(c).

#### 2.4.4.2 Sélection du niveau d'échelle de la coalescence

La précédente méthode de coalescence morphologique est non supervisée car elle ne nécessite pas de préciser le nombre de classes voulues. Cependant, le nombre de clusters dépend fortement du paramètre  $\beta$  de lissage de l'histogramme. Faire varier ce paramètre permet de générer un ensemble de coalescences (non emboîtées) selon une approche espace-échelle. Plus le

lissage de l’histogramme est important, moins le nombre de clusters est important. Plutôt que de fixer *a priori* ce paramètre, nous préférons en déterminer une bonne valeur selon le contenu de l’image à segmenter. Pour cela, nous cherchons à évaluer la partition obtenue de l’image. On notera la différence essentielle entre une partition et une coalescence de l’image : une partition de l’image correspond à un étiquetage des composantes connexes de la coalescence de l’image. Pour régler le paramètre  $\beta$ , nous devons être capable de mesurer la qualité d’une partition. Nous avons choisi de réaliser ceci à l’aide d’une mesure d’énergie [KOEPL94]. Une mesure d’énergie associée à une partition  $\mathcal{M}$  d’une image  $\mathcal{I}$  est définie comme suit [GUIGUE03B, POGGIO85]

$$E : \mathcal{M} \times \mathcal{I} \rightarrow \mathbb{R}^+$$

Comme il n’existe pas de mesure universelle de la qualité d’une partition d’une image, la qualité d’une partition peut être vue comme un compromis entre la qualité et la complexité du modèle (la partition  $\mathcal{M}$ ). La mesure d’énergie est alors décomposée en deux termes :

$$E(\mathcal{M}, \mathcal{I}) = D(\mathcal{M}, \mathcal{I}) + C(\mathcal{M}, \mathcal{I})$$

La qualité de la partition exprime la fidélité aux données initiales et la complexité du modèle décrit son caractère simple ou non. Beaucoup de fonctions d’énergies peuvent se mettre sous cette précédente forme [BOYKOV01, MUMFOR89], la complexité du modèle n’étant pas toujours dépendante de l’image (par exemple la longueur des contours des régions). De manière similaire à PHILIPP-FOLIGUET [PHILIP06], nous proposons une mesure d’énergie permettant de mesurer la qualité d’une partition comme un compromis entre fidélité aux données initiales et complexité du modèle de la partition, ce dernier terme dépendant également de l’image originale. Nous considérons une partition comme un graphe d’adjacence de régions et définissons la mesure d’énergie comme suit :

$$E(\mathcal{M}, \mathcal{I}) = \alpha_d \times \sum_{v \in V} \mathcal{D}(\mathcal{M}, \mathcal{I}, v) + \alpha_c \times \sum_{v \in V} \left[ \frac{1}{|u \sim v|} \sum_{u \sim v} \|f_{\mathcal{M}}(u) - f_{\mathcal{M}}(v)\| \right] \quad (2.17)$$

avec  $\alpha_c = 100 \times \frac{\sqrt{|V|}}{h \times w}$  et  $\alpha_d = \frac{1}{|E| \times |V|}$ .  $h \times w$  désigne la taille de l’image  $\mathcal{I}$ .  $\mathcal{D}(\mathcal{M}, \mathcal{I}, v)$  permet de quantifier la fidélité d’une région  $v$  de la partition  $\mathcal{M}$  par rapport à l’image originale. Ceci est réalisé à l’aide d’un modèle gaussien [GUIGUE03A]. La distance entre deux régions  $\|f_{\mathcal{M}}(u) - f_{\mathcal{M}}(v)\|$  est réalisée à l’aide d’une distance de Mahalanobis.

Le terme de fidélité est similaire à ceux employés pour les « Markov Random Field » [LI95, BOYKOV01] auquel nous avons ajouté le facteur de normalisation de LIU [LIU94] afin de pénaliser les modèles ayant trop de régions. Le terme de complexité est lié à la complexité du graphe d’adjacence de régions relativement à l’image originale. Ces deux termes sont antagonistes et une valeur faible de la mesure d’énergie traduit un bon compromis entre eux. En conséquence, trouver un minimum local de  $E(\mathcal{M}, \mathcal{I})$  permet de trouver une partition qui n’est ni trop fine, ni trop grossière (relativement à l’ensemble des segmentations à évaluer).

A présent que nous disposons d’une mesure de la qualité d’une partition, nous pouvons nous affranchir du choix du paramètre  $\beta$  de la coalescence morphologique. Pour cela, nous générons progressivement des partitions de plus en plus grossières en faisant varier  $\beta$  et nous stoppons dès l’apparition d’un minimum local. Cela permet de trouver automatiquement le « meilleur » niveau de simplification du processus de coalescence grâce à une mesure de qualité effectuée dans le domaine de l’image et non dans le domaine colorimétrique. Pour le cas des images couleur, nous avons trois histogrammes 2D et la détermination du paramètre  $\beta$  est réalisée de manière indépendante pour chaque projection. Ceci permet d’adapter la coalescence à chaque histogramme 2D.

### 2.4.4.3 Intersection et simplification par fusion de régions

A partir d'une image couleur, nous obtenons trois coalescences distinctes et donc trois partitions différentes puisque les histogrammes ne portent pas forcément les mêmes informations du fait de la projection. Nous combinons les trois partitions par une simple intersection [XUE03] (figure 2.31(d)). L'intersection présente évidemment le défaut d'être sur-segmentée. La figure 2.32(e) présente l'intersection obtenue pour une image réelle (figure 2.32(a)) à partir des trois partitions obtenues par coalescence (figure 2.32(b) à 2.32(d)). Puisque la coalescence est effectuée dans le domaine colorimétrique de l'image, elle néglige le contexte spatial et l'intersection est sur-segmentée : des clusters dans l'espace colorimétrique ne correspondent généralement pas à des régions connexes dans l'image. Pour contrer cet effet, il nous faut simplifier la partition obtenue par intersection. Une manière classique de réaliser ceci est de fusionner les régions deux à deux itérativement [MAKROG05], ce qui permet de générer une hiérarchie de partitions modélisée sous la forme d'un arbre binaire de régions [SALEMB00]. Effectuer une fusion de régions nécessite d'exprimer plusieurs points : le modèle associé à chaque région, une mesure de distance entre régions et un critère de terminaison. Le modèle associé à chaque région est un modèle Gaussien et la distance entre deux régions est réalisée par celle de Mahalanobis. Si

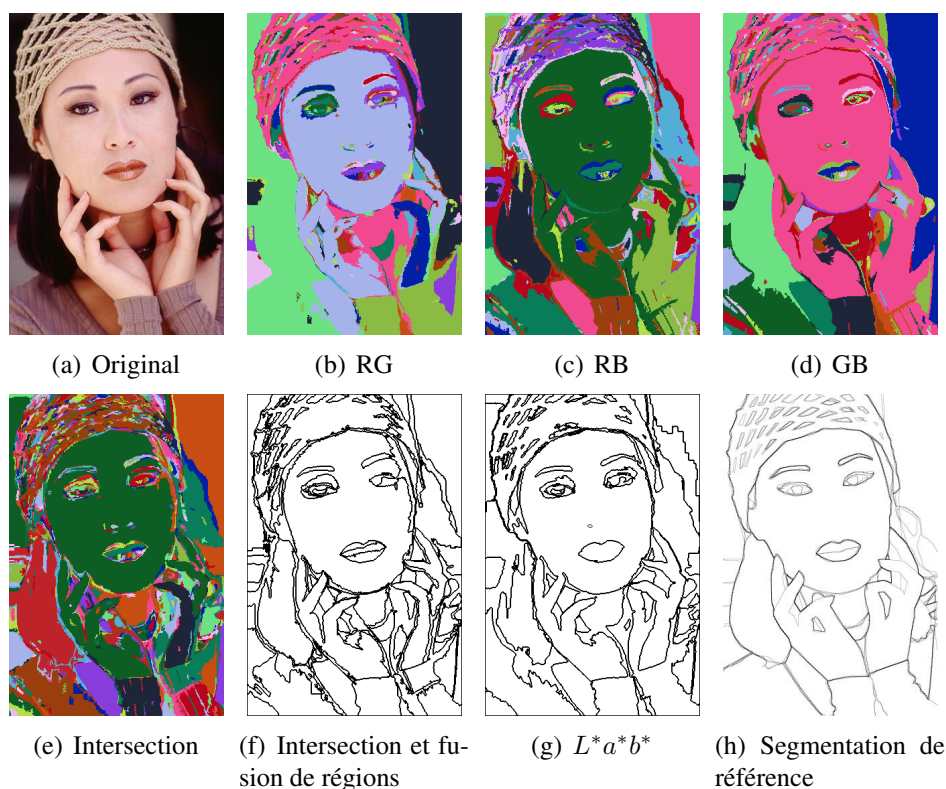


FIG. 2.32 – Illustration des différentes étapes de la méthode proposée sur une image.

l'on veut uniquement générer une hiérarchie de partitions, le critère de terminaison n'est pas nécessaire. Cependant, si l'on veut obtenir une seule partition résultante, il faut déterminer quand stopper le processus de fusion. Plutôt que d'utiliser un seuil arbitraire [ANGULO03], nous proposons de nous baser sur notre précédente formulation d'énergie comme un critère d'arrêt du processus de fusion : après chaque fusion, la mesure d'énergie est évaluée et le processus stoppe dès l'apparition d'un minimum local. Ceci permet d'obtenir alors une segmentation résultant d'un compromis entre attache aux données initiales et complexité du graphe d'adjacence de

régions. Pour accélérer le processus de fusion, nous le réalisons sur l'arbre de recouvrement minimum du graphe d'adjacence où les arêtes sont pondérées par les distances entre les régions. La figure 2.33(a) présente le graphe d'adjacence de la figure 2.32(e) et la figure 2.33(b) son MST. On remarque que le MST conserve la cohérence globale du contenu de l'image tout en simplifiant énormément la topologie du graphe. La figure 2.33(c) présente le graphe d'adjacence après fusion des régions (la partition correspondante est présentée par la figure 2.32(f)). La figure 2.32(f) présente le résultat de la fusion appliquée sur l'image d'intersection présentée par la figure 2.32(e) dans l'espace  $RGB$ . Le même résultat est présenté dans l'espace couleur  $L^*a^*b^*$  par la figure 2.32(g). On remarque que les segmentations obtenues sont très proches de la segmentation de référence (figure 2.32(h)), ce qui atteste de la qualité de la méthode.

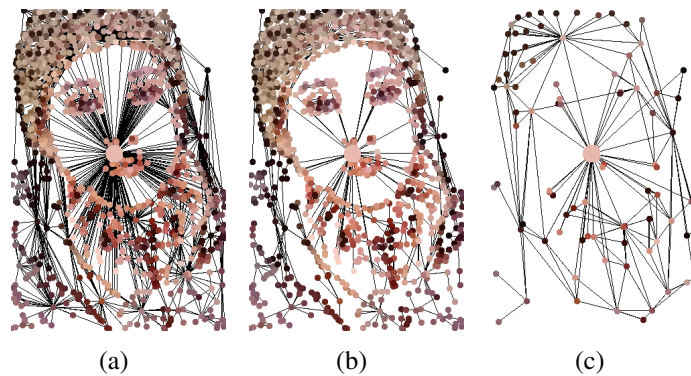


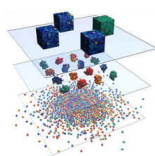
FIG. 2.33 – Le graphe d'adjacence de régions (a) de la partition (figure 2.32(e)), son MST (b) et le graphe d'adjacence correspondant (c) après fusion de région (cela correspond à la segmentation de la figure 2.32(f)).

### 2.4.5 Conclusion et perspectives

Dans cette section, nous avons proposé plusieurs méthodes permettant de générer des hiérarchies de partitions. La première méthode est basée sur une formulation d'un critère connectif nommé les zones homogènes qui, lorsqu'il est appliqué sur graphes, permet d'obtenir une hiérarchie de partitions de précision intermédiaire entre les zones quasi plates et l'algorithme des cascades. Ce critère connectif reposant sur un seuil d'homogénéité basé sur des distances, il peut s'appliquer à des images dont les pixels sont décrits par des vecteurs de dimensions quelconques. La seconde méthode est basée sur une diffusion sur graphe et elle permet naturellement via la diffusion de fusionner les régions similaires au cours des itérations de l'algorithme. Ceci permet d'une part de disposer d'une alternative rapide et simple à la simplification classique d'images et d'autre part de ne pas déplacer les frontières des régions à travers les échelles contrairement à une approche espace échelle. Ceci permet donc de combiner les bénéfices des approches hiérarchiques basées régions et des approches hiérarchiques basées sur un principe espace échelle. La troisième méthode est spécifiquement dédiée au cas des images couleur. Elle repose sur une coalescence morphologique non supervisée d'histogrammes bivariés avec détermination du niveau d'échelle appropriée. L'intersection des cartes de segmentation obtenues est ensuite simplifiée à l'aide d'une fonction d'énergie exprimée sur le graphe d'adjacence des régions qui permet de déterminer le « meilleur » niveau d'échelle suivant un compromis entre complexité de la segmentation et attachement aux données initiales.

Les perspectives de ce travail sont liées à celles des sections précédentes. Premièrement, il nous semble tout d'abord intéressant de reformuler l'ensemble des opérations morphologiques directement sur graphe afin de pouvoir traiter n'importe quel type de données. On retrouve ces directions de recherche dans les travaux récents en morphologie mathématique qui s'attachent soit à reformuler les opérations classiques à l'aide de distances [ANGULO07, NOYEL07] ou bien directement sur graphes [MEYER07]. Deuxièmement, la simplification d'une image à partir d'un graphe d'adjacence de régions représentant une sur-segmentation de l'image a été peu explorée en traitement d'images jusqu'ici. Nous comptons continuer dans cette voie qui consiste à ne plus traiter des images directement au niveau pixel mais au niveau région. Troisièmement, les images fournies par les appareils numériques actuels sont de trop grandes tailles pour pouvoir exploiter facilement les algorithmes classiques. Nous nous orientons donc également vers des approches travaillant non plus sur l'image à une résolution donnée, mais sur une pyramide régulière construite à partir de l'image. Ceci revient à définir des traitement selon une approche top-down, les traitements effectués à une échelle servant de repère pour effectuer de nouveaux traitement à une échelle plus grande.

## Classification de données d'images par apprentissage



Example is the school of mankind,  
and they will learn at no other.

EDMUND BURKE

### Sommaire

<b>3.1 Principes de la classification</b>	<b>58</b>
<b>3.2 Contexte des travaux</b>	<b>61</b>
<b>3.3 Fusion de décisions de classificateurs</b>	<b>62</b>
<b>3.4 Construction de fonctions de décision parcimonieuses</b>	<b>70</b>
<b>3.5 Schémas multi-classes de combinaison de classificateurs binaires</b>	<b>84</b>
<b>3.6 Conclusions et perspectives</b>	<b>95</b>

Ce chapitre présente les recherches que j'ai menées pour proposer des modèles pour la classification de données d'images par apprentissage principalement supervisé. Ces travaux ont été réalisés essentiellement avec Hubert Cardot alors qu'il était Maître de conférences à Caen, mais également après qu'il ait rejoint le LI à Tours en tant que Professeur d'Université. Ceci s'est fait dans le cadre de thèses que nous avons co-encadrées : Cyril Meurie [MEURIE05B], pour la combinaison de classificateurs et Gilles Lebrun [LEBRUN06] pour la construction de fonctions de décisions parcimonieuses à base de SVM. L'utilisation de ces travaux pour l'évaluation de la qualité des images compressées s'est faite avec Christophe Charrier.

Pour clarifier l'exposé de ce chapitre, en voici un plan rapide :

- Dans la Section 3.1, nous exposons le cadre de l'apprentissage artificiel qui regroupe plusieurs étapes du processus de fouille de données. Nous nous intéressons plus précisément au cadre de l'apprentissage supervisé.
- Dans la Section 3.2, nous présentons le contexte de nos travaux de recherches.
- Dans la Section 3.3, nous présentons une approche de la classification de pixels dans des images microscopique basée sur une fusion des décisions de plusieurs classificateurs.
- Dans la Section 3.4, nous montrons qu'il est possible, grâce à une sélection de modèle appropriée, de construire des fonctions de décision parcimonieuses (i.e. de complexités réduites) ayant de bonnes performances en généralisation. En effet, nous nous intéressons au cas des SVM et proposons, à l'aide d'une méthode méta-heuristique à base de recherche avec tabous, de réaliser plusieurs objectifs à savoir : réduire le nombre d'exemples pertinents, réduire le nombre d'attributs pertinents et déterminer les valeurs optimales des hyper-paramètres.

- Dans la Section 3.5, nous nous intéressons au problème de la combinaison de fonctions de décision binaires induites par un schéma de décomposition donné. Nous proposons plusieurs schémas de décodage (i.e. combinaison) et montrons finalement qu'une sélection de modèle globale est plus efficace qu'un ensemble de sélections de modèles indépendantes.

## 3.1 Principes de la classification

### 3.1.1 Généralités

Chez tous les êtres vivants, on observe une certaine aptitude à acquérir de nouveaux comportements à partir d'interactions avec leur environnement. L'objectif de l'apprentissage par ordinateur est d'étudier et de mettre en oeuvre des tels mécanismes sur des systèmes artificiels. Selon la définition qu'en donne MITCHEL [MITCHE97], « Apprendre consiste à améliorer les performances d'un système sur une tâche donnée à partir d'expériences ». Même si d'énormes progrès ont été réalisés dans la construction de modèles d'apprenants artificiels, nous sommes évidemment encore loin des capacités d'apprentissage de la plupart des êtres vivants. IZHIKEVICH a cependant pu effectuer une simulation d'un modèle de neurones impulsionsnels qui a la taille du cerveau humain [IZHIKE06], il en a conclu que l'on devrait pouvoir simuler le cerveau humain en temps réel à partir de 2016. L'apprentissage artificiel par ordinateur se trouve ainsi au carrefour de nombreuses disciplines telles que les mathématiques, l'intelligence artificielle, les sciences sociales, la biologie etc. Il s'inscrit également dans un cadre plus général. En effet, avec l'avènement de la société de l'information, des données sont produites en masses et de nombreux besoins sont apparus pour leur traitement et leur analyse car ces masses de données atteignent des complexités telles, qu'elles sont difficilement analysables par un être humain. L'ensemble de ces éléments d'aide à l'exploitation de masses de données fait référence à un champ de recherche nommé le « Data Mining » ou la « Fouille de Données ». La fouille de données est un terme récent qui résulte d'un mélange d'idées et d'outils provenant de la statistique, de l'intelligence artificielle, de l'informatique, etc. Même si sa définition exacte est encore floue, on peut dire que la fouille de données est un processus d'extraction de connaissances valides et exploitables à partir de masses de données. La fouille de données se décompose habituellement en plusieurs étapes :

1. Compréhension et analyse des objectifs de l'application,
2. Création d'une base de données pour la mise au point de l'application,
3. Pré-traitement et nettoyage des données,
4. Analyse statistique des données (réduction de dimension, etc.),
5. Identification du type de problème et choix d'un algorithme,
6. Evaluation des performances de l'algorithme,
7. Ré-itérer les étapes précédentes si nécessaire,
8. Déploiement de l'application.

La fouille de données se propose donc d'utiliser un ensemble de méthodes pour construire des concepts à partir des données. Ces concepts permettent de grouper des éléments entre eux parce qu'ils ont des descriptions proches. On peut construire des concepts sous forme de modèles globaux (analyse macroscopique des données) ou bien sous forme de motifs locaux (analyse

microscopique des données). Les modèles globaux sont définis par une fonction à optimiser et les modèles locaux par la recherche de motifs fréquents intéressants. La construction de modèles globaux fait alors directement appel aux méthodes d'apprentissage par ordinateur qui constitue une ou plusieurs étapes du processus de fouille de données (pour la classification). Nous nous intéressons uniquement aux concepts inférés sous forme de modèles globaux.

En apprentissage artificiel, on distingue habituellement plusieurs types d'apprentissages :

- Apprentissage non-supervisé,
- Apprentissage supervisé,
- Apprentissage semi-supervisé,

Pour toutes ces méthodes d'apprentissage, nous disposons d'un ensemble d'observations  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathcal{X}$  et d'un nombre de classes à discriminer par le classificateur. Chaque observation est décrite par un certain nombre de caractéristiques. Dans le cadre d'un apprentissage non supervisé, on cherche à construire un modèle directement à partir des données. L'objectif est de décrire comment les données sont organisées et d'en extraire des sous-ensembles homogènes. Parmi les méthodes non supervisées, on peut citer la classification hiérarchique [CORNUÉ02], les cartes auto-organisatrices [DREYFU04], les k-moyennes [MITCHE97], etc. Dans le cadre d'un apprentissage supervisé, on dispose, en plus des observations, de valeurs cibles (ou labels ou classes d'appartenance)  $\{y_i\} \in \mathcal{Y}$  associées à ces observations où  $\mathcal{Y}$  désigne l'ensemble des classes possibles. On cherche alors à construire un modèle qui permet d'estimer les dépendances entre les ensembles  $\mathcal{X}$  et  $\mathcal{Y}$ . On parle d'apprentissage supervisé car les éléments de  $\mathcal{Y}$  permettent de guider le processus d'estimation. Parmi les méthodes supervisées, on peut citer les k-ppv [KUNCHE04], les réseaux de neurones [DREYFU04], les machines à vecteurs de support (SVM) [ABE05, VAPNIK95, CRISTI00], les arbres de décision [QUINLA93], etc. Dans le cadre d'un apprentissage semi-supervisé [ZHOU04, CHAPEL06], parmi les observations, seulement un petit nombre d'entre eux ont un label  $\{y_i\}$ . L'objectif est alors de chercher à classifier les observations non étiquetées. Dans nos travaux, nous nous sommes plus particulièrement intéressés aux méthodes supervisées. Nous donnons donc plus de détails sur leurs principes.

### 3.1.2 Apprentissage supervisé : concepts

Le pouvoir de généralisation d'un algorithme d'apprentissage artificiel est dépendant du processus inductif qu'il réalise et de l'espace des hypothèses  $\mathcal{H}$ . Cet espace correspond à l'ensemble des fonctions de décision réalisables. Le principe inductif permet de sélectionner dans l'espace des hypothèses, à partir d'un ensemble de données, celles explicitant le mieux ces données. Ces concepts représentent le biais d'apprentissage utilisé par l'apprenant artificiel pour produire une fonction de décision avec les meilleures capacités de généralisation [CORNUÉ02]. Soit  $\mathcal{O}$  une population d'objets,  $\mathcal{X}$  l'espace de description associé à ces objets et  $o$  un oracle capable de réaliser une catégorisation des objets issus de  $\mathcal{O}$ , désignée par la fonction  $f_o : \mathcal{O} \rightarrow \mathcal{Y}$ . Soit  $f_d : \mathcal{O} \rightarrow \mathcal{X}$  une fonction qui détermine, pour un objet  $o \in \mathcal{O}$  donné, sa description  $\mathbf{x}$ . A partir de  $\mathcal{O}$ ,  $\mathcal{Y}$ ,  $f_d$  et  $f_o$  il est possible de définir l'espace des exemples  $\mathcal{Z}$ . Un exemple  $z \in \mathcal{Z}$  correspondant à un objet  $o \in \mathcal{O}$  est un couple de données  $(\mathbf{x}, y)$  tels que  $(\mathbf{x}, y) = (f_d(o), f_o(o))$ . En apprentissage supervisé, on cherche une fonction  $f : \mathcal{X} \rightarrow \mathcal{Y}$  qui permet d'estimer la classe  $y$  associée à  $\mathbf{x}$ .  $f$  appartient à l'espace d'hypothèses  $\mathcal{H}$ . Le cas idéal correspond à  $f_o = f_d \circ f$ . La figure 3.1 résume toutes ces notions.



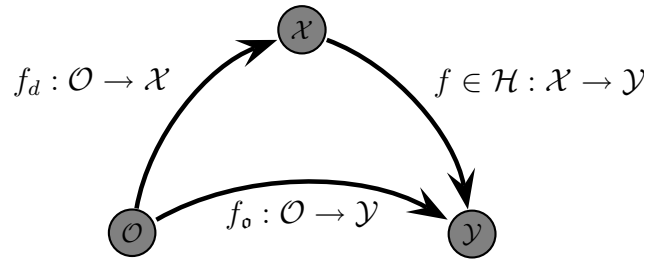


FIG. 3.1 – Pour un problème d'apprentissage donné, un objet  $o \in \mathcal{O}$  représentatif de ce problème est décrit par un vecteur d'attributs  $\mathbf{x} \in \mathcal{X}$  et est identifié par l'oracle  $o$  comme étant de la classe  $y \in \mathcal{Y}$ . Dans ce schéma,  $f_d$  et  $f_o$  représentent respectivement la procédure réalisant la description d'un objet  $o$  et le processus décisionnel de l'oracle  $o$ . L'objectif d'un apprenant est de choisir une hypothèse  $f \in \mathcal{H}$  dont les prédictions sont les plus proches possible de l'oracle.

La problématique de l'apprentissage supervisé est de trouver une fonction  $f$  aussi bonne que possible qui à un vecteur  $\mathbf{x} \in \mathcal{X}$  associe une classe  $y \in \mathcal{Y}$ . L'objectif est alors de choisir  $f^* \in \mathcal{H}$  qui minimise le risque réel correspondant :

$$f^* = \arg \min_{h \in \mathcal{H}} \mathcal{R}_{\text{reel}}(h) \quad (3.1)$$

Pour évaluer la pertinence de la prédiction de  $f$ , on introduit généralement une fonction de coût  $l(f(\mathbf{x}), y)$ . L'estimation du risque lié à une hypothèse  $h$  à partir d'un ensemble  $Z_m \in \mathcal{Z}$  de  $m$  exemples, appelée risque empirique, est égal à :

$$\mathcal{R}_{\text{emp}}(f) = \frac{1}{m} \sum_{i=1}^m l(f(\mathbf{x}_i), y_i) \quad (3.2)$$

Le risque empirique ne permet pas d'évaluer la pertinence d'un modèle car il est possible de choisir  $f$  de manière à ce que le risque empirique soit nul mais que l'erreur en généralisation soit élevée (sur-apprentissage) [CORNUÉ02].

Si l'on replace l'apprentissage supervisé relativement au processus de fouille de données, un système d'apprentissage doit réaliser l'optimisation d'un problème qui regroupe la sélection des exemples, la sélection des attributs, la sélection des hyper-paramètres relatifs à l'exploitation d'algorithmes d'apprentissage supervisé, la structure de la décomposition en sous-problèmes et le principe de décodage mis en œuvre (figure 3.2). La décomposition en sous-problèmes est classique en apprentissage artificiel. On distinguera plusieurs cas. Le premier classique consiste à utiliser plusieurs classificateurs puis à les combiner [KUNCHE04]. Le second consiste à décomposer un problème de classification en  $n_c$  classes en plusieurs sous problèmes plus simples et à les combiner ensuite [HSU02] (approches *un-contre-tous* ou bien *un-contre-un* par exemple). Nous reviendrons plus en détail (voir section 3.5) sur ce second type de décomposition. Chaque sous-problème peut nécessiter une sélection d'exemples et d'attributs qui lui est propre et donner naissance à une nouvelle décomposition. Cela signifie que, pris dans sa globalité, un système d'apprentissage artificiel supervisé devra permettre d'effectuer les étapes 3 à 6 du processus de fouille de données. Nous pouvons regrouper l'ensemble de ces sélections sous le terme générique de *sélection de modèles* (ou *sélection multi-modèle* si nous voulons insister sur la pluralité des sélections à réaliser). La sélection d'un modèle optimisant la qualité du processus décisionnel correspond donc à un problème d'optimisation difficile et deux approches sont possibles

pour réussir à gérer cette complexité. La première consiste à définir un système d'apprentissage en cascade, où chaque étape ne réalise que la sélection d'un des modèles ; l'ensemble de ces étapes réalisant la sélection complète du multi-modèle. Cela correspond classiquement à effectuer de manière séquentielle chacune des étapes 3 à 6 du processus de fouille de données. La seconde approche s'appuie sur l'exploitation d'heuristiques ou de méta-heuristiques pour trouver une solution (multi-modèle) proche qui optimise globalement le problème. Cela correspond à effectuer globalement les étapes 3 à 6 du processus de fouille de données.

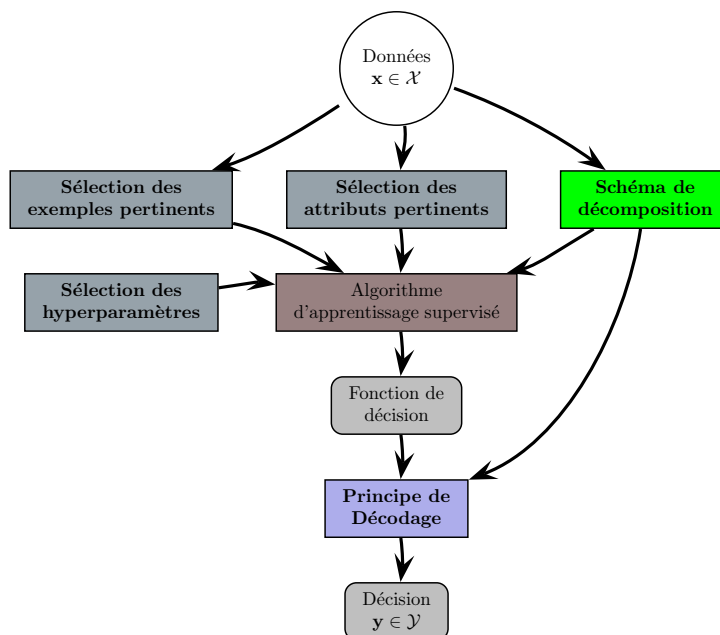


FIG. 3.2 – Les différents composants d'un système d'apprentissage supervisé.

Dans nos travaux, nous avons considéré ces deux approches de la construction d'un système d'apprentissage artificiel supervisé. Nous présentons donc dans la suite comment l'apprentissage supervisé peut être utilisé pour effectuer une classification sur des données issues d'images par la conception d'un système d'apprentissage artificiel basé sur une sélection de modèle en cascade ou bien globale.

## 3.2 Contexte des travaux

Nos travaux s'inscrivent dans un contexte régional et national de conception de systèmes d'aide au diagnostic en Pathologie par le traitement d'images de microscopie. La figure 3.3 présente une vue d'ensemble d'un tel système. Nous travaillons activement depuis l'année 2000 avec le service d'anatomie et de cytologie pathologiques du Centre Hospitalier Public du Cotentin (CHPC) basé à Cherbourg-Octeville. Nous travaillons également avec le groupe régional d'études sur le cancer (GRECAN) du Centre François Baclesse de Caen. Les applications ciblées concernent plus précisément les cytologies (cellules provenant de liquides) des bronches et des séreuses ainsi que l'histologie (coupes de tissus) du cancer du sein. Dans ce cadre, nous avons bénéficié de financements de la ligue contre le cancer, de l'association coeur-cancer, du conseil régional, et plus récemment de l'Agence Nationale de la Recherche. Nos projets se sont faits et

se font dans le cadre de thèses, de projets régionaux du Comité Régional Imagerie et Technologies de l'Information et de la Communication, d'un projet national de l'action de recherche amont masses de données connaissances ambiantes. Nous réalisons actuellement une industrialisation d'un prototype logiciel initié lors de ma thèse nommé ARCTIC (Aide à la Recherche en Cytologie par le Tri Informatique Cellulaire) avec la société ADCIS.

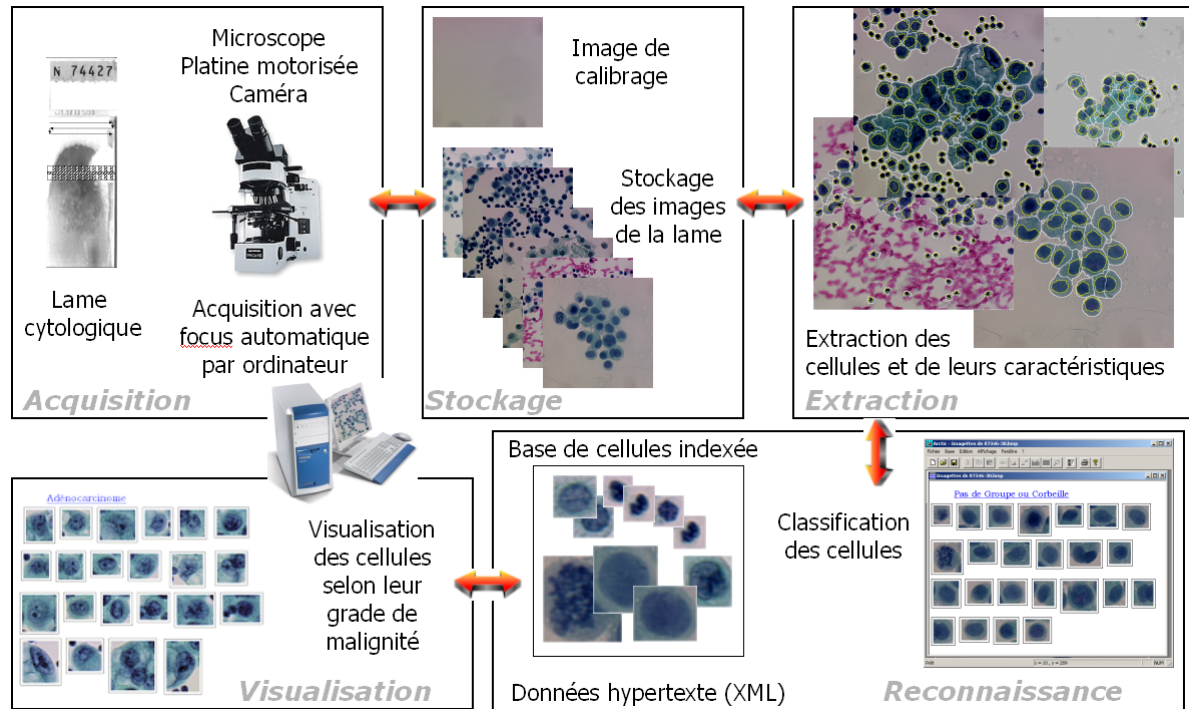


FIG. 3.3 – Vue d'ensemble d'un système d'aide au diagnostic en Pathologie par traitement d'images de microscopie.

### 3.3 Fusion de décisions de classificateurs

☞ **Mes publications associées :** [6, 7, 9, 11, 12, 32, 33, 38–40, 42, 47, 61, 62]

Nos travaux de classification de données provenant d'images concernent principalement la classification de pixels dans des images microscopiques couleur. C'est à partir des travaux que j'ai pu effectuer lors de ma thèse et après que nous avons pris cette direction. C'est dans [6, 21, 32], que nous avons proposé de considérer une classification de pixels non pas comme un résultat final de traitement mais plutôt comme une étape dans un schéma de segmentation. En effet, la classification de pixels négligeant le contexte spatial de l'image, elle ne peut fournir des résultats cohérents dans toutes les zones de l'image. C'est pourquoi, nous considérons la classification de pixels comme une étape d'extraction de germes dans le cadre d'une approche morphologique [DERIVA07, 6]. LEVNER et al. ont récemment proposé [LEVNER07] une approche généraliste de la segmentation morphologique dirigée par la classification en se basant sur nos travaux [6, 21, 32]. Une classification de pixels d'une image n'étant pas une segmentation, on procédera à un étiquetage en composantes connexes suivi d'une ligne de partage des eaux dont les marqueurs sont les zones connexes d'une certaine taille. On notera que nous utilisons une version adaptative

ligne de partage des eaux adaptative qui combine information globales et locales proposée dans ma thèse et améliorée ensuite [6, 7, 33].

### 3.3.1 Évaluation de classifications de pixels

Avant d'aller plus loin, il nous faut aborder un point important qu'est celui de l'évaluation des résultats d'une classification. Classiquement, on peut considérer le taux de réussite global d'un classificateur sur toutes les classes. Ceci n'est cependant pas approprié pour les cas où les quantités de pixels sont très disproportionnées. Le tableau 3.1 illustre ce fait par des proportions extraites de deux bases d'images de microscopie dont les pixels ont été manuellement étiquetés. Ainsi, un classificateur qui classerait tous les pixels comme du fond aura un taux de réussite

Type de base	$Proportion_{fond}$	$Proportion_{cytoplasme}$	$Proportion_{noyau}$
Base d'apprentissage	89,6 %	6,6 %	3,7 %
Base de tests	89 %	7,1 %	3,9 %

TAB. 3.1 – Statistiques de répartitions des pixels par classe sur différentes bases.

proche des 90%, ce qui peut sembler bon mais qui en fait est très mauvais. Un simple pourcentage de taux de classification correcte n'est donc pas du tout approprié. Une façon de pallier ce problème est soit d'utiliser un taux d'erreur balancé soit une mesure d'évaluation spécifique. Un taux d'erreur balancé (BER : Balanced Error Rate) est défini par :

$$e_{BER} = \frac{1}{|\mathcal{Y}|} \sum_{\omega \in \mathcal{Y}} e_{\omega} \quad (3.3)$$

avec  $e_{\omega}$  le taux d'exemples mal classés de la classe  $\omega$ . Nous avons alors proposé, pour le cas des images microscopiques, une mesure d'évaluation spécifique de la classification obtenue. Cette mesure se base sur une vérité terrain. Avec l'aide d'experts pathologistes du Centre Hospitalier Public du Cotentin, nous avons pu constituer de telles vérités terrain (constituées de 8 images pour la cytologie bronchique et de 10 images pour le cytologie des séreuses). Nous pouvons donc comparer des classifications automatiques par rapport à des vérités terrain. En microscopie, on cherche classiquement à distinguer trois classes de pixels : le fond, les cytoplasmes et les noyaux. La méthode que nous avons proposée fournit un indice de qualité sur la reconnaissance du cytoplasme «  $Indice_{cytoplasme}$  » et du noyau «  $Indice_{noyau}$  ». Pour calculer ces indices, nous calculons le nombre de pixels appartenant à chaque classe de l'image. Le taux de commun «  $Com_{cytoplasme}$  » ou «  $Com_{noyau}$  » montre le nombre de pixels correctement identifiés dans l'image de référence. Tandis que le taux de différence «  $Diff_{cytoplasme}$  » ou «  $Diff_{noyau}$  » globalise pour une classe  $\omega_i$  donnée (cytoplasme ou noyau), les deux erreurs suivantes :

1. Un pixel de la classe  $\omega_i$  dans l'image de référence n'est pas reconnu comme étant de la même classe dans l'image segmentée ;
2. Un pixel n'étant pas de la classe  $\omega_i$  dans l'image segmentée est reconnu comme étant de la classe  $\omega_i$  dans l'image de référence.

Cette méthode d'évaluation peut alors être définie par les équations suivantes :

$$Indice_{cytoplasme} = \frac{Com_{cytoplasme} + (1 - Diff_{cytoplasme})}{2} \quad (3.4)$$

$$Indice_{noyau} = \frac{Com_{noyau} + (1 - Dif_{noyau})}{2} \quad (3.5)$$

Les équations (3.4) et (3.5) fournissent respectivement l'indice de qualité sur la reconnaissance du cytoplasme et du noyau, en prenant en compte le nombre de pixels bien classés (1er terme : «  $Com_{noyau}$  ») et le nombre de pixels mal classés (2ème terme : «  $1 - Dif_{noyau}$  »). Il est important de signaler que les indices de qualité sur la reconnaissance des objets cellulaires (i.e. le cytoplasme et le noyau) calculés par cette méthode doivent être considérés à leurs justes valeurs. En effet, les indices sont calculés sur la base du pixel lui-même et une erreur de reconnaissance d'un pixel sur chaque objet entraîne une diminution en pourcentage de l'indice en question. Notons que l'indice de qualité présenté n'est pas un pourcentage, mais pour plus de lisibilité, nous le noterons comme tel. En cytologie bronchique par exemple, il est très difficile voire impossible pour les experts de délimiter précisément à moins de deux pixels près le contour des objets. Ce qui revient à dire que toute classification de pixels ou segmentation produite est considérée comme très correcte à plus ou moins deux pixels sur chaque objet. En conséquence de quoi, les indices approchant les 80 % peuvent être considérés comme représentant une bonne qualité de segmentation. Le tableau 3.2 présente des résultats d'évaluation de segmentations

Image	Liu-Borsotti	Vinet	$Indice_{cytoplasme}$	$Indice_{noyau}$
Figure 3.4(c)	9.4	51	74.8 %	84.4 %
Figure 3.4(d)	9.8	51.8	72.2 %	73.3 %

TAB. 3.2 – Résultats de trois méthodes d'évaluation sur une image segmentée.

en comparaison avec deux autres méthodes classiques d'évaluation [BORSOT98, LIU94, VINET91]. La figure 3.4 présente les images ayant servi à produire ces résultats d'évaluation. On note que sur la Figure 3.4(d) plusieurs noyaux ont disparu et cela se traduit par une moins bonne évaluation bien marquée par notre méthode contrairement aux méthodes classiques.

### 3.3.2 Schémas de classification

Nous nous sommes tout d'abord intéressés à effectuer la classification des pixels d'une image couleur par différents classificateurs. Nous avons considéré des classificateurs supervisés ou non. Pour le cas des classificateurs non supervisés (k-moyennes, C-moyennes floues, Fisher LDA), nous avons proposé un schéma hiérarchique de segmentation qui consiste à discriminer les classes deux à deux de manière hiérarchique. Pour le cas des images microscopiques on souhaite extraire trois classes de pixels : le fond, les cytoplasmes et les noyaux des cellules. Une classification non supervisée effectuée directement sur une image microscopique ne fournit pas de très bons résultats car la proportion de pixels de fond est beaucoup plus importante que pour celles des deux autres classes. Nous effectuons donc une première classification non supervisée visant à discriminer le fond du reste de l'image. Puis, dans les zones qui ne sont pas reconnues comme du fond, nous effectuons une nouvelle classification non supervisée visant à discriminer les cytoplasmes des noyaux. On consultera la thèse de Cyril Meurie [MEURIE05B] pour plus de détails. En particulier, la discrimination du fond contre le reste de l'image peut être améliorée en effectuant une classification non supervisée avec plus de deux classes et en les fusionnant hiérarchiquement ensuite.

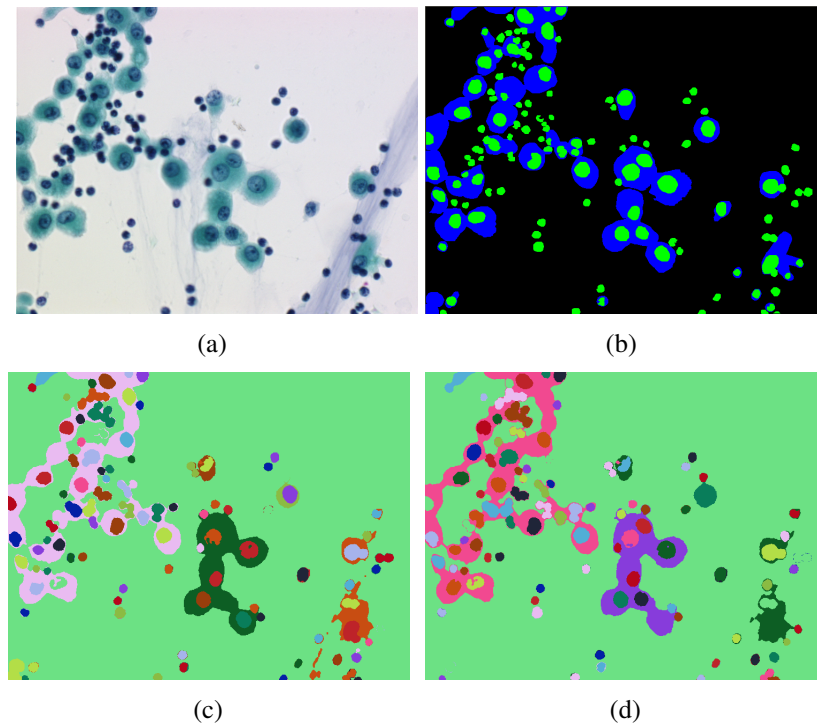


FIG. 3.4 – Résultats de différentes évaluations sur la segmentation d’une image par classification Bayésienne comme germes d’une LPE ((a) l’image initiale, (b) la vérité terrain, (c-d) deux images segmentées).

Pour le cas des classificateurs supervisés (classification bayésienne, k-ppv, réseaux de neurones multi-couches notés RNMC, machines à support de vecteurs notées SVM), la classification supervisée se fait en une seule étape et chaque pixel est étiqueté dans l’une des trois classes possibles. La principale difficulté réside dans la construction des bases de données permettant d’effectuer l’apprentissage du classificateur. La taille de la base d’apprentissage a une influence certaine sur les algorithmes supervisés. Certains, tels que les k-PPV et les SVM, réclament des puissances de calculs importantes et pour des grandes bases d’apprentissage, les temps d’apprentissage deviennent alors prohibitifs. Nous avons pu mettre en évidence que ne retenir qu’un tiers des pixels de chaque classe détériore très peu les résultats et ce, quelque soit l’espace de représentation ou l’algorithme utilisé. En conséquence, nous avons retenu, pour des raisons de temps de calcul, un apprentissage basé sur un tiers de la base d’apprentissage (la différence avec un apprentissage effectué sur toute la base est très faible). Nous reviendrons plus loin sur ce problème de prototypage lorsque nous considérerons la construction de fonctions de décision parcimonieuses. Dans la suite, les paramètres des algorithmes de classification supervisée ont réglés afin d’optimiser notre mesure d’évaluation (nous reviendrons également plus loin sur l’importance de ces paramètres).

Pour chacun des schémas de classification (supervisés ou non), nous avons considéré l’influence de la représentation de chaque pixel. En effet, la couleur d’un pixel peut être représentée autrement que par ses composantes Rouge, Verte et Bleu. Changer de représentation peut permettre d’améliorer les performances de la classification. Nous avons considéré chacun des schéma de classification dans plusieurs espaces couleurs et nous avons retenu la meilleure représentation en évaluant la qualité des classification obtenues sur une base de test (avec notre mesure d’évaluation). Le tableau 3.3 résume les résultats obtenus. On remarque d’une part que le changement d’espace couleur permet d’améliorer les résultats d’un classificateur donné et d’autre part que

les méthodes supervisées sont globalement meilleures que les méthodes non supervisées (ce qui n'est pas vraiment une surprise). Cependant, il est difficile de déterminer quelle est la meilleure méthode parmi toutes celles présentées.

Classificateur	Espace Couleur	$Indice_{cytoplasme}$	$Indice_{noyau}$
k-moyennes	$YCh_1Ch_2$	69.5 %	74.4 %
Fisher G	$RGB$	50.8 %	72.4 %
Fisher $I_1$	$I_1I_2I_3$	57.3 %	71.9 %
Fisher L	$HSL$	59.9 %	69.8 %
SVM	$YCh_1Ch_2$	77.4 %	74.2 %
Bayes	$YCh_1Ch_2$	72.4 %	74.6 %
RNMC	$YC_bC_r$	56.9 %	73 %
k-PPV	$HSL$	79.9 %	70 %

TAB. 3.3 – Résultats des différentes classifications de pixels non supervisées (partie haute) ou non supervisés (partie basse).

La figure 3.5 présente des résultats obtenus sur une image de cytologie bronchique. L'évaluation visuelle est bien en accord avec les résultats fournis par notre mesure d'évaluation.

### 3.3.3 Combinaison de classifications

Même à l'aide de notre mesure d'évaluation, il est difficile de déterminer quelle est la meilleure classification. De plus, une classification peut très bien fonctionner sur une image et moins bien sur d'autres. Pour contourner ce problème du choix du meilleur classificateur, nous avons proposé de fusionner les classifications obtenues par différents classificateurs via une combinaison non supervisée (pas d'apprentissage). Cette combinaison nous permettra de lever les incohérences qui peuvent apparaître entre plusieurs classificateurs. La figure 3.5(c) présente les zones de non-cohérence entre six classifications (figures 3.5(d)-3.5(i)). Comme tous les classificateurs ne fournissent pas une mesure de probabilité d'appartenance à chaque classe, nous ne considérons que des informations de type classe [KUNCHE00, ZOUARI02] (chaque classificateur fournit juste un numéro de classe pour chaque pixel). Nous avons tout d'abord évalué chaque classificateur sur une base de validation à l'aide de notre mesure de qualité de la classification. Ceci permet d'accorder un indice de confiance global  $indice(e_j, \omega_i)$  par classe  $\omega_i$  à chaque classificateur  $e_j$ . Nous pouvons ensuite fusionner les décisions des classificateurs [KITTLE98] afin de prendre une décision finale.

Soit  $E = \{e_1, \dots, e_k\}$  l'ensemble des  $k$  classificateurs utilisés. Chacun des classificateurs associe à un vecteur d'entrée  $\mathbf{x}$  une classe  $\omega_i$ . Nous pouvons donc définir  $E_{\omega_i}(\mathbf{x})$  l'ensemble des classificateurs, qui pour un vecteur d'entrée  $\mathbf{x}$  lui associent tous la classe  $\omega_i$  :

$$E_{\omega_i}(\mathbf{x}) = \{e_j \in E | e_j(\mathbf{x}) = \omega_i\} \quad (3.6)$$

Nous avons évidemment  $\cup\{E_{\omega_i}(x)\} = E$  car un classificateur  $e_j$  ne prend qu'une seule décision de type classe. A chaque ensemble  $E_{\omega_i}(\mathbf{x})$  avec  $\omega_i \in \mathcal{Y}$ , nous pouvons associer l'ensemble des indices de confiance pour chaque classificateur  $e_j$ . Cet indice correspond à la confiance accordée

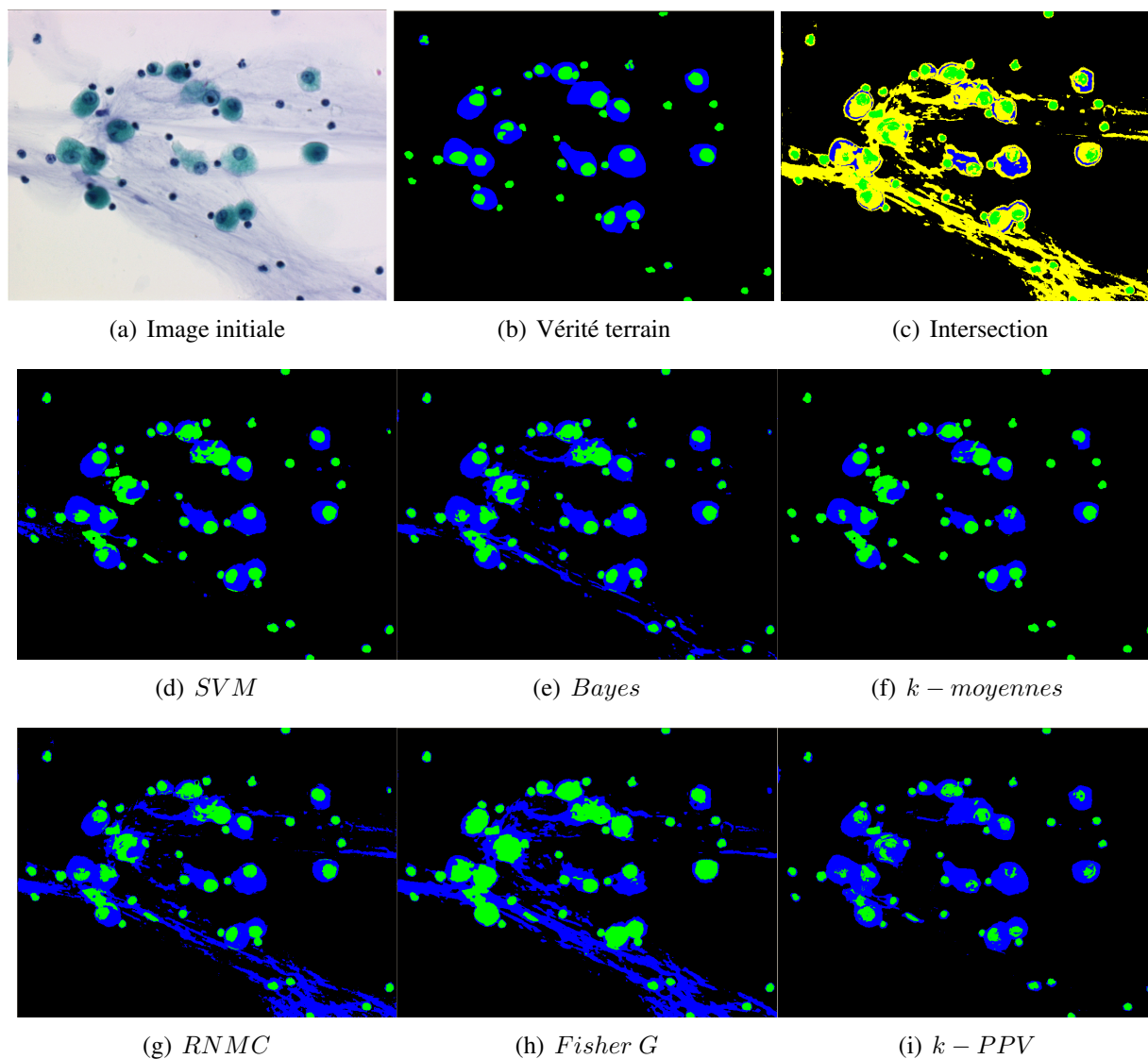


FIG. 3.5 – Résultats de différentes classifications de pixels (lignes du milieu et du bas) pour une image de la base de test. Première ligne : image initiale, vérité terrain, intersection de toutes les classifications (zones de non cohérence en jaune).



à la classification réalisée par le classificateur  $e_j$  lorsqu'il associe à  $\mathbf{x}$  la classe  $\omega_i$ . Soit  $I_{\omega_i}(\mathbf{x})$  l'ensemble de ces indices défini comme suit :

$$I_{\omega_i}(\mathbf{x}) = \{indice(e_j, \omega_i) | e_j \in E_{\omega_i}(\mathbf{x})\} \quad (3.7)$$

L'ensemble  $I_{\omega_i}(\mathbf{x})$  correspond donc aux indices de confiance respectifs des classificateurs qui classent l'entrée  $\mathbf{x}$  comme étant de la classe  $\omega_i$ . A partir de ces ensembles d'indices, nous pouvons calculer la probabilité d'appartenance de  $\mathbf{x}$  à la classe  $\omega_i$  par la relation suivante :

$$p(w = \omega_i | \mathbf{x}) = g(I_{\omega_i}(\mathbf{x})) \quad (3.8)$$

où  $g$  représente une fonction de combinaison parmi les fonctions suivantes : vote majoritaire, minimum, maximum, somme, moyenne que nous pouvons retrouver dans les travaux de KITTLE ET AL. [KITTLE98]. Nous assignons ensuite au pixel  $p$ , décrit par le vecteur d'attributs  $\mathbf{x}$ , la classe  $\omega_k$  telle que :

$$\omega_k = \arg \max_l p(w = \omega_l | \mathbf{x}) \quad (3.9)$$

Comme nous l'avons déjà mentionné, une classification de pixels néglige le contexte spatial de l'image. Combiner plusieurs classifications de pixels obtenues par plusieurs classificateurs pour un même pixel présente le même défaut. Afin d'y pallier, nous intégrons une information de voisinage lors de la fusion des résultats des classificateurs : nous ne combinons pas uniquement  $k$  classifications mais  $9 \times k$  classifications, ce qui revient à prendre en compte toutes les classifications autour du pixel central dans une fenêtre d'analyse  $3 \times 3$  et à les fusionner.

Classificateur	<i>Indice<sub>cytoplasme</sub></i>	<i>Indice<sub>noyau</sub></i>
<b>Règle de combinaison non-supervisée (somme)</b>	<b>78.3 %</b>	<b>74.9 %</b>
Règle de combinaison non-supervisée (vote majoritaire)	78.1 %	74.8 %
Règle de combinaison supervisée (BKS)	78.5 %	74.8 %
Combinaison par Dempster-Shafer	76.3 %	74.7 %

TAB. 3.4 – Résultats des différentes combinaisons de classifications de pixels.

Nous avons pu mettre en évidence, en étudiant toutes les combinaisons possibles des classificateurs du tableau 3.3, que la combinaison des trois meilleurs classificateurs permettait d'obtenir les meilleurs résultats (k-moyennes, SVM et Bayes). Le tableau 3.4 présente plusieurs résultats de combinaison avec la méthode proposée et deux autres méthodes de la littérature (Behaviour Knowledge Space [HUANG95] et une combinaison basée sur la théorie de l'évidence [11]). La combinaison que nous avons proposée permet d'obtenir les meilleurs résultats. Ce résultat de classification est utilisé comme base de départ d'une ligne de partage des eaux couleur adaptative. L'image classifiée est étiquetée en composantes connexes et les frontières des objets sont considérées comme non étiquetées puis érodées. Les composantes connexes constituent alors les marqueurs de la ligne de partage des eaux et les zones non étiquetées les zones d'influence des marqueurs.

Le tableau 3.5 et la figure 3.6 présentent respectivement le résultat de l'étape de segmentation de façon quantitative et le résultat avec les frontières colorées des objets superposés. Dans ce tableau, nous présentons tout d'abord, les résultats de segmentation (classification de pixels et

Classificateur	$Indice_{Cytoplasme}$	$Indice_{Noyau}$
k-moyennes	72.8 %	76.2 %
SVM	73.2 %	75.8 %
Bayes	71.1 %	76.7 %
<b>Règle de combinaison non-supervisée</b>	<b>76.5 %</b>	<b>76.4 %</b>

TAB. 3.5 – Indice de segmentation finale.

LPE) pour les trois meilleurs classificateurs (k-moyennes, SVM et Bayes). Nous présentons ensuite les résultats concernant la combinaison des trois meilleures classifications de pixels suivies d'une LPE. L'étape de raffinement des objets par ligne de partage des eaux permet globalement de mieux détecter les noyaux et les cytoplasmes par rapport à une classification de pixels seule. En outre, la combinaison de classification de pixels suivie d'une croissance de régions par ligne de partage des eaux permet bien d'obtenir globalement de meilleurs résultats. Nous avons également proposé d'autres schémas de fusion par la théorie de l'évidence de plusieurs classificateurs de type SVM effectuant leurs classifications dans différents espaces couleur [12, 47].

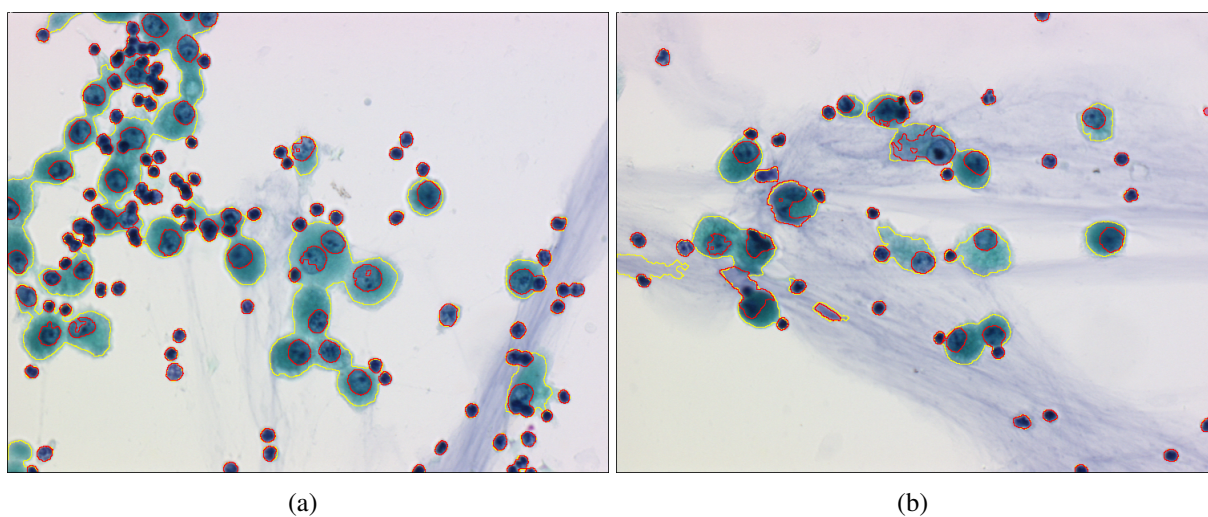


FIG. 3.6 – Deux images de la base de test segmentées.

## 3.4 Construction de fonctions de décision parcimonieuses

☞ Mes publications associées : [13, 14, 43, 44, 48–50, 52, 60, 64, 65, 71]

### 3.4.1 Sélection de modèle et parcimonie

La construction de fonctions de décision simplifiées est essentielle dans les applications ayant des contraintes de temps réel importantes. L'objectif est alors de pouvoir produire des solutions fortement parcimonieuses tout en conservant des capacités en généralisation performantes. Dans nos travaux, nous nous sommes intéressés à la construction de fonctions de décision parcimonieuses pour les SVM [ABE05, CRISTI00]. La réduction de la complexité des fonctions de décision des SVM peut être réalisée en agissant sur deux points essentiels :

- la réduction de la taille de la base d'apprentissage, en ne conservant que les exemples les plus informatifs,
- la sélection des attributs les plus utiles pour un problème donné de classification.

La complexité, en temps d'évaluation de la fonction de décision d'un SVM, est linéairement dépendante du nombre de vecteurs de support. Comme ce nombre est borné par le nombre d'exemples, l'utilisation d'une base d'apprentissage de taille réduite permet de s'assurer une limitation contrôlée de sa complexité. Les SVM ont par nature la possibilité de sélectionner un nombre réduit d'exemples comme vecteurs de support, mais les données d'apprentissage étant bruitées dans la majorité des cas, ce nombre est généralement élevé.

Nous avons vu dans la section précédente qu'il n'est pas toujours nécessaire d'utiliser l'ensemble des données d'apprentissage. En revanche, il est très intéressant de choisir un ensemble réduit d'exemples représentatifs du problème d'apprentissage, pour produire une fonction de décision performante avec l'algorithme des SVM. Bien entendu, plus il y a de données pour décrire un problème d'apprentissage et plus les possibilités pour produire une fonction de décision performante à partir de ces données sont grandes. La sélection d'un sous-ensemble réduit d'exemples (ou d'attributs) est lui-même un processus d'apprentissage et elle doit utiliser un principe d'inférence robuste pour garantir de bonnes capacités en généralisation. Ces deux points se retrouvent également dans le processus de fouille de données.

Pour tous les classificateurs, la sélection d'un modèle a une grande influence sur les capacités en généralisation. Les SVM n'y échappent pas. Il est en effet nécessaire de régler deux hyper-paramètres que sont la largeur de bande du noyau Gaussien  $\sigma$  et la constante de régularisation  $C$ . Si l'algorithme des SVM permet de sélectionner un hyperplan de séparation optimal, lorsque ces différents paramètres sont fixés, la façon de réaliser le choix des valeurs optimales à donner aux paramètres libres d'un modèle reste à définir. Une procédure classique, nommée *grid search* [CHANG01], est suffisante dans le cas où le noyau n'a qu'un paramètre à sélectionner (en plus de la valeur de la constante de régularisation  $C$ ), mais elle devient difficilement exploitable avec plus de deux paramètres. Parmi les paramètres que l'on peut souhaiter régler, on trouve par exemple :

- La recherche d'un sous-ensemble d'exemples pertinents,
- La recherche d'un sous-ensemble d'attributs pertinents,
- Le réglage des hyper-paramètres de l'algorithme ( $\sigma$  et  $C$ ).

Dans la littérature liée à la classification de données provenant d'images, les SVM sont souvent écartés pour plusieurs raisons [MARÉ03, MARTIN04]. Premièrement, si la base d'apprentissage est grande, l'apprentissage des SVM peut s'avérer très long. De plus, si la fonction de

décision produite par un SVM utilise beaucoup de vecteurs de support (ce qui risque d'arriver si les données sont bruitées), le processus de classification sera également long. En outre, les SVM sont souvent utilisés sans réellement prendre garde à optimiser les précédents paramètres que nous avons cités et ceci peut avoir une grande influence sur leurs performances, comme pour tout algorithme d'apprentissage artificiel supervisé. Nous nous attacherons dans la suite à montrer qu'en cherchant à construire des fonctions de décision simplifiées, les SVM peuvent, à l'aide d'une sélection de modèle appropriée, fournir les modèles parcimonieux adaptés à la classification de données d'images.

### 3.4.2 Construction de fonctions de décision simplifiées

Partant des constatations précédentes, nous avons proposé d'exploiter des techniques de classification non-supervisées pour réduire la redondance inutile dans les données, et ainsi découvrir une représentation simplifiée et représentative de la base initiale. Comme les SVM sont sensibles à une sélection efficace des hyper-paramètres, nous avons également pour objectif de réaliser une optimisation de ceux-ci. La recherche d'un modèle performant combinant le choix des hyper-paramètres des SVM, la simplification de la base en nombre d'exemples et en nombre d'attributs utilisés correspond à l'exploration d'un espace de recherche de dimension importante. Pour pallier cet inconvénient, une méthode méta-heuristique à base de recherche avec tabous a été définie. Un nouveau critère de qualité qui correspond à un compromis entre capacité de généralisation et rapidité de la décision est proposé. Il permet à l'utilisateur de spécifier quel taux de dégradation de la qualité prédictive de la fonction de décision il est prêt à accepter, si cela conduit à une amélioration significative des temps de décision. Nous détaillons tous ces principes dans la suite.

#### 3.4.2.1 Prototypage d'une base

Naturellement, nous nous sommes intéressés aux techniques d'apprentissage non supervisé pour produire un ensemble réduit de prototypes à partir d'un ensemble plus conséquent de données d'apprentissage [JAIN99, MEILA05, XU05, TAN06]. Notre but est de produire un ensemble de prototypes suffisamment représentatifs de la base de données. Ces prototypes seront ensuite utilisés pour entraîner un SVM. Les prototypes ne sont pas nécessairement des exemples de la base d'apprentissage, mais des données synthétisées à partir d'elle. Nous avons choisi d'utiliser l'algorithme de quantification vectorielle LBG (Linde, Buzo et Gray [LINDE80]).

Le but de notre méthode de simplification est de produire, pour un niveau donné de simplification, un ensemble réduit de prototypes qui soit fortement représentatif de la répartition des données de chaque classe. Chaque prototype produit par notre méthode de simplification doit résumer le mieux possible un ensemble d'exemples d'une classe donnée de la base d'apprentissage. Pour produire ce résultat, l'algorithme LBG est appliqué, pour une valeur de  $k$  fixée représentant le niveau de simplification, à tous les exemples d'une classe donnée (maximum  $2^k$  prototypes sont obtenus). Une valeur de  $k$  faible correspondra à un niveau de simplification important, et réciproquement. La figure 3.7 présente la construction d'un ensemble de prototypes sur une base d'apprentissage de classification de pixels. Nous avons pu mettre en évidence [LEBRUN06], sur plusieurs bases de référence, que l'apprentissage d'un SVM sur une base prototypée est plus sensible aux choix de ses hyper-paramètres ( $C$  et  $\sigma$ ). Cependant, l'apprentissage sur des prototypes fournit des informations utiles sur les couples d'hyper-paramètres performants. La simplifica-

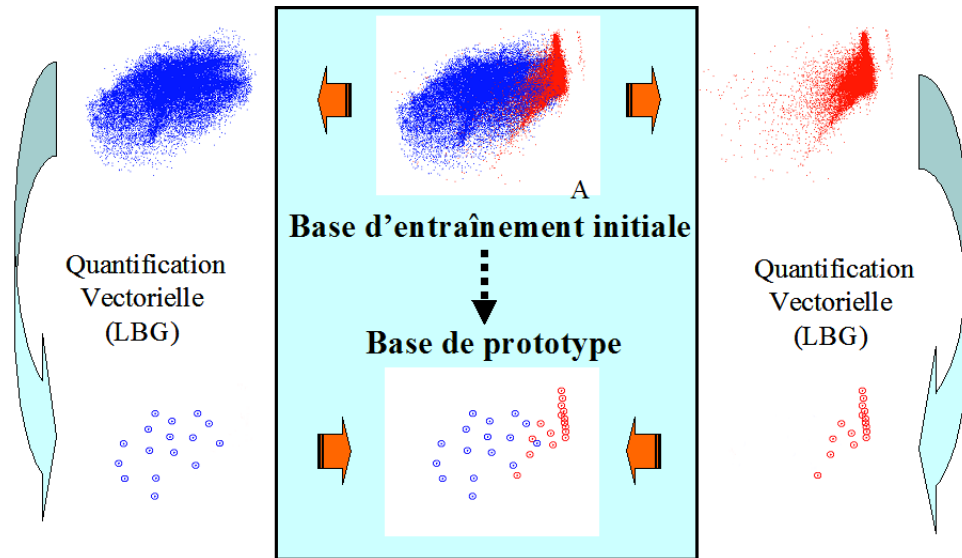


FIG. 3.7 – Synopsis de notre méthode de simplification par quantification vectorielle à partir de l’algorithme LBG. Il y a 16 prototypes par classes ( $k = 4$ ) pour cet exemple. Les données proviennent d’un problème de discrimination de pixels en microscopie bronchique : fond (rouge) ou cellule (bleu) où deux caractéristiques ont été sélectionnées (les composantes bleu et vert d’un espace  $RVB$ ).

tion de la base d’apprentissage par prototypage permet donc une identification plus marquée des hyper-paramètres qui permettent de produire des fonctions de décision qui ont de bonnes capacités de généralisation ainsi qu’un nombre réduit de vecteurs de support pour le niveau de simplification utilisé, mais également pour des niveaux de simplification moins importants. Ce résultat est très important si l’on cherche à optimiser ces deux critères. Il est donc possible de construire des fonctions de décision performantes et parcimonieuses (peu de vecteurs supports). De plus, apprendre sur une base prototypée est évidemment plus rapide. La figure 3.8 présente une illustration de ces remarques. Sont présentés le taux de reconnaissance balancé sur une base de test, le nombre de vecteurs support et le temps d’apprentissage avec différents niveaux de simplification  $k$ . On remarque qu’un ensemble de prototypes bien choisis permet d’améliorer les capacités de généralisation tout en diminuant la complexité des fonctions de décision produites et leur temps d’apprentissage. Ces résultats illustrent le fait que le choix d’un compromis efficace entre performances en généralisation et complexité est difficile à réaliser. De plus, la sélection des hyper-paramètres des SVM peut être plus ou moins dépendante du niveau de simplification, suivant la nature des données utilisées.

### 3.4.2.2 Optimisation globale par recherche avec tabous

Si l’on considère le problème de la construction d’un modèle le plus efficace possible à base de SVM, cela nous ramène au problème de la sélection d’un modèle. Tous les problèmes d’apprentissage correspondent à la recherche d’une fonction de décision  $h^*$  de qualité optimale, suivant un critère  $q$ . L’espace des fonctions de décision admissibles est défini par un ensemble de paramètres ou variables libres  $\theta$ , généralement contraints dans un ensemble de valeurs possibles

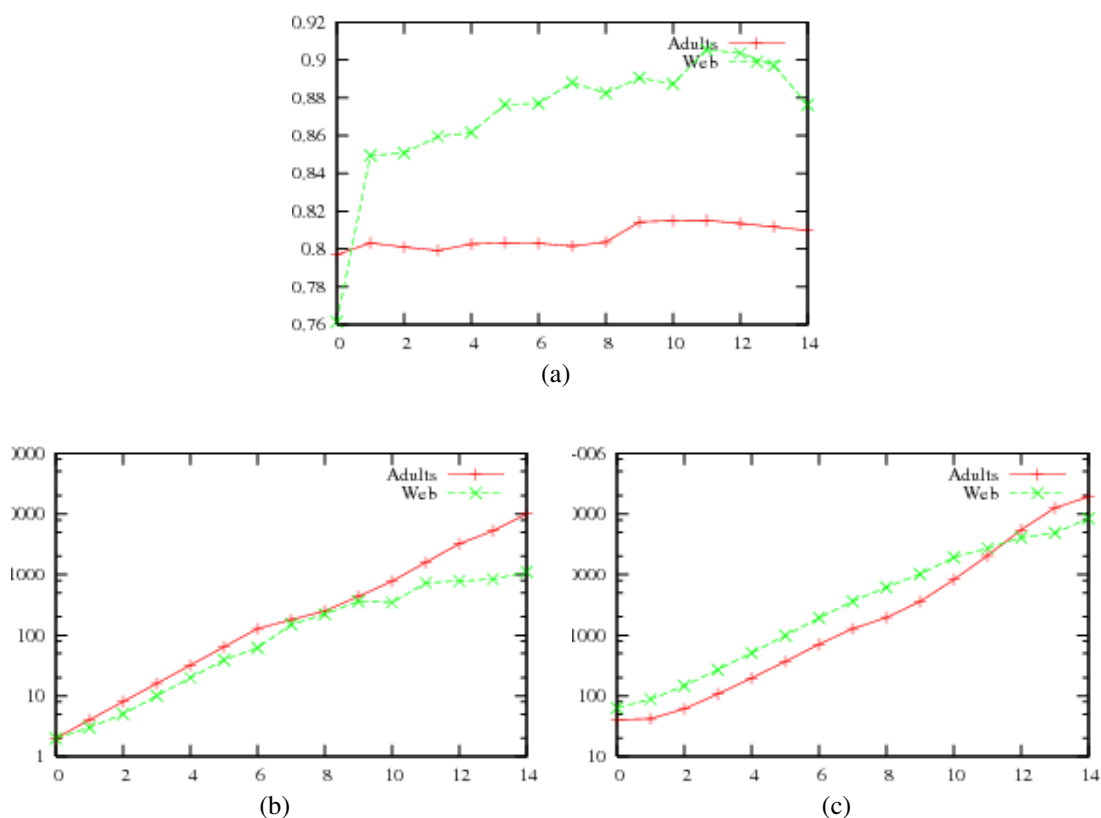


FIG. 3.8 – (a) Taux de reconnaissance sur une base de test et (b) nombre de vecteurs de support de la fonction de décision pour le modèle sélectionné par une procédure de *grid search* pour chaque niveau de simplification  $k$  précisé en abscisse. Le graphique (c) donne la durée totale en secondes de la procédure de *grid search* ( $\log_{\sqrt{2}}(C) \in \{-10, \dots, 20\}$  et  $\log_{\sqrt{2}}(\frac{1}{2\sigma^2}) \in \{-10, \dots, 10\}$ ) pour chaque niveau  $k$  de simplification donné pour ces deux bases. **Adult** et **web** sont les deux bases utilisées [ASUNCI07, PLATT99B].

$\Theta$ . La recherche de la fonction de décision optimale correspond alors à un problème d'optimisation :

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \Theta} q(h_{\boldsymbol{\theta}}) \quad (3.10)$$

Pour le cas des SVM, la sélection d'un modèle efficace peut permettre de produire une fonction de décision performante et de complexité réduite. Cependant, cela nécessite la sélection du niveau de simplification à appliquer à la base initiale d'entraînement, des valeurs des hyper-paramètres associés à l'entraînement d'un SVM et des attributs pertinents. La simplification appliquée dépend exclusivement de la valeur de  $k$  pour la méthode de simplification que nous avons choisie (i.e. LBG). Pour les hyper-paramètres, si on se limite à un noyau gaussien, ce sont les valeurs de  $C$  et de  $\sigma$  qui doivent être déterminées parmi un ensemble de valeurs quantifiées (comme pour un grid search). De plus, nous ajoutons au modèle  $\boldsymbol{\theta}$  la détermination d'un vecteur  $\boldsymbol{\beta}$  d'éléments binaires pour la sélection d'attributs pertinents. Un attribut  $i$  est sélectionné si  $\beta_i = 1$  (dans le cas contraire  $\beta_i = 0$ ).

Un modèle  $\boldsymbol{\theta}$  est donc un vecteur de  $n_{\boldsymbol{\theta}}$  variables entières, notées  $\theta_i$ , qui ont chacune un domaine de définition qui correspond à un intervalle. Pour un modèle  $\boldsymbol{\theta}$  donné, nous notons  $C_{\boldsymbol{\theta}}$ ,  $\sigma_{\boldsymbol{\theta}}$ ,  $k_{\boldsymbol{\theta}}$  et  $\boldsymbol{\beta}_{\boldsymbol{\theta}}$  les valeurs des paramètres qui sont liées à  $\boldsymbol{\theta}$ . Pour prendre en compte la sélection d'attributs réalisé dans l'apprentissage d'un SVM, nous utilisons la fonction noyau  $K_{\boldsymbol{\beta}}$  suivante :

$$K_{\boldsymbol{\beta}}(\mathbf{u}, \mathbf{v}) = \exp \left( \frac{- \sum_{i=1}^n \beta_i (u_i - v_i)^2}{2\sigma^2} \right) \quad (3.11)$$

L'algorithme 5 décrit les étapes d'entraînement d'un SVM pour un modèle  $\boldsymbol{\theta}$ , ainsi que les données utilisées pour déterminer la Qualité  $q$  d'une Fonction de Décision (QFD) produite par un SVM. Pour un problème donné, la phase ECHANTILLONAGE n'est réalisée que pour le premier modèle  $\boldsymbol{\theta}$  exploité. Pour chaque problème d'apprentissage, la simplification de la base d'entraînement  $Z^E$  pour un niveau  $k$  donné est enregistrée (ainsi que toutes celles de niveaux inférieurs à  $k$ ), ceci afin d'éviter d'avoir à la recalculer. Toute nouvelle simplification à réaliser, dont le résultat pour un niveau  $k$  spécifié n'a pas été préalablement enregistré, est produite à partir de la première simplification d'un niveau  $k'$  ( $k' < k$ ) déjà enregistré. On notera que la base d'apprentissage  $Z^A$  est séparée en une base d'entraînement  $Z^E$  qui sert pour l'entraînement d'un SVM et d'une base de validation  $Z^V$  qui sert à évaluer le taux d'erreur balancé. Ceci suppose que la base d'apprentissage est suffisamment grande, sinon il faut mettre en oeuvre des techniques de validation croisée. Nous avons d'ailleurs proposé séparément une technique rapide d'entraînement d'ensemble de SVM pour la validation croisée [50].

```

( $C_\theta, \sigma_\theta, k_\theta, \beta_\theta$ )  $\leftarrow$  EXTRAIREPARAMÈTRES( $\theta$ )
( $Z^E, Z^V$ )  $\leftarrow$  ECHANTILLONAGE( $Z^A$ ) \\\ séparation proportionnelle 2/3 et 1/3
 $\tilde{Z}^E \leftarrow$  SIMPLIFICATION( $Z^E, k_\theta$ )
 $D \leftarrow$  ENTRAÎNEMENTSVM( $\tilde{Z}^E, C_\theta, \sigma_\theta, \beta_\theta$ )
 $e_{\text{BER}} \leftarrow$  TAUXERREURBALANCÉ( $D, Z^V$ )
 $n_{\text{VS}} \leftarrow$  NOMBREDEVECTEURSSUPPORT( $D$ )
 $q \leftarrow$  QFD( $e_{\text{BER}}, n_{\text{SV}}, \beta_\theta$ )
retourner  $q$ 

```

Algorithme 5: Etapes d'entraînement d'un SVM pour un modèle  $\theta$ .

Reste à définir la Qualité d'une Fonction de Décision (QFD) à partir d'un compromis entre le taux d'erreur balancé (BER) et la complexité de la fonction de décision produite. Cette complexité dépend du nombre de vecteurs de support nécessaires à sa définition et du nombre d'attributs utilisés par cette fonction. Notre définition empirique de ce critère de qualité est la suivante :

$$q_{\text{FD}} = (1 - e_{\text{BER}}) - c_{\text{VS}} \log_2(1 + n_{\text{VS}}) - c_\beta \log_2(\text{coût}(\beta)) \quad (3.12)$$

avec  $\text{coût}(\beta) = \sum \beta_i \kappa_i$ . Les constantes  $\kappa_i$  représentent le coût relatif d'extraction des attributs  $i$  si l'on souhaite tenir compte du temps de calcul associé à chaque attribut.

Les valeurs des constantes  $c_{\text{VS}}$  et  $c_\beta$  servent à la pénalisation de la complexité et sont fixées par l'utilisateur. Elles représentent respectivement l'importance de la pénalisation due à l'augmentation du nombre de vecteurs de support et à l'augmentation du nombre d'attributs utilisés ou des coûts associés. L'idée directrice, pour cette définition, est qu'un doublement du nombre de vecteurs de support doit au moins correspondre à une diminution de la valeur de  $c_{\text{VS}}$  pour le taux d'erreur balancé pour que le processus décisionnel correspondant soit considéré comme étant de qualité supérieure. Lorsque les valeurs de  $c_{\text{VS}}$  et  $c_\beta$  sont faibles ( $< 0,01$ ), c'est le principe de parcimonie qui est privilégié.

L'espace de recherche  $\Theta$  a une cardinalité importante lorsque l'on souhaite trouver un modèle  $\theta \in \Theta$  qui réalise la recherche d'une fonction de décision optimale selon les critères que nous avons définis. Il faut alors disposer d'une méthode de sélection qui puisse explorer l'espace  $\Theta$  de manière plus efficace et dans des temps de calcul plus raisonnables qu'une évaluation systématique de tous les modèles possibles. Nous avons choisi une méta-heuristique de recherche avec tabous pour effectuer cette recherche car elle a déjà été utilisée avec succès d'une part pour la sélection d'attributs [KORYC104] et d'hyper-paramètres [CAWLEY01] d'un SVM d'autre part. Le synopsis de la recherche avec tabous est le présenté dans l'algorithme 6.



1. Choisir une solution initiale  $\theta_0 \in \Theta$  ( $\theta^* = \theta_0$ ,  $\bar{\Theta} = \emptyset$ ,  $it = 0$ ),
2. Sélectionner la meilleure solution non taboue  $\theta_{it+1} = \arg \max_{\theta \in \tilde{V}(\theta_{it})} q(\theta)$  dans le voisinage  $\tilde{V}(\theta_{it})$  de  $\theta_{it}$ ,
3. Mettre à jour la meilleure solution actuelle  $\theta^*$  si  $q(\theta_{it+1}) > q(\theta^*)$ ,
4. Mettre à jour la liste des solutions taboues,
5. Répéter les étapes 2 à 5 ( $it = it + 1$ ) jusqu'à ce que le critère d'arrêt soit vrai,
6. Retourner  $\theta^*$ .

Algorithme 6: Synopsis de la recherche tabou.

La recherche avec tabous [DRÉO03, GLOVER97] est basée sur une recherche itérative qui choisit dans un voisinage restreint la meilleure solution même si elle est plus mauvaise que celle de l'itération précédente. Une mémoire à court terme est utilisée pour éviter tout cycle visitant périodiquement le même optimum local. À partir des derniers mouvements mémorisés, un ensemble de solutions est considéré comme tabou. Un mouvement n'est donc réalisable que vers l'une des solutions voisines qui n'est pas taboue. Le choix des solutions taboues peut être tout simplement les dernières solutions visitées, mais en général le critère déterminant les solutions taboues à un moment précis de la recherche dépend du problème à optimiser et de la représentation d'une solution. Nous avons donc proposé d'utiliser la recherche avec tabous pour optimiser l'ensemble des paramètres liés à un SVM (attributs pertinents et hyper-paramètres), ainsi que la sélection du niveau de simplification. L'objectif est de permettre une sélection rapide de bons modèles à partir d'une représentation simplifiée et de n'utiliser les représentations moins simplifiées que si elles diminuent suffisamment le taux d'erreur. Pour obtenir un tel résultat, la recherche avec tabous doit commencer avec une représentation fortement simplifiée de la base d'entraînement (une valeur de  $k$  faible). Ceci permet, dès les premières itérations de la recherche, de produire des solutions qui permettront de réaliser une sélection rapide des attributs et des hyper-paramètres pour identifier des régions d'intérêt dans l'espace de recherche. De plus, nous avons défini des stratégies spécifiques d'intensification et de diversification afin d'améliorer les performances de notre modèle de recherche avec tabous. L'intensification a pour objectif de réaliser une recherche exhaustive dans des zones prometteuses de l'espace de recherche et faiblement explorées. La diversification a pour objectif de réaliser des modifications significatives de la solution pour permettre de quitter rapidement une zone de recherche peu efficace où déjà explorée par intensification afin de découvrir des zones prometteuses de l'espace de recherche.

Nous avons mené de nombreuses expérimentations pour mesurer l'efficacité de la méthode de sélection de modèle proposée pour produire des fonctions de décision efficaces en généralisation et de complexité réduite. Nous n'en montrons ici qu'une petite partie sur la base OpticDigits [ASUNCIO7] à travers le tableau 3.6 où les notations suivantes sont utilisées :

- $T_{\text{app}}$  désigne la durée en secondes nécessaires à notre méthode pour la sélection d'un modèle  $\theta^*$  qui est considéré comme la solution optimale du problème.
- $k$  désigne le niveau de simplification optimal de la base d'entraînement,
- $q_{\text{BER}}$  est égal à  $1 - e_{\text{BER}}$  et l'erreur BER est évaluée à partir de la base de test  $Z^T$ ,
- $n_{\text{VS}}$  correspond au nombre de vecteurs de support utilisés par la fonction de décision,
- $n_{\text{AS}}$  correspond au nombre d'attributs sélectionnés relativement au modèle  $\theta^*$ .

La base OpticDigits contient 3064 et 759 exemples dans respectivement les bases d'apprentissage et de test. Chaque exemple est décrit par un vecteur de 64 attributs. La base est divisée en 10 classes. Les SVM étant des classificateurs binaires, cela impose de choisir un mode de décomposition en sous-problèmes du problème initial (voir section 3.5). Nous avons choisi ici le mode de décomposition *un-contre-tous*, soit une fonction de décision par classe. Ces fonctions de décisions sont optimisées indépendamment les unes des autres. Aux vues des résultats, on remarque que même sans pénalisation de la complexité des fonctions de décision, des versions plus ou moins simplifiées de la base d'apprentissage sont utilisées. L'introduction d'une pénalisation de la complexité permet d'obtenir des fonctions de décision de faibles complexités, de sélection de modèle et d'apprentissage rapides, sans pour autant trop dégrader les performances.

FD	$c_{VS} = c_{\beta} = 10^{-2}$					$c_{VS} = c_{\beta} = 0$				
	$T_{app}$	$k$	$n_{VS}$	$n_{AS}$	$\hat{q}_{BER}$	$T_{app}$	$k$	$n_{VS}$	$n_{AS}$	$\hat{q}_{BER}$
<b>0</b>	212	4	5	16	99,93%	395	4	16	64	100,0%
<b>1</b>	694	4	23	29	99,71%	2406	6	95	21	100,0%
<b>2</b>	288	4	10	18	99,93%	328	4	29	31	100,0%
<b>3</b>	221	2	7	14	98,70%	335	4	19	25	99,28%
<b>4</b>	271	4	10	19	100,0%	531	4	27	62	99,35%
<b>5</b>	242	4	8	14	99,93%	416	5	29	17	100,0%
<b>6</b>	248	2	7	22	100,0%	4169	4	32	64	100,0%
<b>7</b>	254	3	9	23	99,78%	330	4	12	27	99,78%
<b>8</b>	767	6	32	22	97,85%	4597	8	193	59	97,85%
<b>9</b>	563	4	32	29	98,25%	575	4	32	29	98,32%
	$\sum T_{app}$	$1/n_{\omega} \sum k$	$\sum n_{VS}$	$\cup \beta_i$	$\hat{q}_{BER}$	$\sum T_{app}$	$1/n_{\omega} \sum k$	$\sum n_{VS}$	$\cup \beta_i$	$\hat{q}_{BER}$
	3569	3,7	143	59	97,35%	14082	4,7	484	64	97,42%

TAB. 3.6 – Sélection de modèle avec un schéma *un-contre-tous* pour la base **OpticDigits**.

### 3.4.3 Application à la classification de données d'images

#### 3.4.3.1 Classification de pixels

Contrairement à l'approche par fusion de classificateurs que nous avons précédemment exposée (Section 3.3), nous cherchons cette fois à concevoir un classificateur basé sur uniquement des SVM et qui soit globalement optimisé. Comme la complexité des SVM croît avec la taille de la base d'apprentissage, leur application directe sans précautions particulières est difficile pour la classification de pixels. En effet, la base de données de pixels dont nous disposons est de grande taille et très redondante car beaucoup de pixels d'une même classe sont similaires. Même si un SVM permettra d'obtenir de bons taux de classifications, la fonction de décision qu'il réalise devra être interrogée pour la classification d'un grand nombre de pixels lorsque l'on voudra classifier les pixels d'une nouvelle image. Cela peut devenir très handicapant si l'on veut des segmentations qui soient rapides. Nous avons alors mis en oeuvre notre méthode de construction de fonctions de décision de complexités réduites pour la classification de pixels. Comme les SVM sont des classificateurs binaires, nous avons à construire des fonctions de décision selon un schéma de décomposition donné (*un-contre-tous* ou *un-contre-un*). Chaque fonction de décision sera optimisée indépendamment. Pour déterminer le meilleur modèle  $\theta^*$  pour chaque fonction de décision binaire, nous devons intégrer à ce modèle les paramètres que nous souhaitons régler. Pour effectuer une classification de pixels par une fonction de décision, nous devons simplifier la base d'apprentissage (sélection d'exemples pertinents), simplifier la représentation associée aux pixels (sélection d'attributs pertinents) et sélectionner les hyperparamètres optimaux. Cela correspond donc bien au précédent cadre que nous avons présenté.

De part nos travaux en fusion de classification de pixels, nous avons pu mettre en évidence que la représentation associée à un pixel a une influence sur les résultats. Plutôt que de choisir *a priori* un espace couleur donné pour cette représentation des pixels, nous avons choisi de chercher à l'optimiser en considérant ceci comme un problème de sélection d'attributs. En conséquence, chaque pixel est décrit par un vecteur de dimension 25 qui contient les composantes couleur extraites des principaux espaces couleur [TRÉM04] à savoir <sup>1</sup>  $(R, V, B)$ ,  $(X, Y^1, Z)$ ,  $(L^*, a^*, b^*)$ ,  $(L^*, u^*, v^*)$ ,  $(L^1, C, H^1)$ ,  $(Y^2, Ch_1, Ch_2)$ ,  $(I_1, I_2, I_3)$ ,  $(H^2, S, L^2)$ ,  $(Y^3, C_b, C_r)$ . Sélectionner les attributs colorimétriques pertinents dans un espace colorimétrique hybride (provenant de plusieurs espace couleur) de grande dimension a déjà été proposé en utilisant une approche en cascade [VANDEN03] (les attributs sont filtrés avant classification sans optimisation globale) et cela permet d'améliorer les résultats. Nous proposons donc de construire un espace couleur hybride optimal qui permet d'améliorer la qualité des fonctions de décision produites par des SVM. En conséquence, chaque pixel est représenté par un vecteur de dimension 25 correspondant à l'espace couleur hybride initial suivant :

$$E = (R, V, B, X, Y_1, Z, L^*, a^*, b^*, u^*, v^*, C, H^1, Y^2, Ch_1, Ch_2, I_1, I_2, I_3, H^2, S, L^2, Y^3, C_b, C_r) \quad (3.13)$$

Nous avons donc appliqué notre construction de fonctions de décision de complexités réduites par recherche tabou au problème de la classification de pixels en imagerie microscopique couleur. Nous rappelons que nous cherchons à discriminer trois classes (fond, cytoplasmes et noyaux), ce qui nous amène à optimiser indépendamment trois fonctions de décisions pour estimer l'appartenance à chacune de ces classes. Le tableau 3.7 présente les résultats que nous avons obtenu avec une décomposition *un-contre-un* (la classe finale retenue est obtenue par le schéma de décodage de Price [PRICE94], voir la section 3.5).

Nous avons mesuré le taux de reconnaissance du schéma global ( $T_R$ , taux d'erreur balancé mesuré sur une base de test indépendante), le nombre total de vecteurs de support ( $n_{VS}$ ), le temps d'entraînement total (TE), le temps moyen de classification d'une image complète ( $TCT$ ). Pour le nombre de vecteurs support et les temps d'entraînement, ce sont les sommes des quantités associées à chaque fonction de décision binaire. Les résultats sont présentes avec l'utilisation d'un espace couleur hybride (ECH) et d'un espace couleur classique (ECC) pour lesquels aucune sélection d'attributs n'est effectuée lors de l'optimisation (chaque pixel est alors décrit par trois attributs). La première remarque qui s'impose est que même si des schémas basés sur l'exploitation d'espaces couleur classiques peuvent être efficaces, ceux-ci sont globalement plus complexes que ceux construits par notre approche réalisant une sélection d'un espace couleur hybride. En modifiant la pénalité sur le nombre de vecteur supports, nous pouvons obtenir un schéma de classification encore plus parcimonieux et de performances quasiment équivalentes (dernière ligne du tableau 3.7). Un compromis entre efficacité et complexité permet donc bien de sélectionner un modèle performant. La classification rapide de pixels par SVM permet de produire les cartes de probabilité d'appartenance de chaque pixel à l'une des classes. On notera que les SVM ne fournissent pas directement des probabilités, mais la distance à la frontière de décision. PLATT [PLATT99A] a néanmoins proposé une méthode pour produire une estimation de cette probabilité, c'est ce que nous réalisons. Les figures 3.9(b) à 3.9(d) illustrent les résultats obtenus avec l'image microscopique de la figure 3.9(a). La figure 3.9(e) présente le résultat obtenu en affectant chaque pixel à sa classe la plus probable, les probabilités finales étant calculées avec la méthode de Price [PRICE94]. Nous avons, à partir de ces classifications, défini

<sup>1</sup>Nous ajoutons des exposants à certaines composantes couleur pour différencier celles qui ont le même nom dans différents espaces couleur mais pas les mêmes formules de calcul.

ECC	$T_R$	$n_{VS}$	TE	$\overline{TCI}$
$RVB$	86,86	988	968	21,38
$XY^1Z$	87,06	880	1258	20,18
$L^*a^*b^*$	86,69	50	1269	1,97
$L^*u^*v^*$	80,00	13	517	1,07
$LCH^1$	86,24	128	812	3,81
$Y^2Ch_1Ch_2$	84,57	35	429	1,42
$I_1I_2I_3$	86,57	851	1488	19,04
$H^2SL^2$	86,36	146	762	3,81
$Y^3CBCR$	86,28	36	731	1,51
Moyenne	85,63	347	915	8,24
ECH	$T_R$	$n_{VS}$	TE	$\overline{TCI}$
$RVY^1LI_2$	86,97	12	4622	2,13
$Ru^*v^*$ ( $c_{VS} = 0,01$ et $c_\beta = 0,01$ )	86,67	10	1069	1,71

TAB. 3.7 – Taux de reconnaissance global ( $T_R$ ), nombre de vecteurs de support ( $n_{VS}$ ), temps d'entraînement (TE) en seconde, temps de classification moyen d'une image ( $\overline{TCI}$ ) en seconde pour une décomposition *un-contre-un* avec un espace couleur classique (ECC) ou bien un espace couleur hybride (ECH). Les constantes de pénalisation sont  $c_{VS} = 0,003$  et  $c_\beta = 0,01$  sauf pour la dernière ligne.

un schéma de segmentation complet dont les paramètres sont optimisés à nouveau grâce à une recherche tabou à partir d'une mesure d'évaluation de la segmentation (voir [13, 14] pour plus de détails).

### 3.4.3.2 Classification de cellules

A partir d'une segmentation d'images microscopiques, nous obtenons une partition de l'image en régions qui permet d'obtenir, après séparation des cellules accolées, un ensemble de cellules [8]. Nous pouvons alors chercher à classer ces cellules dans différentes classes cellulaires. Lors de ma thèse, j'ai pu constituer une base de cellules extraites par un schéma de segmentation morphologique pour des images microscopiques de la cytologie des séreuses [3, 4, 6, 8, 21]. La figure 3.10 présente le résultat d'une telle segmentation ainsi que l'ensemble des cellules à classer. Ces cellules sont décrites par 46 attributs (de couleur, de forme et de texture) et réparties dans 18 classes grâce à un étiquetage précis effectué par des experts en pathologie du Centre Hospitalier du Cotentin. Deux problématiques peuvent alors être traitées : d'une part la différenciation entre tous les types cellulaires ou bien la différenciation entre deux regroupement de ces types (cellules bénignes et cellules malignes). Lors de ma thèse, j'avais pu montrer qu'avec une architecture de réseaux de neurones basée sur un ensemble de classificateurs neuronaux binaires, de bons taux de classification pouvaient être obtenus : 84.55% pour la distinction entre les 18 types cellulaires (65.8% sans sélection d'attributs) et 96.7% pour la distinction maligne bénigne (taux d'erreurs balancés). Le modèle proposé effectuait également une sélection d'attributs et le réglage d'un hyper-paramètre (le nombre de neurones sur la couche cachée). Ces travaux ont fait l'objet de nombreuses publications lors de ma thèse et un peu après.

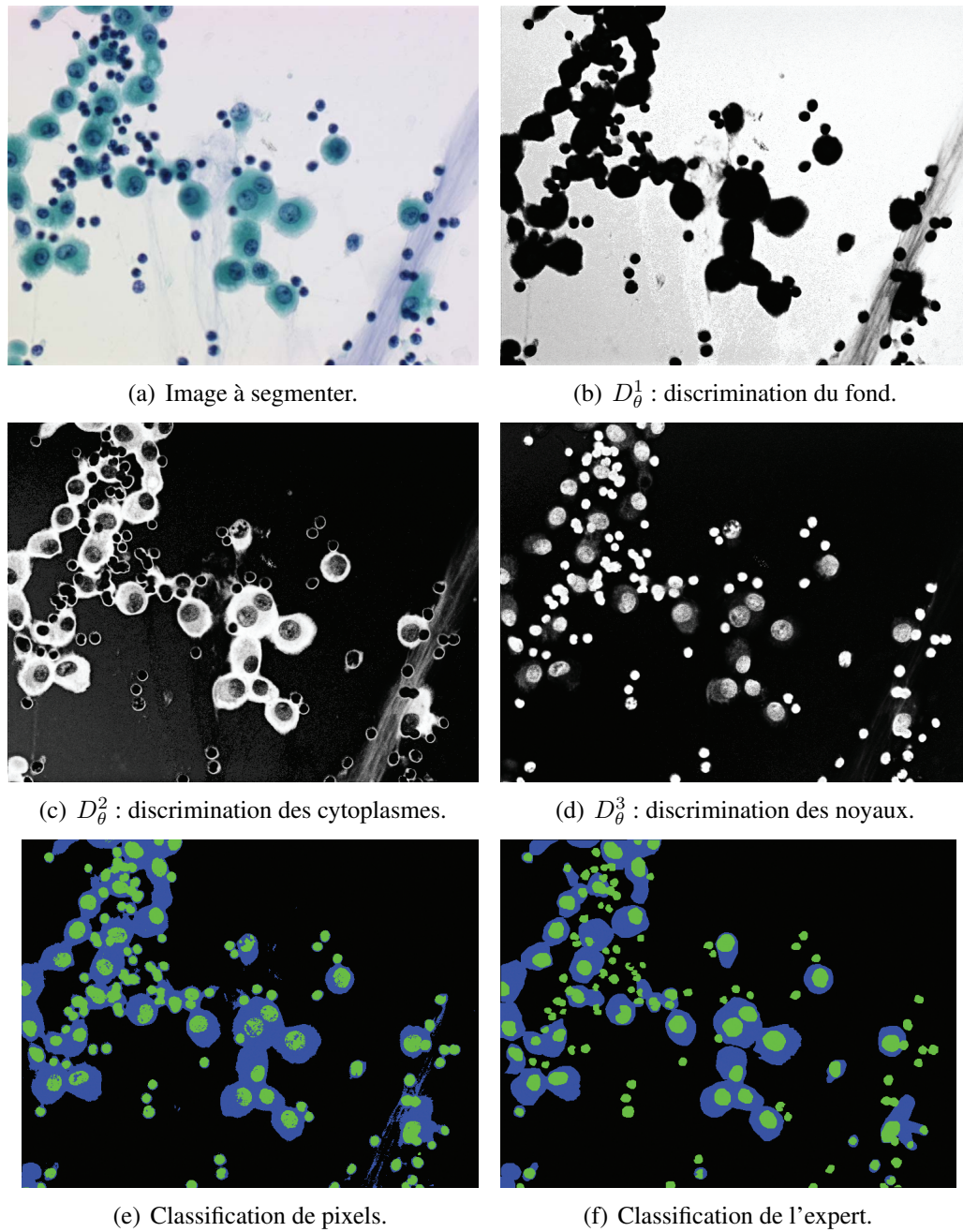


FIG. 3.9 – Classification de pixels produite par la combinaison des fonctions de décision ( $D_{\theta}^1$ ,  $D_{\theta}^2$ ,  $D_{\theta}^3$ ).

☞ **Publications effectuées lors de ma thèse et liées à la microscopie** : [3, 4, 8, 21, 24, 25, 28–30, 59]

Nous avons pu améliorer ces taux grâce à la construction de fonctions de décision simplifiées basées sur des SVM : un taux de 97.4% pour la distinction bénigne/maligne est obtenu. La distinction entre tous les types cellulaires (sans sélection d'attributs) atteint un taux de 66.3%. Les résultats sont donc améliorés par rapport à notre précédente architecture de classification. Le tableau 3.8 présente les taux de reconnaissances obtenus avec différents schémas de décomposition (*un-contre-tous*, *un-contre-un* et une combinaison hybride de ces deux schémas [LEBRUN06]). Les schémas de décomposition et de décodage ont une influence certaine sur les résultats, mais ils sont de toute façon meilleurs que ceux obtenus par notre précédente architecture. Nous reparlerons dans la section 3.5 du problème du choix de la décomposition et du décodage. Notre méthode de construction de fonctions de décisions simplifiées basées sur des SVM est donc à nouveau très efficace (elle pourrait l'être encore plus si nous avions réalisé une sélection des attributs pertinents).

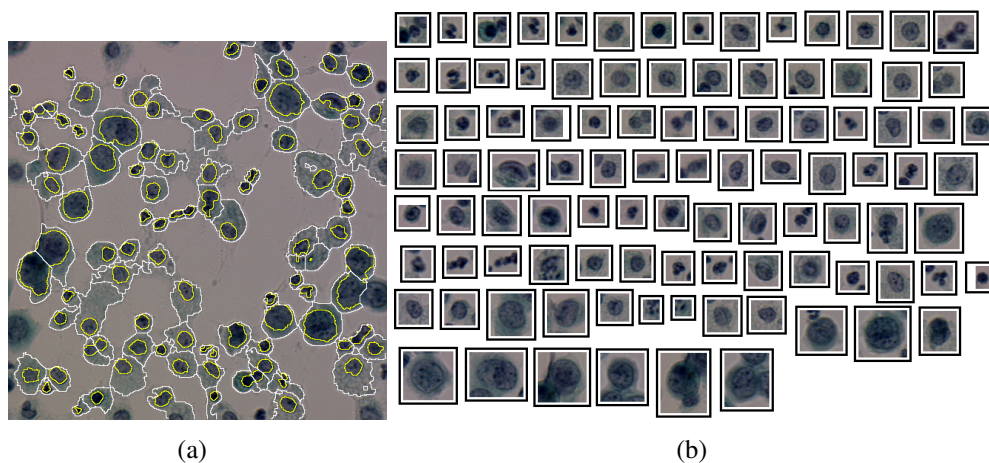


FIG. 3.10 – (a) Une image segmentée de la cytologie des séreuses et (b) les imagettes des noyaux des cellules extraites et à catégoriser.

décomposition	décodage	Taux TC	Taux MB
<i>un-contre-tous</i>	$\max p(\omega_i)$	66,3%	97,4%
<i>un-contre-un</i>	Price	65,9%	97,3%
hybride	hybride	70,4%	97,5%

TAB. 3.8 – Taux de reconnaissance du type cellulaire (Taux TC) et de la nature maligne ou bénigne (Taux MB) obtenus à partir de différents protocoles d'apprentissage. Pour chaque classificateur multi-classes, le type de décomposition et le principe de décodage peuvent être différents (voir la section 3.5 pour plus de précisions sur ces notions).

### 3.4.3.3 Classification d'images

Nous avons également mis en oeuvre notre schéma de classification supervisé pour la mise au point d'une métrique de la qualité d'images compressées [48, 52]. L'UIT a émis une recom-



mandation [UIT00] qui permet à un observateur de classer les images compressées selon leur qualité dans une échelle de cinq classes (présentée par le tableau 3.9) nommée  $Q_{OS}$  (*Quality Opinion Score*). La qualité visuelle d'une image, notée  $Q_{MOS}$  (*Quality Mean Opinion Score*), est alors obtenue par la moyenne des  $Q_{OS}$  de différents experts. La figure 3.11 présente quelques images et leurs scores respectifs. Nous disposons d'une base de 227 images compressées selon

Qualité	
5	Excellente
4	Bonne
3	Relativement bonne
2	Mauvaise
1	Très mauvaise

TAB. 3.9 – Echelle de qualité retenue dans la recommandation UIT-R.

différents niveaux de compression de Jpeg2000 (à partir de 25 images initiales) réparties en deux bases d'apprentissage et de test contenant respectivement 116 et 111 images. Chaque image est décrite par un vecteur de 38 attributs liés à la caractérisation du système visuel humain. Ces images ont été classées par 25 observateurs dans la précédente échelle. Nous avons défini deux



FIG. 3.11 – Quelques images avec leur  $Q_{MOS}$  associé ainsi que la note moyenne ( $\bar{Q}_{OS}$ ), la note maximale ( $Q_{OS}^{max}$ ) et la note minimale ( $Q_{OS}^{min}$ ) attribuées par les observateurs.

mesures de cohérence  $M_1$  et  $M_2$ , la première étant une mesure classique de cohérence et la seconde une mesure de cohérence de rang (une variation d'un rang est acceptée). Les résultats présentés dans le tableau 3.10 montrent que l'opinion des observateurs est très variable et ceci quelle que soit la base utilisée (colonnes 2 à 4). La mesure  $M_2$  montre que la divergence de classification n'excède cependant que très rarement une classe. Nous avons appliqué notre schéma de construction de fonctions de décisions simplifiées pour ce problème. Cette-fois, comme la base d'apprentissage est très réduite, la sélection de modèle utilise une validation croisée de type leave-one-out (effectué avec la nouvelle méthode que nous avons proposée [50]). Nous présentons ici des résultats avec un schéma de décomposition de type *un-contre-tous*. La classification automatique que nous avons proposée permet donc de surpasser les experts humains pris individuellement (mesure  $M_1$ ) et suit globalement le comportement des experts (mesure

Mesure	Moyenne	Min	Max	SVM
$M_1$	$0,529 \pm 0,103$	0,324	0,712	0,693
$M_2$	$0,975 \pm 0,026$	0,909	1,000	0,946

TAB. 3.10 – Statistiques des mesures de cohérence sur la base de test pour les observateurs et notre schéma à base de SVM.

$M_2$ ). C'est le premier modèle d'évaluation de la qualité des images compressées basé sur un apprentissage artificiel.



### 3.5 Schémas multi-classes de combinaison de classificateurs binaires

☞ Mes publications associées : [1, 5, 10, 26, 31, 46, 53]

Pour aborder un problème de classification supervisée avec une base multi-classes deux approches existent [HSU02, RIFKIN04, GUERME05]. La première [GUERME05] correspond à l'optimisation d'un problème unique à partir de l'ensemble des données. La seconde [HSU02, RIFKIN04] correspond à définir un ensemble de sous-problèmes binaires, à réaliser l'optimisation de chacun de ces sous-problèmes et à les combiner ensuite (figure 3.12). La première

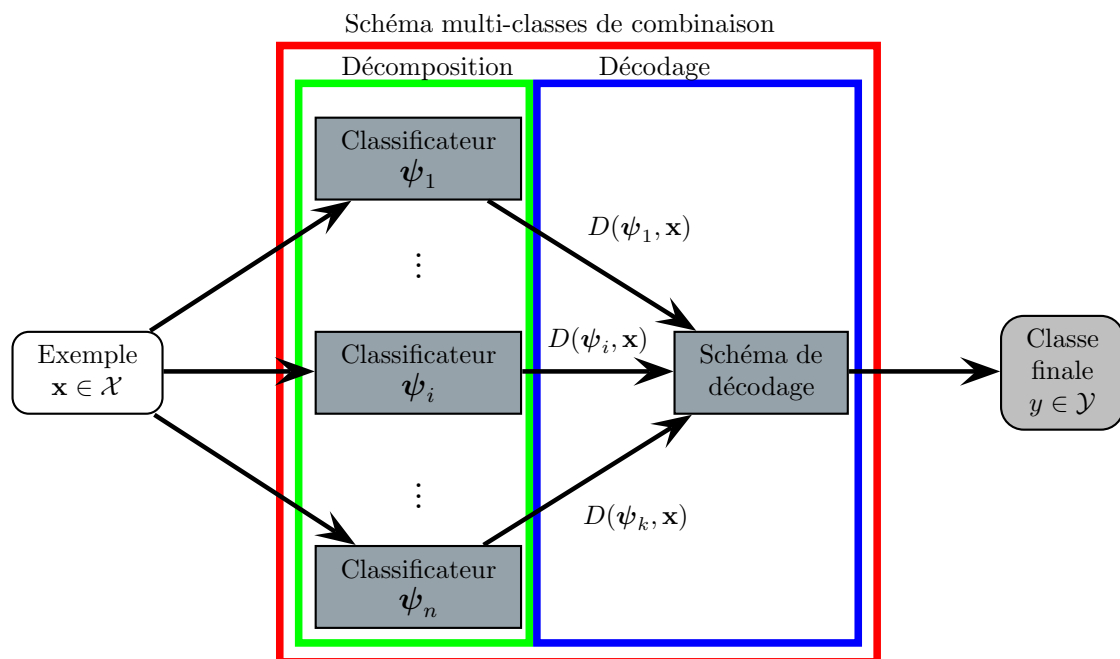


FIG. 3.12 – Les composants d'un schéma multi-classes de combinaison de classificateurs binaires.

approche est la plus simple et est facilement réalisable avec des algorithmes d'apprentissage artificiels qui sont directement applicables pour effectuer une discrimination multi-classes (arbres de décision, réseaux de neurones, k-ppv, etc.). Cependant, certains algorithmes d'apprentissage artificiels, tels que les SVM, ne sont exprimés que pour des problèmes binaires (même s'il existe des extensions multi-classes [CRAMME01, DARCY05], elles sont difficiles à exploiter [RIFKIN04, ABE05]). En conséquence, la seconde approche (de type diviser pour régner) a plusieurs avantages pour les algorithmes permettant uniquement d'effectuer une classification binaire [ALLWEI00, ABE05], mais cela est également le cas pour des algorithmes qui sont directement applicables à des problèmes multi-classes [PRICE94, FÜRNK01, SAVICK03, OU04, 10, 46]. La raison est simple. Résoudre directement un problème multi-classes est difficile car un grand nombre de frontières de décision doivent être apprises simultanément. À l'inverse, les problèmes binaires étant spécialisés pour la discrimination entre deux sous-ensembles, seule une frontière de décision généralement plus simple doit être apprise. Dans [PRICE94] un exemple significatif le montre expérimentalement. Pour discriminer des chiffres manuscrits, un classificateur neuronal multi-classes ainsi qu'un ensemble de classificateurs binaires sont construits.

PRICE remarque alors que la majorité des frontières de décision binaires sont linéaires et la complexité globale du schéma multi-classes basé sur une décomposition binaire est beaucoup moins grande que le schéma multi-classes direct. Le même type de comportement a été constaté pour les arbres de décision [FÜRNK01, FRANK04]. En conséquence, des approches basées sur une décomposition d'un problème en sous-problèmes binaires bénéficie de plusieurs avantages majeurs : simplicité des fonctions de décision, apprentissage rapide et donc une sélection de modèle plus aisée (attributs pertinents, exemples pertinents, hyper-paramètres).

Dans nos travaux, nous nous sommes intéressés à concevoir des schémas multi-classes de combinaison de classificateurs binaires. La figure 3.12 présente le schéma global d'une telle approche. Pour concevoir un tel schéma, deux éléments doivent être déterminés : le schéma de décomposition et le schéma de décodage. Une fois le schéma de décomposition déterminé, nous avons à construire  $k$  fonctions de décision binaires  $D(\psi_i)$ . Une fois l'apprentissage des classificateurs binaires effectués, classer un exemple revient à interroger l'ensemble (ou un sous-ensemble) des classificateurs et à déterminer la classe finale par un schéma de décodage des sorties fournies par les classificateurs. La sortie d'un classificateur  $\psi_i$  pour un exemple  $\mathbf{x} \in \mathcal{X}$  sera notée  $D(\psi_i, \mathbf{x})$ . Un schéma de décodage effectue alors généralement un calcul de confiance d'appartenances aux classes :  $g : \mathbb{R}^k \rightarrow \mathbb{R}^{|\mathcal{Y}|}$  puis une décision est prise à partir de ces dernières par  $h : \mathbb{R}^{|\mathcal{Y}|} \rightarrow \mathcal{Y}$ . Un schéma de décodage effectue donc la transformation suivante :  $h \circ g(D(\psi_1, \mathbf{x}), \dots, D(\psi_k, \mathbf{x}))$ . Nous noterons indifféremment le nombre de classes à discriminer par  $n_c$  ou  $|\mathcal{Y}|$ .

Dans la suite, nous présentons les schémas de décomposition usuels puis les schémas de décodage associés. Nous présentons finalement nos propositions de schémas de décodage.

### 3.5.1 Schémas de décomposition

La décomposition d'un problème multi-classes en  $k$  problèmes binaires est la première étape d'un schéma de multi-classes de combinaison. Elle correspond à définir  $k$  vecteurs  $\psi$ . A partir de ces vecteurs, une matrice  $\Psi$  de codage du problème multi-classes est définie.

$$\Psi = \begin{array}{c} \psi_1 \\ \vdots \\ \psi_k \end{array} \begin{array}{c} \omega_1 \quad \cdots \quad \omega_{n_c} \\ \hline \psi_{1,1} \quad \cdots \quad \psi_{1,n_c} \\ \vdots \quad \ddots \\ \psi_{k,1} \quad \quad \psi_{k,n_c} \end{array} \quad (3.14)$$

Chaque ligne de  $\Psi$  correspond à un vecteur  $\psi_i$  et chaque colonne de  $\Psi$  indique le problème binaire dans lequel les exemples d'une classe donnée sont utilisés. La structure de la matrice  $\Psi$  définit la nature du schéma de combinaison utilisé. Un sous-problème de discrimination binaire issu d'un problème multi-classes est donc défini à partir d'un vecteur  $\psi_i$  (nous désignons également le classificateur par cette même notation). Ce vecteur représente pour chaque classe  $\omega_y$  de la base initiale sa correspondance avec des exemples considérés comme appartenant à la classe positive ( $\psi_{i,y} = +1$ ), à la classe négative ( $\psi_{i,y} = -1$ ) ou non présent ( $\psi_{i,y} = 0$ ) pour le problème binaire induit. La définition d'une base d'apprentissage binaire  $Z(\psi_i)$  à partir d'une base multi-classes  $Z$  et de la transformation binaire induite par  $\psi_i$  est la suivante :

$$Z(\psi_i) = \{(\mathbf{x}, +1) \mid (\mathbf{x}, y) \in Z, \psi_{i,y} = +1\} \cup \{(\mathbf{x}, -1) \mid (\mathbf{x}, y) \in Z, \psi_{i,y} = -1\} \quad (3.15)$$

$D(\psi_i)$  désigne la fonction de décision du classificateur  $\psi_i$  produite à partir d'une base d'entraînement  $Z(\psi_i)^E$ .

Dans la majorité des schémas de combinaison utilisés, cette structure est prédéfinie [DIETTE95, VAPNIK98, KREBEL99]. Pour d'autres méthodes, la structure de la matrice (les différentes décompositions binaires) est produite par un schéma d'apprentissage [PÉREZ-02, ABE05]. Nous ne considérons que des décompositions pré-définies et nous présentons les principales dans la suite ; il en existe d'autres (on consultera par exemple [LEI05, MOREIR98]).

### 3.5.1.1 Un contre tous

La méthode *un-contre-tous* (*One Versus All* [VAPNIK98, RIFKIN04]) est la plus simple et elle produit autant de problèmes binaires que de classes. Chaque problème binaire correspond à la discrimination des exemples d'une classe avec ceux des autres classes. La matrice (3.16) illustre la décomposition *un-contre-tous* pour un problème avec 4 classes.

$$\Psi = \begin{bmatrix} +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & +1 \end{bmatrix} \quad (3.16)$$

Cette méthode de décomposition produit peu de problèmes binaires. Cependant, ceux-ci ont l'inconvénient d'utiliser systématiquement l'ensemble des exemples du problème multi-classes et par conséquent les temps d'apprentissage de chaque sous-problème binaire sont importants [LEI05]. De plus, les fonctions de décision produites sont généralement complexes.

### 3.5.1.2 Un contre un

La méthode *un-contre-un* (*One Versus One* [KREBEL99], *Round Robin* [FÜRNK03], ou *Pair-wise Coupling* [PRICE94]) est sûrement la méthode la plus utilisée. L'idée est que pour chaque sous-problème binaire seules deux classes du problème initial sont retenues. Le nombre de problèmes binaires induits par cette décomposition est de  $n_c(n_c - 1)/2$ . La matrice (3.17) illustre cette décomposition pour un problème avec 4 classes.

$$\Psi = \begin{bmatrix} +1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ +1 & 0 & 0 & -1 \\ 0 & +1 & -1 & 0 \\ 0 & +1 & 0 & -1 \\ 0 & 0 & +1 & -1 \end{bmatrix} \quad (3.17)$$

Avec cette décomposition, chaque problème binaire est de taille plus faible que le problème initial, ce qui réduit les temps d'entraînement liés à chaque problème de discrimination. Ce gain est atténué par l'augmentation du nombre de problèmes binaires à traiter. La discrimination est facilitée, car la frontière de décision ne concerne que deux classes à chaque fois. Pour simplifier les notations, nous noterons  $\psi_{w_i, w_j}$  le classificateur binaire qui discrimine la classe  $w_i$  de la classe  $w_j$ .

### 3.5.1.3 ECOC : Error Correcting Output Codes

Les matrices  $\Psi$  de type *ECOC* comportent  $k$  lignes et le nombre de lignes désigne la longueur du code utilisé [DIETTE95]. Le principe est de produire un grand nombre de lignes (donc

de décompositions binaires) et de choisir la classe la plus cohérente par rapport à l'ensemble de ces prédictions. Le nombre maximum de lignes, sans classificateurs binaires redondants, dépend de la possibilité d'avoir ou non des valeurs nulles dans la matrice  $\Psi$  [DIETTE95, ALLWEI00]. Si le type de matrice  $\Psi$  choisi permet des valeurs nulles, le nombre maximum de lignes (non redondantes) est borné par  $(3^{n_c} - 1)/2$ . Dans le cas contraire, le nombre maximum de lignes est de  $(2^{n_c-1} - 1)$ . La matrice 3.18 présente un code exhaustif sans valeurs nulles pour un problème à quatre classes.

$$\Psi = \begin{bmatrix} +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & +1 & +1 & -1 \end{bmatrix} \quad (3.18)$$

Les décompositions précédentes sont donc des cas particuliers des décompositions de type *ECOC*.

### 3.5.2 Schémas de décodage

Il est possible de distinguer deux familles de principes de décodage (ou combinaison de classificateurs binaires). La première famille correspond à un décodage direct. Il détermine tout d'abord la sortie des  $k$  fonctions de décisions produites par la méthode de décomposition. A partir de l'ensemble de ces prédictions, une valeur de confiance (ou probabilité) est déterminée pour chaque classe du problème initial ( $g : \mathbb{R}^k \rightarrow \mathbb{R}^{|\mathcal{Y}|}$ ) et la plus probable est choisie ( $h : \mathbb{R}^{|\mathcal{Y}|} \rightarrow \mathcal{Y}$ ). La seconde famille correspond à un décodage en cascade et à chaque niveau de ce processus de décodage un sous ensemble restreint de classes candidates est retenu (directement ou par élimination d'autres classes). A la dernière étape de ce processus, une seule classe est finalement retenue. Nous présentons les principaux schémas de décodage de la littérature.

#### 3.5.2.1 Décodage direct

Dans un décodage direct, à partir des sorties des  $k$  classificateurs binaires, on cherche tout d'abord à obtenir des confiances en l'appartenance aux différentes classes possibles ( $g : \mathbb{R}^k \rightarrow \mathbb{R}^{|\mathcal{Y}|}$ ) et la classe la plus probable est choisie ( $h : \mathbb{R}^{|\mathcal{Y}|} \rightarrow \mathcal{Y}$ ). La composition  $h \circ g$  détermine alors le schéma de décodage. On trouve différents décodages directs suivant le schéma de décomposition

##### ❶ Décomposition *un-contre-tous*

Pour la décomposition *un-contre-tous*, le calcul des confiances est direct et ne nécessite qu'une normalisation :

$$p(\omega = \omega_i | \mathbf{x}) = \frac{D(\psi_i, \mathbf{x})}{\sum_{j=1}^{n_c} D(\psi_j, \mathbf{x})} \quad (3.19)$$

On a ensuite  $w(\mathbf{x}) = \arg \max_{\omega_i} p(\omega = \omega_i | \mathbf{x})$  soit  $h = \arg \max$ .

### ⊗ Décomposition *un-contre-un*

Dans une méthode de décomposition *un-contre-un*, on cherche à définir les valeurs des  $p(\omega = \omega_i | \mathbf{x})$  à partir des estimations des  $p(\omega = \omega_i | \mathbf{x}, \omega \in \{\omega_i, \omega_j\}) = D(\boldsymbol{\psi}_{\omega_i, \omega_j}, \mathbf{x})$  (ou  $1 - D(\boldsymbol{\psi}_{\omega_i, \omega_j}, \mathbf{x})$ ) fournies par chacun des classificateurs binaires  $\boldsymbol{\psi}_{\omega_i, \omega_j}$  qui discriminent deux classes  $\omega_i$  et  $\omega_j$ . Idéalement, les estimations  $p(\omega = \omega_i | \mathbf{x})$  doivent satisfaire la relation suivante :

$$\forall \omega_i : p(\omega = \omega_i | \mathbf{x}, \omega \in \{\omega_i, \omega_j\}) = \frac{p(\omega = \omega_i | \mathbf{x})}{p(\omega = \omega_i | \mathbf{x}) + p(\omega = \omega_j | \mathbf{x})} \quad (3.20)$$

Pour ce type de schéma de décomposition, on aura  $w(\mathbf{x}) = \arg \max_{\omega_i} p(\omega = \omega_i | \mathbf{x})$  soit  $h = \arg \max$ .

▷ **Méthode basée sur le vote majoritaire :** Ce principe de décodage est basé sur l'utilisation du vote majoritaire et l'estimation de la probabilité d'appartenance à chaque classe est la suivante :

$$p(\omega = \omega_i | \mathbf{x}) = \frac{1}{N} \sum_{\substack{j=1 \\ j \neq i}}^{n_c} I \left( D \left( \boldsymbol{\psi}_{\omega_i, \omega_j}, \mathbf{x} \right) > 0,5 \right) \quad (3.21)$$

avec  $N = \frac{n_c(n_c-1)}{2}$  un coefficient de normalisation et  $I(v)$  une fonction qui vaut 1 si  $v$  est vrai et 0 sinon.

▷ **Méthode de Price et al :** Price et al [PRICE94] démontrent que :

$$\sum_{\substack{j=1 \\ j \neq i}}^{n_c} p(\omega = \omega_i | \mathbf{x}, \omega \in \{\omega_i, \omega_j\}) - (n_c - 2)p(\omega = \omega_i | \mathbf{x}) = 1 \quad (3.22)$$

A partir de (3.20) et (3.22) ils obtiennent alors la relation suivante :

$$p(\omega = \omega_i | \mathbf{x}) = \frac{1}{\sum_{\substack{j=1 \\ j \neq i}}^{n_c} D \left( \boldsymbol{\psi}_{\omega_i, \omega_j}, \mathbf{x} \right) - (n_c - 2)} \quad (3.23)$$

▷ **Méthode d'Hastie et al :** HASTIE et al [HASTIE97] proposent de choisir les valeurs  $p(\omega = \omega_i | \mathbf{x})$  de façon à ce qu'elles minimisent une mesure d'entropie relative (la distance de KULLBACK-LEIBLER [KULLBA51]) entre les valeurs déduites de l'application de la relation (3.20) et celles prédites par les fonctions de décision binaires  $D \left( \boldsymbol{\psi}_{\omega_i, \omega_j}, \mathbf{x} \right)$ . Notons respectivement ces deux types d'estimation  $p_{i,j}$  et  $D_{i,j}$ . Notons également  $p_i$  les valeurs de  $p(\omega = \omega_i | \mathbf{x})$  et  $\mathbf{p}$  le vecteur contenant l'ensemble des valeurs  $p_i$ . L'objectif est de minimiser la fonctionnelle  $l(\mathbf{p})$  suivante :

$$l(\mathbf{p}) = \sum_{i \neq j} n_{i,j} D_{i,j} \ln \frac{D_{i,j}}{p_{i,j}} \quad (3.24)$$

avec  $p_{i,j} = p_i / (p_i + p_j)$  et  $n_{i,j}$  le nombre d'exemples dans la base d'entraînement binaire correspondante. HASTIE et al proposent alors un algorithme itératif simple pour réaliser la minimisation de (3.24). WU et al [WU04] ont proposé une amélioration de cette méthode.

### ③ Décomposition *ECOC*

La décomposition *ECOC* produit une matrice  $\Psi$  de  $k$  lignes et pour chaque ligne une fonction de décision est produite. Si l'on considère le vecteur des sorties des  $k$  classificateurs  $D(\mathbf{x}) = (D(\psi_1, \mathbf{x}), \dots, D(\psi_k, \mathbf{x}))^T$ , la classe la plus probable est celle dont la colonne  $\Psi_i$  de la matrice  $\Psi$  se rapproche le plus du vecteur  $D(\mathbf{x})$ . Le principe de décodage basé sur les codes correcteurs correspond à choisir pour un exemple  $\mathbf{x}$ , la classe  $\omega_i$  la plus proche par rapport à une distance :

$$\omega(\mathbf{x}) = \arg \min_{\omega_i} d(D(\mathbf{x}), \Psi_i) \quad (3.25)$$

La distance  $d$  est une distance de *HAMMING* [DIETTE95] ou bien une fonction de perte [ALLWEI00], ce qui permet de prendre en compte la valeur des sorties et non juste leurs signes.

#### 3.5.2.2 Décodage en cascade

Les schémas de décodage en cascade dépendent du schéma de décomposition mais leur principe est identique. Ils réalisent une élimination itérative des classes jusqu'à n'en retenir qu'une [KREBEL99, PLATT99B]. Le synopsis général d'une procédure d'élimination est le suivant :

1. Choisir une des fonctions de décision binaire  $\psi_l$ .
2. Eliminer la classe  $\omega$  non prédite par  $\psi_l$ .
3. Répéter les étapes 1 à 2 jusqu'à ce qu'il ne reste plus qu'une classe.

Cela correspond alors à définir un arbre de décision dont les noeuds sont des classificateurs binaires et qui est nommé *DDAG* (Decision Directed Acyclic Graph). La figure 3.13 présente un tel *DDAG* pour une décomposition *un-contre-un*. C'est une approche classique dont le défaut majeur réside dans le choix a priori de la structure de l'arbre. On peut alors chercher à optimiser la structure de l'arbre [TAKAHA03] pour obtenir des *ODAG* (Optimized Directed Acyclic Graph) ou bien modifier sa structure en la basant sur des règles d'élimination en tournoi et l'on obtient les *ADAG* (Adaptive Directed Acyclic Graph) [KIJSIR02].

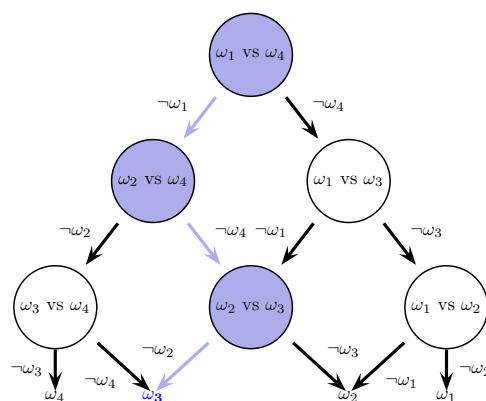


FIG. 3.13 – Arbre de classificateurs binaires dans un schéma de décomposition *un-contre-un*. Les flèches bleues présentent un chemin dans l'arbre.

### 3.5.3 Propositions de schémas de décodage

Les schémas de décodage dépendent du principe de décomposition utilisé et leur influence varie surtout pour des décompositions de type *un-contre-un* (même si l'on peut chercher à améliorer le décodage pour une approche *un-contre-tous* [BEYGEL05]). De plus, les fonctions de décisions produites par ce type de décomposition sont beaucoup moins complexes que celles induites par une décomposition ECOC ou *un-contre-tous* (dans cet ordre de complexité). Ceci a été également constaté par d'autres études [DUAN05, TSUJIN04, HSU02, TAX02]. Cependant, dans une décomposition *un-contre-un*, le nombre important de décompositions et la spécialisation des fonctions de décision rend le décodage plus difficile. Si bien que suivant le schéma de décodage utilisé et différents réglages des paramètres, un schéma de décomposition *un-contre-un* peut être globalement moins performant qu'un schéma de décomposition *un-contre-tous* [RIFKIN04] même si les fonctions de décision du premier schéma sont plus performantes et moins complexes. La principale difficulté dans l'utilisation de schémas de combinaison de classificateurs binaires issus d'une décomposition *un-contre-un* est le problème des inconsistances qui peuvent exister entre l'ensemble des prédictions produites par ces classificateurs binaires. En effet, chaque classificateur est spécialisé dans la discrimination des exemples relativement à deux classes et lorsque qu'un exemple n'appartenant à aucune de ces deux classes est présenté à ce classificateur, la prédiction produite pour cet exemple n'est pas informative. Globalement, il n'y a que  $(n_c - 1)$  classificateurs binaires pertinents pour un exemple donné et  $(n_c - 1)(n_c - 2)/2$  classificateurs binaires non pertinents pour le même exemple. Dans nos travaux, nous nous sommes donc attachés à concevoir des schémas de décodage pour des décompositions de type *un-contre-un*. Nous ne donnons pas ici tous les résultats quantitatifs obtenus car ils ont été obtenus pour des bases de données différentes avec des classificateurs différents et des protocoles expérimentaux parfois différents [LEBRUN06, 1, 5, 10, 26, 31, 46, 53]

#### 3.5.3.1 Décodage direct

Nous avons proposé un nouveau décodage direct [46] basé sur le principe Maximin [DEMYAN90] provenant de la théorie des jeux. Ce principe est basé sur l'hypothèse que la meilleure supposition est celle dont les plus mauvais résultats sont les meilleurs. Dans le cadre d'un schéma de décodage d'une décomposition *un-contre-un*, cela revient à choisir la classe dont la probabilité d'appartenance est la moins mauvaise. Etant donné qu'un classificateur peut être non informatif, une forte probabilité d'appartenance envers une classe donnée ne traduit pas forcément une forte décision envers cette classe. A l'inverse, si la probabilité d'appartenance est faible, il est raisonnable de rejeter cette classe. Nous pouvons alors énoncer ce principe comme suit. Les probabilités d'appartenance à chaque classe sont estimées ainsi :

$$p(\omega = \omega_i | \mathbf{x}) = \min(\{D(\psi_{\omega_a, \omega_b}, \mathbf{x}) | \omega_a = \omega_i \vee \omega_b = \omega_i\}) \quad (3.26)$$

On choisit ensuite la plus probable ( $h = \arg \max$ ). Nous avons comparé expérimentalement ce principe de décodage direct avec les décodages directs classiques (pour des classificateurs de type réseaux de neurones). On ne peut pas conclure qu'une méthode de décodage soit toujours meilleure qu'une autre, cela dépend des bases de données. Cependant, nous pouvons retenir trois méthodes de décodage direct car lors de toutes nos expérimentations, au moins une d'entre elles est toujours la meilleure. Ce sont les décodages par Price, Maximin et ECOC (pour une matrice exhaustive avec valeurs nulles et une fonction de décodage de type perte).



### 3.5.3.2 Décodage en cascade

Initialement, nos propositions de schémas de décodage reposaient tout d'abord sur des schémas en cascade [5] basés sur des DDAG. Les classificateurs que nous avons considérés étaient des réseaux de neurones multi-couches dont les hyper-paramètres ont été optimisés. Nous avons tout d'abord proposé (lors de mes travaux de thèse) de construire une fois pour toutes un DDAG en ordonnant les noeuds de l'arbre selon les performances en généralisation de chaque classificateur mesurées sur une base validation. Puis, nous avons proposé de construire ce DDAG dynamiquement à chaque présentation d'un nouvel exemple à classer [10, 31]. Pour cela, nous avons défini un critère de qualité (QI pour Quality Index) de la décision effectuée par chaque classificateur qui prend en compte la sortie du classificateur  $D(\psi_i, \mathbf{x})$ , une estimation de ses performances en généralisation  $Q(\psi_i)$  sur un ensemble de validation ainsi que la représentativité  $R(Z(\psi_i)^E)$  des données utilisées pour son apprentissage :

$$QI(\psi_i, \mathbf{x}) = D(\psi_i, \mathbf{x}) \times Q(\psi_i) \times R(Z(\psi_i)^E) \quad (3.27)$$

Nous avons appelé ce modèle GNN pour Graph of Neural Networks. La figure 3.14 présente un tel GNN où toutes les possibilités de chemin pour passer d'un classificateur à un autre sont présentées. Deux chemins différents amenant à des décisions différentes sont présentés dans cette figure. Comme pour un DDAG, lorsqu'une classe est éliminée, l'ensemble des classificateurs discriminant cette classe ne sont plus utilisables (cela permet de contraindre à  $n_c - 1$  la profondeur de l'arbre développé). Les GNN correspondent donc à une version alternative des ODAG qui ont été proposés ultérieurement à nos travaux.

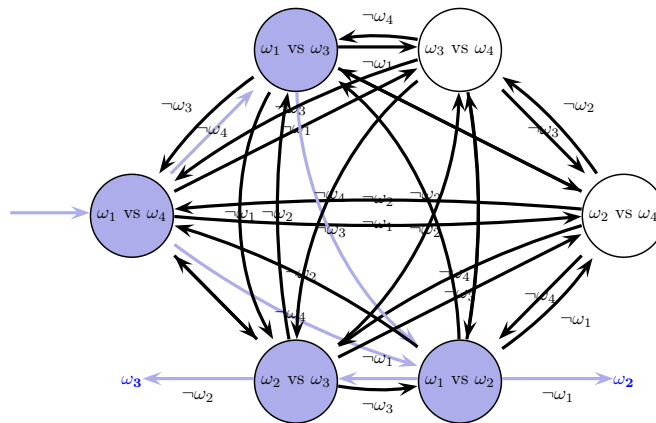


FIG. 3.14 – Graphe de classificateurs binaires.

Même si les GNN nous ont permis d'obtenir de meilleures performances qu'un DDAG, on peut leur faire le même reproche. Si l'un des classificateurs interrogés réalise une erreur, il est impossible d'y pallier. Ceci peut s'en trouver accentué si les classificateurs non pertinents prennent des décisions arbitraires. De plus, si tous les classificateurs sont performants, il devient difficile de choisir leur ordre d'interrogation. Lorsque l'on utilise un DDAG ou un ODAG pour effectuer une classification, cela revient à développer un chemin selon une branche de l'arbre. Puisque ce chemin peut prendre de mauvaises décisions d'élimination, il peut être intéressant de chercher à développer plusieurs branches qui peuvent amener à des décisions différentes et choisir la décision la plus probable (figure 3.14). Nous avons proposé d'associer à chaque

branche développée, un critère qui permet de quantifier la qualité des décisions prises le long de cette branche. Nous l'avons nommé BQI pour Branch Quality Index et il correspond à sommer les QI de chaque classificateur  $\psi_j$  d'une branche  $\phi_i$  explorée dans un DDAG :

$$BQI(\phi_i, \mathbf{x}) = \sum_j^{n_c-1} QI(\psi_j, \mathbf{x}), \psi_j \in \phi_i \quad (3.28)$$

En développant simultanément  $\alpha$  branches dans un DDAG, nous pouvons estimer les confiances en l'appartenance à une classe en sommant les BQI des branches qui ont choisi cette classe et choisir ensuite la plus probable :

$$p(\omega = \omega_i | \mathbf{x}) = \sum_k^\alpha BQI(\phi_k, \mathbf{x}), \phi_k(\mathbf{x}) = \omega_i \quad (3.29)$$

où  $\phi_k(\mathbf{x}) = \omega_i$  signifie que la branche  $\phi_i$  amène à choisir la classe  $\omega_i$ . Cette approche (que nous avons nommée NNIG pour Neural Network Induction Graph) revient donner à combiner les décisions de plusieurs DDAG. Nous avons pu montrer expérimentalement que cela améliore les résultats par rapport à un GNN. Une approche similaire a été proposée dans [FRANK04] pour des classificateurs binaires à base d'arbres de décision.

### 3.5.3.3 Décodage hybride

❶ **Décodage indirect en cascade** Les approches basées sur des DDAG simples donnent généralement des résultats similaires à un décodage direct. Combiner plusieurs décisions provenant d'un DDAG permet d'améliorer les résultats, mais cela augmente la complexité du décodage. Nous avons donc proposé une approche hybride simple (nommée décodage indirect en cascade) qui combine décodage en cascade et décodage direct. Le principe est d'effec-

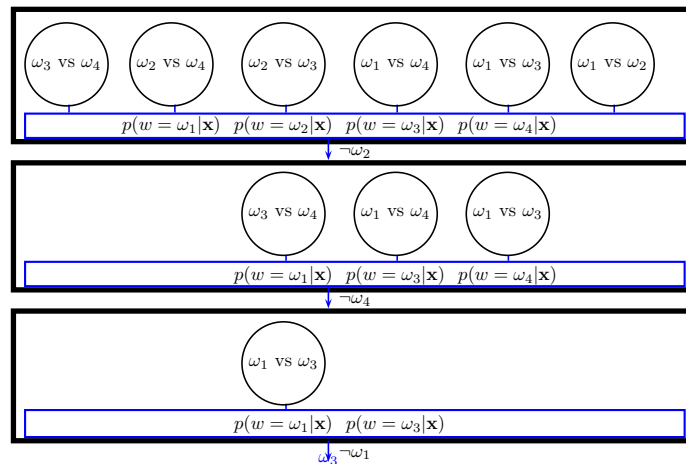


FIG. 3.15 – Décodage indirect en cascade.

tuer un décodage direct et d'éliminer la classe  $\omega_i$  la moins probable (c'est donc un ADAG). Les classificateurs discriminant la classe  $\omega_i$  ne sont alors plus considérés et l'on réitère le principe à partir des classificateurs qui discriminent les classes potentielles restantes. La figure 3.15 illustre ce principe. Nous avons mené plusieurs expérimentations et globalement, ce nouveau

principe de décodage donne de meilleurs résultats qu'un DDAG mais pas qu'un décodage direct. Cependant, cela atténue fortement les différences entre les schémas de décodage direct dont l'influence semble minimale dans notre décodage indirect en cascade. Cela semble cohérent avec les observations de PLATT [PLATT99B] qui précise que l'utilisation d'un DDAG permet de réduire la borne d'erreur du décodage (même si cela ne permet pas de surpasser un décodage direct). Notre décodage indirect en cascade semble réduire encore plus cette borne sans toutefois à nouveau surpasser un décodage direct.

❷ **Décodage direct par stacking** Les schémas de décodage direct permettent d'obtenir de meilleurs résultats que les schémas en cascade simples de type DDAG. Cependant il n'est pas possible de faire un choix entre les décodages par Price, Maximin et ECOC car cela dépend des bases de données. Ceci s'explique facilement car ces schémas de décodage ne peuvent pas s'adapter aux mauvaises configurations qui peuvent apparaître. Par exemple, pour classer un exemple, si des classificateurs non pertinents prédisent une mauvaise classe et si des classificateurs pertinents prédisent cette même mauvaise classe, alors la décision finale risque fort d'être fautive. Savicky et Furnkranz [SAVICK03] ont alors proposé de construire par apprentissage le schéma de décodage direct par stacking [WOLPER92].

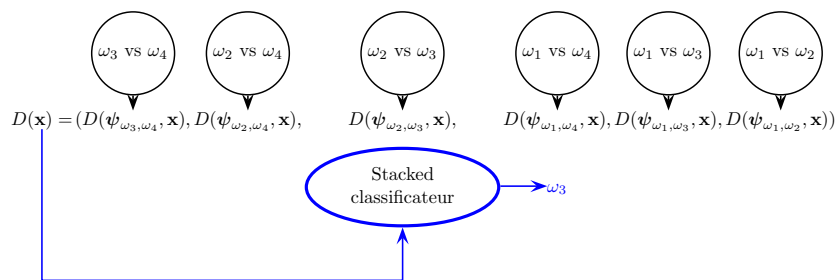


FIG. 3.16 – Décodage direct par Stacking.

Le principe est le suivant, chaque exemple  $x \in Z(\psi_i)^E$  est présenté à tous les classificateurs binaires et l'on obtient un vecteur  $D(\mathbf{x}) = (D(\psi_1, \mathbf{x}), \dots, D(\psi_k, \mathbf{x}))^T$  pour chaque exemple. Ceci permet de construire une nouvelle base de données  $Z = \{(D(\mathbf{x}), y), \forall \mathbf{x} \in \mathcal{X}\}$ . Cette nouvelle base de données peut être alors apprise par un classificateur n-aire. Ce classificateur construit donc une fonction de décision multi-classes qui émule le schéma de décodage. La figure 3.16 résume le fonctionnement de ce schéma de décodage. SAVICKY et FURNKRANZ [SAVICK03] n'ont montré aucune amélioration à l'aide du stacking pour des classificateurs binaires de type arbres de décision, cependant ils utilisaient non pas les sorties des classificateurs mais des prédictions de type classe. Nous avons pu montrer qu'en utilisant les sorties de classificateur neuronaux avec un arbre de décision pour effectuer le stacking, cela permet d'améliorer grandement les résultats [46], ce qui en fait un très bon schéma de décodage. Cependant, en effectuant le stacking par un classificateur n-aire, nous allons exactement à l'inverse de ce qui était voulu avec le schéma de décomposition à savoir simplifier le problème car les arbres de décisions obtenus pour le stacking sont d'une grande complexité. Néanmoins, cela a permis de montrer que les schémas de décodage directs sont limités par leur faible adaptation au cas des classificateurs non pertinents.

❶ **Décodage global par algorithme évolutionnaire** A partir des différents schémas de décomposition de la littérature et nos propositions, nous pouvons globalement faire le classement suivant de leur performances :  $DDAG \leq ADAG \leq \text{Décodage direct} \leq ODAG \leq \text{Décodage par stacking}$ . On voit donc que pour améliorer les performances des schémas de décodage, cela nous amène à prendre en compte le fait que les classificateurs utilisés ne sont pas parfaits et voire même que certains sont parfois non pertinents pour une classification donnée. Jusqu'à présent, toutes les méthodes de la littérature (y compris les nôtres) partent du principe naturel que pour concevoir un schéma multi-classes performant, chacune des fonctions de décision binaires doit être la plus performante possible. C'est ce qui est mis en avant dans les travaux de Rifkin [RIFKIN04] qui montre que si les classificateurs binaires sont bien optimisés, des schémas de décomposition et de décodage différents amènent souvent à des résultats proches. C'est également ce que nous avons pu constater dans nos travaux. Cependant, les schémas de décodage par apprentissage permettent pourtant d'améliorer les performances, ce qui laisse supposer qu'une autre voie est possible. Ceci nous a amené à nous poser la question suivante : produit-on le meilleur schéma multi-classes possible en combinant les meilleurs classificateurs binaires ?

A priori rien ne peut garantir le fait que combiner les fonctions de décision binaires les plus performantes permette de produire le meilleur schéma de décodage exploitant celles-ci. Il suffit de considérer le principe du boosting [SCHAPI03] pour se persuader du contraire : on peut produire un schéma de classification performant à partir de plusieurs classificateurs faibles. Nous nous sommes alors proposés de vérifier ce fait expérimentalement afin de produire des contre-exemples au triptyque habituel à savoir 1) Décomposition 2) Optimisation des classificateurs binaires 3) Décodage. Pour cela, nous avons cherché à optimiser globalement le problème multi-classes par décomposition. Cela revient à effectuer une sélection de modèle globale car l'optimisation d'un classificateur binaire n'est plus effectuée par rapport à ses seules performances en généralisation mais par rapport aux performances en généralisation de l'ensemble du schéma multi-classes. Une telle optimisation globale est un problème difficile car l'espace des modèles potentiels à tester est très grand. Nous avons donc choisi une méta-heuristique à base d'algorithme évolutionnaire pour effectuer cette sélection globale de modèle. Nous ne détaillons pas ici sa mise en oeuvre [LEBRUN06, 53]. Nous avons fixé le schéma de décodage comme étant un décodage direct par la méthode de Price pour une décomposition *un-contre-un*. Le tableau 3.11 présente les résultats que nous avons obtenus sur plusieurs bases avec deux schémas de décomposition différents.

Base	$n_c$	un-contre-tous		un-contre-un	
		$\bar{e}$	$\Delta\bar{e}$	$\bar{e}$	$\Delta\bar{e}$
<b>Satimage</b>	6	11,6±1,0%	+0,1%	11,8 ± 1,0%	-0,0%
<b>USPS</b>	10	8,5±1,6%	-0,3%	8,4±1,6%	-0,5%
<b>Letter</b>	26	19,7±1,8%	-2,4%	18,6±2,1%	-2,8%

TAB. 3.11 – Taux d'erreur moyen  $\bar{e}$  et écart type  $\Delta\bar{e}$  des valeurs estimées de  $e$  par validation croisée avec une sélection de modèles par algorithme évolutionnaire.

La colonne  $\Delta\bar{e}$  correspond à l'écart entre la sélection de modèle global par algorithme évolutionnaire et une sélection de modèle classique (une indépendante par fonction de décision binaire). Les classificateurs binaires sont ici des SVM. Les résultats illustrent bien qu'une sélection multi-modèle est plus performante qu'une sélection classique et cet effet s'accroît lorsque le nombre de classes augmente. Ces résultats remettent totalement en question le triptyque habituel de la conception d'un schéma multi-classes de combinaison de classificateurs binaires.

Ceci nous amène à replacer cette problématique comme un cas particulier du problème général de sélection/combinaison de classificateurs [KUNCHE04] où la diversité de l'ensemble des classificateurs joue un grand rôle.

## 3.6 Conclusions et perspectives

Dans ce chapitre, nous avons proposé différents modèles de classification de données d'images par apprentissage. Nous avons tout d'abord proposé pour des images de microscopie, un schéma de combinaison de classification de pixels optimisés indépendamment. Nous avons ensuite montré qu'il est possible, contrairement à ce qui est souvent présenté dans la littérature, de concevoir des schémas efficaces et performants à base de SVM pour la classification de masses de données issues d'images. Enfin, nous avons présenté nos contributions à la conception de schémas de décodage dans le cadre d'un schéma multi-classes de combinaison de classificateurs binaires. Nous avons proposé plusieurs schémas efficaces et enfin nous avons montré expérimentalement qu'une sélection de modèle globale est plus efficace qu'un ensemble de sélection de modèles indépendantes. Afin de permettre une plus large diffusion de nos travaux, nous avons mis à disposition sur Internet<sup>2</sup>, deux bases d'images microscopiques avec les vérités terrain associées. Ceci permet aux autres chercheurs de comparer leurs résultats avec les nôtres [DUMONT07, MARÉ07].

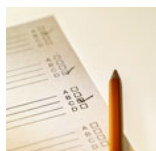
Nos perspectives sont multiples. Il s'agira tout d'abord d'expliquer théoriquement pourquoi, avec une approche par décomposition, une sélection de modèle globale est plus efficace qu'une sélection de modèle classique. Ensuite, nous comptons mener une comparaison exhaustive des différents schémas de décomposition et de décodage et ceci avec différents classificateurs. Ceci n'a jusqu'à présent jamais été mené. Enfin, nous considérerons des algorithmes de sélection/combinaison de classificateurs pour le cas d'une approche par décomposition. Nous comptons également mettre à disposition notre méthode de sélection de modèle à base de recherche tabou pour construire des fonctions de décisions performantes et de complexités réduites et montrer ses performances pour d'autres problèmes de classification de données issues d'images. En particulier, nous chercherons à concevoir par apprentissage des schémas permettant d'effectuer des opérations de traitement d'image classique (filtrage, détection de contour, fusion de régions).

---

<sup>2</sup>Voir <http://users.info.unicaen.fr/~lezoray/Databases.php>



## Conclusion générale et perspectives



L'esprit qui invente est toujours mécontent de ses progrès, parce qu'il voit au-delà.

JEAN LE ROND D'ALEMBERT

### Sommaire

4.1 Synthèse des contributions . . . . .	97
4.2 Perspectives et Projets . . . . .	99

**C**E document a résumé mon activité professionnelle depuis la soutenance de ma thèse. Ce manuscrit m'a permis de faire un bilan complet de mes recherches et je présente dans cette conclusion une synthèse des contributions ainsi que les perspectives de nos travaux.

### 4.1 Synthèse des contributions

Les contributions que nous avons proposées et décrites dans ce mémoire sont les suivantes.

**Chapitre 2.** Ce chapitre a présenté plusieurs voies que nous avons explorées pour un traitement d'images à partir de modèles discrets sur graphes.

**Section 2.2.** Dans cette section, nous avons étudié la problématique des ordres de données vectorielles pour réaliser des opérations de traitement d'images. Dans un cadre discret, nous avons mis en avant que définir une relation d'ordre pour des données multi-variées revient à définir un chemin Hamiltonien sur un graphe complet associé aux données considérées. A partir de cette observation, nous avons proposé deux approches. La première construit un ordre par décimation et exploration d'une hiérarchie d'arbres couvrants minimaux à partir d'un graphe complet. La seconde réalise une réduction de dimension non linéaire à l'aide d'une décomposition en valeurs propres de marches aléatoires. D'une manière générale, ces deux approches peuvent s'appliquer à des données de dimensions quelconques auxquelles un graphe connexe de voisinage est associé.

**Section 2.3.** Dans cette section, nous avons proposé une formulation discrète de la régularisation qui repose sur des graphes pondérés de topologies arbitraires qui représentent des données de dimensions quelconques. Nous avons pu montrer les liens de cette approche avec différentes approches de la littérature. Notre approche dispose d'un certain nombre d'attraits par



rapport aux approches classiques. Premièrement, nous avons une formulation exprimée directement en discret, ce qui permet de considérer des graphes de topologies arbitraires. Comme la topologie peut être arbitraire, nous pouvons naturellement effectuer une régularisation locale, semi-locale ou non locale en modifiant simplement la connexité du graphe. En conséquence, nous pouvons alors effectuer une régularisation discrète non locale sur des variétés, ce qu'aucune autre méthode ne permet d'effectuer actuellement. Deuxièmement, la résolution du problème de régularisation est simple à mettre en oeuvre. De plus, avec une formulation locale, la résolution est rapide, ce qui nous a permis d'appliquer notre formalisme à différentes problématiques (restauration, filtrage, retouche, coloration).

**Section 2.4.** Dans cette section, nous avons proposé plusieurs méthodes permettant de générer des hiérarchies de partitions. La première méthode est basée sur une formulation d'un critère connectif nommé les zones homogènes qui, lorsqu'il est appliqué sur graphes, permet d'obtenir une hiérarchie de partitions de précision intermédiaire entre les zones quasi plates et l'algorithme des cascades. Ce critère connectif reposant sur un seuil d'homogénéité basé sur des distances, il peut s'appliquer à des images dont les pixels sont décrits par des vecteurs de dimensions quelconques. La seconde méthode est basée sur une diffusion sur graphe et elle permet naturellement via la diffusion de fusionner les régions similaires au cours des itérations de l'algorithme. Ceci permet d'une part de disposer d'une alternative rapide et simple à la simplification classique d'images et d'autre part de ne pas déplacer les frontières des régions à travers les échelles contrairement à une approche espace échelle. Ceci permet donc de combiner les bénéfices des approches hiérarchiques basées régions et des approches hiérarchiques basées sur un principe espace échelle. La troisième méthode est spécifiquement dédiée au cas des images couleur. Elle repose sur une coalescence morphologique non supervisée des histogrammes bivariés avec détermination du niveau d'échelle appropriée. L'intersection des cartes de segmentation obtenues est ensuite simplifiée à l'aide d'une fonction d'énergie exprimée sur le graphe d'adjacence des régions qui permet de déterminer le « meilleur » niveau d'échelle suivant un compromis entre complexité de la segmentation et attache aux données initiales.

**Chapitre 3.** Dans ce chapitre nous avons présenté nos contributions à la classification de données issues d'images (des pixels, des régions ou des images) par apprentissage.

**Section 3.3.** Dans cette section, nous nous sommes intéressés à la fusion de décisions de classificateurs pour la classification de pixels d'images de microscopie. Nous avons pu mettre en avant que la fusion permet de résoudre en partie les zones de non-cohérences entre plusieurs classificateurs. De plus, intégré dans un schéma général de segmentation d'images microscopiques, cela permet d'améliorer les résultats. Une mesure de qualité de la segmentation des images microscopiques nous a servi pour établir des confiances en les classifications réalisées.

**Section 3.4.** Dans cette section, nous nous sommes intéressés à la construction de fonctions de décisions parcimonieuses à savoir de complexités réduites et ayant de bonnes performances en généralisation. Nous avons plus particulièrement étudié le cas des classificateurs de type SVM, pour lesquels la complexité des fonctions de décision augmente avec le nombre de vecteurs de support. Nous avons défini un critère de qualité qui correspond à un compromis entre

capacités de généralisation et complexité. Nous avons montré ensuite qu'il est possible d'optimiser ce critère à partir d'une recherche avec tabous qui réalise une sélection multi-modèle comprenant la sélection des exemples pertinents, la sélection des attributs pertinents et la sélection des hyper-paramètres optimaux. L'ensemble de cette approche permet de définir un système d'apprentissage dont les processus décisionnels sont performants et de complexités réduites.

**Section 3.5.** Dans cette section, nous avons présenté nos contributions à la conception de schémas de décodage dans le cadre d'un schéma multi-classes de combinaison de classificateurs binaires. Des décodages direct, en cascade ou hybrides sont présentés. Ils tentent de prendre en compte, à différents niveaux, la notion de classificateur non pertinent. Finalement, nous montrons que les méthodes de sélection de modèles communément utilisées pour l'optimisation des performances de schémas de décomposition ne sont pas optimales. En effet, une sélection de modèle globale est plus efficace qu'un ensemble de sélections de modèles indépendantes et cet effet s'accroît lorsque le nombre de classes à discriminer augmente.

## 4.2 Perspectives et Projets

L'ensemble des activités de recherche que j'ai présentées suscite un certain nombre de perspectives intéressantes. Nous avons déjà indiqué quelques pistes sur la manière dont ses perspectives peuvent être étudiées tant d'un point de vue théorique que pratique. Dans la suite, nous mettons en valeur quelques unes de ces pistes. La figure 4.1 présente un schéma des notions intervenant dans ces perspectives.

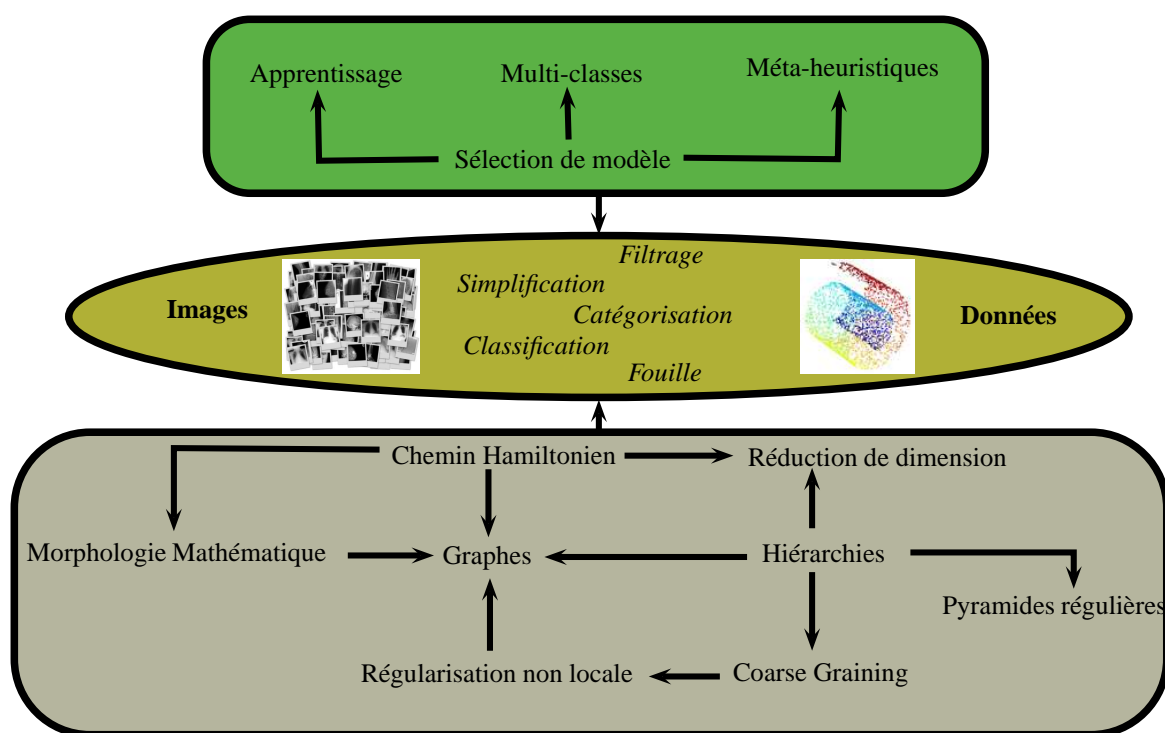


FIG. 4.1 – Schéma des notions intervenant dans les perspectives.

**Traitement d'images et de données multi-variées** Les méthodes d'ordres de données vectorielles que nous avons proposées avaient pour but initial d'étendre facilement les opérateurs de morphologie mathématique au traitement de données multi-variées. Les méthodes que nous avons proposées devront être éprouvées pour des opérateurs morphologiques plus complexes dans le cadre de schémas de traitement morphologiques complets, ceci afin de montrer la validité des approches proposées. Les approches d'ordres de données multi-variées que nous avons proposées vont cependant bien au delà. Puisqu'elles sont exprimées directement pour des graphes de topologies arbitraires dont les noeuds sont décrits par des données multi-variées, elles peuvent également servir comme outil d'analyse de données par morphologie mathématique. Evidemment la morphologie mathématique a déjà été utilisée pour analyser des données, mais ceci principalement sous forme de nuages de points [POSTAI93A]. L'analyse de données par morphologie mathématique nous semble alors une voie prometteuse. Nous reviendrons sur ce point dans les perspectives suivantes. Un autre point concerne l'approche par construction d'un chemin Hamiltonien dans un graphe, nous comptons continuer à étudier ce type d'approche car elle définit en fait une réduction de dimension explicite. Nous nous intéresserons à définir des réductions de dimension qui peuvent être multi-échelle par l'exploitation des hiérarchies de partitions définies lors de la construction du chemin Hamiltonien [HAXHIM07].

**Régularisation non locale de données discrètes** La méthode de régularisation discrète non locale sur graphes pondérés que nous avons proposée offre de nombreuses possibilités pour le traitement non local de variétés. Peu de travaux [HEIN06] ont exploré cette voie jusqu'à présent et nous nous proposons de le faire. Cela est évidemment intéressant pour exploiter les interactions non locales qui peuvent exister dans les images, mais cela présente également un intérêt en fouille de données. En effet, si l'on considère une base de données à laquelle nous associons un graphe de topologie quelconque dont les noeuds sont décrits par des attributs de dimensions quelconques, nous pouvons alors naturellement considérer une régularisation non locale directement pour des données quelconques. Nous avons déjà commencé à explorer cette voie et les résultats sont très prometteurs [?, ?, ?, 73]. Si l'on compare analyse d'images et analyse de données, en analyse d'images on utilise quasi systématiquement une simplification pour faciliter l'analyse. C'est cependant peu le cas en analyse de données et nous pensons qu'une simplification des données permettra d'améliorer une future catégorisation. Il nous faudra dans ce cadre étudier l'influence de la topologie des graphes selon le problème considéré. Les approches hiérarchiques par coarse-graining [LAFON06B] nous semblent une voie qu'il faudra également explorer afin d'avoir une vue hiérarchique des traitements pour éventuellement les accélérer.

**Morphologie mathématique sur graphes** Dans le cadre de la définition d'ordres de données multi-variées, nous avons évoqué l'intérêt que peut avoir la morphologie mathématique pour la fouille de données définies sur graphes. Même disposant d'un ordre de données vectorielles, cela nécessitera de reformuler l'ensemble des opérations morphologiques directement sur des graphes de topologies arbitraires (des travaux récents vont dans ce sens [MEYER07]). Ceci rejoint nos travaux concernant une régularisation exprimée directement en discret sur graphes. Les opérateurs de la morphologie mathématique continue pourront alors être exprimés directement en discret sur graphe, ce qui permettra également de disposer d'opérateurs non locaux.

**Traitement d'images basé régions** Nous avons pu montrer qu'une simplification d'images basée sur le traitement d'un graphe d'adjacence de régions associé à une partition fine

présentait de nombreux intérêts. En effet, à toute image, on peut faire correspondre une partition fine strictement équivalente mais de complexité moindre. On retrouve ce principe dans celui des super-pixels pour des régions [REN03] ou celui des épitomes pour des patches [CHEUNG05], l'intérêt étant de construire une représentation similaire à l'image mais de complexité réduite. Nous nous proposons de continuer dans cette voie en considérant que nombre d'opérations de traitement d'images peuvent être effectuées directement sur des régions et non des pixels.

**Approches top-down pour le traitement d'images** Avec des appareils numériques de plus en plus grandes résolutions, les images dont nous disposons actuellement sont difficilement traitables par les algorithmes actuels. Nous pensons nous orienter vers des approches travaillant non plus sur l'image à une résolution donnée, mais sur une pyramide régulière construite à partir de l'image. Ceci revient à définir des traitements selon une approche top-down (ou multi-grid), les traitements effectués à une échelle servant de repère pour effectuer de nouveaux traitements à une échelle plus grande [HE06].

**Systèmes d'apprentissage pour le traitement d'images** Nous avons pu mettre en évidence qu'une sélection de modèle efficace à l'aide de méta-heuristiques permet de concevoir des fonctions de décisions parcimonieuses et présentant de bonnes performances en généralisation. Une perspective de ce travail consistera à montrer que ce type d'approche est tout à fait envisageable pour d'autres problématiques du traitement d'images. En effet, même si les systèmes à base d'apprentissage présentent de plus en plus d'intérêt pour des tâches de traitement d'images (notamment en filtrage [ELAD06]), cela est toujours envisagé dans le cadre d'une approche en cascade et non selon une optimisation globale. Cependant, dès que l'on considère des systèmes à base d'apprentissage pour effectuer toute ou une partie d'une tâche de traitement d'images, c'est généralement pour introduire de l'adaptation. L'introduction d'adaptation est souvent synonyme d'une augmentation de la complexité qu'il faut absolument contrôler dès lors que l'on s'attache à traiter des masses de données. C'est ce point de vue que nous prenons et nous prévoyons de concevoir des systèmes d'apprentissage parcimonieux et efficaces pour effectuer des tâches bas-niveau de traitement d'images telles que le filtrage, la détection de contours, la fusion de régions, etc. Ceci pourra être en lien avec un traitement d'images basé régions.

**Schémas multi-classes** Nos perspectives porteront tout d'abord sur une comparaison exhaustive des différents schémas de décomposition et de décodage et ceci avec différents classificateurs. Ceci n'a jusqu'à présent jamais été mené. Ensuite, il s'agira d'expliquer théoriquement pourquoi, avec une approche par décomposition, une sélection de modèle globale est plus efficace qu'une sélection de modèle classique. Enfin, nous considérerons des algorithmes de sélection/combinaison de classificateurs pour le cas d'une approche par décomposition. Dans ce cadre, la diversité des classificateurs considérés sera d'une grande importance et les algorithmes génétiques semblent de bons candidats [HADJIT07].

**Applications en microscopie - ANR MDCA FOGRIMMI** Concernant les applications, je coordonne le projet FOGRIMMI (Fouille de GRandes IMages Microscopiques) de l'action de recherche amont MDCA (Masses de Données Connaissances Ambiantes) de l'ANR. Ce projet se fait autour des partenaires suivants : Université de Caen (VAI), ENSICAEN (GREYC), Centre Anti-Cancéreux François Baclesse (GRECAN), Université de Poitiers (SIC), Université

de Bordeaux 1(LABRI). Ce projet vise à concevoir des méthodes de traitement et d'analyse d'Images de Lames Entières (ILE). Ces images permettent d'éviter les biais introduits par une analyse d'image champ par champ par un microscope classique. Cependant, ces images peuvent atteindre jusqu'à 40 Go et ont une résolution pouvant atteindre  $10^{12}$  pixels. Les images d'ILE sont représentées par un seul fichier contenant l'image à pleine résolution et par une séquence d'images codant différentes résolutions de l'image de base (un exemple est donné par la figure 4.2). Ce type d'images a donc la particularité de faire apparaître différentes structures à différentes résolutions. Ces images mêlent de façon intrinsèque une représentation multi-résolution et multi-échelle. Ce projet de recherche est consacré à la conception et à la validation de modèles permettant de détecter et de coder les différentes structures contenues dans de telles images à leur résolution intrinsèque. Nous nous intéresserons plus précisément, dans ce projet, à la détection et la segmentation de structures en cytologie ; cela rejoint la majeure partie de nos perspectives que ce soit en traitement, analyse d'images ou de données.

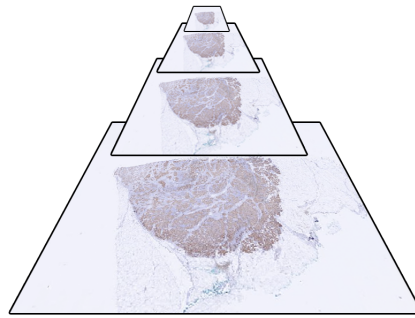


FIG. 4.2 – Pyramide régulière d'une Image de Lame Entière.

## Bibliographie

- [ABE05] S. ABE. « Support Vector Machines for Pattern Classification ». Springer, 2005.
- [ADAMS94] R. ADAMS & L. BISCHOF. « Seeded Region Growing ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), pages 641–647, 1994.
- [AGARWA06] S. AGARWAL. « Ranking on graph data. ». Dans *ICML*, pages 25–32, 2006.
- [ALLWEI00] E. L. ALLWEIN, R. E. SCHAPIRE & Y. SINGER. « Reducing Multiclass to Binary : A Unifying Approach for Margin Classifiers. ». *Journal of Machine Learning Research*, vol. 1, pages 113–141, 2000.
- [ANGULO03] J. ANGULO & J. SERRA. « Morphological coding of color images by vector connected filters ». Dans *Proceedings of ISSPA*, vol. 1, pages 69–72, 2003.
- [ANGULO05A] J. ANGULO. « Morphological color processing based on distances. Application to color denoising and enhancement by centre and contrast operators ». Dans *Proc. of VIIP*, pages 314–319, 2005.
- [ANGULO05B] J. ANGULO. « Unified morphological color processing framework in a lum/sat/hue representation ». Dans *Proc. of ISMM*, pages 387–396, 2005.
- [ANGULO07] J. ANGULO. « Morphological colour operators in totally ordered lattices based on distances : Application to image filtering, enhancement and analysis ». *Computer Vision and Image Understanding*, A paraître, 2007.
- [APTOUL07] E. APTOULA & S. LEFEVRE. « A comparative study on multivariate mathematical morphology ». *Pattern Recognition*, A paraître, 2007.
- [ARBELA04] P. ARBELAEZ & L. COHEN. « Energy Partitions and Image Segmentation ». *Journal of Mathematical Imaging and Vision archive*, vol. 20, pages 43–57, 2004.
- [ASTOLA90] J. ASTOLA, P. HAAVISTO & Y. NEUVO. « Vector median filters ». *Proceedings of the IEEE*, 74(4), pages 678–689, April 1990.
- [ASUNCIO07] A. ASUNCION & D. NEWMAN. « UCI Machine Learning Repository », 2007.
- [AUBERT02] G. AUBERT & P. KORNPROBST. « Mathematical Problems in Image Processing ». Springer-Verlag, 2002.
- [BARNET76] V. BARNETT. « The ordering of multivariate data ». *Journal of the royal society of statistics*, A 139(3), pages 318–355, 1976.
- [BATTIA04] S. BATTIATO, G. GALLO, G. IMPOCO & F. STANCO. « An efficient Re-indexing algorithm for color-mapped images. ». *IEEE Transactions on Image Processing*, 13(11), pages 1419–1423, 2004.

- [BELKIN03] M. BELKIN & P. NIYOGI. « Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. ». *Neural Computation*, 15(6), pages 1373–1396, 2003.
- [BELKIN05] M. BELKIN, P. NIYOGI & V. SINDHWANI. « On Manifold Regularization ». Dans *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, pages 17–24, 2005.
- [BENSOU05] A. BENSOUSSAN & J.-L. MENALDI. « Difference Equations on Weighted Graphs ». *Journal of Convex Analysis*, vol. 12, pages 13–44, 2005.
- [BEUCHE94] S. BEUCHER. « Watershed, hierarchical segmentation and waterfall algorithm ». Dans *Proceedings of ISMM'94*, pages 69–76, 1994.
- [BEYGEL05] A. BEYGELZIMER, J. LANGFORD & B. ZADROZNY. « Weighted One-Against-All ». Dans *Twentieth National Conference on Artificial Intelligence*, pages 720–725, 2005.
- [BORSOT98] M. BORSOTTI, P. CAMPADELLI & R. SCHETTINI. « Quantitative evaluation of color image segmentation results ». *Pattern recognition letters*, vol. 19, pages 741–747, 1998.
- [BOUGLEON] S. BOUGLEUX. « Reconstruction et analyse de formes 3D ». Thèse de doctorat, Université de Caen, en cours de préparation.
- [BOYKOV01] Y. BOYKOV, O. VEKSLER & R. ZABIH. « Fast Approximate Energy Minimization via Graph Cuts ». *IEEE transactions on Pattern Analysis and Machine Intelligence*, 23(11), pages 1222–1239, Novembre 2001.
- [BRONSH97] I. BRONSHTEIN & K. SEMENDYAYEV. « Handbook of Mathematics ». New York : Springer-Verlag, 3rd ed. édition, 1997.
- [BROOK03] A. BROOK, R. KIMMEL & N. SOCHEN. « Variational restoration and edge detection for color images ». *Journal of Mathematical Imaging and Vision*, vol. 18, pages 247–268, 2003.
- [BRUN03] L. BRUN & W. KROPATSCH. « Contraction Kernels and Combinatorial Maps ». *Pattern Recognition Letters*, 24(8), pages 1051–1057, 2003.
- [BUADES05A] A. BUADES, B. COLL & J. MOREL. « A non local algorithm for image denoising ». Dans *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 2, pages 60–65, 2005.
- [BUADES05B] A. BUADES, B. COLL & J. MOREL. « A review of image denoising algorithms, with a new one ». *Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, 4(2), pages 490–530, 2005.
- [BURGET07] B. BURGETH, A. BRUHN, S. DIDAS, J. WEICKERT & M. WELK. « Morphology for matrix data : Ordering versus PDE-based approach. ». *Image and Vision Computing*, 25(4), pages 496–511, 2007.
- [BUSIN04] L. BUSIN, N. VANDENBROUCKE, L. MACAIRE & J.-G. POSTAIRE. « Color space selection for unsupervised color image segmentation by histogram multithresholding ». Dans *Proceedings of the IEEE International Conference on Image Processing (ICIP'04)*, pages 203–206, 2004.
- [CAWLEY01] G. C. CAWLEY. « Model Selection for Support Vector Machines via Adaptive Step-Size Tabu Search ». Dans *ICANNGA*, 2001.

- [CELENK90] M. CELENK. « A color clustering technique for image segmentation ». *Computer Vision Graphics and Image Processing*, vol. 52, pages 145–170, 1990.
- [CHAMBO97] A. CHAMBOLLE & P.-L. LIONS. « Image recovery via total variation minimization and related problems ». *Numerische Mathematik*, 76(2), pages 167–188, 1997.
- [CHAN01A] T. CHAN, S. KANG & J. SHEN. « Total Variation Denoising and Enhancement of Color Images Based on the CB and HSV Color Models ». *J. of Visual Communication and Image Representation*, vol. 12, pages 422–435, 2001.
- [CHAN01B] T. CHAN, S. OSHER & J. SHEN. « The digital TV filter and nonlinear denoising ». *IEEE Transactions on Image Processing*, vol. 10, pages 231–241, 2001.
- [CHAN05B] T. CHAN & J. SHEN. « Image Processing and Analysis - Variational, PDE, wavelet, and stochastic methods ». SIAM, 2005.
- [CHANG01] C.-C. CHANG & C.-J. LIN. « LIBSVM : a library for support vector machines ». Software Available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [CHANUS98] J. CHANUSSOT & P. LAMBERT. « Total ordering based on space filling curves for multivalued morphology ». Dans *Proc. of ISMM*, pages 51–58, 1998.
- [CHAPEL06] O. CHAPPELLE, B. SCHÖLKOPF & A. ZIEN, rédacteurs. « Semi-Supervised Learning ». MIT Press, Cambridge, MA, 2006.
- [CHEUNG05] V. CHEUNG, B. J. FREY & N. JOJIC. « Video Epitomes ». Dans *CVPR*, vol. 1, pages 42–49, 2005.
- [COIFMA05] R. COIFMAN, S. LAFON, A. LEE, M. MAGGIONI, B. NADLER, F. WARNER & S. ZUCKER. « Geometric diffusions as a tool for harmonic analysis and structure definition of data : Diffusion maps ». Dans *Proceedings of the National Academy of Sciences*, 102(21), pages 7426–7431, 2005.
- [COIFMA06] R. COIFMAN & S. LAFON. « Diffusion maps ». *Applied and Computational Harmonic Analysis : Special issue on Diffusion Maps and Wavelets*, vol. 21, pages 5–30, 2006.
- [COMANI02] D. COMANICIU & P. MEER. « Mean Shift : A Robust Approach Toward Feature Space Analysis. ». *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5), pages 603–619, 2002.
- [COMER99] M. COMER & E. DELP. « Morphological operations for colour image processing ». *Journal of electronic imaging*, 8(3), pages 279–289, 1999.
- [CORNUÉ02] A. CORNUÉJOLS & L. MICLET. « Apprentissage artificiel ». Eyrolles, 2002.
- [CRAMME01] K. CRAMMER & Y. SINGER. « On the Algorithmic Implementation of Multi-class Kernel-based Vector Machines. ». *Journal of Machine Learning Research*, vol. 2, pages 265–292, 2001.
- [CRISTI00] N. CRISTIANINI & J. SHAWE-TAYLOR. « An Introduction to Support Vector Machines and other kernel-bases learning methods ». Cambridge University Press, 2000.
- [DARCY05] Y. DARCY & Y. GUERMEUR. « Radius-margin Bound on the Leave-one-out Error of Multi-class SVMs ». Rapport technique, INRIA, 2005.



- [DEMYAN90] V. DEMYANOV & V. MALOZEMOV. « Introduction to Minimax ». Dover, 1990.
- [DERIVA07] S. DERIVAUX, C. WEMMERT, S. LEFEVRE & J. KORCZAK. « Paramétrisation de méthodes de segmentation par utilisation de connaissances et approche génétique ». Dans *Atelier Extraction de Connaissance à partir d'Images (ECOI), Journées Francophones Extraction et Gestion des Connaissances (EGC)*, Namur, Belgium, January 2007.
- [DIESTE05] R. DIESTEL. « Graph Theory », vol. 173. Springer-Verlag, 2005.
- [DIETTE95] T. G. DIETTERICH & G. BAKIRI. « Solving Multiclass Learning Problems via Error-Correcting Output Codes. ». *J. Artif. Intell. Res. (JAIR)*, vol. 2, pages 263–286, 1995.
- [DRÉO03] J. DRÉO, A. PÉROWSKI, P. SIARRY & E. TAILLARD. « Métaheuristiques pour l'optimisation difficile ». Eyrolles, 2003.
- [DREYFU04] G. DREYFUS, M. SAMUELIDES, J. MARTINEZ, M. GORDON, F. BADRAN, S. THIRIA & L. HÉRAULT. « Réseaux de neurones. Méthodologies et applications ». Eyrolles, 2004.
- [DUAN05] K.-B. DUAN & S. S. KEERTHI. « Which Is the Best Multiclass SVM Method? An Empirical Study. ». Dans *Multiple Classifier Systems*, pages 278–285, 2005.
- [DUMONT07] M. DUMONT, R. MARÉE, P. GEURTS & L. WEHENKEL. « Random Sub-windows and Multiple Output Decision Trees for Generic Image Annotation ». Dans *Annual Machine Learning Conference of Belgium and The Netherlands*, 2007.
- [ELAD06] M. ELAD & M. AHARON. « Image Denoising Via Learned Dictionaries and Sparse representation. ». Dans *CVPR*, vol. 1, pages 895–900, 2006.
- [FELZEN04] P. FELZENSVALB & D. HUTTENLOCHER. « Efficient graph-based image segmentation ». *International Journal of Computer Vision*, 59(2), pages 167–181, 2004.
- [FRANK04] E. FRANK & S. KRAMER. « Ensembles of nested dichotomies for multi-class problems. ». Dans *ICML*, pages 305–312, 2004.
- [FÜRNK01] J. FÜRNKRANZ. « Round Robin Rule Learning. ». Dans *ICML*, pages 146–153, 2001.
- [FÜRNK03] J. FÜRNKRANZ. « Round robin ensembles. ». *Intell. Data Anal.*, 7(5), pages 385–403, 2003.
- [GARRID98] L. GARRIDO, P. SALEMBIER & D. GARCIA. « Extensive operators in partition lattices or image sequence analysis ». *Signal Processing*, 6(2), pages 157–180, 1998.
- [GÉRA01] T. GÉRAUD, P. STRUB & J. DARBON. « Color Image Segmentation Based on Automatic Morphological Clustering ». Dans *IEEE International Conference on Image Processing*, vol. 3, pages 70–73, 2001.
- [GILBOA07] G. GILBOA & S. OSHER. « Nonlocal Linear Image Regularization and Supervised Segmentation ». *SIAM - Multiscale Modeling and Simulation (MMS)*, vol. to appear, 2007.
- [GLOVER97] F. GLOVER & M. LAGUNA. « Tabu search ». Kluwer Academic Publishers, Boston MA, 1997.

- [GOUTSI95] J. K. GOUTSIAS, H. J. A. M. HEIJMANS & K. SIVAKUMAR. « Morphological Operators for Image Sequences. ». *Computer Vision and Image Understanding*, 62(3), pages 326–346, 1995.
- [GREENB97] A. GREENBAUM. « Iterative Methods for Solving Linear Systems », vol. 17 de *Frontiers in Applied Mathematics*. SIAM, Philadelphia, 1997.
- [GUERME05] Y. GUERMEUR, A. ELISSEFF & D. ZELUS. « A comparative study of multi-class support vector machines in the unifying framework of large margin classifiers : Research Articles ». *Appl. Stoch. Model. Bus. Ind.*, 21(2), pages 199–214, 2005.
- [GUIGUE03A] L. GUIGUES. « Modèles multi-échelles pour la segmentation d’images ». Thèse de doctorat, Université de Cergy-Pontoise, 2003.
- [GUIGUE03B] L. GUIGUES & H. L. MEN. « Scale-sets image analysis ». Dans *Proceedings of the IEEE International Conference on Image Processing (ICIP’03)*, vol. 2, pages 45–48, 2003.
- [GUIGUE03C] L. GUIGUES, H. L. MEN & J. COCQUEREZ. « The hierarchy of the cocoons of a graph and its application to image segmentation. ». *Pattern Recognition Letters*, 24(8), pages 1059–1066, 2003.
- [GUIGUE06] L. GUIGUES, J. P. COCQUEREZ & H. L. MEN. « Scale-Sets Image Analysis. ». *International Journal of Computer Vision*, 68(3), pages 289–317, 2006.
- [HADJIT07] S. T. HADJITODOROV & L. I. KUNCHEVA. « Selecting diversifying heuristics for cluster ensembles ». Dans *International Workshop on Multiple Classifier Systems*, vol. LNCS 4472, pages 200–209, 2007.
- [HAGEM81] L. HAGEMAN & D. YOUNG. « Applied Iterative Methods ». New York : Academic Press, 1981.
- [HANBUR01] A. HANBURY & J. SERRA. « Mathematical morphology in the HLS colour space ». Dans *Proc. of BMVC’2001*, vol. 2, pages 451–460, 2001.
- [HANBUR02] A. HANBURY. « Mathematical Morphology on the Unit Circle : with applications to hues and to oriented textures ». Thèse de doctorat, Centre de Morphologie Mathématique, Ecole des Mines de Paris, 2002.
- [HASTIE97] T. HASTIE & R. TIBSHIRANI. « Classification by Pairwise Coupling. ». Dans *NIPS*, pages 507–513, 1997.
- [HAXHIM07] Y. HAXHIMUSA, W. G. KROPATSCH, Z. PIZLO, A. ION & A. LEHRBAUM. « Approximating TSP Solution by MST based Graph Pyramid ». Dans F. ESCOLANO & M. VENTO, rédacteurs, *Proceedings of the 6th International Workshop on Graph-based Representation for Pattern Recognition*, vol. LNCS 4538, pages 295–306, 2007.
- [HE06] X. HE, R. S. ZEMEL & D. RAY. « Learning and Incorporating Top-Down Cues in Image Segmentation. ». Dans *ECCV*, vol. 1, pages 338–351, 2006.
- [HEIN05] M. HEIN, J.-Y. AUDIBERT & U. VON LUXBURG. « From Graphs to Manifolds - Weak and Strong Pointwise Consistency of Graph Laplacians. ». Dans *COLT*, pages 470–485, 2005.
- [HEIN06] M. HEIN & M. MAIER. « Manifold Denoising ». Dans *Advances in Neural Information Processing Systems*, vol. 20, pages 1–8, 2006.

- [HOROWI76B] S. HOROWITZ & T. PAVLIDIS. « Picture segmentation by a Tree traversal algorithm ». *J. Assoc. Compt. Math.*, 2(23), pages 368–388, 1976.
- [HSU02] C.-W. HSU & C.-J. LIN. « A Comparison of Methods for Multiclass Support Vector Machines ». Dans *IEEE Transactions in Neural Networks*, 13(2), pages 415–425, 2002.
- [HUANG95] Y. HUANG & C. SUEN. « A method of combining multiple experts for the recognition of unconstrained handwritten numerals ». *IEEE transactions on pattern analysis and machine intelligence*, 17(1), pages 90–94, 1995.
- [IWANOW99] M. IWANOWSKI & J. SERRA. « Morphological Interpolation and Color Images ». Dans *Proceedings of the 10th International Conference on Image Analysis and Processing*, pages 50–, 1999.
- [IZHIKE06] E. M. IZHIKEVICH. « Dynamical Systems in Neuroscience : The Geometry of Excitability and Bursting ». MIT Press, 2006.
- [JAIN99] A. K. JAIN, M. N. MURTY & P. J. FLYNN. « Data clustering : a review ». *ACM Computing Surveys*, 31(3), pages 264–323, 1999.
- [JOLION92] J.-M. JOLION & A. MONTANVERT. « The adaptative pyramid, a framework for 2D image analysis ». *Computer Vision Graphics and Image Processing*, 55(3), pages 339–348, 1992.
- [KARAKO95] D. KARAKOS & P. TRAHANIAS. « Combining vector median and vector directional filters : the directional-distance filters ». Dans *Proceedings of the International Conference on Image Processing*, pages 171–174, 1995.
- [KERVRA06] C. KERVRAAN & J. BOULANGER. « Unsupervised Patch-Based Image Regularization and Representation. ». Dans *ECCV*, vol. 4, pages 555–567, 2006.
- [KIJSIR02] B. KIJSIRIKUL, N. USSIVAKUL & S. MEKNAVIN. « Adaptive Directed Acyclic Graphs for Multiclass Classification. ». Dans *PRICAI*, pages 158–168, 2002.
- [KINCAI02] D. KINCAID & W. CHENEY. « Numerical Analysis : Mathematics of Scientific Computing ». Brooks/Cole, Pacific Grove, 3rd edition édition, 2002.
- [KITTLE98] J. KITTLER, M. HATEF, R.-P. DUIN & J. MATAS. « On combining Classifiers ». *IEEE transactions on PAMI*, 20(3), pages 226–239, Mars 1998.
- [KOEPL94] G. KOEPLER, C. LOPEZ & J.-M. MOREL. « A multiscale algorithm for image segmentation by variational method ». *SIAM Journal on Numerical Analysis*, 31(1), pages 282–299, 1994.
- [KOPPEN99] M. KOPPEN, C. NOWACK & G. ROSEL. « Pareto-Morphology for Color Image Processing ». Dans *The 11th Scandinavian Conference on Image Analysis*, vol. 1, pages 195–202, 1999.
- [KORYCI04] D. KORYCINSKI, M. M. CRAWFORD & J. W. BARNES. « Adaptive feature selection for hyperspectral data analysis ». Dans *SPIE*, vol. 5238, pages 213–225, 2004.
- [KREBEL99] U. H.-G. KREBEL. « Pairwise classification and support vector machines ». *Advances in kernel methods : support vector learning*, pages 255–268, 1999.
- [KROPAT95] W. KROPATSCH. « Building irregular pyramids by dual graph contraction ». Dans *IEE Proceedings of Vision, Image and Signal Processing*, vol. 142, pages 366–374, 1995.

- [KULLBA51] S. KULLBACK & R. A. LEIBLER. « On information and sufficiency ». *Annals of Mathematical Statistics*, vol. 22, pages 79–86, 1951.
- [KUNCHE00] L. KUNCHEVA, C. WHITAKER, C. SHIPP & R. DUIN. « Is independence good for combining classifiers ». Dans *ICPR*, vol. 2, pages 168–171, 2000.
- [KUNCHE04] L. I. KUNCHEVA. « Combining Pattern Classifiers : Methods and Algorithms ». Wiley, 2004.
- [KURUGO01] F. KURUGOLLU, B. SANKUR & A. HARMANCI. « Color image segmentation using histogram multithresholding and fusion ». *Image and Vision Computing*, 19(13), pages 915–928, 2001.
- [LAFON06A] S. LAFON, Y. KELLER & R. R. COIFMAN. « Data Fusion and Multicue Data Matching by Diffusion Maps. ». *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(11), pages 1784–1797, 2006.
- [LAFON06B] S. LAFON & A. LEE. « Diffusion Maps and Coarse-Graining : A Unified Framework for Dimensionality Reduction, Graph Partitioning and Data Set Parameterization. ». *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(9), pages 1393–1403, 2006.
- [LAMBER00A] P. LAMBERT & J. CHANUSSOT. « Extending mathematical morphology to color image processing ». Dans *First Int. Conf. on Color in Graphics and Image Processing (CGIP'2000)*, pages 158–163, 2000.
- [LAMBER02] P. LAMBERT. « Etudes méthodologiques du filtrage et de la segmentation d'images multi-composantes ». Thèse de doctorat, Université de Savoie, 2002.
- [LEBRUN06] G. LEBRUN. « Sélection de modèles pour la classification supervisée avec des Séparateurs à Vaste Marge. Application en traitement et analyse d'images. ». Thèse de doctorat, Université de Caen, Novembre 2006. Directeur : H. Cardot ; Rapporteurs : S. Canu, H. Paugam-Moisy ; Examineurs : C. Charrier, O. Lezoray, M. Revenu.
- [LEI05] H. LEI & V. GOVINDARAJU. « Half-Against-Half Multi-class Support Vector Machines. ». Dans *Multiple Classifier Systems*, pages 156–164, 2005.
- [LEVNER07] I. LEVNER & H. ZHANG. « Classification-Driven Watershed Segmentation ». *IEEE Transactions on Image Processing*, 16(5), pages 1437–1445, 2007.
- [LI95] S. LI. « Markov Random Field Modeling in Computer Vision ». Springer-Verlag, 1995.
- [LINDE80] Y. LINDE, A. BUZO & R. M. GRAY. « An algorithm for vector quantizer design ». Dans *IEEE Trans. Comm.*, vol. COM-28, pages 84–95, 1980.
- [LIU94] J. LIU & Y.-H. YANG. « Multiresolution Color Image Segmentation ». *IEEE transactions on Pattern Analysis and Machine Intelligence*, 16(7), pages 689–700, 1994.
- [LOUVER02] G. LOUVERDIS, M. VARDAVOULIA, I. ANDREADIS & P. TSALIDES. « A new approach to morphological color image processing ». *Pattern recognition*, 35(8), pages 1733–1741, 2002.
- [LUKAC05] R. LUKAC, B. SMOLKA, K. MARTIN, K. PLATANIOTIS & A. VENETSANOPOULOS. « Vector Filtering for Color Imaging ». *IEEE Signal Processing Magazine, Special Issue on Color Image Processing*, 22(1), pages 74–86, 2005.

- [LUKAC06A] R. LUKAC & K. PLATANIOTIS. « Color Image Processing : Methods and Applications », vol. 7. CRC Press, 2006.
- [LUKAC06B] R. LUKAC & K. PLATANIOTIS. « A taxonomy of color image filtering and enhancement solutions ». Dans P. HAWKES, rédacteur, *Advances in Imaging and Electron Physics*, vol. 140, pages 187–264. Elsevier, 2006.
- [MACAIR06] L. MACAIRE, N. VANDENBROUCKE & J.-G. POSTAIRE. « Color image segmentation by analysis of subset connectedness and color homogeneity properties ». *Comput. Vis. Image Underst.*, 102(1), pages 105–116, 2006.
- [MAKROG01] S. MAKROGIANNIS, G. ECONOMOU & S. FOTPOULOS. « A Graph Theory Approach for Automatic Segmentation of Color Images ». Dans *International Workshop on Very Low Bit-rate Video*, pages 162–166, 2001.
- [MAKROG05] S. MAKROGIANNIS, G. ECONOMOU, S. FOTPOULOS & G. BOURBAKIS. « Segmentation of color images using multiscale clustering and graph theoretic region synthesis ». *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 35(2), pages 224–238, 2005.
- [MALLAT98] S. MALLAT. « A Wavelet Tour of Signal Processing ». Academic Press, San Diego, 1998.
- [MARÉ03] R. MARÉE, P. GEURTS, G. VISIMBERGA, J. PIATER & L. WEHENKEL. « An empirical comparison of machine learning algorithms for generic image classification ». Dans *23rd SGAI international conference on innovative techniques and applications of artificial intelligence, Research and development in intelligent systems*, pages 169–182, 2003.
- [MARÉ07] R. MARÉE, M. DUMONT, P. GEURTS & L. WEHENKEL. « Random Subwindows and Randomized Trees for Image Retrieval, Classification, and Annotation ». Dans *15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) and 6th European Conference on Computational Biology (ECCB)*, 2007.
- [MARTIN04] D. MARTIN, C. FOWLKES & J. MALIK. « Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues. ». *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5), pages 530–549, 2004.
- [MATAS95A] J. MATAS. « Colour-based Object Recognition. ». Thèse de doctorat, University of Surrey, 1995.
- [MEER90] P. MEER. « Stochastic image pyramids ». *Computer Vision, Graphics and Image Processing*, 45(3), pages 269–294, 1990.
- [MEILA01] M. MEILA & J. SHI. « A random walks view of spectral segmentation ». Dans *In AI and Statistics (AISTATS)*, 2001.
- [MEILA05] M. MEILA. « Comparing Clustering - An Axiomatic View ». *ICML*, 2005.
- [MEURIE05B] C. MEURIE. « Segmentation d’images couleur par classification pixellaire et hiérarchie de partitions. ». Thèse de doctorat, Université de Caen, Octobre 2005. Directeur : A. Elmoataz ; Rapporteurs : C. Fernandez-Maloigne, P. Gouton ; Examineurs : H. Cardot, O. Lezoray, M. Revenu.
- [MEYER98] F. MEYER. « From connected operators to levelings ». Dans *Proceedings of ISMM’98*, pages 191–199, 1998.

- [MEYER07] F. MEYER & R. LERALLUT. « Morphological Operators for Flooding, Leveling and Filtering Images Using Graphs ». Dans SPRINGER, rédacteur, *Graph based Representation in Pattern Recognition*, vol. LNCS 4538, 2007.
- [MITCHE97] T. MITCHELL. « Machine Learning ». The McGraw-Hill Companies, 1997.
- [MOREIR98] M. MOREIRA & E. MAYORAZ. « Improved Pairwise Coupling Classification with Correcting Classifiers. ». Dans *ECML*, pages 160–171, 1998.
- [MORONE02] N. MORONEY, M. FAIRCHILD, R. HUNT, C. LI, M. LUO & T. NEWMAN. « The CIECAM02 Color Appearance Model. ». Dans *Proceedings of the IS&T/SID 10th Color Imaging Conference*, 2002.
- [MUMFOR89] D. MUMFORD & J. SHAH. « Optimal Approximations by piecewise smooth functions and associated variational problems ». *Comm. On Pure and Applied Math.*, 17(4), pages 577–685, 1989.
- [NOCK04] R. NOCK & F. NIELSEN. « Statistical region merging ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11), pages 1452–1458, 2004.
- [NOYEL07] G. NOYEL, J. ANGULO & D. JEULIN. « On distances, paths and connections for hyperspectral image segmentation ». Dans *ISMM*, 2007.
- [ORTIZ00] F. ORTIZ, F. TORRES, S. PUENTE, F. CANDELAS & P. GIL. « Use of the Hue/Saturation/Intensity Color Spaces to the Morphological ». Dans *Proceedings de First International Conference on Color in Graphics and Image Processing*, pages 219–224, 2000.
- [ORTIZ01] F. ORTIZ, F. TORRES, J. ANGULO & S. PUENTE. « Comparative study of vectorial morphological operations in different color spaces ». Dans *Proc. of Intelligent Robots and Computer Vision XX : Algorithms, Techniques, and Active Vision*, vol. 4572, pages 259–268, 2001.
- [OU04] G. OU, Y. L. MURPHEY & L. A. FELDKAMP. « Multiclass Pattern Classification Using Neural Networks. ». Dans *ICPR (4)*, pages 585–588, 2004.
- [PÉREZ-02] F. PÉREZ-CRUZ & A. ARTÉS-RODRÍGUEZ. « Puncturing Multi-class Support Vector Machines. ». Dans *ICANN*, pages 751–756, 2002.
- [PERONA90] P. PERONA & J. MALIK. « Scale-space and edge detection using anisotropic diffusion ». *IEEE Transactions on Pattern Analysis Machine Intelligence.*, vol. 12, pages 629–639, 1990.
- [PETERS97] R. A. PETERS. « Mathematical morphology for angle valued images ». Dans *SPIE International Conference on Electronic Imaging*, 1997.
- [PEYR07] G. PEYRÉ. « Manifold Models for Signals and Images ». preprint 17, cere-made, 2007.
- [PHILIP06] S. PHILIPP-FOLIGUET & L. GUIGUES. « New criteria for evaluating image segmentation results ». Dans *ICASSP*, 2006.
- [PINHO04] A. J. PINHO & A. J. R. NEVES. « A survey on palette reordering methods for improving the compression of color-indexed images. ». *IEEE Transactions on Image Processing*, 13(11), pages 1411–1418, 2004.
- [PITAS91] I. PITAS & P. TSALIDES. « Multivariate Ordering in Color Image Filtering ». *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, pages 247–259, 1991.

- [PLATAN00] K. PLATANIOTIS & A. VENETSANOPOULOS. « Color Image Processing and Applications ». Springer Verlag, 2000.
- [PLATT99A] J. PLATT. « Probabilistic outputs for support vector machines and comparison to regularized likelihood methods ». Dans A. J. SMOLA, P. BARTLETT, B. SCHOELKOPF & D. SCHUURMANS, rédacteurs, *Advances in Large Margin Classifiers*, pages 61–74, 1999.
- [PLATT99B] J. C. PLATT, N. CRISTIANINI & J. SHAWE-TAYLOR. « Large Margin DAGs for Multiclass Classification. ». Dans *NIPS*, pages 547–553, 1999.
- [POGGIO85] T. POGGIO, V. TORRE & C. KOCH. « Computational Vision and Regularization Theory ». *Nature*, vol. 317, pages 314–319, 1985.
- [POSTAI93A] J.-G. POSTAIRE, R. D. ZHANG & C. LECOCQ-BOTTE. « Cluster Analysis by Binary Morphology. ». *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(2), pages 170–180, 1993.
- [POSTAI93B] J. POSTAIRE, R. ZHANG & C. LECOCQ-BOTTE. « Cluster Analysis by Binary Morphology ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2), pages 170–180, 1993.
- [PRICE94] D. PRICE, S. KNERR, L. PERSONNAZ & G. DREYFUS. « Pairwise Neural Network Classifiers with Probabilistic Outputs. ». Dans *NIPS*, pages 1109–1116, 1994.
- [QUINLA93] J. R. QUINLAN. « C4.5 : Programs for Machine Learning ». Morgan Kaufmann, 1993.
- [REGAZZ97] C. REGAZZONI & A. TESCHIONI. « A New Approach to Vector Median Filtering based on Space Filling Curves ». *IEEE Trans. on Image Processing*, vol. 6, pages 1025–1037, 1997.
- [REN03] X. REN & J. MALIK. « Learning a Classification Model for Segmentation. ». Dans *ICCV*, pages 10–17, 2003.
- [RIFKIN04] R. M. RIFKIN & A. KLAUTAU. « In Defense of One-Vs-All Classification. ». *Journal of Machine Learning Research*, vol. 5, pages 101–141, 2004.
- [ROBLES04] A. ROBLES-KELLY & E. R. HANCOCK. « String Edit Distance, Random Walks And Graph Matching. ». *IJPRAI*, 18(3), pages 315–327, 2004.
- [RONSE90] C. RONSE. « Why mathematical mophology needs complete lattices ». *Signal Processing*, 21(2), pages 129–154, 1990.
- [RUDIN92] L. RUDIN, S. OSHER & E. FATEMI. « Nonlinear total variation based noise removal algorithms ». *Physica D*, vol. 60, pages 259–268, 1992.
- [SALEMB97] P. SALEMBIER, L. GARRIDO & D. GARCIA. « Image sequence analysis and Merging algorithm ». Dans *International Workshop on Very Low Bit-rate Video*, pages 1–8, 1997.
- [SALEMB00] P. SALEMBIER & L. GARRIDO. « Binary partition tree as an efficient representation for image processing, segmentation and information retrieval ». *IEEE Transactions on Image Processing*, 9(4), pages 561–576, 2000.
- [SAVICK03] P. SAVICKÝ & J. FÜRNRANZ. « Combining Pairwise Classifiers with Stacking. ». Dans *IDA*, pages 219–229, 2003.

- [SCHAPI03] R. E. SCHAPIRE. « The boosting approach to machine learning : An overview ». Dans D. D. DENISON, M. H. HANSEN, C. HOLMES, B. MALLICK & B. YU, rédacteurs, *Nonlinear Estimation and Classification*. Springer, 2003.
- [SERRA88] J. SERRA. « Image Analysis and Mathematical Morphology : Theoretical Advances », vol. 2. Academic Press, London, 1988.
- [SERRA02] J. SERRA. « Connectivity for sets and functions ». *Fundamenta Informaticae*, vol. 41, pages 147–186, 2002.
- [SERRA03] J. SERRA. « Connexions et segmentations d’images ». *Traitement du signal*, 2003.
- [SHI00] J. SHI & J. MALIK. « Normalized Cuts and Image Segmentation. ». *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8), pages 888–905, 2000.
- [SMITH97] S. M. SMITH & J. M. BRADY. « SUSAN—A New Approach to Low Level Image Processing ». *Int. J. Comput. Vision*, 23(1), pages 45–78, 1997.
- [SOILLE96] P. SOILLE. « Morphological partitioning of multispectral images ». *Journal of Electronic Imaging*, 18(4), pages 252–265, 1996.
- [SOILLE04] P. SOILLE. « Morphological Image Analysis : Principles and Applications ». Springer-Verlag, 2004.
- [SORIA-02] A. SORIA-FRISCH & M. KOPPEN. « The Fuzzy Integral as Similarity Measure for a New Color Morphology ». Dans *Proc. of CGIV*, pages 523–526, 2002.
- [STRING99] E. STRINGA, A. TESCHIONI & C. S. REGAZZONI. « A classical morphological approach to color image filtering based on space filling curves. ». Dans *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP’99)*, pages 351–354, 1999.
- [SUMENG05] B. SUMENGEN & B. S. MANJUNATH. « Multi-scale Edge Detection and Image Segmentation ». Dans *European Signal Processing Conference (EU-SIPCO)*, pages CD–proceedings, 2005.
- [SZLAM06] A. SZLAM, M. MAGGIONI & R. COIFMAN. « A general framework for adaptive regularization based on diffusion processes on graphs ». Rapport technique, Yale, 2006.
- [TAKAHA03] F. TAKAHASHI & S. ABE. « Optimizing directed acyclic graph support vector machines ». Dans *Proc. Artificial Neural Networks*, pages 166–170, 2003.
- [TALBOT98] H. TALBOT, C. EVANS & R. JONES. « Complete Ordering and Multivariate Mathematical ». Dans *Proceedings of the fourth international symposium on Mathematical morphology and its applications to image and signal processing*, pages 27–34, 1998.
- [TAN06] P. N. TAN, M. STEINBACH & V. KUMAR. « Introduction to Data Mining ». Addison-Wesley, 2006.
- [TANG01] B. TANG, G. SAPIRO & V. CASELLES. « Color image enhancement via chromaticity diffusion ». *IEEE Transactions on Image Processing*, vol. 10, pages 701–707, 2001.
- [TAON] V. TA. « Traitement et Organisation de grandes masses de données images : applications en cancérologie ». Thèse de doctorat, Université de Caen, en cours de préparation.



- [TAX02] D. TAX & R. DUIN. « Using two-class classifiers for multiclass classification ». Dans *International Conference on Pattern Recognition (ICPR)*, vol. 2, pages 124–127, 2002.
- [THEOHA05] C. THEOHARATOS, G. ECONOMOU & S. FOTOPOULOS. « Color edge detection using the minimal spanning tree ». *Pattern recognition*, vol. 38, pages 603–606, 2005.
- [TOMASI98] C. TOMASI & R. MANDUCHI. « Bilateral Filtering for Gray and Color Images. ». Dans *ICCV*, pages 839–846, 1998.
- [TRAHAN96] P. TRAHANIAS, D. KARAKOS & A. VENETSANOPOULOS. « Directional processing of color images : Theory and experimental results ». *IEEE Transactions on Image Processing*, vol. 5, pages 868–880, 1996.
- [TRÉ00] A. TRÉMEAU & P. COLANTONI. « Regions adjacency graph applied to color image segmentation ». *IEEE Transactions on Image Processing*, 9(4), pages 735–744, 2000.
- [TRÉM04] A. TRÉMEAU, C. FERNANDEZ-MALOIGNE & P. BONTON. « Images Couleur : de l’acquisition au traitement ». Collection Science Sup. Dunod, 2004.
- [TROYAN00] M. TROYANOV. « Solving the p-Laplacian on manifolds ». Dans *Proc. Amer. Math. Soc.*, vol. 128, pages 541–545, 2000.
- [TSALID02] P. TSALIDES, M. I. VARDAVOULIA & I. ANDREADIS. « Vector Ordering and Morphological Operations for Colour Image Processing ». *Pattern Anal. Appl*, 5(3), pages 271–287, 2002.
- [TSCHUM05] D. TSCHUMPERLÉ & R. DERICHE. « Vector-Valued Image Regularization with PDEs : A Common Framework for Different Applications ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(4), pages 506–517, 2005.
- [TSUJIN04] D. TSUJINISHI, Y. KOSHIBA & S. ABE. « Why Pairwise Is Better than One-against-All or All-at-Once ». Dans *International Joint Conference on Neural Networks*, vol. 1, pages 693–698, 2004.
- [UIT00] UIT. « Méthodologie d’évaluation subjective de la qualité des images de télévision ». UTI-R Recommendation BT.500-10, 2000.
- [VANDEN03] N. VANDENBROUCKE, L. MACAIRE & J.-G. POSTAIRE. « Color image segmentation by pixel classification in an adapted hybrid color space. Application to soccer image analysis. ». *Computer Vision and Image Understanding*, 90(2), pages 190–216, 2003.
- [VANHAM03] I. VANHAMEL, I. PRATIKAKIS & H. SAHLI. « Multiscale Gradient Watersheds of Color Images ». *IEEE Transactions on Image Processing*, 12(6), pages 617–626, 2003.
- [VAPNIK95] V. N. VAPNIK. « The nature of statistical learning theory ». Springer-Verlag, 1995.
- [VAPNIK98] V. VAPNIK. « Statistical Learning Theory ». John Wiley, 1998.
- [VARDAV01] M. I. VARDAVOULIA, I. ANDREADIS & P. TSALIDES. « A new vector median filter for colour image processing ». *Pattern Recognition Letters*, 22(6), pages 1733–1741, 2001.

- [VERTAN96] C. VERTAN, M. MALCIU, T. ZAHARIA & V. BUZULOIU. « A Clustering Approach to Vector Mathematical Morphology ». Dans *Proc. of the IEEE International Conference on Electronics, Circuits and Systems ICECS 96*, vol. 1, pages 187–190, 1996.
- [VINET91] L. VINET. « Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques ». Thèse de doctorat, Université de Paris IX Dauphine, 1991.
- [VONLUX06] U. VON LUXBURG. « A Tutorial on Spectral Clustering ». Rapport technique 149, Max Planck Institute for Biological Cybernetics, 2006. To appear in *Statistics and Computing*.
- [WEICKE99] J. WEICKERT. « Coherence-enhancing diffusion of colour images. ». *Image Vision Comput.*, 17(3-4), pages 201–212, 1999.
- [WOLPER92] D. WOLPERT. « Stacked generalization ». *Neural Networks*, 5(2), pages 241–260, 1992.
- [WU04] T.-F. WU, C.-J. LIN & R. C. WENG. « Probability Estimates for Multi-class Classification by Pairwise Coupling. ». *Journal of Machine Learning Research*, vol. 5, pages 975–1005, 2004.
- [XU05] R. XU & D. I. I. WUNSCH. « Survey of Clustering Algorithms ». *IEEE Transactions on neural networks*, 16(3), pages 645–678, 2005.
- [XUE03] H. XUE, T. GÉRAUD & A. DURET-LUTZ. « Multi-band Segmentation using Morphological Clustering and Fusion Application to Color Image Segmentation ». Dans *Proceedings of the IEEE International Conference on Image Processing (ICIP'03)*, vol. 1, pages 353–356, 2003.
- [YAROSL85] L. P. YAROSLAVSKY. « Digital Picture Processing — an Introduction ». Springer, Berlin., 1985.
- [ZHOU03] D. ZHOU, J. WESTON, A. GRETTON, O. BOUSQUET & B. SCHOLKOPF. « Ranking on Data Manifolds. ». Dans *NIPS*, 2003.
- [ZHOU04] D. ZHOU & B. SCHOLKOPF. « A Regularization Framework for Learning from Graph Data ». Dans *ICML Workshop on Statistical Relational Learning and Its Connections to Other Fields*, pages 132–137, 2004.
- [ZOUARI02] H. ZOUARI, L. HEUTTE, Y. LECOURTIER & A. ALIMI. « Un panorama des méthodes de combinaison de classifieurs en reconnaissance des formes ». Dans *RFIA*, vol. 2, pages 499–508, 2002. Angers, France.
- [ZUCKER76] S. W. ZUCKER. « Region growing : Childhood and adolescence ». *Comput. Graph. Image Process.*, vol. 5, pages 382–399, 1976.



## Curriculum Vitae

---



### **Olivier LÉZORAY**

Né le 12 Septembre 1972 à Lannion (Côtes d'Armor).

Marié, deux enfants

Maître de Conférences à l'IUT Cherbourg Manche,  
Département Services et Réseaux de Communication  
LUSAC EA 2607, Équipe Vision et Analyse d'Images,  
120 Rue de l'exode, F-50000 SAINT-LÔ

Téléphone : 0233775517

Fax : 0233771166

Courriel : [olivier.lezoray@unicaen.fr](mailto:olivier.lezoray@unicaen.fr)

<http://www.info.unicaen.fr/~lezoray>

### **A.1 Diplômes**

**2000 Doctorat de l'Université de Caen Basse Normandie** mention très honorable avec les félicitations du jury. Titre : « *Segmentation d'images par morphologie mathématique et classification de données par réseaux de neurones : Application à la classification de cellules en cytologie des séreuses* ». Directeur : M. Revenu ; Rapporteurs : P. Bonton, A. Faure ; Examineurs : H. Cardot, H. Elie, A. Elmoataz.

**1995 DEA d'Intelligence Artificielle et Algorithmique.** Université de Caen. Mention Assez Bien.

**1994 Maîtrise d'Informatique.** Université de Caen. Mention Assez Bien.

**1993 Licence d'Informatique.** Université de Caen. Mention Assez Bien.

**1992 DEUG A Sciences et Structure de la matière.** Université de Caen.

### **A.2 Carrière**

J'ai pris mes fonctions de Maître de Conférences en Informatique à l'IUT Cherbourg Manche le 1er Septembre 2000. Depuis, je travaille dans l'équipe Vision et Analyse d'Images du Laboratoire Universitaire des Sciences Appliquées de Cherbourg (LUSAC EA 2607). Je fais partie de cette équipe depuis ma thèse.

- **Prime d'Encadrement Doctoral et de Recherche** : obtention en Septembre 2006
- **Prime de Responsabilités Pédagogiques** : depuis Septembre 2000

- **Depuis Septembre 2000** : Maître de Conférences en Informatique à l'IUT Cherbourg Manche, Antenne de Saint-Lô, Département Services et Réseaux de Communication (SRC).
- **2000** : Qualification aux fonctions de Maître de Conférences dans les sections CNU 27 et 61.
- **Septembre 1999 à Août 2000** : ATER en Informatique au département d'informatique de l'Université de Caen Basse Normandie.

## A.3 Responsabilités pédagogiques et administratives

### A.3.1 Enseignements et responsabilités pédagogiques

Depuis que je suis en poste à l'IUT Cherbourg Manche (Antenne de Saint-Lô), Département Services et Réseaux de Communication (SRC), j'y effectue la majorité de mes enseignements. J'ai intégré le département SRC lors de son ouverture, et j'étais le seul enseignant d'informatique affecté sur site. Les premières années ont demandé beaucoup d'investissement afin de mettre en place ce nouveau département. J'ai donc conçu intégralement la majorité des enseignements en informatique du département SRC du moins pour les premières années. Être nommé dans un nouveau département d'IUT délocalisé n'est pas chose facile, mais cela est très enrichissant. Avant que le département SRC n'atteigne une taille d'équipe pédagogique suffisante, j'ai dû assurer un service bien supérieur à celui imposé par mon statut et conserver une activité de recherche dans ces conditions n'est pas simple. De plus, avec une équipe pédagogique restreinte, j'ai dû assurer la tâche administrative lourde de directeur des études. Depuis, j'ai pu enseigner dans d'autres structures que l'IUT Cherbourg Manche et à d'autres niveaux que le DUT. Tous ces éléments sont décrits dans la section suivante.

#### A.3.1.1 Enseignements

- Niveau DEUG :
  - Programmation Orientée Objet (Java)
  - Programmation fonctionnelle (Scheme)
- Niveau DUT :
  - Outils Bureautiques,
  - Systèmes d'exploitation,
  - Algorithmique, Programmation Orientée Objet (Java),
  - Systèmes d'information et bases de données (Php, SQL),
  - Programmation Multimédia (XML, XSLT),
  - Systèmes d'information avancés (UML, Servlets, JSP),
  - Programmation hypermédia (XHTML, CSS, Javascript),
  - Images numériques,
  - Réseaux Internet et services.
- Niveau Licence Professionnelle :
  - Mathématiques pour la synthèse d'images,
  - Traitement d'images couleur.
- Niveau École d'ingénieur (2ème année)
  - Systèmes Intelligents.
- Niveau Master Recherche

- Classification et réseaux de neurones.
- Conventions de Formation CORTECHS
- Systèmes d'information.

Le tableau A.1 récapitule les enseignements que j'ai effectués depuis ma nomination. La figure A.1 donne une vue synthétique de ces enseignements en terme de volume par an, de volume par type de cours et de niveau. On remarque que, pour certaines années, le volume d'enseignements est élevé. Ceci correspond soit à la mise en place de nouveaux programmes pédagogiques (2005-2006), soit aux premières années de mise en place du département SRC. Depuis l'obtention de la PEDR, je peux consacrer beaucoup plus de temps à mes activités de recherche.

Heures Eq. TD/ Matière	2006-2007		2005-2006		2004-2005		2003-2004		2003-2002		2002-2001		2000-2001
	SRC1	SRC2	SRC1	SRC2	SRC1	SRC2	SRC1	SRC2	SRC1	SRC2	SRC1	SRC2	SRC1
DUT													
Outils Bureautiques	19,50		19,50		11,00		16,50		24,00		32,00		18,00
Systèmes d'exploitation	29,25		29,25	6,75	27,75	21,75	39,25	21,75	40,00	24,75	41,00	22,75	57,75
Algorithmique	34,50		35,50		42,75		42,75		54,75		54,75		76,25
Prog. Orientée Objet	9,00		7,50	18,50		29,50		23,50		20,00		50,00	
Prog. Multimédia		28,75		26,25		32,25		41,25		42,75		46,75	
Syst. Info. Avancés		28,75		15,75		15,75							
Prog. Hypermédia				7,50		7,50		18,50		31,25		12,00	25,00
Images numériques		6,00											
Réseaux Internet		11,00		9,00									28,00
Bases de données	6,00		15,00	15,00		15,00	18,00		18,00	24,00	9,00	12,00	50,25
Systèmes d'information		17,50		15,50		15,00		24,00		24,00			
Licence		L1		L1									
Introduction Traitement d'images		4,50		4,50									
Licence Professionnelle				Lpro		Lpro		Lpro					
Traitement d'images couleur				15,00		15,00		15,00					
Mathématiques pour l'image				15,00									
ENSICAEN		2ème année		2ème année									
Systèmes Intelligents		23,33		23,33									
Master ou DEA		M2R ESCI		M2R ESCI		DEA ESCI & IAA		DEA ESCI & IAA					
Réseaux de neurones		15		15		15		15					

TAB. A.1 – Récapitulatif des enseignements effectués depuis 2000.

### A.3.1.2 Encadrements

- Niveau DUT :
  - Environ 4 projets tuteurés par an,
  - Environ 6 stages par an.
- Niveau Ecole d'ingénieurs :
  - 1 Projet de deuxième année par an (2004-)

### A.3.1.3 Responsabilités pédagogiques

Je participe à la vie de l'IUT en m'occupant des tâches pédagogiques et administratives courantes dans un IUT : encadrements et soutenances de projets tuteurés, encadrements et soutenances de stages, examen des dossiers d'admission, participation aux jurys, participation aux réunions pédagogiques, responsabilité de modules, surveillance d'examens, etc. J'ai également supervisé la mise en place des nouveaux programmes pédagogiques nationaux en 2005 et 2006. Les responsabilités pédagogiques que j'ai occupées et que j'occupe sont les suivantes :

- Directeur des études du département Services et Réseaux de Communications de l'IUT Cherbourg Manche (depuis 2000) : Conception des emplois du temps à la semaine (2000-2006), Gestion des Notes (2000, 2005-2007), Gestion des vacataires (2000-2006), Gestion des absences (2000, 2005-2007), Gestion des planning annuels (2000-2006),

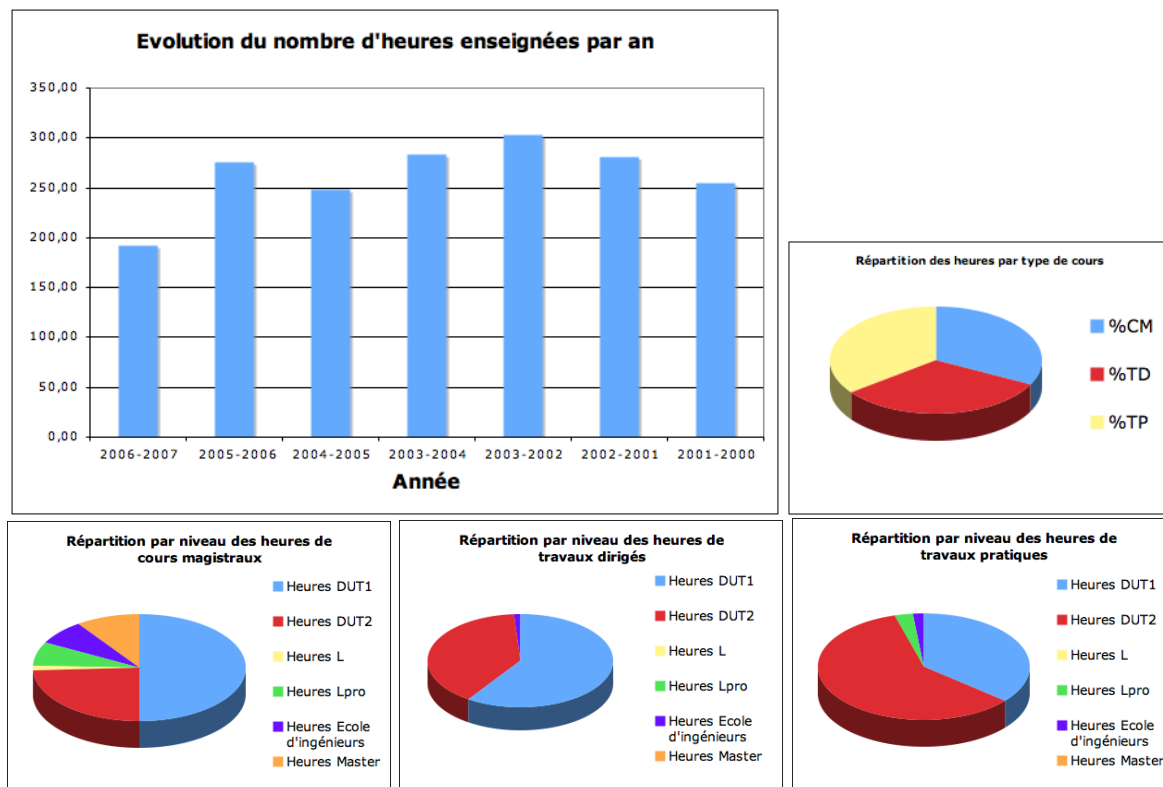


FIG. A.1 – Illustrations de la répartition des enseignements.

- Responsable technique et sécurité du site Internet de l'IUT Cherbourg Manche (2000-2004)
- Administration réseau du département SRC (2000-2002).
- Conception des logiciels pédagogiques (Intranet pédagogique, Extranet pédagogique, gestion des projets, gestion des emplois du temps, gestion des absences, gestion des notes, gestion des planning, gestion des feuilles de service) : depuis 2000. Cette activité de développement logiciel au sein du département est pour moi essentielle. En effet, même si toutes ces gestions sont effectuées dans tous les départements d'IUT, rien ne nous est fourni par le ministère de l'éducation nationale pour les réaliser efficacement. J'ai choisi d'informatiser un maximum ces tâches car elles sont très phagocytantes et cela permet une coordination plus efficace des membres de l'équipe pédagogique tout en simplifiant les tâches pédagogiques.

### A.3.2 Responsabilités administratives

- Président de Jury de Bac STT ACA, 2001 à Saint-Lô,
- Trésorier de l'association SRCProd de promotion des projets tuteurés du département SRC de l'IUT Cherbourg Manche,
- Président de commission de recrutement de PRAG à l'IUT Cherbourg-Manche (2006),
- Président de commission de recrutement d'ATER à l'IUT Cherbourg-Manche (2001-2005),
- Membre titulaire du conseil du LUSAC (2002-2005),
- Membre titulaire et assesseur de la commission de spécialistes en Informatique de l'Université de Caen Basse Normandie (2002-2006),

- Vice-Président de la commission de spécialistes en Informatique de l'Université de Caen Basse Normandie (2007-),
- Membre du jury du Master Electronique, Systèmes, Capteurs, Images (depuis 2003),
- Membre des jurys de première et deuxième année de DUT SRC,
- Membre des grands-jury de l'IUT Cherbourg-Manche (2005-2006),

## A.4 Activités liées à la recherche

Etant situé sur un site délocalisé (Saint-Lô est à 100 Kms de Cherbourg et à 60 Kms de Caen), aucune réelle infrastructure de recherche n'y était présente. Avec les responsabilités pédagogiques et les nombreux enseignements à effectuer, une présence quasi quotidienne est nécessaire sur le site de Saint-Lô. Nous nous sommes donc attachés, avec mon collègue Christophe Charrier, à développer une infrastructure dans les locaux même de l'IUT Cherbourg Manche sur l'antenne de Saint-Lô. Depuis 2004, nous avons réussi à obtenir une salle dédiée à nos activités et à rapatrier nos matériels depuis Cherbourg. Ceci facilite grandement nos activités notamment pour l'accueil de stagiaires, de contractuels et de doctorants.

### A.4.1 Thèmes de recherche

- ⇒ Modèles discrets pour le traitement d'images,
  - Graphes, Morphologie Mathématique, Couleur, régularisation discrète, hiérarchies de partitions,
  - Applications pour le filtrage, la simplification et la segmentation.
- ⇒ Classification de données d'images par apprentissage,
  - Fusion de classificateurs, Apprentissage supervisé/semi-supervisé, Complexité des fonctions de décision,
  - Applications en fouille de données pour la classification, l'indexation et la catégorisation.

### A.4.2 Activités de recherche

Mon activité de recherche se fait, depuis mon doctorat, au sein du LUSAC EA 2607 (Laboratoire Universitaire des Sciences Appliquées de Cherbourg), équipe Vision et Analyse d'Images (VAI) et mes axes de recherche portent sur :

1. Les modèles et algorithmes pour le traitement et l'analyse d'images,
2. La classification, l'apprentissage et le traitement de l'information.

Mon premier axe de recherche concerne les modèles et algorithmes pour le traitement et l'analyse d'images dont la finalité principale est l'analyse du contenu des images à travers différentes échelles. L'objectif est de segmenter les principaux composants visuels d'une image pour en extraire les caractéristiques par le biais de descripteurs ou modèles compacts pour une meilleure compréhension, représentation et interprétation du contenu. Je m'intéresse plus particulièrement aux approches variationnelles et morphologiques ainsi qu'aux méthodes multi-échelles pour le filtrage, la segmentation et l'indexation d'images couleur. Les modèles mathématiques essentiellement étudiés sont formalisés dans le cadre d'une représentation par graphe



afin de faciliter et d'améliorer le traitement et l'analyse. Plusieurs de nos travaux sont innovants : traitement d'images multi-variées par réduction de dimension sur graphes, régularisation discrète non locale sur graphes pondérés de topologies arbitraires, nouvelles méthodes morphologiques, pyramides irrégulières combinant approches espace-échelle et approches régions.

Mon second axe de recherche concerne la classification, l'apprentissage et le traitement de l'information. L'objectif est de définir des systèmes de décision performants et de complexités réduites. Les études théoriques et pratiques portent essentiellement sur les réseaux de neurones et les Séparateurs à Vaste Marge (SVM) comme classificateurs. Les bases de données des systèmes informatiques actuels sont de plus en plus volumineuses et par conséquent leurs redondances intrinsèques augmentent avec leurs tailles. Je m'intéresse donc à la définition de nouveaux algorithmes d'apprentissage permettant la réduction de cette redondance pour optimiser les temps d'apprentissage et de décision des classificateurs utilisés, ainsi que leurs taux de reconnaissance. Je m'attache également à la création de systèmes de décision multi-classes optimaux à partir de la combinaison d'un ensemble de classificateurs binaires. Les points forts de mon activité de recherche portent d'une part sur l'élaboration de modèles adaptés à l'apprentissage et à la classification de données quelconques tout en s'affranchissant d'un maximum de paramètres et d'autre part sur les méthodes de combinaison de classificateurs binaires pour lesquels j'ai proposé plusieurs modèles très performants.

Pour mes deux axes de recherche, la majorité des applications portent sur la microscopie médicale et le multimédia.

### A.4.3 Co-encadrement de thèses et de DEA

J'ai participé aux encadrements de plusieurs doctorants ainsi que de stages de master recherche (voir-ci-dessous). Ces thèses ont été et sont encadrées sur plusieurs sites universitaires (Saint-Lô, Cherbourg, Caen ainsi que Tours pour la thèse de G. Lebrun), ce qui nécessite de ma part une très grande mobilité mais également pour certains étudiants dont la thèse se déroule à Cherbourg. J'ai également encadré deux étudiants durant leur stage de DEA. Travailler sur un site délocalisé ne facilite pas l'attraction d'étudiants motivés de Master Recherche. Avec mon intégration au sein du laboratoire GREYC UMR CNRS 6072, cette situation devrait s'améliorer car je pourrais encadrer ces stages sur Caen.

#### A.4.3.1 Thèses soutenues

- Thèse en Informatique de Cyril Meurie (Octobre 2001 à Octobre 2005, financement de type MNESR). Titre : « *Segmentation d'images couleur par classification pixellaire et hiérarchie de partitions.* ». Directeur : A. Elmoataz (50%), Co-encadrant : O. Lezoray (50%), Rapporteurs : C. Fernandez-Maloigne, P. Gouton, Examineurs : M. Revenu, H. Cardot.
- Thèse en Informatique de Gilles Lebrun (Octobre 2002 à Novembre 2006, financement association Coeur-Cancer et conseil régional de Basse Normandie) . Titre : « *Sélection de modèles pour la classification supervisée avec des Séparateurs à Vaste Marge. Application en traitement et analyse d'images.* ». Directeur : Hubert Cardot (20%), Co-encadrants : C.

Charrier (40%), O. Lezoray (40%), Rapporteurs : S. Canu, H. Paugam-Moisy, Examineurs : M. Revenu.

#### A.4.3.2 Thèses en cours

- Thèse en Informatique de Vinh Tong Ta (Septembre 2006 - ). Titre : « *Traitement et Organisation de grandes masses de données images : applications en cancérologie.* ». Directeur : A. Elmoataz (50%), Co-encadrant : O. Lezoray (50%). Financement : Bourse Conseil régional et association Coeur-Cancer de Cherbourg.

#### A.4.3.3 DEA soutenus

- Vinh Thong Ta. Titre : « *Représentation d'images par arbres binaires de régions.* », DEA Intelligence Artificielle et Algorithmique. Encadrement : O. Lezoray et A. Elmoataz, Université de Caen, 2004.
- T. Reuste. Titre : « *Indexation d'images et recherche par points d'intérêt.* », DEA Electronique, Systèmes, Capteurs, Images. Encadrement : O. Lezoray, R. Clouard, Université de Caen, 2004.

## A.5 Animation de la recherche

### A.5.1 Participation à des réseaux

- Membre du GDR ISIS, du GDR Stic-Santé,
- Membre de sociétés savantes internationales : IAPR, EURASIP, IS&T,
- Membre de sociétés savantes nationales : AFIA, AFRIF, GFINC.

### A.5.2 Activités éditoriales

- Editeur invité de la revue EURASIP Journal on Advances in Signal Processing : organisateur d'un numéro spécial à paraître en 2008 sur le thème "Machine Learning in Image Processing".

### A.5.3 Membre de comités de lecture, de programme

- Revues internationales : IEEE transactions on Image Processing, IEEE transactions on Medical Imaging, IEEE transactions on Information Technology in Biomedicine, IEEE transactions on Multimedia, International Journal of Neural Systems, Integrated Computer-Aided Systems, Journal of Real-Time Imaging, International Journal of Artificial Intelligence In Medicine, International Journal of Robotics and Automation, Image and Vision Computing, Computer Vision and Image Understanding, Pattern Recognition Letters.

- Revues nationales : Traitement du signal
- Conférences internationales : EUSIPCO (2007), KES (2007), IAPR Workshop on Graph-Based Representations (2005, 2007), International Conference on Human System Learning (2005-2008), International Conference on Visualization Imaging and Image Processing (2005), International Conference on Image Processing (2006).
- Conférences nationales : COMpression et REprésentation des Signaux Audiovisuels (2005, 2006)

#### **A.5.4 Organisation de Journées, de Conférences**

- Chairman d'une session spéciale "Image Mass Data Analysis" lors de la conférence ICISP 2008,
- Program Committee Chairman de la conférence ICISP 2008, Cherbourg,
- Chairman d'une session posters à la conférence CORESA 2006, Caen,
- Co-organisation avec C. Charrier de la conférence nationale « Compression et Représentation des Signaux Audiovisuels (CORESA) » à Caen en 2006.
- Co-Organisation avec C. Charrier d'une session spéciale « Color Image Processing and Analysis for Machine Vision », lors de la conférence Complex Systems Intelligence and Modern Technological Applications (CSIMTA) en 2004, à Cherbourg-Octeville. Cette session spéciale a donné lieu à un numéro spécial de la revue International Journal of Robotics and Automation, Vol 20, N° 2, ISSN 0826-8185, 2005.
- Co-Organisation avec C. Charrier de la première journée « Imagerie couleur : indexation et perception » à Cherbourg en 2002. Cette journée, parrainée par le Groupe Français de l'Imagerie Numérique Couleur (GFINC) et le GDR-PRC ISIS, a été initiée par C. Charrier et moi-même, elle est depuis reconduite tous les ans sous l'intitulé « Traitements perceptifs couleur dédiés à l'image et à la vidéo ».

### **A.6 Projets de recherche et valorisation**

#### **A.6.1 Projets nationaux ANR**

**Projet FOGRIMMI.** Ce projet, dont je suis le porteur scientifique, s'inscrit dans le cadre de l'appel à projets Masses de Données Connaissances Ambiantes 2006 de l'ANR. Il implique les laboratoires LUSAC, GREYC, GRECAN, SIC et LABRI. Ce projet de recherche est consacré à la conception et à la validation d'un modèle permettant de coder à leur résolution intrinsèque les différentes structures contenues dans des images de lames microscopiques entières. Coût total du projet : 286000 euros, durée : 36 mois.

#### **A.6.2 Projets régionaux du Comité Régional Imagerie et Technologies de l'Information et de la Communication (CRITIC)**

- 2003-2005 : « Génération dynamique d'objets visuels pour une cartographie interactive multimédia sur le web », porté par le LUSAC et la société CHI Webdesign, coût total du projet : 120000 euros, durée : 18 mois, (responsable : A. Elmoataz)

- 2006-2007 : « Industrialisation d'un ensemble de composants logiciels pour la cytopathologie », porté par le LUSAC et la société ADCIS, coût total du projet : 120000 euros, durée : 18 mois. Je suis responsable scientifique de ce projet.

### **A.6.3 Collaborations Industrielles**

#### **A.6.3.1 Diplôme de Recherche Technologique**

2005-2006 : « Etude et caractérisation de l'intérieur de tubes des parois d'une zone hostile dans une autre zone hostile », avec COGEMA site de la Hague, Beaumont Hague. Montant 4000 euros. Responsables : A. Elmoataz, D. Houivet, O. Lezoray. Ce contrat accompagne un diplôme de recherche technologique (DRT) dont l'objectif est le recensement des moyens de caractérisation à distance de l'intérieur d'un tube situé dans une zone inaccessible.

#### **A.6.3.2 Projets de l'incubateur bas-normand d'entreprises technologiques**

2004-2005 : « Développement d'une application logicielle pour la mise en place de services innovants autour de la photographie numérique », porté par la société Logipix, le CERTIC, le LUSAC et le GREYC, coût total du projet : 4000 euros, durée : 24 mois. Je suis responsable scientifique de ce projet.

### **A.6.4 Consultances**

2002-2005 : Consultant en systèmes d'information et membre du conseil d'administration de la société Mangrove Systems basée à Granville, Manche. Cette société a reçu le soutien de l'Anvar en 2001 dans le cadre du concours national d'aide à la création d'entreprises de technologies innovantes et elle a été rachetée par la société américaine Neoware en Janvier 2005.



---

# Publications

---

## Publications : Tableau récapitulatif <sup>1</sup>

PUBLICATIONS	NOMBRE
<i>Chapitres de Livre</i> [1, 2]	2
<i>Revue internationale avec comité de lecture</i> [3–20]	18
<i>Revue nationale avec comité de lecture</i> [21–23]	3
<i>Conférences internationales avec actes et comité de lecture</i> [24–58]	35
<i>Conférences nationales avec actes et comité de lecture</i> [59–71]	13

## Chapitres de livres

- [1] G. Lebrun, O. Lezoray, C. Charrier, and H. Cardot. An EA multi-model selection for svm multi-class schemes. In Alejandro Pazos, Juan R. Rabuñal, Julián Dorado, editor, *Encyclopedia of Artificial Intelligence*. Information Science Reference, 2008.
- [2] O. Lezoray, G. Lebrun, C. Meurie, C. Charrier, A. Elmoataz, M. Lecluse, and H. Elie. Machine learning in morphological segmentation of microscopic images. In Themis P. Exarchos, Athanasios Papadopoulos, and Dimitrios I. Fotiadis, editors, *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*. Information Science Reference, 2008.

---

<sup>1</sup>La majorité des publications est disponible à <http://www.info.unicaen.fr/~lezoray/Publications.php>

## Revue Internationale

- [3] O. Lezoray, A. Elmoataz, H. Cardot, G. Gougeon, M. Lecluse, H. Elie, and M. Revenu. Segmentation of cytological images using color and mathematical morphology. *Acta Stereologica*, 18(1) :1–14, 1999. Selected papers of the 7th European Congress of Stereology.
- [4] O. Lezoray, A. Elmoataz, H. Cardot, and M. Revenu. Segmentation of color images from serous cytology for an automatic cellular classification. *Analytical and quantitative cytology and histology*, 22(4) :311–323, 2000.
- [5] O. Lezoray and H. Cardot. A neural network architecture for data classification. *International Journal of Neural Systems*, 11(1) :33–42, 2001.
- [6] O. Lezoray and H. Cardot. Cooperation of color pixel classification schemes and color watershed : a study for microscopical images. *IEEE transactions on Image Processing*, 11(7) :783–789, 2002.
- [7] O. Lezoray. Supervised automatic histogram clustering and watershed segmentation. application to microscopic medical images. *Image Analysis and Stereology*, 22 :113–120, 2003.
- [8] O. Lezoray, A. Elmoataz, and H. Cardot. A color object recognition scheme : application to cellular sorting. *Machine Vision and Applications*, 14 :166–171, 2003.
- [9] C. Meurie, G. Lebrun, O. Lezoray, and A. Elmoataz. A comparison of supervised pixels-based color image segmentation methods. application in cancerology. *WSEAS Transactions on Computers*, 2(3) :739–744, jul 2003. ISSN 1109-2750.
- [10] O. Lezoray, D. Fournier, and H. Cardot. Neural network induction graph for pattern recognition. *Neurocomputing*, 57 :257–274, 2004.
- [11] C. Meurie, O. Lezoray, C. Charrier, and A. Elmoataz. Combination of multiple pixel classifiers for microscopic image segmentation. *IJRA (IASTED International Journal of Robotics and Automation)*, 20(2) :63–69, 2005. Special issue on Colour Image Processing and Analysis for Machine Vision, ISSN 0826-8185.
- [12] C. Charrier, G. Lebrun, and O. Lezoray. Evidential segmentation of microscopic color images with pixel classification posterior probabilities. *Journal of Multimedia*, 2(3) :57–65, 2007.
- [13] G. Lebrun, C. Charrier, O. Lezoray, and H. Cardot. A fast and efficient segmentation scheme for cell microscopic image. *Cellular and Molecular Biology*, 53, Biomedical Signal and Image Processing Series(2) :51–61, 2007.
- [14] G. Lebrun, C. Charrier, O. Lezoray, and H. Cardot. Tabu search model selection for SVM. *International Journal of Neural Systems*, A paraître, 2007.
- [15] O. Lezoray and C. Charrier. Color image segmentation using morphological clustering and fusion with color image segmentation using morphological clustering and fusion with automatic scale selection. *Pattern Recognition Letters*, à paraître, 2007.
- [16] O. Lezoray, A. Elmoataz, and S. Bougleux. Graph regularization for color image processing. *Computer Vision and Image Understanding (CVIU)*, 107(1-2) :38–55, 2007.
- [17] O. Lezoray and M. Lecluse. Automatic segmentation and classification of cells from broncho alveolar lavage. *Image Analysis and Stereology*, 26 :111–119, 2007.

- [18] O. Lezoray, C. Meurie, and A. Elmoataz. Graph-based ordering scheme for color image filtering. *International Journal of Image and Graphics*, à paraître, 2007.
- [19] A. Elmoataz, O. Lezoray, and S. Bougleux. Nonlocal discrete regularization on weighted graphs : a framework for image and manifold processing. *IEEE transactions on Image Processing*, à paraître.
- [20] O. Lezoray and H. Cardot. Comparing combination rules of pairwise neural networks classifiers. *Neural Processing Letter*, à paraître.

## Revue Nationale

- [21] O. Lezoray, A. Elmoataz, H. Cardot, and M. Revenu. Segmentation d'images couleur : applications en microscopie cellulaire. *Traitement du signal*, 17(1) :33–45, 2000.
- [22] O. Lezoray and H. Cardot. Segmentation of color images by clustering 2D histogram and merging regions. *Moroccan Journal of Control Computer Science and Signal Processing*, 2003.
- [23] O. Lezoray and C. Charrier. Segmentation d'images couleur par coalescence non supervisée d'histogrammes 2D et fusion de régions selon la théorie de dempster-shafer. *Traitement du signal*, 21(6) :605–622, 2005.

## Conférences Internationales

- [24] O. Lezoray, A. Elmoataz, H. Cardot, G. Gougeon, M. Lecluse, H. Elie, and M. Revenu. Segmentation of cytological images using color and mathematical morphology. In *Abstract Book of the 7th European Congress of Stereology*, page 52, 1998.
- [25] O. Lezoray, A. Elmoataz, H. Cardot, and M. Revenu. A.R.C.T.I.C : Un système de tri cellulaire par analyse d'images. In *Vision Interface*, pages 312–319, 1999.
- [26] O. Lezoray, H. Cardot, D. Fournier, and M. Revenu. MONNA : a Multiple Ordinate Neural Network Architecture. In *EIS (Engineering of Intelligent Systems)*, pages 47–53, 2000.
- [27] O. Lezoray, A. Elmoataz, H. Cardot, and M. Revenu. A color morphological segmentation. In *CGIP (Color in Graphics and Image Processing)*, pages 170–175, 2000.
- [28] O. Lezoray, A. Elmoataz, and H. Cardot. Automatic cellular recognition in serous cytology. In *Abstract Book of the 7th ESACP, Analytical and Cellular Pathology*, volume 22, page 70, 2001.
- [29] O. Lezoray, M. Lecluse, and H. Elie. Study of the texture of cells in serous cytology by image analysis. In *Abstract Book of the 7th ESACP, Analytical and Cellular Pathology*, volume 22, page 74, 2001.
- [30] J. Thuillier, B. Plancoulaine, O. Lezoray, P. Herlin, and D. Bloyet. Building of a microscopical acquisition device dedicated to clinical oncology. In *Abstract Book of the 7th ESACP, Analytical and Cellular Pathology*, volume 22, page 77, 2001.



- [31] H. Cardot and O. Lezoray. Graph of neural networks for pattern recognition. In *ICPR (International Conference on Pattern Recognition)*, volume 2, pages 873–876, Août 2002.
- [32] O. Lezoray and H. Cardot. Bayesian marker extraction for color watershed in segmenting microscopic images. In *ICPR (International Conference on Pattern Recognition)*, volume 1, pages 739–742, Août 2002.
- [33] O. Lezoray and H. Cardot. Histogram and watershed based segmentation of color images. In *CGIV (Color in Graphics, Imaging and Vision)*, pages 358–362, Avril 2002.
- [34] C. Charrier and O. Lezoray. Influence of JPEG2000 standard compression on color image segmentation quality. In *ICISP (IAPR International Conference on Image and Signal Processing)*, volume 2, pages 389–396, juin 2003.
- [35] O. Lezoray. Unsupervised 2d multiband histogram clustering and region merging for color image segmentation. In *ISSPIT (IEEE International Symposium on Signal Processing and Information Technology)*, pages 267–270, 2003.
- [36] O. Lezoray and H. Cardot. Hybrid color image segmentation using 2D histogram clustering and region merging. In *ICISP (IAPR International Conference on Image and Signal Processing)*, volume 1, pages 22–29, juin 2003.
- [37] O. Lezoray and A. Elmoataz. Graph based smoothing and segmentation of color images. In *ISSPA (International Symposium on Signal Processing and its Applications)*, volume 1, pages 22–29, 2003.
- [38] C. Meurie, G. Lebrun, O. Lezoray, and A. Elmoataz. A comparison of supervised pixels-based color image segmentation methods. application in cancerology. In *SSIP (WSEAS International Conference On Signal, Speech and Image Processing)*, 2003.
- [39] C. Meurie, G. Lebrun, O. Lezoray, and A. Elmoataz. A supervised segmentation scheme for cancerology color images. In *ISSPIT (IEEE International Symposium on Signal Processing and Information Technology)*, pages 664–667, 2003.
- [40] C. Meurie, O. Lezoray, H. Cardot, and A. Elmoataz. Comparaison de classifieurs non-supervisés pour la segmentation d’images couleur : Application en imagerie biomédicale. In *ICISP (IAPR International Conference on Image and Signal Processing)*, volume 1, pages 30–37, juin 2003.
- [41] O. Lezoray. An unsupervised color image segmentation based on morphological 2D clustering and fusion. In *CGIV (Color in Graphics, Imaging and Vision)*, pages 173–177, 2004.
- [42] C. Meurie, O. Lezoray, and A. Elmoataz. Multiple pixel classifier combination for bronchial tumors image segmentation. In *CSIMTA (International Conference on Complex Systems Intelligence and Modern Technological Applications)*, volume 1, pages 305–310, 2004. special session on Color Image Processing and Analysis for Machine Vision.
- [43] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot. Fast pixel classification by SVM using vector quantization, tabu search and hybrid color space. In A. Gagalo-wicz and W. Philips, editors, *CAIP (International Conference on Computer Analysis of Images and Patterns)*, volume LNCS 3691 of *Computer Science*, pages 685–692. Springer, 2005. ISBN 3-540-28969-0.

- [44] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot. A fast and efficient segmentation scheme for cell microscopic image. In *World Congress of Cellular and Molecular Biology*, pages 35–36, 2005.
- [45] O. Lezoray, C. Meurie, and A. Elmoataz. A graph approach to color mathematical morphology. In *ISSPIT (IEEE International Symposium on Signal Processing and Information Technology)*, pages 856–861, dec 2005.
- [46] Olivier Lezoray and Hubert Cardot. Combining multiple pairwise neural networks classifiers : A comparative study. In *ANNIIP (International Workshop on Artificial Neural Networks and Intelligent Information Processing)*, pages 52–61, 2005.
- [47] C. Charrier, G. Lebrun, and O. Lezoray. Fusion of SVM-based microscopic color images through colorimetric transformation. In *ICASSP (IEEE International Conference on Acoustics, Speech, and Signal Processing)*, volume 2, pages 1112–1115, 2006.
- [48] C. Charrier, G. Lebrun, and O. Lezoray. A machine learning-based color image quality metric. In *CGIV (Color in Graphics, Imaging and Vision)*, pages 251–256, 2006.
- [49] G. Lebrun, O. Lezoray, C. Charrier, and H. Cardot. A new model selection method for SVM. In *IDEAL (Intelligent Data Engineering and Automated Learning)*, volume LNCS 4424, pages 99–107, 2006.
- [50] G. Lebrun, O. Lezoray, C. Charrier, and H. Cardot. Speed-up LOO-CV with SVM classifier. In *IDEAL (Intelligent Data Engineering and Automated Learning)*, volume LNCS 4424, pages 108–115, 2006.
- [51] O. Lezoray, C. Meurie, P. Belhomme, and A. Elmoataz. Multi-scale image segmentation in a hierarchy of partitions. In *EUSIPCO (European Signal Processing Conference)*, CD Proceedings, 2006.
- [52] C. Charrier, G. Lebrun, and O. Lezoray. Selection of features by a machine learning expert to design a color image quality metric. In ISBN 09774740, editor, *Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale*, CD Proceedings, 2007.
- [53] G. Lebrun, O. Lezoray, C. Charrier, and H. Cardot. An EA multi-model selection for SVM multiclass schemes. In *International Work Conference on Artificial Neural Networks, Special Session on Kernel Methods*, volume LNCS 4507, pages 257–264, 2007.
- [54] O. Lezoray, S. Bougleux, and A. Elmoataz. Parameterless discrete regularization on graphs for color image filtering. In *ICIAR (International Conference on Image Analysis and Recognition)*, volume Lecture Notes in Computer Science 4633, pages 46–57, 2007.
- [55] O. Lezoray, C. Meurie, and A. Elmoataz. Mathematical morphology in any color space. In *IAPR International Conference on Image Analysis and Processing, Computational Color Imaging Workshop*. IEEE, 2007.
- [56] C. Meurie and O. Lezoray. A new method of morphological hierarchical segmentation. In *VIE (IEEE and IET Visual Information Engineering)*, 2007.
- [57] V. Ta, O. Lezoray, and A. Elmoataz. Graphs regularization for data sets and images : Filtering and semi-supervised classification. In *PASCAL Workshop : Learning from and with graphs, 6th IAPR -TC-15 Workshop on Graph based Representations in Pattern Recognition*, CD Proceedings, 2007.

- [58] V.T. Ta, O. Lezoray, and A. Elmoataz. Graph based semi and unsupervised classification and segmentation of microscopic images. In *ISSPIT*, 2007.

## Conférences Nationales

- [59] O. Lezoray, A. Elmoataz, H. Cardot, G. Gougeon, M. Lecluse, H. Elie, and M. Revenu. Segmentation d'images couleur : application à la cytologie des séreuses en vue d'un tri informatique cellulaire. In *RFIA (Reconnaissance des Formes et Intelligence Artificielle)*, volume 2, pages 287–295, 1998.
- [60] G. Lebrun, C. Charrier, O. Lezoray, and H. Cardot. Réduction du temps d'apprentissage des SVM par quantification vectorielle. In *CORESA (COmpression et REprésentation des Signaux Audiovisuels)*, pages 223–226, 2004.
- [61] C. Charrier, G. Lebrun, and O. Lezoray. Fusion de segmentation d'images microscopiques par SVM dans différents espaces couleur selon la théorie de l'évidence. In *Colloque Gretsi sur le traitement du signal et des images*, pages 913–916, sep 2005.
- [62] C. Charrier and O. Lezoray. Segmentation d'images couleur par la théorie de l'évidence. In *CORESA (COmpression et REprésentation des Signaux Audiovisuels)*, pages 291–296, 2005.
- [63] L. Garnier, C. Meurie, and O. Lezoray. Superquartiques de dupin joignant un superellipsoïde de révolution et un plan. In *Groupe de Travail en Modélisation Géométrique*, pages 77–91, 2005.
- [64] G. Lebrun, C. Charrier, O. Lezoray, and H. Cardot. Construction de fonctions de décision performantes et de complexités réduites avec des SVM. In *RJCIA (Rencontres Jeunes Chercheurs en Intelligence Artificielle)*, pages 85–89, mai 2005.
- [65] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot. Classification rapide de pixels par SVM dans un espace couleur hybride. In *CORESA (COmpression et REprésentation des Signaux Audiovisuels)*, pages 231–236, nov 2005.
- [66] O. Lezoray and M. Lecluse. Stratégie d'analyse d'images de lavages broncho alvéolaires. In *JSTIM (Journées Sciences, Technologies et Imagerie pour la Médecine)*, pages 118–119, 2005.
- [67] O. Lezoray, C. Meurie, P. Belhomme, and A. Elmoataz. Hiérarchies de partitions pour la simplification et la segmentation d'images couleur. In *CORESA (COmpression et REprésentation des Signaux Audiovisuels)*, pages 231–236, nov 2005.
- [68] O. Lezoray, A. Elmoataz, and S. Bougleux. Régularisation sur graphes pour le traitement d'images couleur. In *CORESA (COmpression et REprésentation des Signaux Audiovisuels)*, pages 236–241, nov 2006.
- [69] O. Lezoray, C. Meurie, and A. Elmoataz. Ordre de couleurs : une approche par graphe. In *CORESA (COmpression et REprésentation des Signaux Audiovisuels)*, pages 64–69, nov 2006.
- [70] C. Meurie and O. Lezoray. Influence de l'espace couleur pour la segmentation hiérarchique d'images. In *CORESA (COmpression et REprésentation des Signaux Audiovisuels)*, pages 242–247, nov 2006.

- [71] G. Lebrun, C. Charrier, and O. Lezoray. Schéma d'apprentissage basé sur la sélection automatique d'exemples pertinents. In *CORESA (COmpression et REprésentation des Signaux Audiovisuels)*, 2007.

## Articles en cours de soumission

- [72] V.T. Ta, O. Lezoray, and A. Elmoataz. Discrete regularization : a common graph based framework for different applications. Technical report, 2007.
- [73] A. Elmoataz, O. Lezoray, and S. Bogleux. Nonlocal discrete p-laplacian driven image and manifold processing. *Comptes Rendus de l'Académie des Sciences - Mécanique*, soumis.
- [74] V.T. Ta, S. Bogleux, A. Elmoataz, and O. Lezoray. Nonlocal anisotropic discrete regularization for image, data filtering and clustering. *Computer Vision and Image Understanding (CVIU)*, soumis, 2007.



---

# RÉSUMÉ

---

L'ensemble des travaux présentés dans la cadre de cette Habilitation à diriger des recherches concerne le traitement d'images par modèles discrets sur graphes et la classification de données d'images par apprentissage.

Dans un premier temps, nous présentons nos travaux concernant le traitement d'images et de données d'images à partir d'une représentation basée sur des graphes de topologies arbitraires. Nous nous intéressons tout d'abord au problème de la définition d'un ordre de données vectorielles pour le traitement d'images (principalement pour le filtrage médian ou morphologique). Deux approches sont présentées. La première définit un chemin Hamiltonien sur un graphe complet associé aux données vectorielles. Le principe repose sur une décimation et une exploration d'une hiérarchie d'arbres couvrants minimaux. La seconde définit une réduction de dimension non linéaire par décomposition en valeurs propres de marches aléatoires. Nous nous intéressons ensuite au problème de la régularisation discrète de données. Nous proposons un cadre général pour effectuer cette régularisation discrète sur graphes pondérés de topologie arbitraires. En modifiant la topologie du graphe, les poids des arêtes et les données associées aux noeuds, notre approche généralise un certain nombre d'approches de la littérature et permet naturellement une formulation non locale. Enfin, nous présentons plusieurs méthodes permettant de générer des hiérarchies de partitions par méthodes morphologiques, par diffusion et fusion de régions dans un graphe d'adjacence de régions et par coalescence non supervisée morphologique d'histogrammes bi-variés et fusion de régions.

Dans un second temps, nous présentons nos travaux concernant la classification de données d'images (pixels, régions ou images) par apprentissage. Nous nous intéressons tout d'abord à la fusion de décisions de classificateurs pour la classification de pixels d'images microscopiques. Nous nous intéressons ensuite à la production de fonctions de décisions de complexités réduites ayant de bonnes performances en généralisation. Nous étudions plus particulièrement le cas des machines à vecteurs de support et proposons une recherche avec tabous permettant d'aboutir à un compromis entre complexité et capacités de généralisation tout en effectuant une sélection des exemples pertinents, une sélection des attributs pertinents et une sélection des hyper-paramètres optimaux. Enfin, nous présentons nos contributions à la conception de schémas de décodage dans le cadre d'un schéma multi-classes de combinaison de classificateurs binaires.