



HAL
open science

Modélisation et apprentissage de relations spatiales pour la reconnaissance et l'interprétation d'images

Michaël Clément

► To cite this version:

Michaël Clément. Modélisation et apprentissage de relations spatiales pour la reconnaissance et l'interprétation d'images. Traitement des images [eess.IV]. Université Sorbonne Paris Cité, 2017. Français. NNT : 2017USPCB024 . tel-02111681

HAL Id: tel-02111681

<https://theses.hal.science/tel-02111681>

Submitted on 26 Apr 2019

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

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

THÈSE

présentée pour obtenir le grade de

Docteur de l'Université Paris Descartes

Spécialité : Informatique

École doctorale Informatique, Télécommunications et Électronique (Paris)

Michaël Clément

Modélisation et apprentissage de relations spatiales pour la reconnaissance et l'interprétation d'images

Soutenue publiquement le 26 septembre 2017

Membres du jury

<i>Directeur de thèse</i>	Laurent Wendling	Professeur, Université Paris Descartes
<i>Rapporteurs</i>	Isabelle Debled-Rennesson	Professeur, Université de Lorraine
	Christophe Garcia	Professeur, INSA de Lyon
<i>Examineurs</i>	Valérie Gouet-Brunet	Directeur de recherche, IGN
	Rolf Ingold	Professeur, Université de Fribourg
	Camille Kurtz	Maître de Conférences, Université Paris Descartes
	Cédric Wemmert	Professeur, Université de Strasbourg

Remerciements

Je tiens à remercier mon directeur de thèse, Laurent Wendling, ainsi que mon encadrant, Camille Kurtz, pour m'avoir permis de réaliser ces travaux de thèse dans d'excellentes conditions humaines et scientifiques. Vos conseils et votre soutien durant de ces trois années m'ont été très bénéfiques. Merci pour votre attitude positive et bienveillante, même lors des moments de doute. Merci pour les répétitions et relectures interminables que j'ai pu vous faire subir.

J'adresse mes remerciements à Isabelle Debled-Rennesson et Christophe Garcia, pour leur relecture attentive de ce manuscrit. Je remercie également Valérie Gouet-Brunet, Rolf Ingold et Cédric Wemmert d'avoir accepté de participer à mon jury de thèse.

Je tiens ensuite à exprimer toute ma gratitude à Adrien Poulenard, avec qui j'ai eu l'opportunité de collaborer au début de ma thèse, et qui est à l'origine d'une partie de ces travaux.

Un grand merci à Georges Stamon, figure paternelle de l'équipe SIP, pour sa gentillesse réconfortante, ainsi qu'aux autres membres permanents, Florence Cloppet, Nicolas Loménie et Nicole Vincent, qui ont constitué pour moi des référents et des repères pendant ces trois ans. Un grand merci à tous les doctorants, postdoctorants et stagiaires qui ont passé un séjour au laboratoire, et avec qui j'ai pu partager de bons moments. J'ai notamment une pensée amicale pour mes éternels camarades de bureau, Héloïse et Cuong.

Je remercie également du fond du cœur l'ensemble de mes proches, famille et amis, qui ont vécu avec moi ces années de thèse, m'ont soutenu et supporté. Merci, en particulier, à ma mère, qui m'a toujours aidé et encouragé dans mes choix, parfois difficiles, tout au long de mon parcours universitaire.

Enfin, mes remerciements les plus chaleureux vont à Lucie, qui a toujours été à mes côtés, et sans qui cette thèse n'aurait pas pu voir le jour.

Résumé

Ces dernières années, la quantité de données visuelles produites par divers types de capteurs est en augmentation permanente. L'interprétation et l'indexation automatique de telles données constituent des défis importants pour les domaines liés à la reconnaissance de formes et la vision par ordinateur. Dans ce contexte, la position relative des différents objets d'intérêt composant les images représente une information particulièrement importante pour interpréter leur contenu. Les relations spatiales sont en effet porteuses d'une sémantique riche, qui est fortement liée à la perception humaine. Les travaux de recherche présentés dans cette thèse proposent ainsi d'explorer différentes approches génériques de description de l'information spatiale, en vue de les intégrer dans des systèmes de reconnaissance et d'interprétation d'images de haut niveau.

Tout d'abord, nous présentons une approche pour la description de configurations spatiales complexes, où les objets peuvent être imbriqués les uns dans les autres. Cette notion est formalisée par deux nouvelles relations spatiales, nommées enlacement et entrelacement. Nous proposons un modèle qui permet de décrire et de visualiser ces configurations avec une granularité directionnelle. Ce modèle est validé expérimentalement pour des applications en imagerie biomédicale, en télédétection et en analyse d'images de documents. Ensuite, nous présentons un cadre d'apprentissage de relations spatiales composites à partir d'ensembles d'images. Inspirée des approches par sacs de caractéristiques visuelles, cette stratégie permet de construire des vocabulaires de configurations spatiales apparaissant dans les images, à différentes échelles. Ces caractéristiques structurelles peuvent notamment être combinées avec des descriptions locales, conduisant ainsi à des représentations hybrides et complémentaires. Les résultats expérimentaux obtenus sur différentes bases d'images structurées permettent d'illustrer l'intérêt de cette approche pour la reconnaissance et la classification d'images.

Abstract

In recent years, the amount of visual data produced by various types of sensors has been continuously increasing. The automatic interpretation and indexation of such data constitute an important challenge in the fields of pattern recognition and computer vision. In this context, the relative position of the different objects of interest depicted in images represents particularly important information for the interpretation of their content. Spatial relations indeed carry rich semantics that are strongly tied with human perception. The research work presented in this thesis thus proposes to explore different generic approaches to the description of spatial information, in order to integrate them in high-level image recognition and understanding systems.

First, we present an approach for the description of complex spatial configurations, where objects can be imbricated in each other. This notion is formalized by two novel spatial relations, namely enlacement and interlacement. We propose a model to describe and to visualize these configurations with directional granularity. This model is experimentally validated for applications in biomedical imaging, remote sensing and document image analysis. Then, we present a framework for learning composite spatial relations from image datasets. Inspired by bags of visual features approaches, this strategy allows to build vocabularies of spatial configurations occurring across images, at different scales. These structural features can notably be combined with local descriptions, leading to hybrid and complementary representations. Experimental results obtained for different datasets of structured images highlight the interest of this approach for image recognition and classification tasks.

Table des matières

Introduction	1
Contexte de la thèse	3
Problématique	4
Verrous scientifiques et axes de recherche	5
Organisation du document	6
Partie I Relations spatiales en analyse d’images	9
1 Relations spatiales : définitions et méthodes classiques	11
1.1 Relations spatiales élémentaires	12
1.1.1 Propositions linguistiques spatiales	12
1.1.2 Relations spatiales de Freeman	12
1.1.3 Familles de relations spatiales	13
1.2 Méthodes classiques et limitations	14
1.2.1 Approches qualitatives	14
1.2.2 Approches quantitatives	17
1.2.3 Limitations et synthèse	19
2 Dualité entre relations spatiales et position relative	21
2.1 Relations spatiales et position relative	21
2.2 Descripteurs de position relative	22
2.2.1 Relations spatiales directionnelles	23
2.2.2 Combinaisons avec d’autres familles de relations spatiales	24
2.3 Paysages de relations spatiales	26
2.3.1 Paysages flous morphologiques	27
2.3.2 Champs de forces	28
2.4 Synthèse	29
3 Vers des relations spatiales complexes et composites	31
3.1 Relations spatiales complexes	32
3.1.1 La relation « entouré par »	32
3.1.2 La relation « entre »	33
3.1.3 Autres relations complexes	35

3.2	Relations spatiales composites	36
3.2.1	Descripteurs composites de relations spatiales	36
3.2.2	Apprentissage de relations spatiales	37
3.3	Synthèse et positionnement des travaux	39
Partie II Modélisation de relations spatiales complexes		41
4	Descripteurs directionnels d'enlacement et d'entrelacement	43
4.1	Relations d'enlacement et d'entrelacement	44
4.2	Enlacement directionnel entre objets	45
4.2.1	Notation et définitions préliminaires	45
4.2.2	Enlacement le long d'une droite	47
4.2.3	Enlacement global entre objets	49
4.2.4	Propriétés d'invariance	52
4.3	Histogrammes d'enlacement et d'entrelacement	52
4.3.1	Représentations par histogrammes	52
4.3.2	Exemples illustratifs	54
4.3.3	Complexité algorithmique	55
4.4	Bilan scientifique	56
5	Validations expérimentales : reconnaissance de configurations spatiales	57
5.1	Protocole expérimental général	58
5.1.1	Méthodes comparatives de référence	58
5.1.2	Classification et évaluation des résultats	59
5.2	Caractérisation de vaisseaux sanguins en ophtalmologie	61
5.2.1	Jeux de données	61
5.2.2	Protocole de validation	63
5.2.3	Résultats expérimentaux et discussion	65
5.3	Reconnaissance de lettrines décoratives	66
5.3.1	Jeu de données	66
5.3.2	Protocole de validation	67
5.3.3	Résultats expérimentaux et discussion	68
5.4	Recherche et classification d'objets géographiques urbains	68
5.4.1	Jeu de données	68
5.4.2	Protocole de validation	69
5.4.3	Résultats expérimentaux et discussion	73
5.5	Bilan scientifique	74
6	Évaluations des relations d'enlacement et d'entrelacement	75
6.1	Évaluations à partir de descripteurs	76
6.1.1	Approche par mise en correspondance d'ensembles flous	76

6.1.2	Approche par agrégation directionnelle	77
6.1.3	Résultats illustratifs	78
6.2	Paysages flous directionnels d'enlacement	80
6.2.1	Définitions et évaluations floues	80
6.2.2	Résultats illustratifs	84
6.2.3	Vers les paysages flous d'entrelacement	85
6.3	Applications à la relation spatiale « entouré par »	86
6.3.1	Nouvelles modélisations à partir de l'enlacement	86
6.3.2	Exemples caractéristiques	88
6.3.3	Résultats expérimentaux et discussion	89
6.4	Bilan scientifique	93
Partie III Apprentissage de relations spatiales composites		95
7	Graphes relationnels attribués de relations spatiales	97
7.1	Décomposition en histogrammes de forces	98
7.1.1	Histogrammes de forces	98
7.1.2	Graphes relationnels attribués d'histogrammes de forces	100
7.1.3	Stratégies de comparaison et de mise en correspondance	103
7.1.4	Méthode de décomposition	106
7.2	Descripteurs FHD hiérarchiques	107
7.2.1	ARG hiérarchiques et contraction d'arêtes	108
7.2.2	Méthode de décomposition hiérarchique	109
7.3	Bilan scientifique	110
8	Apprentissage par sacs de relations spatiales	113
8.1	Sacs de caractéristiques : état de l'art synthétique	114
8.1.1	Vocabulaires de caractéristiques visuelles locales	114
8.1.2	Vers les sacs de relations spatiales	115
8.2	Désassemblage du descripteur FHD hiérarchique	116
8.3	Vocabulaires structurels de relations spatiales et de formes	117
8.3.1	Construction des vocabulaires	117
8.3.2	Stratégies d'encodage et d'agrégation	119
8.3.3	Compatibilité avec les caractéristiques visuelles locales	123
8.4	Bilan scientifique	123
9	Validations expérimentales : reconnaissance d'images structurées	125
9.1	Protocole expérimental général	126
9.1.1	Méthodes comparatives de référence	126
9.1.2	Classification et évaluation des résultats	126
9.2	Reconnaissance d'objets structurés	127

9.2.1	Jeux de données	127
9.2.2	Protocole de validation	129
9.2.3	Résultats expérimentaux et discussion	130
9.3	Reconnaissance d’images de scènes d’extérieur	132
9.3.1	Jeux de données	132
9.3.2	Protocole de validation	133
9.3.3	Résultats expérimentaux et discussion	134
9.4	Bilan scientifique	136
10 Vers l’apprentissage de descripteurs locaux d’enlacement		137
10.1	Contexte applicatif et motivations	138
10.1.1	Reconnaissance de lettrines décoratives anciennes	138
10.1.2	Configurations spatiales complexes locales	139
10.2	Approche par sacs d’enlacement local	140
10.2.1	Décomposition de lettrines en couches structurelles	140
10.2.2	Descripteurs locaux d’enlacement	142
10.2.3	Apprentissage par sacs d’enlacement	143
10.3	Validations expérimentales	144
10.3.1	Jeux de données	144
10.3.2	Protocole de validation	145
10.3.3	Résultats expérimentaux et discussion	146
10.4	Bilan scientifique	148
Conclusion		150
Bilan et contributions		153
Perspectives		154
Bibliographie		157
Annexes		171
A	Liste des publications	171
B	Démonstrations des propriétés du descripteur d’enlacement	173
C	Développements logiciels	177

Table des figures

1.1	Illustration de l’ambiguïté du positionnement spatial entre deux objets	14
1.2	Illustration des relations spatiales topologiques RCC8	16
1.3	Ensembles flous trapézoïdaux pour les relations spatiales cardinales	18
2.1	Illustration du principe de calcul des histogrammes d’angles et de forces	23
2.2	Illustration des treize intervalles temporels d’Allen	25
2.3	Définition d’une région de référence à partir de caractéristiques topologiques	26
2.4	Paysages flous directionnels pour la relation spatiale « <i>à droite de</i> »	28
2.5	Exemple de \mathcal{F} - <i>template</i> obtenu pour un objet de référence	29
3.1	Visibilité et de couverture angulaire pour la relation spatiale « <i>entouré par</i> »	32
3.2	Relation spatiale « <i>entre</i> » à partir de l’enveloppe convexe de l’union des objets	34
3.3	Relation spatiale « <i>entre</i> » pour un maillage de points d’une image histologique	34
3.4	Détection d’objets géographiques alignés dans une image de télédétection	35
3.5	Illustration schématique de la décomposition en histogrammes de forces	37
3.6	Exemples de méta-modèles spatiaux générés à partir de plusieurs exemples	38
4.1	Exemples caractéristiques de configurations d’enlacement ou d’entrelacement	44
4.2	Illustration d’une droite orientée $\Delta^{(\theta,\rho)}$ définie dans le plan	46
4.3	Différentes coupes longitudinales d’une objet binaire A	47
4.4	Illustration de l’enlacement entre des points le long de droites orientées	48
4.5	Illustration de l’enlacement entre des segments le long de droites orientées	49
4.6	Illustration du calcul de l’enlacement pour deux objets A et B	50
4.7	Évolution des profils d’enlacement E_{AB} et E_{BA} sur une série d’exemples	51
4.8	Représentations polaires des descripteurs d’enlacement et d’entrelacement	54
5.1	Exemples d’images de fonds d’œil des jeux de données DRIVE et CHASEDB1	61
5.2	Exemples d’images de fonds d’œil du jeu de données HRF	62
5.3	Exemples de réseaux vasculaires segmentés et bruités	64
5.4	Courbes de précision-rappel pour la classification du jeu de données HRF	65
5.5	Exemples d’images provenant du jeu de données de lettrines décoratives	66
5.6	Exemples de binarisations pour différentes images de lettrines décoratives	67
5.7	Exemple d’une zone urbaine issues de l’image satellitaire THRS	69

5.8	Niveaux de classification de l'image satellitaire	70
5.9	Exemples de blocs urbains issus de l'image satellitaire THRS	71
5.10	Courbes de précision-rappel obtenues pour la classification de blocs urbains	72
5.11	Requêtes sémantiques d'enlacement et d'entrelacement entre des objets urbains	73
6.1	Exemples de profils flous directionnels sinusoïdaux	77
6.2	Représentations polaires des descripteurs d'enlacement et d'entrelacement (rappel)	78
6.3	Exemples de paysages flous d'enlacement dans différentes directions	81
6.4	Paysages flous d'enlacement d'un objet pour différentes amplitudes	82
6.5	Paysages flous d'enlacement d'un objet pour différentes directions	83
6.6	Paysage flou d'enlacement global d'une spirale	84
6.7	Exemples de paysages flous d'entrelacement	85
6.8	Exemples caractéristiques de situations d'entourement entre deux objets	88
6.9	Exemple de situation d'entourement pour une image satellitaire	89
6.10	Évaluations directionnelles de l'entourement pour des exemples caractéristiques	91
6.11	Évaluations directionnelles de l'entourement pour le bassin d'Arcachon	92
6.12	Paysages flous d'enlacement pour l'image du bassin d'Arcachon	93
7.1	Illustration du calcul d'un histogramme de forces	99
7.2	Exemples illustratifs d'ARG complets pour différentes images	102
7.3	Exemple d'isomorphisme entre deux graphes complets	105
7.4	Méthode de décomposition d'un objet en ses sous-parties structurelles	107
7.5	Illustration schématique d'un descripteur FHD hiérarchique	109
7.6	Illustration d'une décomposition hiérarchique d'un objet structuré	110
8.1	Illustration schématique du désassemblage d'un descripteur FHD hiérarchique	116
8.2	Exemples de mots de relations spatiales pour des images de papillons	118
8.3	Exemples de mots de formes pour des images de papillons	119
8.4	Exemples de mots de relations spatiales pour des images de scènes	120
8.5	Exemples de mots de formes pour des images de scènes	121
9.1	Exemples d'images des jeux de données <i>Peale</i> , <i>Motorbikes</i> et <i>Patisboul</i>	128
9.2	Taux de reconnaissance pour le jeu de données <i>Peale</i>	129
9.3	Taux de reconnaissance pour le jeu de données <i>Motorbikes</i>	130
9.4	Taux de reconnaissance pour le jeu de données <i>Patisboul</i>	131
9.5	Exemples d'images du jeu de données <i>8 scènes</i>	132
9.6	Taux de reconnaissance pour le jeu de données <i>8 scènes</i>	133
9.7	Courbes de précision-rappel comparatives	134
9.8	Matrices de confusion comparatives	136
10.1	Exemples d'images de lettrines décoratives	138
10.2	Graphe de Zipf d'une image de lettrine décorative	141
10.3	Exemples de décompositions de Zipf pour des images de lettrines	142

Table des figures

10.4 Exemples de mots d'enlacement obtenus pour des images de lettrines	143
10.5 Courbes de précision-rappel pour la classification des images de lettrines	147
10.6 Matrices de confusion comparatives	148

Liste des tableaux

5.1	Étude de la robustesse au bruit sur les jeux de données DRIVE et CHASEDB1	64
5.2	Taux de reconnaissance pour la classification des images de lettrines	68
6.1	Évaluations de l'enlacement par mise en correspondance floue	79
6.2	Évaluations de l'enlacement par agrégation directionnelle	79
6.3	Synthèse des résultats pour l'évaluation de la relation « entouré par »	90
9.1	Comparatif des taux de reconnaissance moyens pour les trois jeux de données	131
9.2	Taux de reconnaissance par classe pour différentes forces	135
10.1	Taux de reconnaissance moyens pour la classification des images de lettrines	146

Introduction

Introduction

Contexte de la thèse

Les mécanismes de perception du système visuel humain nous permettent d'interpréter le monde qui nous entoure et de nous repérer dans l'espace, mais aussi de détecter et de reconnaître différents objets présents dans notre environnement. La simulation de ces capacités complexes par des méthodes automatisées constitue un défi important pour le domaine de la science informatique. Ainsi, les communautés spécialisées en vision par ordinateur et en reconnaissance de formes s'intéressent au développement d'algorithmes permettant d'analyser et d'interpréter des données visuelles telles les images et les vidéos.

En effet, au cours de ces dernières années, nous observons une augmentation considérable de la quantité de ce type de données visuelles dans de nombreux domaines. Ce contexte technologique nouveau est la conséquence de plusieurs facteurs. D'une part, l'augmentation de la puissance de calcul des ordinateurs, ainsi que des capacités de stockage, ont rendu possible le développement de méthodes de traitement permettant de manipuler efficacement des grands volumes de données numériques. D'autre part, le déploiement des réseaux et des protocoles de communication à grande échelle, ainsi que la mise à disposition généralisée des capteurs optiques de plus en plus performants, rendent les données visuelles largement accessibles auprès de diverses communautés. Cette multiplication de la quantité d'images et de vidéos génère alors des ensembles de données complexes et hétérogènes, et il existe alors un besoin important de méthodes informatisées permettant de les analyser et de les interpréter.

Dans ce contexte, l'équipe Systèmes Intelligents de Perception (SIP) du Laboratoire d'Informatique de Paris Descartes (LIPADE) est spécialisée dans diverses problématiques liées à l'analyse d'images et la reconnaissance de formes. Les activités de recherche de l'équipe s'articulent autour de plusieurs axes spécifiques, comme, par exemple, la segmentation d'images guidée par les connaissances, la caractérisation et le suivi d'objets en mouvement, ou encore la description des relations spatiales entre objets d'intérêt dans les images. Ces travaux trouvent des applications dans différents domaines tels que l'imagerie biologique et médicale, l'analyse de documents manuscrits et imprimés, ou encore l'étude d'images satellitaires. Cette thèse s'inscrit donc dans la continuité des travaux de recherche de l'équipe SIP, et se focalise en particulier sur la modélisation des relations spatiales entre objets, ainsi que sur l'exploitation de tels modèles pour des méthodologies de reconnaissance et d'interprétation d'images.

Problématique

Une des difficultés majeures dans le domaine de la vision par ordinateur réside dans la nature même des données à traiter, c'est-à-dire les pixels des images. D'une part, il existe une perte d'information inévitable lors de la retranscription d'une scène du monde réel par un format numérique bidimensionnel, où une portion souvent non négligeable de l'information de profondeur disparaît. D'autre part, les valeurs numériques associées aux pixels correspondent à des données brutes, qui sont particulièrement éloignées des concepts sémantiques généralement utilisés pour désigner ou décrire des objets dans notre langage naturel.

Ce décalage sémantique est d'autant plus important que la définition même d'un objet, dans un sens générique, reste fondamentalement ambiguë et subjective. En effet, selon le contexte envisagé, le terme objet peut désigner par exemple un concept relativement abstrait (l'objet *voiture*, l'objet *tomate*, l'objet *lettre a*, etc.), indépendamment de son apparence au sein des images, ou bien au contraire justement, désigner précisément une entité en fonction de son apparence visuelle (un objet *rond et rouge*, par exemple). Cet aspect du problème est souvent désigné dans la littérature par le terme de *fossé sémantique* [Smeulders *et al.*, 2000].

À plus fortes raisons, la difficulté du problème est accrue dès lors que les objets d'intérêt sont représentés dans des images complexes, comme par exemple lorsque les frontières entre les objets sont difficilement identifiables, ou encore lorsque les objets sont partiellement occultés. De plus, les objets à reconnaître peuvent aussi être eux-mêmes complexes : une voiture peut être considérée comme un objet structuré, lui-même composé de sous-objets (elle comporte des portes, des fenêtres, des roues, etc.). On parle alors du problème de l'échelle. En ce sens, les approches qui consistent à extraire de l'information à partir du contenu des images sont certes particulièrement adaptées aux traitements informatisés, qui permettent une exploitation à la fois automatique et objective des données, mais il reste difficile de mettre en lumière de manière robuste les contenus sémantiques portés par les données visuelles. C'est alors dans cette opposition apparente que résident les principaux enjeux des travaux de recherche dans ces domaines.

Face à ces défis, une stratégie communément employée par la communauté scientifique consiste à bâtir des représentations intermédiaires du contenu des images, à partir de l'extraction de différents types de caractéristiques visuelles (reposant par exemple sur la forme, les contours, ou encore sur des informations de texture) qui aspirent généralement à être robustes par rapport aux diverses variations ou dégradations apparaissant dans les images. Couplées à des algorithmes issus du domaine de l'apprentissage automatique, ces représentations intermédiaires permettent de réduire partiellement le fossé entre les données brutes et les concepts de plus haut niveau nécessaires à leur interprétation. En particulier, depuis le début des années 2000, les approches d'apprentissage par sacs de mots visuels, ou sacs de caractéristiques (inspirées des sacs de mots employés pour l'indexation de documents textuels) ont permis des avancées significatives pour la reconnaissance d'objets et la classification d'images [Everingham *et al.*, 2010]. Plus récemment, les stratégies dites d'apprentissage profond, reposant sur l'utilisation de réseaux de neurones convolutifs, ont également illustré des performances remarquables pour diverses tâches liées à l'interprétation de données visuelles [LeCun *et al.*, 2015].

Si ces approches permettent d'obtenir des résultats satisfaisants dans de nombreuses situations, elles négligent alors bien souvent la structure spatiale entre les objets représentés dans les images. Pourtant, bien qu'appréhendées de manière implicite, les relations spatiales entre les différents objets d'intérêt composant une scène jouent un rôle particulièrement important dans notre perception de celle-ci. Les relations spatiales sont ainsi porteuses d'une riche sémantique, mais sont définies de manière imprécise et ambiguë, et leur exploitation dans des processus de reconnaissance automatisés demeure délicate. Dans cette thèse, nous nous intéressons donc à cette problématique d'exploitation de la richesse de l'information spatiale pour différentes tâches d'analyse et de reconnaissance d'images.

Verrous scientifiques et axes de recherche

Au cours de cette thèse, nous nous attacherons à explorer des approches de modélisation des relations spatiales, partant du constat que celles-ci sont généralement sous-exploitées. En effet, bien qu'elles constituent des caractéristiques essentielles, elles sont parfois difficiles à prendre en compte explicitement et sans ambiguïtés. Dans ce contexte, nos recherches sont motivées par un certain nombre de verrous scientifiques.

Tout d'abord, nous avons constaté que certains types de configurations spatiales entre des objets demeurent difficiles à décrire pour la plupart des modèles existants. C'est notamment le cas lorsqu'il s'agit de distinguer le fait qu'un objet est entouré par un autre, et de manière plus générale, lorsque les objets sont imbriqués les uns dans les autres. Pourtant, la caractérisation de ce type de configurations peut être utile dans de nombreux cadres applicatifs, comme par exemple pour mesurer la complexité d'un réseau vasculaire en imagerie biomédicale, ou bien la densité d'habitation de blocs urbains en télédétection. De manière analogue, la position relative d'objets étant constitués de plusieurs composantes connexes reste également difficile à caractériser.

Ensuite, un second verrou découle du fait que les objets d'intérêt peuvent être présents dans une quantité variable selon les images, et peuvent également apparaître à différents niveaux d'échelle. Par ailleurs, la comparaison de tels ensembles d'objets à travers de nombreuses images induit alors des problématiques de mise en correspondance de nature combinatoire, qui peuvent être particulièrement coûteuses d'un point de vue algorithmique.

Les travaux réalisés dans le cadre de cette thèse se sont donc articulés autour de deux axes de recherche transversaux et complémentaires, en fonction de ces verrous scientifiques. D'une part, le premier axe consiste à étudier la modélisation de ces configurations spatiales dites complexes, où les objets peuvent être imbriqués les uns dans les autres. Notre objectif est de proposer des modèles robustes permettant de décrire quantitativement et de visualiser ce type de configurations. D'autre part, dans le second axe, nous nous intéressons alors à la représentation de la position relative entre plusieurs objets au sein des images. Nous proposons d'étudier la mise en place de descripteurs composites de relations spatiales, avec l'objectif de pouvoir les intégrer efficacement dans des processus d'apprentissage supervisé à partir d'exemples.

Organisation du document

Ce manuscrit de thèse est structuré en trois parties principales, accompagnées de trois annexes. La première partie présente un état de l'art sur la modélisation des relations spatiales en analyse d'images, tandis que les parties suivantes développent respectivement les deux axes de recherches étudiés durant cette thèse. Chaque partie est divisée en plusieurs chapitres. Un résumé synthétique des différentes parties, ainsi que des annexes, est énoncé ci-dessous. La manuscrit est ensuite conclu par un bilan et un résumé des perspectives de recherche envisagées.

Partie I : Relations spatiales en analyse d'images

Nous présentons le contexte général des relations spatiales, et nous illustrons comment ces concepts peuvent être exploités dans les domaines de l'analyse d'images et de la reconnaissance de formes. Nous dressons un état de l'art des différentes approches ayant été proposées dans la littérature pour modéliser les relations spatiales entre objets, en mettant l'accent sur la dualité qui existe entre le concept de relation spatiale et celui de position relative. Nous illustrons également la nécessité de mettre en place des modèles spécifiques pour certains types de configurations spatiales complexes, et nous examinons plusieurs approches récentes qui ont proposé des stratégies d'apprentissage de relations spatiales composites. Finalement, suite à cet état de l'art, nous énonçons le positionnement général de ces travaux de thèse.

Partie II : Modélisation de relations spatiales complexes

Nous nous intéressons à la modélisation de configurations spatiales complexes, et en particulier aux situations où les objets peuvent être imbriqués les uns dans les autres. Cette notion est formalisée par deux nouvelles relations spatiales, que nous nommons enlacement et entrelacement. Nous proposons alors un modèle générique qui permet de décrire et d'évaluer l'enlacement et l'entrelacement entre objets bidimensionnels, avec une granularité directionnelle. Ce modèle permet de distinguer efficacement différents types de configurations spatiales complexes, et présente des propriétés d'invariance bien souvent utiles en reconnaissance de formes. Cette approche est validée expérimentalement dans différents contextes applicatifs, tels que la caractérisation de vaisseaux sanguins en imagerie ophthalmologique, la reconnaissance de lettrines décoratives pour l'analyse de documents anciens, et la classification d'objets géographiques urbains en télé-détection. Par ailleurs, nous proposons également une stratégie d'évaluation et de visualisation de ces nouvelles relations spatiales, à partir de la définition de paysages flous directionnels.

Partie III : Apprentissage de relations spatiales composites

Nous proposons un cadre général de description et d'apprentissage de relations spatiales composites entre objets d'intérêt représentés dans des images. Tout d'abord, nous présentons une stratégie de description du contenu d'une image à partir d'une représentation structurelle, permettant de caractériser les relations spatiales et les formes des objets d'intérêt. Puis, inspirés des approches par sacs de caractéristiques visuelles communément employées en vision par

Introduction

ordinateur, nous proposons une approche permettant d'apprendre des vocabulaires de configurations spatiales apparaissant à travers des images, potentiellement à différents niveaux d'échelle. Les résultats expérimentaux obtenus pour plusieurs bases d'images composées d'objets structurés permettent d'illustrer l'intérêt de cette approche pour la reconnaissance et la classification d'images. Nous présentons également comment intégrer les descripteurs d'enlacement au sein de cette nouvelle stratégie d'apprentissage, dans le contexte applicatif de l'analyse d'images de lettrines décoratives.

Annexes

L'[Annexe A](#) récapitule les différentes publications scientifiques produites dans le cadre de ces travaux de thèse. Puis, l'[Annexe B](#) présente les démonstrations mathématiques des propriétés d'invariance du descripteur d'enlacement (présenté dans le [Chapitre 4](#)). Enfin, l'[Annexe C](#) dresse un aperçu de l'environnement technologique utilisé pour les développements logiciels réalisés au cours de cette thèse.

Partie I

Relations spatiales en analyse d'images

Résumé

Nous présentons le contexte général des relations spatiales, et nous illustrons comment ces concepts peuvent être exploités dans les domaines de l'analyse d'images et de la reconnaissance de formes. Nous dressons un état de l'art des différentes approches ayant été proposées dans la littérature pour modéliser les relations spatiales entre objets, en mettant l'accent sur la dualité qui existe entre le concept de relation spatiale et celui de position relative. Nous illustrons également la nécessité de mettre en place des modèles spécifiques pour certains types de configurations spatiales complexes, et nous examinons plusieurs approches récentes qui ont proposé des stratégies d'apprentissage de relations spatiales composites. Finalement, suite à cet état de l'art, nous énonçons le positionnement général de ces travaux de thèse.

Chapitre 1

Relations spatiales : définitions et méthodes classiques

Sommaire

1.1 Relations spatiales élémentaires	12
1.1.1 Propositions linguistiques spatiales	12
1.1.2 Relations spatiales de Freeman	12
1.1.3 Familles de relations spatiales	13
1.2 Méthodes classiques et limitations	14
1.2.1 Approches qualitatives	14
1.2.2 Approches quantitatives	17
1.2.3 Limitations et synthèse	19

Dans ce chapitre, nous proposons une introduction générale au domaine des relations spatiales telles qu'elles seront envisagées dans cette thèse. En [Section 1.1](#), la notion de relation spatiale est tout d'abord introduite sous l'angle de la perception humaine, en soulignant ses origines dans les domaines de la linguistique et de la psychologie. Nous présentons également les différentes relations élémentaires qui ont été proposées par Freeman [[Freeman, 1975](#)] pour leur modélisation au sein de systèmes informatisés, qui peuvent être regroupées en trois grandes familles : relations directionnelles, topologiques et métriques. Ensuite, en [Section 1.2](#), nous présentons un historique synthétique des méthodes qui ont été classiquement proposées dans la littérature pour modéliser les relations spatiales. On retrouve généralement deux grandes familles d'approches : les approches qualitatives et les approches quantitatives. Nous concluons alors le chapitre par une synthèse de ces approches classiques.

1.1 Relations spatiales élémentaires

1.1.1 Propositions linguistiques spatiales

Les êtres humains sont capables de se repérer dans l'espace grâce à différents concepts inhérents à leurs mécanismes de perception. Bien que ces facultés existent de manière innée, notre interprétation de la position des objets est d'abord le fruit d'un apprentissage empirique. Ainsi, c'est par exemple en observant et en essayant d'attraper différents objets plus ou moins éloignés de lui, qu'un nouveau-né commence à faire l'expérience des distances ou des directions. Cependant, notre appréhension du positionnement des objets ne se limite pas uniquement à cette expérience sensorielle immédiate. En effet, nous sommes également capables de conceptualiser ces notions spatiales au-delà des objets qui sont directement dans notre champ de vision. Les travaux de [Kuipers, 1978] en sciences cognitives envisagent ces capacités humaines sous la forme de « cartes mentales » qui nous permettent de nous orienter dans l'espace. C'est grâce à ces capacités d'abstraction mentales sur l'environnement qu'un individu peut, par exemple, se représenter le trajet jusqu'à son domicile ou jusqu'à son prochain lieu de rendez-vous.

Par ailleurs, dans la vie de tous les jours, nous sommes également en mesure de décrire et d'interpréter la localisation de différents objets, à l'aide de représentations symboliques qui sont propres au langage. Par exemple, « à gauche », « à l'intérieur » ou « loin » constituent des propositions linguistiques, ou *relations spatiales*, communément employées pour décrire le positionnement spatial d'objets. De nombreuses études ont été proposées pour analyser l'impact de la linguistique dans notre psychologie de la perception, en fonction des différentes langues. On pourra notamment citer les travaux de [Landau et Jackendoff, 1993] sur les mécanismes linguistiques pour décrire des lieux ou des objets, ou encore la thèse de Laure Vieu [Vieu, 1991] sur la sémantique des relations spatiales en langue française.

1.1.2 Relations spatiales de Freeman

Parallèlement, au milieu des années 1970, Freeman a tracé les premières lignes directrices vers un système informatisé permettant d'évaluer et de décrire le positionnement spatial d'objets représentés dans le plan [Freeman, 1975] (c'est-à-dire, par exemple, dans des images numériques). Dans ses travaux, Freeman proposa notamment une liste de relations spatiales dites *élémentaires*. Celles-ci sont représentées par les treize propositions linguistiques suivantes :

- LEFT OF : « à gauche de » ;
- RIGHT OF : « à droite de » ;
- BESIDE (*alongside, next to*) : « à côté de » ;
- ABOVE : « au-dessus de » ;
- BELOW : « en-dessous de » ;
- BEHIND (*in back of*) : « derrière » ;
- IN FRONT OF : « devant » ;
- NEAR (*close to, next to*) : « proche de » ;
- FAR : « loin de » ;

1.1. Relations spatiales élémentaires

- TOUCHING : « *touche* » ;
- BETWEEN : « *entre* » ;
- INSIDE (*within*) : « *à l'intérieur de* » ;
- OUTSIDE : « *à l'extérieur de* ».

Par exemple, pour deux objets A et B , on peut représenter la proposition linguistique « *l'objet A est au-dessus de l'objet B* » par la relation spatiale ABOVE, instanciée pour l'objet A par rapport à l'objet B . Dans ce cas, B est l'*objet de référence*, et A est l'*objet cible* ou *argument*. On notera que toutes les relations proposées sont binaires (c'est-à-dire définies pour deux objets), à l'exception de la relation BETWEEN qui est ternaire. Par ailleurs, nous pouvons également constater que les deux relations BEHIND et IN FRONT OF ne sont pas directement applicables dans le cas d'objets représentés dans un plan bidimensionnel.

Selon Freeman, cette liste de relations spatiales élémentaires se veut exhaustive, dans le sens où celles-ci pourraient être combinées pour obtenir d'autres relations potentiellement plus complexes. Par exemple, la relation spatiale « *entouré par* » pourrait être obtenue à partir de la combinaison des quatre relations cardinales « *à gauche* », « *à droite* », « *au-dessus* » et « *en-dessous* ».

1.1.3 Familles de relations spatiales

À partir de cette liste proposée par Freeman, on peut généralement distinguer trois grandes familles de relations spatiales :

- Les relations spatiales directionnelles, qui font appel aux angles et aux orientations (par exemple LEFT OF, RIGHT OF ou ABOVE) ;
- Les relations spatiales topologiques, où des notions de théorie des ensembles, d'adjacence et de voisinage interviennent (par exemple INSIDE, OUTSIDE ou TOUCHING) ;
- Les relations spatiales de distance, qui font intervenir les notions de proximité ou d'éloignement entre les objets (par exemple NEAR et FAR).

Par ailleurs, Freeman a également mis en évidence que les relations mathématiques classiques (c'est-à-dire selon une logique binaire) ne permettent pas de modéliser correctement le comportement des relations spatiales, même lorsqu'il s'agit d'étudier des objets géométriques élémentaires. Il préconisa ainsi l'emploi de la théorie des ensembles flous [Zadeh, 1965]. Formellement, un ensemble flou A défini sur un espace \mathcal{S} est défini par sa fonction d'appartenance μ_A telle que $\forall x \in \mathcal{S}, \mu_A(x) \in [0, 1]$. Ce type de modélisation permet de prendre en compte de manière graduelle les aspects imprécis inhérents aux relations spatiales (par exemple le passage progressif de la relation NEAR à la relation FAR).

Afin d'illustrer cet aspect, la [Figure 1.1](#) présente quelques exemples de configurations spatiales entre des objets simples, où l'on peut visualiser cette imprécision liée aux relations spatiales. Nous pouvons notamment constater que la taille et la forme des objets interviennent de manière importante dans l'appréhension des relations spatiales directionnelles, en particulier lorsque les objets sont proches. Ces paramètres interviennent d'ailleurs également pour les autres catégories de relations spatiales, notamment pour les relations topologiques.

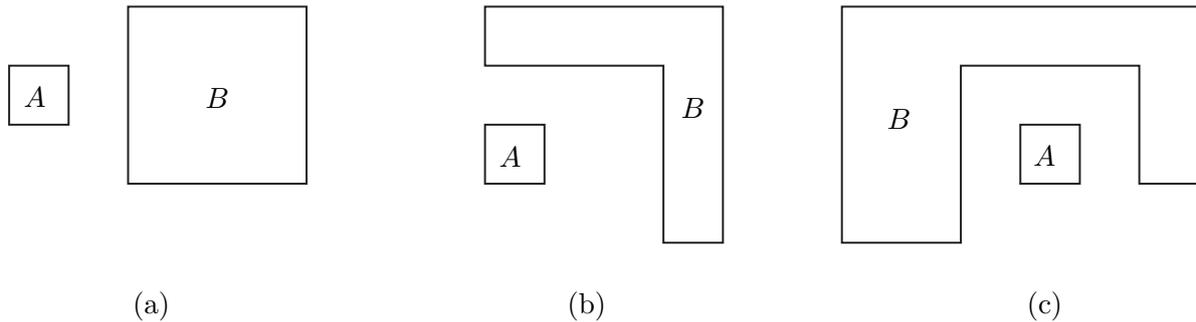


Figure 1.1 – Illustration de l’ambiguïté du positionnement spatial entre deux objets. Dans la configuration (a), l’objet A peut être considéré comme à gauche de l’objet B . Dans la configuration (b), A peut toujours être considéré comme étant à gauche de B , mais aussi en-dessous. La configuration (c) est plus ambiguë, puisqu’une grande partie de B est maintenant située à gauche de A . Illustration reprise de [Dehak, 2002], d’après [Rosenfeld et Kak, 1982].

Nous avons donc présenté le caractère particulièrement imprécis et ambigu de la description de la position relative d’objets. Nous avons également montré que différentes propositions linguistiques, regroupées en familles de relations, peuvent servir de support à de telles descriptions. Dans la suite de ce chapitre, nous dressons alors un historique synthétique des approches proposées par différentes communautés spécialisées en informatique, qui s’inscrivent généralement dans la continuité de cette première systématisation des relations spatiales par Freeman.

1.2 Méthodes classiques et limitations

Nous présentons ici de manière synthétique un historique des deux grandes familles d’approches qui ont été proposées pour modéliser le positionnement spatial d’objets représentés dans le plan. Dans la littérature, on retrouve traditionnellement des approches plutôt d’ordre qualitatives, qui utilisent généralement des représentations symboliques, et des approches quantitatives, qui cherchent à mesurer l’imprécision des relations spatiales au sein de la modélisation. Nous soulignons cependant le fait que la distinction entre ces deux familles d’approches n’est pas toujours bien définie, et que leurs points de vue peuvent bien souvent être complémentaires.

1.2.1 Approches qualitatives

Les approches qualitatives pour modéliser les relations spatiales ont majoritairement été proposées au sein de communautés spécialisées en *raisonnement qualitatif* (et plus particulièrement en raisonnement qualitatif *spatial* [Freksa, 1991 ; Cohn et Hazarika, 2001]), qui constitue une branche de l’intelligence artificielle. Ces approches ont généralement pour objectif de représenter des données en ayant recours à des valeurs symboliques, ou à des espaces discrets partiellement ordonnés. L’utilisation de telles primitives, qui pourraient au premier abord paraître comme simplificatrices, s’inscrit en réalité dans une démarche de raisonnement de plus haut niveau sur

1.2. Méthodes classiques et limitations

les données (que l'on pourrait notamment associer à une volonté de réduction du fossé sémantique). Les approches s'intéressent alors aux représentations symboliques les plus appropriées pour effectuer des raisonnements sur des phénomènes physiques, initialement mesurés de manière quantitative. Par ailleurs, certains travaux ont proposé de combiner ces raisonnements qualitatifs avec des mécanismes issus de la logique floue [Dutta, 1991]. Notamment, ces approches par représentations symboliques de données spatiales sont souvent employées dans les Systèmes d'Information Géographique (SIG), et de manière générale pour l'analyse d'images de télédétection [Forestier *et al.*, 2012]. D'autres domaines d'application sont également possibles, comme la mise en place de raisonnements automatisés au sein de systèmes experts.

Plusieurs travaux sur la représentation qualitative des relations spatiales se sont inspirés d'études sur le raisonnement temporel [Vieu, 1997], en particulier sur l'algèbre des intervalles temporels d'Allen [Allen, 1983]. Le principe consiste à étendre les raisonnements temporels unidimensionnels à des entités de dimension deux (par exemple, entre autres stratégies, par projection des objets sur les axes horizontaux et verticaux). Nous verrons d'ailleurs par la suite que ce principe a également inspiré certaines approches quantitatives.

En outre, les approches qualitatives ont aussi été beaucoup employées pour décrire les relations spatiales topologiques entre des objets. De manière générale, ces travaux s'inscrivent alors souvent dans le domaine dit de la méréotopologie [Smith, 1996], qui vise à réunir les concepts liés aux ontologies, et ceux liés à la topologie. Plusieurs articles ont proposé des états de l'art sur les différentes approches proposées dans ce domaine, comme par exemple la récente synthèse de [Chen *et al.*, 2015]. Ci-après, nous proposons un aperçu des principales approches proposées.

Dans [Egenhofer et Herring, 1990] et [Egenhofer et Franzosa, 1991], un formalisme plus étendu que les relations élémentaires de Freeman est proposé pour décrire les relations topologiques entre des objets. Dans ce formalisme, un objet A est considéré à partir de sa zone intérieure A° , son contour ∂A ainsi que sa zone extérieure A^- . Pour deux objets A et B , les différentes intersections possibles de ces trois régions sont alors considérées. La configuration topologique entre les deux objets est alors représentée par neuf valeurs binaires, représentées synthétiquement par une matrice 3×3 :

$$R(A, B) = \begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{bmatrix} \quad (1.1)$$

Ce formalisme pour décrire la configuration topologique de deux objets, sous la forme de cette matrice de valeurs qualitatives, est alors appelé modèle *9-intersections*.

Parallèlement, les auteurs de [Randell *et al.*, 1992] ont proposé un autre ensemble de relations topologiques qui peuvent être organisées de manière hiérarchique, sous la forme d'un treillis. Les différentes configurations topologiques possibles de deux objets sont ainsi modélisées par une hiérarchie descendante de relations, dont les relations les plus spécifiques sont mutuellement exclusives et collectivement exhaustives (on parle de relations *JEPD*, pour *jointly exhaustive and pairwise disjoint* en anglais). Ces relations, au nombre de huit, sont illustrées schématiquement

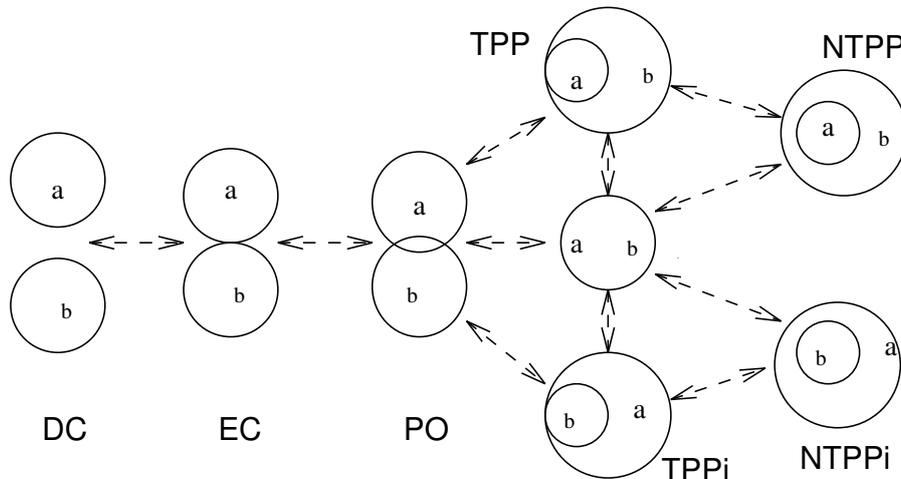


Figure 1.2 – Illustration des relations spatiales topologiques RCC8. Les acronymes du schéma sont les suivants : *DC* pour *disconnected*, *EC* pour *externally connected*, *PO* pour *partially overlaps*, *TPP* et *TPPi* pour *tangential proper part*, *NTPP* et *NTPPi* pour *nontangential proper part*, et *EQ* pour *equals* (acronyme non affiché sur le schéma, correspondant au cas où $a = b$). Schéma repris de [Randell *et al.*, 1992].

en Figure 1.2. Ces travaux furent alors popularisés sous le nom des relations topologiques *RCC8* (pour *Region Connection Calculus*). Ils ont plus tard été étendus afin de distinguer plus de configurations topologiques, notamment dans le cas où les objets peuvent présenter des concavités [Cohn *et al.*, 1997]. Dans cette extension, de nouvelles relations sont introduites pour détecter une situation d’imbrication d’objets (mais sans chevauchement des objets). Plus formellement, il s’agit des configurations spatiales où les objets peuvent être concaves et où l’intersection de leurs enveloppes convexes n’est pas nulle. Pour décrire qualitativement ces situations, ce nouveau modèle, nommé *RCC23*, est alors composé de 23 relations : les 8 relations initiales du modèle *RCC8*, auxquelles s’ajoutent 15 nouvelles relations pour l’imbrication d’objets. Ce modèle a notamment été utilisé sous la forme de graphes relationnels attribués par [Inglada et Michel, 2009] pour la détection d’objets structurés au sein d’images de télédétection, ou encore pour l’analyse d’images histologiques dans [Randell *et al.*, 2013].

En somme, les différentes approches évoquées proposent donc des représentations symboliques de l’information spatiale, principalement pour l’étude de configurations qui font intervenir la topologie entre les objets. Pour la plupart, elles étendent donc les relations élémentaires initialement proposées par Freeman. Néanmoins, bien que ces approches qualitatives soient en mesure de détecter un large panel de configurations, elles permettent difficilement de modéliser le caractère graduel de ces relations spatiales sous la forme de mesures.

1.2.2 Approches quantitatives

Si les approches qualitatives permettent de représenter les relations spatiales de manière symbolique, les méthodes quantitatives de représentation des relations spatiales ont généralement recours, quant à elles, à des concepts issus de la théorie des ensembles flous. En effet, comme cela a pu être mis en évidence par [Freeman, 1975], les raisonnements flous semblent particulièrement adaptés pour modéliser l'imprécision inhérente aux relations spatiales. Ces mesures d'imprécision peuvent alors être employées pour quantifier, de manière graduelle, les relations spatiales entre les objets. Nous ferons ainsi un résumé historique des principales méthodes quantitatives qui ont été proposées. Certaines de ces méthodes seront ensuite présentées plus en détail dans le chapitre suivant.

Pour mesurer le positionnement spatial entre deux objets de manière quantitative, une des approches les plus intuitives consiste à procéder à une simplification de la forme des objets, afin de réduire l'ambiguïté dans leur positionnement. Par exemple, dans [Peuquet et Ci-Xiang, 1987], les objets sont ramenés à leurs barycentres respectifs. Les auteurs distinguent différentes situations selon l'éloignement des objets. En fonction de ces différentes situations, les relations spatiales directionnelles sont estimées soit à partir d'un cône de vision depuis le barycentre (dans le cas où les objets sont éloignés), soit par une demi-droite orientée dans la direction considérée (lorsque les objets sont proches). Huit directions discrètes sont considérées : les quatre relations cardinales, ainsi que les diagonales correspondantes. Bien que particulièrement simple à mettre en œuvre, cette approche ne permet pas de distinguer les cas où les objets sont concaves, ou bien lorsqu'il présentent des élongations significatives dans certaines directions. De plus, l'approche est sensible au choix de l'objet de référence (par rapport à l'objet cible), et peut alors ne pas respecter la notion d'inverse sémantique (si A est à gauche de B , alors B est à droite de A).

Une autre manière communément employée pour simplifier la forme des objets consiste à les projeter le long des axes horizontaux et verticaux. Ils deviennent alors semblables à des entités unidimensionnelles, et des traitements numériques analogues aux intervalles temporels d'Allen peuvent être effectués [Allen, 1983]. C'est par exemple le cas dans les travaux proposés par [Kóczy, 1988], où les projections d'objets flous sur l'axe des abscisses sont utilisées pour indiquer si un objet est à gauche ou à droite d'un autre (et respectivement sur l'axe des ordonnées pour les directions verticales), en fonction de l'aire des objets le long de ces projections. Cependant, cette réduction aux axes orthogonaux ne permet pas de prendre en compte des objets ayant des formes arbitraires, et se limite aux seules quatre directions cardinales.

Le principe des *2D-strings*, initialement proposé par [Chang et al., 1987] repose également sur cette représentation simplifiée des objets. Dans ce type d'approche structurale, la position relative des différents objets représentés dans une image est décrite par une chaîne de caractères. Par exemple, la chaîne de caractères $s = (A < B = C < D)$ peut indiquer que A est à gauche de B et C , qui sont eux-même à gauche de D . Bien que présentant l'avantage de pouvoir décrire la configuration spatiale de plusieurs objets simultanément, cette approche a également recouru à une simplification grossière de la forme des objets, en les réduisant à leurs seuls barycentres et en projetant ces derniers le long des axes orthogonaux. Elle est donc encline à de fortes

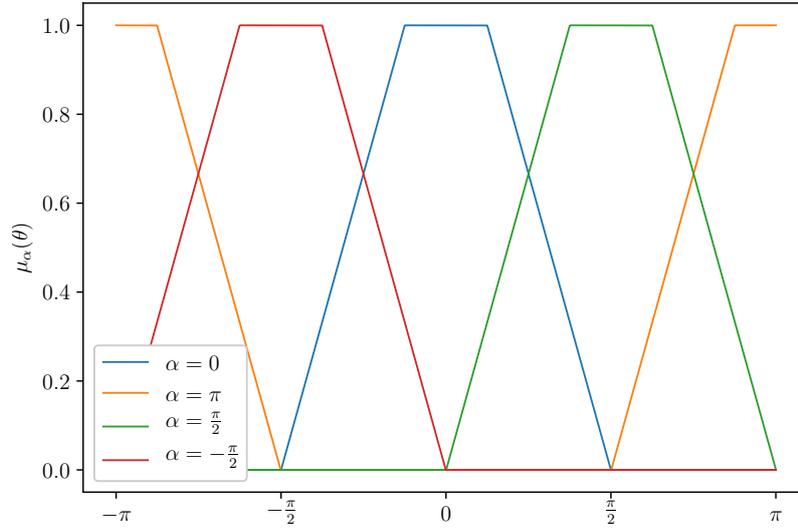


Figure 1.3 – Exemples d’ensembles flous trapézoïdaux pouvant être utilisés pour modéliser les relations spatiales cardinales.

ambiguïtés lors de la description, et ne peut prendre en compte que les directions cardinales. Ces travaux furent par la suite étendus par [Lee et Hsu, 1992] et [Chang et Lee, 1995]. Les objets sont alors représentés en fonction de leurs extrémités le long des axes orthogonaux, en considérant la boîte englobante des objets. Ces extensions permettent d’améliorer la capacité de description du modèle dans certains cas, mais se limite toujours aux directions cardinales, et ne peut pas gérer efficacement les objets présentant des concavités. De plus, la représentation des configurations spatiales sous la forme de chaînes de caractères induit alors des problématiques de mise en correspondance et de comparaison de telles chaînes, qui doivent alors prendre en compte la sémantique sous-jacente des relations spatiales.

Les auteurs de [Krishnapuram *et al.*, 1993] ont adopté un point de vue différent. Dans ces travaux, les objets sont considérés comme des ensembles de points discrets, et tous les couples de points entre les deux objets sont considérés. Un formalisme flou est alors proposé pour modéliser les différentes relations élémentaires de Freeman. Notamment, les relations spatiales directionnelles sont évaluées en mesurant l’angle θ formé par chaque couple de points avec l’axe des abscisses. Par exemple, la relation « à gauche de » est mesurée à l’aide d’un ensemble flou trapézoïdal défini de la manière suivante :

$$\mu_{\text{LEFT}}(\theta) = \begin{cases} 0 & \text{si } |\theta| > \frac{\pi}{2}, \\ \frac{(\pi/2)-|\theta|}{(\pi/2)(1-a)} & \text{si } a\frac{\pi}{2} \leq |\theta| \leq \frac{\pi}{2}, \\ 1 & \text{si } |\theta| < a\frac{\pi}{2}. \end{cases} \quad (1.2)$$

Ici, le paramètre $a \in [0, 1]$ permet de contrôler l’aspect graduel de la relation spatiale. Cette formulation peut alors être généralisée pour les quatre directions cardinales, comme illustré par

1.2. Méthodes classiques et limitations

exemple en [Figure 1.3](#). Les évaluations obtenues pour chaque couple de points $(a_i, b_j) \in A \times B$ sont alors agrégées selon des opérateurs d’agrégation flous proposés par [\[Dubois et Jaulent, 1987\]](#). Des formulations similaires sont présentées pour les autres relations spatiales directionnelles, ainsi que par exemple pour les relations de distance (afin de modéliser de manière graduelle le fait que des objets soient proches ou éloignés). Cette approche quantitative présente l’avantage de prendre en compte tous les points des objets, sans avoir recours à des simplifications grossières de leurs formes. Cependant, elle reste coûteuse d’un point de vue algorithmique, et ceci d’autant plus lorsqu’il s’agit de traiter des objets eux-mêmes définis comme des ensembles flous.

Selon un point de vue similaire, la méthode proposée par [\[Miyajima et Ralescu, 1994\]](#) se focalise sur les relations directionnelles. La méthode propose de calculer l’angle $\angle(a_i \vec{b}_j, \vec{\mu}_0)$ formé par les couples de points $(a_i, b_j) \in A \times B$ et le vecteur unitaire $\vec{\mu}_\theta = [\cos \theta, \sin \theta]$ orienté selon l’axe horizontal. Ces angles sont alors accumulés au sein d’un histogramme directionnel, nommé *histogramme d’angles*, qui constitue alors une représentation synthétique de la position de A par rapport à B . Ainsi, plutôt que de considérer uniquement les directions cardinales comme la plupart des approches présentées précédemment, cette méthode permet de considérer toutes les directions dans l’intervalle $[0, 2\pi]$. Ensuite, une stratégie de mise en correspondance est appliquée : l’histogramme est comparé avec différents profils flous correspondant aux relations spatiales directionnelles élémentaires, afin d’obtenir des évaluations du positionnement de A par rapport à B (sous la forme d’un degré de validité, indiquant par exemple dans quelle mesure A est « à gauche » de B). D’un point de vue algorithmique, le principe de cette méthode est similaire à celle de [\[Krishnapuram et al., 1993\]](#) présentée juste avant. Elle présente donc également l’avantage de prendre explicitement en compte la forme des objets, puisque tous les points sont considérés. Cependant, le temps de calcul peut aussi vite devenir prohibitif. Par ailleurs, l’histogramme d’angles est particulièrement sensible à la discrétisation des objets, notamment selon le pas de directions choisi pour l’intervalle $[0, 2\pi]$.

Nous avons ainsi présenté différentes approches qualitatives et quantitatives pour modéliser les relations spatiales. Ces deux familles d’approches peuvent généralement être envisagées de manière complémentaire, certaines approches s’inspirant les unes des autres (en particulier celles reposant sur les intervalles d’Allen).

1.2.3 Limitations et synthèse

Dans ce chapitre, nous avons proposé une introduction à la modélisation des relations spatiales entre des couples d’objets représentés dans le plan. Tout d’abord, nous avons mis en relief les origines sémantiques et conceptuelles des relations spatiales dans des domaines tels que les sciences cognitives, la psychologie ou encore la linguistique. Nous avons ensuite présenté un historique synthétique des modèles informatisés traditionnellement proposés, qui se répartissent généralement selon des points de vue qualitatifs ou quantitatifs. La présentation de ces différentes approches a notamment permis d’illustrer la nécessité de prendre en compte la forme des objets pour modéliser correctement les imprécisions liées aux relations spatiales. En particulier,

nous avons pu constater que lorsque les objets présentent des concavités, et que l'intersection de leurs enveloppes convexes n'est pas nulle, alors les configurations spatiales sont particulièrement chargées d'ambiguïtés.

Dans le cadre de cette thèse, nos travaux de recherche approfondiront principalement l'étude de modèles quantitatifs de relations spatiales, car ceux-ci sont plus naturellement utilisables dans des systèmes de reconnaissance et d'apprentissage à partir de bases d'images, en particulier lorsque les objets d'intérêt peuvent être complexes, ou constitués de plusieurs composantes connexes. Néanmoins, nous soulignons également la complémentarité des approches qualitatives avec les représentations quantitatives, qui peuvent être particulièrement intéressantes dans certains contextes applicatifs. En ce sens, dans le chapitre suivant, nous présenterons de manière plus approfondie des approches quantitatives, qui prennent explicitement en compte la forme des objets, tout en se nourrissant pour certaines de représentations qualitatives.

Chapitre 2

Dualité entre relations spatiales et position relative

Sommaire

2.1	Relations spatiales et position relative	21
2.2	Descripteurs de position relative	22
2.2.1	Relations spatiales directionnelles	23
2.2.2	Combinaisons avec d'autres familles de relations spatiales	24
2.3	Paysages de relations spatiales	26
2.3.1	Paysages flous morphologiques	27
2.3.2	Champs de forces	28
2.4	Synthèse	29

Dans ce chapitre, nous présentons de manière plus approfondie les différentes méthodes quantitatives qui ont été proposées pour modéliser les relations spatiales entre des objets représentés dans des images. Tout d'abord, dans la [Section 2.1](#), nous présentons en une catégorisation des méthodes qui consiste à distinguer le concept de relation spatiale de celui de position relative. Cette catégorisation est notamment inspirée du point de vue présenté dans les travaux de [[Matsakis *et al.*, 2010](#)]. Nous présentons alors, d'une part les différents descripteurs de position relative proposés dans la littérature en [Section 2.2](#), et d'autre part les approches par paysages de relations spatiales en [Section 2.3](#). Enfin, le chapitre est conclu par une synthèse de ces différentes approches en [Section 2.4](#).

2.1 Relations spatiales et position relative

Historiquement, la modélisation des relations spatiales entre objets représentés dans des images a été souvent envisagée à partir de propositions linguistiques bien définies. Par exemple, étant donné deux objets A et B , les approches classiques cherchent généralement à modéliser

dans quelle mesure A est « à gauche » de B . Il s'agit d'évaluer une relation spatiale. Selon cette logique, il est alors possible de représenter plus précisément une configuration spatiale en combinant les évaluations de différentes relations élémentaires. Par exemple, on peut évaluer les relations spatiales « à gauche » et « au dessus » pour indiquer si un objet se trouve au nord-ouest d'un autre. Nous remarquons alors que des évaluations de différentes relations spatiales sont souvent nécessaires pour *décrire* la position d'un objet par rapport à un autre. C'est ainsi qu'apparaît le principe de dualité que nous souhaitons mettre en avant dans ce chapitre : une description de la position relative entre objets fait intervenir différentes relations spatiales, tandis que les relations spatiales entre objets servent à décrire leur position relative.

Par ailleurs, nous avons vu dans le chapitre précédent que des évaluations de relations spatiales élémentaires peuvent être chargées d'ambiguïtés, en particulier lorsque la forme des objets peut être arbitrairement complexe. Dans ce cas, une description linguistique en termes de combinaisons de relations spatiales élémentaires peut ne pas être suffisante pour certaines applications qui nécessitent une description précise de la position relative d'objets. Ceci suggère l'idée de décrire la position d'un objet par rapport à un autre par le biais d'une représentation quantitative à part entière, de la même manière que différents types de descripteurs sont proposés en reconnaissance de formes, comme par exemple pour caractériser la forme géométrique d'un objet [Zhang et Lu, 2004]. Nous parlons alors d'un *descripteur de position relative*.

En somme, la modélisation des relations spatiales en analyse d'images peut généralement être envisagée selon deux points de vue transverses et complémentaires. Le premier point de vue consiste à prendre pour repère initial la *relation spatiale*, dans le but de fournir des évaluations de cette relation pour différentes instances de configurations spatiales. Plusieurs évaluations peuvent alors être combinées pour obtenir une vision d'ensemble de la situation. Le second point de vue consiste, lui, à prendre pour repère initial les *objets*, et se pose la question de leur *position relative*. Celle-ci est représentée au sein d'un descripteur. Il est alors possible, si nécessaire, d'extraire des évaluations de différentes relations spatiales à partir de ce descripteur.

Notre objectif ici n'est pas de mettre en concurrence ces différents points de vue, mais, au contraire, de souligner l'importance de cette distinction conceptuelle, qui illustre d'autant plus la nature complexe et la richesse de l'information spatiale. Selon le type d'application considéré, certaines approches peuvent être alors plus ou moins adaptées. Dans la suite de ce chapitre, nous illustrons ce principe de dualité en présentant, d'une part des approches par descripteurs de position relative, et d'autre part des approches par paysages de relations spatiales, qui permettent de représenter une relation spatiale directement dans l'espace image. De plus amples détails sur certaines de ces méthodes peuvent être trouvés dans différents états de l'art proposés dans la littérature [Bloch et Ralescu, 2003 ; Bloch, 2005 ; Naem et Matsakis, 2015].

2.2 Descripteurs de position relative

Récemment, un état de l'art très complet des différents descripteurs de position relative qui ont été proposés dans la littérature a été présenté dans [Naem et Matsakis, 2015]. Ce travail offre un bon panorama des avantages et inconvénients des différents types de descripteurs, ainsi

2.2. Descripteurs de position relative

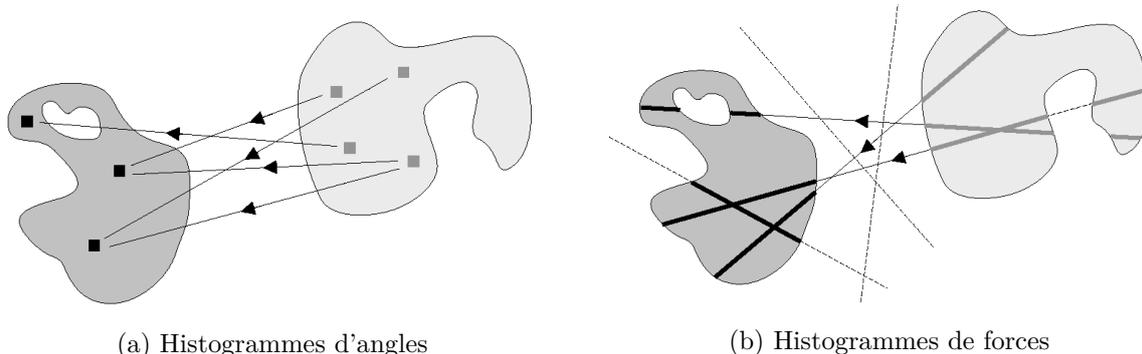


Figure 2.1 – Illustration du principe de calcul (a) des histogrammes d’angles [Miyajima et Ralescu, 1994] et (b) des histogrammes de forces [Matsakis et Wendling, 1999]. Schémas repris de [Matsakis et Wendling, 1999].

que des détails sur leurs propriétés d’invariance et leurs complexités algorithmiques. Dans ce qui suit, nous proposons un descriptif synthétique de quelques uns de ces descripteurs.

2.2.1 Relations spatiales directionnelles

L’un des premiers descripteurs de position relative proposé dans la littérature fut l’histogramme d’angles [Miyajima et Ralescu, 1994], déjà brièvement introduit dans le chapitre précédent. Il permet de décrire la position d’un objet A par rapport à un objet B , en considérant l’angle $\angle(a_i\vec{b}_j, \vec{\mu}_0)$ formé par chacun des couples de points $(a_i, b_j) \in A \times B$ des deux objets avec le vecteur unitaire orienté selon l’axe des abscisses. Ces angles sont accumulés au sein d’un histogramme directionnel \mathcal{H}_{AB} où, pour un ensemble de directions discrètes $\theta \in [0, 2\pi]$, la valeur de $\mathcal{H}_{AB}(\theta)$ correspond au nombre d’occurrences de couples de points (a_i, b_j) tels que $\angle(a_i\vec{b}_j, \vec{\mu}_0) = \theta$. Dans ces travaux, l’histogramme d’angles n’est pas utilisé à proprement parler comme un descripteur, mais plutôt comme une représentation intermédiaire (une fois normalisé par le nombre d’occurrences maximal, l’histogramme peut être assimilé à un ensemble flou) servant à extraire des évaluations floues des différentes relations directionnelles. Cette approche présente l’avantage de considérer les objets dans leur ensemble, ainsi que sur tout un panel de directions (et non pas seulement les directions cardinales). Elle modélise de façon synthétique la position d’un objet par rapport à un autre. Cependant, l’algorithme permettant de calculer l’histogramme d’angles est d’une complexité importante, de l’ordre de $\mathcal{O}(n^2)$ où n désigne le nombre de points (ou pixels) composant les objets. De plus, comme nous l’avons indiqué dans le chapitre précédent, l’approche est particulièrement sensible à la discrétisation des objets, par rapport aux différentes directions considérées.

Afin de pallier ces limitations, les auteurs de [Matsakis et Wendling, 1999] ont proposé une généralisation de l’histogramme d’angles, alors nommé *histogramme de forces* et noté \mathcal{F}_{AB} . Cette méthode préconise de manipuler les objets non plus comme des ensemble de points, mais comme des sections longitudinales formant des segments de droites (d’une manière analogue au

formalisme de la transformée de Radon). Une illustration de cette différence de manipulation des objets pour les histogrammes d'angles et de forces est présentée en [Figure 2.1](#). Cette stratégie algorithmique permet de grandement réduire le temps de calcul de l'histogramme : il passe d'une complexité algorithmique en $\mathcal{O}(n^2)$ à un calcul de l'ordre de $\mathcal{O}(k n \sqrt{n})$ où n représente le nombre de points des objets, et k correspond au nombre de directions discrètes considérées dans l'intervalle $[0, 2\pi]$. Cette approche présente également l'avantage d'être moins sensible aux erreurs de discrétisation, tout en considérant toutes les directions de la même façon (on dit alors que l'histogramme de forces est isotrope). La granularité de description de la configuration spatiale peut être contrôlée en choisissant le nombre de directions discrètes. Par ailleurs, le modèle permet également de décrire le positionnement relatif d'objets définis de manière vectorielle, sous la forme de polygones.

Le calcul de l'histogramme de forces repose également sur la définition d'une force d'attraction élémentaire entre couples de points. On considère alors l'intégrale de cette force le long de toutes les droites parallèles traversant les objets, dans une direction θ . Pour deux points $a, b \in \mathbb{R}^2$, cette force d'attraction est définie en fonction de leur distance $d = \|\vec{ab}\|_2$ par $\varphi_r(d) = 1/d^r$. Le paramètre r permet alors de contrôler l'influence de la distance sur le résultat. Lorsque $r = 0$, la distance entre les points n'est pas explicitement prise en compte, et l'histogramme de forces est alors équivalent à l'histogramme d'angles, tout en restant plus robuste à la discrétisation. Dans ce cas particulier, il est également possible de calculer l'histogramme de forces entre un objet et lui-même, résultant alors en une description directionnelle de la forme de cet objet. Cet histogramme est alors nommé \mathcal{F} -signature de l'objet. La \mathcal{F} -signature a notamment été exploitée plus en détail dans [\[Tabbone et al., 2003\]](#), pour la reconnaissance de symboles graphiques. Une autre valeur particulière est $r = 2$. La force est alors qualifiée de gravitationnelle, et l'on donne alors plus d'importance aux couples de points qui sont proches les uns des autres.

Les histogrammes de forces ont été étudiés de manière plus approfondie dans des travaux ultérieurs. Leur comportement par rapport aux transformations affines à été mis en évidence dans [\[Matsakis et al., 2004\]](#). Des optimisations algorithmiques [\[Ni et Matsakis, 2010\]](#) ont montré que les histogrammes de forces pouvaient être calculés en $\mathcal{O}(n \log n)$ dans certaines conditions (notamment lorsqu'il n'y a pas de chevauchement entre les objets). Les histogrammes de forces ont également été appliqués dans de nombreux travaux. Dans [\[Matsakis et al., 2001\]](#), une approche d'apprentissage par réseaux de neurones est proposée pour générer des descriptions linguistiques de configurations spatiales entre objets. Ils ont également été utilisés en robotique [\[Skubic et al., 2003\]](#), pour la recherche d'images par le contenu [\[Tabbone et Wendling, 2003\]](#), la mise en correspondance de scènes géographiques [\[Buck et al., 2013\]](#), ou encore la détection de motifs répétitifs dans des images de documents [\[Debled-Rennesson et Wendling, 2015\]](#).

2.2.2 Combinaisons avec d'autres familles de relations spatiales

Les auteurs de [\[Wang et Makedon, 2003; Wang et al., 2004\]](#) ont proposé une extension du principe des histogrammes d'angles et de forces, sous le nom de *R*-histogrammes*. Dans cette approche, ils proposent d'ajouter une information explicite concernant la distance, mais

2.2. Descripteurs de position relative

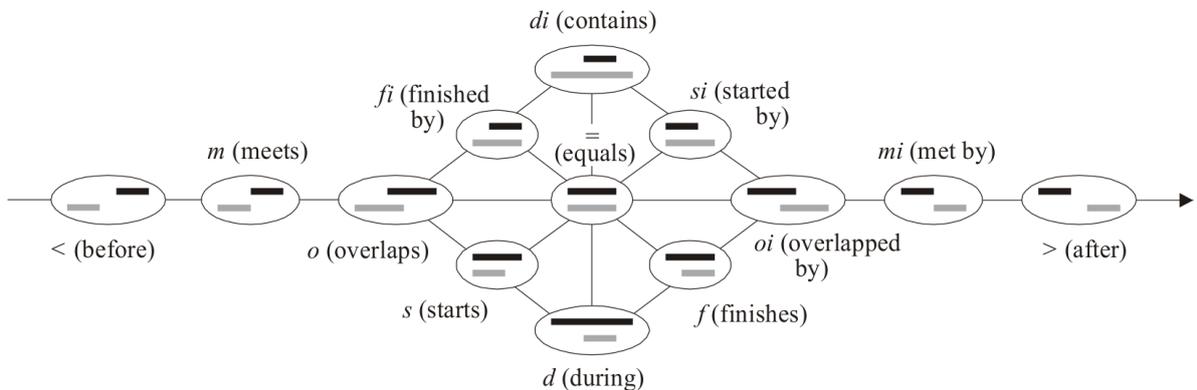


Figure 2.2 – Illustration des treize intervalles temporels d’Allen [Allen, 1983]. Schéma repris de [Matsakis et Nikitenko, 2005].

également une variable qualitative concernant le chevauchement des objets. Ces ajouts résultent alors en un histogramme bidimensionnel, combinant à la fois relations spatiales directionnelles et topologiques au sein d’une même représentation. Cependant, les expérimentations proposées ne permettent pas d’illustrer l’expressivité de l’approche par rapport au chevauchement d’objets.

Un autre descripteur de position relative est celui proposé par [Matsakis et Nikitenko, 2005]. Ces travaux proposent une extension de l’histogramme de forces, où la force d’attraction élémentaire entre couples de points est remplacée par un processus de calcul en fonction des intervalles temporels d’Allen [Allen, 1983]. Une représentation de ces intervalles est illustrée en Figure 2.2. Le descripteur devient alors un ensemble de treize histogrammes, chacun d’entre eux représentant un intervalle d’Allen de manière directionnelle. Pour appliquer efficacement ces intervalles sur les sections longitudinales des objets, les auteurs proposent d’appliquer différents types d’opérateurs flous afin de lisser certains problèmes de continuité, qui peuvent survenir lorsque les objets sont fragmentés (c’est-à-dire lors du passage d’un intervalle à un autre).

Dans les travaux de [Santosh *et al.*, 2009], le positionnement spatial des objets n’est plus envisagé de manière relative (c’est-à-dire pour objet par rapport à un autre), mais à partir d’un point de référence. Ce point de référence est défini comme le barycentre d’une région déterminée selon des caractéristiques topologiques entre les deux objets, obtenues à partir du formalisme *9-intersections* proposé par [Egenhofer et Herring, 1990]. La définition de cette région pour différentes configurations d’objets est illustrée en Figure 2.3. À partir de ce point de référence, une couverture angulaire est calculée pour chacun des deux objets, à partir d’un faisceau de droites le long d’un ensemble de directions discrètes. Ce descripteur de positionnement spatial est alors appliqué pour caractériser la structure de symboles graphiques présents au sein de documents techniques [Santosh *et al.*, 2012].

Récemment, les auteurs de [Matsakis *et al.*, 2015] ont présenté le Φ -descripteur, un nouveau descripteur de position relative qui s’inspire notamment de celui reposant sur les intervalles d’Allen [Matsakis et Nikitenko, 2005]. Ce descripteur incorpore des informations spatiales di-

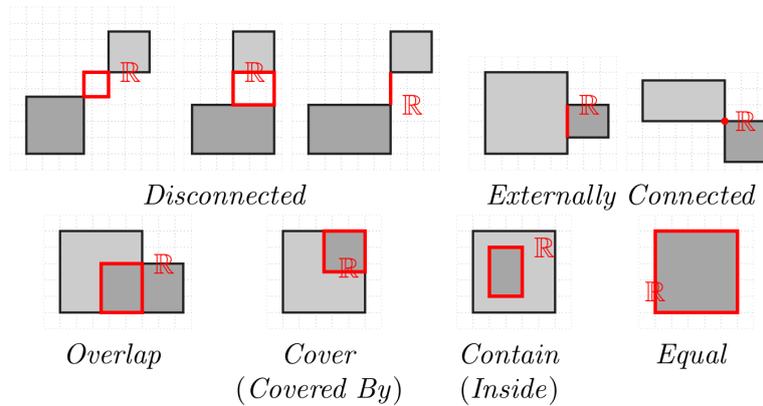


Figure 2.3 – Définition d’une région de référence entre deux objets à partir de caractéristiques topologiques. Illustration reprise de [Santosh, 2011].

rectionnelles, topologiques et de distance au sein de treize histogrammes qui peuvent être organisés en plusieurs groupes. Les auteurs ont également étudié les propriétés de ce descripteur par rapport aux transformations affines [Matsakis, 2016], et introduit différents types d’opérateurs flous pour effectuer des évaluations de relations topologiques [Matsakis et Naeem, 2016]. Il s’agit d’un modèle puissant qui peut exprimer un grand nombre de configurations spatiales différentes. Cependant, la dimensionnalité de la représentation générée peut la rendre délicate à exploiter, notamment lorsqu’il s’agit de considérer le positionnement spatial de plusieurs objets simultanément.

De manière générale, les descripteurs de position relative proposés dans la littérature ont pour objectif de synthétiser l’information spatiale au sein d’une même représentation, prenant généralement la forme de vecteurs de caractéristiques. Ces descripteurs sont alors particulièrement adaptés pour des tâches de reconnaissance ou de classification. Cependant, certaines applications nécessitent l’évaluation plus précise de relations spatiales prédéfinies.

2.3 Paysages de relations spatiales

Nous étudions ici les approches qui s’appuient plutôt sur le point de vue de l’évaluation de relations spatiales. Le principe de ces approches consiste généralement à modéliser, en chaque point de l’espace image, le degré de validité d’une relation spatiale donnée. Ces types d’approches permettent alors de visualiser la relation spatiale sous la forme d’une zone d’acceptation graduelle, qui s’adapte en fonction de la forme des objets considérés. Ce principe d’évaluation et de visualisation de l’information spatiale trouve notamment ses racines dans des travaux issus de la psychologie et de la cognition, en particulier ceux proposés par [Logan et Sadler, 1996].

2.3. Paysages de relations spatiales

2.3.1 Paysages flous morphologiques

La première application de cette idée a été proposée par [Bloch, 1999]. Dans ces travaux, une approche inspirée de la morphologie mathématique est présentée pour modéliser les relations spatiales directionnelles directement dans l'espace image. Étant donné une relation spatiale directionnelle, par exemple « à gauche de », le principe consiste à calculer en chaque point de l'espace image (c'est-à-dire pour chaque pixel, dans le cas d'images discrètes) le degré de validité de cette relation, par rapport à un objet de référence.

Formellement, un paysage flou directionnel est obtenu en effectuant une dilatation morphologique floue de l'objet de référence, à partir d'un élément structurant bien défini. Cet élément structurant doit modéliser la sémantique que l'on souhaite associer à la relation spatiale étudiée. Dans le cas des relations directionnelles, il s'agit généralement d'une fonction décroissante de la direction considérée. Dans [Bloch, 1999], un tel élément structurant est défini de la manière suivante. Soit $\theta \in [0, 2\pi]$ la direction que l'on souhaite évaluer, et soit u_θ le vecteur unitaire associé à cette direction. L'élément structurant dans la direction θ est alors défini par :

$$\forall x \in \mathcal{S}, \nu_\theta(x) = \max \left(0, 1 - \frac{2}{\pi} \arccos \frac{o\vec{x} \cdot \vec{\mu}_\theta}{\|o\vec{x}\|} \right), \quad (2.1)$$

où \mathcal{S} désigne l'espace image, et le point o désigne l'origine de l'élément structurant ν_θ . Notons que d'autres formulations pourraient être employées pour contrôler le comportement graduel de la relation spatiale considérée. Soit A un objet représenté par un ensemble flou μ_A sur \mathcal{S} . La dilatation morphologique floue de l'objet A par l'élément structurant ν_θ est alors définie par :

$$\forall x \in \mathcal{S}, \delta_{\nu_\theta}(\mu_A)(x) = \sup_{u \in \mathcal{S}} \min(\nu_\theta(x - u), \mu_A(x)). \quad (2.2)$$

Notons également ici que l'opérateur \min pourrait être remplacé par un autre opérateur d'intersection floue (t -norme). En Figure 2.4, nous pouvons observer quelques paysages flous directionnels générés par cette approche pour modéliser la relation spatiale « à droite de ». Il s'agit ainsi du résultat de la dilatation morphologique floue avec l'élément structurant ν_0 pour différents objets bidimensionnels.

À partir d'un tel paysage flou défini par rapport à un objet de référence A , il est alors possible d'évaluer le degré de validité de la relation spatiale pour un objet cible B . Ceci peut être effectué en projetant cet objet cible sur le paysage flou, puis en appliquant différents types de mesures floues sur cette projection. Les mesures dites de nécessité et de possibilité sont alors couramment utilisées. Dans ce contexte, elles sont données par :

$$\Pi(\mu_B, \delta_{\nu_\theta}(A)) = \sup_{x \in \mathcal{S}} \min(\mu_B(x), \delta_{\nu_\theta}(A)(x)), \quad (2.3)$$

$$N(\mu_B, \delta_{\nu_\theta}(A)) = \inf_{x \in \mathcal{S}} \max(1 - \mu_B(x), \delta_{\nu_\theta}(A)(x)), \quad (2.4)$$

où B désigne l'objet cible et $\delta_{\nu_\theta}(A)$ désigne le paysage flou directionnel de A dans la direction θ . La nécessité correspond à une évaluation pessimiste, tandis que la possibilité est une mesure

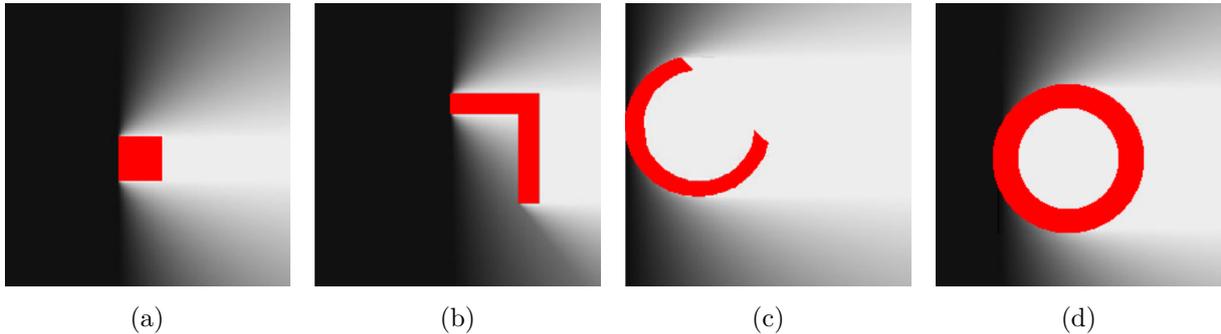


Figure 2.4 – Exemples de paysages flous directionnels permettant d’évaluer localement la relation spatiale « à droite de » (dans la direction $\theta = 0$). Le paysage flou s’adapte à la forme des différents objets. Exemples repris de [Delaye, 2011], d’après [Bloch et Ralescu, 2003].

optimiste. Ces valeurs représentent alors des degrés de confiance que l’on peut avoir pour la proposition spatiale « B est dans la direction θ de A ».

Les paysages flous morphologiques constituent donc un modèle particulièrement adapté à la visualisation et l’évaluation de relations spatiales prédéfinies, qui permet de prendre en compte explicitement la forme des objets. Le formalisme reposant sur la logique flou permet alors de générer des évaluations selon différents points de vue. Ce modèle a notamment été employé pour la segmentation d’images de cerveau en IRM [Colliot *et al.*, 2006], ou pour la détection de bâtiments en imagerie satellitaire [Ok *et al.*, 2013].

2.3.2 Champs de forces

Une autre approche pour générer des paysages de relations spatiales est celle des \mathcal{F} -*templates*, que l’on peut également nommer *champs de forces*. Ce concept a été proposé initialement par [Matsakis *et al.*, 2006] (puis étendu aux objets flous par [Xin *et al.*, 2006] et généralisé dans [Matsakis *et al.*, 2010]) comme une version duale du modèle des histogrammes de forces, selon le point de vue que nous avons précédemment exposé en Section 2.1. Pour une relation spatiale \mathcal{R} et un objet B donnés, l’approche consiste à calculer en chaque point de l’image le degré de validité de la relation par rapport à l’objet. Un champ de forces est donc fondamentalement similaire à un paysage flou directionnel morphologique, comme l’illustre l’exemple présenté en Figure 2.5. Il est noté $\mathcal{F}^{\theta B}$, où θ désigne la direction du paysage et B l’objet de référence.

Les principales différences entre les deux approches viennent alors essentiellement du formalisme employé et de l’algorithme de calcul du paysage. Ici, plutôt que d’appliquer des opérateurs morphologiques telles que la dilatation à partir d’un élément structurant, l’approche considère des sections longitudinales à travers des droites orientées dans différentes directions, d’une manière analogue au modèle des histogrammes de forces. Enfin, de la même manière que l’aspect graduel des paysages flous directionnels peut être régulé par la définition de l’élément structurant, dans le cas des champs de forces il peut également être contrôlé par la définition de la force d’attraction entre les points, comme pour les histogrammes de forces.

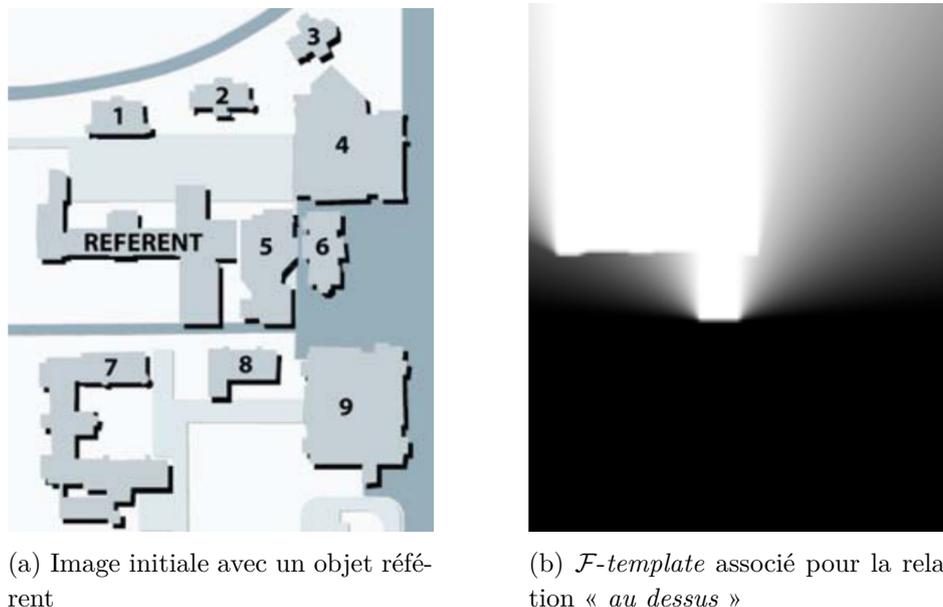


Figure 2.5 – Exemple de \mathcal{F} -template obtenu pour un objet de référence et pour la relation spatiale « au dessus de ». Images reprises de [Delaye, 2011], d’après [Xin et al., 2006].

Finalement, pour évaluer dans quelle mesure un objet A se trouve dans la direction θ de B , le même type d’opérateurs flous que ceux présentés pour les paysages flous morphologiques peuvent être utilisés (en d’autres termes, les mesures de nécessité et de possibilité peuvent être appliquées sur l’objet A projeté dans le champ de forces $\mathcal{F}^{\theta B}$).

2.4 Synthèse

Dans ce chapitre, nous avons présenté un certain nombre d’approches permettant de décrire les relations spatiales entre objets représentés dans des images. Pour présenter ces différentes approches, nous avons souligné l’existence d’une dualité entre le concept de relation spatiale et celui de position relative. Nous avons, d’une part présenté différents descripteurs de position relative, et d’autre part des approches dédiées à la modélisation de relations spatiales directement dans l’espace image. Si les descripteurs permettent de résumer synthétiquement l’information spatiale au sein d’une même représentation, les approches par paysages permettent, quant à elles, une meilleure visualisation d’une relation spatiale donnée.

En somme, ces deux types d’approches permettent d’envisager la modélisation des relations spatiales selon deux points de vue bien souvent complémentaires. Elles permettent ainsi de représenter efficacement un large panel de configurations qui s’intègrent dans les trois principales familles de relations spatiales définies dans le chapitre précédent (directionnelles, topologiques et métriques). Néanmoins, elles peuvent alors exposer leurs limites pour décrire ou évaluer des situations complexes, notamment lorsque les propositions linguistiques spatiales correspondantes

sont chargées d'ambiguïtés (comme par exemple l'*entourement*, l'alignement, ou encore l'imbrication d'objets). Par ailleurs, ces approches se focalisent sur la caractérisation de l'information spatiale uniquement entre des couples d'objets au sein d'une image.

Dans le chapitre suivant, nous nous intéresserons alors aux approches qui ont cherché à modéliser ces relations spatiales plus complexes. Nous étudierons également comment prendre en compte l'information spatiale au sein d'images composées de plusieurs objets d'intérêt, mais également à travers de multiples images, à partir de stratégies d'apprentissage.

Chapitre 3

Vers des relations spatiales complexes et composites

Sommaire

3.1 Relations spatiales complexes	32
3.1.1 La relation « <i>entouré par</i> »	32
3.1.2 La relation « <i>entre</i> »	33
3.1.3 Autres relations complexes	35
3.2 Relations spatiales composites	36
3.2.1 Descripteurs composites de relations spatiales	36
3.2.2 Apprentissage de relations spatiales	37
3.3 Synthèse et positionnement des travaux	39

Dans ce chapitre, nous proposons d’envisager la modélisation des relations spatiales en analyse d’images au-delà de la catégorisation par descripteurs de position relative et de paysages de relations spatiales suggérée précédemment. D’une part, nous illustrons dans la [Section 3.1](#), à travers plusieurs exemples de travaux, qu’il existe des relations spatiales complexes qui dépassent le cadre des familles de relations élémentaires proposées initialement par Freeman, et que des modélisations spécifiques sont souvent nécessaires pour les représenter efficacement. D’autre part, en [Section 3.2](#), nous présentons plusieurs études qui proposent des représentations composites de relations spatiales. Par le terme composite, nous entendons ainsi à la fois la caractérisation de l’information spatiale entre plusieurs objets au sein d’une image (à partir de descriptions par graphes), et à la fois la construction des prototypes de configurations spatiales à travers des ensembles d’images (à partir de stratégies d’apprentissage). Enfin, en [Section 3.3](#), nous concluons le chapitre par une synthèse globale de cette partie du manuscrit. Nous y énonçons notamment notre positionnement général, à partir duquel les propositions méthodologiques de cette thèse seront fondées.

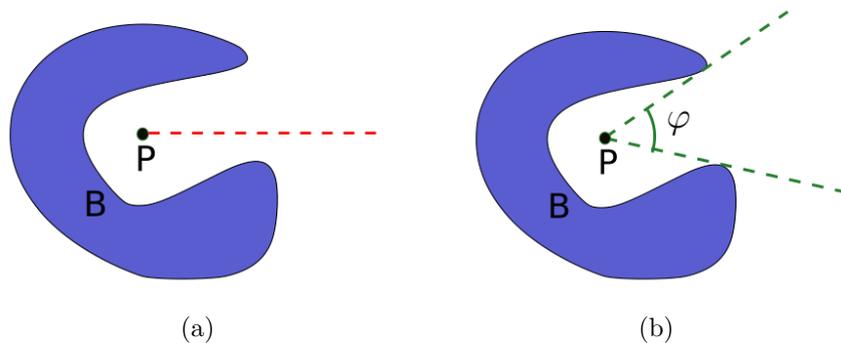


Figure 3.1 – Notions de visibilité et de couverture angulaire pour la relation spatiale « entouré par ». (a) Les points sur la demi-droite rouge sont considérés comme visibles depuis le point P . (b) La couverture angulaire pour laquelle des points sont visibles depuis P permet d’évaluer le degré de validité de la relation. Illustrations reprises de [Vanegas *et al.*, 2009].

3.1 Relations spatiales complexes

Nous proposons ici d’explorer différentes études qui se sont intéressées à des relations spatiales dites complexes, dans le sens où elles ne pourraient pas être simplement déduites des relations spatiales élémentaires, mais au contraire, nécessiteraient des modélisations spécifiques pour prendre en compte leur complexité.

3.1.1 La relation « entouré par »

Initialement, la relation spatiale « entouré par » avait été envisagée par Freeman comme une combinaison des quatre relations cardinales élémentaires : un objet A est « entouré par » un objet B si ce dernier se situe simultanément « à gauche », « à droite », « au-dessus » et « en-dessous » de celui-ci. Selon ce point de vue, cette relation spatiale pourrait donc être déduite des relations élémentaires. En pratique, même si cette idée semble au premier abord conforme à l’intuition, la modélisation quantitative de cette relation s’avère souvent difficile. Ceci est principalement dû au fait que la sémantique de cette relation est particulièrement subjective. D’une part, comme nous l’avons évoqué, elle peut être définie de manière directionnelle, et d’autre part, elle fait intervenir des notions topologiques, car il s’agit de détecter une certaine forme d’inclusion.

Une des premières approches qui a été proposée pour modéliser cette relation spatiale est celle de [Rosenfeld et Klette, 1985]. Dans cette approche, une notion de visibilité depuis un point est définie : un point est dit visible depuis un autre point si l’on peut tracer une droite entre ces points sans traverser un objet (voir par exemple la Figure 3.1). L’entourement depuis un point est alors défini comme la quantité de points de l’autre objet qui sont visibles lorsque l’on effectue un balayage angulaire dans toutes les directions. Formellement, le degré de validité de

3.1. Relations spatiales complexes

la proposition « a est entouré par B » où a est un point et B un objet est donné par :

$$s(a, B) = \frac{1}{2\pi} \int_0^{2\pi} r_\theta(a, B) d\theta \quad (3.1)$$

où r_θ est une fonction binaire telle que $r_\theta(a, B) = 1$ lorsque la demi-droite partant du point a dans la direction θ traverse l'objet B , et $r_\theta(a, B) = 0$ sinon. La valeur globale d'entournement pour A par rapport à B peut alors être obtenue par :

$$s(A, B) = \min_{a_i \in A} s(a_i, B). \quad (3.2)$$

Dans [Vanegas *et al.*, 2011], une extension de cette approche par visibilité angulaire est présentée. Les auteurs proposent de restreindre le calcul de la visibilité angulaire aux seuls points qui sont localisés sur les concavités de l'objet, partant de l'hypothèse que seules les zones concaves peuvent intervenir dans l'entournement d'un autre objet. Ils proposent également de pondérer cette évaluation par une fonction de distance à l'objet référent. Ce type d'approche permet d'obtenir des résultats qui sont généralement conformes à l'intuition, mais qui sont particulièrement sensibles à la discrétisation des objets en fonction des différentes directions, de la même manière que pour l'histogramme d'angles.

Une autre approche repose sur l'utilisation de l'histogramme de forces pour évaluer cette relation [Matsakis et Andréfouët, 2002]. La même idée de couverture angulaire est appliquée, mais cette fois-ci à partir de l'histogramme de forces. La définition proposée repose sur la comptabilisation de zones de l'histogramme de forces qui sont inférieures à un certain seuil. Cette définition est étendue aux α -coupes $\mathcal{F}_{AB}^\alpha = \{\theta, \mathcal{F}_{AB}(\theta) \geq \alpha\}$ de l'histogramme \mathcal{F}_{AB} normalisé, pour différentes valeurs de $\alpha \in [0, 1]$, afin d'introduire une notion de flou. Ceci permet alors d'évaluer la relation selon différents points de vue. Cependant, cette approche souffre d'un problème majeur : elle ne peut pas distinguer si c'est A qui est entouré par B , ou si c'est B qui est entouré par A . En effet, lorsque A est entouré par B , l'histogramme \mathcal{F}_{BA} indiquera que B se trouve dans toutes les directions depuis A , mais l'histogramme opposé \mathcal{F}_{AB} indiquera alors que A se trouve également dans toutes les directions de B .

3.1.2 La relation « *entre* »

La relation spatiale « *entre* » est une des relations élémentaires proposée initialement par Freeman. Pendant un certain temps, peu de travaux de recherche ont porté un intérêt à cette relation, notamment du fait qu'il s'agit d'une relation ternaire (elle fait intervenir trois objets, par exemple, un objet C peut être entre un objet A et un objet B). En 2006, les auteurs de [Bloch *et al.*, 2006] ont présenté une étude approfondie de cette relation spatiale, proposant un panel de méthodes permettant de modéliser sa complexité dans différentes situations. La première méthode proposée repose sur l'idée de définir la région « *entre* » deux objets comme l'enveloppe convexe de l'union de ces deux objets, comme l'illustre par exemple la Figure 3.2. Cette idée intuitive peut notamment être étendue aux objets flous. La seconde méthode proposée repose

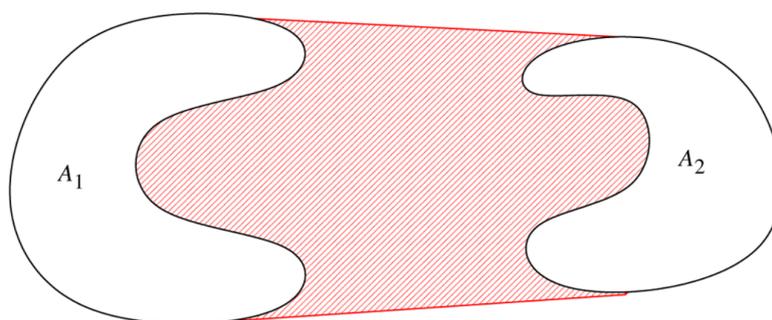


Figure 3.2 – Zone de validité de la relation spatiale « *entre* » à partir de l’enveloppe convexe de l’union des deux objets. Illustration reprise de [Bloch *et al.*, 2006].

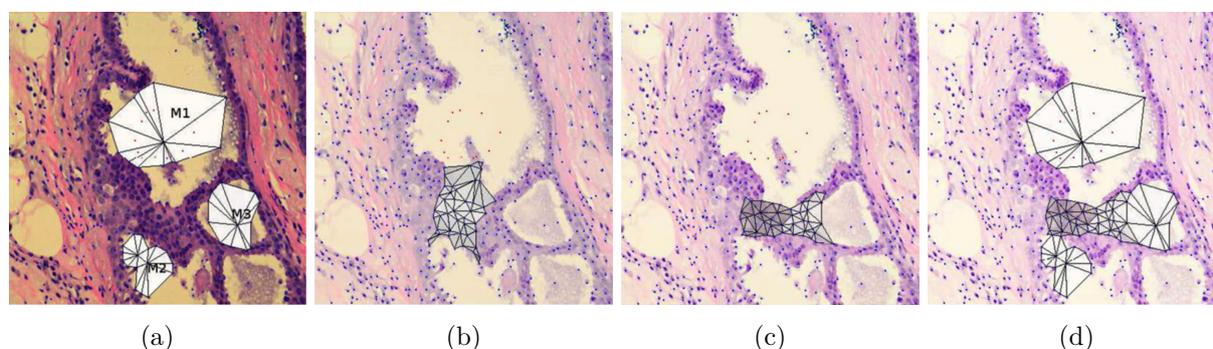


Figure 3.3 – Définition de la relation « *entre* » pour un maillage de points d’une image histologique. (a) Trois régions d’intérêt, (b) région « *entre* » M_1 et M_2 , (c) extension à la région M_3 et (d) affichage des régions en surbrillance. Images issues de [Loménie et Racoceanu, 2012].

sur un formalisme de morphologie mathématique, qui peut être combiné à un histogramme d’angles pour détecter les directions dominantes. Le principe consiste à appliquer des dilations directionnelles successives des objets afin de détecter une zone d’intersection. Cette zone peut alors être étendue, par exemple à partir d’un calcul de distances géodésiques appliqué sur l’union des objets. Enfin, la troisième approche proposée dans ces travaux repose sur la définition de la notion de *visibilité* floue. L’idée de l’enveloppe convexe de l’union des objets est alors restreinte aux seuls points depuis lesquels les deux objets sont visibles, ceci afin d’obtenir des régions qui sont plus conformes à l’intuition.

Plus récemment, [Loménie et Racoceanu, 2012] ont proposé une modélisation de la relation spatiale « *entre* » dans des espaces topologiques discrets, où les objets peuvent être représentés par des maillages (comme par exemple la triangulation de Delaunay). Différents opérateurs morphologiques sont définis pour raisonner sur ce type de données, afin de déterminer la région « *entre* » pour plusieurs objets. Ces opérateurs sont notamment appliqués pour l’analyse d’images histologiques. Quelques exemples sont ainsi illustrés en Figure 3.3.

3.1. Relations spatiales complexes

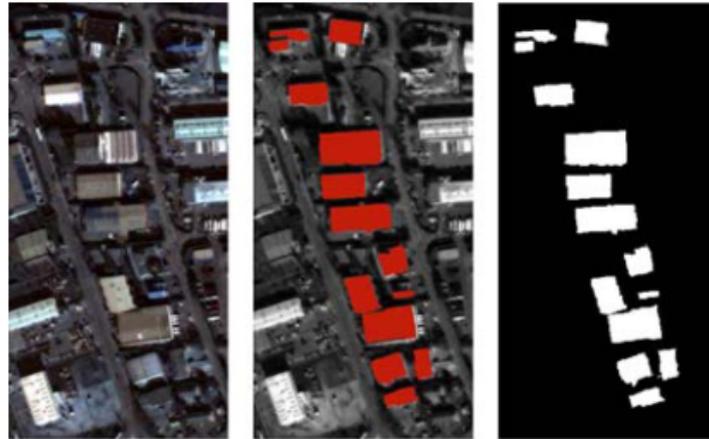


Figure 3.4 – Détection d’objets géographiques globalement alignés dans une image de télédétection. Images reprises de [Vanegas *et al.*, 2013].

3.1.3 Autres relations complexes

Alignement et parallélisme

Un autre type de relations spatiales complexes consiste à étudier l’alignement et le parallélisme d’objets. La complexité vient ici notamment du fait qu’il faut considérer plusieurs objets simultanément (et non pas deux ou trois objets comme pour la plupart des autres relations). Dans leurs travaux, les auteurs de [Vanegas *et al.*, 2013] ont proposé une approche qui permet de détecter des séries d’objets qui seraient *alignés* ou *parallèles*. Pour détecter l’alignement, un histogramme d’orientation similaire à l’histogramme d’angles (avec une période de π au lieu de 2π) est couplé à une mesure de similarité floue. Les auteurs proposent également la définition de graphes de voisinages qui permettent de détecter à la fois des alignements locaux (entre trois objets voisins) et globaux (par fusion de groupes d’alignements locaux).

Dans le même article, il est proposé une approche pour évaluer le parallélisme de groupes d’objets, là aussi d’un point de vue local ou global. Ces évaluations sont effectuées à partir de paysages flous directionnels orientés dans la direction normale à l’axe principal des objets, assimilés alors à un champ visuel. Différents opérateurs flous sont définis pour déterminer si des groupes d’objets alignés sont parallèles. Ces travaux sont principalement appliqués pour la détection d’objets urbains dans des images de télédétection. La Figure 3.4 illustre un exemple d’une série de bâtiments alignés qui a été détectée par cette approche.

La relation « à travers »

Une étude de la relation spatiale « à travers » (*to go across* en anglais) a été proposée par [Vanegas *et al.*, 2009]. Afin de mieux comprendre la sémantique de cette relation, les auteurs ont effectué un sondage sur un échantillon de personnes, celui-ci leur demandant leur avis quant à des configurations spatiales où un objet linéique L passe à travers un autre objet R de différentes

façons. Ce sondage leur a permis de dégager plusieurs significations de la relation. La première signification est la moins contraignante, elle consiste simplement à détecter si L passe par R . Pour cela, il s'agit de vérifier que L intersecte bien R , et que les extrémités de L sont à l'extérieur de R . La seconde signification ajoute la contrainte géométrique que l'objet linéique L rentre dans l'objet R par un côté, puis ressort du côté opposé. Pour détecter cette situation, la méthode considère une fonction floue de la distance angulaire entre les vecteurs tangents à la frontière de l'objet R . Enfin, la troisième signification fait l'hypothèse que l'objet linéique L traverse profondément l'objet R . Pour prendre en compte cet aspect, une fonction floue de la distance de l'objet L à un point de l'érodé ultime de R est calculée.

La relation « *le long de* »

Dans [Takemura *et al.*, 2012], la relation spatiale « *le long de* » (*along* en anglais) est étudiée. Cette étude propose de définir cette relation à partir de la région de l'espace qui se situe « *entre* » deux objets. L'hypothèse consiste à dire que la région le long de deux objets doit être de forme allongée, orientée dans la direction des objets, et proche de chacun d'eux. Pour cela, l'étude se décline en plusieurs stratégies, qui reposent chacune sur l'utilisation des approches proposées par [Bloch *et al.*, 2006] pour modéliser la relation « *entre* ». Comme pour la relation « *à travers* », différentes définitions sont proposées, permettant ainsi de prendre en compte la distance entre les objets, ou bien de manipuler également des objets flous.

Nous avons ainsi présenté différentes méthodes visant à caractériser des relations spatiales plus complexes que les familles de relations élémentaires de Freeman. Néanmoins, ces méthodes partent généralement de définitions conceptuelles *a priori* sur les relations étudiées, qui sont alors appliquées à l'image sur différents objets.

3.2 Relations spatiales composites

Dans cette section, nous examinons un certain nombre de travaux qui ont proposé d'exploiter les relations spatiales d'une manière que nous qualifions de composite. Par ce terme, nous entendons d'une part la caractérisation de configurations spatiales entre plusieurs objets d'intérêt présents dans une image, et d'autre part la possibilité d'apprendre des prototypes de relations spatiales à partir d'une base d'images d'apprentissage.

3.2.1 Descripteurs composites de relations spatiales

Suite au développement de descripteurs de position relative efficaces (voir la Section 2.2 du Chapitre 2), plusieurs études ont alors été portées sur la description du contenu d'images composées de plusieurs objets d'intérêt. Pour cela, une représentation par graphes relationnels attribués s'est montrée particulièrement adaptée. L'idée consiste à relier les différents objets d'intérêt d'une image (celle-ci ayant par exemple pu être segmentée préalablement) à l'aide d'un graphe relationnel où des attributs (ou descripteurs) peuvent être assignés aux nœuds et aux

3.2. Relations spatiales composites

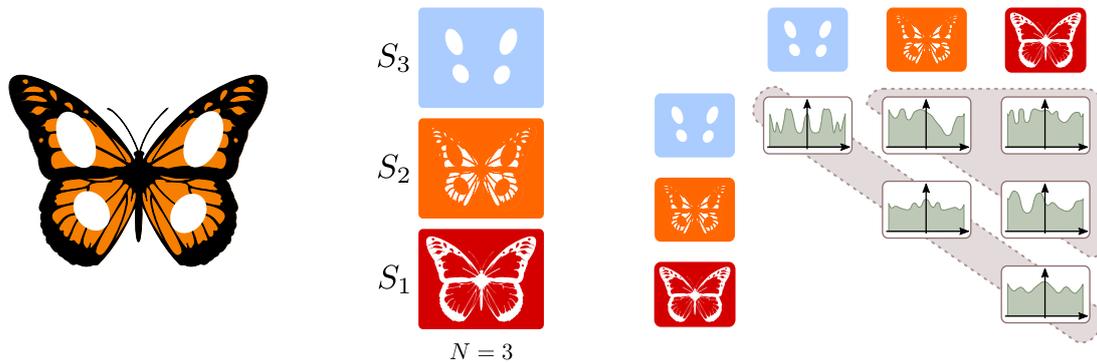


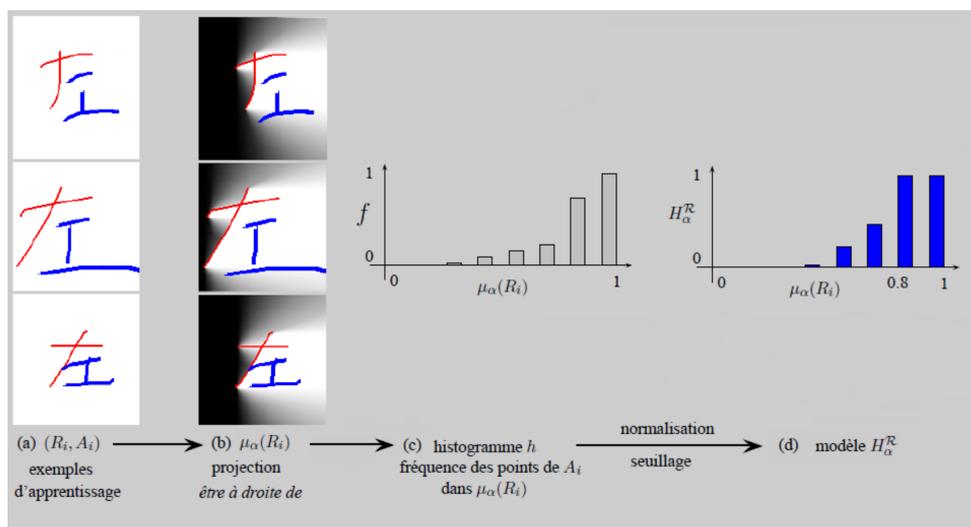
Figure 3.5 – Illustration schématique de la décomposition en histogrammes de forces. Figure reprise de [Garnier *et al.*, 2012].

arêtes. Dans ce contexte, on pourra citer notamment les travaux de [Tabbone et Wendling, 2004] pour la recherche d’images par le contenu, ceux de [Cesar *et al.*, 2005] pour la reconnaissance de visages, et enfin ceux de [Santosh *et al.*, 2012] pour la reconnaissance de symboles techniques.

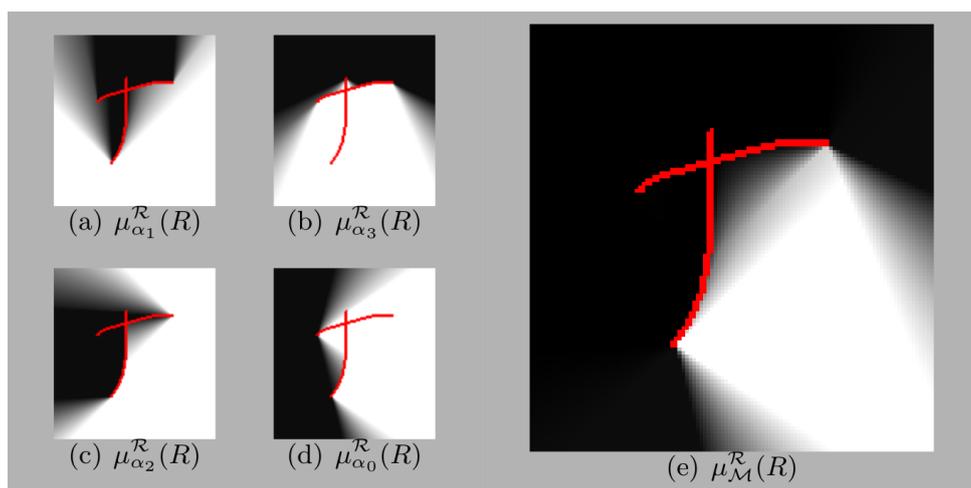
Une des approches les plus abouties pour la description des relations spatiales composites au sein d’images est celle proposée récemment par [Garnier *et al.*, 2012]. Une illustration schématique de cette chaîne de traitement est présentée en Figure 3.5. Elle commence par la segmentation des images en différentes régions à partir de leurs lignes de niveaux, séparées en différents quantiles. Puis, les relations spatiales entre chaque couple de régions issues de la segmentation sont caractérisées à partir du modèle des histogrammes de forces [Matsakis et Wendling, 1999]. Cette approche propose également d’exploiter la \mathcal{F} -signature pour décrire la forme des régions, résultant en une description homogène et structurée du contenu d’une image. Le descripteur peut alors être visualisé sous la forme d’une représentation matricielle similaire à la matrice d’adjacence d’un graphe, mais où les entrées sont les histogrammes de forces correspondant aux différents nœuds et arêtes. Une stratégie de comparaison de tels descripteurs est également proposée, et des résultats préliminaires sur des bases d’objets structurés ont pu illustrer l’intérêt de cette approche. Cependant, cette approche reste fortement dépendante de la méthode de segmentation associée, et il est nécessaire de fixer *a priori* le nombre de régions des images pour pouvoir les comparer. Par ailleurs, des problématiques de nature combinatoire interviennent pour la mise en correspondance de ces descripteurs, rendant alors le passage à l’échelle et l’apprentissage de telles représentations des tâches non triviales.

3.2.2 Apprentissage de relations spatiales

Les travaux présentés par Adrien Delaye dans sa thèse [Delaye, 2011] figurent parmi les premiers à avoir proposé un apprentissage de prototypes de relations spatiales à partir de données, alors nommés méta-modèles spatiaux. L’approche proposée repose principalement sur l’utilisation du modèle des paysages flous morphologiques [Bloch, 1999], et a pour objectif une tâche dite de *localisation* : il s’agit de déterminer une région de l’espace qui serait la plus appro-



(a) Étape d'apprentissage à partir de plusieurs couples d'objets



(b) Modèles spatiaux obtenus à partir de l'apprentissage

Figure 3.6 – Exemples de méta-modèles spatiaux générés à partir de plusieurs exemples d'apprentissage. Illustrations issues de [Delaye, 2011].

priée pour modéliser un ensemble de configurations spatiales considérées comme similaires. Pour cela, il est proposé d'apprendre un prototype de relation spatiale \mathcal{R} à partir d'un ensemble de couples d'objets d'apprentissage $\{(R_i, A_i), i = 1, \dots, N\}$. Le principe de cette approche est illustré en Figure 3.6. Les paysages flous directionnels $\mu_\alpha(R_i)$ pour les quatre directions cardinales $\alpha \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ sont calculés. Un histogramme H_α^R est construit pour les différentes directions, comptabilisant les fréquences d'apparition des degrés de validité pour les objets cibles A_i par rapport à chaque paysage. Ces histogrammes constituent alors le modèle qui aura été appris à partir des exemples. Ensuite, en situation de test avec un couple d'objets (R, A) , les pay-

3.3. Synthèse et positionnement des travaux

sages flous $\mu_\alpha^{\mathcal{R}}(R) = H_\alpha^{\mathcal{R}} \circ \mu_\alpha(R)$ sont calculés pour les différentes directions, appliquant ainsi les histogrammes appris aux paysages flous de l'objet R . L'intersection de ces modèles permet alors d'obtenir un paysage flou $\mu_{\mathcal{M}}^{\mathcal{R}}(R)$ qui représente la région de l'espace la plus conforme au méta-modèle appris sur la base d'apprentissage. Différentes déclinaisons de cette méthode ont également été proposées, notamment afin de prendre également en compte la distance entre les objets de l'ensemble d'apprentissage. Cette approche a été principalement appliquée à l'analyse de tracés manuscrits [Delaye et Anquetil, 2014], où les objets d'intérêt sont généralement des structures linéiques (par exemple pour la reconnaissance d'expressions mathématiques).

Les auteurs de [Santosh *et al.*, 2014b] ont proposé une autre méthode, inspirée des stratégies d'apprentissage par sacs de caractéristiques communément employées en vision par ordinateur. Cette méthode consiste à construire des vocabulaires de configurations spatiales apparaissant dans une base d'images d'apprentissage, en utilisant des descripteurs de position relative présentés précédemment [Santosh *et al.*, 2012]. L'approche a été appliquée dans le contexte applicatif de la reconnaissance de symboles techniques, à partir de primitives visuelles bien contrôlées (comme par exemple les traits, les cercles, les coins et les extrémités), limitant ainsi les problématiques de mise en correspondance de graphes lors du calcul de similarité entre les représentations.

3.3 Synthèse et positionnement des travaux

Au cours de la première partie de ce manuscrit, nous avons exposé un état de l'art des différentes approches proposant de modéliser et d'exploiter les relations spatiales entre des objets représentés dans des images, en envisageant progressivement différents points de vue. Ainsi, dans le premier chapitre, nous avons proposé une introduction au domaine des relations spatiales en posant quelques définitions, et en présentant des approches classiques reposant sur des modélisations qualitatives et quantitatives. Nous avons notamment mis en exergue la nécessité de prendre en considération la forme des objets de manière explicite pour modéliser correctement l'information spatiale. Ensuite, durant le second chapitre, nous avons approfondi notre étude des relations spatiales selon le point de vue quantitatif, en mettant en avant l'importance de la dualité conceptuelle entre les approches par paysages de relations spatiales et celles par descripteurs de position relative. Enfin, dans ce dernier chapitre, nous avons illustré l'importance de proposer des modèles spécifiques pour certaines relations spatiales complexes. Par ailleurs, nous avons également examiné différentes méthodes qui permettent de caractériser l'ensemble des relations spatiales composites qui apparaissent entre plusieurs objets d'intérêt dans les images, en employant notamment des stratégies d'apprentissage.

Les travaux de recherche présentés dans la suite de ce manuscrit s'articulent ainsi autour de deux aspects majeurs. Tout d'abord, dans la [Partie II](#), nous proposons une nouvelle catégorie de relations spatiales complexes. Nous nous intéressons spécifiquement aux situations où les objets peuvent être imbriqués les uns dans les autres. D'ailleurs, nous avons vu que certaines approches qualitatives comme celle de [Cohn *et al.*, 1997] permettent de détecter ce type de situations. Cependant, à notre connaissance, aucune approche quantitative n'a été proposée pour décrire efficacement ce type de configurations spatiales avec des objets de formes arbitraires.

Pourtant, la caractérisation de telles configurations peut répondre à de nombreux besoins dans différents contextes applicatifs tels que l'imagerie biomédicale, l'analyse de documents ou encore la télédétection. Nous proposons ainsi une modélisation originale de ce type de configurations, en explorant d'une part le volet des descripteurs de position relative, et d'autre part celui des paysages de relations spatiales.

Ensuite, dans la [Partie III](#), nous nous intéressons de manière plus approfondie à l'exploitation de modèles de relations spatiales composites pour des tâches de reconnaissance et de classification d'images. Si les approches proposées par [[Delaye et Anquetil, 2014](#)] et [[Santosh *et al.*, 2014b](#)] constituent des avancées significatives pour l'apprentissage de configurations spatiales à partir de bases d'images, elles restent néanmoins dédiées à des domaines d'application spécifiques (reconnaissance de tracés manuscrits et de symboles techniques) et sont alors difficilement généralisables dans différents contextes. Nous proposons ainsi un cadre générique d'apprentissage de configurations spatiales apparaissant à travers différents niveaux d'échelle dans les images. Cette approche est notamment naturellement combinable avec d'autres représentations efficaces en vision par ordinateur comme les approches par sacs de caractéristiques visuelles. Les résultats expérimentaux que nous présentons permettent d'attester de l'intérêt de cette approche pour différents types d'applications.

Partie II

Modélisation de relations spatiales complexes

Résumé

Nous nous intéressons à la modélisation de configurations spatiales complexes, et en particulier aux situations où les objets peuvent être imbriqués les uns dans les autres. Cette notion est formalisée par deux nouvelles relations spatiales, que nous nommons enlacement et entrelacement. Nous proposons alors un modèle générique qui permet de décrire et d'évaluer l'enlacement et l'entrelacement entre objets bidimensionnels, avec une granularité directionnelle. Ce modèle permet de distinguer efficacement différents types de configurations spatiales complexes, et présente des propriétés d'invariance bien souvent utiles en reconnaissance de formes. Cette approche est validée expérimentalement dans différents contextes applicatifs, tels que la caractérisation de vaisseaux sanguins en imagerie ophtalmologique, la reconnaissance de lettres décoratives pour l'analyse de documents anciens, et la classification d'objets géographiques urbains en télédétection. Par ailleurs, nous proposons également une stratégie d'évaluation et de visualisation de ces nouvelles relations spatiales, à partir de la définition de paysages flous directionnels.

Chapitre 4

Descripteurs directionnels d'enlacement et d'entrelacement

Sommaire

4.1	Relations d'enlacement et d'entrelacement	44
4.2	Enlacement directionnel entre objets	45
4.2.1	Notation et définitions préliminaires	45
4.2.2	Enlacement le long d'une droite	47
4.2.3	Enlacement global entre objets	49
4.2.4	Propriétés d'invariance	52
4.3	Histogrammes d'enlacement et d'entrelacement	52
4.3.1	Représentations par histogrammes	52
4.3.2	Exemples illustratifs	54
4.3.3	Complexité algorithmique	55
4.4	Bilan scientifique	56

Dans ce chapitre, nous nous intéressons à la description de configurations spatiales complexes, où les objets peuvent être imbriqués les uns dans les autres. Tout d'abord, dans la [Section 4.1](#), nous proposons de formaliser ces notions par deux nouvelles relations spatiales, nommées enlacement et entrelacement. En [Section 4.2](#), après avoir posé certaines notations sur la manipulation d'objets bidimensionnels, nous présentons le modèle proposé pour caractériser l'enlacement relatif d'objets avec une granularité directionnelle. Nous énonçons également les différentes propriétés d'invariance de ce modèle. Dans la [Section 4.3](#), nous présentons les descripteurs d'enlacement et d'entrelacement obtenus à partir du modèle de calcul de l'enlacement directionnel. Enfin, le chapitre est conclu par un bilan scientifique en [Section 4.4](#).

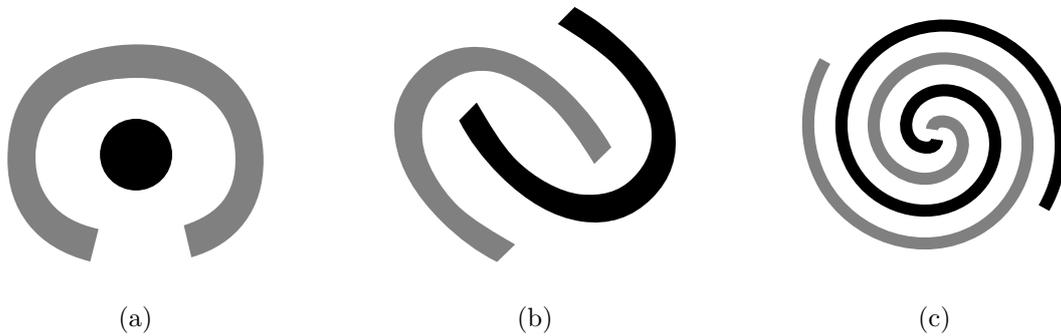


Figure 4.1 – Exemples caractéristiques de situations où la configuration spatiale des deux objets présente une certaine forme d'enlacement ou d'entrelacement. Pour chaque exemple, A correspond à l'objet noir et B correspond à l'objet gris.

4.1 Relations d'enlacement et d'entrelacement

Dans ce chapitre, nous nous attachons à la modélisation de nouvelles relations spatiales, que nous proposons de nommer *enlacement* et *entrelacement*. Ces nouveaux concepts ont pour objectif de désigner des configurations spatiales où les objets peuvent être imbriqués les uns dans les autres. Quelques exemples caractéristiques de ce type de configurations sont illustrés en Figure 4.1. En dépit du fait que ces configurations spatiales soient appréhendées relativement facilement par la perception humaine, elles restent, pour la plupart du temps, définies d'une manière vague et ambiguë, ce qui les rend complexes à caractériser d'un point de vue informatique. Par conséquent, nous proposons de clarifier et de fixer le sens de ces termes pour désigner ces concepts sémantiques, ceci afin de guider notre modélisation quantitative.

La relation spatiale d'*enlacement* caractérise et généralise l'idée qu'un objet A (objet *argument*) est, d'une certaine façon, entouré ou capturé par un autre objet B (objet *réfèrent*). Ainsi, une situation d'enlacement se présente à partir du moment où un objet se situe dans une concavité formée par un autre objet. Selon la configuration spatiale envisagée, la proposition « A est enlacé par B » peut être valide ou non, et ceci avec un certain degré. Il s'agit d'une relation spatiale relative, en ce sens qu'elle est toujours définie pour un objet par rapport à un autre objet. Notons également que cette relation n'est pas symétrique : l'objet A peut être enlacé par l'objet B , mais l'inverse n'est pas nécessairement vrai.

Compte tenu de cette définition, la relation spatiale d'*entrelacement* désigne alors un enlacement *mutuel* de deux objets. L'entrelacement englobe donc l'idée que deux objets sont imbriqués, ou bien s'entourent l'un et l'autre, mais pas nécessairement respectivement de la même façon. Cette fois-ci, il ne s'agit pas d'une définition relative, mais plutôt d'une mesure absolue de la configuration spatiale des deux objets. On utilisera donc plutôt des propositions du type « A et B sont entrelacés », selon une certaine appréciation qui reste à préciser.

Pour mieux saisir ces définitions, nous pouvons considérer à nouveau les exemples illustratifs de la Figure 4.1. Pour la configuration (a), la proposition « A est enlacé par B » serait ainsi

4.2. Enlacement directionnel entre objets

valide à un certain degré, mais l'opposé n'est pas vérifié. Les objets ne sont donc pas entrelacés. Il s'agit typiquement d'une situation dite d'*entourement*, que l'on peut considérer comme un cas particulier de l'enlacement. Nous reviendrons notamment sur ce cas particulier dans la suite de ce manuscrit. Dans la situation (b), l'enlacement entre les deux objets est cette fois-ci mutuel : certaines zones de l'objet A sont imbriquées dans l'objet B , et réciproquement. Par conséquent, selon les définitions sémantiques proposées, nous pourrions donc considérer que les objets sont entrelacés, dans une certaine mesure. Enfin, pour la configuration (c), nous pourrions dire que les deux objets sont très fortement entrelacés, en cela qu'ils s'enlacent mutuellement dans toutes les directions et avec plusieurs degrés de profondeur.

Notre objectif ici est donc de décrire de manière précise des configurations spatiales d'enlacement et d'entrelacement, pour des objets bidimensionnels ayant des formes géométriques arbitraires. Ces objets peuvent notamment être constitués de plusieurs composantes connexes. De telles descriptions devraient notamment permettre de discriminer différents types de configurations d'enlacement et d'entrelacement. Par conséquent, nous proposons d'appréhender le problème d'une manière similaire à d'autres tâches classiques de reconnaissance de formes, induisant par exemple des contraintes d'invariance par rapport à différentes transformations géométriques telles que les translations, les rotations ou les homothéties. Notre but est donc de construire un descripteur de position relative, conforme à l'intuition humaine, rapide à calculer et qui soit applicable pour des objets arbitrairement complexes. De telles exigences peuvent être utiles dans de nombreuses tâches liées à l'analyse et l'interprétation d'images, et constituent la base sur laquelle notre approche est construite.

4.2 Enlacement directionnel entre objets

Dans cette section, nous introduisons tout d'abord le formalisme utilisé pour représenter des objets dans le plan. À partir de ce formalisme, nous présentons ensuite les définitions proposées pour quantifier l'enlacement relatif de deux objets d'un point de vue directionnel, en spécifiant d'abord le calcul le long d'une droite orientée, puis en l'intégrant ce calcul le long de toutes les droites orientées traversant les objets. Enfin, nous énonçons les différentes propriétés d'invariance liées à ces définitions.

4.2.1 Notation et définitions préliminaires

Objets bidimensionnels

Dans un espace Euclidien, un objet A bidimensionnel peut être défini comme un ensemble non-vide et borné de points $A \subset \mathbb{R}^2$. De manière équivalente, cet objet A peut être défini par une fonction caractéristique $\mathbb{1}_A : \mathbb{R}^2 \rightarrow \{0, 1\}$ qui associe à chaque point $(x, y) \in \mathbb{R}^2$ du plan une valeur binaire représentant l'appartenance du point à l'objet. Cette définition d'un objet peut naturellement être étendue à une modélisation floue, où un objet A est alors défini par un sous-ensemble flou représenté par une fonction d'appartenance $\mu_A : \mathbb{R}^2 \rightarrow [0, 1]$. De manière plus générale, il est possible de modéliser les objets comme des fonctions à supports compacts

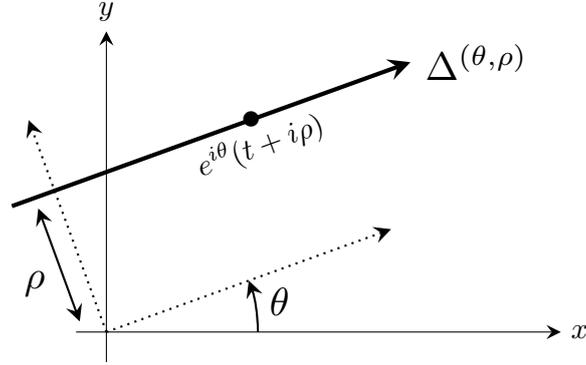


Figure 4.2 – Illustration d'une droite orientée $\Delta^{(\theta, \rho)}$ définie dans le plan par un angle d'orientation θ et une distance à l'origine ρ . Un point localisé sur une telle droite peut être défini par un nombre complexe $e^{i\theta}(t + i\rho)$ où $t \in \mathbb{R}$ représente sa coordonnée relative le long de $\Delta^{(\theta, \rho)}$.

de \mathbb{R}^2 vers \mathbb{R} , permettant ainsi des manipulations analytiques plus aisées. Par conséquent, dans un souci de généralité, dans la suite de ce chapitre, un objet A sera modélisé par sa fonction représentative notée $f_A : \mathbb{R}^2 \rightarrow \mathbb{R}$.

Coupes longitudinales

Dans la suite, le plan réel \mathbb{R}^2 et le plan complexe \mathbb{C} sont utilisés de manière équivalente par l'isomorphisme d'espaces vectoriels $(x, y) \in \mathbb{R}^2 \mapsto (x + iy) \in \mathbb{C}$.

Soit $\theta \in \mathbb{R}$ un angle et $\rho \in \mathbb{R}$ une distance depuis l'origine. Nous définissons une droite orientée d'angle θ et à l'altitude ρ par l'ensemble suivant :

$$\Delta^{(\theta, \rho)} = \{e^{i\theta}(t + i\rho), t \in \mathbb{R}\}, \quad (4.1)$$

où $t \in \mathbb{R}$ représente la coordonnée d'un point relativement à la droite $\Delta^{(\theta, \rho)}$. Un exemple de droite orientée définie de cette façon est proposé en Figure 4.2.

Soit $\Delta^{(\theta, \rho)}$ une droite orientée, et A un objet. Le sous-ensemble $A \cap \Delta^{(\theta, \rho)}$ représente une coupe unidimensionnelle de l'objet A , que nous appelons *coupe longitudinale*. Dans le cas d'objets binaires, une telle coupe longitudinale de A est soit vide (lorsque la droite orientée ne traverse pas l'objet) soit composée d'un nombre fini de segments, comme l'illustre par exemple la Figure 4.3. Dans le cas continu, et de manière similaire à la notation fonctionnelle des objets introduite précédemment, une coupe longitudinale de A le long de la droite orientée $\Delta^{(\theta, \rho)}$ peut être définie à l'aide d'une fonction notée $f_A^{(\theta, \rho)}$ telle que :

$$\begin{aligned} f_A^{(\theta, \rho)} : \mathbb{R} &\longrightarrow \mathbb{R} \\ t &\longmapsto f_A(e^{i\theta}(t + i\rho)), \end{aligned} \quad (4.2)$$

où t représente un point unidimensionnel sur la droite $\Delta^{(\theta, \rho)}$ et $f_A^{(\theta, \rho)}(t)$ donne la valeur représentative pour ce point par rapport à l'objet A .

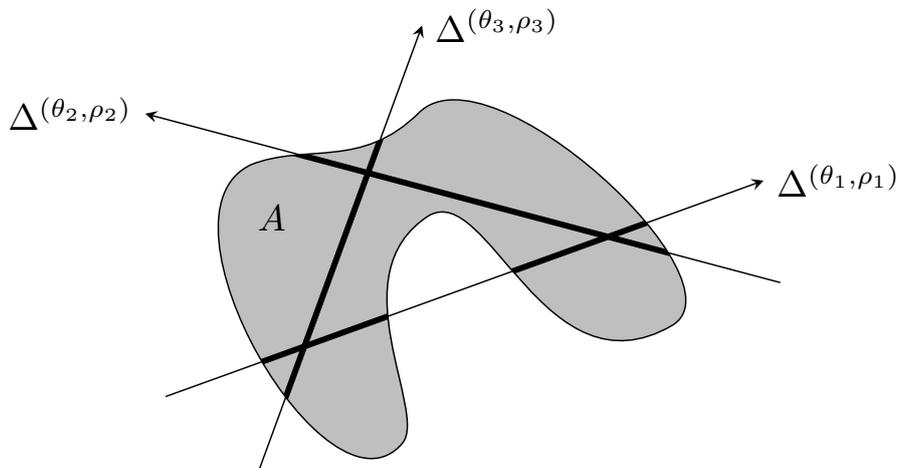


Figure 4.3 – Différentes coupes longitudinales d’une objet binaire A . Une droite orientée $\Delta^{(\theta, \rho)}$ découpe l’objet soit en un ensemble vide, soit en un ensemble fini de segments.

4.2.2 Enlacement le long d’une droite

Soit (A, B) un couple d’objets. Notre objectif est de décrire quantitativement dans quelle mesure l’objet A est enlacé par l’objet B . Afin de décrire une telle configuration, nous proposons de considérer la disposition spatiale des différents points $a_i \in A$ et $b_i \in B$ des objets comme des votes, ou des arguments, en faveur de la proposition « A est enlacé par B ». Dans notre cas, il s’agit en fait de comptabiliser les occurrences de points de l’objet A qui sont situés *entre* des points de l’objet B . Cependant, il peut être difficile d’identifier précisément de telles occurrences dans le plan. Ainsi, nous proposons de manipuler les points des objets dans un contexte unidimensionnel, où la notion *entre* peut être définie avec moins d’ambiguïtés. Cette idée de manipuler des données unidimensionnelles est notamment inspirée du modèle de calcul des histogrammes de forces [Matsakis et Wendling, 1999].

Afin d’illustrer plus intuitivement le principe de cette stratégie, nous nous focalisons dans un premier temps sur le cas où les objets sont binaires. Soit $\Delta^{(\theta, \rho)}$ une droite orientée. Comme expliqué précédemment, une telle droite coupe les objets en deux sous-ensembles de points de A et B , où chaque point peut être défini par sa coordonnée relative le long de la droite $\Delta^{(\theta, \rho)}$. Ainsi, une occurrence élémentaire de l’enlacement de A par B peut être définie comme un triplet ordonné de points $(b_i, a_k, b_j) \in B \times A \times B$ avec $b_i < a_k < b_j$ le long de la droite.

Ce principe est notamment illustré par la Figure 4.4 pour deux configurations différentes. Pour la première droite orientée $\Delta^{(\theta, \rho)}$, le seul argument qui peut être mis en faveur de la proposition « A est enlacé par B » est le triplet ordonné (b_1, a_1, b_2) . Remarquons par la même occasion qu’il n’y a pas d’argument qui puisse être mis en faveur de l’enlacement de B par A . Pour la deuxième droite, qui est orientée dans la direction opposée, plusieurs occurrences élémentaires de l’enlacement de A par B peuvent être identifiées. L’ensemble des triplets ordonnés $\{(b_1, a_1, b_2), (b_1, a_1, b_3), (b_1, a_2, b_3), (b_2, a_2, b_3)\}$ constitue l’ensemble des arguments à mettre en

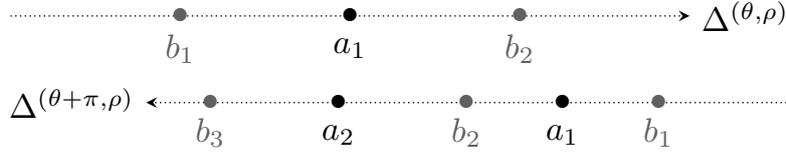


Figure 4.4 – Illustration de l'enlacement entre des points le long de droites orientées. Le long de $\Delta^{(\theta, \rho)}$, le triplet ordonné (b_1, a_1, b_2) est le seul argument qui puisse être mis en faveur de la proposition « A est enlacé par B ». Le long de la droite $\Delta^{(\theta+\pi, \rho)}$, les arguments qui peuvent être mis en faveur de la proposition sont représentés par l'ensemble des triplets ordonnés $\{(b_1, a_1, b_2), (b_1, a_1, b_3), (b_1, a_2, b_3), (b_2, a_2, b_3)\}$.

faveur de « A est enlacé par B » le long de cette droite. Pour la relation opposée « B est enlacé par A », on retrouve notamment l'argument (a_1, b_2, a_2) .

Par conséquent, afin de mesurer globalement l'enlacement de A par B le long d'une droite orientée $\Delta^{(\theta, \rho)}$, nous proposons de comptabiliser le nombre d'occurrences élémentaires de points de A situés entre des points de B sur cette droite. En d'autres termes, nous souhaitons capturer les parties A qui sont localisées simultanément *avant* et *après* des parties de B .

Nous reprenons maintenant les notations fonctionnelles introduites précédemment pour manipuler les coupes longitudinales. Ainsi, la quantité de l'objet A qui est localisée *après* un point x le long d'une droite orientée $\Delta^{(\theta, \rho)}$ est donnée par :

$$\int_x^{+\infty} f_A^{(\theta, \rho)}(y) dy. \quad (4.3)$$

Géométriquement parlant, si l'objet A est binaire, cette intégrale représente la somme des longueurs des segments de A sur l'intervalle $[x, +\infty[$ de la coupe longitudinale considérée. Par conséquent, la quantité de l'objet A localisée *après* des points d'un objet B est donnée par :

$$\int_{-\infty}^{+\infty} f_B^{(\theta, \rho)}(x) \int_x^{+\infty} f_A^{(\theta, \rho)}(y) dy dx. \quad (4.4)$$

De manière symétrique, la quantité de l'objet A localisée *avant* des points de B est obtenue par :

$$\int_{-\infty}^{+\infty} f_A^{(\theta, \rho)}(x) \int_x^{+\infty} f_B^{(\theta, \rho)}(y) dy dx. \quad (4.5)$$

Afin de prendre en considération à la fois les parties de B qui sont situées *avant* et *après* des parties de A , il convient de combiner ces deux quantités. Ceci nous conduit à introduire la définition suivante.

Définition 1. Soit f et g deux fonctions mesurables bornées à support compact de \mathbb{R} dans \mathbb{R} . L'enlacement de f par g est défini par :

$$E(f, g) = \int_{-\infty}^{+\infty} g(x) \int_x^{+\infty} f(y) \int_y^{+\infty} g(z) dz dy dx. \quad (4.6)$$

4.2. Enlacement directionnel entre objets

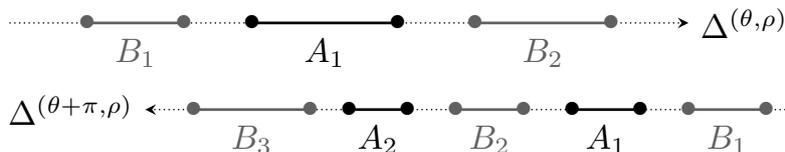


Figure 4.5 – Illustration de l'enlacement entre des segments le long de droites orientées (les A_i et B_i désignent les longueurs respectives des segments). Le long de la première droite, l'enlacement des segments de A par les segments de B dans la direction θ est donné par le produit $B_1A_1B_2$. Le long de la deuxième droite, l'enlacement des segments de A par les segments de B dans la direction $\theta + \pi$ est donné par $(B_1A_1B_2) + (B_1A_1B_3) + (B_1A_2B_3) + (B_2A_2B_3)$.

Cette définition générique entre deux fonctions réelles peut ainsi être appliquée à toutes coupes longitudinales $f_A^{(\theta, \rho)}$ et $f_B^{(\theta, \rho)}$. Ainsi, la valeur $E(f_A^{(\theta, \rho)}, f_B^{(\theta, \rho)})$ représente l'enlacement global de A par B le long de la droite orientée $\Delta^{(\theta, \rho)}$. Dans le cas d'objets binaires, cela correspond au nombre total de triplets orientés qui peuvent être envisagés comme des arguments à mettre en faveur de la proposition « A est enlacé par B » dans la direction θ .

Afin d'illustrer cette définition, la Figure 4.5 étend les configurations de la Figure 4.4 à des segments plutôt que des points (les A_i et B_i désignent alors les longueurs respectives des segments). Ainsi, l'enlacement des segments de A par les segments de B le long de la première droite $\Delta^{(\theta, \rho)}$ est donné par $B_1A_1B_2$. Le long de la deuxième droite, l'enlacement des segments de A par les segments de B dans la direction $\theta + \pi$ est donné par $(B_1A_1B_2) + (B_1A_1B_3) + (B_1A_2B_3) + (B_2A_2B_3)$. Symétriquement, l'enlacement des segments de B par les segments de A est donné par $A_1B_2A_2$.

4.2.3 Enlacement global entre objets

Dans la section précédente, nous avons introduit la notion d'enlacement entre fonctions réelles, qui peut être appliquée aux fonctions caractéristiques d'objets prises le long de coupes longitudinales. Dans la suite, nous détaillons comment cette définition peut être étendue pour manipuler les objets dans leur ensemble.

Soit $\theta \in \mathbb{R}$ un angle d'orientation. L'ensemble des droites orientées parallèles $\{\Delta^{(\theta, \rho)}, \rho \in \mathbb{R}\}$ dans cette direction traverse un objet en un ensemble de fonctions de coupes longitudinales. Ainsi, pour mesurer l'enlacement d'un objet par rapport à un autre dans cette direction, il est possible d'agréger les valeurs d'enlacement obtenues pour chacune de ces coupes longitudinales. Ce principe est exprimé par la définition suivante.

Définition 2. Soit $\theta \in \mathbb{R}$ un angle d'orientation, et soit A et B deux objets. L'enlacement global de A par B dans la direction θ est défini par :

$$E_{AB}(\theta) = \int_{-\infty}^{+\infty} E(f_A^{(\theta, \rho)}, f_B^{(\theta, \rho)}) d\rho. \quad (4.7)$$

Le principe de cette définition peut être illustré par la Figure 4.6. Dans le cas binaire, cette

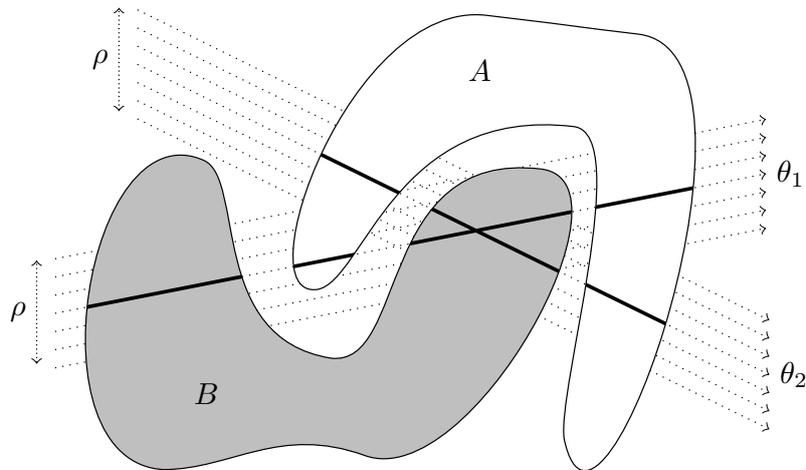


Figure 4.6 – Illustration du calcul de l'enlacement pour deux objets A et B . L'intégration des valeurs d'enlacement unidimensionnelles le long de l'ensemble des droites parallèles orientées traversant les objets permet de mesurer l'enlacement global dans cette direction.

définition correspond alors au nombre total de triplets ordonnés de points que l'on peut mettre en faveur de « A est enlacé B » le long des coupes longitudinales traversant les objets dans cette direction. De manière intuitive, nous pouvons l'interpréter comme la quantité de l'objet B que l'objet A devrait traverser si ce dernier était translaté dans la direction θ , et ceci par rapport à la quantité de l'objet B localisée dans la direction opposée.

Par conséquent, la valeur scalaire qui résulte de cette définition est essentiellement directionnelle, et dépend de la configuration spatiale entre les deux objets. Elle est fondée sur la manière dont se répartissent les points des différents objets dans le plan, selon le point de vue des droites parallèles orientées dans la direction θ . Par exemple, si aucun point de A n'est localisé entre des points de B dans la direction θ , alors nous aurons $E_{AB}(\theta) = 0$. Au contraire, si de larges segments de A sont capturés par des segments de B le long de la direction θ , alors la valeur de $E_{AB}(\theta)$ augmentera.

Afin de mieux illustrer le principe de la Définition 2, la Figure 4.7 montre les profils d'enlacement obtenus pour différentes directions $\theta \in [0, 2\pi]$ sur une série d'exemples synthétiques. Les objets de cette série d'exemples sont constitués de plusieurs composantes connexes, ce qui permet d'illustrer le comportement des profils E_{AB} et E_{BA} par rapport à des configurations de plus en plus complexes. Par souci de simplicité, les configurations présentées dans ces exemples sont essentiellement horizontales : les profils sont nuls dans la plupart des directions, avec deux pics isolés et centrés autour des angles $\theta = 0$ et $\theta = \pi$. La situation d'enlacement la plus simple est celle représentée en (a) : seulement A est enlacé par B dans les directions 0 et π (tandis que E_{AB} est nul pour toutes les directions). En (b), l'enlacement des deux objets est maintenant mutuel : une des composantes connexes de B est enfermée entre deux composantes connexes de A , et réciproquement. Pour cette raison, les deux profils d'enlacement E_{AB} et E_{BA} sont ici identiques. En (c), une composante connexe de B est située entre trois composantes connexes

4.2. Enlacement directionnel entre objets

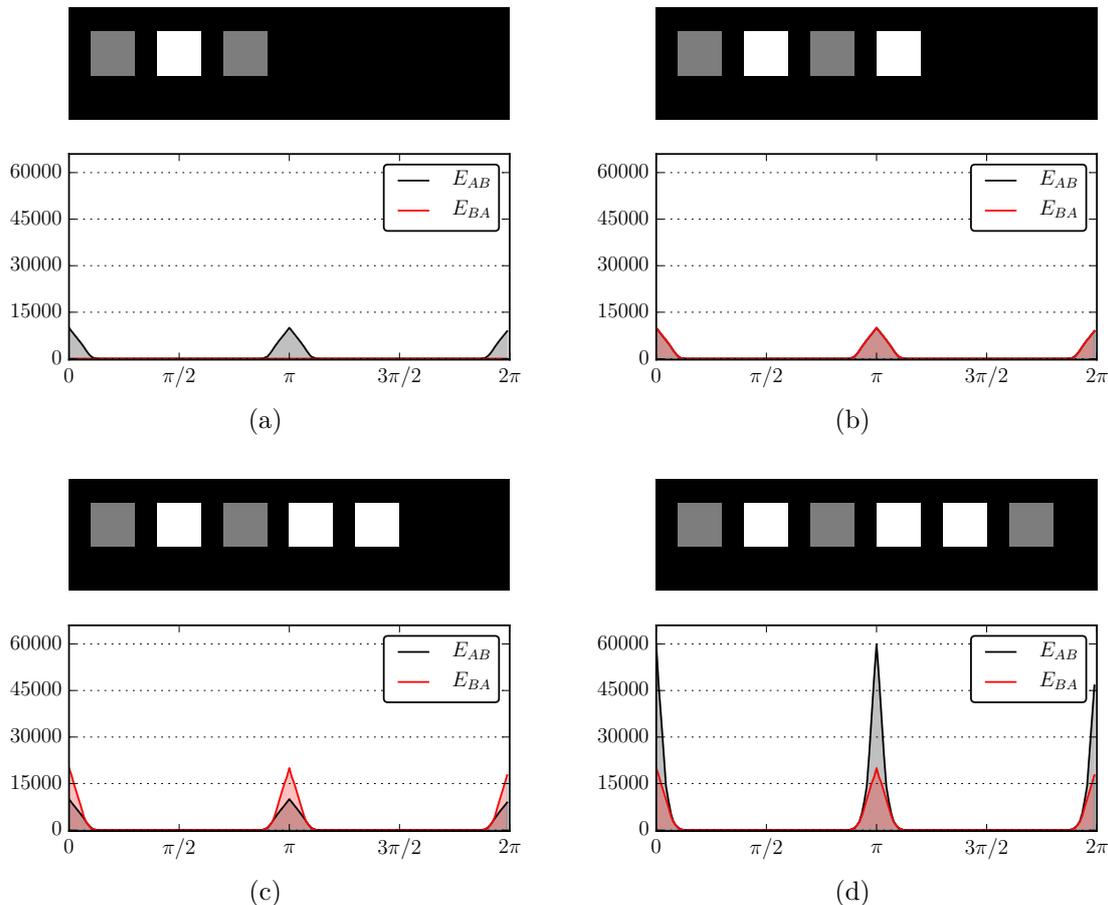


Figure 4.7 – Évolution des profils d’enlacement directionnels E_{AB} et E_{BA} sur une série d’exemples simples où les objets sont constitués de plusieurs composantes connexes (en blanc : objet A ; en gris : objet B). Ces exemples illustrent des configurations spatiales de plus en plus complexes dans la direction horizontale. Sur ces exemples, chaque carré a une aire de 10×10 pixels. En (b), les deux courbes se confondent.

de A , résultant en un profil E_{BA} plus élevé que E_{AB} . Enfin, pour la configuration (d), le pic plus élevé du profil E_{AB} est dû au fait que les trois composantes connexes de A sont entre les composantes de B qui sont sur les extrémités. Nous pouvons également remarquer que ces deux extrémités n’ont pas d’influence sur E_{BA} , et par conséquent E_{BA} est identique à la configuration précédente, car la même composante de B est enlacée par A .

Ces exemples permettent ainsi de souligner la nécessité de considérer les deux profils d’enlacement E_{AB} et E_{BA} pour décrire précisément et discriminer les différentes configurations spatiales possibles entre un couple d’objets (A, B). Ceci nous conduit à la notion d’entrelacement entre objets, qui sera étudiée plus en détail dans la [Section 4.3](#).

4.2.4 Propriétés d'invariance

Nous énonçons ici les différentes propriétés d'invariance des profils d'enlacement directionnels qui peuvent être obtenus à partir de la Définition 2. Les démonstrations mathématiques de ces propriétés peuvent être trouvées en [Annexe B](#) de ce manuscrit.

Propriété 1 (Périodicité). *Soit A et B deux objets. L'enlacement directionnel E_{AB} est périodique et de période π :*

$$\forall \theta \in \mathbb{R}, E_{AB}(\theta) = E_{AB}(\theta + k\pi), k \in \mathbb{Z}. \quad (4.8)$$

Propriété 2 (Translation). *Soit A et B deux objets. L'enlacement directionnel E_{AB} est invariant par rapport aux translations. En notant T_v la translation par un vecteur $v \in \mathbb{C} \simeq \mathbb{R}^2$, alors nous avons :*

$$E_{T_v(A)T_v(B)}(\theta) = E_{AB}(\theta). \quad (4.9)$$

Propriété 3 (Rotation). *Soit A et B deux objets. L'enlacement directionnel E_{AB} est co-variant par rapport aux rotations. Soit $\alpha \in \mathbb{R}$ un angle et soit R_α la rotation d'angle α associée, alors nous avons :*

$$E_{R_\alpha(A)R_\alpha(B)}(\theta) = E_{AB}(\theta - \alpha). \quad (4.10)$$

Propriété 4 (Homothétie 1). *Soit A et B deux objets. L'enlacement directionnel E_{AB} est quasi-invariant par rapport aux homothéties. Soit $\lambda \in \mathbb{R}$ et S_λ l'homothétie de facteur λ associée, alors nous avons :*

$$E_{S_\lambda(A)S_\lambda(B)}(\theta) = \lambda^4 E_{AB}(\theta). \quad (4.11)$$

La π -périodicité de $E_{AB}(\theta)$ soulignée par la Propriété 1 est intéressante d'un point de vue algorithmique, car cela nous permet de restreindre le calcul au seul intervalle $[0, \pi]$ pour considérer toutes les directions. Les autres propriétés garantissent que l'enlacement directionnel est invariant par rapport aux translations, que les rotations induisent seulement des décalages circulaires des profils, et que leur comportement est linéaire par rapport aux facteurs d'homothéties. Ces différentes propriétés d'invariance sont bien souvent utiles dans de nombreuses tâches applicatives liées à la reconnaissance de formes.

4.3 Histogrammes d'enlacement et d'entrelacement

Dans la suite, nous présentons comment généraliser le principe de la Définition 2, dans le but de construire des descripteurs d'enlacement et d'entrelacement, qui pourront alors être utilisés dans différentes tâches d'analyse et d'interprétation d'images.

4.3.1 Représentations par histogrammes

Afin de mieux refléter l'enlacement global d'un objet par rapport à un autre, la valeur d'enlacement obtenue pour une direction donnée doit être relative à l'aire globale des deux objets considérés. De cette façon, il est alors possible de comparer l'enlacement de différents

4.3. Histogrammes d'enlacement et d'entrelacement

couples d'objets ayant la même aire, mais des formes différentes. Nous notons alors $\|A\|_1$ l'aire d'un objet A , donnée par :

$$\|A\|_1 = \iint_{\mathbb{R}^2} |f_A(x, y)| dx dy. \quad (4.12)$$

Le descripteur directionnel d'enlacement est alors exprimé par la définition suivante.

Définition 3. *Soit A et B deux objets. Le descripteur directionnel d'enlacement de A par rapport à B est défini par la fonction suivante :*

$$\begin{aligned} \mathcal{E}_{AB} : \mathbb{R} &\longrightarrow \mathbb{R} \\ \theta &\longmapsto \frac{E_{AB}(\theta)}{\|A\|_1 \|B\|_1}. \end{aligned} \quad (4.13)$$

Ainsi, le descripteur d'enlacement \mathcal{E}_{AB} prend la forme d'une fonction, qui associe à chaque angle θ la valeur d'enlacement de A par B dans cette direction, normalisée par l'aire des objets. Par construction, chaque valeur résultante de cette fonction correspond au nombre total d'occurrences d'enlacement élémentaires, obtenues en considérant les différentes coupes longitudinales des objets dans la direction θ . Le descripteur peut alors être interprété comme un histogramme circulaire, que nous proposons de nommer *\mathcal{E} -histogramme*.

Ce facteur de normalisation permet de mieux représenter l'idée intuitive de départ selon laquelle l'enlacement global est mesuré en agrégeant les occurrences locales d'enlacement entre des points le long de droites orientées. Il permet également d'introduire la propriété d'invariance suivante (dont la démonstration est donnée en [Annexe B](#)), qui découle notamment de la [Propriété 4](#).

Propriété 5 (Homothétie 2). *Soit deux objets A et B . L'histogramme d'enlacement \mathcal{E}_{AB} est invariant par rapport aux homothéties. Soit $\lambda \in \mathbb{R}$ et S_λ l'homothétie de facteur λ associée, alors nous avons :*

$$\mathcal{E}_{S_\lambda(A)S_\lambda(B)}(\theta) = \mathcal{E}_{AB}(\theta). \quad (4.14)$$

Les [Propriétés 1](#), [2](#) et [3](#) précédemment énoncées pour l'enlacement directionnel E_{AB} sont naturellement conservées pour le descripteur d'enlacement \mathcal{E}_{AB} . Avec la [Propriété 5](#) déduite de la [Propriété 4](#), nous pouvons conclure que le \mathcal{E} -histogramme entre deux objets est invariant par rapport aux translations et aux homothéties, et quasi-invariant par rapport aux rotations (une rotation d'angle α implique un décalage circulaire de α dans le descripteur).

L'histogramme d'enlacement \mathcal{E}_{AB} permet de décrire quantitativement dans quelle mesure un objet A est enlacé par un autre objet B . Il s'agit donc d'un *descripteur de position relative* (il décrit, d'une certaine façon, la position d'un objet par rapport à un autre). Cependant, le descripteur \mathcal{E}_{AB} seul ne permet pas de décrire complètement la configuration spatiale entre le couple d'objets (A, B) . En effet, comme nous l'avons déjà souligné avec la [Figure 4.7](#), l'autre descripteur \mathcal{E}_{BA} est complémentaire de \mathcal{E}_{AB} et doit être considéré également pour caractériser toute la configuration. Par ailleurs, une combinaison de ces deux descripteurs peut fournir une description de l'entrelacement des deux objets, défini comme un enlacement mutuel. Nous proposons donc de les combiner, à partir de la définition suivante.

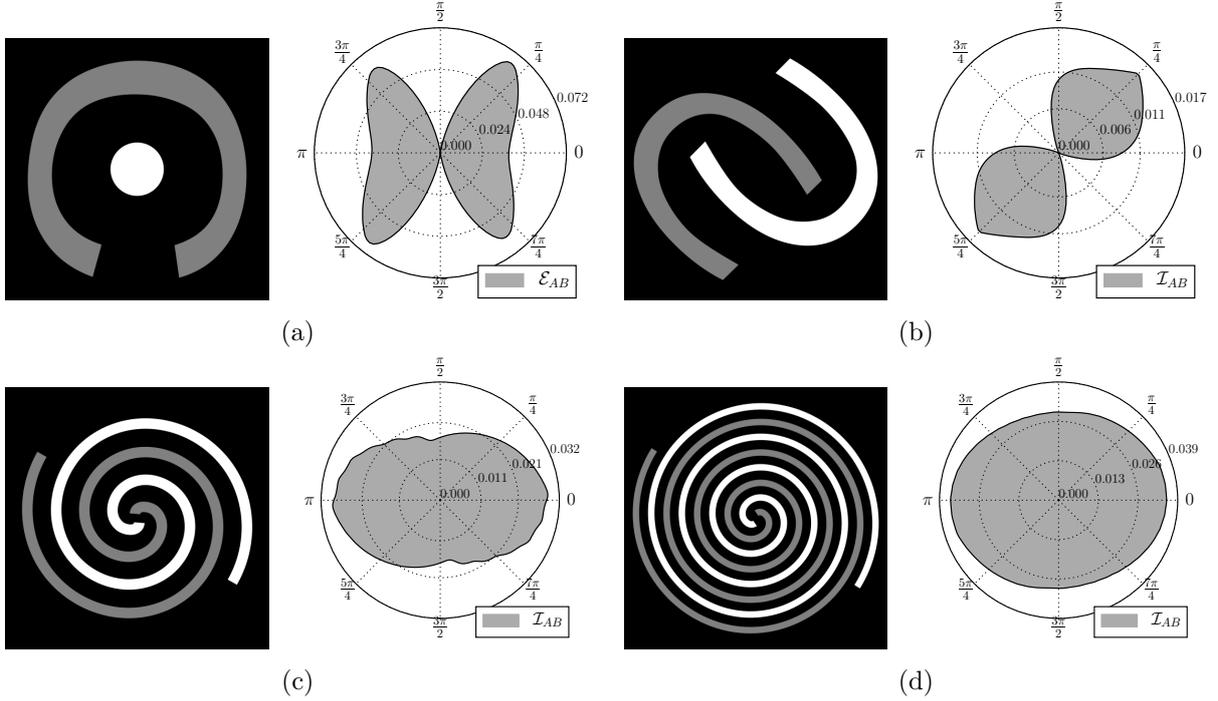


Figure 4.8 – Représentations polaires des descripteurs directionnels (a) d'enlacement \mathcal{E}_{AB} , et (b, c, d) d'entrelacement \mathcal{I}_{AB} , obtenus pour différentes configurations spatiales entre deux objets (en blanc : objet A ; en gris : objet B).

Définition 4. Soit (A, B) un couple d'objets. Le descripteur directionnel d'entrelacement de (A, B) est donné par la moyenne harmonique entre leur couple de descripteurs d'enlacement respectifs $(\mathcal{E}_{AB}, \mathcal{E}_{BA})$:

$$\mathcal{I}_{AB}(\theta) = \frac{2 \mathcal{E}_{AB}(\theta) \mathcal{E}_{BA}(\theta)}{\mathcal{E}_{AB}(\theta) + \mathcal{E}_{BA}(\theta)}. \quad (4.15)$$

De la même manière que pour le descripteur d'enlacement, le descripteur d'entrelacement peut être interprété comme un histogramme circulaire, alors nommé \mathcal{I} -histogramme. Enfin, par construction, les différentes propriétés d'invariance du descripteur d'enlacement sont conservées pour le descripteur d'entrelacement.

4.3.2 Exemples illustratifs

La Figure 4.8 illustre des représentations polaires de descripteurs directionnels d'enlacement et d'entrelacement pour les différentes configurations précédemment introduites en Figure 4.1. Ces représentations polaires permettent de visualiser la distribution de l'enlacement (ou de l'entrelacement) en fonction des différentes directions. Dans un souci de visualisation, les descripteurs sont représentés ici sur tout le cercle trigonométrique, bien qu'en pratique la moitié du cercle soit suffisante car les descripteurs sont π -périodiques.

4.3. Histogrammes d'enlacement et d'entrelacement

Pour la configuration (a), le descripteur d'enlacement \mathcal{E}_{AB} est représenté (\mathcal{E}_{BA} est nul partout). Nous pouvons constater que l'objet A est enlacé dans quasiment toutes les directions, sauf autour de l'axe vertical. D'un point de vue intuitif, ceci traduit le fait que l'objet A peut s'échapper de la situation sans avoir à traverser B le long de l'axe vertical (c'est-à-dire en se déplaçant vers le bas). Par ailleurs, le fait que le descripteur \mathcal{E}_{BA} soit nul partout correspond bien au fait que B n'est pas enlacé par A . Pour les configurations (b), (c) et (d), le descripteur d'entrelacement \mathcal{I}_{AB} est représenté (c'est-à-dire la moyenne harmonique des deux descripteurs \mathcal{E}_{AB} et \mathcal{E}_{BA}). En (b), nous pouvons constater une orientation dominante sur l'axe diagonal. Le long de l'autre diagonale, l'entrelacement est nul, ce qui est conforme au fait que les deux objets peuvent s'échapper le long de cet axe sans avoir à se traverser. Selon cette logique, on obtient un entrelacement globalement élevé pour les spirales de la configuration (c), et un entrelacement encore plus élevé pour la configuration (d).

4.3.3 Complexité algorithmique

Les descripteurs directionnels d'enlacement et d'entrelacement ont été définis dans les sections précédentes dans un cadre continu, en utilisant des notations analytiques pour manipuler les objets. En particulier cela nous a permis de généraliser la notion d'objet, et de mettre l'accent sur les différentes propriétés d'invariance des descripteurs. Ci-après, nous présentons quelques considérations algorithmiques pour calculer les descripteurs d'enlacement et d'entrelacement dans un contexte discret, où les objets sont représentés par des ensembles de pixels dans des images numériques de tailles finies.

Considérons une image carrée composée de N pixels (donc de taille $\sqrt{N} \times \sqrt{N}$), où A et B sont définis comme des objets binaires. Dans une direction θ fixée, le calcul de $\mathcal{E}_{AB}(\theta)$ nécessite de générer l'ensemble des droites parallèles $\{\Delta^{(\theta,\rho)}, \rho \in \mathbb{R}\}$ traversant l'image. En pratique, une droite orientée $\Delta^{(\theta,\rho)}$ initiale est tracée à partir de l'algorithme de géométrie discrète de Bresenham [Bresenham, 1965]. Les droites parallèles suivantes sont obtenues en effectuant des décalages de cette droite initiale, tout en s'assurant que chaque pixel de l'image soit bien considéré une et une seule fois.

Le long d'une droite de l'image, le nombre de pixels évolue en $\mathcal{O}(\sqrt{N})$, les pires cas étant obtenus par exemple pour la diagonale de l'image (pour la direction $\theta = \frac{\pi}{4}$) ou pour les directions horizontales et verticales. La triple intégrale de la Définition 1 est calculée en distribuant de manière appropriée les produits de longueurs de segments le long des coupes longitudinales des deux objets. Cette distribution de produits de longueurs de segments le long d'une droite orientée induit un complexité cubique, et donc un calcul en $\mathcal{O}(N\sqrt{N})$.

En pratique, afin de garantir l'isotropie des descripteurs, les longueurs des segments obtenus par les coupes longitudinales sont calculées en normalisant le nombre de pixels d'un segment par l'angle d'orientation (de sorte qu'un segment de n pixels orienté dans la direction $\frac{\pi}{4}$ ait une longueur de $n\sqrt{2}$). Ainsi, un segment composé de n pixels orienté dans une direction $\theta \in [0, \pi]$

a une longueur l_θ donnée par :

$$l_\theta = \begin{cases} n \cos^2(\theta) & \text{si } \theta \leq \frac{\pi}{4} \text{ ou } \theta \geq \frac{3\pi}{4}, \\ n \sin^2(\theta) & \text{sinon.} \end{cases} \quad (4.16)$$

Une fois la complexité le long d'une droite définie, le nombre de droites parallèles traversant l'image est borné par \sqrt{N} (ce pire cas étant obtenu pour les directions horizontales et verticales). Par conséquent, le calcul de $\mathcal{E}_{AB}(\theta)$ a une complexité en $\mathcal{O}(N^2)$. Enfin, le descripteur d'enlacement \mathcal{E}_{AB} calculé le long d'un ensemble de k directions discrètes a alors une complexité en $\mathcal{O}(kN^2)$. Notons qu'il s'agit d'une complexité en pire des cas, qui reste difficilement atteignable en pratique, car cela impliquerait des objets extrêmement distordus avec un nombre maximal d'alternances de pixels entre les objets A et B (par exemple selon un motif de damier).

Nous avons choisi de fixer le nombre de directions discrètes selon une stratégie similaire à celle qui est proposée pour les histogrammes de forces de [Matsakis et Wendling, 1999], c'est-à-dire à $k = 180$. Rappelons également que la π -périodicité du descripteur permet d'optimiser le calcul en considérant uniquement l'intervalle de directions $[0, \pi]$.

En utilisant cette stratégie de discrétisation, le temps d'exécution réel pour calculer le descripteur d'enlacement est d'environ 1 à 2 secondes pour une image de taille 500×500 , avec un processeur Intel Core i7 et une implémentation en Python sans optimisations particulières. Remarquons également que le calcul du descripteur est hautement parallélisable, en *multithreading*, ou dans un environnement distribué. Enfin, les objets définis par des fonctions d'appartenance floues peuvent être traités en considérant un nombre fini m de valeurs d'appartenance, puis en utilisant les schémas classiques de fusion d'ensembles flous, tels que la stratégie de la simple somme [Dubois et Jaulent, 1987] ou de la double somme [Krishnapuram *et al.*, 1993].

4.4 Bilan scientifique

Dans ce chapitre, nous nous sommes intéressés à la modélisation de configurations spatiales complexes, afin de décrire des situations où les objets peuvent être imbriqués les uns dans les autres. Ces situations sont généralement chargées d'ambiguïtés et souvent peu traitées d'un point de vue quantitatif. Afin de lever une part de cette ambiguïté, nous avons conceptualisé ces notions par de nouvelles relations spatiales. D'une part, la relation d'enlacement permet de généraliser le cas où un objet est plus ou moins localisé au sein des concavités formées par un autre objet. D'autre part, la relation d'entrelacement est alors définie comme un enlacement mutuel de deux objets. Nous avons alors proposé un modèle algorithmique efficace permettant de caractériser ce type de configurations spatiales à partir d'histogrammes directionnels. Ce modèle a été illustré à partir de plusieurs exemples, et nous avons également démontré les différentes propriétés d'invariance des descripteurs d'enlacement et d'entrelacement.

Le modèle proposé dans ce chapitre a été valorisé par la publication d'un article dans une revue internationale [Clément *et al.*, 2017a]. Ces travaux ont également été présentés dans une version préliminaire lors d'une conférence nationale [Clément *et al.*, 2016c].

Chapitre 5

Validations expérimentales : reconnaissance de configurations spatiales

Sommaire

5.1	Protocole expérimental général	58
5.1.1	Méthodes comparatives de référence	58
5.1.2	Classification et évaluation des résultats	59
5.2	Caractérisation de vaisseaux sanguins en ophtalmologie	61
5.2.1	Jeu de données	61
5.2.2	Protocole de validation	63
5.2.3	Résultats expérimentaux et discussion	65
5.3	Reconnaissance de lettrines décoratives	66
5.3.1	Jeu de données	66
5.3.2	Protocole de validation	67
5.3.3	Résultats expérimentaux et discussion	68
5.4	Recherche et classification d'objets géographiques urbains	68
5.4.1	Jeu de données	68
5.4.2	Protocole de validation	69
5.4.3	Résultats expérimentaux et discussion	73
5.5	Bilan scientifique	74

Ce chapitre présente les différentes validations expérimentales permettant d'illustrer l'intérêt et l'efficacité des descripteurs d'enlacement et d'entrelacement proposés dans le chapitre précédent. Ces expérimentations ont été menées dans trois domaines d'application différents liés à l'analyse d'images et la reconnaissance de formes. Tout d'abord, dans la [Section 5.1](#), nous détaillons le protocole expérimental mis en place pour les trois applications expérimentales développées dans ce chapitre. En [Section 5.2](#), une première application est proposée en imagerie

biomédicale pour la caractérisation de vaisseaux sanguins en ophtalmologie de rétines. Dans la [Section 5.3](#), la deuxième application concerne la reconnaissance de lettrines décoratives dans des images de documents anciens. Enfin, en [Section 5.4](#), la troisième application se focalise sur la classification d'objets urbains à partir d'images de télédétection. Nous concluons alors le chapitre par un bilan de ces expérimentations dans la [Section 5.5](#).

5.1 Protocole expérimental général

Cette première section permet d'introduire quelques aspects communs aux différentes applications expérimentales qui sont présentées dans la suite de ce chapitre. Plus spécifiquement, nous présentons de manière synthétique les méthodes utilisées pour comparer et juger de l'efficacité des descripteurs d'enlacement et d'entrelacement. Nous détaillons également le protocole de classification employé ainsi que les différents indices de performance calculés pour évaluer quantitativement la qualité des résultats obtenus.

5.1.1 Méthodes comparatives de référence

Nous présentons ici les différentes méthodes comparatives utilisées dans ces expérimentations. Nous proposons ainsi de comparer la performance des descripteurs d'enlacement et d'entrelacement avec deux autres types de descripteurs : un descripteur de formes (le descripteur GFD) qui peut permettre de reconnaître des formes complexes, et un descripteur de position relative (l'histogramme de forces) qui caractérise les relations spatiales directionnelles. Ces deux approches peuvent être considérées ici comme des méthodes de référence pour comparer la performance des descripteurs d'enlacement et d'entrelacement proposés. Soulignons ici que, du fait de la nouveauté des relations spatiales étudiées, il n'existe pas de méthodes dédiées pour la description de l'enlacement et de l'entrelacement.

Descripteurs GFD

Les descripteurs GFD (pour *Generic Fourier Descriptor*) sont des descripteurs de formes proposés par [\[Zhang et Lu, 2002\]](#). Ils sont notamment utilisés dans le standard MPEG-7. Les descripteurs GFD sont calculés en appliquant une transformation de Fourier 2D sur une image ayant été convertie en coordonnées polaires. Cette conversion en coordonnées polaires se fait en plaçant l'origine du repère au niveau du barycentre de l'image considérée, tout en normalisant les distances par le rayon maximum de la forme à partir de ce barycentre. Ceci permet d'obtenir une invariance aux translations ainsi qu'aux homothéties. La transformation de Fourier 2D, pour une image I définie dans le cas discret, est alors formulée de la manière suivante :

$$PF(\rho, \theta) = \sum_x \sum_y I(x, y) \exp \left(2i\pi \left(\frac{r(x, y)}{R} \rho + v(x, y) \theta \right) \right), \quad (5.1)$$

5.1. Protocole expérimental général

où $r(x, y)$ et $v(x, y)$ désignent respectivement le rayon et l'angle des coordonnées polaires du point (x, y) de l'image. Les paramètres de cette transformation sont fixés par $0 \leq \rho < R$ et $0 \leq \theta < T$, où R et T désignent respectivement la résolution radiale et la résolution angulaire. Suite à cette transformation, le descripteur GFD est alors donné par :

$$GFD(m, n) = \left\{ \frac{|PF(0, 0)|}{M_{11}}, \frac{|PF(0, 1)|}{|PF(0, 0)|}, \dots, \frac{|PF(m, n)|}{|PF(0, 0)|} \right\} \quad (5.2)$$

où m et n sont respectivement les fréquences radiales et angulaires, et où M_{11} désigne le moment d'ordre 1 de l'image considérée. Dans les expérimentations qui suivent, nous avons fixé les résolutions à $R = 4$ et $T = 9$ (conformément aux paramètres proposés par les auteurs), résultant finalement en des descripteurs GFD composés de 36 valeurs.

Histogrammes de forces

Les descripteurs utilisés comme deuxième méthode comparative sont les histogrammes de forces [Matsakis et Wendling, 1999] (notés également F-Histogrammes dans la suite). Les F-Histogrammes sont des descripteurs de position relative, déjà présentés durant la [Partie I](#) de ce manuscrit, qui sont dédiés à la caractérisation des relations spatiales directionnelles. Dans le cadre des expérimentations de ce chapitre, les histogrammes de forces ont été calculés avec une force constante (c'est-à-dire avec $r = 0$) et le long d'un ensemble de 180 directions discrètes dans l'intervalle $[0, 2\pi]$, qui correspond à un pas de 2 degrés.

Les histogrammes de forces constituent une référence de comparaison particulièrement pertinente pour juger de l'efficacité des histogrammes d'enlacement et d'entrelacement. En effet, les histogrammes de forces sont appropriés pour décrire les relations spatiales directionnelles usuelles, cependant, ils ne sont pas conçus pour distinguer les configurations spatiales plus complexes où les objets peuvent s'entrelacer.

5.1.2 Classification et évaluation des résultats

Dans la plupart des expérimentations qui seront présentées dans ce chapitre, nous avons mis en place un protocole de classification d'images pour évaluer la performance des descripteurs proposés dans différents contextes applicatifs. Dans un souci de comparabilité de ces résultats expérimentaux, nous avons choisi d'utiliser le même protocole de classification et d'évaluation pour les différentes applications. Celui-ci est explicité ci-après.

Protocole de classification

Nous considérons que l'on dispose d'un ensemble d'images, où chacune est associée à une étiquette correspondant à différentes classes ou catégories. Ces classes associées aux images constituent ce que l'on appelle la vérité terrain.

La méthode de classification employée ici est celle des k -plus-proches-voisins (abrégiée k -NN dans la suite, pour k -nearest neighbors). Nous avons choisi la distance euclidienne pour calculer

la dissimilarité entre les descripteurs. Pour classifier l'échantillon testé, nous avons fixé le nombre de voisins considérés par défaut à $k = 5$ (pour certaines évaluations, plus de voisins pourront être considérés). Ensuite, parmi les 5 plus proches voisins d'un échantillon de test, un vote pondéré en fonction de la distance euclidienne est effectué pour déterminer la classe qui lui est assignée.

Le protocole de validation croisée utilisé est le *leave-one-out* (*LOO*) : chaque échantillon de test est comparé au reste du jeu de données considéré comme base d'apprentissage, et ceci successivement pour tous les échantillons du jeu de données. Ce type de validation croisée est envisageable ici car les jeux de données qui seront utilisés ont une taille qui est raisonnable par rapport aux temps de calculs induits.

Évaluation des résultats

Une fois la validation croisée effectuée, nous calculons alors différents indicateurs de performance pour évaluer la qualité des résultats de classification obtenus. Pour chaque classe i du jeu de données considéré, les indicateurs suivants sont calculés :

- Le nombre de *vrais positifs* V_i^P (nombre d'échantillons bien reconnus de la classe i) ;
- Le nombre de *faux positifs* F_i^P (nombre d'échantillons faussement assignés à la classe i) ;
- Le nombre de *faux négatifs* F_i^N (nombre d'échantillons de la classe i qui n'ont pas été reconnus).

Nous pouvons alors calculer le taux de reconnaissance global qui correspond au nombre d'échantillons bien reconnus par rapport à l'ensemble des échantillons des n classes :

$$T_R = \frac{\sum_{i=1}^n V_i^P}{\sum_{i=1}^n (V_i^P + F_i^N)}. \quad (5.3)$$

Par ailleurs, pour chaque classe i du jeu de données considéré, nous pouvons également considérer les scores de précision P_i et de rappel R_i :

$$P_i = \frac{V_i^P}{V_i^P + F_i^P} \quad \text{et} \quad R_i = \frac{V_i^P}{V_i^P + F_i^N}. \quad (5.4)$$

Nous pouvons alors obtenir les scores globaux de précision P et de rappel R comme les moyennes des P_i et R_i sur l'ensemble des n classes :

$$P = \frac{1}{n} \sum_{i=1}^n P_i \quad \text{et} \quad R = \frac{1}{n} \sum_{i=1}^n R_i. \quad (5.5)$$

Enfin, nous proposons également de calculer des courbes de précision-rappel pour évaluer de manière plus détaillée la qualité des résultats. Afin de calculer ces courbes, nous faisons varier le seuil du nombre de voisins considérés par l'algorithme de classification k -NN, et pour chaque seuil, nous calculons les scores de rappel R et de précision $P(R)$ correspondants. Nous fixons alors un certain nombre de seuils de rappels uniformément espacés $R \in \{0, 0.1, \dots, 1\}$, et nous

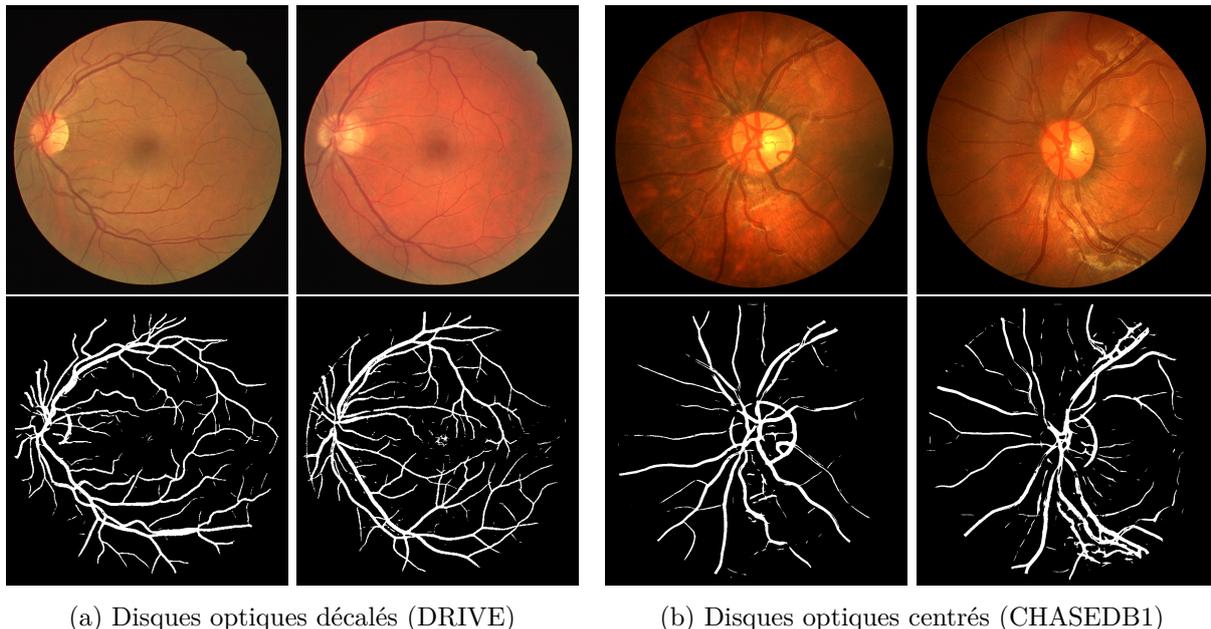


Figure 5.1 – Exemples d’images de fonds d’œil provenant des jeux de données DRIVE [Staal *et al.*, 2004] et CHASEDB1 [Fraz *et al.*, 2012], avec leurs segmentations automatiques respectives obtenues par l’algorithme de [Orlando *et al.*, 2017]. En (a), les images proviennent du jeu de données DRIVE (les disques optiques sont décalés par rapport au centre), tandis qu’en (b), les images proviennent du jeu de données CHASEDB (les disques optiques sont centrés).

calculons alors pour chaque seuil le score de précision interpolée défini comme suit :

$$P_{\text{interp}}(R) = \max_{R' \leq R} P(R'). \quad (5.6)$$

5.2 Caractérisation de vaisseaux sanguins en ophtalmologie

Cette section présente les expérimentations menées dans un contexte applicatif lié à l’imagerie ophtalmologique. Nous considérons donc ici des images issues d’examens du fond d’œil, permettant de visualiser la rétine et son réseau vasculaire.

5.2.1 Jeux de données

Dans le contexte de ces expérimentations, nous avons utilisé trois jeux de données d’images ophtalmologiques de fonds d’œil :

- *CHASEDB1* [Fraz *et al.*, 2012], qui est composé de 20 images où le disque optique est positionné au centre de l’image ;
- *DRIVE* [Staal *et al.*, 2004], qui est composé de 20 images où le disque optique est décalé du centre de l’image (vers la gauche ou la droite) ;

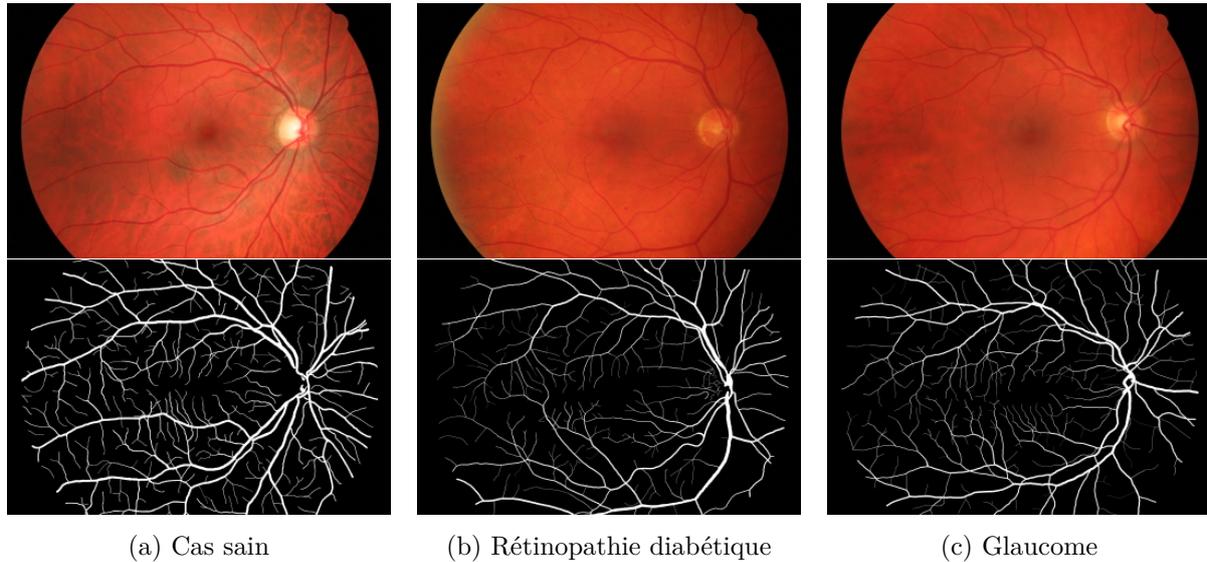


Figure 5.2 – Exemples d’images de fonds d’œil du jeu de données HRF [Odstreilik *et al.*, 2013], avec en dessous leurs segmentations de référence des vaisseaux sanguins. De gauche à droite, les images correspondent à différentes situations pathologiques : cas sain, rétinopathie diabétique et glaucome.

- *High-Resolution Fundus* (HRF) [Odstreilik *et al.*, 2013], qui est composé de 45 images haute résolution où les fonds d’œil représentés correspondent à différentes situations pathologiques. On y retrouve ainsi 15 images provenant de patients sains, 15 correspondant à une rétinopathie diabétique, et 15 présentant des symptômes de glaucome.

Quelques exemples illustratifs d’images provenant de ces trois jeux de données sont présentés en Figure 5.1 (pour DRIVE et CHASEDB1) et en Figure 5.2 (pour HRF).

Généralement, ces jeux de données sont utilisés dans des contextes d’analyse d’images pour évaluer la qualité d’algorithmes de segmentation dédiés aux structures tubulaires telles que les vaisseaux sanguins [Merveille *et al.*, 2017]. Ils sont d’ailleurs également proposés avec des segmentations de référence des vaisseaux sanguins, effectuées préalablement par des ophtalmologues, qui permettent ainsi d’évaluer la qualité de ces algorithmes de segmentation.

Dans le contexte de ces validations expérimentales, plutôt que de se focaliser sur la segmentation automatique de telles structures, nous proposons d’utiliser les descripteurs d’enlacement et d’entrelacement proposés pour caractériser la disposition spatiale des vaisseaux sanguins dans l’image, par rapport au reste du fond d’œil. Ainsi, pour obtenir la localisation des vaisseaux sanguins dans les images, nous utilisons directement les segmentations de référence fournies par les ophtalmologues afin de disposer de localisations optimales. Par ailleurs, pour étudier la robustesse de nos descripteurs par rapport à une segmentation préalable, nous utilisons également les résultats d’un algorithme de segmentation proposé récemment par [Orlando *et al.*, 2017]. Il s’agit d’une approche de segmentation supervisée qui repose sur l’utilisation de champs aléatoires conditionnels. Par la suite, d’autres algorithmes de segmentation pourraient être envisagés.

5.2.2 Protocole de validation

Afin de valider l'intérêt des descripteurs d'enlacement et d'entrelacement présentés dans le [Chapitre 4](#), dans ce contexte d'imagerie ophtalmologique, nous proposons deux applications expérimentales distinctes. La première consiste à étudier la robustesse des descripteurs d'enlacement par rapport au bruit et aux potentielles erreurs de segmentation des images. Cette étude sera réalisée sur les jeux de données DRIVE et CHASEDB1. La seconde propose, à partir des descripteurs, de classifier automatiquement des images de fonds d'œil en fonction de la situation pathologique qu'ils représentent. L'étude sera alors effectuée à partir du jeu de données HRF.

Robustesse à la segmentation et au bruit

L'objectif de cette première application est d'illustrer la robustesse des descripteurs d'enlacement proposés lorsque des erreurs de segmentation des vaisseaux sanguins interviennent. Pour cela, nous proposons de discriminer les images provenant des jeux de données DRIVE et CHASEDB1. Il s'agit donc de catégoriser les fonds d'œil représentés en deux classes : celles présentant un disque optique centré (CHASEDB1) et celles dont le disque optique est décalé (DRIVE), comme nous avons pu le voir sur la [Figure 5.1](#).

Pour cette application, la localisation des vaisseaux sanguins est obtenue à partir de l'algorithme de segmentation proposé par [[Orlando et al., 2017](#)]. À partir de ces résultats de segmentation, nous proposons alors d'ajouter graduellement différents types de bruits aléatoires sur les images segmentées, simulant ainsi la génération d'erreurs de segmentation de plus en plus fortes.

Les différents types de bruits aléatoires appliqués sont les suivants :

- Le bruit *Sel et poivre*, qui consiste à ajouter aléatoirement à l'image des pixels noirs ou blancs en fonction d'une loi uniforme, avec une certaine probabilité permettant de contrôler la quantité de bruit ;
- Le bruit *Gaussien*, qui ajoute des pixels bruités à l'image en fonction de la distribution d'une loi normale. Le paramètre de variance de cette loi permet alors de contrôler la quantité de bruit ajouté ;
- Le bruit *Speckle*, qui est un bruit multiplicatif sur l'image défini également en fonction d'une loi normale. L'aspect multiplicatif a pour effet de dégrader localement la forme considérée (en l'occurrence, le réseau vasculaire segmenté) sans ajouter de bruit en dehors de celle-ci.

En [Figure 5.3](#), l'effet de ces différents types de bruits aléatoires sur une image de vaisseaux sanguins peut être visualisé. Notons que pour le cas du bruit *Gaussien* et du bruit *Speckle*, nous procédons à une binarisation de l'image segmentée suite à l'application du bruit aléatoire.

Finalement, après la segmentation automatique et l'application du bruit aléatoire, nous extrayons de l'image un objet A bruité représentant les vaisseaux sanguins, ainsi qu'un objet B qui correspond au masque circulaire du fond de l'œil (auquel on a soustrait l'objet A). Nous calculons alors le descripteur d'entrelacement \mathcal{I}_{AB} pour caractériser la complexité du réseau vasculaire. Comme comparaison de référence, l'histogramme de forces \mathcal{F}_{AB} est calculé, ainsi que le descripteur GFD de l'objet A .

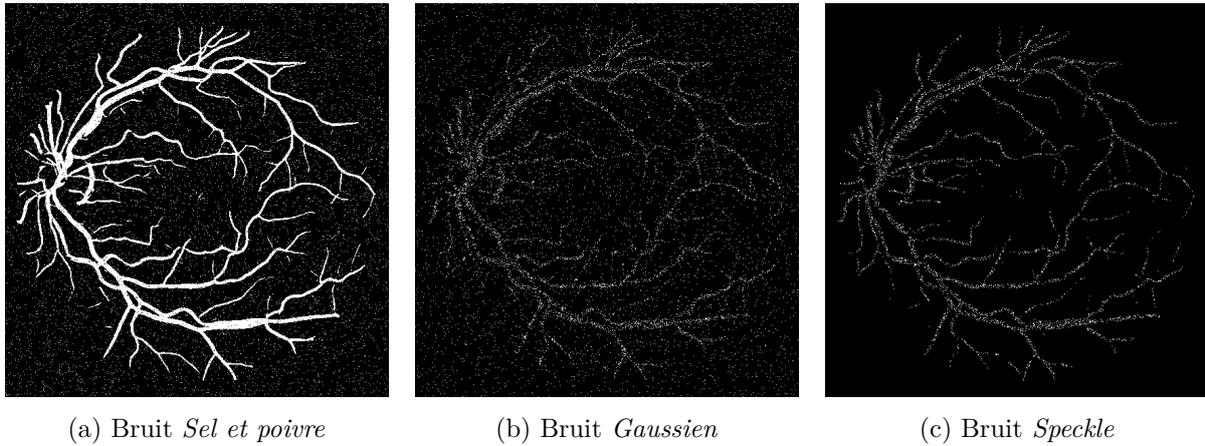


Figure 5.3 – Exemples de réseaux vasculaires segmentés auxquels différents types de bruits aléatoires ont été appliqués.

Classification de pathologies rétiniennes

Cette deuxième application expérimentale consiste à classifier les images du jeu de données HRF selon les pathologies qu’elles représentent. En effet, les vaisseaux sanguins pathologiques peuvent être affectés par des artefacts tels que des néovascularisations, des anévrismes ou des hémorragies. Par conséquent, la structure et la configuration spatiale de leurs réseaux peuvent être affectées [Abràmoff *et al.*, 2010]. Il convient toutefois de noter que ces différents types d’artefacts au sein du réseau vasculaire ne constituent pas les seuls symptômes caractéristiques des pathologies qui sont étudiées ici, et que d’autres biomarqueurs peuvent être utilisés pour mieux les caractériser. Cependant, nous insistons sur le fait que l’objectif de ces expérimentations est avant tout d’illustrer que les descripteurs proposés peuvent être utilisés pour caractériser la complexité de ces structures spatiales, au-delà du contexte applicatif sous-jacent.

De la même manière que pour l’étude de la robustesse au bruit, l’objet A représente ici le réseau vasculaire et l’objet B représente le masque correspondant au reste du fond d’œil, auquel on a soustrait A . Nous considérons alors le descripteur d’entrelacement \mathcal{I}_{AB} , l’histogramme de forces \mathcal{F}_{AB} et le descripteur GFD de l’objet A .

Tableau 5.1 – Comparaison des taux de reconnaissance obtenus pour l’étude de la robustesse au bruit sur les jeux de données DRIVE et CHASEDB1.

	GFD	F-Histogrammes	Entrelacement
Bruit <i>Speckle</i>	81,00 ± 0,43	96,50 ± 0,03	97,75 ± 0,03
Bruit <i>Gaussien</i>	93,75 ± 0,47	99,00 ± 0,05	99,25 ± 0,03
Bruit <i>Sel et poivre</i>	94,25 ± 0,10	96,25 ± 0,09	98,75 ± 0,02

5.2. Caractérisation de vaisseaux sanguins en ophtalmologie

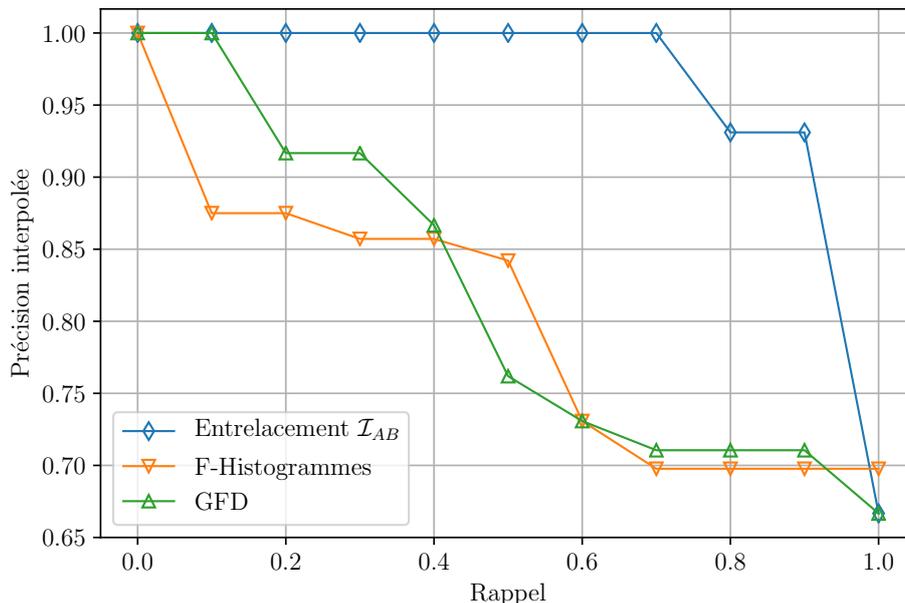


Figure 5.4 – Courbes de précision-rappel obtenues pour la classification des images de fonds d’œil du jeu de données HRF, pour les descripteurs d’entrelacement \mathcal{I}_{AB} , ainsi que pour les descripteurs GFD et les histogrammes de forces.

5.2.3 Résultats expérimentaux et discussion

Le [Tableau 5.1](#) présente les résultats obtenus pour l’étude sur la robustesse à la segmentation et au bruit pour les jeux de données DRIVE et CHASEDB1. Dans ce tableau, nous rapportons les taux de reconnaissance moyens obtenus pour la classification binaire des images (en fonction du centre optique décalé ou non), et ceci pour les différents types de bruits ajoutés aux segmentations initiales des images. Le bruit *Sel et poivre* a été appliqué avec des taux croissants, pour 10 niveaux de probabilités dans l’intervalle $[0, 0.10]$. De manière similaire, le bruit *Gaussien* et le bruit *Speckle* ont été appliqués pour 10 valeurs de variances croissantes dans l’intervalle $[0, 2]$. Ces résultats de classification permettent d’illustrer la robustesse des descripteurs d’entrelacement par rapport aux erreurs de segmentation. Ils sont globalement plus efficaces que les descripteurs GFD pour discriminer les cas où le disque optique est décalé ou centré, et ils sont légèrement meilleurs que les histogrammes de forces.

La [Figure 5.4](#) montre les courbes de précision-rappel obtenues pour la deuxième application expérimentale proposée, où nous effectuons une classification des pathologies représentées par les images de fonds d’œil. Nous avons procédé ici à une classification binaire, dans le but de discerner les réseaux vasculaires provenant de patients sains de ceux présentant des pathologies (rétinopathie diabétique ou glaucome). À partir de ces résultats, nous pouvons observer la capacité des descripteurs d’entrelacement à discriminer les images de fonds d’œil émanant de patients malades, de celles de patients sains. Ils permettent ainsi de caractériser efficacement les artefacts symptomatiques du réseau vasculaire, par opposition aux cas sains.

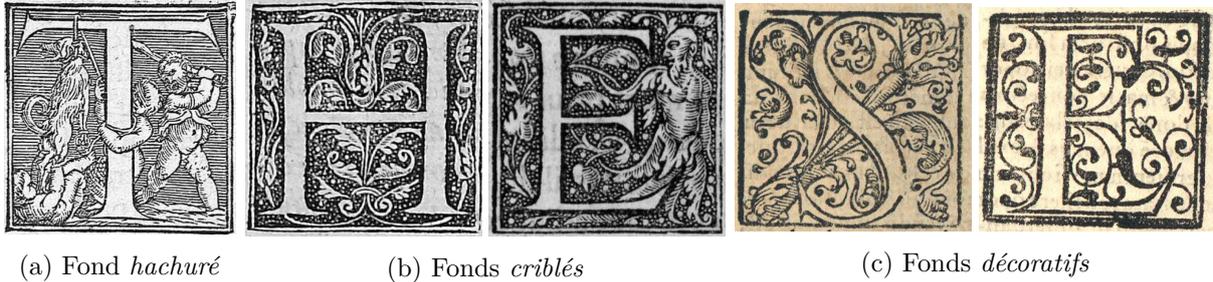


Figure 5.5 – Exemples d’images provenant du jeu de données de lettres décoratives. La première image a un fond de style *hachuré*, les deux suivantes ont un fond *criblé*, et les deux dernières ont un fond *décoratif*.

5.3 Reconnaissance de lettres décoratives

Dans cette section, nous proposons un deuxième type de validations expérimentales pour illustrer l’intérêt des descripteurs d’enlacement et d’entrelacement. Nous nous positionnons ici dans le contexte applicatif de l’analyse d’images de documents anciens.

5.3.1 Jeu de données

Le jeu de données utilisé pour cette application est un ensemble d’images de lettres décoratives qui ont été extraites de documents anciens. Cette base d’images a été mise en place dans le contexte du projet de recherche NaviDoMass¹ [Jouili *et al.*, 2010]. L’analyse et l’indexation automatique de telles images de lettres représente un grand intérêt pour l’archivage de documents historiques et la conservation numérique du patrimoine, par exemple pour des établissements tels que les bibliothèques et les musées. En effet, la caractérisation des différents types de lettres peut servir à identifier par exemple les auteurs ou les périodes d’impression. Pour autant, ce processus d’analyse automatisé constitue une tâche difficile, car ces images présentent une grande variété de motifs texturés, ainsi que des décorations ornementales complexes. De plus, les documents dont les images sont extraites peuvent avoir subi divers types de dégradations physiques au cours de leurs cycles de vie [Coustaty *et al.*, 2011].

Dans ces travaux, nous avons extrait un sous-ensemble de 120 images de ce jeu de données. Ces images ont été manuellement annotées et catégorisées par des historiens, notamment en fonction de leurs différents styles de fonds, qui peuvent être utiles pour caractériser par exemple une époque particulière, ou un certain type d’imprimerie. Quelques exemples d’images de lettres issues de ce jeu de données, illustrant ces différents styles de fonds, sont présentées en Figure 5.5. Indépendamment de la lettre représentée dans l’image, nous distinguons trois types de styles de fonds : les fonds *hachurés*, les fonds *criblés*, et les fonds *décoratifs*.

1. <http://navidomass.univ-lr.fr/>

5.3. Reconnaissance de lettrines décoratives

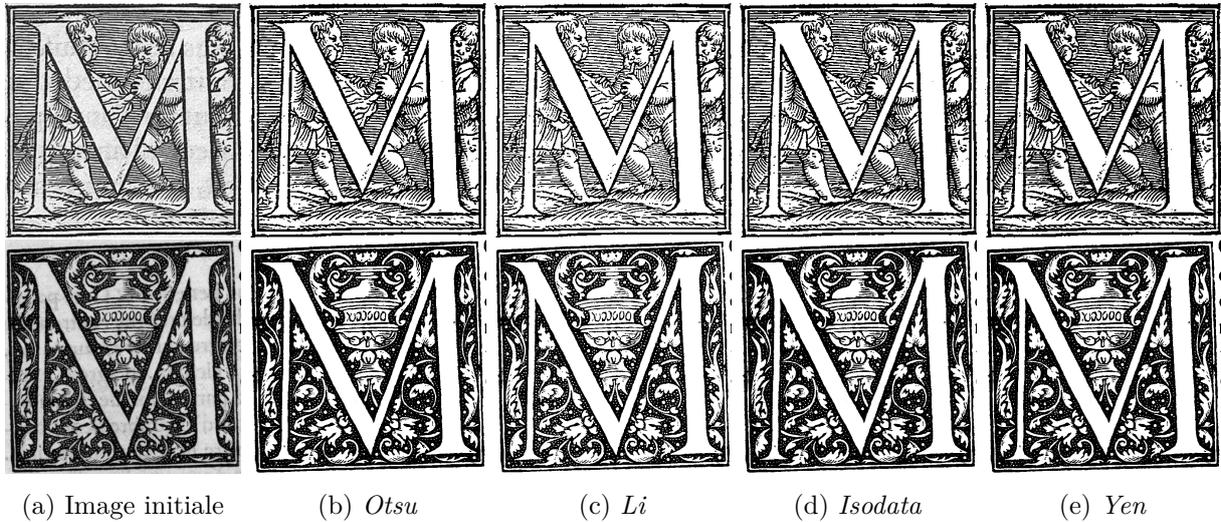


Figure 5.6 – Exemples de binarisations obtenues pour différentes images de lettrines décoratives.

5.3.2 Protocole de validation

Pour cette application expérimentale, notre objectif consiste à classifier les différentes images de lettrines en fonction de leurs trois styles de fonds, indépendamment des lettres qui sont représentées dans les images. L’hypothèse sous-jacente pour cette application est que ces différents types de fonds peuvent être caractérisés par des motifs d’enlacement et d’entrelacement caractéristiques à chaque style. Nous nous rapprochons donc ici d’une analyse spatiale de la texture du fond des lettrines, envisagée ici de manière globale sur les images.

Dans un premier temps, les images du jeu de données sont segmentées en deux couches. Cette étape de segmentation vise à séparer les pixels correspondant à de l’encre de ceux correspondant à du papier. Pour cela, nous avons utilisé plusieurs algorithmes de binarisation classiques en traitement d’images [Sezgin et Sankur, 2004], dans le but d’évaluer la stabilité de l’approche par rapport à différents résultats de segmentation. Les méthodes de binarisation utilisées pour ces expérimentations sont les suivantes :

- *Otsu* (minimisation de la variance intra-classe) ;
- *Li* (minimisation de l’entropie croisée) ;
- *Isodata* (seuil fixé à la moyenne des moyennes des deux classes) ;
- *Yen* (maximum de corrélation avec une fonction de coût).

Le résultat de ces différentes techniques de binarisation peut être visualisé en Figure 5.6 pour deux images de lettrines.

Pour chaque image, une fois l’étape de segmentation effectuée, nous calculons alors le descripteur d’entrelacement \mathcal{L}_{AB} entre la couche d’encre (objet A) et la couche papier (objet B). Ceci nous permet d’obtenir un ensemble de vecteurs de caractéristiques, qui sont alors utilisés pour effectuer la classification des images de lettrines en fonction de leurs styles de fonds, selon le protocole expérimental détaillé dans la Section 5.1.

Tableau 5.2 – Taux de reconnaissance obtenus pour la classification des images de lettrines en fonction de leurs différents styles de fonds, pour différentes méthodes de segmentation, ainsi que pour les différentes méthodes comparatives considérées.

	GFD	F-Histogrammes	Entrelacement
<i>Otsu</i>	80,00	93,33	95,83
<i>Isodata</i>	79,17	95,00	95,83
<i>Li</i>	81,67	96,67	97,50
<i>Yen</i>	78,33	94,17	93,33
Moyenne	79,79 ± 1,23	94,79 ± 1,23	95,62 ± 1,49

5.3.3 Résultats expérimentaux et discussion

Le [Tableau 5.2](#) présente les taux de reconnaissance obtenus pour la classification des images de lettrines en fonction de leurs différents styles de fonds. Nous pouvons constater, à partir de ces résultats, que les descripteurs d’entrelacement sont en mesure de caractériser efficacement les différents styles de fonds représentés dans ce jeu de données. Les descripteurs d’entrelacement obtiennent de meilleurs résultats que les descripteurs GFD pour toutes les méthodes de segmentation, et obtiennent des résultats similaires aux histogrammes de forces, tout en capturant une information spatiale d’une autre nature. En effet, du fait de la complexité des couches d’images à analyser, les histogrammes de forces caractérisent principalement la répartition globale des objets, tandis que l’approche proposée prend en compte les différents motifs d’entrelacement présents dans les images, comme par exemple les hachures qui se caractérisent par des alternances directionnelles d’encre et de papier.

Ces résultats nous ont permis d’illustrer l’intérêt des descripteurs d’entrelacement pour ce type d’images de documents, en les appliquant directement sur les images à une échelle globale. Nous reviendrons de manière plus approfondie sur ce contexte applicatif à la fin du manuscrit, dans le [Chapitre 10](#), où nous proposons une stratégie d’apprentissage de descripteurs d’enlacement calculés à une échelle locale sur les images de lettrines.

5.4 Recherche et classification d’objets géographiques urbains

Dans cette section, nous présentons un troisième type de validations expérimentales, cette fois-ci dans le contexte applicatif de l’analyse d’images de télédétection.

5.4.1 Jeu de données

Le jeu de données utilisé pour cette application est une image satellitaire à Très Haute Résolution Spatiale (*THRS*) provenant du satellite Pléiades². Cette image comporte un nombre

2. Projet ANR COCLICO (ANR Modèles Numériques, ANR-12-MN-001-COCLICO, 2012-2016). <http://icube-coclico.unistra.fr/index.php/Coclico>

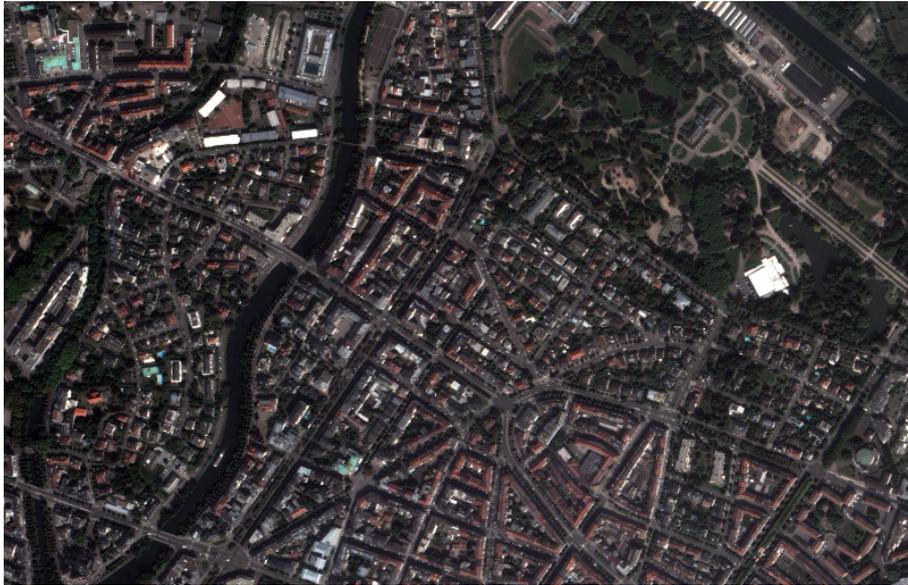


Figure 5.7 – Exemple d'une zone urbaine issues de l'image satellitaire THRS utilisée pour valider les descripteurs d'enlacement et d'entrelacement.

de pixels de l'ordre de 10^8 , et représente une zone urbaine issue de la région de Strasbourg, en Alsace (France). Un exemple de zone urbaine provenant de cette image satellitaire est présenté en [Figure 5.7](#), où l'on peut y observer une forte hétérogénéité, à la fois radiométrique et géométrique, des différents objets urbains représentés [[Kurtz et al., 2012](#)].

Dans le cadre de ces validations expérimentales, nous disposons également de deux cartes de vérité terrain indiquant des informations d'emprise au sol de cette image satellitaire. Un extrait de ces cartes est illustré en [Figure 5.8](#). En (a), la première carte indique la localisation des différents objets géographiques, à une échelle élémentaire. Nous retrouvons ainsi quatre classes sémantiques d'objets : *bâtiments*, *routes*, *végétation* et *eau* (alors notées *B*, *R*, *V* et *E* respectivement). En (b), la seconde carte indique la localisation des blocs urbains dans l'image. Nous nous intéresserons ici en particulier aux blocs dits *continus* (avec une forte densité d'habitation) et aux blocs dits *discontinus* (avec une densité d'habitation plus faible). Ceux-ci sont représentés en rouge et en bleu respectivement dans cette carte de référence.

5.4.2 Protocole de validation

À partir de ce jeu de données, nous présentons ici deux applications expérimentales. La première application consiste à classifier les différents blocs urbains continus et discontinus de l'image satellitaire, en fonction de la disposition spatiale des objets élémentaires qui les composent, à partir des descripteurs d'enlacement et d'entrelacement. Dans la seconde application, nous proposons de mettre en place un système permettant d'effectuer des requêtes de zones géographiques en fonction de l'enlacement ou de l'entrelacement entre différents types d'objets urbains élémentaires.

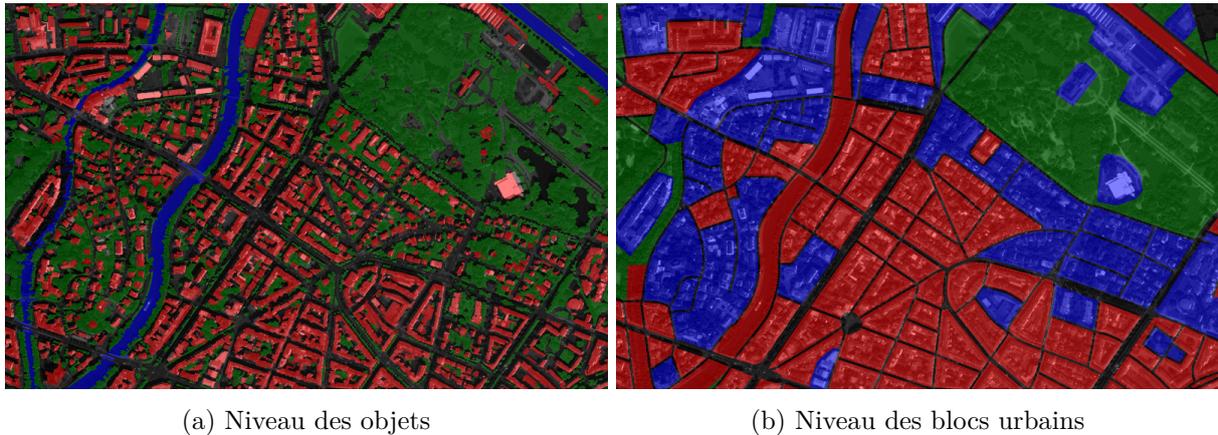


Figure 5.8 – Illustration des niveaux de classification, affichés par transparence sur une coupe de l’image satellitaire utilisée.

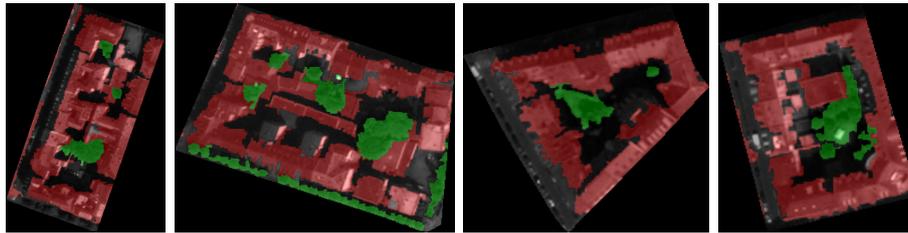
Classification de blocs urbains continus et discontinus

Pour cette application, notre objectif est de discerner des blocs urbains continus et discontinus en fonction de la structure spatiale de la végétation et des bâtiments représentés au sein de ces blocs. L’intuition de départ est que, indépendamment de leurs formes ou de leurs tailles, ces différents types de blocs présentent des dispositions spatiales caractéristiques de végétation et de bâtiments. Par exemple, dans le cas de blocs discontinus, l’on peut s’attendre à ce que les bâtiments soient fortement enlacés par la végétation.

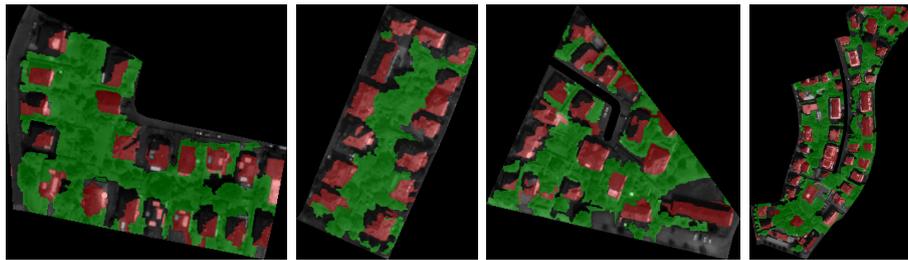
Pour mettre en place ce processus de classification, nous avons effectué un certain nombre de pré-traitements. Tout d’abord, l’image originale a été segmentée à partir d’un algorithme de segmentation multirésolution ascendant dédié aux images satellitaires THRS [Baatz et Schape, 2000]. Les paramètres de cet algorithme ont été fixés de telle sorte à obtenir une segmentation fine de l’image, où les segments obtenus peuvent correspondre aux objets urbains élémentaires. Ces segments ont ensuite été étiquetés à partir d’une approche de classification supervisée, en fonction des quatre classes d’objets présentes dans la carte de référence (bâtiments, végétation, routes et eau). Nous avons alors découpé cette image segmentée en fonction des différents blocs continus et discontinus apparaissant dans la carte de référence des blocs urbains. Ainsi, pour chaque composante connexe correspondant à un bloc continu ou discontinu, nous générons l’image correspondant aux objets bâtiments (objet B) et végétation (objet V) présents dans cette composante. Nous obtenons alors 617 images contenant différentes répartitions de bâtiments et de végétation. Chaque image correspond donc à un bloc continu ou discontinu, et fait la taille du rectangle englobant de ce bloc. Quelques exemples de blocs obtenus par cette procédure sont illustrés en Figure 5.9.

Ensuite, pour chacune de ces images de blocs urbains, nous avons calculé les descripteurs d’enlacement (\mathcal{E}_{BV} et \mathcal{E}_{VB}) et d’entrelacement (\mathcal{I}_{BV}) entre les *bâtiments* (objet B) et la *végétation* (objet V). Comme pour les autres applications expérimentales de ce chapitre, les compa-

5.4. Recherche et classification d'objets géographiques urbains



(a) Blocs urbains *contigus* (denses)



(b) Blocs urbains *discontinus* (peu denses)

Figure 5.9 – Exemples de blocs urbains issus de l'image satellitaire THRS. La première ligne présente des blocs urbains *contigus*, tandis que la deuxième ligne présente des blocs urbains *discontinus*. Les objets représentant les bâtiments (objet B) et la végétation (objet V) à l'intérieur de ces blocs sont affichés en transparence, en rouge et en vert respectivement.

raisons de référence sont effectuées avec les histogrammes de forces correspondants et avec les descripteurs GFD. Pour cette application, deux descripteurs GFD sont calculés (un pour l'objet B et un pour l'objet V) puis concaténés au sein d'un même vecteur. Enfin, le même protocole de classification est utilisé : classification par k -NN avec validation croisée de type *leave-one-out*.

Requêtes sémantiques de configurations spatiales

La deuxième application expérimentale proposée est un système permettant d'effectuer des requêtes sémantiques de configurations d'enlacement ou d'entrelacement entre différents couples d'objets urbains. Une requête sémantique typique pourrait être formulée en langage naturel, par exemple, retrouver des zones urbaines qui présentent un « *entrelacement fort* entre les *bâtiments* et la *route* », ou bien les zones où « *l'eau est fortement enlacée* par la *route* ».

Afin de conserver une plus grande diversité de configurations spatiales entre objets urbains dans l'image satellitaire, nous n'avons pas utilisé les différents blocs urbains de l'application précédentes, car ils sont composés quasi-exclusivement de bâtiments et de végétation. Nous avons plutôt choisi d'utiliser l'intégralité de l'image satellitaire initiale, que nous avons alors découpée en fenêtres carrées de taille 500×500 pixels, sans chevauchement. Nous obtenons un total de 396 fenêtres. Nous utilisons alors la carte de référence des objets urbains élémentaires pour obtenir la localisation des différents types d'objets présents au sein de chaque fenêtre. Ainsi, chaque fenêtre peut être composée des quatre types d'objets élémentaires (E, V, B et R)

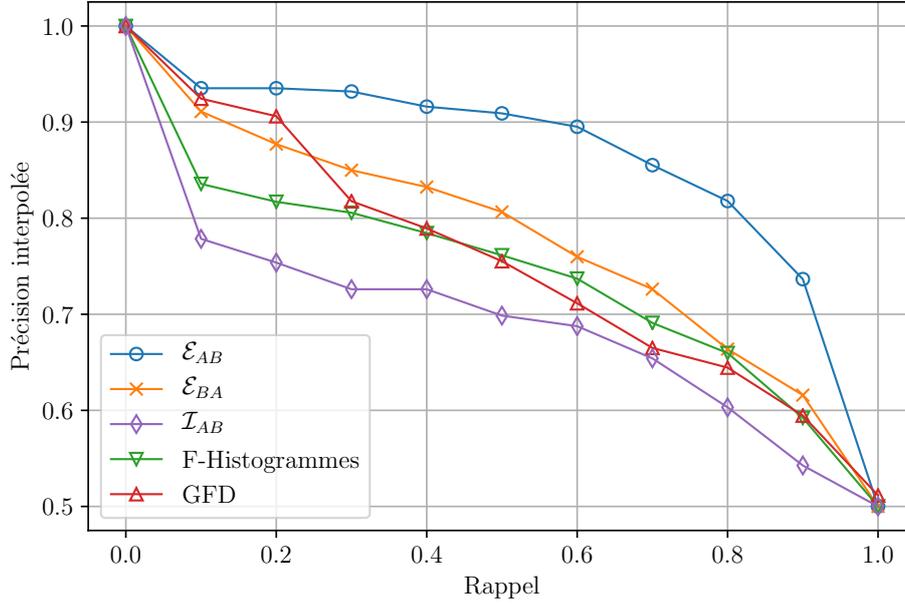


Figure 5.10 – Courbes de précision-rappel obtenues pour la classification de blocs urbains *continus* et *discontinus*, pour les descripteurs d’enlacement et d’entrelacement, ainsi que pour les descripteurs GFD et les histogrammes de forces.

disposés selon différentes configurations spatiales.

Pour chacune de ces fenêtres, nous avons calculé les descripteurs d’enlacement \mathcal{E}_{AB} et \mathcal{E}_{BA} ainsi que le descripteur d’entrelacement \mathcal{I}_{AB} entre chacun des couples d’objets urbains possibles (dénotés respectivement par EV , EB , ER , VB , VR et BR). Nous obtenons ainsi un total de 7128 descripteurs.

Ensuite, nous avons employé une stratégie d’agrégation sur chaque descripteur pour l’ensemble des directions, ceci dans le but d’obtenir pour chaque couple d’objets une quantité globale d’enlacement (ou d’entrelacement) représentée par une valeur scalaire. Cette quantité est ainsi calculée de la manière suivante :

$$\tilde{\mathcal{H}}_{AB} = \int_0^\pi \mathcal{H}_{AB}(\theta) d\theta \quad (5.7)$$

où \mathcal{H}_{AB} désigne soit le descripteur d’enlacement \mathcal{E}_{AB} (ou \mathcal{E}_{BA}), soit le descripteur d’entrelacement \mathcal{I}_{AB} entre un couple d’objets (correspondant alors aux objets A et B) au sein d’une fenêtre de l’image satellitaire, en fonction de la requête effectuée. Ces valeurs scalaires, représentant des quantités globales d’enlacement ou d’entrelacement, sont ainsi calculées pour toutes les fenêtres, et pour tous les couples d’objets. Elles sont alors triées afin de pouvoir procéder à la formulation de requêtes de configurations spatiales pour différents couples d’objets.

5.4. Recherche et classification d'objets géographiques urbains

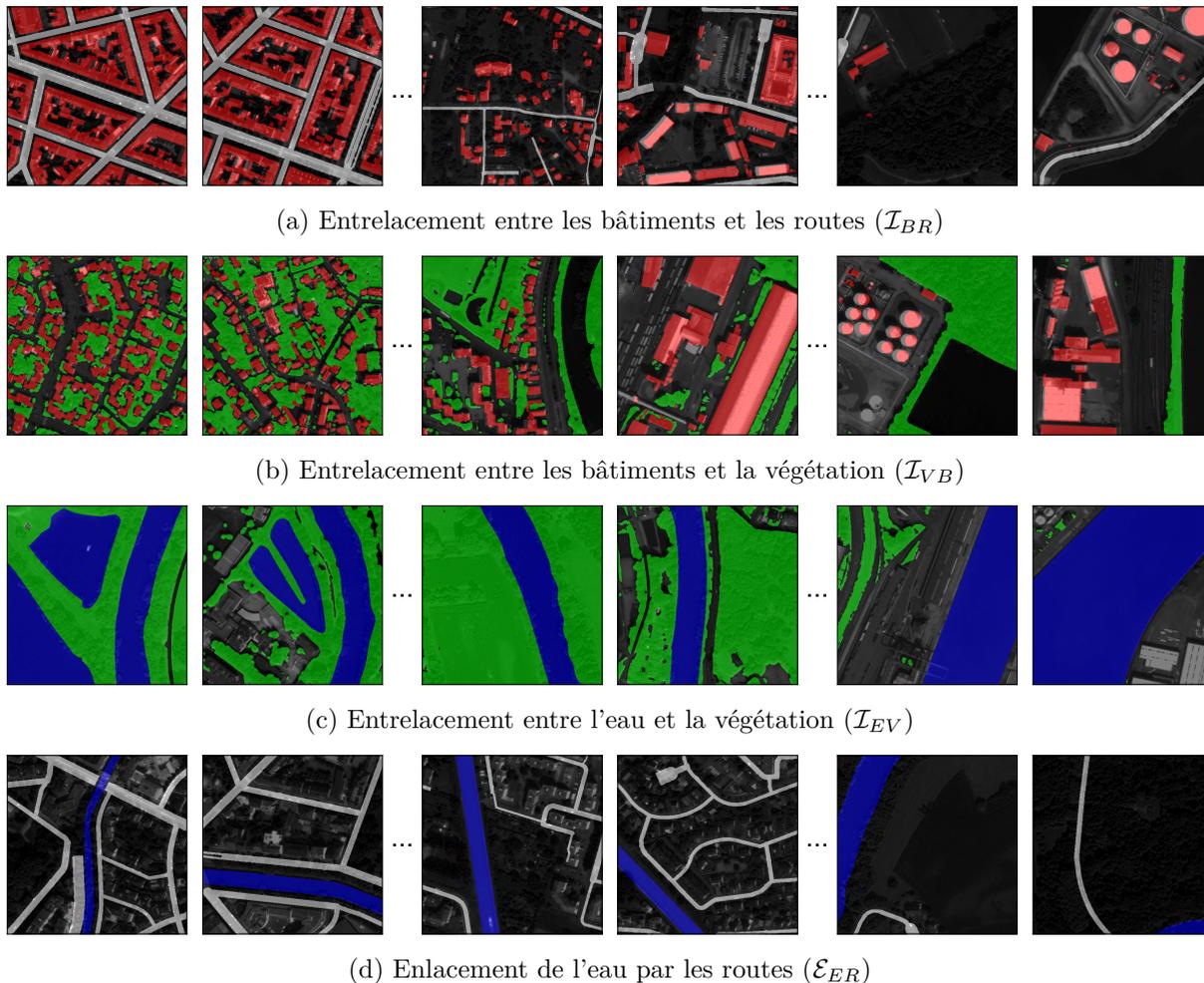


Figure 5.11 – Requêtes sémantiques d'enlacement et d'entrelacement entre différents couples d'objets urbains (chaque ligne correspond à une requête différente). Pour chaque requête, différents niveaux d'enlacement ou d'entrelacement sont affichés.

5.4.3 Résultats expérimentaux et discussion

La Figure 5.10 présente les courbes de précision-rappel obtenues pour la première application expérimentale, c'est-à-dire la classification des différentes images de blocs urbains continus ou discontinus. Nous pouvons y visualiser les résultats pour les descripteurs d'enlacement (\mathcal{E}_{BV} et \mathcal{E}_{VB}) et d'entrelacement (\mathcal{I}_{BV}), ainsi que pour les descripteurs GFD et les F-Histogrammes. À partir de ces courbes, nous pouvons observer que les meilleurs résultats sont obtenus pour les descripteurs d'enlacement \mathcal{E}_{BV} , qui caractérisent dans quelle mesure les *bâtiments* sont enlacés par la *végétation*. Ceci semble confirmer l'hypothèse selon laquelle les bâtiments sont plus fortement enlacés par la végétation dans le cas de blocs urbains discontinus, tandis que cet enlacement serait plus faible pour les blocs continus (voir la Figure 5.9). Au contraire, les résultats suggèrent que les descripteurs d'entrelacement \mathcal{I}_{BV} sont ici moins efficaces pour cette application

spécifique. Ceci peut être dû au fait que de nombreux blocs urbains ne présentent finalement pas un enlacement mutuel de leurs objets, mais plutôt un enlacement relatif. Ces résultats viennent donc étayer l'idée selon laquelle les relations spatiales d'enlacement et d'entrelacement peuvent être utilisées en tant que concepts différents, indépendamment les uns des autres, notamment en fonction des besoins de l'application considérée.

La Figure 5.11 présente les résultats de plusieurs requêtes sémantiques de configurations d'enlacement et d'entrelacement effectuées pour différents couples d'objets urbains. Chacune des lignes correspond à une requête différente (enlacement ou entrelacement) pour différents couples d'objets. Pour chaque requête, nous affichons ici les deux premiers résultats (c'est-à-dire les configurations pour lesquelles la quantité d'enlacement ou d'entrelacement est la plus élevée), deux résultats intermédiaires, et les deux derniers résultats (quantité d'enlacement ou d'entrelacement la plus faible). Par ailleurs, les images ne contenant pas les objets demandés par la requête sont rejetées. Pour la première requête, nous pouvons remarquer que les premiers résultats correspondent à des zones urbaines très denses avec de nombreux blocs d'immeubles. La seconde requête retrouve alors des zones plus résidentielles, avec des toits de maisons au sein de larges espaces de végétation. La troisième requête permet notamment de retrouver des zones de ramifications des cours d'eau en présence de végétation, qui se traduisent par un fort entrelacement entre ces deux types d'objets. Enfin, la quatrième requête permet de retrouver des zones où les cours d'eau passent à travers le réseau routier.

Dans ce contexte d'analyse d'images de télédétection, les résultats de cette étude suggèrent donc que les descripteurs d'enlacement et d'entrelacement sont efficaces pour caractériser différents types de configurations spatiales complexes.

5.5 Bilan scientifique

Dans ce chapitre, nous avons présenté plusieurs types de validations expérimentales permettant d'illustrer l'intérêt et la généralité des descripteurs d'enlacement et d'entrelacement. Nous avons proposé une première application en imagerie ophtalmologique, pour caractériser l'entrelacement des vaisseaux sanguins représentés dans des images de fonds d'œil. Cette application nous a permis d'attester de la robustesse des descripteurs par rapport aux erreurs de segmentation et au bruit, et d'illustrer leur intérêt pour la classification de pathologies rétinienne. Dans la seconde application, nous avons employé les descripteurs d'entrelacement pour reconnaître les différents styles de fonds présents au sein d'images de lettrines décoratives. Nous reviendrons notamment de manière plus approfondie sur ce contexte applicatif dans le Chapitre 10. Enfin, dans une troisième application, les descripteurs d'enlacement et d'entrelacement ont été appliqués dans le domaine de la télédétection, pour la classification de blocs urbains ainsi que pour la formulation de requêtes sémantiques de configurations spatiales entre objets géographiques.

Ces résultats expérimentaux ont été valorisés au sein de l'article de revue internationale présentant le modèle de l'enlacement et de l'entrelacement [Clément *et al.*, 2017a]. Une partie préliminaire des résultats sur la caractérisation du réseau vasculaire d'images de fonds d'œil a également été présentée dans l'article de conférence nationale [Clément *et al.*, 2016c].

Chapitre 6

Évaluations des relations d'enlacement et d'entrelacement

Sommaire

6.1	Évaluations à partir de descripteurs	76
6.1.1	Approche par mise en correspondance d'ensembles flous	76
6.1.2	Approche par agrégation directionnelle	77
6.1.3	Résultats illustratifs	78
6.2	Paysages flous directionnels d'enlacement	80
6.2.1	Définitions et évaluations floues	80
6.2.2	Résultats illustratifs	84
6.2.3	Vers les paysages flous d'entrelacement	85
6.3	Applications à la relation spatiale « entouré par »	86
6.3.1	Nouvelles modélisations à partir de l'enlacement	86
6.3.2	Exemples caractéristiques	88
6.3.3	Résultats expérimentaux et discussion	89
6.4	Bilan scientifique	93

Dans ce chapitre, nous présentons comment exploiter le modèle d'enlacement et d'entrelacement du point de vue de l'évaluation de relations spatiales, par opposition au point de vue des descripteurs de position relative précédemment considéré. Dans un premier temps, nous introduisons une approche permettant de générer des évaluations des relations d'enlacement et d'entrelacement à partir des descripteurs du [Chapitre 4](#) ([Section 6.1](#)). Ensuite, nous proposons une extension de notre modèle, permettant la construction de paysages flous directionnels d'enlacement et d'entrelacement, inspirés par les travaux de [\[Bloch, 1999\]](#) pour les relations spatiales directionnelles classiques ([Section 6.2](#)). Ces paysages flous peuvent alors être utilisés pour effectuer des évaluations quantitatives et visuelles. Enfin, nous proposons une application spécifique de ces approches pour évaluer la relation spatiale complexe « entouré par » ([Section 6.3](#)). Nous dressons alors un bilan des contributions scientifiques du chapitre ([Section 6.4](#)).

6.1 Évaluations à partir de descripteurs

Dans cette section, notre objectif est d'étudier comment les descripteurs d'enlacement et d'entrelacement proposés dans le [Chapitre 4](#) peuvent être utilisés pour générer des évaluations de relations spatiales dans le contexte de configurations spatiales complexes. Le but est d'illustrer que ces descripteurs sont exploitables dans un cadre classique d'évaluation de relations spatiales, tout en soulignant l'intérêt de la prise en compte de l'aspect directionnel pour mesurer l'enlacement et l'entrelacement.

Dans la suite, l'histogramme \mathcal{H}_{AB} désigne, pour deux objets A et B , soit le descripteur d'enlacement \mathcal{E}_{AB} , soit le descripteur d'entrelacement \mathcal{I}_{AB} , en fonction de l'évaluation considérée. Cet histogramme est alors normalisé par sa valeur maximum, de sorte à ce qu'il puisse être considéré comme un sous-ensemble flou de \mathbb{R} dans $[0, 1]$. Une évaluation désigne alors un nombre réel dans l'intervalle $[0, 1]$ dont l'objectif est de correspondre à un degré de validité de la relation étudiée pour des objets donnés. Nous proposons ci-après deux approches distinctes permettant de générer de telles évaluations.

6.1.1 Approche par mise en correspondance d'ensembles flous

Une méthode traditionnellement utilisée pour réaliser des évaluations de relations spatiales est la mise en correspondance d'ensemble flous (*fuzzy pattern matching* en anglais), proposée initialement par [\[Dubois et al., 1988\]](#) dans un cadre plus général de raisonnement à partir d'informations floues. L'idée principale de cette approche consiste à comparer un ensemble flou à un autre ensemble flou de référence (aussi appelé *profil flou* dans la suite) qui est censé modéliser le résultat attendu.

Dans [\[Krishnapuram et al., 1993\]](#), l'ensemble flou directionnel modélisant la relation « être dans la direction θ de » est donné par $\mu_\alpha(\theta) = \mu_0(\theta - \alpha)$, avec pour formulation de référence :

$$\mu_0(\theta) = \begin{cases} \cos^2(\theta) & \text{si } -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}, \\ 0 & \text{sinon.} \end{cases} \quad (6.1)$$

Quelques exemples de profils flous directionnels pour différents angles α sont ainsi illustrés en [Figure 6.1](#). Ces profils flous permettent de modéliser le degré de validité d'une direction de manière graduelle. Dans notre contexte, du fait de la π -périodicité des descripteurs d'enlacement et d'entrelacement, il est plus pertinent d'utiliser un profil sans restriction sur l'intervalle $[-\frac{\pi}{2}, \frac{\pi}{2}]$, c'est-à-dire $\mu_0(\theta) = \cos^2(\theta)$ car l'enlacement à gauche et à droite correspondent à la même quantité. Notons que d'autres types de profils flous auraient pu être envisagés pour cette approche, comme par exemple des profils gaussiens ou trapézoïdaux.

Ces ensembles flous de référence correspondent à des profils directionnels qui modélisent le fait qu'un objet soit *enlacé* par un autre (ou que deux objets soient *entrelacés*) dans une direction particulière. Considérons deux objets A et B et leur histogramme \mathcal{H}_{AB} d'enlacement ou d'entrelacement, ainsi qu'une direction α . Une évaluation floue de la relation spatiale « A est enlacé par B dans la direction α » (ou de « A et B sont entrelacés dans la direction α »)

6.1. Évaluations à partir de descripteurs

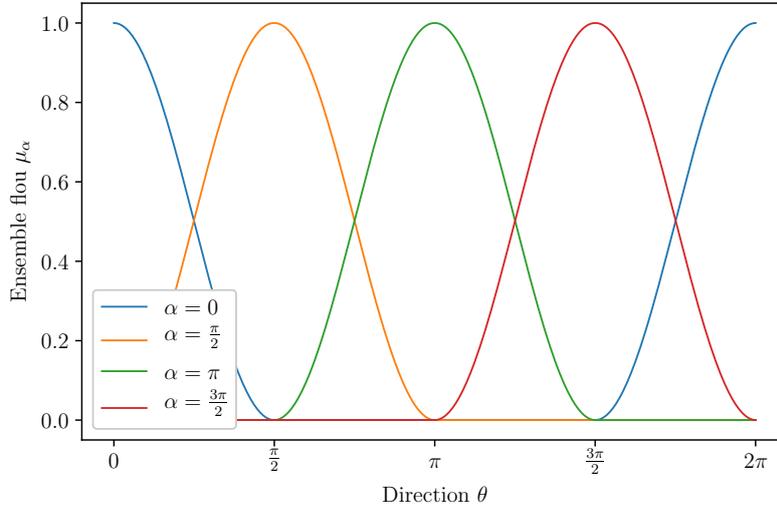


Figure 6.1 – Exemples de profils flous directionnels sinusoidaux pour différents angles α .

peut être obtenue par les mesures floues de possibilité Π_α et de nécessité N_α , données par les formules suivantes :

$$\Pi_\alpha(\mathcal{H}_{AB}) = \sup_{\theta \in [0, \pi]} \min(\mathcal{H}_{AB}(\theta), \mu_\alpha(\theta)), \quad (6.2)$$

$$N_\alpha(\mathcal{H}_{AB}) = \inf_{\theta \in [0, \pi]} \max(1 - \mathcal{H}_{AB}(\theta), \mu_\alpha(\theta)). \quad (6.3)$$

Notons ici également que les opérateurs min et max pourraient être remplacés par d'autres t -normes et t -conormes floues, en fonction de la sémantique souhaitée pour modéliser la relation spatiale. La nécessité N_α représente une évaluation pessimiste, tandis que la possibilité Π_α correspond à une évaluation optimiste. L'intervalle $[N_\alpha, \Pi_\alpha]$ permet alors de visualiser et d'évaluer le degré de confiance que l'on peut avoir dans le fait que la relation spatiale soit valide pour la direction et les deux objets considérés.

6.1.2 Approche par agrégation directionnelle

Nous proposons ici une deuxième approche pour générer des évaluations des relations d'enlacement et d'entrelacement à partir des descripteurs directionnels. Cette approche repose sur une stratégie d'agrégation directionnelle des descripteurs.

Soit $\alpha \in [0, \pi]$ un angle d'orientation pour lequel nous souhaitons effectuer une évaluation. Le principe de cette approche consiste à calculer l'aire sous la courbe du descripteur \mathcal{H}_{AB} , le long d'un intervalle de directions centré autour de α . De cette façon, une évaluation directionnelle sur l'intervalle $[\alpha - \frac{\omega}{2}, \alpha + \frac{\omega}{2}]$, $\omega \in \mathbb{R}$, est donnée par :

$$\frac{1}{\omega} \int_{\alpha - \frac{\omega}{2}}^{\alpha + \frac{\omega}{2}} \mathcal{H}_{AB}(\theta) d\theta. \quad (6.4)$$

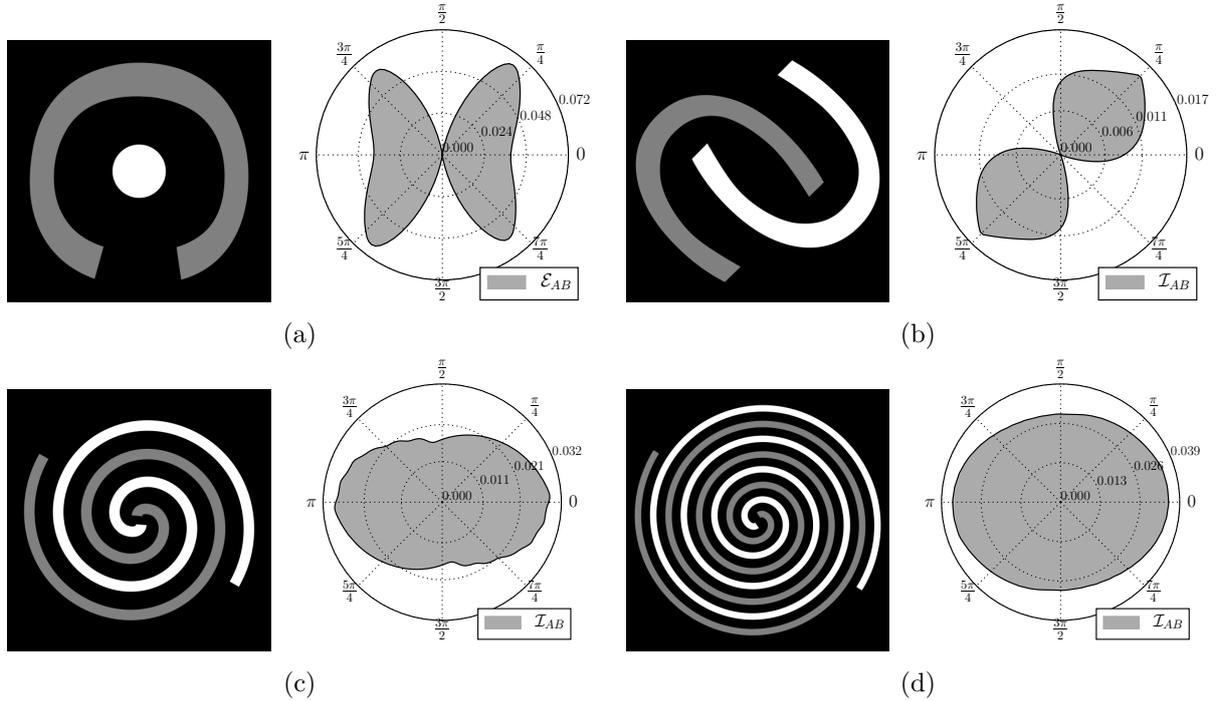


Figure 6.2 – Représentations polaires des descripteurs directionnels (a) d'enlacement \mathcal{E}_{AB} , et (b, c, d) d'entrelacement \mathcal{I}_{AB} , obtenus pour différentes configurations spatiales entre deux objets (en blanc : objet A ; en gris : objet B). Rappel de la Figure 4.8.

La valeur de α correspond à la direction à laquelle on s'intéresse en particulier, et le paramètre ω permet de contrôler la taille de l'intervalle, et donc la finesse de l'évaluation. Cette formulation permet alors de se focaliser spécifiquement sur une direction, ou bien d'effectuer une évaluation plus globale.

Par exemple, une évaluation sur l'intervalle $[0, \frac{\pi}{2}]$ permet d'évaluer la relation selon l'axe diagonal, selon le quadrant supérieur droit (qui correspond également au quadrant inférieur gauche, de par la symétrie des descripteurs d'enlacement et d'entrelacement). Une autre possibilité à partir de cette approche est de considérer tout l'intervalle $[0, \pi]$, de manière globale. On évalue alors la quantité globale d'enlacement ou d'entrelacement au sein de la configuration spatiale étudiée, sans se focaliser sur une direction particulière.

6.1.3 Résultats illustratifs

Le Tableau 6.1 montre les intervalles de nécessité-possibilité obtenus avec l'approche par mise en correspondance d'ensembles flous, pour les configurations spatiales de la Figure 6.2. Nous présentons les résultats obtenus pour quatre orientations α dans l'intervalle $[0, \pi]$, du fait de la π -périodicité des descripteurs. Pour la configuration (a), le descripteur d'enlacement \mathcal{E}_{AB} est utilisé, tandis que pour les configurations (b), (c) et (d), nous utilisons le descripteur d'entrelacement \mathcal{I}_{AB} . Ces résultats sont cohérents avec le modèle : les valeurs de possibilité permettent

6.1. Évaluations à partir de descripteurs

Tableau 6.1 – Évaluations de l’enlacement et de l’entrelacement entre objets dans différentes directions α , pour l’approche par mise en correspondance d’ensembles flous (intervalles $[N_\alpha; \Pi_\alpha]$). Les lignes (a), (b), (c) et (d) font référence aux configurations spatiales de la Figure 6.2.

	$\alpha = 0$	$\alpha = \frac{\pi}{4}$	$\alpha = \frac{\pi}{2}$	$\alpha = \frac{3\pi}{4}$
(a) \mathcal{E}_{AB}	[0,23 ; 0,76]	[0,06 ; 0,99]	[0,24 ; 0,78]	[0,01 ; 0,94]
(b) \mathcal{I}_{AB}	[0,17 ; 0,82]	[0,43 ; 1,00]	[0,19 ; 0,83]	[0,00 ; 0,58]
(c) \mathcal{I}_{AB}	[0,28 ; 1,00]	[0,11 ; 0,89]	[0,00 ; 0,72]	[0,11 ; 0,89]
(d) \mathcal{I}_{AB}	[0,14 ; 1,00]	[0,07 ; 0,95]	[0,00 ; 0,86]	[0,05 ; 0,93]

d’indiquer la confiance que l’on a dans la présence d’enlacement ou d’entrelacement dans une direction particulière, alors que les valeurs de nécessité vérifient uniquement si le descripteur correspond bien au profil directionnel. Par exemple, pour la configuration (b), nous obtenons une valeur de nécessité de 0,43 dans la direction $\alpha = \frac{\pi}{4}$ car le descripteur d’entrelacement est très similaire au profil directionnel dans cette direction. Selon ce modèle, les valeurs de nécessité sont faibles pour toutes les directions des configurations (c) et (d), car les histogrammes d’entrelacement correspondants ont des valeurs relativement élevées dans toutes les directions : il ne se dégage donc pas d’orientation dominante. Cette approche produit donc des évaluations cohérentes, mais elle ne permet pas de mesurer convenablement une quantité globale d’enlacement ou d’entrelacement qui serait présente dans une direction.

Le Tableau 6.2 présente les évaluations obtenues pour l’approche par agrégation directionnelle, pour les configurations d’objets de la Figure 6.2. Pour chacune des configurations, les évaluations sont calculées pour des intervalles de directions successifs centrés en $\alpha \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi\}$, avec une amplitude fixée à $\omega = \frac{\pi}{2}$. Les directions et l’amplitude correspondent alors aux différentes colonnes du tableau, représentées par l’intervalle $[\alpha - \frac{\omega}{2}, \alpha + \frac{\omega}{2}]$. Nous calculons également une évaluation globale sur tout l’intervalle $[0, \pi]$. Pour la configuration (a), l’évaluation est celle pour le descripteur d’enlacement \mathcal{E}_{AB} . Nous pouvons observer un enlacement plus élevé pour les directions horizontales (c’est-à-dire pour l’intervalle $[-\frac{\pi}{4}, \frac{\pi}{4}]$), ainsi qu’un enlacement global de 0,61. La valeur d’enlacement plus faible pour l’orientation verticale est ainsi due à l’ouverture

Tableau 6.2 – Évaluation de l’enlacement et de l’entrelacement d’objets dans différentes directions, pour l’approche par agrégation directionnelle (Équation 6.4). Les lignes (a), (b), (c) et (d) font référence aux configurations spatiales de la Figure 6.2.

	$[-\frac{\pi}{4}, \frac{\pi}{4}]$	$[0, \frac{\pi}{2}]$	$[\frac{\pi}{4}, \frac{3\pi}{4}]$	$[\frac{\pi}{2}, \pi]$	$[0, \pi]$
(a) \mathcal{E}_{AB}	0,71	0,62	0,50	0,59	0,61
(b) \mathcal{I}_{AB}	0,47	0,81	0,50	0,15	0,48
(c) \mathcal{I}_{AB}	0,90	0,80	0,66	0,76	0,78
(d) \mathcal{I}_{AB}	0,96	0,91	0,85	0,89	0,90

de l'objet B située vers le bas, traduisant ainsi le fait que A est moins enlacé le long de cet axe. Nous présenterons plus tard dans la [Section 6.3](#) une approche spécifique pour évaluer avec plus de finesse ce type de configurations d'*entourage*. Pour la configuration (b), nous retrouvons bien une valeur d'entrelacement élevée pour l'intervalle $[0, \frac{\pi}{2}]$, traduisant le fait que les objets sont entrelacés selon l'axe diagonal, tandis que l'entrelacement est plus faible pour la diagonale opposée. Finalement, les spirales des configurations (c) et (d) présentent un fort entrelacement global (0,78 et 0,90 respectivement, pour l'intervalle $[0, \pi]$), ainsi que des valeurs plus élevées le long de l'axe horizontal. Cette orientation dominante est principalement due aux terminaisons extérieures des spirales, qui entraînent alors un niveau d'entrelacement supplémentaire entre les deux objets dans cette direction.

6.2 Paysages flous directionnels d'enlacement

Dans cette section, nous proposons une autre approche d'évaluation des relations d'enlacement et d'entrelacement. Cette approche est inspirée des travaux de [\[Bloch, 1999\]](#) sur les paysages flous directionnels de relations spatiales. Nous définissons les paysages flous directionnels d'enlacement, qui permettent de visualiser directement dans l'espace image les degrés de validité de cette relation. Ces paysages peuvent alors être utilisés pour effectuer des évaluations quantitatives locales de configurations d'enlacement.

6.2.1 Définitions et évaluations floues

Nous cherchons ici à caractériser dans quelle mesure chaque région de l'espace est enlacée par un objet A . Pour cela nous proposons de construire différents paysages flous d'enlacement qui permettent d'évaluer la quantité d'enlacement de l'objet A en chaque point localisé en dehors de cet objet. Le modèle d'enlacement proposé précédemment dans le [Chapitre 4](#) étant essentiellement directionnel, nous proposons alors naturellement de définir des paysages flous d'enlacement selon le même point de vue.

Définitions

Nous nous appuyons ici sur la définition de l'enlacement le long d'une droite, introduite dans le [Chapitre 4](#). Nous proposons d'adapter cette définition afin de calculer l'enlacement directionnel en un point donné. Soit A un objet binaire du plan, représenté par sa fonction caractéristique $f_A : \mathbb{R}^2 \rightarrow \{0, 1\}$. Dans une direction $\theta \in \mathbb{R}$, et pour un point en dehors de A localisé aux coordonnées (ρ, t) dans le plan après rotation d'angle θ , l'enlacement directionnel local peut être défini par :

$$\mathcal{E}_A(\theta)(\rho, t) = \frac{1}{\|A\|_1} \int_t^{+\infty} f_A^{(\theta, \rho)}(x) dx \int_{-\infty}^t f_A^{(\theta, \rho)}(x) dx. \quad (6.5)$$

Ainsi, l'image correspondant à $\mathcal{E}_A(\theta)$ peut être interprétée comme un paysage représentant les valeurs d'enlacement directionnel en tout point situé en dehors de l'objet A . Cette image peut alors être normalisée par son point de valeur maximum afin d'obtenir des valeurs dans l'intervalle

6.2. Paysages flous directionnels d'enlacement

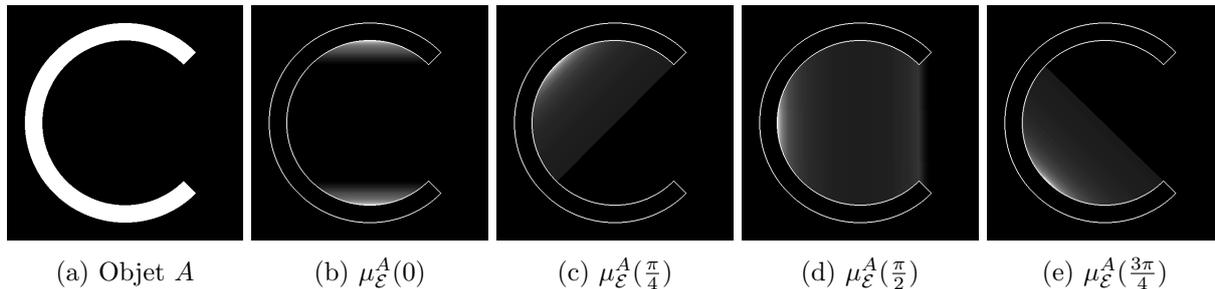


Figure 6.3 – Exemples de paysages flous directionnels d'enlacement (*Fuzz-DELs*) d'un objet binaire A orientés dans différentes directions. En (b), (c), (d) et (e), les contours de l'objet A sont affichés en blanc.

$[0, 1]$, et ainsi être considérée comme un ensemble flou de \mathbb{R}^2 . En ce sens, nous définissons le paysage flou directionnel d'enlacement $\mu_{\mathcal{E}}^A(\theta)$ (alors abrégé « *Fuzz-DEL* », pour *fuzzy directional enlacement landscape*) de l'objet A par :

$$\mu_{\mathcal{E}}^A(\theta)(\rho, t) = \frac{\mathcal{E}_A(\theta)(\rho, t)}{\max_{\rho, t} \mathcal{E}_A(\theta)(\rho, t)}. \quad (6.6)$$

Afin d'illustrer ces définitions, la [Figure 6.3](#) présente les paysages flous d'enlacement obtenus pour un objet binaire A , pour différentes directions. De tels paysages flous permettent donc de visualiser et d'évaluer dans quelle mesure chaque point est enlacé par l'objet A dans une direction θ donnée. Nous pouvons notamment remarquer que les valeurs non nulles d'un *Fuzz-DEL* se situent nécessairement dans les zones de concavités de l'objet. Ceci est particulièrement intéressant d'un point de vue algorithmique, car cela permet de limiter le calcul aux seuls points situés dans $\text{CH}(A) \setminus A$, où $\text{CH}(A)$ désigne l'enveloppe convexe de l'objet A . Un autre point à mettre en évidence ici est que les *Fuzz-DELs* sont symétriques et π -périodiques, c'est-à-dire que l'on a $\mu_{\mathcal{E}}^A(\theta + \pi) = \mu_{\mathcal{E}}^A(\theta)$.

Nous avons donc défini un *Fuzz-DEL* pour une direction θ fixée, permettant ainsi d'évaluer la relation d'enlacement dans une direction bien spécifique. Cependant, comme cela a été présenté dans les sections précédentes pour la génération d'évaluations à partir des descripteurs, il peut également être intéressant de disposer d'un contrôle sur la granularité directionnelle pour analyser les configurations spatiales. De plus, le fait de ne considérer qu'une seule direction au sein de cette approche travaillant directement dans l'image, peut donner lieu à des erreurs de discrétisation liées à la grille de pixels. En ce sens, nous proposons d'agréger plusieurs paysages flous d'enlacement focalisés sur différentes directions. Soit $\theta \in [0, \pi]$ et $\omega \in \mathbb{R}$. Le *Fuzz-DEL* agrégé sur l'intervalle $[\theta - \frac{\omega}{2}, \theta + \frac{\omega}{2}]$ est alors défini par :

$$\mu_{\mathcal{E}}^A(\theta, \omega)(\rho, t) = \frac{1}{\omega} \int_{\theta - \frac{\omega}{2}}^{\theta + \frac{\omega}{2}} \mu_{\mathcal{E}}^A(\alpha)(\rho, t) d\alpha, \quad (6.7)$$

où θ représente la direction autour de laquelle le *Fuzz-DEL* est centré, tandis que ω contrôle

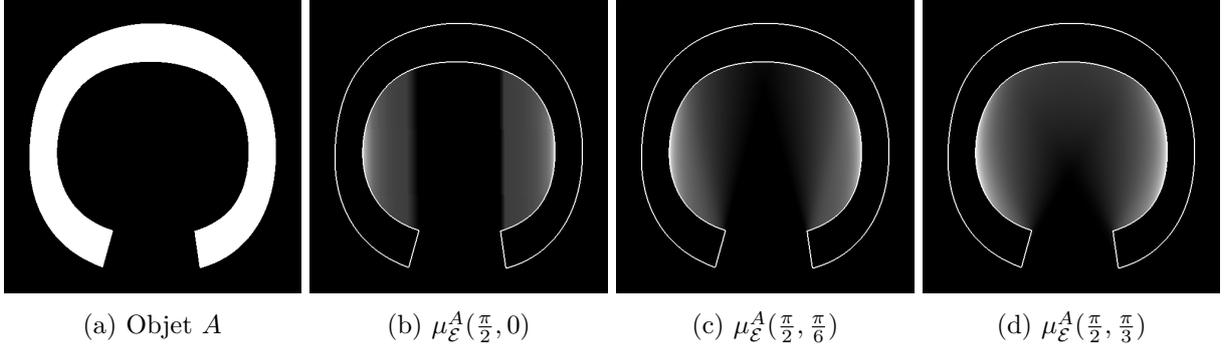


Figure 6.4 – Paysages flous directionnels d'enlacement (*Fuzz-DELS*) d'un objet binaire A , pour une direction fixée à $\theta = \frac{\pi}{2}$ et pour des amplitudes croissantes $\omega \in \{0, \frac{\pi}{6}, \frac{\pi}{3}\}$. En (b), (c) et (d), les contours de l'objet A sont affichés en blanc.

l'amplitude de l'angle de vue. Ceci permet d'analyser la situation d'un point de vue très spécifique (pour une valeur de ω faible) ou bien d'un point de vue global (pour une valeur de ω élevée). En particulier, nous proposons de définir le *Fuzz-DEL* qui agrège toutes les directions sur l'intervalle $[0, \pi]$, que nous désignons par $\tilde{\mu}_{\mathcal{E}}^A = \mu_{\mathcal{E}}^A(\frac{\pi}{2}, \pi)$.

Dans le but d'illustrer ce formalisme d'agrégation directionnelle, la Figure 6.4 et la Figure 6.5 montrent différents *Fuzz-DELS* obtenus pour deux objets. D'une part, la Figure 6.4 permet d'illustrer l'influence du paramètre ω qui contrôle l'amplitude de l'intervalle pour une direction donnée, ici la direction verticale avec $\theta = \frac{\pi}{2}$. Notons que les paysages flous pour la direction opposée en $\theta = \frac{3\pi}{2}$ seraient identiques du fait de la propriété de symétrie. Pour une amplitude de $\omega = 0$ (ce qui correspond alors aux mêmes paysages que ceux de la Figure 6.4), nous pouvons visualiser les valeurs d'enlacement nulles localisées sur la zone centrale de l'image. Ces valeurs nulles traduisent le fait que si un autre objet était localisé dans cette zone de l'image, il ne serait pas enlacé par l'objet A selon l'axe vertical (il pourrait s'échapper de la situation par le bas sans avoir à traverser l'objet A). Nous pouvons ensuite observer l'influence du paramètre d'amplitude ω , qui provoque un effet de lissage progressif sur le paysage, lorsque l'on considère un éventail de plus en plus large de directions. D'autre part, la Figure 6.5 montre les paysages flous d'enlacement obtenus pour différentes directions, avec une amplitude fixée à $\omega = \frac{\pi}{3}$. Sur ces exemples, nous pouvons observer que les paysages flous d'enlacement permettent de visualiser les concavités de l'objet d'un point de vue directionnel. Dans la direction horizontale (pour $\theta = 0$), les valeurs d'enlacement locales sont relativement élevées (les valeurs des pixels sont proches du blanc), et le sont de plus en plus au fur et à mesure que l'on s'enfonce au sein de la forme. À l'inverse, le paysage flou d'enlacement pour la direction verticale (pour $\theta = \frac{\pi}{2}$) est nul presque partout, sauf pour quelques concavités de l'objet.

6.2. Paysages flous directionnels d'enlacement

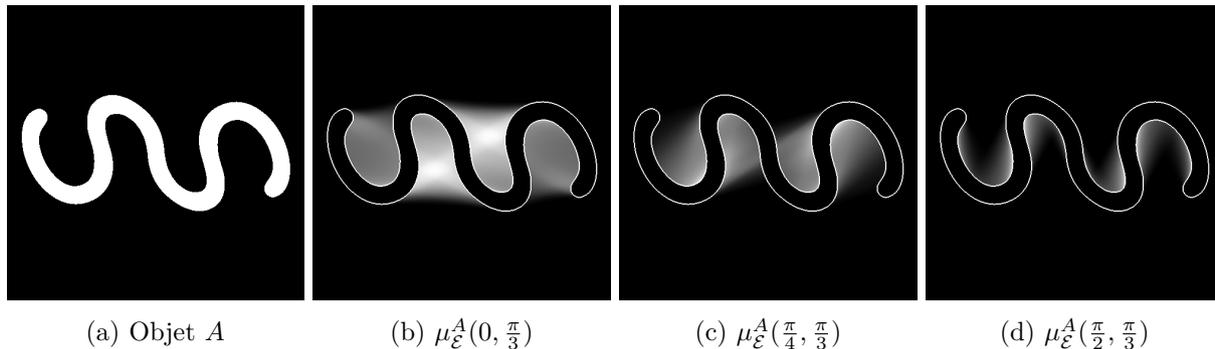


Figure 6.5 – Paysages flous directionnels d'enlacement (*Fuzz-DELS*) d'un objet binaire A , pour différentes directions $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}\}$ et pour une amplitude fixée à $\omega = \frac{\pi}{3}$. En (b), (c) et (d), les contours de l'objet A sont affichés en blanc.

Évaluations floues

Dans les définitions des paysages flous d'enlacement directionnels présentées ci-avant, un objet binaire A est considéré, et le paysage flou associé permet alors d'évaluer, en chaque point, l'enlacement local selon une certaine amplitude de directions. Dans ce qui suit, nous montrons comment exploiter ces paysages flous pour évaluer à quel point un autre objet binaire B (objet argument) est enlacé par l'objet A (objet de référence), en utilisant notamment différents opérateurs flous classiques. En ce sens, l'évaluation peut être effectuée de manière similaire à l'approche d'évaluation à partir des descripteurs, la différence étant qu'ici les ensembles flous considérés sont des paysages dans \mathbb{R}^2 (tandis que les descripteurs étaient assimilés à des ensembles flous de \mathbb{R}).

Soit μ_A et μ_B deux ensembles flous de \mathbb{R}^2 . Nous rappelons la formulation des mesures de nécessité-possibilité :

$$\Pi(\mu_A, \mu_B) = \sup_{x,y} t(\mu_A(x, y), \mu_B(x, y)), \quad (6.8)$$

$$N(\mu_A, \mu_B) = \inf_{x,y} T(1 - \mu_A(x, y), \mu_B(x, y)), \quad (6.9)$$

où t désigne un opérateur d'intersection floue (*t-norme*) et T est un opérateur d'union floue (*t-conorme*). De la même manière que pour l'approche d'évaluation à partir des descripteurs, nous avons choisi les opérateurs min et max comme *t-norme* et *t-conorme* respectives.

Dans notre contexte, cette mesure d'appariement floue est appliquée pour évaluer à quel point un objet argument B (représenté par sa fonction d'appartenance μ_B) est compatible avec un *Fuzz-DEL* $\mu_{\mathcal{E}}^A(\theta, \omega)$ issu d'un objet de référence A . L'intervalle de nécessité-possibilité $[N(\mu_{\mathcal{E}}^A(\theta, \omega), \mu_B), \Pi(\mu_{\mathcal{E}}^A(\theta, \omega), \mu_B)]$ constitue donc une évaluation floue de la proposition « B est enlacé par A dans la direction θ ». La nécessité représente un point de vue dit pessimiste, tandis que la possibilité constitue un point de vue optimiste. La valeur moyenne $M(\mu_{\mathcal{E}}^A(\theta, \omega), \mu_B)$ (au

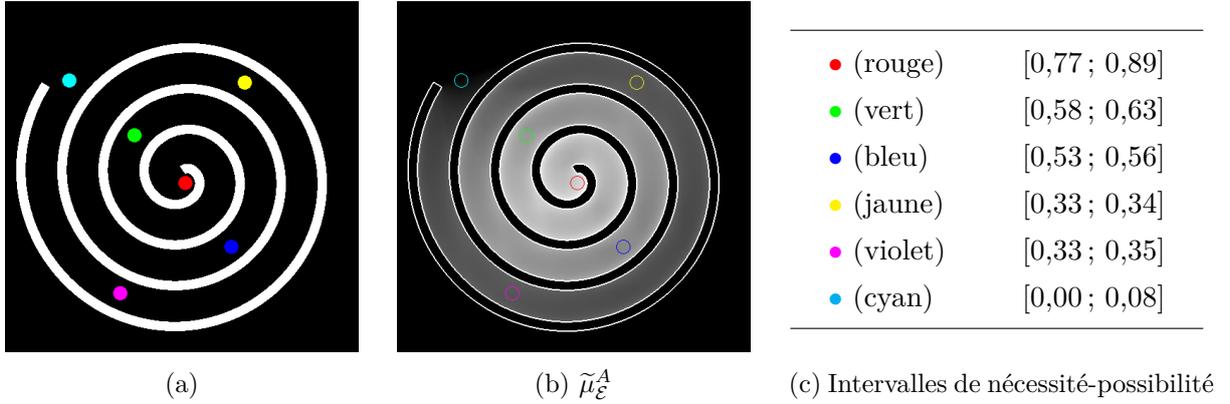


Figure 6.6 – Paysage flou d'enlacement global d'une spirale (objet de référence A) et évaluations par nécessité-possibilité pour différents objets arguments placés à l'intérieur de la spirale (représentés en différentes couleurs).

sens de la t -norme min) peut également être considérée, celle-ci étant définie par :

$$M(\mu_{\mathcal{E}}^A(\theta, \omega), \mu_B) = \frac{\sum_{\rho, t} \min(\mu_{\mathcal{E}}^A(\theta, \omega)(\rho, t), \mu_B(\rho, t))}{\sum_{\rho, t} \mu_B(\rho, t)}. \quad (6.10)$$

Ces mesures peuvent donc être utilisées pour évaluer quantitativement à quel point un objet argument B est enlacé par un objet A de référence, en fonction de différentes directions.

6.2.2 Résultats illustratifs

La Figure 6.6 présente un exemple illustratif d'évaluation de la relation spatiale « être enlacé par » pour une spirale considérée comme objet de référence A , et pour différents cercles colorés placés à différentes positions à l'intérieur de la spirale, considérés alors comme des objets cibles. Ici, la relation spatiale est envisagée de manière globale : le $Fuzz-DEL$ calculé pour la spirale est celui qui agrège toutes les directions (c'est-à-dire le paysage flou $\tilde{\mu}_{\mathcal{E}}^A$). En particulier, cet exemple nous permet d'illustrer la capacité du modèle à capturer de multiples degrés de profondeur dans les concavités d'un objet.

À partir du paysage flou $\tilde{\mu}_{\mathcal{E}}^A$ représenté en Figure 6.6 (b), nous pouvons en effet observer un motif décroissant de l'enlacement local lorsque l'on part du centre de la spirale et que l'on suit un chemin circulaire (les valeurs des pixels passent progressivement du blanc au noir). Afin d'évaluer quantitativement ce comportement, nous avons calculé les intervalles de nécessité-possibilité $[N(\tilde{\mu}_{\mathcal{E}}^A, \mu_B), \Pi(\tilde{\mu}_{\mathcal{E}}^A, \mu_B)]$ pour chacun des différents cercles de couleurs placés à différents emplacements à l'intérieur de la spirale (chacun étant considéré comme un objet B par rapport au paysage flou global). Les intervalles de nécessité-possibilité obtenus sont répertoriés en Figure 6.6 (c). Nous pouvons ainsi observer un enlacement élevé au centre de la spirale (cercle rouge), puis un comportement décroissant des intervalles au fur et à mesure que l'on s'éloigne du centre, jusqu'à obtenir des valeurs proches de 0 pour le cercle cyan.

6.2. Paysages flous directionnels d'enlacement

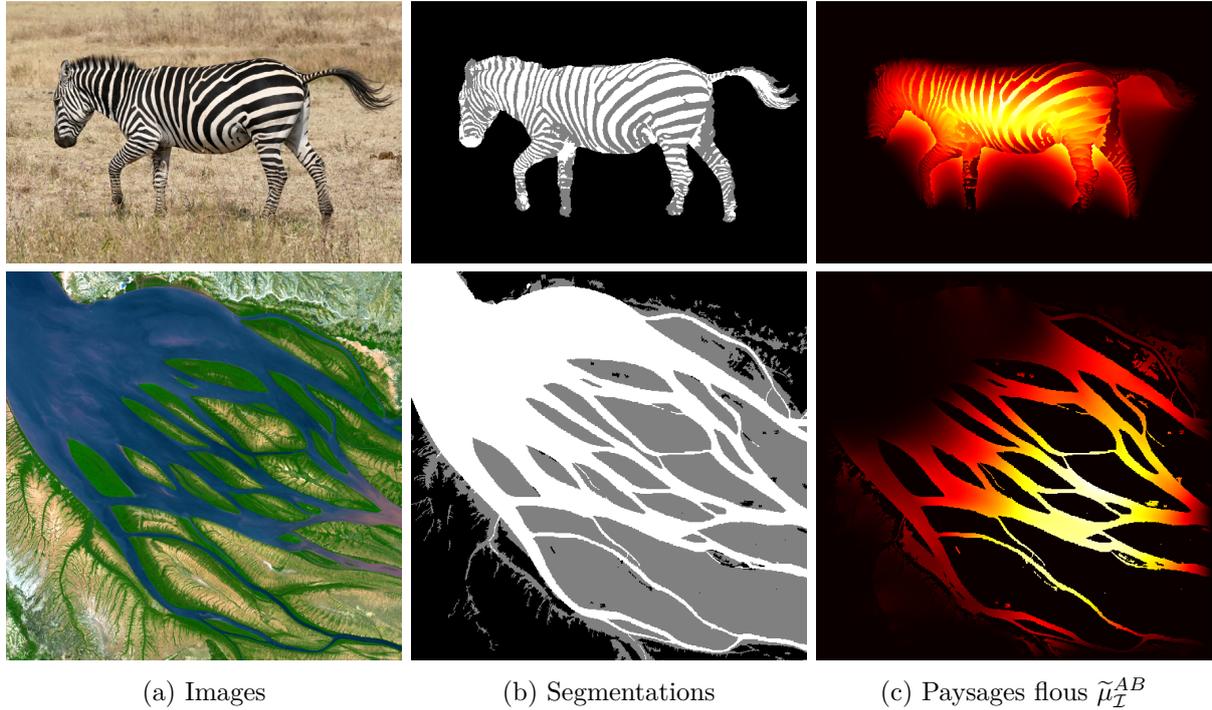


Figure 6.7 – Exemples de paysages flous d’entrelacement pour deux images illustratives. Les paysages d’entrelacement $\tilde{\mu}_T^{AB}$ sont affichés sous la forme de cartes de chaleur.

Notons ici que les méthodes classiques d’évaluations de relations spatiales ne seraient pas en mesure de prendre en compte ces différents niveaux de profondeur au sein de la spirale. Par exemple, l’approche proposée par [Vanegas *et al.*, 2011] pour évaluer la relation « entouré par » ne considère que les points de l’objet de référence qui sont visibles depuis un objet argument. Elle ne permet donc pas de considérer la situation dans son ensemble. Dans la Section 6.3, nous utiliserons notamment cette approche comme élément de comparaison pour évaluer cette relation spécifique.

6.2.3 Vers les paysages flous d’entrelacement

Nous proposons ici quelques résultats préliminaires pour la définition de paysages flous d’entrelacement. En agrégeant le long de toutes les directions, un paysage flou d’entrelacement entre deux objets A et B peut être obtenu par :

$$\tilde{\mu}_T^{AB} = \tilde{\mu}_E^A + \tilde{\mu}_E^B. \quad (6.11)$$

La Figure 6.7 présente quelques exemples de paysages flous d’entrelacement obtenus pour deux images illustratives. Chaque image a été segmentée en trois classes, avec en noir le fond (qui est alors ignoré), puis en blanc et en gris les objets A et B respectivement.

Le premier paysage d’entrelacement est obtenu à partir de la segmentation d’une image re-

présentant un zèbre dont la robe est composée de rayures alternées. Nous pouvons ainsi observer des valeurs d'entrelacement élevées au centre de la robe de cet animal. Le deuxième paysage d'entrelacement est obtenu à partir d'une image satellitaire ASTER¹ représentant un large delta de rivière. Nous pouvons alors observer que l'entrelacement est principalement localisé au niveau des multiples ramifications entre la rivière (en blanc) et les marais (en gris). De telles visualisations de l'entrelacement entre ces types d'objets peuvent être utiles par exemple pour le suivi écologique et la surveillance de l'évolution des paysages.

6.3 Applications à la relation spatiale « entouré par »

Dans cette section, nous proposons d'appliquer les différentes approches que nous avons présentées dans ce chapitre pour l'évaluation spécifique de la relation spatiale « entouré par ». Comme nous l'avons explicité dans la [Partie I](#) du manuscrit, cette relation est intrinsèquement difficile à évaluer quantitativement, bien qu'elle soit appréhendée relativement facilement par la perception humaine. En effet, celle-ci peut être catégorisée à la limite entre différents types de relations spatiales : d'une part, il s'agit d'une relation spatiale se rapprochant de notions topologiques, et d'autre part, elle est fortement reliée à des concepts directionnels.

6.3.1 Nouvelles modélisations à partir de l'enlacement

Entourement à partir des descripteurs d'enlacement

Dans un premier temps, nous proposons de dériver une évaluation de la proposition « *A est entouré par B* » à partir du descripteur directionnel d'enlacement \mathcal{E}_{AB} . Une telle évaluation peut être représentée par un nombre réel $\mathcal{S}_{AB} \in [0, 1]$ qui correspond au degré de validité accordé à cette proposition.

Pour cette approche, de la même façon qu'en [Section 6.1](#), le descripteur \mathcal{E}_{AB} est tout d'abord normalisé par sa valeur maximum, et peut alors être interprété comme une répartition de la quantité d'enlacement de *A* par *B* le long des différentes directions de l'intervalle $\theta \in [0, \pi]$. Pour $\alpha \in [0, 1]$, nous définissons alors une α -coupe de \mathcal{E}_{AB} par $\mathcal{E}_{AB}^\alpha(\theta) = \min(\alpha, \mathcal{E}_{AB}(\theta))$. Le principe consiste alors à calculer l'aire globale sous la courbe d'une α -coupe \mathcal{E}_{AB}^α , la valeur de α pouvant être interprétée comme un seuil de tolérance. Nous définissons le degré de validité accordé à la proposition « *A est entouré par B* » avec une tolérance α par :

$$\mathcal{S}_{AB}(\alpha) = \begin{cases} \frac{1}{\pi} \int_0^\pi 1 - \delta(\mathcal{E}_{AB}(\theta)) \, d\theta & \text{si } \alpha = 0, \\ \frac{1}{\alpha\pi} \int_0^\pi \mathcal{E}_{AB}^\alpha(\theta) \, d\theta & \text{sinon,} \end{cases} \quad (6.12)$$

où δ correspond à une fonction prenant la valeur 1 en 0, et valant 0 partout ailleurs.

Ainsi, pour une tolérance de $\alpha = 0$, l'objet argument est considéré comme entouré dans une direction θ à partir du moment où $\mathcal{E}_{AB}(\theta) > 0$, et seules les directions pour lesquelles

1. U.S./Japan ASTER Science Team, NASA/GSFC/METI/ERSDAC/JAROS.

6.3. Applications à la relation spatiale « entouré par »

l'enlacement est nul peuvent diminuer la valeur d'entourement $\mathcal{S}_{AB}(0)$. De manière générale, pour $0 < \alpha \leq 1$, dès lors que $\mathcal{E}_{AB}(\theta) \geq \alpha$, l'objet A est considéré comme entouré dans cette direction. Afin d'enrichir cette approche et de disposer de différents points de vue pour analyser la configuration, nous proposons de calculer $\mathcal{S}_{AB}(\alpha)$ pour plusieurs seuils de tolérance $\alpha \in [0, 1]$. En effet, l'évolution des valeurs d'entourement $\mathcal{S}_{AB}(\alpha)$ permet d'interpréter la situation de différentes façons (d'un point de vue pessimiste à optimiste), et ainsi de mieux prendre en compte la complexité de cette relation spatiale.

Entourement à partir des paysages flous d'enlacement

La deuxième approche proposée pour évaluer la relation spatiale « entouré par » est fondée sur une application spécifique du modèle des paysages flous d'enlacement présentés dans la [Section 6.2](#).

Pour cette application, l'objet argument A est projeté sur un *Fuzz-DEL* de l'objet B de référence, puis normalisé en un ensemble flou. Une telle projection est définie par :

$$\mu_{\mathcal{E}}^{AB}(\theta, \omega) = \frac{\min_{\rho, t} (\mu_{\mathcal{E}}^A(\theta, \omega)(\rho, t), \mu_B(\rho, t))}{\max_{\rho, t} \mu_{\mathcal{E}}^A(\theta, \omega)(\rho, t)}. \quad (6.13)$$

Cette projection est effectuée pour différentes directions $\theta \in [0, \pi]$ afin d'obtenir des mesures de nécessité $N(\mu_{\mathcal{E}}^{AB}(\theta, \omega), \mu_B)$ et possibilité $\Pi(\mu_{\mathcal{E}}^{AB}(\theta, \omega), \mu_B)$. Ces valeurs peuvent alors être interprétées comme des profils directionnels, respectivement pessimistes et optimistes, qui sont alors exploités pour dériver des évaluations globales de la proposition « A est entouré par B ». Ces évaluations globales sont obtenues par :

$$N_S^{AB} = \frac{1}{\pi} \int_0^\pi N(\mu_{\mathcal{E}}^{AB}(\theta, \omega), \mu_B) d\theta, \quad (6.14)$$

$$\Pi_S^{AB} = \frac{1}{\pi} \int_0^\pi \Pi(\mu_{\mathcal{E}}^{AB}(\theta, \omega), \mu_B) d\theta. \quad (6.15)$$

Pour la suite de cette étude, nous avons fixé l'amplitude des paysages flous directionnels à une valeur faible de $\frac{\pi}{36}$ (correspondant à 5 degrés) afin de considérer les différentes directions individuellement, tout en lissant les problèmes de discrétisation liés à la grille de pixels.

Approches comparatives

Nous avons proposé deux approches spécifiques pour évaluer la relation « entouré par », l'une fondée sur le descripteur d'enlacement avec différents seuils de tolérance, l'autre exploitant le modèle des paysages flous directionnels d'enlacement. Nous détaillons ici les deux méthodes de référence employées afin de comparer les résultats obtenus par les approches que nous proposons. Ces deux méthodes comparatives ont notamment été évoquées dans le [Chapitre 3](#).

La première méthode comparative est celle proposée par [[Matsakis et Andréfouët, 2002](#)], qui est fondée sur l'histogramme de forces. Dans cette approche, l'histogramme de forces \mathcal{F}_{AB}

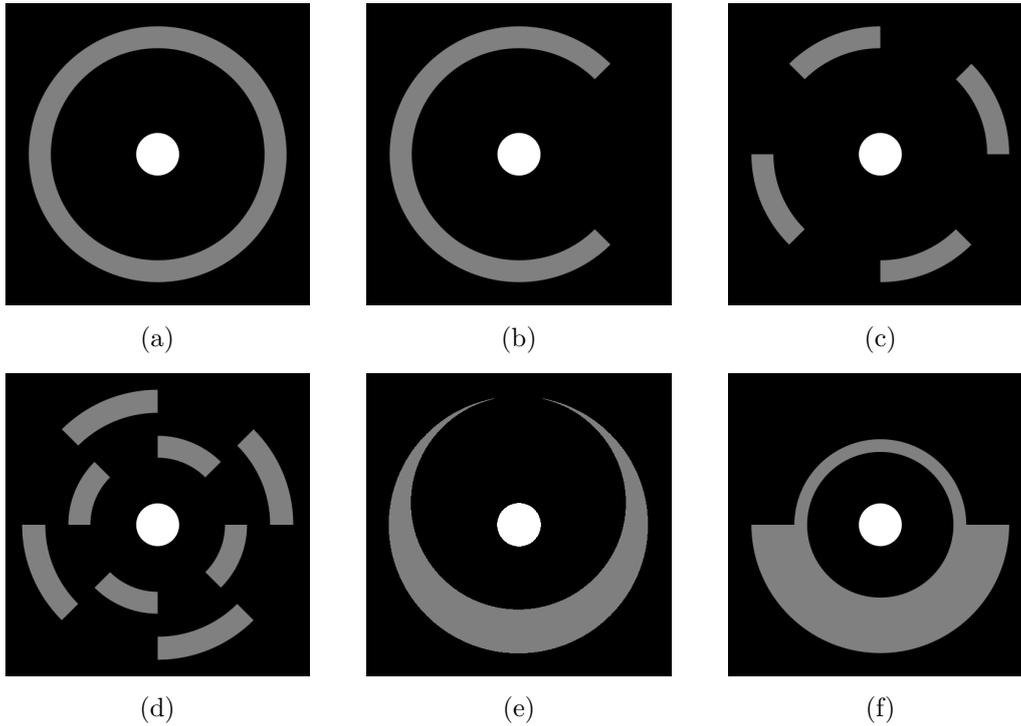


Figure 6.8 – Exemples caractéristiques de situations d'entourment entre deux objets (en gris : objet de référence B ; en blanc : objet cible A).

résumant la position de A par rapport à B est calculé, et une mesure floue est extraite de l'histogramme en fonction de l'amplitude d'angles la plus grande où l'objet B ne se trouve pas par rapport à A . Cette mesure est effectuée pour différents niveaux d' α -coupes de l'histogramme de forces, puis une étape d'agrégation par simple somme [Dubois et Jaulent, 1987] vient résumer la situation en une évaluation notée $s(A, B) \in [0, 1]$.

La deuxième approche comparative considérée est celle proposée par [Vanegas et al., 2011]. Cette approche est inspirée des paysages flous de [Bloch, 1999] pour les relations spatiales directionnelles, ainsi que des travaux de [Rosenfeld et Klette, 1985]. L'approche propose de construire un paysage flou dédié à la relation d'entourment, noté μ_{surround} , en considérant en chaque point la couverture angulaire pour laquelle l'objet de référence est visible. L'évaluation pour un objet cible est effectuée en projetant cet objet sur le paysage d'entourment, puis en utilisant un intervalle de nécessité-possibilité $[N, \Pi]$ ainsi qu'une mesure moyenne M .

6.3.2 Exemples caractéristiques

La Figure 6.8 présente quelques exemples caractéristiques de configurations spatiales où un objet est entouré par un autre. Pour chaque image, l'objet B est en gris et correspond à l'objet de référence, et l'objet A est en blanc et correspond à l'objet argument. Nous cherchons donc à évaluer dans quelle mesure « A est entouré par B » pour chacune des configurations.

6.3. Applications à la relation spatiale « entouré par »

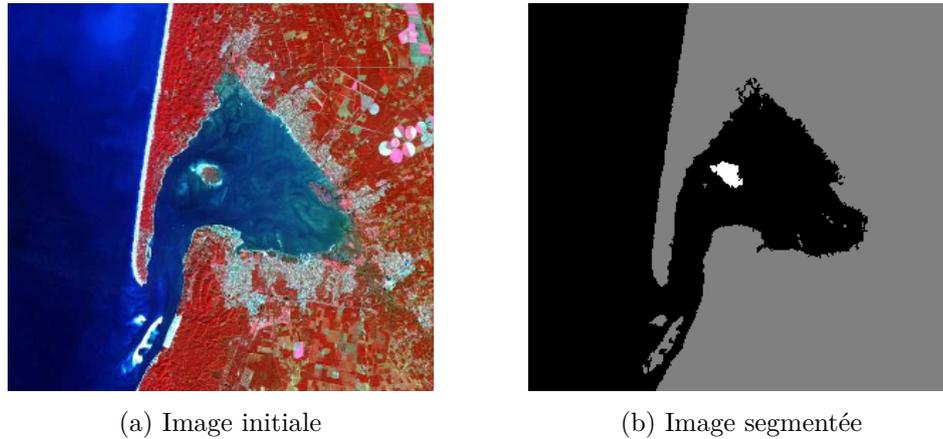


Figure 6.9 – Exemple de situation d’entourement pour une image satellitaire représentant le bassin d’Arcachon, dans la région de la Gironde (France). En (b), l’objet B en gris correspond aux régions côtières et l’objet argument A en blanc correspond à l’Île aux oiseaux.

Ces situations, bien que relativement simples à interpréter pour un être humain, sont riches d’ambiguïtés et peuvent être analysées selon différents points de vue. Nous verrons que les approches d’évaluation proposées dans les sections précédentes permettent justement de capturer cette pluralité des points de vue envisageables pour analyser ces configurations spatiales.

La Figure 6.9 présente un exemple concret pour l’analyse de données satellitaires. Nous illustrons ici une image représentant la région du bassin d’Arcachon en Gironde (France). Cette image satellitaire a été segmentée afin de délimiter un objet correspondant à l’Île aux oiseaux (objet A , en blanc) et un autre correspondant à la région côtière (objet B , en gris). Nous nous intéressons alors à l’entourement de cette île par rapport à la côte.

6.3.3 Résultats expérimentaux et discussion

Nous présentons ici les différents résultats expérimentaux obtenus pour les exemples caractéristiques de la Figure 6.8 (configurations (a), (b), (c), (d), (e) et (f)) et de la Figure 6.9 (notée configuration (g) dans la suite). L’objectif consiste donc à évaluer dans quelle mesure l’objet A est entouré par l’objet B pour ces différentes configurations spatiales. De telles évaluations étant fondamentalement subjectives, il est difficile de construire des critères de comparaison et de validation qui permettraient de juger objectivement de l’efficacité d’une méthode par rapport à une autre, comme par exemple lorsque l’on dispose d’une vérité terrain. En ce sens, nous chercherons plutôt ici à illustrer la richesse d’information portée par les méthodes proposées, et à discuter les particularités de chaque configuration.

En ce sens, le Tableau 6.3 propose une synthèse des résultats obtenus pour les différentes configurations spatiales, et pour toutes les approches d’évaluation proposées. Chaque ligne du tableau correspond à une configuration spatiale, et les colonnes désignent les différentes approches considérées :

Tableau 6.3 – Synthèse des résultats obtenus pour l'évaluation de la relation spatiale « être entouré par » à partir des approches proposées et des méthodes comparatives. Les lignes (a), (b), (c), (d), (e) et (f) correspondent aux configurations de la Figure 6.8, et la ligne (g) correspond à la configuration de la Figure 6.9.

	\mathcal{F}_{AB}	μ_{surround}^B		$\mathcal{S}_{AB}(\alpha)$		$\mu_{\mathcal{E}}^{AB}(\theta, \omega)$	
	$s(A, B)$	$[N; \Pi]$	M	$[\min; \max]$	moy \pm std	$[N; \Pi]$	M
(a)	0,99	[1,00; 1,00]	1,00	[1,00; 1,00]	1,00 \pm 0,00	[0,94; 1,00]	0,96
(b)	0,50	[0,70; 0,79]	0,76	[0,50; 0,63]	0,55 \pm 0,03	[0,36; 0,64]	0,48
(c)	0,74	[0,50; 0,54]	0,52	[0,40; 0,49]	0,45 \pm 0,03	[0,25; 0,53]	0,39
(d)	0,94	[0,93; 1,00]	0,97	[0,75; 1,00]	0,95 \pm 0,07	[0,54; 1,00]	0,79
(e)	0,16	[0,76; 0,90]	0,85	[0,50; 0,99]	0,70 \pm 0,12	[0,27; 1,00]	0,53
(f)	0,25	[0,94; 1,00]	0,99	[0,48; 1,00]	0,82 \pm 0,19	[0,77; 1,00]	0,85
(g)	0,51	[0,69; 0,85]	0,79	[0,35; 1,00]	0,62 \pm 0,21	[0,63; 1,00]	0,80

- \mathcal{F}_{AB} (mesure $s(A, B)$) correspond à l'approche de [Matsakis et Andréfouët, 2002];
- μ_{surround}^B (mesures $[N; \Pi]$ et M) correspond à l'approche de [Vanegas et al., 2011];
- \mathcal{S}_{AB} (mesures $[\min; \max]$ et moy \pm std) correspond à notre approche proposée pour évaluer la relation à partir du descripteur d'enlacement;
- $\mu_{\mathcal{E}}^{AB}(\theta, \omega)$ (mesures $[N; \Pi]$ et M) correspond à notre approche proposée pour évaluer la relation à partir des paysages flous d'enlacement.

En parallèle de ce tableau synthétique, la Figure 6.10 (pour les configurations de (a) à (f)) et la Figure 6.11 (pour la configuration (g)) présentent les profils directionnels d'entourement à partir desquels les mesures de notre approche fondée sur les paysages d'enlacement $\mu_{\mathcal{E}}^{AB}(\theta, \omega)$ ont été obtenues. Ces profils correspondent aux mesures de nécessité, possibilité et moyennes obtenues pour les différentes directions de l'intervalle $[0, \pi]$. Ils permettent de visualiser avec plus de granularité le comportement de l'approche en fonction des orientations. Nous commentons alors l'ensemble de ces résultats pour les différentes configurations spatiales considérées.

Les configurations de (a) à (d) peuvent être considérées comme similaires, en ce sens que la forme des objets est relativement uniforme. Dans la situation (a), l'objet est complètement entouré, et, dans le Tableau 6.3, les approches s'accordent toutes à donner des évaluations proches de 1. Dans la situation (b), l'objet est partiellement entouré, et les approches considérées expriment alors des points de vue différents, qui gravitent globalement autour de 0,5. En Figure 6.10, les profils directionnels correspondant permettent de visualiser que l'entourement est faible autour de l'axe horizontal. L'objet est également partiellement entouré pour la configuration (c), où la moitié de l'objet de référence a été supprimée. En ce sens, on retrouve des valeurs d'entourement nulles le long des axes diagonaux sur les profils directionnels de la

6.3. Applications à la relation spatiale « entouré par »

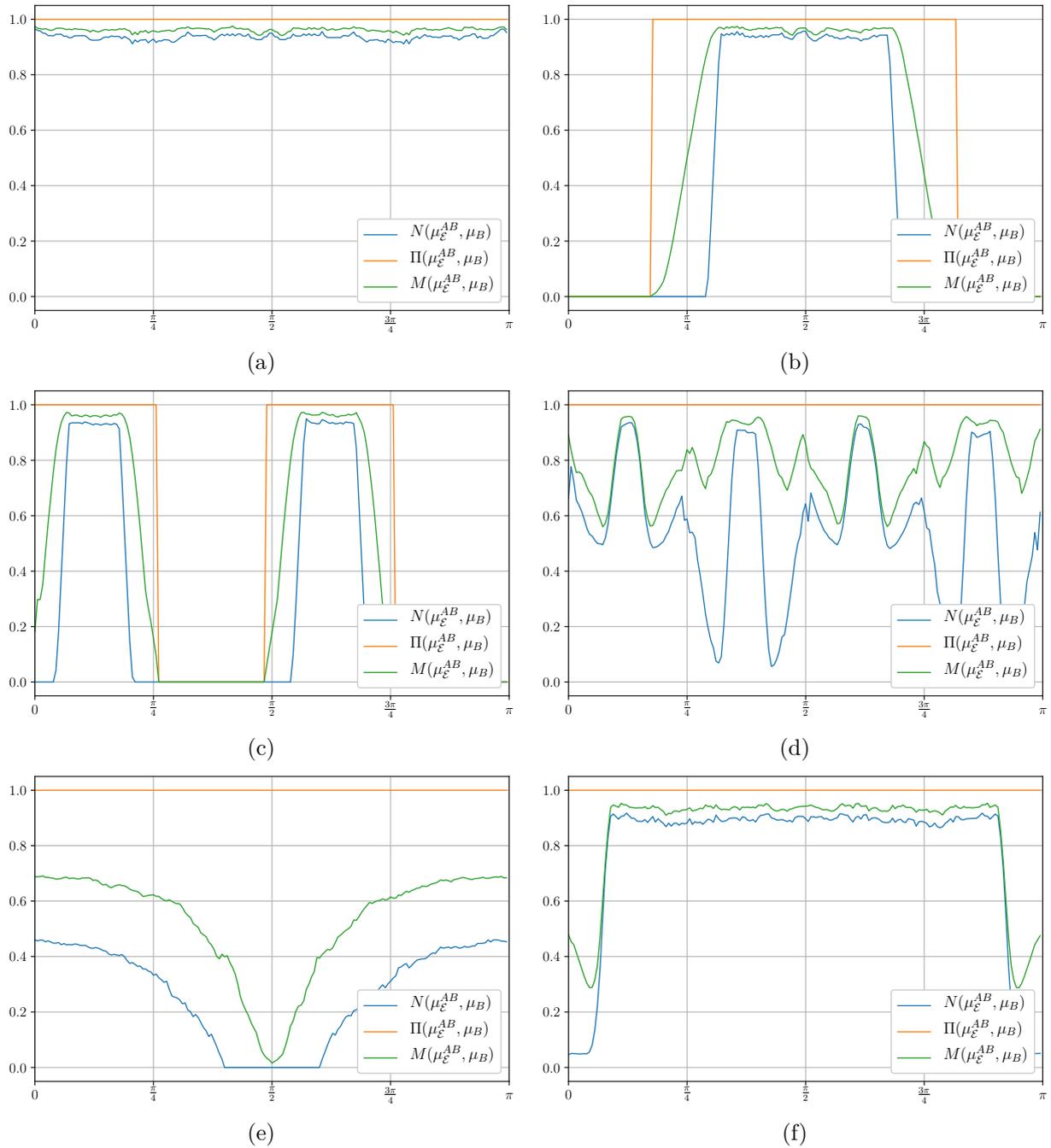


Figure 6.10 – Profils directionnels de nécessité-possibilité (en bleu et orange) et moyens (en vert) d’entouré pour les configurations de Figure 6.8.

Figure 6.10. Pour ces deux configurations (b) et (c), nous pouvons également remarquer que les intervalles de points de vue sont généralement assez restreints, avec des différences peu élevées entre le point de vue pessimiste et le point de vue optimiste. Ceci est notamment dû à l’uniform-

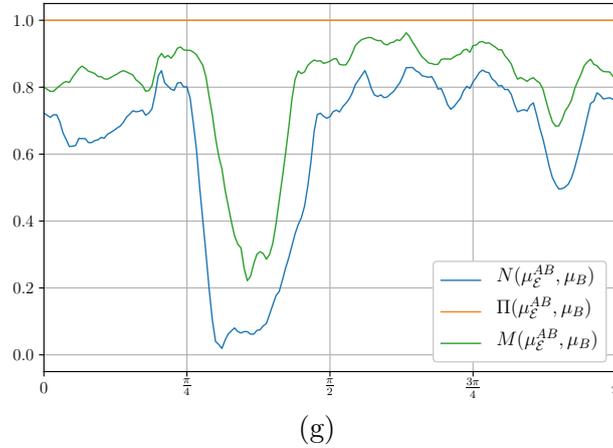


Figure 6.11 – Profils directionnels de nécessité-possibilité (en bleu et orange) et moyens (en vert) d'entouragement pour l'image du bassin d'Arcachon présentée en Figure 6.9.

mité de ces configurations, où l'épaisseur des objets ne varie pas en fonction des directions. La configuration (d) est alors plus complexe, car des portions de cercle ont été ajoutés, empêchant l'objet blanc de s'échapper sans traverser l'objet de référence. Cela se traduit par des points de vue plus variables pour nos approches (variance plus élevée pour \mathcal{S}_{AB} et intervalle plus large pour $\mu_{\mathcal{E}}^{AB}(\theta, \omega)$), tandis que les méthodes de référence s'accordent à dire que l'objet est entouré (les valeurs sont proches de 1 pour \mathcal{F}_{AB} et μ_{surround}^B). Le modèle d'enlacement est donc ici plus à même de décrire la complexité de cette configuration.

Les configurations (e) et (f) sont différentes, car la forme de l'objet de référence n'est plus uniforme. Dans la situation (e), les points de vue pessimistes de nos approches s'accordent à dire que l'objet est entouré par l'objet de référence. En effet, l'objet ne peut pas s'échapper de la situation sans traverser au moins de petites parties de l'autre objet. Cependant, il aurait à traverser une faible quantité de l'objet en se dirigeant vers le haut. De ce fait, les points de vue de nos approches sont beaucoup plus espacés. Les profils directionnels correspondants dans la Figure 6.10 traduisent bien cet aspect graduel, avec des valeurs faibles pour les orientations verticales. Enfin, pour la configuration (f), la plupart des approches indiquent que l'objet est complètement entouré. Pourtant, nos approches reposant sur le modèle d'enlacement permettent de prendre en compte le fait qu'une partie de l'objet de référence est moins épaisse, et ont donc là-aussi des points de vue plus espacés.

L'image de la Figure 6.9 représente alors un exemple réel sur une image satellitaire, désignée comme la configuration (g) dans le Tableau 6.3, ainsi qu'en Figure 6.11 pour les profils directionnels. Nous pouvons constater des valeurs d'entouragement faibles pour l'intervalle $[\frac{\pi}{4}, \frac{\pi}{2}]$, correspondant à l'axe diagonal le long duquel l'île est la moins entourée. Afin de mieux illustrer cette situation, nous présentons également en Figure 6.12 plusieurs paysages flous d'enlacement obtenus pour cette image satellitaire. En particulier, le paysage flou (c) orienté dans la direction $\frac{\pi}{3}$ correspond à la direction où l'entouragement est le plus faible.

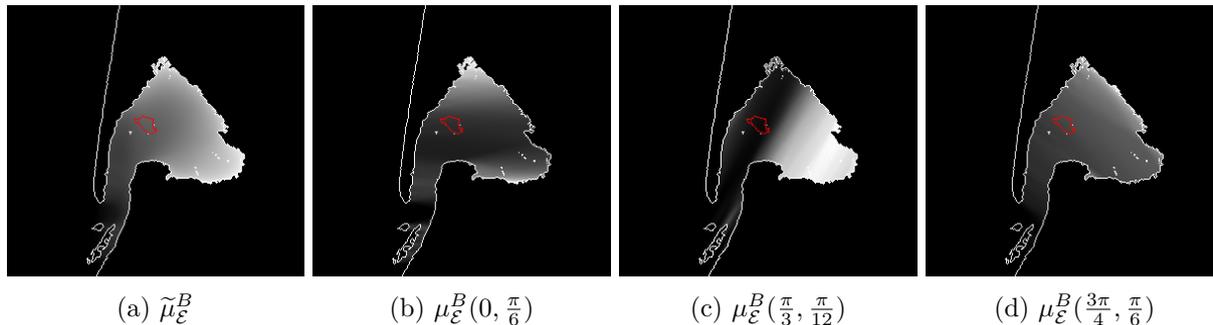


Figure 6.12 – Paysages flous d’enlacement pour l’image satellitaire de la Figure 6.9 représentant le bassin d’Arcachon.

6.4 Bilan scientifique

Dans ce chapitre, nous avons proposé un ensemble d’approches permettant de générer des évaluations quantitatives de configurations spatiales complexes à partir du modèle de l’enlacement. Tout d’abord, nous avons proposé une approche permettant d’évaluer la quantité d’enlacement ou d’entrelacement dans une direction à partir des descripteurs du Chapitre 4. Nous avons ensuite introduit les paysages flous directionnels d’enlacement, qui permettent de visualiser les degrés de validité de cette relation directement dans l’espace image. Enfin, dans le but de souligner l’intérêt de ces approches, celles-ci ont été appliquées à l’évaluation de la relation spatiale spécifique « *entouré par* », que nous envisageons comme un cas particulier de la relation d’enlacement. Ces expérimentations nous ont permis d’illustrer le potentiel descriptif ainsi que la généralité de nos approches.

Les travaux présentés dans ce chapitre ont été valorisés par différentes publications. La partie proposant des évaluations à partir des descripteurs d’enlacement a été présentée dans l’article de revue internationale [Clément *et al.*, 2017a]. Les travaux sur les paysages flous directionnels d’enlacement et d’entrelacement ont été présentés dans un article de conférence internationale [Clément *et al.*, 2017c].

Partie III

Apprentissage de relations spatiales composites

Résumé

Nous proposons un cadre général de description et d'apprentissage de relations spatiales composites entre objets d'intérêt représentés dans des images. Tout d'abord, nous présentons une stratégie de description du contenu d'une image à partir d'une représentation structurelle, permettant de caractériser les relations spatiales et les formes des objets d'intérêt. Puis, inspirés des approches par sacs de caractéristiques visuelles communément employées en vision par ordinateur, nous proposons une approche permettant d'apprendre des vocabulaires de configurations spatiales apparaissant à travers des images, potentiellement à différents niveaux d'échelle. Les résultats expérimentaux obtenus pour plusieurs bases d'images composées d'objets structurés permettent d'illustrer l'intérêt de cette approche pour la reconnaissance et la classification d'images. Nous présentons également comment intégrer les descripteurs d'enlacement au sein de cette nouvelle stratégie d'apprentissage, dans le contexte applicatif de l'analyse d'images de lettres décoratives.

Chapitre 7

Graphes relationnels attribués de relations spatiales

Sommaire

7.1	Décomposition en histogrammes de forces	98
7.1.1	Histogrammes de forces	98
7.1.2	Graphes relationnels attribués d’histogrammes de forces	100
7.1.3	Stratégies de comparaison et de mise en correspondance	103
7.1.4	Méthode de décomposition	106
7.2	Descripteurs FHD hiérarchiques	107
7.2.1	ARG hiérarchiques et contraction d’arêtes	108
7.2.2	Méthode de décomposition hiérarchique	109
7.3	Bilan scientifique	110

Dans ce chapitre, nous nous intéressons à l’exploitation des relations spatiales pour la description d’images qui ont été décomposées en différentes régions d’intérêt, dans le but de mettre en place des systèmes de reconnaissance et de classification. Premièrement, nous présentons la décomposition en histogrammes de forces (Section 7.1), un descripteur structural de relations spatiales entre couples de régions dans une image. Les travaux liés à ce descripteur ont été initiés dans la thèse de Mickaël Garnier [Garnier, 2014] et poursuivis dans le cadre d’un stage de Master qui a précédé la présente thèse. Nous proposons maintenant une définition originale de ce descripteur reposant sur un formalisme de graphes relationnels attribués. Nous présentons notamment différentes stratégies de comparaison et de mise en correspondance de ce descripteur. Ensuite, nous proposons une extension hiérarchique de ce descripteur (Section 7.2), afin de rendre possible la caractérisation de relations spatiales entre régions d’intérêt apparaissant à différentes échelles dans une image. Nous illustrons notamment cette extension à l’aide d’une approche de segmentation hiérarchique ascendante. Enfin, nous dressons un bilan scientifique de ces travaux (Section 7.3).

7.1 Décomposition en histogrammes de forces

Dans cette section, nous détaillons dans un premier temps le modèle des histogrammes de forces, déjà introduit de manière plus synthétique dans la [Partie I](#) de ce manuscrit. Nous présentons ensuite le descripteur FHD (pour *Force Histograms Decomposition*) fondé sur un graphe relationnel attribué d’histogrammes de forces, ainsi que différentes stratégies de comparaison pour ce descripteur. Nous proposons alors une approche de segmentation qui peut être associée à la construction de ce descripteur.

7.1.1 Histogrammes de forces

Un histogramme de forces [[Matsakis et Wendling, 1999](#)] (noté également *F-Histogramme* dans la suite) entre deux objets est un histogramme circulaire le long des directions dans l’intervalle $[0, 2\pi[$. Pour deux objets binaires A et B , chaque valeur de l’histogramme correspond à un angle θ , et mesure le degré de validité (ou le poids) que l’on pourrait attribuer à la proposition « A est dans la direction θ de B ». L’ensemble de l’histogramme résume de manière globale le positionnement spatial de A par rapport à B , selon un ensemble de directions.

Force d’attraction entre points

Le calcul d’un histogramme de forces est fondé sur la définition d’une force d’attraction entre couples de points, calculée par une fonction φ . Soit deux points du plan situés à une distance d l’un de l’autre, leur force d’attraction est alors donnée par :

$$\forall d \in \mathbb{R}_+^*, \varphi_r(d) = \frac{1}{d^r} \quad (7.1)$$

où r caractérise la force d’attraction utilisée. Si $r = 0$, il s’agit d’une force d’attraction constante où tous les points des objets sont traités avec la même importance. Lorsque $r \in [1, +\infty[$, on donne alors plus de poids aux points proches les uns des autres. Inversement, lorsque $r \in]-\infty, -1]$, il s’agit d’une force répulsive, et l’on donne alors plus d’importance aux points qui sont éloignés les uns des autres. Une valeur particulière est $r = 2$, qui correspond à la force d’attraction gravitationnelle.

Extension aux segments de droites

Pour calculer la force d’attraction globale d’un objet A par rapport à un objet B dans une direction θ donnée, il faut alors considérer la force d’attraction entre l’ensemble des couples de points composant ces deux objets, dans cette direction. Plutôt que de directement considérer ces couples de points, ce qui serait particulièrement coûteux en termes de temps de calcul, le modèle repose sur le traitement de couples de segments. Soit deux segments I et J portés par

7.1. Décomposition en histogrammes de forces

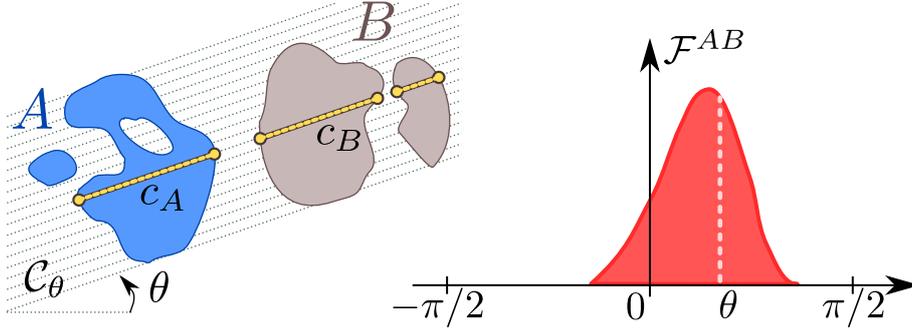


Figure 7.1 – Illustration du calcul d'un histogramme de forces. La force d'attraction globale entre deux objets A et B dans une direction θ , correspondant à la valeur $\mathcal{F}_{AB}(\theta)$ de l'histogramme de forces, est calculée en considérant l'ensemble \mathcal{C}_θ composé de toutes les coupes longitudinales \mathcal{C}_A et \mathcal{C}_B qui balayent les deux objets dans cette direction.

une droite orientée d'angle θ , la force d'attraction entre ces deux segments est alors définie par :

$$f(|I|, D_{IJ}^\theta, |J|) = \int_{D_{IJ}^\theta}^{|I|+D_{IJ}^\theta+|J|} \int_0^{|J|} \varphi_r(u-v) dv du \quad (7.2)$$

où $|I|$ et $|J|$ désignent les longueurs des segments, et D_{IJ}^θ la distance entre ces segments.

Par ailleurs, l'intersection entre un objet A et une droite orientée δ_θ d'angle θ correspond à un ensemble de segments, noté $\mathcal{C}_A = \delta_\theta \cap A$. Cet ensemble peut être vide (la droite ne traverse pas l'objet), ou contenir un ou plusieurs segments (l'objet peut être concave, constitué de plusieurs composantes connexes ou posséder des trous). L'ensemble \mathcal{C}_A constitue donc une *coupe longitudinale* d'angle θ de l'objet A . Ainsi, la fonction f de l'Équation 7.2 est étendue à une fonction F qui considère alors la force d'attraction entre deux coupes longitudinales :

$$F(\theta, \mathcal{C}_A, \mathcal{C}_B) = \sum_{I \in \mathcal{C}_A} \sum_{J \in \mathcal{C}_B} f(|I|, D_{IJ}^\theta, |J|). \quad (7.3)$$

Calcul global sur deux objets

Par la suite, l'ensemble \mathcal{C}_θ des droites parallèles orientées δ_θ est considéré. Cet ensemble de droites traverse entièrement les deux objets sous la forme de coupes longitudinales d'angle θ . La force d'attraction globale de l'objet A par rapport à l'objet B dans la direction θ est alors donnée par :

$$\mathcal{F}_{AB}(\theta) = \sum_{\delta_\theta \in \mathcal{C}_\theta} F(\theta, \mathcal{C}_A, \mathcal{C}_B) \quad (7.4)$$

où $\mathcal{C}_A = \delta_\theta \cap A$ et $\mathcal{C}_B = \delta_\theta \cap B$ sont les coupes longitudinales engendrées par chaque droite orientée. Enfin, le calcul de $\mathcal{F}_{AB}(\theta)$ pour un ensemble de directions discrètes $\theta \in [0, 2\pi[$ représente l'histogramme de forces et constitue une description directionnelle de la position relative de A

par rapport à B . La Figure 7.1 illustre le calcul d'un histogramme de forces d'un objet A par rapport à un objet B .

Un histogramme de forces entre deux objets peut être interprété comme un descripteur, ou vecteur de caractéristiques, au même titre que d'autres types de descripteurs utilisés classiquement en reconnaissance de formes. En ce sens, dans la suite de ce chapitre, ainsi que dans les chapitres suivants, nous utiliserons une notation vectorielle pour représenter les histogrammes de forces : le F-Histogramme de A par rapport à B calculé pour K directions discrètes pourra également être désigné par le vecteur \mathbf{f}_{AB} de dimension K .

Par ailleurs, ce modèle algorithmique permet également le calcul de l'histogramme de forces entre un objet et lui-même. Cet histogramme correspond alors à une description directionnelle de la forme de l'objet. Pour un objet A , cet histogramme est noté \mathbf{f}_{AA} et est appelé *F-signature* de l'objet [Wendling *et al.*, 2002].

Propriétés et complexité

Le modèle des histogrammes de forces présente plusieurs propriétés d'invariance intéressantes et souvent attendues pour diverses tâches de reconnaissance [Matsakis, 1998] :

- *Translation* : les objets sont considérés indépendamment de leur localisation dans l'image ;
- *Homothétie* : l'invariance est obtenue (aux erreurs de discrétisation près) en normalisant l'histogramme par sa valeur maximale, ou bien par l'aire des objets considérés ;
- *Isotropie* : toutes les directions sont considérées de manière équivalente ;
- *Rotation* : il est possible d'estimer la rotation en effectuant des décalages circulaires ;
- *Symétrie* : on peut considérer que $\mathbf{f}_{AB}(\theta) = \mathbf{f}_{BA}(\theta + \pi)$, cette propriété est aussi désignée sous le nom d'*inverse sémantique*. En outre, la F-signature est elle π -périodique, on a $\mathbf{f}_{AA}(\theta) = \mathbf{f}_{AA}(\theta + \pi)$.

De plus, la méthode de calcul des histogrammes de forces permet aussi de traiter des objets représentés dans des images vectorielles. Un algorithme permettant d'étendre le calcul aux objets 3D a également été proposé par [Ni *et al.*, 2004].

Par ailleurs, les histogrammes de forces peuvent être calculés en $\mathcal{O}(K n \sqrt{n})$ où K désigne le nombre de directions discrètes considérées, et n représente le nombre de pixels de l'image. Il a également été montré que les histogrammes de forces peuvent être calculés avec une complexité en $\mathcal{O}(K n \log n)$ [Ni et Matsakis, 2010] lorsque les objets ne se chevauchent pas et qu'une force constante est utilisée.

7.1.2 Graphes relationnels attribués d'histogrammes de forces

Notre objectif ici est de décrire le contenu d'une image, ou bien, de manière équivalente un objet d'intérêt représenté dans une image (selon l'application considérée), en exploitant les relations spatiales entre ses différentes sous-parties structurelles. Nous faisons ici l'hypothèse que l'image (ou l'objet d'intérêt étudié) a été partitionnée en N régions ou sous-parties structurelles initiales (avec $N \geq 1$). Ces sous-parties peuvent provenir par exemple du résultat d'un algorithme de segmentation, ou bien peuvent avoir été obtenues par des annotations préalables

7.1. Décomposition en histogrammes de forces

de différentes régions d'intérêt. Ainsi, étant donnée cette hypothèse, une sous-partie structurelle peut être considérée comme un objet binaire représentant un ensemble de pixels de l'image. Notons qu'un tel objet binaire peut potentiellement présenter des trous, ou être constitué de plusieurs composantes connexes.

Graphes relationnels attribués

Un graphe est défini par un couple (V, E) où V représente l'ensemble des nœuds du graphe et E correspond à l'ensemble des arêtes entre les nœuds. Un graphe relationnel attribué (noté *ARG* pour *Attributed Relational Graph*) est un graphe où différents types d'attributs peuvent être associés aux nœuds et aux arêtes. Un attribut peut être n'importe quel type de données, comme par exemple une étiquette textuelle ou un vecteur de \mathbb{R}^n . Dans notre contexte de description du contenu d'une image, les nœuds d'un ARG correspondent aux N sous-parties structurelles initiales de l'image, et tous les couples de nœuds de l'image sont reliés entre eux par des arêtes. Un tel ARG est donc un graphe complet non orienté, noté dans la suite \mathcal{F}_N , qui est composé de N nœuds et de $N(N - 1)/2$ arêtes.

À des fins de visualisation, la [Figure 7.2](#) présente quelques exemples de graphes relationnels attribués complets associés à différentes images où les sous-parties structurelles ont été préalablement détournées (les attributs ne sont pas représentés ici, mais chaque nœud et chaque arête peuvent être valués). Ces images sont issues d'un jeu de données représentant des scènes d'extérieur, qui sera présenté plus en détail dans le [Chapitre 9](#).

Nous pouvons noter ici que d'autres structures de graphes auraient pu être considérées, comme par exemple le graphe d'adjacence des régions (*RAG* pour *Region Adjacency Graph*, qu'il ne pas confondre avec l'acronyme ARG), qui utilise le même ensemble de nœuds mais ne construit une arête entre deux nœuds que si les régions correspondantes sont adjacentes spatialement (en fonction de la connexité choisie). Dans notre contexte, nous souhaitons caractériser les relations spatiales entre tous les couples de sous-parties de l'image, indépendamment de leur adjacence ou de leur connexité. En effet, nous partons ici de l'hypothèse que toute l'information contenue dans l'image peut être utile pour la caractériser. C'est la raison pour laquelle nous avons choisi une structure de graphe relationnel complet, où toutes les régions sont reliées entre elles. Notons cependant qu'il est possible d'ajouter des critères de pondération (ou de suppression d'arêtes), par exemple en fonction de la distance entre les sous-parties, dans le cas où l'on souhaiterait effectuer des descriptions plus locales de l'image.

Histogrammes de forces comme attributs de l'ARG

À partir de cette structure d'ARG complet, nous proposons d'utiliser les histogrammes de forces comme vecteurs de caractéristiques qui seront attachés comme attributs des nœuds et des arêtes du graphe. Pour deux nœuds $u, v \in V$ d'un ARG, correspondant à deux sous-parties structurelles de l'image, l'histogramme de forces \mathbf{f}_{uv} , résumant la position de u par rapport à v , est assigné comme attribut à l'arête $(u, v) \in E$ du graphe. De la même façon, les histogrammes de forces \mathbf{f}_{uu} et \mathbf{f}_{vv} décrivant la forme respective des sous-parties u et v sont assignés comme

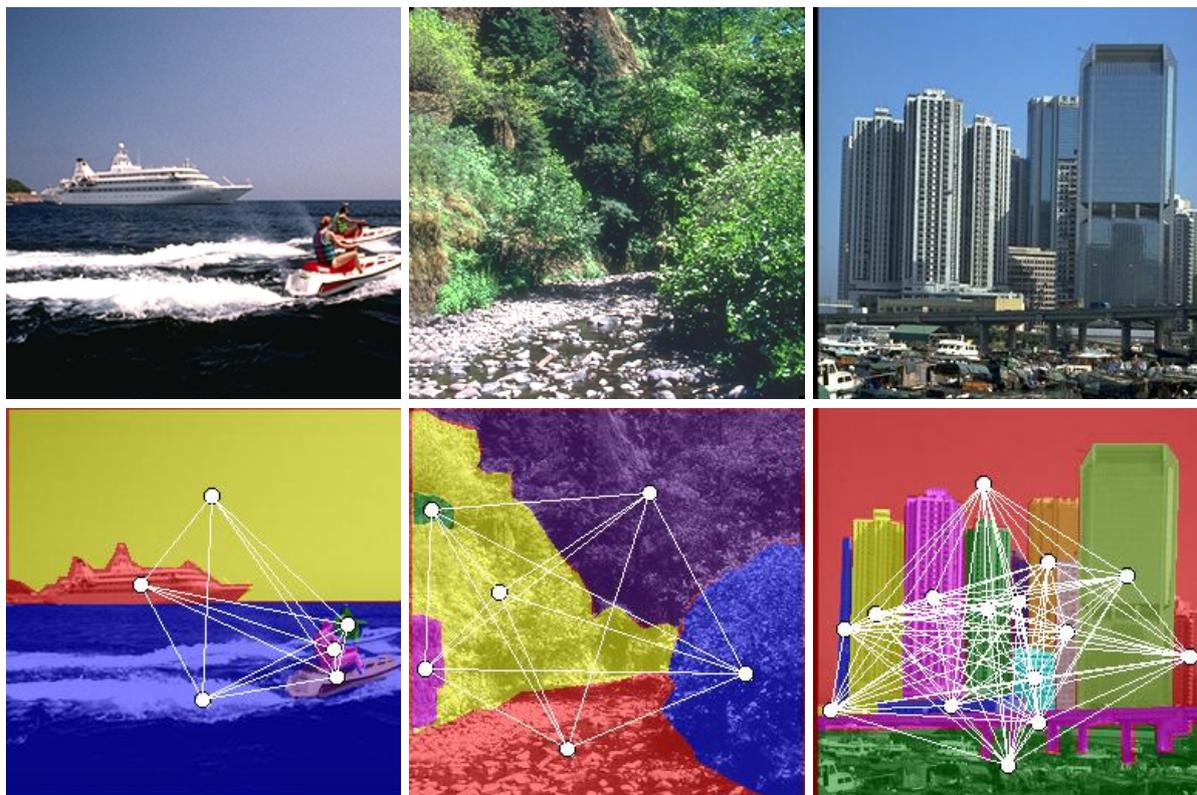


Figure 7.2 – Exemples illustratifs de graphes relationnels complets, pour différentes images qui ont préalablement été décomposées en sous parties structurales (représentées ici par des zones colorées).

attributs des nœuds u et v . Ainsi, le descripteur composé de N sous-parties peut être représenté par $\mathcal{F}_N = (V, E)$, où $V = \{(v_i, \mathbf{f}_{v_i v_i}), i \in 1, \dots, N\}$ correspond à l'ensemble des nœuds attribués, et $E = \{((v_i, v_j), \mathbf{f}_{v_i v_j}), i, j \in 1, \dots, N, j > i\}$ correspond à l'ensemble des arêtes attribuées.

Cet ARG représente donc une description structurale du contenu de l'image, caractérisant à la fois les formes de ses sous-parties, et leurs relations spatiales mutuelles. Cette représentation est uniquement composée d'histogrammes de forces associés aux nœuds et aux arêtes du graphe, ce qui résulte en un descripteur d'image particulièrement homogène, que l'on nomme Décomposition en Histogrammes de Forces (noté *FHD* pour *Force Histograms Decomposition*).

Paramètres et propriétés

Le descripteur FHD implique la définition des paramètres suivants :

- Le nombre de directions discrètes K pour le calcul des F-Histogrammes, et donc par extension la dimensionnalité des attributs de l'ARG ;
- Le type de forces considéré (c'est-à-dire la valeur de r dans la fonction φ_r de l'Équation 7.1), qui détermine le comportement des F-Histogrammes par rapport aux distances qui sé-

7.1. Décomposition en histogrammes de forces

parent les sous-parties structurelles. Notons ici que pour les F-Histogrammes de formes, seule la force constante $r = 0$ est possible. Il en est de même lorsque les objets se chevauchent ;

- Le nombre de sous-parties N , qui représente la cardinalité de l'ARG. Ce nombre est usuellement fixé *a priori*, par exemple par la méthode de segmentation sous-jacente.

Par ailleurs, le descripteur FHD regroupe un certain nombre de propriétés globales par rapport à l'image, héritées directement de celles des F-Histogrammes :

- Symétrie : du fait de la notion d'inverse sémantique liée aux relations spatiales directionnelles, les F-Histogrammes sont 2π -périodiques, et un graphe non-orienté suffit à décrire la situation ;
- Invariance par rapport aux translations : une translation des sous-parties structurelles de l'image ne modifie pas le descripteur ;
- Invariance par rapport aux homothéties : aux erreurs de discrétisation près, un changement d'échelle de l'image ne modifie pas le descripteur (à condition que l'ensemble des F-Histogrammes soient normalisés) ;
- Quasi-invariance par rapport aux rotations : une rotation de l'image implique seulement un décalage circulaire global de tous les F-Histogrammes de l'ARG.

7.1.3 Stratégies de comparaison et de mise en correspondance

Dans le but d'utiliser les descripteurs FHD dans un processus de reconnaissance ou de classification d'images, il est important de définir une stratégie de comparaison de ces descripteurs. Nous proposons donc ici de construire une mesure de dissimilarité adaptée pour de tels descripteurs. Par ailleurs, nous étudions également différentes stratégies de mise en correspondance dans le but d'améliorer la comparaison des descripteurs FHD.

Comparaison de descripteurs FHD

Étant donné deux descripteurs FHD de même taille N (correspondant au nombre de sous-parties structurelles de chaque objet), une manière de les comparer consiste à calculer la distance entre chacun de leurs histogrammes de forces pris deux-à-deux. Il existe de nombreuses mesures de distance entre histogrammes [Cha et Srihari, 2002 ; Rabin *et al.*, 2008]. En se basant sur les résultats initialement présentés par [Garnier *et al.*, 2012], nous avons choisi d'utiliser ici la distance du χ^2 pour comparer les histogrammes de forces deux-à-deux. La distance entre deux histogrammes de forces \mathbf{f}_1 et \mathbf{f}_2 le long de K directions discrètes est alors donnée par :

$$d_{\chi^2}(\mathbf{f}_1, \mathbf{f}_2) = \sum_{\theta=0}^K \frac{((\mathbf{f}_1)_\theta - (\mathbf{f}_2)_\theta)^2}{(\mathbf{f}_1)_\theta + (\mathbf{f}_2)_\theta}. \quad (7.5)$$

Par ailleurs, comme nous avons pu le voir précédemment, le descripteur FHD comprend à la fois des descripteurs de formes (F-Histogrammes associés aux nœuds de l'ARG) et des descripteurs de relations spatiales (F-Histogrammes associés aux arêtes de l'ARG). Nous pro-

posons alors de considérer séparément la distance entre ces deux types d'information. Soit deux descripteurs FHD de taille N , désignés par \mathcal{F}_N^A et \mathcal{F}_N^B , et composés des ensembles de nœuds $\{u_1, \dots, u_N\}$ et $\{v_1, \dots, v_N\}$ respectivement. La distance globale entre \mathcal{F}_1 et \mathcal{F}_2 est définie par :

$$\mathcal{D}_\alpha(\mathcal{F}_N^A, \mathcal{F}_N^B) = \alpha \mathcal{D}_{shape}(\mathcal{F}_N^A, \mathcal{F}_N^B) + (1 - \alpha) \mathcal{D}_{spatial}(\mathcal{F}_N^A, \mathcal{F}_N^B), \quad (7.6)$$

où α est un nombre réel entre 0 et 1 permettant d'affecter un poids variable à la distance \mathcal{D}_{shape} entre les descripteurs de formes et à la distance $\mathcal{D}_{spatial}$ entre les descripteurs de relations spatiales. Ces distances sont elles-mêmes définies de la manière suivante :

$$\mathcal{D}_{shape}(\mathcal{F}_N^A, \mathcal{F}_N^B) = \frac{1}{N} \sum_{i=1}^N d_{\chi^2}(\mathbf{f}_{u_i u_i}, \mathbf{f}_{v_i v_i}), \quad (7.7)$$

$$\mathcal{D}_{spatial}(\mathcal{F}_N^A, \mathcal{F}_N^B) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N d_{\chi^2}(\mathbf{f}_{u_i u_j}, \mathbf{f}_{v_i v_j}). \quad (7.8)$$

Une telle pondération des distances $\mathcal{D}_{spatial}$ et \mathcal{D}_{shape} permet également d'éviter que les descripteurs de relations spatiales ne deviennent prédominants par rapport aux descripteurs de formes lorsque $N > 3$ (car il y a alors plus d'histogrammes de forces pour les relations spatiales que pour les formes).

Mise en correspondance de descripteurs FHD

La mesure de distance présentée ci-avant ne peut être discriminante que s'il existe une correspondance entre les sous-parties structurelles composant les deux images comparées. En effet, il est possible que les régions issues de l'étape de décomposition pour deux objets ne soient pas organisées de la même façon au sein du graphe, même si ceux-ci sont très similaires. Ce cas est notamment susceptible d'apparaître dans le cas d'images en couleurs : il n'y a alors pas d'ordre naturel entre les sous-parties des objets, celles-ci étant associées à des valeurs d'intensité dans l'espace RVB (rouge, vert, bleu). De ce fait, la distance entre les deux descripteurs risque d'être anormalement élevée, car calculée entre des histogrammes de forces qui ne correspondent pas deux-à-deux. L'erreur se répercute à la fois sur les histogrammes de forces décrivant la forme et sur ceux décrivant les relations spatiales, la rendant d'autant plus importante.

Par conséquent, pour comparer des descripteurs FHD de manière efficace, il est nécessaire d'avoir recours à une stratégie de mise en correspondance des sous-parties structurelles des objets, en amont du calcul de distance. Il s'agit donc ici d'un problème de recherche d'isomorphisme de graphes, qui est de nature combinatoire : pour deux objets décomposés en N régions, il existe $N!$ possibilités d'appariement de leurs régions. La [Figure 7.3](#) illustre le principe d'un isomorphisme entre deux graphes. Pour N relativement petit, il est possible d'énumérer toutes les combinaisons possibles d'appariement des régions, mais cette stratégie n'est pas envisageable d'un point de vue calculatoire dans le cas général, le problème étant NP-difficile.

Par ailleurs, il est à noter que nous restons fortement dépendants de la qualité du résultat de segmentation préliminaire, où certaines sous-parties obtenues peuvent parfois être assimilées

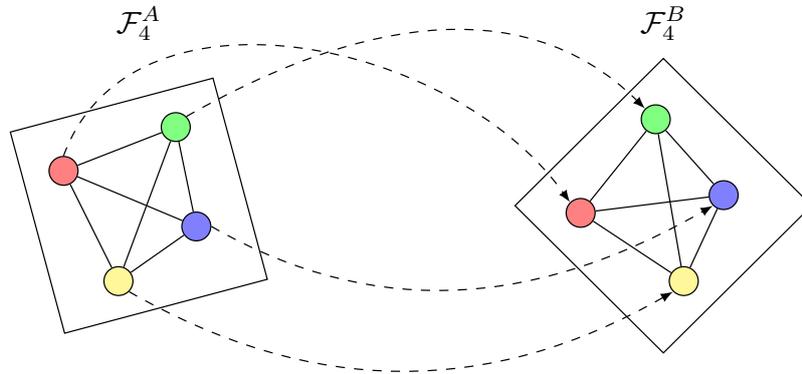


Figure 7.3 – Exemple d’isomorphisme entre deux graphes relationnels complets \mathcal{F}_4^A et \mathcal{F}_4^B composés de 4 nœuds.

à du bruit ou autres erreurs. En ce sens, il n’existe pas toujours une solution optimale à ce problème d’appariement, d’autant plus que les nœuds des graphes à mettre en correspondance sont multivalués (on peut alors parler d’appariement inexact de graphes).

Afin de procéder à cet appariement entre deux descripteurs FHD, nous proposons ici d’employer des heuristiques se basant sur des hypothèses concernant les sous-parties des objets. Ceci permet d’approximer une solution tout en réduisant la complexité du problème. Nous proposons ainsi deux stratégies d’appariement, dont le principe est expliqué ci-après.

Alignement par luminance

La première stratégie repose sur l’hypothèse que les régions correspondantes entre deux objets seront similaires en termes de colorimétrie. Afin de disposer d’un ordre total entre les couleurs associées aux sous-parties des objets, nous calculons leurs valeurs de luminance respectives dans l’espace $L^*u^*v^*$ (luminance et chrominance). Pour chaque objet, les régions sont alors triées par valeur de luminance, de manière décroissante. Les descripteurs FHD sont calculés à partir de ces régions triées, et les histogrammes de forces peuvent être comparés deux-à-deux.

Appariement selon la forme

La seconde stratégie part de l’hypothèse que les régions correspondantes entre deux objets sont similaires du point de vue de leurs formes. Ainsi, chaque sous-partie est appariée en fonction des histogrammes de forces décrivant la forme (c’est-à-dire ceux associés aux nœuds des ARG). Pour mettre en place cette stratégie, il est possible de l’assimiler à un problème d’optimisation dont l’objectif est de minimiser la distance globale $\mathcal{D}_{shape}(Q, T)$ entre les descripteurs de formes respectifs des deux objets. Nous proposons alors deux approches. La première consiste à énumérer les $N!$ possibilités d’appariement entre les régions des deux objets : il s’agit d’une solution dite *optimale*, qui est cependant inappropriée pour des descripteurs FHD com-

posés d'un nombre élevé de régions. Une approximation de cette solution consiste à mettre en correspondance les régions successivement en faisant à chaque fois le meilleur choix possible : il s'agit d'une stratégie dite *gloutonne* (*greedy* en anglais). D'un point de vue algorithmique, cette approximation peut être considérée comme un bon compromis nécessitant une complexité quadratique en fonction de N .

Discussion

Le choix d'une stratégie de mise en correspondance est fortement dépendant de l'application envisagée et chacune présente des avantages et des inconvénients. D'un côté, la stratégie d'alignement par luminance est relativement simple et peut être considérée comme une étape de pré-traitement, cependant, elle est vouée à échouer si la luminance n'est pas une caractéristique discriminante (par exemple, si deux objets ont la même structure mais des couleurs différentes). D'un autre côté, la stratégie d'appariement selon la forme paraît plus appropriée, mais implique d'être appliquée à chaque comparaison de deux descripteurs. De plus, cette stratégie est également plus disposée à commettre des erreurs, notamment lorsque certaines régions peuvent être assimilées à du bruit. Il pourrait alors être intéressant de combiner ces deux stratégies.

7.1.4 Méthode de décomposition

Afin de décrire la structure spatiale interne d'un objet complexe représenté dans une image, il est tout d'abord nécessaire de le décomposer en différentes sous-parties, ou régions d'intérêt. Ici, nous proposons une méthode reposant sur l'utilisation d'un algorithme de segmentation d'images. Pour autant, nous pouvons souligner que la méthodologie de description d'images présentée ci-avant a un caractère générique, et que d'autres stratégies de décomposition d'objets pourraient être envisagées.

La segmentation d'images constitue un vaste champ de recherche à part entière, et l'efficacité des méthodes existantes reste bien souvent fortement dépendante de l'application envisagée [Gonzalez et Woods, 2002]. On retrouve généralement deux grandes familles de méthodes : les approches dites *basées contours* et les approches dites *basées régions*. Les approches basées contours sont fondées sur la détection de fortes variations d'intensité entre les pixels dans les images (correspondant généralement aux contours des régions d'intérêt), tandis que les approches basées régions ont plutôt recours à des méthodes de regroupement de pixels connexes. Il est à noter que quelle que soit l'approche envisagée, les méthodes de segmentation sont très souvent particulièrement dépendantes de leur paramétrisation. Dans le contexte de ces travaux, nous nous sommes plutôt intéressés à une approche basée régions, que nous avons jugée plus adaptée pour manipuler des objets complexes représentés dans des images en couleurs, où les frontières ne sont pas toujours bien définies.

Dans le cadre de ces travaux, nous avons choisi d'appliquer l'algorithme *Mean Shift*, proposé par [Comaniciu et Meer, 2002], qui est adapté à la segmentation d'images en couleurs. *Mean Shift* est un algorithme issu de l'apprentissage automatique, qui considère l'espace des caractéristiques comme une densité de probabilité, et qui cherche à en estimer les modes. Nous

7.2. Descripteurs FHD hiérarchiques

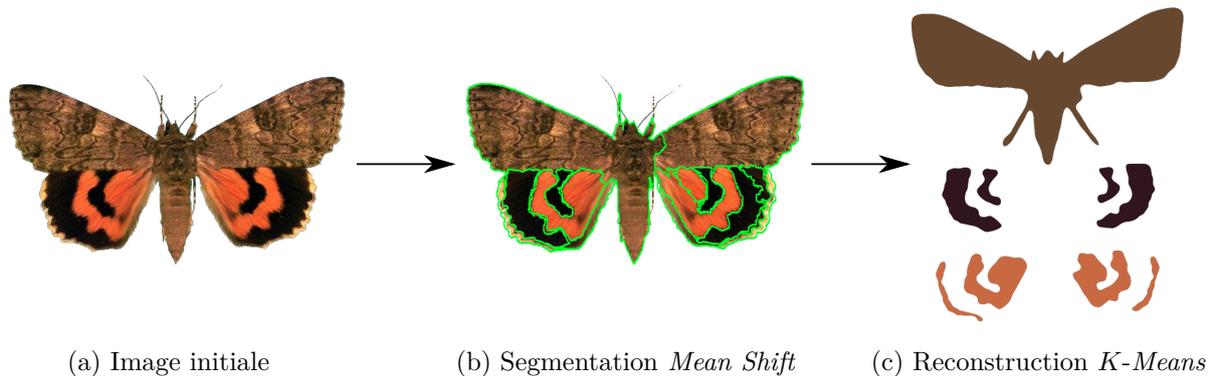


Figure 7.4 – Méthode de décomposition d'un objet en ses sous-parties structurales.

résumons ici son fonctionnement pour la segmentation d'images en couleurs. Une image est tout d'abord convertie dans l'espace $L^*u^*v^*$ (luminance et chrominance) qui permet de répartir les couleurs de manière plus uniforme par rapport à la perception humaine. Pour chaque pixel, l'algorithme construit un voisinage dont il calcule la valeur moyenne. Ce processus est itéré plusieurs fois jusqu'à convergence, c'est-à-dire jusqu'à ce que tous les pixels soient assignés aux valeurs moyennes obtenues. La procédure du *Mean Shift* produit une partition de l'image en un certain nombre de régions présentant des propriétés colorimétriques homogènes, et supposées délimiter les différentes régions d'intérêt composant l'image. Un résultat de segmentation d'un objet dans une image par l'algorithme *Mean Shift* est présenté en Figure 7.4 (b).

Cependant, par définition, la segmentation d'une image ne permet pas d'obtenir des régions constituées de plusieurs composantes connexes. Or, les sous-parties structurales d'un objet peuvent parfois être effectivement déconnectées (comme par exemple les tâches orange du papillon de la Figure 7.4 (b)). Pour pallier ce problème, nous appliquons alors l'algorithme de classification non-supervisée *K-Means* [MacQueen, 1967] sur les pixels de l'image auxquels nous avons affecté la valeur d'intensité colorimétrique moyenne obtenue à l'issue de la segmentation *Mean Shift*. Ceci permet de reconnecter de manière pertinente les différentes sous-parties de l'objet étudié, en fonction de leurs similarités colorimétriques. Par ailleurs, cela permet, par la même occasion, de fixer *a priori* le nombre de sous-parties de l'objet, afin de contrôler la taille des descripteurs FHD issus de cette décomposition. L'algorithme *K-Means* est initialisé avec N graines afin de produire un regroupement des régions de l'image en N sous-parties structurales. Par exemple, dans la Figure 7.4 (c), les régions formées par l'étape de segmentation ont été regroupées afin d'obtenir un total de trois sous-parties structurales de l'objet.

7.2 Descripteurs FHD hiérarchiques

Sur la base du descripteur FHD présenté dans la section précédente, nous présentons maintenant une stratégie hiérarchique qui permet une description multiniveau du contenu d'une image ou d'un objet. En effet, l'approche présentée ci-avant permet uniquement de construire

des descripteurs de tailles prédéfinies, avec un nombre fixe de sous-parties structurelles. Or, bien souvent, les régions d'intérêt d'une image ou d'un objet peuvent apparaître à différents niveaux d'échelle, et peuvent également être présentes en quantités variables d'une image à une autre. Le but de la représentation hiérarchique que nous proposons est donc de capturer les différents types de configurations spatiales qui apparaissent dans une image au sein de différents niveaux de décomposition.

7.2.1 ARG hiérarchiques et contraction d'arêtes

La stratégie multiniveau que nous présentons est fondée sur l'opération élémentaire qui consiste à contracter une arête d'un graphe, fusionnant ainsi les deux nœuds qui y sont reliés. Elle repose notamment sur la propriété de bilinéarité du modèle des histogrammes de forces. En effet, soit les objets A_1, A_2, B_1, B_2 , nous avons alors :

$$\mathbf{f}_{(A_1 \cup A_2)(B_1 \cup B_2)} = \mathbf{f}_{A_1 B_1} + \mathbf{f}_{A_1 B_2} + \mathbf{f}_{A_2 B_1} + \mathbf{f}_{A_2 B_2}. \quad (7.9)$$

Cette propriété indique donc que le F-Histogramme de l'union de différents objets peut être obtenu linéairement à partir des F-Histogrammes élémentaires associés à ces objets. Dans notre contexte, cette propriété va notamment permettre d'effectuer des contractions d'arêtes au sein du descripteur FHD, d'une manière efficace d'un point de vue computationnel, en mettant à jour les attributs de l'ARG conformément aux nœuds qui seront fusionnés. En ce sens, nous proposons de procéder à des contractions successives au sein du descripteur FHD, afin d'obtenir un descripteur hiérarchique multiniveau.

L'opération de contraction d'une arête de l'ARG est définie de la manière suivante. Soit $\mathcal{F}_N = (V, E)$ un ARG complet correspondant à un descripteur FHD composé de N sous-parties structurelles initiales (avec $N \geq 2$). Soit $(u, v) \in E$ une arête entre les nœuds $u, v \in V$ (avec $u \neq v$), que l'on souhaite contracter. Cette arête peut être choisie, par exemple, en utilisant en parallèle un algorithme de *clustering* hiérarchique [Ward, 1963] ou de segmentation hiérarchique [Salembier et Garrido, 2000 ; Soille, 2008]. La contraction de l'arête (u, v) conduit à un nouveau graphe complet \mathcal{F}_{N-1} composé de $N-1$ nœuds : les mêmes nœuds que \mathcal{F}_N , à l'exception de u et v qui sont fusionnés en un nouveau nœud w . Toutes les arêtes qui étaient précédemment adjacentes aux nœuds u et v dans le graphe \mathcal{F}_N sont alors reliées au nœud w dans \mathcal{F}_{N-1} .

D'une part, du point de vue de l'image, cette contraction d'arête correspond à l'union de deux objets binaires, formant ainsi un nouvel objet. D'autre part, du point de vue de l'ARG, les attributs correspondants aux nœuds fusionnés et aux arêtes adjacentes sont mis à jour conformément à la propriété de bilinéarité de l'Équation 7.9. Chaque contraction d'arête successive produit donc un nouveau descripteur FHD où deux sous-parties ont été fusionnées.

À partir d'une décomposition initiale d'une image en N sous-parties, les contractions successives produisent un ensemble d'ARG multiniveau $\mathcal{F} = \{\mathcal{F}_i\}_{i=1}^N$ que nous nommons alors *descripteur FHD hiérarchique*. Cet ensemble d'ARG, qui peut être visualisé schématiquement sur la Figure 7.5, permet effectivement de capturer de manière homogène les relations spatiales et les différentes formes qui apparaissent à différents niveaux d'échelle dans une image.

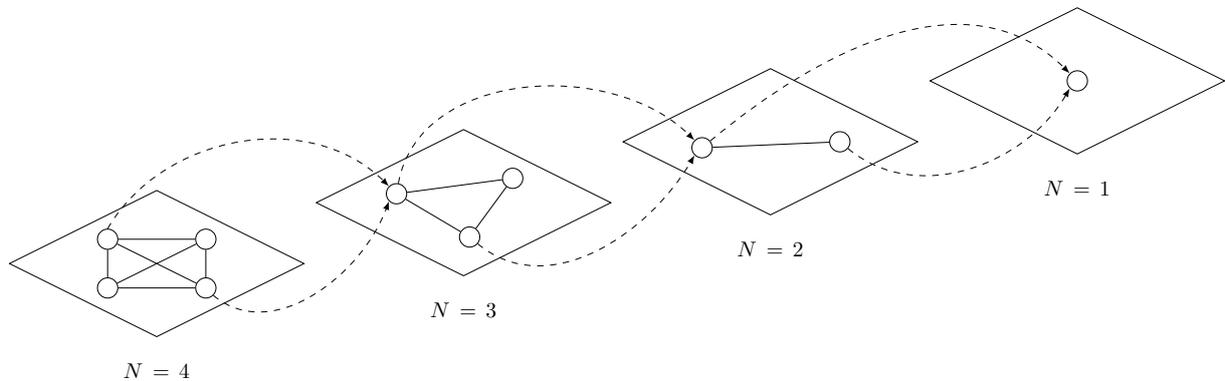


Figure 7.5 – Illustration schématique d'un descripteur FHD hiérarchique à quatre étages.

7.2.2 Méthode de décomposition hiérarchique

La méthodologie générique de description multiniveau présentée dans la section précédente repose en partie sur un processus de décomposition hiérarchique associé. Dans cette section, nous proposons une méthode de décomposition hiérarchique d'une image, qui peut être vue comme une extension de la méthode de décomposition présentée dans la [Sous-section 7.1.4](#).

La décomposition hiérarchique proposée démarre avec une sur-segmentation initiale de l'image à partir de l'algorithme *Mean Shift*. L'image est ainsi partitionnée en un ensemble de s segments S_i , délimitant différentes régions exposant des propriétés colorimétriques homogènes. Nous considérons alors ces segments S_i comme les régions initiales de l'image.

À partir de ces régions initiales, nous appliquons l'algorithme de *clustering* hiérarchique ascendant (*AHC* pour *Ascendant Hierarchical Clustering*) initialement proposé par [Ward, 1963], afin de reconstruire progressivement les sous-parties structurelles pertinentes de l'image. Cet algorithme de *clustering* hiérarchique a été appliqué sur les valeurs RVB moyennes des régions initiales de l'image. La hiérarchie de *clusters* est obtenue en utilisant la distance euclidienne entre les individus et un critère de fusion de type lien moyen (*average linkage*). Un des avantages d'utiliser un algorithme de *clustering*, par opposition à un algorithme de segmentation hiérarchique classique avec des contraintes spatiales [Salembier et Garrido, 2000; Arbelaez et al., 2011], est que cela permet de construire des sous-parties structurelles de l'image qui peuvent être constituées de plusieurs composantes connexes.

Ainsi, l'algorithme AHC permet d'établir une hiérarchie de sous-parties de l'image, par fusions binaires successives de régions similaires d'un point de vue colorimétrique, mais pas nécessairement adjacentes. Cette succession de fusions de régions peut être représentée sous la forme d'un arbre binaire, communément appelé le *dendrogramme* associé au *clustering* hiérarchique. Les s régions initiales de l'image correspondent aux feuilles du dendrogramme, qui est alors composé de s niveaux représentant l'ensemble des fusions successives, jusqu'à la racine qui contient une seule sous-partie structurelle représentant toute l'image (ou bien tout l'objet d'intérêt considéré, si le fond a été préalablement supprimé). Une telle décomposition hiérarchique d'un objet

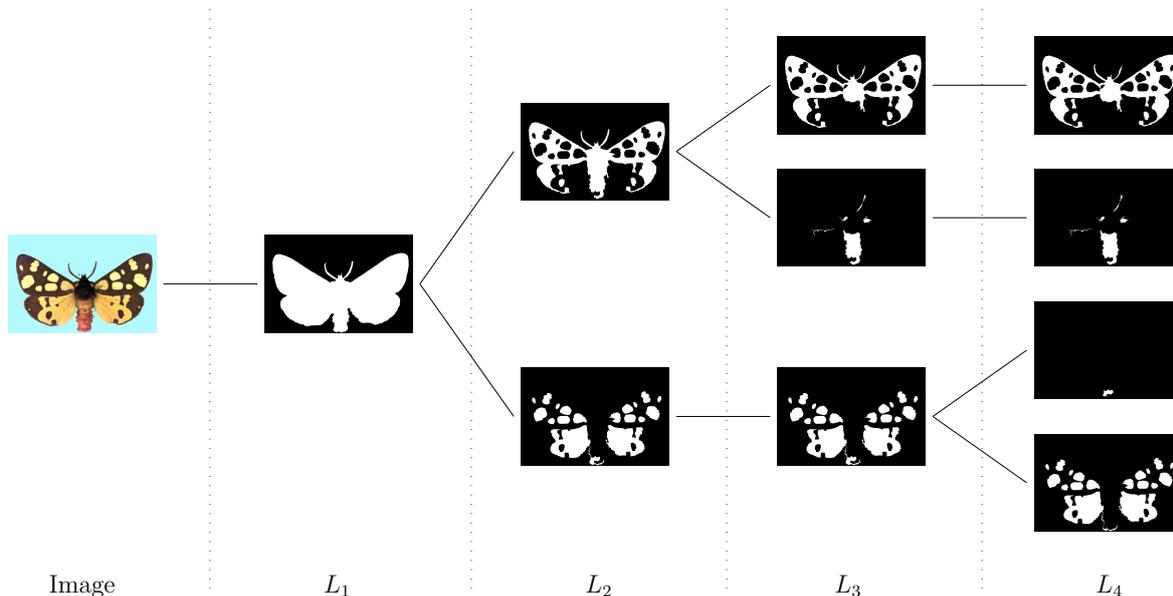


Figure 7.6 – Illustration d’une décomposition hiérarchique d’un objet structuré. Seuls les quatre premiers niveaux de la hiérarchie sont représentés ici.

représenté dans une image est illustrée en Figure 7.6, où seuls les quatre niveaux L_1, \dots, L_4 les plus hauts du dendrogramme sont affichés. Chaque niveau intermédiaire L_i (avec $i \in [1, s]$) de la hiérarchie correspond à une partition de l’image composée de $|L_i| = i$ sous-parties structurales.

Étant donnée cette décomposition hiérarchique d’une image, nous calculons le descripteur FHD hiérarchique à partir d’un niveau intermédiaire L_N du dendrogramme, avec $N \leq s$. Ce niveau intermédiaire peut être vu comme la profondeur maximale du descripteur FHD hiérarchique. Le descripteur est calculé pour les N sous parties structurales initiales, puis les niveaux supérieurs du descripteur sont obtenus de manière ascendante grâce à la propriété de bilinéarité des F-Histogrammes. Nous obtenons ainsi le descripteur FHD hiérarchique $\mathcal{F} = \{\mathcal{F}_i\}_{i=1}^N$ où chaque niveau \mathcal{F}_i correspond à un étage intermédiaire L_i du dendrogramme.

7.3 Bilan scientifique

Dans ce chapitre, nous avons présenté un ensemble de méthodologies génériques visant à décrire le contenu d’images en couleurs, à partir des relations spatiales entre leurs sous-parties structurales. Nous avons proposé une représentation par graphes relationnels attribués, en utilisant le modèle des histogrammes de forces comme attributs des graphes. Cette représentation est alors nommée descripteur FHD. Nous avons également présenté différentes stratégies de comparaison et d’appariement de ces descripteurs, ainsi qu’une méthode de décomposition d’images en ses sous-parties à partir d’une segmentation initiale. Des expérimentations préliminaires ont été menées pour illustrer le potentiel descriptif de cette approche. Néanmoins, celle-ci a été par

7.3. Bilan scientifique

la suite généralisée, comme nous le verrons dans le [Chapitre 8](#). Des validations expérimentales plus complètes de cette version étendue seront alors présentées dans le [Chapitre 9](#). Ainsi, dans un souci de cohérence, les premières expérimentations sur le descripteur FHD initial ne sont pas présentées dans ce manuscrit. De plus amples détails sur ces résultats préliminaires peuvent être consultés dans les articles mentionnés ci-après.

Nous avons ensuite proposé une extension de ce descripteur, afin de pouvoir caractériser les relations spatiales apparaissant à différents niveaux d'échelle dans les images. Cette extension repose notamment sur la propriété de bilinéarité des F-Histogrammes, qui permet alors un calcul optimisé au sein d'un processus de *clustering* hiérarchique ascendant. Dans le [Chapitre 8](#), nous présentons alors une stratégie d'apprentissage permettant d'exploiter l'information contenue dans ce type de descripteurs hiérarchiques, afin de pouvoir réaliser diverses tâches de classification ou de reconnaissance d'images.

Les travaux présentés dans ce chapitre ont été valorisés par plusieurs publications. La version initiale des descripteurs FHD pour des images en couleurs a été présentée dans un article de conférence internationale [[Clément *et al.*, 2015a](#)], lors d'un atelier d'une conférence nationale [[Clément *et al.*, 2015b](#)], ainsi que dans un numéro spécial d'une revue nationale [[Clément *et al.*, 2016b](#)]. Enfin, la version hiérarchique des descripteurs FHD a été valorisée dans un article de conférence internationale [[Clément *et al.*, 2016a](#)].

Chapitre 8

Apprentissage par sacs de relations spatiales

Sommaire

8.1 Sacs de caractéristiques : état de l’art synthétique	114
8.1.1 Vocabulaires de caractéristiques visuelles locales	114
8.1.2 Vers les sacs de relations spatiales	115
8.2 Désassemblage du descripteur FHD hiérarchique	116
8.3 Vocabulaires structurels de relations spatiales et de formes	117
8.3.1 Construction des vocabulaires	117
8.3.2 Stratégies d’encodage et d’agrégation	119
8.3.3 Compatibilité avec les caractéristiques visuelles locales	123
8.4 Bilan scientifique	123

Ce chapitre présente une méthode d’apprentissage de type sacs de caractéristiques, fondée sur l’exploitation des attributs des descripteurs FHD hiérarchiques. Tout d’abord, nous dressons un état de l’art synthétique des approches par sacs de caractéristiques communément employées en vision par ordinateur (Section 8.1). Ensuite, nous présentons une stratégie de désassemblage des descripteurs FHD hiérarchiques, afin qu’ils soient organisés comme des ensembles non-ordonnés d’attributs, et ainsi exploitables pour un apprentissage par sacs de caractéristiques (Section 8.2). Nous décrivons alors la construction des vocabulaires de relations spatiales et de formes, ainsi que les représentations structurelles d’images qui en découlent (Section 8.3). Nous illustrons notamment la compatibilité de ces représentations avec celles produites par les sacs de caractéristiques classiques, conduisant alors à des vocabulaires hybrides. Enfin, un bilan scientifique de ces travaux est établi pour conclure le chapitre (Section 8.4).

8.1 Sacs de caractéristiques : état de l'art synthétique

Dans cette section, nous présentons un état de l'art synthétique des approches par sacs de caractéristiques qui ont été proposées dans la littérature pour diverses tâches de reconnaissance et de classification d'images. Ces approches sont traditionnellement inspirées des stratégies de représentation par sacs de mots (*bags of words* en anglais) qui ont été employées avec succès dans les domaines de la recherche d'information et de la classification de textes.

8.1.1 Vocabulaires de caractéristiques visuelles locales

Ces dernières années, une attention particulière a été accordée aux stratégies d'apprentissage par sacs de caractéristiques dites visuelles [Zhang *et al.*, 2007]. En effet, ces approches ont conduit à des progrès significatifs dans le domaine de la vision par ordinateur, en témoignent certains résultats récents pour la détection et la reconnaissance d'objets représentés dans de grandes bases d'images [Everingham *et al.*, 2010].

Popularisées par les travaux de [Sivic et Zisserman, 2003], ces approches reposent généralement sur la construction de vocabulaires de primitives visuelles. Afin de générer ces vocabulaires, différents types de caractéristiques locales sont calculées sur les images. Parmi celles-ci, on pourra notamment citer les descripteurs SIFT [Lowe, 2004], HOG [Dalal et Triggs, 2005] ou SURF [Bay *et al.*, 2008] qui reposent sur la construction d'histogrammes de gradients au sein de fenêtres locales ; mais aussi les descripteurs LBP [Ojala *et al.*, 1996] qui caractérisent plutôt les motifs binaires ; ou encore les MSER [Matas *et al.*, 2004] qui détectent de petites régions d'intérêt stables dans les images. Ces caractéristiques sont généralement calculées à partir de la détection de points d'intérêt ou de régions saillantes [Mikolajczyk et Schmid, 2005], ou bien de manière dense, c'est-à-dire au sein de fenêtres locales recouvrant toute l'image [Felzenszwalb *et al.*, 2010].

À partir d'une base d'images d'apprentissage, une quantification vectorielle de ces caractéristiques visuelles est alors effectuée, généralement par le biais d'un algorithme de *clustering* comme *K-Means* [MacQueen, 1967] ou bien par un modèle de mélange gaussien (*GMM* pour *Gaussian Mixture Model*). Cette étape permet d'obtenir un ensemble de points moyens (ou prototypes), qui est alors interprété comme un vocabulaire de mots visuels associé au modèle. Chaque image est ensuite représentée par un histogramme de composition qui correspond à la distribution de ses caractéristiques locales par rapport au vocabulaire de mots visuels appris. Un tel histogramme est aussi appelé multiensemble, ou sac (d'où le terme *bags of features*). La classification des images peut alors être effectuée à l'aide de classifieurs supervisés, comme les machines à vecteurs de support (*SVM* pour *Support Vector Machines*) [Cortes et Vapnik, 1995] ou les forêts aléatoires (*Random Forests*) [Breiman, 2001].

Une des limitations majeures des stratégies classiques par sacs de caractéristiques est leur manque de prise en compte d'informations spatiales et structurelles. Ce manque est inhérent au fait que les images sont représentées par des ensembles non-ordonnés de caractéristiques locales. Afin de pallier cette limitation, plusieurs travaux ont été proposés pour intégrer une part d'information spatiale dans ces modèles. Par exemple, à l'échelle des descripteurs, les caractéristiques par corrélogrammes [Savarese *et al.*, 2006] permettent de prendre en compte les co-occurrences

8.1. Sacs de caractéristiques : état de l’art synthétique

spatiales au sein de la représentation. Les auteurs de [Hoàng *et al.*, 2010] proposent d’intégrer l’aspect spatial en considérant les relations triangulaires qui existent entre triplets de points d’intérêt (ou de régions saillantes). Dans les travaux de [Sudderth *et al.*, 2005], la position relative de points d’intérêt détectés est considérée à partir d’un modèle probabiliste génératif. Les auteurs de [Niebles et Fei-Fei, 2007] ont également proposé un modèle génératif hiérarchique qui incorpore des contraintes spatio-temporelles entre les caractéristiques pour la reconnaissance d’actions humaines à partir de vidéos. Du point de vue des modèles discriminants, les travaux de [Lazebnik *et al.*, 2006] proposent une mise en correspondance des caractéristiques à partir de pyramides spatiales, c’est-à-dire en subdivisant récursivement l’image en quadrants de plus en plus petits. Plus récemment, l’approche de [Penatti *et al.*, 2014] propose également de considérer l’information spatiale avec des quadrants, mais ce ceux-ci sont définis à partir de points d’intérêt localisés. Cette approche permet alors de caractériser la disposition spatiale des différents mots visuels du vocabulaire.

Plusieurs travaux de recherche ont donc proposé différentes stratégies pour intégrer efficacement des aspects spatiaux au sein d’approches par vocabulaires de mots visuels. Cependant, elles reposent toutes sur l’utilisation de caractéristiques locales en tant que primitives visuelles, et incorporent alors *a posteriori* des informations spatiales entre ces caractéristiques. Dans ce chapitre, nous proposons de considérer les relations spatiales elles-mêmes en tant que caractéristiques primitives des images, rendant alors possible la construction de vocabulaires de configurations spatiales à partir de données d’apprentissage. Par ailleurs, les relations spatiales sont envisagées ici d’un point de vue global, c’est-à-dire entre les différentes régions structurelles composant les images, et non pas à partir de points d’intérêt locaux.

8.1.2 Vers les sacs de relations spatiales

Dans le domaine applicatif de la reconnaissance de symboles techniques, des travaux récents [Santosh *et al.*, 2014a; Santosh *et al.*, 2014b] ont introduit le concept des *sacs de relations* (*BoR* pour *bags of relations*). Il s’agit d’une approche originale permettant de produire des vocabulaires composites de configurations spatiales, d’un point de vue structurel sur les images considérées (donc par opposition au point de vue local employé par les approches sacs de caractéristiques classiques). Cette approche a été appliquée avec un ensemble de primitives structurelles spécifiques au domaine d’application des symboles techniques, comme par exemple les cercles, les coins ou les extrémités.

La méthodologie d’apprentissage présentée dans ce chapitre s’inspire donc de cette famille d’approches. Nous proposons de tirer parti de ce concept de sacs de relations spatiales, avec pour objectif de l’étendre à des tâches plus génériques de reconnaissance d’objets complexes représentés dans des images en couleurs. Au lieu de reposer sur des primitives pré-définies, nous nous concentrons sur l’apprentissage de vocabulaires structurels de configurations spatiales (et de formes), qui pourront notamment être combinés avec des vocabulaires visuels issus de caractéristiques locales.

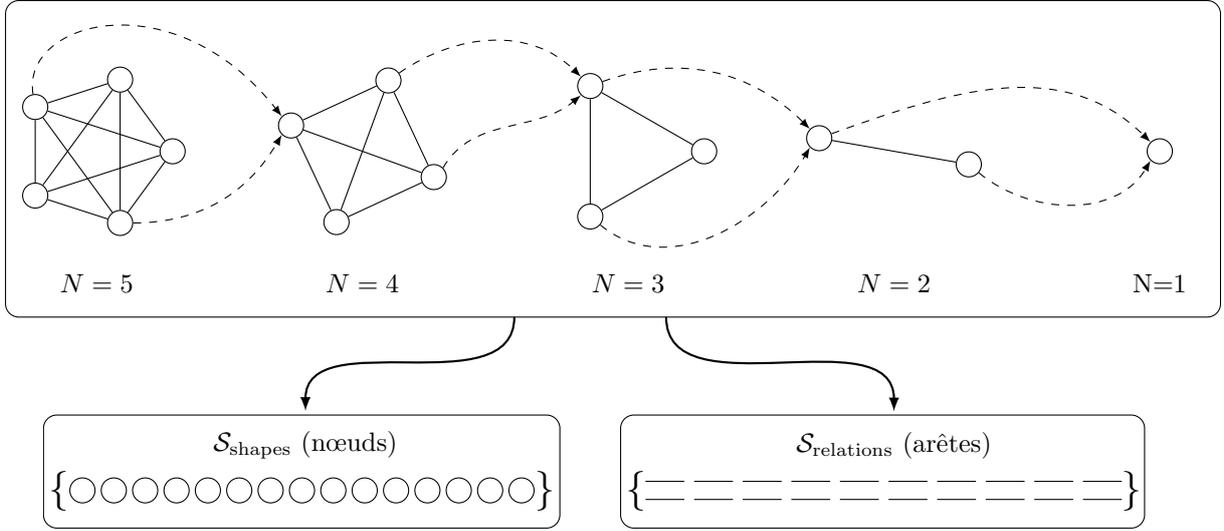


Figure 8.1 – Illustration schématique de l’étape de désassemblage d’un descripteur FHD hiérarchique à 5 niveaux, en un ensemble $\mathcal{S}_{\text{shapes}}$ de 15 attributs de formes (les nœuds) et un ensemble $\mathcal{S}_{\text{relations}}$ de 20 attributs de relations spatiales (les arêtes).

8.2 Désassemblage du descripteur FHD hiérarchique

Afin de combiner la définition des descripteurs FHD du [Chapitre 7](#) à une approche d’apprentissage de type sacs de caractéristiques, notre proposition consiste à relâcher la contrainte structurelle de ces descripteurs, en effectuant une opération dite de désassemblage des graphes. Cette étape a pour but de former, à partir des attributs associés aux descripteurs FHD, des ensembles non-ordonnés de F-Histogrammes. Étant donné une image et son descripteur FHD hiérarchique $\mathcal{F} = \{\mathcal{F}_i\}_{i=1}^N$ correspondant, cette opération de désassemblage produit deux ensembles non-ordonnés d’attributs :

- L’ensemble $\mathcal{S}_{\text{shapes}}$ qui est composé de tous les F-Histogrammes attribués aux nœuds des différents ARG du descripteur FHD hiérarchique, décrivant la forme des sous-parties de l’image. On a alors $|\mathcal{S}_{\text{shapes}}| = \sum_{i=1}^N i = \frac{N(N+1)}{2}$ (chaque niveau \mathcal{F}_i comporte i nœuds) ;
- L’ensemble $\mathcal{S}_{\text{relations}}$ qui est composé de tous les F-Histogrammes attribués aux arêtes des différents ARG du descripteur FHD hiérarchique, décrivant la position relative des sous-parties de l’image les unes par rapport aux autres. On a alors $|\mathcal{S}_{\text{relations}}| = \sum_{i=2}^N \frac{i(i-1)}{2}$ (chaque niveau \mathcal{F}_i comporte $\frac{i(i-1)}{2}$ arêtes).

Une illustration schématique de cette opération de désassemblage d’un descripteur FHD hiérarchique est illustrée en [Figure 8.1](#). Nous pouvons y visualiser un descripteur FHD hiérarchique à 5 niveaux, alors désassemblé en un ensemble $\mathcal{S}_{\text{shapes}}$ composé de 15 attributs de formes (les nœuds) et un ensemble $\mathcal{S}_{\text{relations}}$ composé de 20 attributs de relations spatiales (les arêtes).

Dans notre contexte, cette opération de désassemblage apporte plusieurs avantages. D’une part, elle permet de surmonter les limitations liées à la modélisation statique par graphes, en

particulier par rapport aux problématiques de comparaison et de mise en correspondance. Ceci est particulièrement pertinent lorsque l'on considère des images composées d'un nombre variable de sous-parties, rendant alors difficile la comparaison des descripteurs FHD. D'autre part, une telle représentation non-ordonnée de caractéristiques est bien souvent plus adaptée pour des applications dans des approches par apprentissage supervisé usuellement utilisées dans des tâches de reconnaissance et de classification.

8.3 Vocabulaires structurels de relations spatiales et de formes

Nous considérons ici un ensemble d'images où chacune a été décomposée en ses sous-parties structurelles, et a ensuite été caractérisée par son descripteur FHD hiérarchique correspondant. La première étape de l'approche d'apprentissage que nous proposons consiste à construire des vocabulaires de configurations spatiales et de formes apparaissant entre les parties structurelles d'un sous-ensemble d'images, considéré comme base d'apprentissage. La seconde étape consiste ensuite à construire une représentation synthétique de chaque image à partir des vocabulaires construits lors de l'étape précédente. Ces représentations pourront alors être employées dans des processus de reconnaissance ou de classification d'images.

8.3.1 Construction des vocabulaires

Pour procéder à la construction des vocabulaires de formes et de relations spatiales, nous considérons ici M images issues d'une base d'apprentissage. Le descripteur FHD hiérarchique de chaque image est alors désassemblé selon la stratégie détaillée précédemment. Nous obtenons ainsi M ensembles d'attributs de relations spatiales $\mathcal{S}_{\text{relations}}$ et M ensembles d'attributs de formes $\mathcal{S}_{\text{shapes}}$. Les ensembles de relations spatiales et ceux de formes sont alors considérés séparément.

Afin de construire les vocabulaires de relations spatiales et de formes, nous appliquons une stratégie de quantification vectorielle afin de représenter ces ensembles d'attributs par un nombre de points réduits, qui représenteront les mots de nos vocabulaires. Cette quantification vectorielle est effectuée ici à l'aide de l'algorithme de *clustering K-Means* [MacQueen, 1967].

Nous appliquons l'algorithme deux fois :

- D'une part avec les attributs de relations spatiales, afin de produire $K_{\text{relations}}$ *clusters*, ceux-ci étant représentés par l'ensemble des points moyens $\mathcal{V}_{\text{relations}} = \{\mathbf{w}_1, \dots, \mathbf{w}_{K_{\text{relations}}}\}$;
- D'autre part avec les attributs de formes, afin de produire K_{shapes} *clusters*, ceux-ci étant représentés par l'ensemble des points moyens $\mathcal{V}_{\text{shapes}} = \{\mathbf{w}_1, \dots, \mathbf{w}_{K_{\text{shapes}}}\}$.

Ces points moyens (aussi appelés centroïdes) constituent les différents mots des vocabulaires de formes et de relations spatiales, synthétisant la diversité des formes et des configurations spatiales apparaissant au sein des différentes images de la base d'apprentissage, et ceci potentiellement à différents niveaux d'échelle. Les paramètres $K_{\text{relations}}$ et K_{shapes} correspondent donc respectivement au nombre de *mots de relations spatiales* (ou *mots spatiaux*) et de *mots de formes* contenus dans les vocabulaires. Ces nouveaux types de mots forment alors des vocabulaires que

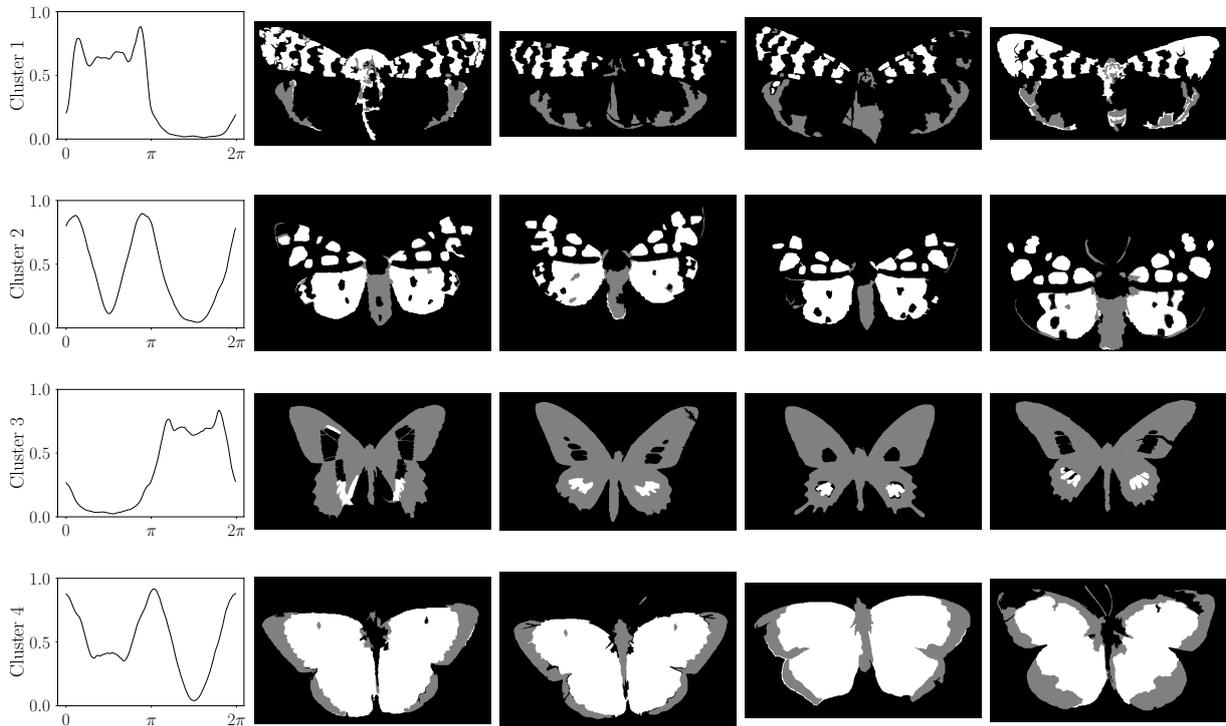


Figure 8.2 – Exemples illustratifs de mots de relations spatiales obtenus pour des images de papillons. Chaque ligne correspond à un *cluster* différent. À gauche : le centroïde du *cluster* représenté par une courbe. À droite : quelques exemples de couples d’objets dont l’histogramme de forces à été rattaché à ce centroïde (en blanc : objet *A*; en gris : objet *B*).

l’on peut qualifier de *structurels*, par analogie aux vocabulaires *visuels* produits par le *clustering* de caractéristiques visuelles locales.

Afin d’illustrer cette étape, nous présentons quelques exemples issus de la construction de vocabulaires pour deux jeux de données. Ceux-ci seront présentés plus en détail au cours des validations expérimentales du [Chapitre 9](#). La [Figure 8.2](#) et la [Figure 8.3](#) présentent respectivement quelques exemples de mots de relations spatiales et de mots de formes issus de vocabulaires construits à partir d’images représentant des papillons. Ces images ont été décomposées en sous-parties structurelles à l’aide de l’approche présentée précédemment dans la [Sous-section 7.2.2](#). De la même façon, la [Figure 8.4](#) et la [Figure 8.5](#) présentent respectivement des mots de relations spatiales et des mots de formes obtenus à partir d’un jeu de données représentant des images de scènes d’extérieur. Les sous-parties structurelles ont été obtenus à partir d’annotations fournies avec le jeu de données. Nous pouvons constater la capacité de cette approche de *clustering* à regrouper des formes caractéristiques, ainsi que des configurations spatiales similaires, et ce, indépendamment de la nature des objets. Par exemple, le *cluster* 1 de la [Figure 8.4](#) pourrait être associé à la relation spatiale « à gauche de », et les couples d’objets au sein de ce *cluster* représentent différents objets (ici, des arbres, des personnes ou des bâtiments).

8.3. Vocabulaires structurels de relations spatiales et de formes

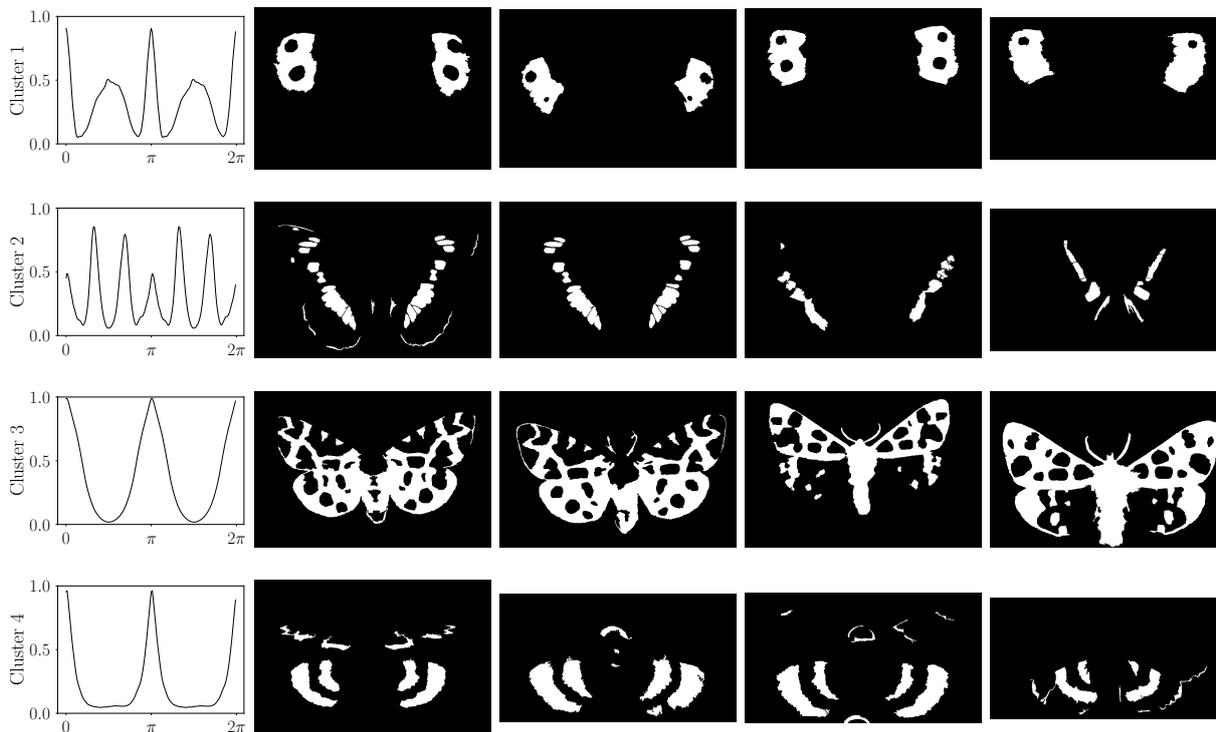


Figure 8.3 – Exemples illustratifs de mots de formes obtenus pour des images de papillons. Chaque ligne correspond à un *cluster* différent. À gauche : le centroïde du *cluster* représenté par une courbe. À droite : quelques exemples de formes dont l’histogramme de forces à été rattaché à ce centroïde.

8.3.2 Stratégies d’encodage et d’agrégation

Après la construction des vocabulaires de relations spatiales et de formes, la prochaine étape consiste à représenter synthétiquement chaque image par un vecteur de caractéristiques, en fonction des différents mots des vocabulaires. Pour construire ce vecteur de caractéristiques, le processus se déroule de la manière suivante. Dans un premier temps, les attributs de l’image (c’est-à-dire les F-Histogrammes des ensembles $\mathcal{S}_{relations}$ et \mathcal{S}_{shapes} de l’image) vont être *encodés* à partir des différents mots des vocabulaires structurels précédemment appris. Ces encodages des différents attributs sont ensuite *agrégés* au sein d’un même vecteur (étape dite de *pooling*) qui peut alors être interprété comme *multiensemble*, c’est-à-dire un ensemble fini où chaque élément peut apparaître plusieurs fois.

De nombreuses stratégies d’encodage et d’agrégation ont été proposées dans la littérature, en fonction des contextes et des bases d’images employées [Chatfield *et al.*, 2011]. Nous pouvons ici faire référence par exemple aux stratégies dites d’encodage lissé (*soft encoding*) proposées par [Philbin *et al.*, 2008] et par [Gemert *et al.*, 2010], ou encore aux approches reposant sur l’utilisation du noyau de Fisher [Jégou *et al.*, 2012 ; Sánchez *et al.*, 2013]. On pourra également citer l’approche de [Jiu *et al.*, 2012] qui propose de relier l’étape de construction du vocabulaire

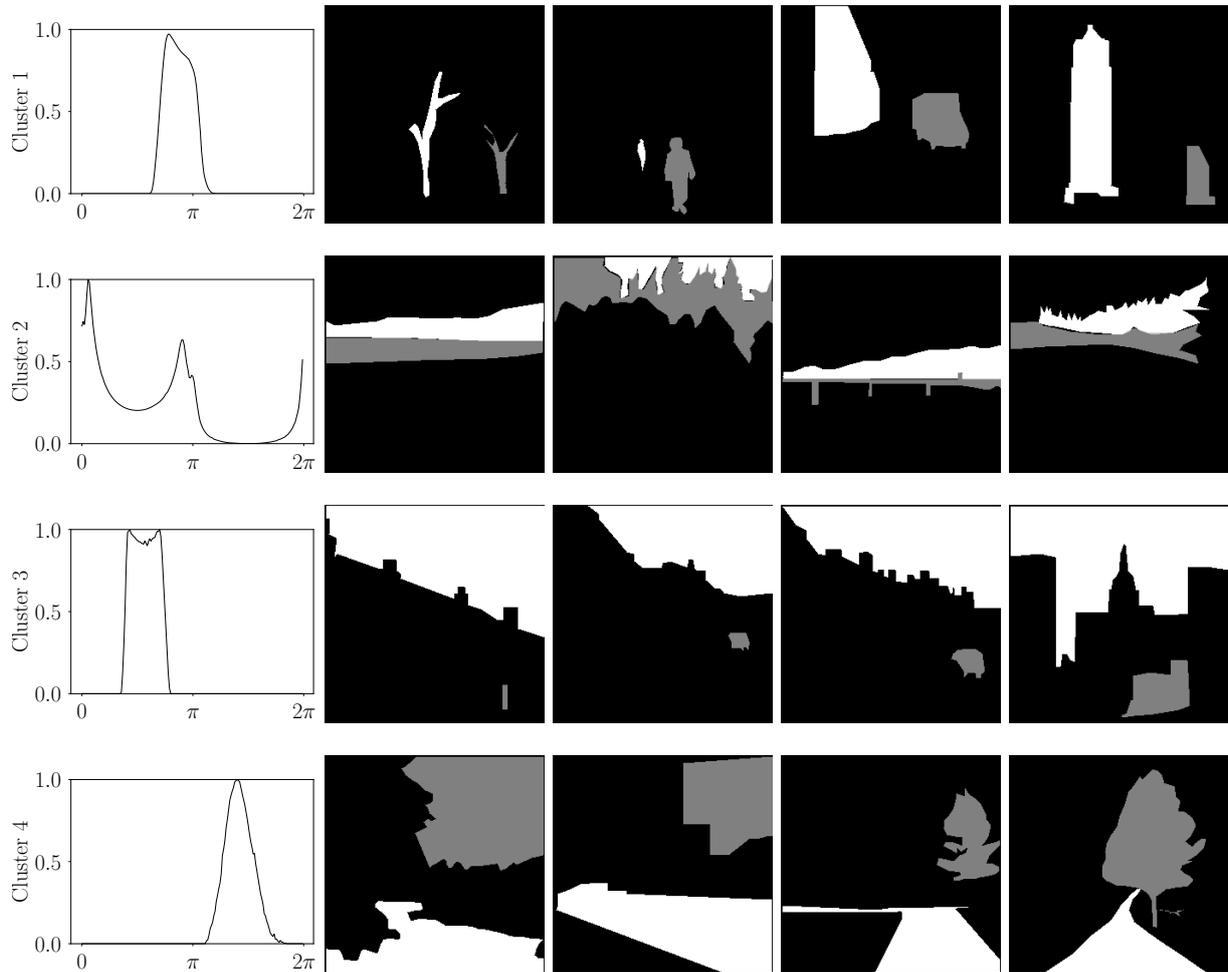


Figure 8.4 – Exemples illustratifs de mots de relations spatiales obtenus pour des images de scènes d’extérieur. Chaque ligne correspond à un *cluster* différent. À gauche : le centroïde du *cluster* représenté par une courbe. À droite : quelques exemples de couples d’objets dont l’histogramme de forces à été rattaché à ce centroïde (en blanc : objet *A* ; en gris : objet *B*).

avec celle de l’encodage de manière supervisée à partir de réseaux de neurones.

Ces différentes stratégies permettent généralement d’améliorer le caractère discriminant des représentations fondées sur des vocabulaires visuels. Dans le contexte de ces travaux, notre objectif n’est pas de proposer des optimisations avancées du modèle d’apprentissage par sacs de caractéristiques, mais plutôt d’illustrer le comportement de ce modèle lorsqu’il est appliqué d’un point de vue structural, à partir de caractéristiques telles que les descripteurs de position relative proposés. En ce sens, nous avons étudié les deux stratégies d’encodage les plus communément utilisées, à savoir l’encodage binaire classique et l’encodage lissé. Ces stratégies ont alors été combinées avec une agrégation dite par moyenne (*average pooling* en anglais), comme il sera explicité ci-après.

8.3. Vocabulaires structurels de relations spatiales et de formes

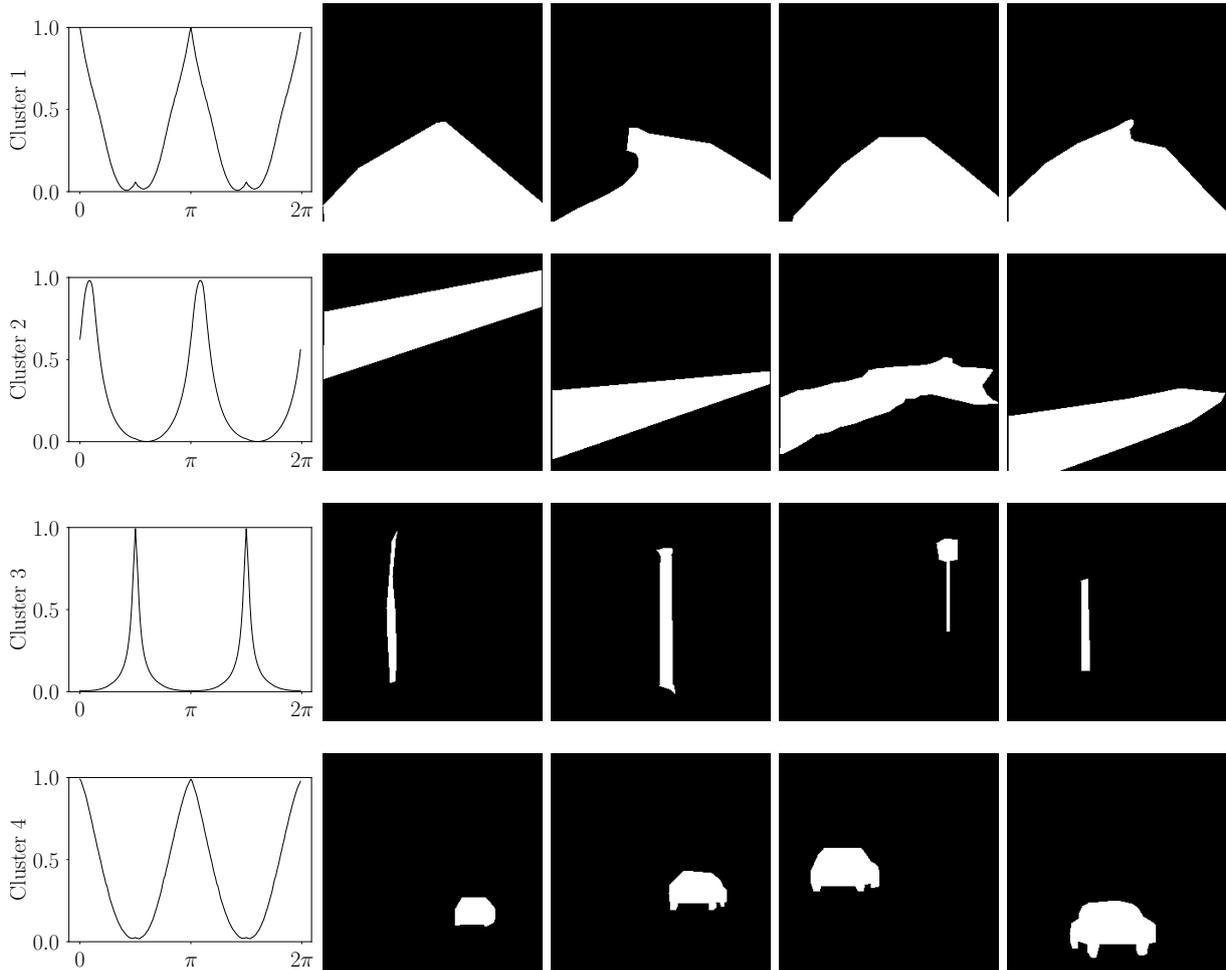


Figure 8.5 – Exemples illustratifs de mots de formes obtenus pour des images de scènes d’extérieur. Chaque ligne correspond à un *cluster* différent. À gauche : le centroïde du *cluster* représenté par une courbe. À droite : quelques exemples de formes dont l’histogramme de forces à été rattaché à ce centroïde.

Encodage binaire

La façon la plus courante de procéder à l’encodage des attributs consiste à assigner chaque attribut de l’image à son centroïde le plus proche, d’une manière binaire. Cette stratégie est souvent désignée dans la littérature par le terme *hard assignment*. Dans notre contexte, les F-Histogrammes sont ainsi respectivement assignés à leur centroïde le plus proche $\mathbf{w}_k \in \mathcal{V}$, où $\mathcal{V} = \{\mathbf{w}_1, \dots, \mathbf{w}_K\}$ désigne soit l’ensemble $\mathcal{V}_{\text{shapes}}$ pour les attributs de formes, soit l’ensemble $\mathcal{V}_{\text{relations}}$ pour les attributs de relations spatiales. L’image est alors décrite en comptabilisant le nombre d’occurrences d’assignations d’un attribut à chaque mot du vocabulaire. Ainsi, pour un mot représenté par le centroïde \mathbf{w}_k , l’encodage binaire suivi de l’agrégation par moyenne peuvent

être formulés de la manière suivante :

$$z_k = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1 & \text{si } \mathbf{w}_k = \underset{\mathbf{w} \in \mathcal{V}}{\operatorname{argmin}} \|\mathbf{f}_i - \mathbf{w}\| \\ 0 & \text{sinon} \end{cases} \quad (8.1)$$

où $\mathcal{S} = \{\mathbf{f}_i\}_{i=1}^n$ désigne l'ensemble d'attributs $\mathcal{S}_{\text{shapes}}$ ou $\mathcal{S}_{\text{relations}}$ de l'image. Les valeurs z_k sont calculées pour tous les mots $\mathbf{w}_k \in \mathcal{V}$, et concaténées au sein du vecteur de caractéristiques $\mathbf{z} = [z_1, \dots, z_K]$, qui peut être interprété comme un histogramme qui comptabilise le nombre d'occurrences de chaque mot du vocabulaire. Ces procédures d'encodage et d'agrégation sont donc réalisées indépendamment pour les deux vocabulaires $\mathcal{V}_{\text{shapes}}$ et $\mathcal{V}_{\text{relations}}$, produisant deux ensembles de vecteurs de caractéristiques résultant d'apprentissages dans des espaces différents.

Encodage lissé

La deuxième approche considérée pour l'encodage des attributs consiste à effectuer l'assignation des attributs aux différents mots d'une manière pondérée ou lissée. Cette approche est souvent désignée dans la littérature par le terme *soft assignment*. Le but est de prendre en compte la distance des attributs aux différents mots des vocabulaires, puis d'effectuer un lissage de ces distances à l'aide par exemple d'un noyau gaussien, d'une manière similaire aux stratégies d'estimation par noyau utilisées en statistique.

Nous considérons donc un vocabulaire de mots structurels $\mathcal{V} = \{\mathbf{w}_1, \dots, \mathbf{w}_K\}$ (de relations spatiales ou de formes), et une image représentée par un ensemble d'attributs $\mathcal{S} = \{\mathbf{f}_i\}_{i=1}^n$ (ici aussi, de relations spatiales ou de formes selon le vocabulaire associé). Pour un mot \mathbf{w}_k du vocabulaire, l'encodage lissé suivi de l'agrégation par moyenne sont obtenus de la façon suivante :

$$z_k = \frac{1}{n} \sum_{i=1}^n K_\sigma(\mathbf{f}_i, \mathbf{w}_k) \quad (8.2)$$

où K_σ désigne le noyau gaussien entre deux échantillons, donné par :

$$K_\sigma(\mathbf{x}_1, \mathbf{x}_2) = \exp - \frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{2\sigma^2} \quad (8.3)$$

où la variance σ contrôle la largeur du noyau. L'image est alors décrite par le vecteur de caractéristiques $\mathbf{z} = [z_1, \dots, z_K]$ qui résume l'agrégation des encodages lissés pour tous les mots du vocabulaire.

Cette stratégie d'encodage lissé présente l'avantage de pondérer la contribution de chaque attribut de l'image à travers les différents mots des vocabulaires, par opposition à la stratégie d'encodage binaire. En effet, l'algorithme de *clustering K-Means* peut générer des centroïdes qui sont proches les uns des autres, et les attributs peuvent parfois être positionnés à la frontière de plusieurs *clusters*, introduisant une certaine ambiguïté [Gemert *et al.*, 2010].

8.4. Bilan scientifique

Sacs de relations spatiales et de formes

Étant donné que nous construisons deux vocabulaires distincts, nous produisons deux types de vecteurs de caractéristiques pour les images : un ensemble pour les formes et un autre pour les relations spatiales. Nous nommons ainsi *sacs de relations spatiales* (*BoR* pour *bags of relations*) et *sacs de formes* (*BoS* pour *bags of shapes*) les vecteurs de caractéristiques respectifs des relations spatiales ou des formes seulement. La combinaison de ces vecteurs de caractéristiques par concaténation est alors nommée *sacs de relations spatiales et de formes* (*BoSR*). Ces différentes représentations pourront alors être utilisées dans des processus de classification et de reconnaissance d'images.

8.3.3 Compatibilité avec les caractéristiques visuelles locales

Un des avantages de cette stratégie d'apprentissage par sacs de relations spatiales et de formes est qu'elle est naturellement compatible avec les approches classiques par sacs de mots visuels, qui utilisent des caractéristiques locales extraites à partir des images. En effet, bien que les caractéristiques initialement calculées sur les images soient de natures différentes, les multiensembles obtenus à l'issue de l'apprentissage par ces deux types d'approches peuvent alors être naturellement combinés par concaténation de vecteurs. Ceci permet d'obtenir des représentations hybrides originales qui assemblent descriptions structurelles et descriptions locales, capturant ainsi des informations complémentaires à différentes échelles et à différents niveaux d'analyse. Nous verrons notamment dans les validations expérimentales du [Chapitre 9](#) que cette nouvelle stratégie de combinaison peut conduire à des améliorations significatives des taux de reconnaissance obtenus pour certaines tâches de classification d'images.

8.4 Bilan scientifique

Dans ce chapitre, nous avons présenté une méthodologie d'apprentissage permettant d'exploiter efficacement l'information structurelle contenue dans les descripteurs FHD hiérarchiques. Cette méthodologie, inspirée des approches par sacs de caractéristiques, permet de construire des vocabulaires structurels, dont les mots correspondent à des prototypes de configurations spatiales et de formes apparaissant à différents niveaux d'échelle dans les images. Ces nouvelles représentations structurelles sont alors naturellement compatibles avec les vocabulaires visuels classiques, reposant sur l'extraction de caractéristiques locales à partir des images. Le chapitre suivant présente alors les validations expérimentales réalisées sur plusieurs jeux de données pour illustrer l'intérêt de cette approche.

Cette méthodologie a été valorisée par la publication d'un article au sein d'une conférence internationale [Clément *et al.*, 2016a], ainsi que d'une conférence nationale [Clément *et al.*, 2017e]. Un article de revue internationale, présentant également l'ensemble des validations expérimentales du chapitre suivant, est actuellement en cours de révision [Clément *et al.*, 2017b].

Chapitre 9

Validations expérimentales : reconnaissance d'images structurées

Sommaire

9.1	Protocole expérimental général	126
9.1.1	Méthodes comparatives de référence	126
9.1.2	Classification et évaluation des résultats	126
9.2	Reconnaissance d'objets structurés	127
9.2.1	Jeux de données	127
9.2.2	Protocole de validation	129
9.2.3	Résultats expérimentaux et discussion	130
9.3	Reconnaissance d'images de scènes d'extérieur	132
9.3.1	Jeu de données	132
9.3.2	Protocole de validation	133
9.3.3	Résultats expérimentaux et discussion	134
9.4	Bilan scientifique	136

Ce chapitre présente les différentes validations expérimentales menées pour illustrer l'intérêt des descripteurs FHD hiérarchiques ([Chapitre 7](#)), couplés à la stratégie d'apprentissage par sacs de relations spatiales ([Chapitre 8](#)), au sein de processus de reconnaissance d'images. Pour cela, nous proposons deux applications expérimentales principales. Tout d'abord, nous décrivons d'un point de vue général le protocole expérimental mis en œuvre pour ces deux applications ([Section 9.1](#)). La première application consiste à caractériser et reconnaître des images représentant des objets structurés en considérant une décomposition hiérarchique ([Section 9.2](#)). Nous employons pour cela trois jeux de données composés respectivement de différents types d'objets : des papillons, des motos et des pâtisseries. Dans la seconde application, notre approche est alors exploitée pour la reconnaissance d'images de scènes d'extérieur ([Section 9.3](#)), en considérant des régions d'intérêt correspondant à des objets sémantiques. Le chapitre est alors conclu par un bilan de ces expérimentations ([Section 9.4](#)).

9.1 Protocole expérimental général

Pour valider l'intérêt des descripteurs FHD hiérarchiques ainsi que l'apprentissage par sacs de relations spatiales, nous proposons deux applications expérimentales, qui diffèrent notamment par la nature des jeux de données et par les méthodes employées pour décomposer les images en sous-parties structurelles. Dans ce chapitre, les histogrammes de forces attribués aux descripteurs FHD sont calculés sur un ensemble de 180 directions discrètes, et avec par défaut une force constante ($r = 0$). L'influence d'autres types de forces sera notamment étudiée expérimentalement. Ci-après, nous présentons les approches comparatives utilisées pour juger de l'efficacité de notre méthode de description structurelle. Nous détaillons également le protocole de classification employé et les différents indices d'évaluation calculés pour évaluer quantitativement la qualité des résultats obtenus.

9.1.1 Méthodes comparatives de référence

Pour comparer les résultats obtenus par notre approche de description structurelle et d'apprentissage par sacs de relations spatiales, nous avons décidé de mettre en place une approche classique par sacs de caractéristiques, à partir de descripteurs locaux calculés sur différents points des images. Ceci nous permet à la fois de disposer de résultats de référence pour les jeux de données étudiés, mais aussi d'évaluer la stratégie de combinaison de ces deux types d'approches au sein de représentations hybrides.

Dans ces expérimentations, nous avons utilisé les histogrammes de gradients orientés (notés *HOG* pour *Histograms of Oriented Gradients*) [Dalal et Triggs, 2005] comme caractéristiques locales sur les images. Le principe du descripteur HOG consiste à accumuler au sein d'un histogramme les orientations des gradients au sein d'une fenêtre locale, d'une manière similaire au descripteur SIFT [Lowe, 2004]. Ici, les descripteurs HOG ont été calculés selon une grille dense, c'est-à-dire sur toute l'image à partir de points régulièrement espacés. Chaque image est donc subdivisée en un ensemble de fenêtres locales de tailles 32×32 pixels, sans chevauchements entre les fenêtres. Les orientations des gradients au sein de chaque fenêtre sont interpolées selon 9 directions discrètes, résultant donc en un vecteur de caractéristiques de dimension 9.

Nous suivons alors le même processus d'apprentissage que pour les sacs de relations spatiales, mais avec les caractéristiques locales HOG : pour un ensemble d'images issues d'une base d'apprentissage, un vocabulaire de mots visuels est construit à partir de l'algorithme *K-Means*, et chaque image est représentée par un vecteur de caractéristiques résultant de l'encodage (binaire ou lissé) et l'agrégation de ses caractéristiques locales, en fonction de ce vocabulaire.

9.1.2 Classification et évaluation des résultats

Dans les expérimentations qui suivent, nous mettons en place un protocole de classification d'images à partir des approches par sacs de relations spatiales et de formes (désignées par BoR, BoS et BoSR dans la suite).

Pour procéder à la classification des images, nous avons utilisé les machines à vecteurs de

9.2. Reconnaissance d'objets structurés

support (*SVM* pour *Support Vector Machines*) en fournissant comme données d'entrée les vecteurs de caractéristiques résultant de la procédure d'apprentissage par vocabulaires. Un noyau linéaire a été utilisé comme mesure de similarité entre les vecteurs caractéristiques. D'autres noyaux communément employés pour ce type de tâches ont été expérimentés (comme le noyau Gaussien et celui du χ^2), mais dans ce contexte ils n'ont pas permis d'améliorer significativement les résultats obtenus, tout en augmentant largement les temps de calcul. La classification est effectuée selon différentes méthodes de validation croisée, en fonction de l'application considérée. Les classifieurs SVM ont été mis en place selon une stratégie *one-against-all* pour la classification multiclassées : un classifieur SVM binaire est créé pour chaque classe (dont tous les échantillons sont considérés comme positifs, et tous les autres sont négatifs), puis les résultats de ces classifieurs binaires sont combinés. Par ailleurs, l'hyperparamètre dit de marge souple des SVM est optimisé par validation croisée sur la base d'apprentissage considérée, selon un découpage de type *5-fold*.

Par ailleurs, pour la stratégie d'encodage lissé lors de l'affectation des attributs aux mots de vocabulaire (dans le [Chapitre 8](#), voir l'[Équation 8.2](#) et l'[Équation 8.3](#)), nous avons utilisé un noyau Gaussien adaptatif $K_{\tilde{\sigma}}$ où la variance $\tilde{\sigma}$ est fixée à la valeur moyenne de l'ensemble des distances Euclidiennes entre les attributs d'apprentissage et leurs mots de vocabulaires associés.

Les résultats de classification sont alors évalués quantitativement à l'aide des mêmes indicateurs statistiques présentés dans le [Chapitre 5](#) pour les validations sur les descripteurs d'enlacement : le taux de reconnaissance, la précision et le rappel.

9.2 Reconnaissance d'objets structurés

9.2.1 Jeux de données

Pour ces validations expérimentales, nous avons employé trois jeux de données d'images en couleurs représentant des objets structurés. La [Figure 9.1](#) présente quelques exemples d'images issues de ces trois jeux de données : *Peale*, *Motorbikes* et *Patisboul*.

Le jeu de données *Peale* est un ensemble d'images issues d'une collection entomologique de l'*Academy of Natural Sciences* à Philadelphie¹, initiée par Titian Ramsay Peale au XIX^e siècle. Au sein de cette collection, nous avons sélectionné un sous-ensemble de 114 images en couleurs, chacune représentant un papillon vu du dessus. Parmi ces images, 11 espèces de papillons sont représentées, avec un nombre variable d'échantillons pour chaque espèce (on retrouve de 5 à 20 images par classe). Cette base d'images a été sélectionnée car les papillons sont des objets typiques où la forme, ainsi que la disposition spatiale des différents motifs apparaissant sur les ailes, représentent des caractéristiques déterminantes pour distinguer les différentes espèces.

Le jeu de données *Motorbikes* est un ensemble d'images représentant des motos, initialement proposé par *Caltech* (*California Institute of Technology*) au début des années 2000². Pour les besoins de ces validations expérimentales, nous avons décidé de restreindre ce jeu de données, en

1. <http://clade.ansp.org/entomology>

2. <http://vision.caltech.edu/archive.html>



(a) Images du jeu de données *Peale*



(b) Images du jeu de données *Motorbikes*



(c) Images du jeu de données *Patisboul*

Figure 9.1 – Exemples d'images provenant des jeux de données *Peale*, *Motorbikes* et *Patisboul*, employés pour la reconnaissance d'objets structurés. Pour chaque jeu de données, les images illustrées correspondent chacune à une classe différente.

ne gardant que des images où les motos représentées sont vues de côté et avec un fond uniforme. Notre sous-ensemble est composé de 265 images. Nous avons alors manuellement étiqueté les images en 5 classes distinctes : *cross*, *cruiser*, *mobyettes*, *routières* et *sportives*. Comme pour les papillons, les motos sont des objets où la configuration spatiale entre leurs sous-parties structurelles (roues, moteur, carénages, etc.) sont des éléments distinctifs des différents types de modèles.

Enfin, le jeu de données *Patisboul* est un ensemble d'images représentant des pâtisseries. Il a été confectionné à partir d'un moteur de recherche en ligne, et est composé d'un total de 170 images répertoriant 17 types de pâtisseries (chaque classe est composée de 10 échantillons). Là aussi, la forme et les relations spatiales entre les différentes régions constituantes des objets peuvent permettre d'identifier leur nature. Il convient cependant de noter que ce jeu de données est plus « difficile » que les précédents, notamment car les objets sont représentés sous différents angles de vues.

9.2. Reconnaissance d'objets structurés

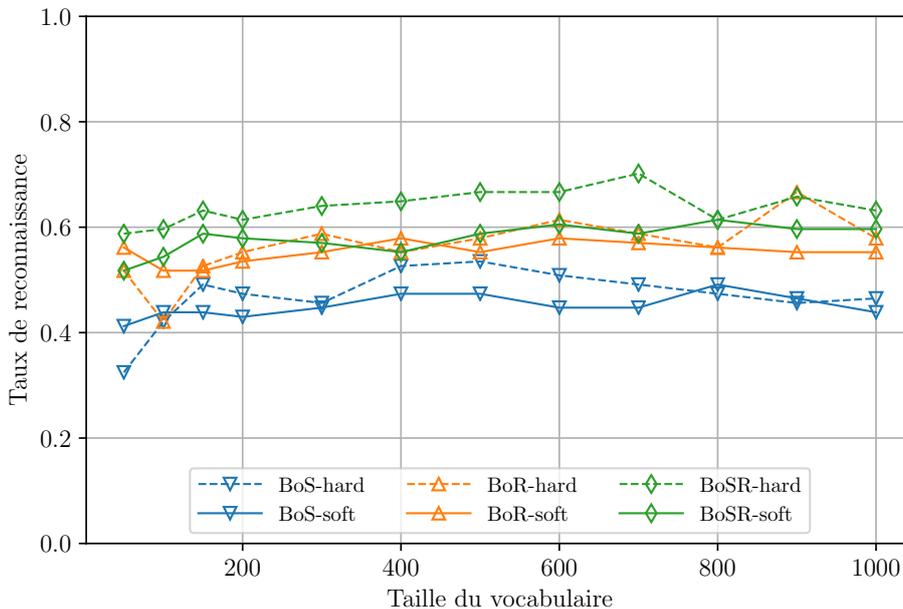


Figure 9.2 – Taux de reconnaissance obtenus pour la classification des images du jeu de données *Peale*, pour différentes tailles de vocabulaires, et pour les différentes stratégies d'encodage (encodage binaire : *hard*; encodage lissé : *soft*).

9.2.2 Protocole de validation

Pour chacun de ces jeux de données, les images sont décomposées en suivant la stratégie de segmentation hiérarchique présentée dans la [Sous-section 7.2.2](#) du [Chapitre 7](#). Les fonds des images étant homogènes, ils ont été supprimés préalablement et n'entrent pas en compte dans ce processus de segmentation. Pour l'algorithme *Mean Shift*, les paramètres de fenêtre spatiale et d'intensité ont été fixés par les valeurs $(h_s, h_r) = (8, 8)$, avec une taille minimum de régions de $M = 100$. Ceci permet généralement d'obtenir une sur-segmentation des images. À partir de cette sur-segmentation, le *clustering* hiérarchique ascendant est appliqué à partir des segments initiaux pour regrouper progressivement les régions au sein du dendrogramme, jusqu'à la racine qui correspond alors à l'ensemble de l'objet (le fond ayant déjà été ignoré). Dans le cadre de ces expérimentations, les descripteurs FHD hiérarchiques ont alors été calculés à partir d'un niveau de profondeur initial fixé à $k = 5$. La propriété de bilinéarité des F-Histogrammes permet alors d'obtenir les étages supérieurs des descripteurs par fusion des régions. De cette façon, chaque image est alors représentée par un descripteur FHD à 5 niveaux, chaque niveau i étant composé de i sous-parties.

Après désassemblage de ce descripteur, dans le but d'appliquer ensuite la stratégie d'apprentissage par sacs de relations spatiales, nous obtenons alors un total de $|S_{\text{shapes}}| = \frac{5 \times 6}{2} = 15$ F-Histogrammes de formes, et $|S_{\text{relations}}| = \sum_{i=2}^5 \frac{i(i-1)}{2} = 20$ F-Histogrammes de relations spatiales par image.

Par ailleurs, puisque les trois jeux de données proposés pour cette application ne sont pas

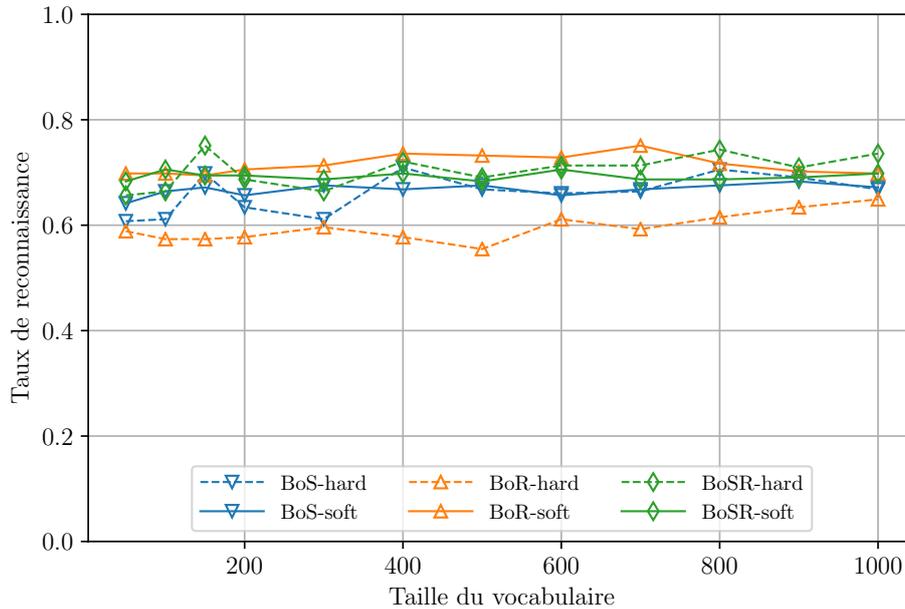


Figure 9.3 – Taux de reconnaissance obtenus pour la classification des images du jeu de données *Motorbikes*, pour différentes tailles de vocabulaires, et pour les différentes stratégies d'encodage (encodage binaire : *hard* ; encodage lissé : *soft*).

de tailles trop importantes, nous avons mis en place un protocole de validation croisée de type *5-fold* pour effectuer la classification des images de ces trois jeux de données.

9.2.3 Résultats expérimentaux et discussion

La Figure 9.2, la Figure 9.3 et la Figure 9.4 présentent les taux de reconnaissance obtenus pour la classification des images des jeux de données *Peale*, *Motorbikes* et *Patisboul* respectivement. Ces résultats de classification ont été obtenus pour différentes tailles de vocabulaires allant de 50 à 1000 mots pour les approches BoS (sacs de formes), BoR (sacs de relations) et BoSR (combinaison des deux), ainsi que pour les stratégies d'encodage binaire et lissé. Le paramètre de taille de vocabulaire correspond à la valeur de K pour l'application de l'algorithme de *clustering K-Means* sur les F-Histogrammes durant la phase d'apprentissage. Il s'agit donc d'un paramètre important de l'approche : des vocabulaires composés de trop peu de mots peuvent donner lieu à de mauvaises capacités de discrimination ; tandis que des vocabulaires trop grands peuvent conduire à des situations de surapprentissage ainsi qu'à des temps de calcul prohibitifs en raison de la grande dimensionnalité des vecteurs de caractéristiques obtenus.

À partir de ces résultats, nous pouvons dans un premier temps observer une relative robustesse des taux de reconnaissance obtenus par rapport aux différentes tailles de vocabulaires expérimentées. Cette stabilité suggère qu'un faible nombre de mots structurels sont généralement suffisants pour capturer l'essentiel de l'information spatiale discriminante au sein des images de ces jeux de données. Par ailleurs, nous pouvons également remarquer que, selon le jeu de don-

9.2. Reconnaissance d'objets structurés

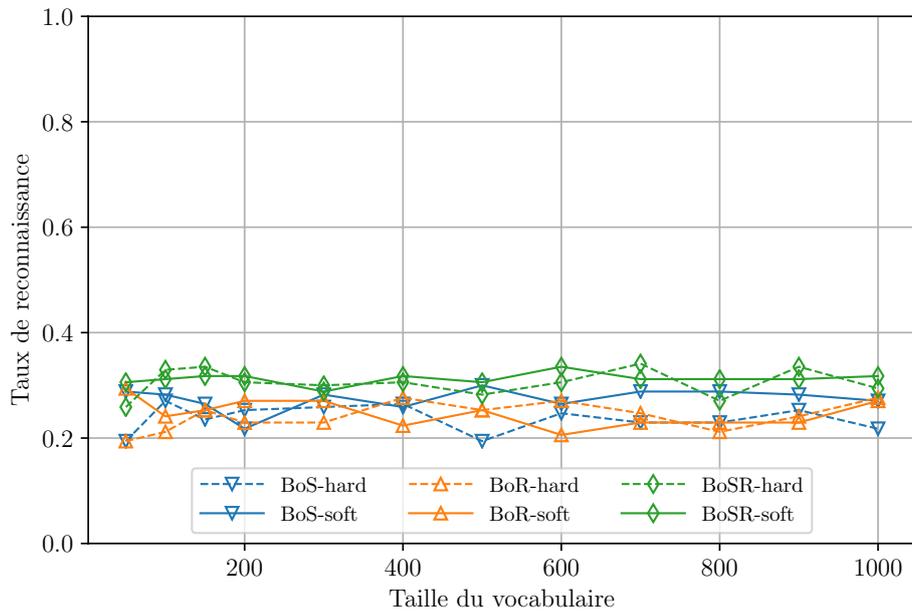


Figure 9.4 – Taux de reconnaissance obtenus pour la classification des images du jeu de données *Patisboul*, pour différentes tailles de vocabulaires, et pour les différentes stratégies d'encodage (encodage binaire : *hard* ; encodage lissé : *soft*).

nées, à la fois les formes et les relations spatiales peuvent être utiles pour reconnaître ce type d'objets structurés.

Dans le [Tableau 9.1](#), nous rapportons les résultats comparatifs globaux pour ces trois jeux de données. Ces résultats globaux correspondent aux moyennes des taux de reconnaissance obtenus pour les différentes tailles de vocabulaires allant de 50 à 1000, avec la stratégie d'encodage lissé. Dans l'ensemble, nous pouvons remarquer que notre approche structurale par sacs de relations spatiales et de formes conduit à des résultats comparables à ceux obtenus avec l'approche classique par sacs de caractéristiques locales. En effet, selon les jeux de données, les résultats des

Tableau 9.1 – Comparatif des taux de reconnaissance moyens pour les trois jeux de données considérés. Les résultats correspondent aux moyennes des taux de reconnaissance pour les différentes tailles de vocabulaire, avec un encodage lissé.

	<i>Peale</i>	<i>Motorbikes</i>	<i>Patisboul</i>
BoS	45,03	66,73	27,40
BoR	55,26	71,45	24,75
BoSR	57,82	69,28	31,27
HOG	43,13	78,02	29,17
BoSR+HOG	61,55	73,18	42,84



Figure 9.5 – Exemples d'images provenant du jeu de données *8 scènes*. Chaque image illustre une scène d'une classe différente.

approches BoS, BoR et BoSR sont sensiblement meilleurs ou légèrement en dessous de ceux obtenus pour l'approche HOG. En outre, à l'exception du jeu de données *Motorbikes*, l'approche BoSR+HOG qui combine informations structurelles et locales permet d'obtenir de meilleurs résultats que lorsque ces approches sont considérées séparément. Ainsi, ces résultats expérimentaux suggèrent que de telles représentations hybrides peuvent être utiles pour améliorer la reconnaissance de ce type d'objets complexes.

9.3 Reconnaissance d'images de scènes d'extérieur

9.3.1 Jeu de données

Le jeu de données *8 scènes* est un ensemble d'images représentant des scènes d'extérieur³. Il est composé de 2686 images de taille 256×256 distribuées de manière homogène en 8 classes correspondant à différents types de scènes identifiées par les étiquettes suivantes : *coast*, *forest*, *highway*, *insidacity*, *mountain*, *opencountry*, *street* et *tallbulding*. Quelques exemples d'images provenant de ce jeu de données sont présentées en Figure 9.5.

Ce jeu de données est également fourni avec des annotations de référence qui spécifient la localisation des différents objets sémantiques représentés dans les images (comme par exemple le ciel, la route, les personnes ou encore les bâtiments). Ces annotations permettent de construire des masques binaires des différents objets, qui peuvent alors être interprétés comme les différentes sous-parties structurelles de chaque image (voir par exemple la Figure 7.2). En effet, ces objets

3. <http://people.csail.mit.edu/torralba/code/spatialenvelope/>

9.3. Reconnaissance d'images de scènes d'extérieur

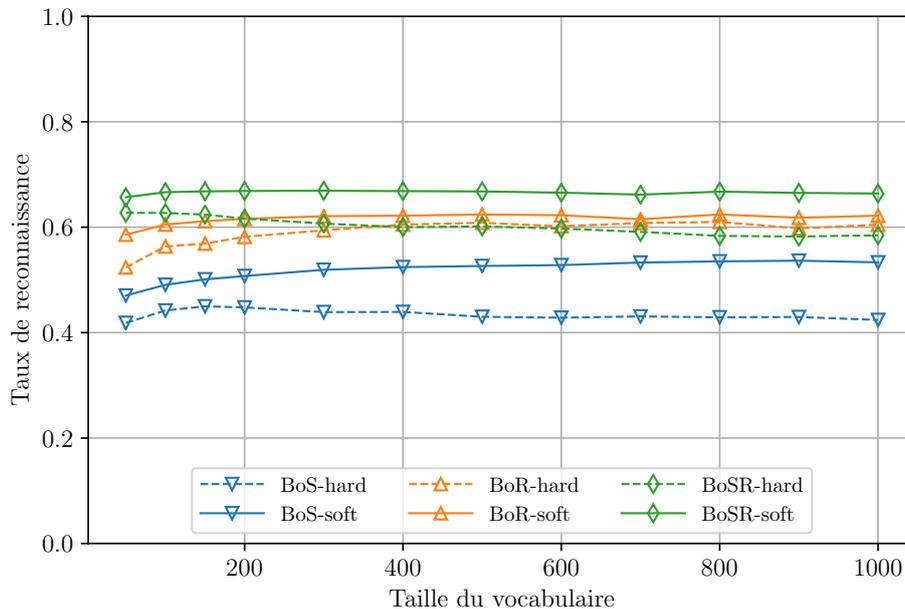


Figure 9.6 – Taux de reconnaissance pour la classification des images du jeu de données *8 scènes*, pour différentes tailles de vocabulaires, et pour les stratégies d’encodage binaire et lissé.

et ces surfaces constituent des composants distinctifs pour reconnaître la nature d’une scène. Les masques d’objets nous permettent donc d’appliquer notre approche de description structurale, sans avoir recours à une étape préalable de segmentation difficile pour ce type d’images, et qui serait susceptible d’introduire des erreurs. Du point de vue de ces validations expérimentales, cela nous permet d’évaluer de manière plus objective notre approche, et d’illustrer le potentiel descriptif contenu dans l’information spatiale. En effet, en utilisant les masques binaires directement comme sous-parties structurales, l’approche ne considère pas les valeurs des pixels des images mais uniquement les formes et les configurations spatiales entre les différents objets.

9.3.2 Protocole de validation

Comme ce jeu de données est significativement plus grand que ceux utilisés dans l’application expérimentale précédente, nous avons mis en place un protocole expérimental légèrement différent pour l’étape de classification. Plutôt que d’effectuer une validation croisée de type *5-fold*, le jeu de données est découpé en un ensemble d’apprentissage représentant 25% de sa taille (671 images d’apprentissage) et un ensemble de test correspondant aux 75% restants (2015 images de test). Ensuite, afin de mieux évaluer la robustesse de l’approche, ce processus de classification est répété 10 fois avec différentes initialisations aléatoires de ces ensembles d’apprentissage (en préservant à chaque fois la distribution des différentes classes du jeu de données). Nous évaluons ensuite les résultats de classification à partir du taux de reconnaissance, de la précision et du rappel pour chaque itération, et nous reportons alors les scores moyens pour l’ensemble des itérations.

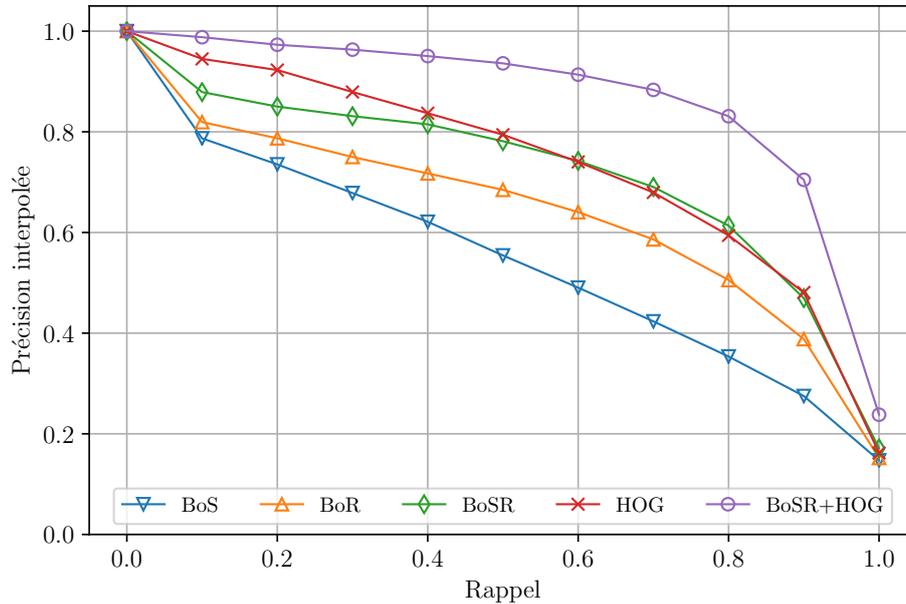


Figure 9.7 – Courbes de précision-rappel interpolées obtenues pour le jeu de données *8 scènes*, pour les différentes approches proposées (BoS, BoR et BoSR), pour l’approche comparative (HOG), et pour la combinaison de ces approches (BoSR+HOG).

9.3.3 Résultats expérimentaux et discussion

De la même manière que pour les jeux de données de l’application expérimentale précédente, la Figure 9.6 présente l’évolution des taux de reconnaissance obtenus pour la classification des images du jeu de données *8 scènes*, pour les différentes approches proposées (BoS, BoR et BoSR), pour différentes tailles de vocabulaires allant de 50 à 1000, ainsi que pour les deux stratégies d’encodage (binaire et lissé). À partir de ces résultats, nous pouvons de nouveau remarquer une stabilité globale des taux de reconnaissance obtenus par rapport aux différentes tailles de vocabulaires. Ceci suggère que même pour ce jeu de données composé d’un plus grand nombre d’images, l’essentiel de l’information structurale est capturée par des vocabulaires de petites tailles. Il s’agit d’un résultat intéressant d’un point de vue computationnel, car ces représentations synthétiques induisent des temps de calcul plus courts et des capacités de passage à l’échelle plus importantes. Par ailleurs, nous pouvons également observer que la stratégie d’encodage lissé permet ici d’améliorer sensiblement les résultats de classification pour toutes les approches. Après prise en considération de ces résultats, les expérimentations plus détaillées présentées dans la suite de cette section sont alors rapportées pour une taille de vocabulaire fixée à $K = 100$ et pour la stratégie d’encodage lissé.

La Figure 9.7 montre les courbes de précision-rappel moyennes obtenues pour les différentes approches proposées (BoS, BoR et BoSR), pour l’approche comparative (HOG), et pour la combinaison de ces approches (BoSR+HOG). Nous pouvons ainsi observer que notre approche BoSR permet d’obtenir des résultats comparables à ceux obtenus par l’approche HOG fondée

9.3. Reconnaissance d'images de scènes d'extérieur

Tableau 9.2 – Évolution des taux de reconnaissance par classe (jeu de données *8 scènes*), pour différents types de forces appliquées aux F-Histogrammes (approche BoR).

	$r = 0$	$r = 2 (\Delta\%)$	$r = -2 (\Delta\%)$
<i>coast</i>	21,2	+14,1	+1,7
<i>forest</i>	66,9	-1,3	+0,5
<i>highway</i>	74,0	-11,3	+0,9
<i>insidicity</i>	71,7	+9,6	+0,2
<i>mountain</i>	70,2	+2,5	-1,3
<i>opencountry</i>	52,4	-10,8	+0,3
<i>street</i>	86,1	+3,8	-3,1
<i>tallbuilding</i>	53,0	+16,3	+0,1

sur des caractéristiques locales. En outre, la combinaison de ces deux approches conduit alors à une augmentation manifeste des résultats de classification. Ces observations viennent donc à nouveau confirmer l'idée selon laquelle des descriptions locales et structurelles peuvent être envisagées comme des informations complémentaires pour améliorer l'efficacité de systèmes de reconnaissance d'images.

Afin d'obtenir des résultats plus détaillés par rapport aux relations spatiales, le [Tableau 9.2](#) propose une étude de l'influence du type de forces utilisé pour les F-Histogrammes des descripteurs FHD (c'est-à-dire l'influence de la valeur de r dans l'[Équation 7.2](#) du [Chapitre 7](#)) sur les résultats de classification de l'approche BoR. Ce paramètre permet de prendre explicitement en compte la distance entre les objets lors du calcul des histogrammes de forces. Nous considérons la force constante $r = 0$ comme une base de référence, et nous observons comment les taux de reconnaissance pour les différentes classes du jeu de données évoluent lorsque l'on considère une force gravitationnelle (avec $r = 2$) ou une force répulsive (avec $r = -2$). Nous pouvons observer que pour la force gravitationnelle, certaines classes sont alors mieux reconnues (par exemple les classes *coast*, *insidicity* et *tallbuilding*). Au contraire, la force répulsive ne semble pas altérer significativement les résultats de classification pour cette base d'images.

Enfin, la [Figure 9.8](#) présente les matrices de confusion moyennes obtenues pour les différentes approches considérées, permettant ainsi d'évaluer leurs performances relatives en fonction des différentes classes du jeu de données. À partir de ces résultats, nous pouvons notamment observer une confusion entre les classes *coast* et *opencountry* pour notre approche. Ce résultat peut être expliqué par le fait que ces images représentent généralement des scènes de paysages qui sont similaires en termes de structures spatiales. Par ailleurs, nous pouvons remarquer qu'approximativement la moitié des classes du jeu de données sont mieux reconnues par l'approche locale HOG, tandis que l'autre moitié est mieux reconnue par notre approche structurelle BoSR. Par exemple, notre approche semble plus efficace pour la reconnaissance de scènes spatialement complexes comme les classes *insidicity* et *street*, où l'on retrouve généralement de nombreux objets

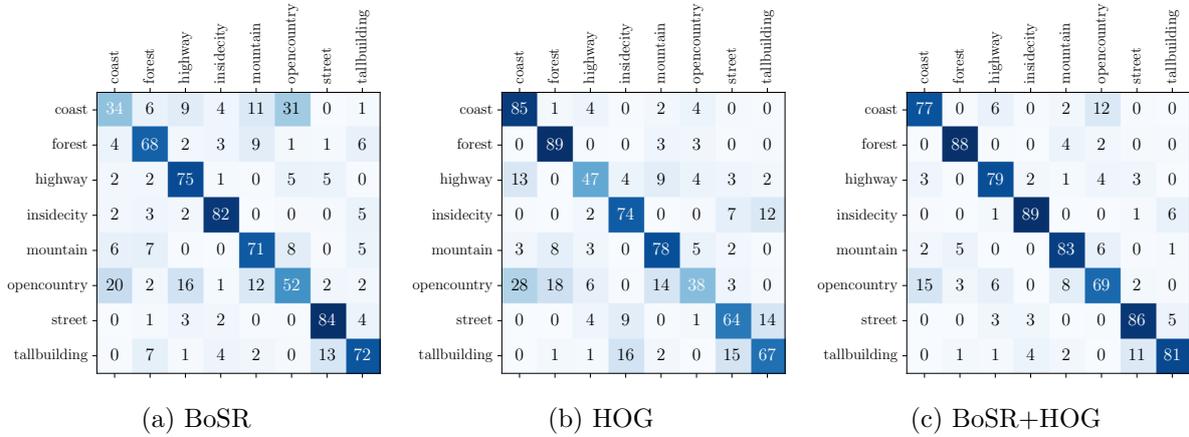


Figure 9.8 – Comparaison des matrices de confusion moyennes obtenues pour le jeu de données *8 scènes*, pour les approches (a) BoSR (sacs de relations spatiales et de formes), (b) HOG (sacs de caractéristiques locales) et (c) BoSR+HOG (combinaison des deux approches).

disposés différemment. Finalement, lorsque ces deux approches sont combinées, nous observons bien qu'elles tirent parti l'une de l'autre, permettant ainsi d'obtenir des taux de classification plus élevés et plus homogènes pour les différentes classes de ce jeu de données.

9.4 Bilan scientifique

Au cours de ce chapitre, nous avons présenté un ensemble des validations expérimentales permettant d'attester de l'intérêt des descripteurs FHD couplés à la stratégie d'apprentissage par sacs de relations spatiales et de formes. Ces expérimentations ont été divisées en deux parties principales. Dans une première partie, trois jeux de données d'images représentant des objets structurés (des papillons, des motos et des pâtisseries) ont été utilisés pour illustrer notamment l'intérêt de l'aspect hiérarchique des descripteurs FHD. Dans une seconde partie, un jeu de données d'images de scènes d'extérieur a été considéré pour illustrer la capacité de généralisation des sacs de relations spatiales par rapport à une plus grande base d'images, où les régions d'intérêt correspondent à des objets sémantiques. Par ailleurs, nous avons pu constater la complémentarité de nos représentations structurelles et spatiales avec les vocabulaires visuels classiques reposant sur des caractéristiques locales.

Les validations expérimentales présentées dans ce chapitre ont été valorisées au sein de plusieurs publications. Des résultats préliminaires obtenus pour les jeux de données *Peale* et *Motobikes* ont été présentés dans l'article de conférence internationale [Clément *et al.*, 2016a]. Les résultats obtenus pour la reconnaissance de scènes d'extérieur ont été présentés dans l'article de conférence nationale [Clément *et al.*, 2017e]. Enfin, l'ensemble des résultats expérimentaux obtenus pour ces différents jeux de données sont présentés dans l'article de revue internationale [Clément *et al.*, 2017b] qui est en cours de révision.

Chapitre 10

Vers l'apprentissage de descripteurs locaux d'enlacement

Sommaire

10.1 Contexte applicatif et motivations	138
10.1.1 Reconnaissance de lettrines décoratives anciennes	138
10.1.2 Configurations spatiales complexes locales	139
10.2 Approche par sacs d'enlacement local	140
10.2.1 Décomposition de lettrines en couches structurelles	140
10.2.2 Descripteurs locaux d'enlacement	142
10.2.3 Apprentissage par sacs d'enlacement	143
10.3 Validations expérimentales	144
10.3.1 Jeu de données	144
10.3.2 Protocole de validation	145
10.3.3 Résultats expérimentaux et discussion	146
10.4 Bilan scientifique	148

Dans ce chapitre, nous proposons de faire le lien entre le modèle de description de l'enlacement présenté dans le [Chapitre 4](#) et la stratégie d'apprentissage par sacs de relations spatiales du [Chapitre 8](#). Pour cela, nous nous intéressons de manière plus approfondie à la problématique de reconnaissance d'images de lettrines décoratives anciennes, dans le cadre d'une collaboration scientifique avec Mickaël Coustaty, à l'Université de La Rochelle. L'approche proposée ici repose ainsi sur la définition de vocabulaires de configurations spatiales complexes, obtenus à partir des descripteurs d'enlacement. Dans la [Section 10.1](#), nous décrivons tout d'abord le contexte applicatif et les motivations sous-jacentes à cette approche. Ensuite, en [Section 10.2](#), nous présentons une approche par sacs d'enlacement, où les descripteurs d'enlacement sont calculés à partir de fenêtres locales afin de capturer une information spatiale proche de la texture. Dans la [Section 10.3](#), nous reportons les résultats expérimentaux obtenus pour valider l'intérêt de cette approche. Enfin, le chapitre est conclu en [Section 10.4](#) par un bilan scientifique de ces travaux.



Figure 10.1 – Exemples d’images de lettrines décoratives, illustrant les différents styles de fonds, indépendamment de la lettre représentée. La première image a un fond de style *hachuré*, la deuxième a un fond *criblé*, et les deux dernières ont un fond *décoratif*, dont l’une avec une lettre de couleur inversée.

10.1 Contexte applicatif et motivations

Notre objectif est d’illustrer que les modèles développés dans la [Partie II](#) de ce manuscrit peuvent être intégrés au sein de la stratégie d’apprentissage proposée dans la [Partie III](#). Plus spécifiquement, nous proposons d’utiliser les descripteurs d’enlacement comme caractéristiques élémentaires au sein de l’apprentissage par sacs de relations spatiales. Afin de parvenir à cet objectif, nous nous focalisons sur le contexte applicatif complexe de la reconnaissance d’images de lettrines décoratives.

10.1.1 Reconnaissance de lettrines décoratives anciennes

Le patrimoine culturel de l’Europe constitue un bien public unique qui représente notre mémoire collective. L’analyse et la préservation d’un tel patrimoine a motivé le développement d’industries liées à la numérisation de contenus culturels et historiques, comme par exemple des œuvres d’art ou des textes manuscrits. En effet, la numérisation et l’analyse automatisée de telles données peuvent permettre de dater historiquement les périodes dont les documents proviennent, ou encore d’identifier leurs auteurs ou imprimeurs. Des traitements peuvent également être effectués pour restaurer la qualité d’anciens documents, qui subissent inexorablement diverses dégradations liées à leurs cycles de vie (encre ternie, jaunissement ou déchirement des pages, par exemple). Le contexte applicatif étudié ici est donc celui de l’analyse de tels documents anciens, ceux-ci présentant généralement un contenu hétérogène et faiblement structuré. Ces types de données sont bien souvent complexes et soulèvent des problématiques de recherche et d’extraction d’information, ainsi que de navigation parmi des corpus de documents. Ces problématiques ont notamment fait l’objet, ces dernières années, d’un projet ANR nommé NaviDoMass¹, qui a permis de réaliser des avancées notables dans ce domaine [[Jouili et al., 2010](#)].

Plus spécifiquement, les travaux présentés dans ce chapitre se concentrent sur l’analyse d’images historiques représentant des lettrines décoratives. Quelques exemples illustratifs de

1. <http://l3i.univ-larochelle.fr/NaviDoMass>

10.1. Contexte applicatif et motivations

telles images sont présentés en [Figure 10.1](#) (il s’agit du même type d’images que celles utilisées dans la [Section 5.3](#) du [Chapitre 5](#), pour valider l’intérêt des descripteurs d’entrelacement). Ces images sont mises à disposition par le Centre d’Études Supérieures de la Renaissance (CESR)² de l’Université François Rabelais de Tours. Sur ces exemples, nous pouvons constater que les images de lettrines peuvent généralement être décrites à partir de plusieurs caractéristiques : la lettre de l’alphabet qui est représentée, la couleur de cette lettre par rapport au fond, et les différents types de motifs présents sur le fond. L’ensemble de ces caractéristiques rendent ainsi ces images particulièrement complexes à analyser.

Dans le but de faire face à cette complexité, les auteurs de [[Naegel et Wendling, 2009](#) ; [Naegel et Wendling, 2010](#)] ont défini une méthode de binarisation fondée sur l’utilisation d’opérateurs connexes définis sur l’arbre morphologique des formes. Ces travaux permettent d’extraire des zones pertinentes de ces images, et se focalisent alors sur la reconnaissance automatique des lettres. Cependant, l’inconvénient principal de cette approche vient du fait qu’il est nécessaire de fixer *a priori* la couleur de la lettre (c’est-à-dire fond noir sur fond blanc, ou inversement) pour fonctionner correctement. Pour pallier ces limitations, les auteurs de [[Coustaty et al., 2011](#)] ont proposé une méthode de décomposition structurelle des images de lettrines reposant sur la décomposition de Meyer pour filtrer les images, et la loi de Zipf pour caractériser les différentes régions en fonction de la fréquence d’apparition de motifs à niveaux de gris. En plus de pouvoir extraire efficacement la lettre indépendamment du fond dans de nombreuses situations, cette approche permet d’obtenir des couches structurelles qui peuvent permettre d’analyser les détails des motifs représentés au niveau du fond des images de lettrines.

10.1.2 Configurations spatiales complexes locales

La plupart des travaux de recherche réalisés pour l’analyse des lettrines se sont focalisés sur la segmentation et l’extraction de la lettre représentée dans l’image. Dans ce contexte, nous nous intéressons ici plutôt à la caractérisation des types de motifs présents en arrière-plan de ces images, indépendamment de la lettre représentée. La [Figure 10.1](#), évoquée ci-avant, illustre notamment les styles de fonds qui peuvent apparaître au sein de différentes images de lettrines. Dans le [Chapitre 5](#), nous avons proposé de caractériser ces styles de fonds à l’aide du descripteur d’entrelacement, à partir d’une binarisation globale des images. Les résultats obtenus sur une base d’images restreinte nous ont permis d’illustrer de l’intérêt de ces descripteurs pour reconnaître ces styles de fonds. Nous proposons donc d’approfondir ces travaux, notamment à partir d’un jeu de données plus conséquent où une stratégie d’apprentissage peut être bénéfique pour répondre à des contraintes de passage à l’échelle. Par ailleurs, inspirés en partie des travaux de [[Fan et al., 2012](#)] et de [[Wang et al., 2015](#)] sur la définition de caractéristiques locales entre couches de pixels structurelles, nous proposons ici d’appliquer les descripteurs d’enlacement à une échelle locale sur les images de lettrines. L’objectif est donc de capturer la disposition spatiale de ces tracés, alors assimilables à des motifs texturaux, et de les intégrer au sein de vocabulaires de relations spatiales complexes.

2. <http://cesr.univ-tours.fr>

10.2 Approche par sacs d'enlacement local

Dans un premier temps, nous détaillons de manière synthétique le principe de la décomposition des images de lettrines à partir de la loi de Zipf [Coustaty *et al.*, 2011], qui permet d'obtenir différentes couches structurelles résumant leurs zones homogènes, leurs contours et leurs détails. Nous présentons ensuite comment étendre l'approche par sacs de relations spatiales afin qu'elle soit compatible avec des histogrammes d'enlacement calculés à une échelle locale, à partir de cette décomposition.

10.2.1 Décomposition de lettrines en couches structurelles

Comme nous l'avons évoqué dans la section précédente, les images de lettrines sont généralement composées de différentes couches d'information. La détection et l'extraction de ces couches structurelles constituent donc des tâches importantes pour analyser et interpréter le contenu de ce type d'images. En s'appuyant notamment sur les travaux proposés par Mickaël Coustaty dans sa thèse sur l'analyse d'images de lettrines [Coustaty, 2012], nous avons décidé d'utiliser la stratégie de segmentation à partir de la loi de Zipf comme un prétraitement à notre approche de description spatiale. Pour plus de détails sur cette étape de décomposition, il est également possible de se référer aux travaux de [Coustaty *et al.*, 2011].

La loi de Zipf [Zipf, 1949] est une loi de probabilités empirique qui repose notamment sur la loi de puissance. Cette loi formule une hypothèse sur la distribution des fréquences d'apparition de motifs topologiquement ordonnés. Elle affirme que les fréquences d'apparition de tels motifs sont fonction des rangs d'apparition de ces motifs lorsqu'ils sont triés par fréquences d'apparition. La loi de Zipf est ainsi formulée par :

$$N_{\sigma}(i) = k \times i^a \quad (10.1)$$

où $N_{\sigma}(i)$ correspond au nombre d'occurrences du motif classé au rang i , k est une constante liée à la taille du motif étudié, et a est un paramètre de la loi. L'application du logarithme sur cette équation permet d'obtenir une relation linéaire entre ces termes, et une régression linéaire permet alors d'estimer la valeur de l'exposant a .

Dans notre contexte d'analyse d'images de lettrines, les motifs considérés correspondent à l'ensemble des combinaisons de niveaux de gris possibles sur des masques de taille 3×3 pixels. Du fait que les images de lettrines soient essentiellement de nature binaire (elles représentent généralement de l'encre sur du papier), les différents niveaux de gris présents dans les images sont tout d'abord réduits au nombre de trois à partir d'un *clustering* par *K-Means*. Il existe alors 27 motifs possibles pour ces masques de 3×3 pixels. Par ailleurs, cette réduction du nombre de niveaux de gris est essentielle, car un nombre trop élevé de combinaisons possibles rendrait la fréquence d'apparition de nombreux motifs nulle ou très faible. Ces fréquences d'apparition ne pourraient alors pas être modélisées convenablement par la loi de Zipf.

Afin d'illustrer cette modélisation, la Figure 10.2 présente un graphe de Zipf obtenu pour une image de lettrine. L'approximation a été effectuée par la méthode des moindres carrés appliquée

10.2. Approche par sacs d'enlacement local

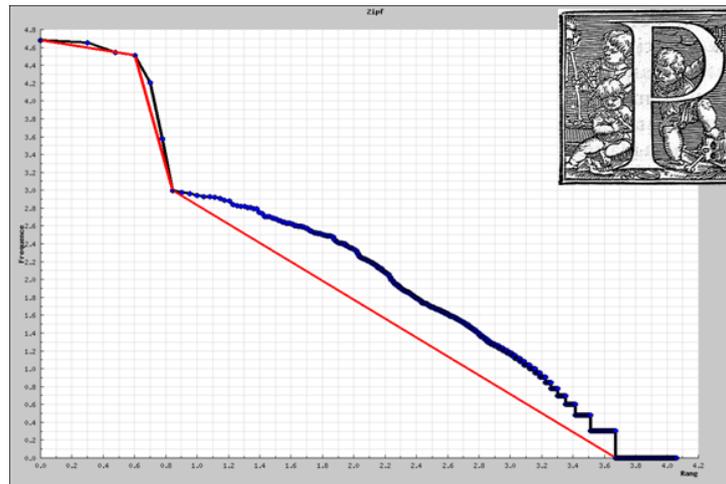


Figure 10.2 – Graphe de Zipf d’une image de lettrine décorative, d’après [Coustaty *et al.*, 2011]. La courbe bleue correspond à l’évolution des fréquences d’apparition des différents motifs de taille 3×3 . La courbe rouge correspond à la loi de Zipf associée, composée de trois parties linéaires distinctes.

sur les fréquences d’apparition triées des différents motifs de niveaux de gris de l’image. Nous pouvons observer que ces fréquences d’apparition semblent être raisonnablement modélisées par trois segments linéaires distincts de la loi de Zipf.

Selon cette modélisation, la segmentation est effectuée en fonction de ces trois parties qui sont détectées pour chaque image. La Figure 10.3 présente les décompositions obtenues par cette méthode pour différentes images de lettrines. Nous obtenons ainsi trois couches structurales pour chaque image, correspondant aux trois parties de la courbe de Zipf associée aux fréquences d’apparition des différents motifs. Le contenu de ces différentes couches peut être interprété de la manière suivante :

- La première couche, représentée en noir, est celle issue des motifs les plus fréquents dans l’image. Elle correspond généralement aux régions « plates » avec des valeurs de pixels relativement homogènes sur les masques 3×3 considérés ;
- La seconde couche, représentée en rouge, correspond au milieu de la courbe de Zipf, et donc aux motifs moyennement fréquents. On y retrouve généralement des zones de contours, avec des transitions du foncé vers le clair (ou inversement) ;
- La troisième couche, représentée en vert, est associée aux motifs les moins fréquents. Elle est généralement composée des détails de l’image.

Ces trois couches, que nous nommons respectivement la couche *homogène* (H), les *contours* (O pour *outline*) et les *détails* (D), constituent la segmentation de l’image. Dans la suite, elles seront utilisées pour caractériser leurs organisations spatiales au sein de fenêtre locales, dans le but de distinguer les différents types de fonds, en faisant alors abstraction de la lettre. Cette décomposition permet, en effet, de conserver certains détails caractéristiques comme les hachures ou les petits points des fonds criblés.

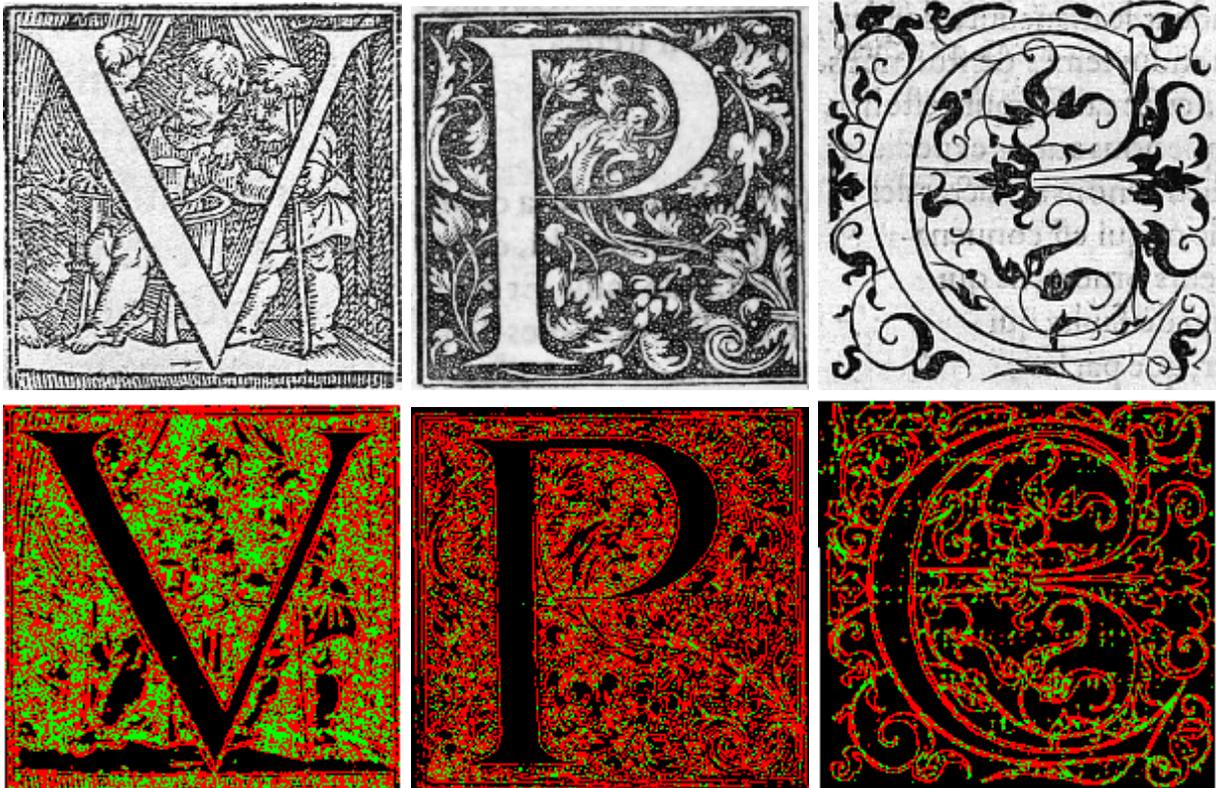


Figure 10.3 – Exemples de décompositions de Zipf pour des images de lettrines décoratives. La première ligne montre des lettrines avec différents styles de fonds (de gauche à droite : fonds *hachurés*, *criblés* et *décoratifs*). La seconde ligne montre les décompositions de Zipf correspondantes (en noir : zones *homogènes*, en rouge : *contours*, en vert : *détails*).

10.2.2 Descripteurs locaux d'enlacement

Afin de capturer plus efficacement les structures spatiales apparaissant entre les différentes couches des images à une échelle fine (et non pas entre des objets sémantiques définis de manière globale), nous proposons de calculer les descripteurs directionnels d'enlacement sur des fenêtres locales au sein des images.

Dans un premier temps, les images sont segmentées à partir de la décomposition de Zipf présentée ci-avant. Pour chaque image, nous obtenons trois couches binaires correspondant respectivement aux régions homogènes (couche H), aux contours (couche O) et aux détails (couche D). Nous considérons alors tous les couples de couches possibles, dénotés par HO, HD et OD. Chacune des images de ces couples est subdivisée en fenêtres locales carrées de tailles 32×32 pixels, sans chevauchement des fenêtres. Par exemple, pour une image de lettrine de taille 200×200 pixels, nous obtenons alors un total de $36 \times 3 = 108$ fenêtres locales (36 fenêtres pour chaque couple HO, HD et OD).

Pour chacune de ces fenêtres, nous calculons les deux descripteurs d'enlacement \mathcal{E}_{AB} et

10.2. Approche par sacs d'enlacement local

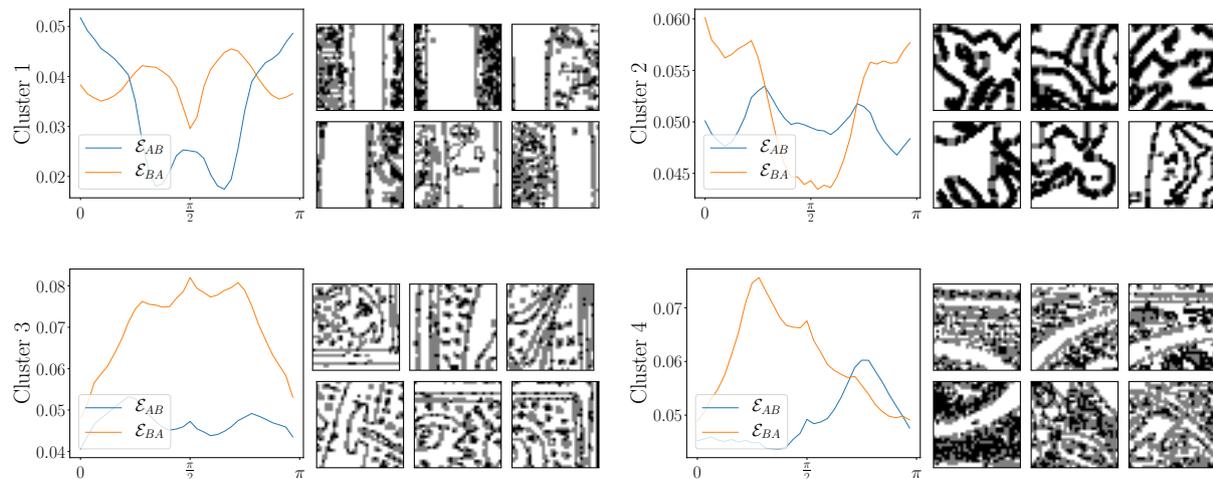


Figure 10.4 – Exemples de mots d’enlacement obtenus pour des images de lettrines décoratives. Ces mots d’enlacement correspondent à des fenêtres locales pour les couches HO (régions homogènes et contours). Pour chacun de ces quatre *clusters* illustratifs, le centroïde est affiché, avec à côté quelques exemples de fenêtres locales rattachées à ce centroïde. Les courbes \mathcal{E}_{AB} et \mathcal{E}_{BA} sont affichées ici séparément, mais sont concaténées au sein d’un même vecteur lors du *clustering*.

\mathcal{E}_{BA} entre les objets binaires du couple considéré (par exemple pour une fenêtre locale au sein de la couche HO, A et B correspondent aux pixels des régions homogènes et des contours, respectivement). Comme nous avons pu le voir dans la [Partie II](#), ces deux descripteurs sont complémentaires pour décrire pleinement la configuration spatiale. Ils sont ici concaténés au sein d’un même vecteur de caractéristiques.

10.2.3 Apprentissage par sacs d’enlacement

Nous proposons ensuite d’appliquer la stratégie d’apprentissage par sacs de relations spatiales, d’une manière adaptée aux descripteurs locaux d’enlacement. Nous cherchons donc à établir des vocabulaires de configurations complexes d’enlacement entre les différentes couches des images de lettrines, à une échelle locale.

Pour construire ces vocabulaires, l’algorithme *K-Means* est utilisé pour regrouper les descripteurs locaux d’enlacement au sein de *clusters* similaires. Nous construisons trois vocabulaires distincts : un pour chaque couple de couches HO, HD et OD. Si chaque vocabulaire est composé de K mots, nous obtenons donc $3 \times K$ *mots d’enlacement* qui correspondent à des configurations spatiales caractéristiques entre les différentes couches de la décomposition de Zipf.

À des fins de visualisation, la [Figure 10.4](#) présente quelques exemples de mots d’enlacement obtenus suite au *clustering* de descripteurs issus de fenêtres locales des couches HO (régions homogènes et contours). Nous pouvons notamment observer les différents motifs distinctifs des fonds de lettrines qui sont capturés par ces exemples de mots d’enlacement.

À la suite de cette procédure de construction de vocabulaires d'enlacement, nous appliquons les étapes d'encodage et d'agrégation pour représenter chaque image par sa composition de mots d'enlacement. Pour ces travaux, nous avons choisi d'utiliser une stratégie d'encodage lissé par un noyau gaussien, couplée à une agrégation par moyenne.

Soit un vocabulaire d'enlacement $\mathcal{V} = \{\mathbf{w}_1, \dots, \mathbf{w}_K\}$ et soit une image caractérisée par un ensemble de descripteurs locaux d'enlacement $\{\mathbf{e}_i\}_{i=1}^n$, nous rappelons que l'encodage lissé suivi de l'agrégation pour un mot \mathbf{w}_k est donné par :

$$z_k = \frac{1}{n} \sum_{i=1}^n K_\sigma(\mathbf{e}_i, \mathbf{w}_k) \quad (10.2)$$

où K_σ désigne le noyau gaussien :

$$K_\sigma(\mathbf{x}_1, \mathbf{x}_2) = \exp -\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{2\sigma^2}. \quad (10.3)$$

Le paramètre σ du noyau gaussien permet de contrôler le niveau de lissage lors de l'encodage des descripteurs locaux d'enlacement. Finalement, l'image est représentée par le vecteur de caractéristiques $\mathbf{z} = [z_1, \dots, z_K]$ correspondant à l'agrégation pour tous les mots du vocabulaire. Nous désignons alors cette approche par le terme *BoE*, pour *bags of enlacement*. Cette procédure est suivie indépendamment pour les trois vocabulaires correspondant aux couples de couches HO, HD et OD. Nous obtenons donc pour chaque image trois vecteurs de caractéristiques, qui peuvent être considérés indépendamment, ou bien combinés au sein d'un même vecteur avant de procéder à la classification des images. L'influence de la prise en compte de ces différentes couches sera notamment étudiée expérimentalement dans la section suivante.

10.3 Validations expérimentales

Dans cette section, nous présentons les expérimentations quantitatives menées pour valider l'intérêt de la méthodologie d'apprentissage par descripteurs locaux d'enlacement, dans le but de reconnaître les styles de fonds d'images de lettrines décoratives.

10.3.1 Jeu de données

Le jeu de données employé pour ces expérimentations est une extension de celui utilisé dans le [Chapitre 5](#) de ce manuscrit. Cette extension est composée au total de 636 images représentant chacune une lettrine décorative. Ces images de lettrines ont été classifiées par des historiens en fonction de leurs différents styles de fonds. On retrouve alors trois catégories de styles de fonds : les fonds *hachurés* (253 images), les fonds *criblés* (214 images) et les fonds *décoratifs* (169 images). Ces différents types de fonds ont notamment été illustrés précédemment dans la [Figure 10.1](#), ainsi que la [Figure 10.3](#) pour les résultats de segmentation. En particulier, la classe de fonds *décoratifs* est composée de motifs variés, avec notamment plusieurs images dites inversées (c'est-à-dire avec une lettre et des motifs sombres sur un fond clair, et inversement).

10.3. Validations expérimentales

Afin de maintenir une certaine homogénéité au sein de ce jeu de données, toutes les images ont été sous-échantillonnées selon une taille maximum de 200×200 pixels.

10.3.2 Protocole de validation

Dans un premier temps, les images du jeu de données sont segmentées à partir de la stratégie de décomposition de Zipf présentée ci-avant. Pour une image donnée, nous obtenons donc trois couches de décomposition (H, O et D), et les trois couples de couches HO, HD et OD sont alors considérés. Chaque image d'un couple est divisée en fenêtres locales de taille 32×32 pixels. Ensuite, nous calculons les descripteurs locaux d'enlacement pour chacune de ces fenêtres. Les descripteurs d'enlacement sont calculés sur un ensemble de 32 directions discrètes, uniformément réparties le long de l'intervalle $[0, \pi]$. La concaténation des descripteurs \mathcal{E}_{AB} et \mathcal{E}_{BA} se traduit alors par des descripteurs de taille 64. À la suite de ces phases de segmentation et de description des images, nous obtenons un total de 21503 descripteurs locaux d'enlacement sur l'ensemble du jeu de données.

Le jeu de données est alors divisé en deux parties : un ensemble d'apprentissage qui correspond à 25% de la base d'images (159 images), tandis que les 75% restants sont utilisés comme ensemble de test (477 images). L'apprentissage par sacs d'enlacement (BoE) est alors appliqué sur les images de l'ensemble d'apprentissage pour les trois combinaisons possibles de couches HO, HD et OD, et pour différentes tailles de vocabulaires (c'est-à-dire le nombre de *clusters* pour l'algorithme *K-Means*). Pour l'encodage lissé (Équation 10.2 et Équation 10.3), nous avons utilisé un noyau gaussien $K_{\tilde{\sigma}}$ où la valeur de $\tilde{\sigma}$ a été fixée de manière adaptative à la moyenne des distances euclidiennes entre les descripteurs locaux d'enlacement et les centroïdes des *clusters* formés, de la même façon que pour les expérimentations du Chapitre 9.

Les vecteurs de caractéristiques obtenus suite à l'encodage et l'agrégation par moyenne sont alors utilisés comme données d'entrées de machines à vecteurs de support (SVM) afin de classifier les images de lettrines en fonction de leurs styles de fonds. Dans ces expérimentations, nous avons utilisé les classifieurs SVM avec un noyau linéaire, et selon une stratégie *one-versus-all* pour la classification multiclassées. L'hyperparamètre de marge souple est optimisé par validation croisée sur la base d'apprentissage, selon un découpage de type *5-fold*. Nous avons également effectué des expérimentations avec d'autres noyaux pour les SVM (noyau gaussien et χ^2), mais les temps de calcul étaient particulièrement augmentés, et nous n'avons pas observé d'impact significatif sur les résultats de classification obtenus.

Par ailleurs, afin de mieux évaluer la robustesse de notre approche, ce protocole de classification est répété dix fois avec différentes bases d'apprentissage choisies aléatoirement. Chaque base d'apprentissage aléatoire générée garantit la même distribution des différentes classes du jeu de données. Nous évaluons alors quantitativement les résultats de classification obtenus à partir du taux de reconnaissance, de la précision et du rappel. Ces scores sont calculés pour chaque itération du protocole, et nous reportons alors les scores moyens obtenus à l'issue des dix itérations.

Tableau 10.1 – Taux de reconnaissance moyens obtenus pour la classification des images de lettrines en fonction de leurs styles de fonds. Les scores sont présentés pour différentes tailles de vocabulaires, et pour toutes les combinaisons de couples de couches issues de la décomposition de Zipf.

Couches / K	100	200	500
DH	79,81 \pm 2,15	80,04 \pm 1,63	80,00 \pm 1,55
OD	80,69 \pm 1,82	80,96 \pm 1,88	81,30 \pm 1,62
OH	80,96 \pm 1,83	81,03 \pm 1,68	81,43 \pm 1,64
DH+OH	81,49 \pm 1,87	81,70 \pm 1,62	81,43 \pm 1,38
OH+OD	82,68 \pm 1,62	82,68 \pm 1,65	82,49 \pm 1,99
DH+OD	84,84 \pm 1,35	83,98 \pm 2,39	84,53 \pm 1,56
DH+OH+OD	85,07 \pm 2,05	84,32 \pm 2,33	85,26 \pm 1,10

10.3.3 Résultats expérimentaux et discussion

Combinaisons de couches et taille des vocabulaires

Le [Tableau 10.1](#) présente les taux de reconnaissance obtenus pour toutes les combinaisons possibles de couches HO, HD et OD, ainsi que pour différentes tailles de vocabulaires de l'approche par sacs d'enlacement. Les taux de reconnaissance rapportés correspondent aux moyennes obtenues suite aux dix itérations avec des bases d'apprentissage aléatoires. Les trois premières lignes correspondent aux couples de couches HO, HD et OD considérés individuellement, chacun d'entre eux pour K mots d'enlacement (avec ici $K \in \{100, 200, 500\}$). Les lignes suivantes désignent alors les combinaisons de ces couples de couches, obtenues par concaténation directe des vecteurs de caractéristiques issus de l'apprentissage par sacs d'enlacement. Notons que ces concaténations augmentent naturellement la dimensionnalité des représentations.

À partir des résultats de ce tableau, nous pouvons dans un premier temps observer la stabilité globale des taux de reconnaissance obtenus au fil des dix itérations du protocole de classification (les valeurs d'écart-type sont relativement faibles). Ceci permet notamment d'attester de la capacité de généralisation de la méthodologie d'apprentissage proposée. Par ailleurs, en étudiant les trois premières lignes, nous pouvons constater une légère augmentation des taux de reconnaissance lorsque la couche O est considérée. Ensuite, les résultats semblent progressivement augmenter au fur et à mesure que l'on combine les différents couples de couches, les meilleurs taux de reconnaissance étant obtenus pour la combinaison des trois couples HO+HD+OD. Ces résultats suggèrent donc que les trois couches d'information sont pertinentes pour caractériser la diversité des structures spatiales apparaissant dans ces images de lettrines. En ce qui concerne la taille des vocabulaires, nous pouvons observer que les taux de reconnaissance semblent rester relativement stables pour les différentes valeurs testées. Cela suggère que pour chaque couple de couches, la plupart de l'information spatiale pertinente est synthétisée dans un nombre re-

10.3. Validations expérimentales

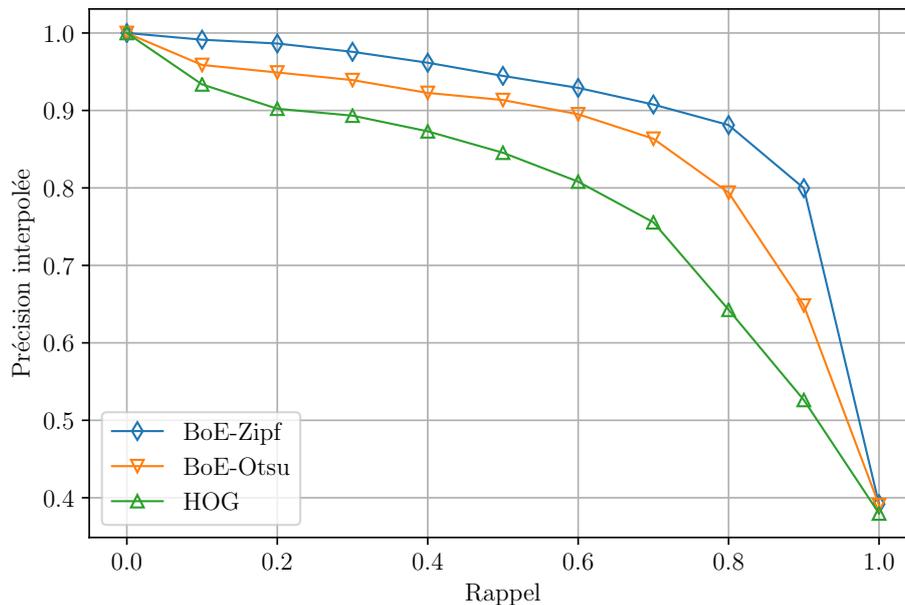


Figure 10.5 – Courbes de précision-rappel obtenues pour la classification d’images de lettres décoratives en fonction de leurs styles de fonds. Les résultats correspondant à l’approche proposée (BoE et Zipf) et aux méthodes comparatives de référence considérées (seuillage Otsu à la place de la décomposition de Zipf, et sacs de caractéristiques HOG).

lativement faible de mots d’enlacement. Nous avons, notamment, constaté un comportement similaire lors des expérimentations du [Chapitre 9](#) à partir des descripteurs FHD.

Résultats comparatifs

La méthodologie proposée pour la reconnaissance de ces images de lettres est composée de deux étapes principales : la décomposition des images par la loi de Zipf, et l’apprentissage par sacs d’enlacement pour la classification. Nous proposons ici des résultats comparatifs pour ces deux étapes de l’approche.

D’une part, pour la décomposition de Zipf, nous comparons les résultats obtenus avec la même approche par sacs d’enlacement, mais appliquée directement sur des binarisations des images de lettres à partir d’un seuillage par la méthode d’Otsu. Cette approche est désignée par BoE-Otsu. D’autre part, pour les sacs d’enlacement, nous comparons les résultats de notre approche avec un apprentissage par sacs de caractéristiques classique (c’est-à-dire le même que celui présenté dans le [Chapitre 9](#)). Cette approche utilise les caractéristiques HOG [Dalal et Triggs, 2005] extraites directement depuis les images de lettres, sans la décomposition de Zipf. Nous avons utilisé les mêmes fenêtres locales de taille 32×32 pixels. Ces caractéristiques locales sont alors agrégées selon la même stratégie que les descripteurs d’enlacement, avec un encodage lissé adaptatif. L’approche est alors désignée par HOG.

Dans ce contexte d’étude comparative, notre approche (décomposition de Zipf et sacs d’enla-

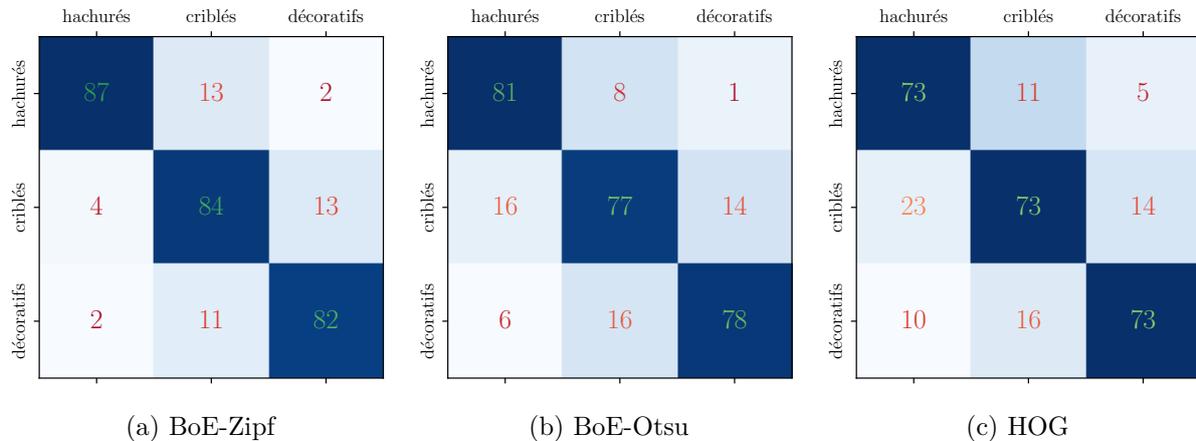


Figure 10.6 – Matrices de confusion moyennes pour les approches proposées et pour les méthodes comparatives de référence.

ement) est désignée par BoE-Zipf. Elle a été appliquée pour des vocabulaires de taille $K = 100$ et pour la combinaison de toutes les couches (HD+HO+OD), totalisant ainsi 300 mots d'enlacement. À des fins de comparabilité, les approches BoE-Otsu et HOG sont elles aussi appliquées avec $K = 300$ mots, et avec le même protocole de classification.

La Figure 10.5 présente les courbes de précision-rappel interpolées obtenues pour l'approche BoE-Zipf et pour les méthodes comparatives BoE-Otsu et HOG. Nous pouvons constater que l'approche par sacs d'enlacement BoE (c'est-à-dire BoE-Otsu ou BoE-Zipf) permet d'obtenir de meilleurs résultats que l'approche par sacs de caractéristiques HOG. Ces résultats semblent donc confirmer l'intérêt de considérer les descripteurs d'enlacement à une échelle locale, à la place des caractéristiques classiques basées sur les orientations de gradients. Nous pouvons également observer que la décomposition de Zipf permet d'obtenir une amélioration notable sur le jeu de données testé, par rapport au simple seuillage des images par la méthode d'Otsu.

Finalement, afin de mieux interpréter la performance relative de ces méthodes comparatives, la Figure 10.6 présente les matrices de confusion moyennes obtenues pour notre approche BoE-Zipf et pour les approches BoE-Otsu et HOG. Ces matrices de confusion permettent de visualiser les scores de classification pour les différentes classes du jeu de données. Nous pouvons notamment constater que notre approche semble bien reconnaître les lettrines avec des fonds *hachurés*. En effet, les descripteurs d'enlacement sont particulièrement adaptés pour caractériser des motifs directionnels qui s'alternent tels que les hachures.

10.4 Bilan scientifique

Dans ce chapitre, nous avons proposé une approche permettant de combiner les descripteurs d'enlacement avec la stratégie d'apprentissage par sacs de relations spatiales. Cette approche a été appliquée au domaine spécifique de l'analyse d'images de lettrines décoratives, où nous nous sommes focalisés particulièrement sur la reconnaissance des différents styles de fonds apparais-

10.4. Bilan scientifique

sant dans ces images. Afin de mieux caractériser la complexité des motifs de ces styles de fonds, les images sont décomposées en différentes couches structurelles, puis les descripteurs d'enlacement ont été calculés à une échelle locale, permettant de capturer une information spatiale proche de la texture. Dans ce contexte applicatif, les résultats expérimentaux obtenus nous ont permis d'illustrer l'intérêt de cette représentation par vocabulaires de configurations spatiales complexes, obtenues à partir de fenêtres locales.

Les travaux présentés dans ce chapitre ont été valorisés par la publication d'un article dans une conférence internationale [Clément *et al.*, 2017d].

Conclusion

Conclusion

Bilan et contributions

Les travaux réalisés durant cette thèse se sont inscrits dans une problématique générale d'analyse et d'interprétation de données visuelles. En particulier, nous nous sommes intéressés à l'exploitation de la richesse de l'information spatiale pour l'élaboration de nouvelles méthodologies de reconnaissance et de classification d'images. Nos travaux se sont alors articulés autour de deux axes de recherche. Nous établissons ici un bilan synthétique des différentes contributions scientifiques apportées pour chacun de ces deux axes étudiés.

Le premier axe de recherche a porté sur la modélisation de configurations spatiales complexes entre des objets bidimensionnels. Nous avons formalisé le concept d'imbrication d'objets en définissant deux nouvelles relations spatiales. D'abord, la relation d'enlacement permet de généraliser la pluralité de situations où un objet est entouré ou capturé par un autre, selon un point de vue directionnel. À partir de cette définition, la relation d'entrelacement désigne alors un enlacement mutuel de deux objets. Nous avons ainsi proposé un modèle algorithmique conforme à ces définitions sémantiques, qui permet de décrire quantitativement ces nouvelles relations spatiales par le biais d'histogrammes directionnels. Ce modèle caractérise de manière graduelle les différents niveaux d'enlacement et d'entrelacement entre des objets, là où les descripteurs de position relative classiques ne permettent généralement pas de démêler les ambiguïtés inhérentes à ces configurations. Les descripteurs proposés sont applicables pour des objets arbitrairement complexes, et présentent des propriétés d'invariance qui sont souvent nécessaires pour divers types d'applications liées à la reconnaissance de formes. Par ailleurs, guidés par le principe de dualité entre descripteurs de position relative et évaluations de relations spatiales, nous avons également proposé une stratégie de visualisation de ces nouvelles relations, sous la forme de paysages flous directionnels d'enlacement. Ces paysages permettent d'évaluer, directement dans l'espace image, différents types de configurations spatiales complexes, comme notamment l'entournement d'objets. Des validations expérimentales sur plusieurs exemples caractéristiques, ainsi que dans différents contextes applicatifs tels que l'imagerie biomédicale, la télédétection et l'analyse de documents, nous ont permis d'illustrer l'intérêt et la généralité de ces approches.

Dans le second axe de recherche, nous nous sommes alors intéressés à la mise en place de méthodologies de description de relations spatiales composites, c'est-à-dire entre plusieurs objets et au sein de plusieurs images. Pour cela, nous avons tout d'abord proposé un descripteur hiérarchique structurel, prenant en compte les formes et les relations spatiales composant une image, à

partir d'objets apparaissant à différents niveaux d'échelle. Ce descripteur repose sur la définition de graphes relationnels attribués entre les régions d'intérêt d'une image, où les attributs associés sont des descripteurs de position relative comme les histogrammes de forces. Ensuite, afin de lever les contraintes de nature combinatoire liées à ces représentations par graphes relationnels, nous avons proposé une stratégie de désassemblage de ce descripteur afin d'obtenir une représentation de la structure d'une image sous la forme d'ensembles non-ordonnés d'attributs. À partir de cette représentation, nous avons établi une stratégie d'apprentissage par sacs de caractéristiques, alors nommée sacs de relations spatiales et de formes. Cette stratégie permet de bâtir des vocabulaires structurels, composés de prototypes de configurations spatiales qui émergent au sein des images d'une base d'apprentissage. Les validations expérimentales conduites sur plusieurs jeux de données ont permis d'attester du potentiel descriptif de l'information structurelle spatiale pour des tâches de classification d'images. Par ailleurs, ces expérimentations ont également montré que ces représentations structurelles peuvent être combinées efficacement avec des approches par sacs de caractéristiques reposant sur des descripteurs locaux. Elles ont ainsi illustré la complémentarité entre les informations globales et locales.

Afin d'appréhender l'aspect transverse des deux axes de recherche qui ont guidé nos travaux, nous avons présenté une approche visant à relier ces derniers, que nous avons alors appliquée à la reconnaissance d'images de lettrines décoratives. En effet, cette approche permet d'intégrer les descripteurs d'enlacement au sein de vocabulaires de configurations spatiales complexes, donnant ainsi lieu à des représentations par sacs d'enlacement. De plus, ces configurations sont caractérisées à une échelle locale, c'est-à-dire au sein de fenêtres de petites tailles, permettant ainsi de capturer une information spatiale d'une autre nature, plus proche de la texture.

En définitive, à travers l'ensemble des approches proposées dans cette thèse, nous avons cherché à apporter des représentations d'images originales, en lien avec la perception humaine, et en s'attachant tout particulièrement à la caractérisation des relations spatiales d'un point de vue structurel, dans le but d'en extraire leur richesse sémantique sous-jacente.

Perspectives de recherche

Les différents travaux de recherche présentés dans cette thèse ont, certes, permis de lever certains verrous concernant l'exploitation des relations spatiales pour l'analyse d'images, mais ils ont également ouvert la voie à de nombreuses pistes de réflexion. Nous distinguons ci-après un éventail de perspectives, qui pourront être envisagées à plus ou moins long terme, pour les deux principaux axes de recherche explorés.

Dans la continuité des travaux entrepris sur l'enlacement et l'entrelacement entre objets, nous envisageons plusieurs extensions du modèle proposé. Un premier aspect serait de rendre possible le calcul des descripteurs d'enlacement pour des objets qui se chevauchent (c'est-à-dire dans le cas où les objets ont une intersection non nulle), afin de pouvoir caractériser un panel encore plus large de configurations spatiales complexes. Pour ce faire, une solution serait de décomposer le calcul le long de chaque droite orientée, et de définir une valeur d'enlacement qui serait maximale pour les portions de segments qui se chevauchent. Selon le même objec-

Conclusion

tif, des expérimentations sur le calcul des descripteurs sur des objets définis par des fonctions d'appartenance floues permettraient d'étendre d'autant plus le champ d'application du modèle. Nous envisageons également d'étudier l'application du descripteur d'enlacement entre un objet et son complémentaire, permettant ainsi d'obtenir une mesure directionnelle de ses concavités, notamment pour des tâches de description de formes. Par ailleurs, une extension de notre modèle algorithmique à la manipulation d'objets tridimensionnels pourra également être envisagée. Finalement, à travers une réflexion plus globale, il nous semblerait intéressant de réfléchir à une formalisation générique de nouvelles familles de relations spatiales complexes, en intégrant par exemple des mesures d'espacement entre les objets, et potentiellement d'autres types de relations qui seraient à catégoriser.

Du point de vue de l'apprentissage de relations spatiales composites, de nombreuses perspectives peuvent également être considérées. De manière générale, il pourrait être bénéfique d'interpréter de façon plus approfondie la sémantique associée aux vocabulaires structurels de configurations spatiales. Il s'agirait alors de générer des descriptions, en langage naturel, des relations composites contenues dans les vocabulaires, et d'explorer la possibilité d'organiser ces relations sous la forme de familles, par exemple selon une structure hiérarchique de type fine à grossière. Ceci permettrait de contrôler la granularité de représentation des vocabulaires, en fonction du contexte applicatif considéré. Des travaux préliminaires dans cette direction ont été entrepris dans le cadre de projets encadrés de recherche avec des étudiants de Master 1, et mériteraient d'être approfondis. Un autre champ exploitable digne d'intérêt consisterait à intégrer d'autres descripteurs de position relative dans cette approche, dans la continuité de ce qui a été entamé avec les vocabulaires d'enlacement pour la reconnaissance des styles de lettres. En particulier, il pourrait s'avérer intéressant d'exploiter les capacités de description du Φ -descripteur proposé récemment par [Matsakis *et al.*, 2015], qui permet de caractériser une large variété de configurations spatiales.

Enfin, nous considérons que l'intégration de connaissances sémantiques, par exemple dans la continuité des travaux proposés par [Hudelot *et al.*, 2008] sur la définition d'ontologies de relations spatiales, pourraient permettre d'améliorer les capacités d'interprétation des images. L'objectif consisterait alors en la mise en place de systèmes d'indexation et de recherche d'informations visuelles de haut niveau, à partir desquels des requêtes complexes en langage naturel pourraient être formulées. En ce sens, dans le cadre de perspectives plus larges, il serait particulièrement intéressant d'explorer comment de tels mécanismes d'interprétation de la structure spatiale des images pourraient être intégrés à des approches par apprentissage profond, qui travaillent notamment à l'échelle des régions [Girshick *et al.*, 2016]. De telles perspectives pourraient permettre d'injecter des connaissances spatiales de haut niveau sémantique au sein de processus de segmentation et de classification à partir de grands volumes de données.

Bibliographie

- [Abràmoff *et al.*, 2010] Abràmoff, M. D., Reinhardt, J. M., Russell, S. R., Folk, J. C., Mahajan, V. B., Niemeijer, M. et Quellec, G. Automated early detection of diabetic retinopathy. *Ophthalmology*, 117(6):1147–1154, 2010 (cité en page 64).
- [Allen, 1983] Allen, J. F. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983 (cité en pages 15, 17, 25).
- [Arbelaez *et al.*, 2011] Arbelaez, P., Maire, M., Fowlkes, C. et Malik, J. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011 (cité en page 109).
- [Baatz et Schape, 2000] Baatz, M. et Schape, A. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. In *Angewandte Geographische Informationsverarbeitung XII*, pp. 12–23, 2000 (cité en page 70).
- [Bay *et al.*, 2008] Bay, H., Ess, A., Tuytelaars, T. et Van Gool, L. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008 (cité en page 114).
- [Bloch, 1999] Bloch, I. Fuzzy relative position between objects in image processing: A morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):657–664, 1999 (cité en pages 27, 37, 75, 80, 88).
- [Bloch, 2005] Bloch, I. Fuzzy spatial relationships for image processing and interpretation: A review. *Image and Vision Computing*, 23(2):89–110, 2005 (cité en page 22).
- [Bloch *et al.*, 2006] Bloch, I., Colliot, O. et Cesar, R. M. On the ternary spatial relation "between". *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(2):312–327, 2006 (cité en pages 33, 34, 36).
- [Bloch et Ralescu, 2003] Bloch, I. et Ralescu, A. Directional relative position between objects in image processing: A comparison between fuzzy approaches. *Pattern Recognition*, 36(7):1563–1582, 2003 (cité en pages 22, 28).
- [Breiman, 2001] Breiman, L. Random forests. *Machine learning*, 45(1):5–32, 2001 (cité en page 114).

- [Bresenham, 1965] Bresenham, J. E. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30, 1965 (cité en page 55).
- [Buck *et al.*, 2013] Buck, A. R., Keller, J. M. et Skubic, M. A memetic algorithm for matching spatial configurations with the histograms of forces. *IEEE Transactions on Evolutionary Computation*, 17(4):588–604, 2013 (cité en page 24).
- [Cesar *et al.*, 2005] Cesar, R. M., Bengoetxea, E., Bloch, I. et Larrañaga, P. Inexact graph matching for model-based recognition: Evaluation and comparison of optimization algorithms. *Pattern Recognition*, 38(11):2099–2113, 2005 (cité en page 37).
- [Cha et Srihari, 2002] Cha, S.-H. et Srihari, S. N. On measuring the distance between histograms. *Pattern Recognition*, 35(6):1355–1370, 2002 (cité en page 103).
- [Chang et Lee, 1995] Chang, C.-C. et Lee, C.-F. Relative coordinates oriented symbolic string for spatial relationship retrieval. *Pattern Recognition*, 28(4):563–570, 1995 (cité en page 18).
- [Chang *et al.*, 1987] Chang, S.-K., Shi, Q.-Y. et Yan, C.-W. Iconic indexing by 2-D strings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(3):413–428, 1987 (cité en page 17).
- [Chatfield *et al.*, 2011] Chatfield, K., Lempitsky, V., Vedaldi, A. et Zisserman, A. The devil is in the details: An evaluation of recent feature encoding methods. In *British Machine Vision Conference (BMVC)*, 2011 (cité en page 119).
- [Chen *et al.*, 2015] Chen, J., Cohn, A. G., Liu, D., Wang, S., Ouyang, J. et Yu, Q. A survey of qualitative spatial representations. *The Knowledge Engineering Review*, 30(1):106–136, 2015 (cité en page 15).
- [Cohn *et al.*, 1997] Cohn, A. F., Bennett, B., Gooday, J. et Gotts, N. M. Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 1(3):275–316, 1997 (cité en pages 16, 39).
- [Cohn et Hazarika, 2001] Cohn, A. G. et Hazarika, S. M. Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 46(1-2):1–29, 2001 (cité en page 14).
- [Colliot *et al.*, 2006] Colliot, O., Camara, O. et Bloch, I. Integration of fuzzy spatial relations in deformable models - Application to brain MRI segmentation. *Pattern recognition*, 39(8):1401–1414, 2006 (cité en page 28).
- [Comaniciu et Meer, 2002] Comaniciu, D. et Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002 (cité en page 106).
- [Cortes et Vapnik, 1995] Cortes, C. et Vapnik, V. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995 (cité en page 114).

Bibliographie

- [Coustaty, 2012] Coustaty, M. *Contribution à l'analyse complexe de documents anciens, application aux letrines*. Thèse de doctorat. Université de La Rochelle (La Rochelle, France) (cité en page 140).
- [Coustaty *et al.*, 2011] Coustaty, M., Pareti, R., Vincent, N. et Ogier, J.-M. Towards historical document indexing: Extraction of drop cap letters. *International Journal on Document Analysis and Recognition*, 14(3):243–254, 2011 (cité en pages 66, 139–141).
- [Dalal et Triggs, 2005] Dalal, N. et Triggs, B. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886–893, 2005 (cité en pages 114, 126, 147).
- [Debled-Rennesson et Wendling, 2015] Debled-Rennesson, I. et Wendling, L. Extraction of successive patterns in document images by a new concept based on force histogram and thick discrete lines. In *International Conference on Image Analysis and Processing (ICIAP)*, vol. 1, pp. 387–397, 2015 (cité en page 24).
- [Dehak, 2002] Dehak, S. M. R. *Inférence quantitative des relations spatiales directionnelles*. Thèse de Doctorat. École Nationale Supérieure des Télécommunications (Paris, France) (cité en page 14).
- [Delaye, 2011] Delaye, A. *Méta-modèles de positionnement spatial pour la reconnaissance de tracés manuscrits*. Thèse de Doctorat. INSA Rennes (Rennes, France) (cité en pages 28, 29, 37, 38).
- [Delaye et Anquetil, 2014] Delaye, A. et Anquetil, É. Learning of fuzzy spatial relations between handwritten patterns. *International Journal on Data Mining, Modelling and Management*, 6(2):127–147, 2014 (cité en pages 39, 40).
- [Dubois et Jaulent, 1987] Dubois, D. et Jaulent, M.-C. A general approach to parameter evaluation in fuzzy digital pictures. *Pattern Recognition Letters*, 6(4):251–259, 1987 (cité en pages 19, 56, 88).
- [Dubois *et al.*, 1988] Dubois, D., Prade, H. et Testemale, C. Weighted fuzzy pattern matching. *Fuzzy Sets and Systems*, 28(3):313–331, 1988 (cité en page 76).
- [Dutta, 1991] Dutta, S. Approximate spatial reasoning: Integrating qualitative and quantitative constraints. *International Journal of Approximate Reasoning*, 5(3):307–330, 1991 (cité en page 15).
- [Egenhofer et Franzosa, 1991] Egenhofer, M. J. et Franzosa, R. D. Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2):161–174, 1991 (cité en page 15).

- [Egenhofer et Herring, 1990] Egenhofer, M. J. et Herring, J. R. *Categorizing binary topological relations between regions, lines, and points in geographic databases*. Rapp. tech. (cité en pages 15, 25).
- [Everingham *et al.*, 2010] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. et Zisserman, A. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338, 2010 (cité en pages 4, 114).
- [Fan *et al.*, 2012] Fan, B., Wu, F. et Hu, Z. Rotationally invariant descriptors using intensity order pooling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10):2031–2045, 2012 (cité en page 139).
- [Felzenszwalb *et al.*, 2010] Felzenszwalb, P. F., Girshick, R. B., McAllester, D. et Ramanan, D. Object Detection with Discriminatively Trained Part Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010 (cité en page 114).
- [Forestier *et al.*, 2012] Forestier, G., Puissant, A., Wemmert, C. et Gançarski, P. Knowledge-based region labeling for remote sensing image interpretation. *Computers, Environment and Urban Systems*, 36(5):470–480, 2012 (cité en page 15).
- [Fraz *et al.*, 2012] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G. et Barman, S. A. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9):2538–2548, 2012 (cité en page 61).
- [Freeman, 1975] Freeman, J. The modelling of spatial relations. *Computer Graphics and Image Processing*, 4(2):156–171, 1975 (cité en pages 11, 12, 17).
- [Freksa, 1991] Freksa, C. « Qualitative spatial reasoning ». In *Cognitive and linguistic aspects of geographic space*, Springer Netherlands, pp. 361–372 (cité en page 14).
- [Garnier, 2014] Garnier, M. *Modèles descriptifs de relations spatiales pour l'aide au diagnostic d'images biomédicales*. Thèse de Doctorat. Université Paris Descartes (Paris, France) (cité en page 97).
- [Garnier *et al.*, 2012] Garnier, M., Hurtut, T. et Wendling, L. Object description based on spatial relations between level-sets. In *International Conference on Digital Image Computing Techniques and Applications (DICTA)*, pp. 1–7, 2012 (cité en pages 37, 103).
- [Gemert *et al.*, 2010] Gemert, J. C. V., Veenman, C. J., Smeulders, A. W. M. et Geusebroek, J. M. Visual word ambiguity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1271–1283, 2010 (cité en pages 119, 122).
- [Girshick *et al.*, 2016] Girshick, R., Donahue, J., Darrell, T. et Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):142–158, 2016 (cité en page 155).

Bibliographie

- [Gonzalez et Woods, 2002] Gonzalez, R. C. et Woods, R. E. *Digital image processing*. Prentice Hall (cité en page 106).
- [Hoàng *et al.*, 2010] Hoàng, N. V., Gouet-Brunet, V., Rukoz, M. et Manouvrier, M. Embedding spatial information into image content description for scene retrieval. *Pattern Recognition*, 43(9):3013–3024, 2010 (cité en page 115).
- [Hudelot *et al.*, 2008] Hudelot, C., Atif, J. et Bloch, I. Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, 159(15):1929–1951, 2008 (cité en page 155).
- [Hunter, 2007] Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, 9(3):90–95, 2007 (cité en page 178).
- [Inglada et Michel, 2009] Inglada, J. et Michel, J. Qualitative spatial reasoning for high-resolution remote sensing image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2):599–612, 2009 (cité en page 16).
- [Jégou *et al.*, 2012] Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Pérez, P. et Schmid, C. Aggregating local image descriptors into compact codes. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 34(9):1704–1716, 2012 (cité en page 119).
- [Jiu *et al.*, 2012] Jiu, M., Wolf, C., Garcia, C. et Baskurt, A. Supervised Learning and Codebook Optimization for Bag-of-Words Models. *Cognitive Computation*, 4(4):409–419, 2012 (cité en page 119).
- [Jouili *et al.*, 2010] Jouili, S., Coustaty, M., Tabbone, S. et Ogier, J.-M. NAVIDOMASS: Structural-based Approaches Towards Handling Historical Documents. In *International Conference on Pattern Recognition (ICPR)*, pp. 946–949, 2010 (cité en pages 66, 138).
- [Kóczy, 1988] Kóczy, L. T. On the description of relative position of fuzzy patterns. *Pattern Recognition Letters*, 8(1):21–28, 1988 (cité en page 17).
- [Krishnapuram *et al.*, 1993] Krishnapuram, R., Keller, J. M. et Ma, Y. Quantitative analysis of properties and spatial relations of fuzzy image regions. *IEEE Transactions on Fuzzy Systems*, 1(3):222–233, 1993 (cité en pages 18, 19, 56, 76).
- [Kuipers, 1978] Kuipers, B. Modeling spatial knowledge. *Cognitive science*, 2(2):129–153, 1978 (cité en page 12).
- [Kurtz *et al.*, 2012] Kurtz, C., Passat, N., Gançarski, P. et Puissant, A. Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology. *Pattern Recognition*, 45(2):685–706, 2012 (cité en page 69).
- [Landau et Jackendoff, 1993] Landau, B. et Jackendoff, R. Whence and whither in spatial language and spatial cognition? *Behavioral and Brain Sciences*, 16(2):217–238, 1993 (cité en page 12).

- [Lazebnik *et al.*, 2006] Lazebnik, S., Schmid, C. et Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 2169–2178, 2006 (cité en page 115).
- [LeCun *et al.*, 2015] LeCun, Y., Bengio, Y. et Hinton, G. Deep learning. *Nature*, 521(7553):436–444, 2015 (cité en page 4).
- [Lee et Hsu, 1992] Lee, S.-Y. et Hsu, F.-J. Spatial reasoning and similarity retrieval of images using 2D C-string knowledge representation. *Pattern Recognition*, 25(3):305–318, 1992 (cité en page 18).
- [Logan et Sadler, 1996] Logan, G. D. et Sadler, D. D. A computational analysis of the apprehension of spatial relations. In *Language and Space*, MIT Press (cité en page 26).
- [Loménie et Racoceanu, 2012] Loménie, N. et Racoceanu, D. Point set morphological filtering and semantic spatial configuration modeling: Application to microscopic image and biostructure analysis. *Pattern Recognition*, 45(8):2894–2911, 2012 (cité en page 34).
- [Lowe, 2004] Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004 (cité en pages 114, 126).
- [MacQueen, 1967] MacQueen, J. B. Some methods for classification and analysis of multivariate observations. In *Berkeley Symposium on Mathematical Statistics and Probability (BSMSP)*, pp. 281–297, 1967 (cité en pages 107, 114, 117).
- [Matas *et al.*, 2004] Matas, J., Chum, O., Urban, M. et Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004 (cité en page 114).
- [Matsakis, 1998] Matsakis, P. *Relations spatiales structurelles et interprétation d’images*. Thèse de Doctorat. Université Paul Sabatier de Toulouse (Toulouse, France) (cité en page 100).
- [Matsakis, 2016] Matsakis, P. Affine properties of the relative position PHI-descriptor. In *International Conference on Pattern Recognition (ICPR)*, pp. 1941–1946, 2016 (cité en page 26).
- [Matsakis et Andréfouët, 2002] Matsakis, P. et Andréfouët, S. The fuzzy line between among and surround. In *IEEE International Conference on Fuzzy Systems (FUZ-IEEE)*, vol. 2, pp. 1596–1601, 2002 (cité en pages 33, 87, 90).
- [Matsakis *et al.*, 2004] Matsakis, P., Keller, J. M., Sjahputera, O. et Marjamaa, J. The use of force histograms for affine-invariant relative position description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):1–18, 2004 (cité en page 24).
- [Matsakis *et al.*, 2001] Matsakis, P., Keller, J. M., Wendling, L., Marjamaa, J. et Sjahputera, O. Linguistic description of relative positions in images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 31(4):573–588, 2001 (cité en page 24).

Bibliographie

- [Matsakis et Naeem, 2016] Matsakis, P. et Naeem, M. Fuzzy models of topological relationships based on the PHI-descriptor. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1096–1104, 2016 (cité en page 26).
- [Matsakis *et al.*, 2015] Matsakis, P., Naeem, M. et Rahbarnia, F. Introducing the Φ -descriptor - A most versatile relative position descriptor. In *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, pp. 87–98, 2015 (cité en pages 25, 155).
- [Matsakis *et al.*, 2006] Matsakis, P., Ni, J. et Wang, X. Object localization based on directional information: Case of 2D raster data. In *International Conference on Pattern Recognition (ICPR)*, pp. 142–146, 2006 (cité en page 28).
- [Matsakis et Nikitenko, 2005] Matsakis, P. et Nikitenko, D. Combined extraction of directional and topological relationship information from 2D concave objects. In *Fuzzy Modeling with Spatial Information for Geographic Problems*, Springer Berlin Heidelberg, pp. 15–40 (cité en page 25).
- [Matsakis et Wendling, 1999] Matsakis, P. et Wendling, L. A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):634–643, 1999 (cité en pages 23, 37, 47, 56, 59, 98).
- [Matsakis *et al.*, 2010] Matsakis, P., Wendling, L. et Ni, J. A general approach to the fuzzy modeling of spatial relationships. In *Methods for Handling Imperfect Spatial Information*, Springer, pp. 49–74 (cité en pages 21, 28).
- [Merveille *et al.*, 2017] Merveille, O., Talbot, H., Najman, L. et Passat, N. Curvilinear structure analysis by ranking the orientation responses of path operators. *IEEE Transactions in Pattern Analysis and Machine Intelligence*. Sous presse, 2017 (cité en page 62).
- [Mikolajczyk et Schmid, 2005] Mikolajczyk, K. et Schmid, C. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005 (cité en page 114).
- [Miyajima et Ralescu, 1994] Miyajima, K. et Ralescu, A. Spatial organization in 2D segmented images: Representation and recognition of primitive spatial relations. *Fuzzy Sets and Systems*, 65(2):225–236, 1994 (cité en pages 19, 23).
- [Naeem et Matsakis, 2015] Naeem, M. et Matsakis, P. Relative position descriptors - A review. In *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, pp. 286–295, 2015 (cité en page 22).
- [Naegel et Wendling, 2009] Naegel, B. et Wendling, L. Combining shape descriptors and component-tree for recognition of ancient graphical drop caps. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 2, pp. 297–302, 2009 (cité en page 139).

-
- [Naegel et Wendling, 2010] Naegel, B. et Wendling, L. A document binarization method based on connected operators. *Pattern Recognition Letters*, 31(11):1251–1259, 2010 (cité en page 139).
- [Ni et Matsakis, 2010] Ni, J. et Matsakis, P. An equivalent definition of the histogram of forces: Theoretical and algorithmic implications. *Pattern Recognition*, 43(4):1607–1617, 2010 (cité en pages 24, 100).
- [Ni et al., 2004] Ni, J., Matsakis, P. et Wawrzyniak, L. Quantitative representation of the relative position between 3D objects. In *IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP)*, pp. 452–289, 2004 (cité en page 100).
- [Niebles et Fei-Fei, 2007] Niebles, J. C. et Fei-Fei, L. A hierarchical model of shape and appearance for human action classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007 (cité en page 115).
- [Odstrcilik et al., 2013] Odstrcilik, J., Kolar, R., Budai, A., Hornegger, J., Jan, J., Gazarek, J., Kubena, T., Cernosek, P., Svoboda, O. et Angelopoulou, E. Retinal vessel segmentation by improved matched filtering: Evaluation on a new high-resolution fundus image database. *IET Image Processing*, 7(4):373–383, 2013 (cité en page 62).
- [Ojala et al., 1996] Ojala, T., Pietikäinen, M. et Harwood, D. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, 1996 (cité en page 114).
- [Ok et al., 2013] Ok, A. O., Senaras, C. et Yuksel, B. Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3):1701–1717, 2013 (cité en page 28).
- [Orlando et al., 2017] Orlando, J., Prokofyeva, E. et Blaschko, M. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Transactions on Biomedical Engineering*, 64(1):16–27, 2017 (cité en pages 61–63).
- [Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. et Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011 (cité en page 178).
- [Penatti et al., 2014] Penatti, O. A. B., Silva, F. B., Valle, E., Gouet-Brunet, V. et Torres, R. d. S. Visual word spatial arrangement for image retrieval and classification. *Pattern Recognition*, 47(2):705–720, 2014 (cité en page 115).
- [Pérez et Granger, 2007] Pérez, F. et Granger, B. E. IPython: A system for interactive scientific computing. *Computing in Science and Engineering*, 9(3):21–29, 2007 (cité en page 178).

Bibliographie

- [Peuquet et Ci-Xiang, 1987] Peuquet, D. J. et Ci-Xiang, Z. An algorithm to determine the directional relationship between arbitrarily-shaped polygons in the plane. *Pattern Recognition*, 20(1):65–74, 1987 (cité en page 17).
- [Philbin *et al.*, 2008] Philbin, J., Chum, O., Isard, M., Sivic, J. et Zisserman, A. Lost in quantization: Improving particular object retrieval in large scale image databases. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008 (cité en page 119).
- [Rabin *et al.*, 2008] Rabin, J., Delon, J. et Gousseau, Y. Circular Earth Mover’s Distance for the comparison of local features. In *International Conference on Pattern Recognition (ICPR)*, pp. 1–4, 2008 (cité en page 103).
- [Randell *et al.*, 1992] Randell, D. A., Cui, Z. et Cohn, A. G. A spatial logic based on regions and connection. In *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pp. 165–176, 1992 (cité en pages 15, 16).
- [Randell *et al.*, 2013] Randell, D. A., Landini, G. et Galton, A. Discrete mereotopology for spatial reasoning in automated histological image analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):568–581, 2013 (cité en page 16).
- [Rosenfeld et Kak, 1982] Rosenfeld, A. et Kak, A. C. *Digital picture processing*. Academic Press (cité en page 14).
- [Rosenfeld et Klette, 1985] Rosenfeld, A. et Klette, R. Degree of adjacency or surroundedness. *Pattern Recognition*, 18(2):169–177, 1985 (cité en pages 32, 88).
- [Salembier et Garrido, 2000] Salembier, P. et Garrido, L. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing*, 9(4):561–576, 2000 (cité en pages 108, 109).
- [Sánchez *et al.*, 2013] Sánchez, J., Perronnin, F., Mensink, T. et Verbeek, J. Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision*, 105(3):222–245, 2013 (cité en page 119).
- [Santosh, 2011] Santosh, K. *Graphics recognition using spatial relations and shape analysis*. Thèse de Doctorat. Institut National Polytechnique de Lorraine (Nancy, France) (cité en page 26).
- [Santosh *et al.*, 2012] Santosh, K., Lamiroy, B. et Wendling, L. Symbol recognition using spatial relations. *Pattern Recognition Letters*, 33(3):331–341, 2012 (cité en pages 25, 37, 39).
- [Santosh *et al.*, 2014a] Santosh, K., Lamiroy, B. et Wendling, L. Integrating vocabulary clustering with spatial relations for symbol recognition. *International Journal on Document Analysis and Recognition*, 17(1):61–78, 2014 (cité en page 115).

- [Santosh *et al.*, 2009] Santosh, K., Wendling, L. et Lamiroy, B. Unified pairwise spatial relations: An application to graphical symbol retrieval. In *Graphics Recognition. Achievements, Challenges, and Evolution (GREC)*, pp. 163–174, 2009 (cité en page 25).
- [Santosh *et al.*, 2014b] Santosh, K., Wendling, L. et Lamiroy, B. BoR: Bag-of-Relations for symbol retrieval. *International Journal of Pattern Recognition and Artificial Intelligence*, 28(6), 2014 (cité en pages 39, 40, 115).
- [Savarese *et al.*, 2006] Savarese, S., Winn, J. et Criminisi, A. Discriminative object class models of appearance and shape by correlators. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2033–2040, 2006 (cité en page 114).
- [Sezgin et Sankur, 2004] Sezgin, M. et Sankur, B. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168, 2004 (cité en page 67).
- [Sivic et Zisserman, 2003] Sivic, J. et Zisserman, A. Video Google: A text retrieval approach to object matching in videos. In *IEEE International Conference on Computer Vision (ICCV)*, vol. 2, pp. 1470–1477, 2003 (cité en page 114).
- [Skubic *et al.*, 2003] Skubic, M., Matsakis, P., Chronis, G. et Keller, J. Generating multi-level linguistic spatial descriptions from range sensor readings using the histogram of forces. *Autonomous Robots*, 14(1):51–69, 2003 (cité en page 24).
- [Smeulders *et al.*, 2000] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A. et Jain, R. Content-based image retrieval at the end of the early years. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000 (cité en page 4).
- [Smith, 1996] Smith, B. Mereotopology: A theory of parts and boundaries. *Data & Knowledge Engineering*, 20(3):287–303, 1996 (cité en page 15).
- [Soille, 2008] Soille, P. Constrained connectivity for hierarchical image partitioning and simplification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1132–1145, 2008 (cité en page 108).
- [Staal *et al.*, 2004] Staal, J. J., Abramoff, M. D., Niemeijer, M., Viergever, M. A. et Van Ginneken, B. Ridge based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004 (cité en page 61).
- [Sudderth *et al.*, 2005] Sudderth, E. B., Torralba, A., Freeman, W. T. et Willsky, A. S. Learning hierarchical models of scenes, objects, and parts. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 1331–1338, 2005 (cité en page 115).
- [Tabbone *et al.*, 2003] Tabbone, S., Wendling, L. et Tombre, K. Matching of graphical symbols in line-drawing images using angular signature information. *International Journal on Document Analysis and Recognition*, 6(2):115–125, 2003 (cité en page 24).

Bibliographie

- [Tabbone et Wendling, 2003] Tabbone, S. et Wendling, L. Color and grey level object retrieval using a 3D representation of force histogram. *Image and Vision Computing*, 21(6):483–495, 2003 (cité en page 24).
- [Tabbone et Wendling, 2004] Tabbone, S. et Wendling, L. Retrieving images by content from strong relational graph matching. In *International Conference on Pattern Recognition (ICPR)*, vol. 2, pp. 951–954, 2004 (cité en page 37).
- [Takemura *et al.*, 2012] Takemura, C. M., Cesar, R. M. et Bloch, I. Modeling and measuring the spatial relation along: Regions, contours and fuzzy sets. *Pattern Recognition*, 45(2):757–766, 2012 (cité en page 36).
- [Vanegas *et al.*, 2009] Vanegas, M. C., Bloch, I. et Inglada, J. Fuzzy spatial relations for high resolution remote sensing image analysis: The case of “to go across”. In *IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 4, pp. 773–776, 2009 (cité en pages 32, 35).
- [Vanegas *et al.*, 2011] Vanegas, M. C., Bloch, I. et Inglada, J. A fuzzy definition of the spatial relation “surround” - Application to complex shapes. In *European Society for Fuzzy Logic and Technology (EUSFLAT)*, pp. 844–851, 2011 (cité en pages 33, 85, 88, 90).
- [Vanegas *et al.*, 2013] Vanegas, M. C., Bloch, I. et Inglada, J. Alignment and parallelism for the description of high-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(6):3542–3557, 2013 (cité en page 35).
- [Vieu, 1991] Vieu, L. *Semantique des relations spatiales et inferences spatio-temporelles : Une contribution a l'étude des structures formelles de l'espace en langage naturel*. Thèse de Doctorat. Université Paul Sabatier de Toulouse (Toulouse, France) (cité en page 12).
- [Vieu, 1997] Vieu, L. Spatial representation and reasoning in artificial intelligence. In *Spatial and temporal reasoning*, Springer Netherlands, pp. 5–41 (cité en page 15).
- [Walt *et al.*, 2011] Walt, S. van der, Colbert, S. C. et Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, 13(2):22–30, 2011 (cité en page 177).
- [Walt *et al.*, 2014] Walt, S. van der, Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E. et Yu, T. scikit-image: image processing in Python. *PeerJ*, 2:e453, 2014 (cité en page 177).
- [Wang et Makedon, 2003] Wang, Y. et Makedon, F. R-Histogram: Quantitative representation of spatial relations for similarity-based image retrieval. In *ACM International Conference on Multimedia*, pp. 323–326, 2003 (cité en page 24).
- [Wang *et al.*, 2004] Wang, Y., Makedon, F. et Chakrabarti, A. R*-Histograms : Efficient representation of spatial relations between objects of arbitrary topology. In *ACM International Conference on Multimedia*, pp. 356–359, 2004 (cité en page 24).

- [Wang *et al.*, 2015] Wang, Z., Fan, B., Wang, G. et Wu, F. Exploring local and overall ordinal information for robust feature description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11):2198–2211, 2015 (cité en page 139).
- [Ward, 1963] Ward, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963 (cité en pages 108, 109).
- [Wendling *et al.*, 2002] Wendling, L., Tabbone, S. et Matsakis, P. Fast and robust recognition of orbit and sinus drawings using histograms of forces. *Pattern Recognition Letters*, 23(14):1687–1693, 2002 (cité en page 100).
- [Xin *et al.*, 2006] Xin, W., Ni, J. et Matsakis, P. Fuzzy object localization based on directional (and distance) information. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 256–263, 2006 (cité en pages 28, 29).
- [Zadeh, 1965] Zadeh, L. A. Fuzzy Sets. *Information and Control*, 8(3):338–353, 1965 (cité en page 13).
- [Zhang et Lu, 2002] Zhang, D. et Lu, G. Shape-based image retrieval using generic Fourier descriptor. *Signal Processing: Image Communication*, 17(10):825–848, 2002 (cité en page 58).
- [Zhang et Lu, 2004] Zhang, D. et Lu, G. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004 (cité en page 22).
- [Zhang *et al.*, 2007] Zhang, J., Marszalek, M., Lazebnik, S. et Schmid, C. Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision*, 73(2):213–238, 2007 (cité en page 114).
- [Zipf, 1949] Zipf, G. K. *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley (cité en page 140).

Annexes

Annexe A

Liste des publications

Articles en revues internationales

- [Clément *et al.*, 2017a] Clément, M., Poulenard, A., Kurtz, C. et Wendling, L. Directional enlacement histograms for the description of complex spatial configurations between objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Sous presse, 2017 (cité en pages 56, 74, 93).
- [Clément *et al.*, 2017b] Clément, M., Kurtz, C. et Wendling, L. Learning spatial relations and shapes for structural object description and scene recognition. *Pattern Recognition*. Pré-publication, article en cours de révision, 2017 (cité en pages 123, 136).

Articles dans les actes de conférences internationales

- [Clément *et al.*, 2015a] Clément, M., Garnier, M., Kurtz, C. et Wendling, L. Color object recognition based on spatial relations between image layers. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 1, pp. 427–434, 2015 (cité en page 111).
- [Clément *et al.*, 2016a] Clément, M., Kurtz, C. et Wendling, L. Bags of spatial relations and shapes features for structural object description. In *International Conference on Pattern Recognition (ICPR)*, pp. 1994–1999, 2016 (cité en pages 111, 123, 136).
- [Clément *et al.*, 2017c] Clément, M., Kurtz, C. et Wendling, L. Fuzzy directional enlacement landscapes. In *International Conference on Discrete Geometry for Computer Imagery (DGCI)*, pp. 171–182, 2017 (cité en page 93).
- [Clément *et al.*, 2017d] Clément, M., Coustaty, M., Kurtz, C. et Wendling, L. Local enlacement histograms for historical drop caps style recognition. In *International Conference on Document Analysis and Recognition (ICDAR)*, 2017 (cité en page 149).

Articles en revues nationales

[Clément *et al.*, 2016b] Clément, M., Kurtz, C. et Wendling, L. Description d'objets en couleurs à partir des relations spatiales entre régions structurelles. *Revue des Nouvelles Technologies de l'Information. Fouille de Données Complexes*, RNTI-E-31:57–72, 2016 (cité en page 111).

Articles dans les actes de conférences nationales

[Clément *et al.*, 2015b] Clément, M., Kurtz, C. et Wendling, L. Descripteurs de relations spatiales entre régions structurelles pour la reconnaissance d'objets en couleurs. In *Atelier Fouille de Données Complexes (FDC), Conférence Internationale sur l'Extraction et la Gestion des Connaissances (EGC)*, 2015 (cité en page 111).

[Clément *et al.*, 2016c] Clément, M., Kurtz, C. et Wendling, L. Descripteurs directionnels d'enlacement et d'entrelacement entre objets. In *Congrès national sur la Reconnaissance des Formes et l'Intelligence Artificielle (RFIA)*, 2016 (cité en pages 56, 74).

[Clément *et al.*, 2017e] Clément, M., Kurtz, C. et Wendling, L. Sacs de relations spatiales et de formes pour la reconnaissance d'images de scènes naturelles. In *Journées francophones des jeunes chercheurs en vision par ordinateur (ORASIS)*, 2017 (cité en pages 123, 136).

Annexe B

Démonstrations des propriétés du descripteur d'enlacement

Cette annexe répertorie les démonstrations des différentes propriétés d'invariance des descripteurs d'enlacement et d'entrelacement présentée dans le [Chapitre 4](#) de ce manuscrit. Ces démonstrations sont notamment le fruit d'un travail collaboratif avec Adrien Poulenc, qui a effectué un stage de mathématiques appliquées au sein de l'équipe SIP du laboratoire LIPADE durant l'été de l'année 2014. Les énoncés des propriétés sont rappelés, puis sont suivis de leurs démonstrations respectives.

Propriété 1 (Périodicité). *Soit A et B deux objets. L'enlacement directionnel E_{AB} est périodique et de période π :*

$$\forall \theta \in \mathbb{R}, E_{AB}(\theta) = E_{AB}(\theta + k\pi), k \in \mathbb{Z}. \quad (\text{B.1})$$

Démonstration. Par définition, pour tout $\theta, t \in \mathbb{R}$,

$$f_A^{(\theta+\pi, \rho)}(t) = f_A^{(\theta, -\rho)}(-t). \quad (\text{B.2})$$

En utilisant le changement de variables $(u, v, w, h) = -(x, y, z, \rho)$ dans l'expression de $E_{AB}(\theta + \pi)$ nous obtenons :

$$\begin{aligned} E_{AB}(\theta + \pi) &= \int_{\mathbb{R}^2} f_B^{(\theta, h)}(u) \int_{-\infty}^u f_A^{(\theta, h)}(v) \int_{-\infty}^v f_B^{(\theta, h)}(w) dw dv du dh \\ &= \int_{\mathbb{R}^4} f_B^{(\theta, h)}(u) f_A^{(\theta, h)}(v) f_B^{(\theta, h)}(w) \mathbb{1}_{\{u \geq v \geq w\}} dw dv du dh. \end{aligned} \quad (\text{B.3})$$

En effectuant alors le changement de variables $(u, v, w) = (z, y, x)$, on a alors :

$$\begin{aligned} E_{AB}(\theta + \pi) &= \int_{\mathbb{R}^4} f_B^{(\theta, h)}(z) f_A^{(\theta, h)}(y) f_B^{(\theta, h)}(x) \mathbb{1}_{\{z \geq y \geq x\}} dx dy dz dh \\ &= \int_{\mathbb{R}^2} f_B^{(\theta, h)}(x) \int_x^{+\infty} f_A^{(\theta, h)}(y) \int_y^{+\infty} f_B^{(\theta, h)}(z) dz dy dx dh \\ &= E_{AB}(\theta). \end{aligned} \quad \square$$

Propriété 2 (Translation). *Soit A et B deux objets. L'enlacement directionnel E_{AB} est invariant par rapport aux translations. En notant T_v la translation par un vecteur $v \in \mathbb{C} \simeq \mathbb{R}^2$, nous avons :*

$$E_{T_v(A)T_v(B)}(\theta) = E_{AB}(\theta). \quad (\text{B.4})$$

Démonstration. Pour toute fonction $C : \mathbb{R}^2 \rightarrow \mathbb{R}$, une translation T_v est définie par :

$$T_v(C)(x) = C(x - v).$$

Ainsi, pour tout $\theta, \rho, t \in \mathbb{R}$, on a :

$$\begin{aligned} f_{T_v(C)}^{(\theta, \rho)}(t) &= T_v(C)(e^{i\theta}(t + i\rho)) \\ &= C(e^{i\theta}(t + i\rho) - v) \\ &= C(e^{i\theta}(t + i\rho - e^{-i\theta}v)) \\ &= C(e^{i\theta}(t - \operatorname{Re}(e^{-i\theta}v) + i(\rho - \operatorname{Im}(e^{-i\theta}v)))) \\ &= f_C^{(\theta, \rho - \operatorname{Im}(e^{-i\theta}v))}(t - \operatorname{Re}(e^{-i\theta}v)). \end{aligned} \quad (\text{B.5})$$

Par conséquent, on obtient :

$$\begin{aligned} \int_{\mathbb{R}} E(f_{T_v(A)}^{(\theta, \rho)}, f_{T_v(B)}^{(\theta, \rho)}) d\rho &= \int_{\mathbb{R}} E(f_A^{(\theta, \rho - \operatorname{Im}(e^{-i\theta}v))}, f_B^{(\theta, \rho - \operatorname{Im}(e^{-i\theta}v))}) d\rho \\ &= \int_{\mathbb{R}} E(f_A^{(\theta, \rho)}, f_B^{(\theta, \rho)}) d\rho. \end{aligned} \quad \square$$

Propriété 3 (Rotation). *Soit A et B deux objets. L'enlacement directionnel E_{AB} est co-variant par rapport aux rotations. Soit $\alpha \in \mathbb{R}$ un angle et soit R_α la rotation d'angle α associée, alors on a :*

$$E_{R_\alpha(A)R_\alpha(B)}(\theta) = E_{AB}(\theta - \alpha). \quad (\text{B.6})$$

Démonstration. Par définition, pour toute fonction $C : \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$R_\alpha(C) = C(e^{-i\alpha} \bullet). \quad (\text{B.7})$$

Par conséquent, pour tout $\theta, \rho \in \mathbb{R}$ nous vérifions que :

$$f_{R_\alpha(C)}^{(\theta, \rho)} = f_C^{(\theta - \alpha, \rho)}. \quad (\text{B.8})$$

En ce sens :

$$\begin{aligned} E_{R_\alpha(A)R_\alpha(B)}(\theta) &= \int_{\mathbb{R}} E(f_{R_\alpha(A)}^{(\theta, \rho)}, f_{R_\alpha(B)}^{(\theta, \rho)}) d\rho \\ &= \int_{\mathbb{R}} E(f_A^{(\theta - \alpha, \rho)}, f_B^{(\theta - \alpha, \rho)}) d\rho \\ &= E_{AB}(\theta - \alpha). \end{aligned} \quad \square$$

Propriété 4 (Homothétie 1). *Soit A et B deux objets. L'enlacement directionnel E_{AB} est quasi-invariant par rapport aux homothéties. Soit $\lambda \in \mathbb{R}$ et S_λ l'homothétie de facteur λ associée, alors nous avons :*

$$E_{S_\lambda(A)S_\lambda(B)}(\theta) = \lambda^4 E_{AB}(\theta). \quad (\text{B.9})$$

Démonstration. Soit $C : \mathbb{R}^2 \rightarrow \mathbb{R}$, nous avons pour tout $\theta, \rho \in \mathbb{R}$ et pour tout $t \in \mathbb{R}$:

$$\begin{aligned} f_{S_\lambda(C)}^{(\theta, \rho)}(t) &= S_\lambda(C)(e^{i\theta}(t + i\rho)) \\ &= C(e^{i\theta}(\frac{t}{\lambda} + i\frac{\rho}{\lambda})) \\ &= f_C^{(\theta, \rho/\lambda)}(\frac{t}{\lambda}). \end{aligned} \quad (\text{B.10})$$

Par conséquent, on a :

$$E(f_{S_\lambda(A)}^{(\theta, \rho)}, f_{S_\lambda(B)}^{(\theta, \rho)}) = E(f_A^{(\theta, \rho/\lambda)}(\frac{\bullet}{\lambda}), f_B^{(\theta, \rho/\lambda)}(\frac{\bullet}{\lambda})). \quad (\text{B.11})$$

Soit $f, g : \mathbb{R} \rightarrow \mathbb{R}$ des fonctions mesurables bornées à support compact. Avec le changement de variable $w = \frac{z}{\lambda}, v = \frac{y}{\lambda}$ et $u = \frac{x}{\lambda}$, nous obtenons :

$$\begin{aligned} E(f(\frac{\bullet}{\lambda}), g(\frac{\bullet}{\lambda})) &= \int_{\mathbb{R}} f(\frac{x}{\lambda}) \int_x^{+\infty} f(\frac{y}{\lambda}) \int_y^{+\infty} g(\frac{z}{\lambda}) dz dy dx \\ &= \lambda \int_{w=\frac{z}{\lambda}} f(\frac{x}{\lambda}) \int_x^{+\infty} f(\frac{y}{\lambda}) \int_{\frac{y}{\lambda}}^{+\infty} g(w) dw dy dx \\ &= \lambda^2 \int_{v=\frac{y}{\lambda}} f(\frac{x}{\lambda}) \int_{\frac{x}{\lambda}}^{+\infty} f(v) \int_y^{+\infty} g(w) dw dv dx \\ &= \lambda^3 \int_{u=\frac{x}{\lambda}} f(u) \int_{\frac{x}{\lambda}}^{+\infty} f(v) \int_y^{+\infty} g(w) dw dv du \\ &= \lambda^3 E(f, g). \end{aligned} \quad (\text{B.12})$$

En appliquant ce résultat à $f = f_A^{(\theta, \rho/\lambda)}$ et $g = f_B^{(\theta, \rho/\lambda)}$ nous avons alors :

$$\begin{aligned} E_{S_\lambda(A)S_\lambda(B)}(\theta) &= \int_{\mathbb{R}} E(f_{S_\lambda(A)}^{(\theta, \rho)}, f_{S_\lambda(B)}^{(\theta, \rho)}) d\rho \\ &= \int_{\mathbb{R}} E(f_A^{(\theta, \rho/\lambda)}(\frac{\bullet}{\lambda}), f_B^{(\theta, \rho/\lambda)}(\frac{\bullet}{\lambda})) d\rho \\ &= \lambda^3 \int_{\mathbb{R}} E(f_A^{(\theta, \rho/\lambda)}, f_B^{(\theta, \rho/\lambda)}) d\rho \\ &= \lambda^4 \int_{h=\frac{\rho}{\lambda}} E(f_A^{(\theta, h)}, f_B^{(\theta, h)}) dh \\ &= \lambda^4 E_{AB}(\theta). \quad \square \end{aligned}$$

Propriété 5 (Homothétie 2). *Soit deux objets A et B . L'histogramme d'enlacement \mathcal{E}_{AB} est invariant par rapport aux homothéties. Soit $\lambda \in \mathbb{R}$ et S_λ l'homothétie de facteur λ associée, alors*

Annexe B. Démonstrations des propriétés du descripteur d'enlacement

on a :

$$\mathcal{E}_{S_\lambda(A)S_\lambda(B)}(\theta) = \mathcal{E}_{AB}(\theta). \quad (\text{B.13})$$

Démonstration. Soit $\lambda \in \mathbb{R}$ un facteur d'homothétie. Par définition, pour tout objet A , nous avons :

$$\|S_\lambda(A)\|_1 = \lambda^2 \|A\|_1. \quad (\text{B.14})$$

Ainsi, pour tout objets A et B :

$$\begin{aligned} \mathcal{E}_{S_\lambda(A)S_\lambda(B)}(\theta) &= \frac{E_{S_\lambda(A)S_\lambda(B)}(\theta)}{\|S_\lambda(A)\|_1 \|S_\lambda(B)\|_1} \\ &= \frac{\lambda^4 E_{AB}(\theta)}{\lambda^4 \|A\|_1 \|B\|_1} \\ &= \mathcal{E}_{AB}(\theta). \end{aligned} \quad \square$$

Annexe C

Développements logiciels

Les travaux de recherche réalisés dans le contexte de cette thèse ont donné lieu à de nombreux développements logiciels. Dans cette annexe, nous présentons un aperçu du cadre de développement dans lequel ces travaux ont été réalisés. De manière générale, nous nous sommes attachés à la mise en place de protocoles de validation rendant les résultats expérimentaux obtenus reproductibles. À l’avenir, nous envisageons de distribuer sous licence libre le code source des différents développements logiciels de cette thèse.

Python 3

La plupart des développements réalisés durant cette thèse ont été effectués avec le langage de programmation *Python 3*, qui propose un environnement scientifique particulièrement performant, et dont le code source est sous licence libre. Nous dressons ci-après une liste synthétique des bibliothèques logicielles et outils employés pour développer les modèles proposés, ainsi que pour la réalisation des différentes validations expérimentales de ces travaux de thèse.

NumPy

La librairie *NumPy* [Walt *et al.*, 2011] est au cœur de l’environnement de développement scientifique en *Python*. Elle offre un cadre général pour la manipulation de données numériques multidimensionnelles, ainsi qu’un ensemble de fonctions mathématiques utilitaires, comme par exemple pour l’algèbre linéaire ou la génération de nombres aléatoires. Optimisée par des routines de bas-niveau développées en langage *C*, la bibliothèque *NumPy* permet d’effectuer des calculs autrement plus rapides que s’ils étaient réalisés au sein de la bibliothèque standard de *Python*, ce dernier étant un langage interprété.

Scikit-image et Scikit-learn

La librairie *Scikit-image* [Walt *et al.*, 2014] propose de nombreuses méthodes utilitaires pour le traitement d’images. Elle permet ainsi, par exemple, d’appliquer différents types d’opérateurs de morphologie mathématique, de filtrer ou de redimensionner des images. Cette librairie met également à disposition un ensemble d’algorithmes pour diverses tâches liées à la vision par

ordinateur. Par exemple, les caractéristiques locales HOG utilisées dans les expérimentations de cette thèse ont été calculées à partir de cette librairie. Nous avons également utilisé différentes méthodes de binarisation d'images qui y sont proposées.

Parallèlement, la librairie *Scikit-learn* [Pedregosa *et al.*, 2011] propose un ensemble d'algorithmes liés au domaine de l'apprentissage automatique. On y retrouve différents algorithmes de *clustering*, de classification, ainsi que des méthodes de régression. Par ailleurs, cette librairie propose également des méthodes utilitaires pour calculer divers types d'indicateurs permettant d'évaluer quantitativement les résultats de tels algorithmes, ainsi que pour la mise en place de protocoles de validation (comme par exemple différentes méthodes de validation croisée).

Matplotlib et IPython

La librairie *Matplotlib* [Hunter, 2007] permet de générer et de visualiser différents types de graphiques et de figures à partir de données, selon une syntaxe pratique et accessible. Nous en avons ainsi fait une utilisation intensive.

Par ailleurs, le travail de recherche scientifique avec le langage *Python* est également rendu plus efficace par l'utilisation de l'environnement *IPython* [Pérez et Granger, 2007], en particulier grâce au *Notebook* (récemment intégré au sein d'un projet plus large nommé *Jupyter*). Le *Notebook* permet, en effet, d'exécuter à la volée des morceaux de code en *Python* organisés sous la forme de cellules personnalisables, au sein d'une interface web. Il représente ainsi un moyen particulièrement utile pour tester différents prototypes de modèles, ainsi que pour présenter et partager des résultats de recherche.

Modélisation et apprentissage de relations spatiales pour la reconnaissance et l'interprétation d'images

Ces dernières années, la quantité de données visuelles produites par divers types de capteurs est en augmentation permanente. L'interprétation et l'indexation automatique de telles données constituent des défis importants pour les domaines liés à la reconnaissance de formes et la vision par ordinateur. Dans ce contexte, la position relative des différents objets d'intérêt composant les images représente une information particulièrement importante pour interpréter leur contenu. Les relations spatiales sont en effet porteuses d'une sémantique riche, qui est fortement liée à la perception humaine. Les travaux de recherche présentés dans cette thèse proposent ainsi d'explorer différentes approches génériques de description de l'information spatiale, en vue de les intégrer dans des systèmes de reconnaissance et d'interprétation d'images de haut niveau.

Tout d'abord, nous présentons une approche pour la description de configurations spatiales complexes, où les objets peuvent être imbriqués les uns dans les autres. Cette notion est formalisée par deux nouvelles relations spatiales, nommées enlacement et entrelacement. Nous proposons un modèle qui permet de décrire et de visualiser ces configurations avec une granularité directionnelle. Ce modèle est validé expérimentalement pour des applications en imagerie biomédicale, en télédétection et en analyse d'images de documents. Ensuite, nous présentons un cadre d'apprentissage de relations spatiales composites à partir d'ensembles d'images. Inspirée des approches par sacs de caractéristiques visuelles, cette stratégie permet de construire des vocabulaires de configurations spatiales apparaissant dans les images, à différentes échelles. Ces caractéristiques structurelles peuvent notamment être combinées avec des descriptions locales, conduisant ainsi à des représentations hybrides et complémentaires. Les résultats expérimentaux obtenus sur différentes bases d'images structurées permettent d'illustrer l'intérêt de cette approche pour la reconnaissance et la classification d'images.

Modeling and Learning Spatial Relations for Image Recognition and Understanding

In recent years, the amount of visual data produced by various types of sensors has been continuously increasing. The automatic interpretation and indexation of such data constitute an important challenge in the fields of pattern recognition and computer vision. In this context, the relative position of the different objects of interest depicted in images represents particularly important information for the interpretation of their content. Spatial relations indeed carry rich semantics that are strongly tied with human perception. The research work presented in this thesis thus proposes to explore different generic approaches to the description of spatial information, in order to integrate them in high-level image recognition and understanding systems.

First, we present an approach for the description of complex spatial configurations, where objects can be imbricated in each other. This notion is formalized by two novel spatial relations, namely enlacement and interlacement. We propose a model to describe and to visualize these configurations with directional granularity. This model is experimentally validated for applications in biomedical imaging, remote sensing and document image analysis. Then, we present a framework for learning composite spatial relations from image datasets. Inspired by bags of visual features approaches, this strategy allows to build vocabularies of spatial configurations occurring across images, at different scales. These structural features can notably be combined with local descriptions, leading to hybrid and complementary representations. Experimental results obtained for different datasets of structured images highlight the interest of this approach for image recognition and classification tasks.