



HAL
open science

Méta-modèles de positionnement spatial pour la reconnaissance de tracés manuscrits

Adrien Delaye

► **To cite this version:**

Adrien Delaye. Méta-modèles de positionnement spatial pour la reconnaissance de tracés manuscrits. Traitement du texte et du document. INSA de Rennes, 2011. Français. <NNT : >. <tel-00651446>

HAL Id: tel-00651446

<https://theses.hal.science/tel-00651446v1>

Submitted on 13 Dec 2011

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.



HAL Authorization

Thèse



THÈSE INSA Rennes
sous le sceau de l'Université Européenne de Bretagne
pour obtenir le grade de
DOCTEUR DE L'INSA DE RENNES
Spécialité : informatique

présentée par

Adrien Delaye

ÉCOLE DOCTORALE : MATISSE
LABORATOIRE : IRISA – UMR6074

Méta-modèles de positionnement spatial pour la reconnaissance de tracés manuscrits

Thèse soutenue le 09/12/2011

devant le jury composé de :

Christian Viard-Gaudin

Professeur à l'université de Nantes / *Président*

Hubert Cardot

Professeur à l'université de Tours / *Rapporteur*

Laurent Wendling

Professeur à l'université Paris Descartes / *Rapporteur*

Laurent Heutte

Professeur à l'université de Rouen / *Examineur*

Eric Petit

Ingénieur de recherche France Télécom / *Examineur*

Eric Anquetil

Professeur à l'INSA de Rennes / *Directeur de thèse*

Remerciements

Je tiens à remercier en premier lieu l'ensemble des membres de mon jury pour avoir accepté de juger les travaux présentés dans cette thèse. Je remercie en particulier Christian Viard-Gaudin, qui m'a fait l'honneur de présider ce jury, ainsi que Hubert Cardot et Laurent Wendling, qui ont accepté de la tâche de rapporteur pour ce document. Merci également à Laurent Heutte et à Eric Petit pour avoir joué le rôle d'examineurs de la soutenance.

Ma reconnaissance et ma gratitude vont ensuite à Eric pour la qualité de l'encadrement dont j'ai bénéficié pendant ces trois années de thèse. Son enthousiasme communicatif, sa disponibilité, mais aussi la grande confiance qu'il m'a accordée ont profondément contribué à l'aboutissement de ces travaux.

J'adresse toute ma sympathie aux membres de l'équipe Imadoc, à mes collègues passés et présents, et tout particulièrement à Joseph et Abdullah, avec qui j'ai partagé bien plus qu'un bureau : de la bonne humeur, des échanges amicaux, des discussions scientifiques et des projets en commun.

Je remercie ma famille et mes amis pour leur chaleureux soutien tout au long de ces mois de travail. Merci en particulier à mon père pour le temps et l'attention qu'il a portée à relire ce document. Merci enfin à Adeline, qui a toujours été à mes côtés, malgré la distance.

Table des matières

Table des matières	ii
Table des figures	vi
Liste des tableaux	ix
1 Introduction	1
2 Etat de l'art : description du positionnement relatif	7
2.1 Représentation par descripteurs	8
2.1.1 Positionnement de primitives manuscrites	10
2.1.1.1 Approches quantitatives	10
2.1.1.2 Approches qualitatives	15
2.1.1.3 Synthèse	19
2.1.2 Positionnement d'objets manuscrits complexes	19
2.1.2.1 Approches quantitatives	20
2.1.2.2 Approches qualitatives	23
2.1.2.3 Synthèse	24
2.1.3 Positionnement générique d'objets dans des images	26
2.1.3.1 Approches quantitatives	27
2.1.3.2 Approches qualitatives	30
2.1.4 Synthèse des approches par descripteur	30
2.2 Représentations spatiales	31
2.2.1 Théorie des modèles spatiaux pour l'appréhension de relations spatiales	33
2.2.1.1 Principes	33
2.2.1.2 Expressivité	34
2.2.1.3 Gradualité, imprécision et formalisation floue	34
2.2.1.4 Usages : vérification et localisation	35
2.2.2 Modèles spatiaux simples pour l'analyse du manuscrit	35
2.2.3 Modèles spatiaux morphologiques	37
2.2.3.1 Définition par dilatation morphologique floue	37
2.2.3.2 Exemple : paysages flous directionnels	38
2.2.3.3 Comparaison d'un objet avec le modèle spatial	40
2.2.3.4 Exemple : évaluation de positions directionnelles	41
2.2.3.5 Expressivité	41

Table des matières

2.2.4	Modèles spatiaux par l'approche \mathcal{F} -template	43
2.2.5	Synthèse des approches par modèles spatiaux	45
2.3	Positionnement de nos travaux	47
2.3.1	Apprentissage de relations spatiales pour l'analyse du manuscrit	47
2.3.2	Vers l'apprentissage de méta-modèles spatiaux	48
3	Méta-modèles spatiaux	51
3.1	Principes	52
3.1.1	Schéma général et interprétation	52
3.1.2	Morphologie mathématique et apprentissage	54
3.1.3	Combinaison de points de vue directionnels	56
3.2	Formalisation et mise en œuvre	58
3.2.1	Apprentissage de points de vue	58
3.2.1.1	Formalisation	58
3.2.1.2	Apprentissage	59
3.2.1.3	Exemples	63
3.2.2	Sélection des points de vue directionnels	66
3.2.2.1	Stratégie générique	66
3.2.2.2	Stratégies alternatives	68
3.2.3	Conjonction des points de vue	70
3.3	Prise en compte de la distance	71
3.3.1	Stratégie globale	72
3.3.2	Stratégie directionnelle	74
3.4	Prise en compte de la bipolarité de l'information spatiale	77
3.4.1	Définitions	78
3.4.2	Morphologie bipolaire floue	78
3.4.3	Méta-modèles spatiaux bipolaires	80
3.4.3.1	Apprentissage des fonctions des points de vue bipolaires	80
3.4.3.2	Construction du modèle spatial bipolaire global	83
3.5	Intégration d'un modèle d'étendue	84
3.5.1	Objectif	85
3.5.2	Etendue d'un objet vu comme un nuage de points	87
3.5.3	Etendue d'un objet par considération de ses points singuliers	91
3.5.4	Synthèse	95
3.6	Résultats qualitatifs	96
3.6.1	Modélisation de relations entre primitives simples	96
3.6.1.1	Primitives de tracé distantes	97
3.6.1.2	Concavités et recouvrement de boîtes englobantes	98
3.6.1.3	Primitives adjacentes	99
3.6.1.4	Objets avec intersection	100
3.6.2	Modélisation de relations entre objets manuscrits complexes	101
3.6.3	Problème de localité du contexte de positionnement	102
3.7	Synthèse	104
4	Représentation de formes par <i>Patrons Structurés Flous</i>	107
4.1	Démarche proposée	108

4.1.1	Stabilité et importance de l'information spatiale	108
4.1.2	Représentation de symboles par <i>Patron Structurel Flou</i> (PSF)	109
4.1.3	Segmentation guidée par les modèles spatiaux	111
4.2	<i>Patrons Structurels Flous</i> pour la représentation de symboles	112
4.2.1	Définition du référentiel de positionnement	112
4.2.1.1	Référentiel concret	112
4.2.1.2	Choix d'un référentiel global	113
4.2.1.3	Optimisation du pouvoir descriptif du référentiel	114
4.2.1.4	Positionnement absolu du référentiel par <i>Référence Virtuelle Absolue</i> (RVA)	115
4.2.2	Formalisation des PSF	116
4.2.2.1	Correspondance d'un segment avec un composant du PSF	118
4.2.2.2	Fusion globale	120
4.2.2.3	Comparaison de scores bipolaires	125
4.2.3	Exploitation des PSF	126
4.2.3.1	Hypothèses de références	126
4.2.3.2	Segmentation guidée par les modèles spatiaux	128
4.2.3.3	Recherche de l'assignement optimal des segments	131
4.2.3.4	Synthèse	133
4.2.4	Apprentissage des PSF	133
4.2.4.1	Initialisation	134
4.2.4.2	Itérations pour la recherche de références	137
4.2.4.3	Itérations pour l'apprentissage du PSF complet	138
4.3	Validation qualitative	139
4.4	Synthèse	142
5	Expérimentations	145
5.1	Classification de relations spatiales entre objets manuscrits	145
5.1.1	Jeux de données	146
5.1.1.1	Accents et gestes d'édition en-ligne	146
5.1.1.2	Primitives de caractères chinois	147
5.1.2	Protocole expérimental	149
5.1.2.1	Processus de validation omni-scripteur	149
5.1.2.2	Apprentissage et optimisation du classifieur	150
5.1.2.3	Significativité des comparaisons	150
5.1.3	Expérimentations	151
5.1.3.1	Apport des méta-modèles de positionnement	151
5.1.3.2	Apport de la distance	154
5.1.3.3	Apport de la bipolarité	157
5.1.3.4	Apport de la modélisation de l'étendue	158
5.1.4	Généralisation à d'autres jeux de données	160
5.1.4.1	Lettres en contexte de mots manuscrits hors-ligne	160
5.1.4.2	Symboles mathématiques en-ligne	163
5.1.4.3	Résultats	164
5.1.5	Synthèse	167
5.2	Evaluation des PSF pour la représentation de symboles structurés	167

Table des matières

5.2.1	Jeux de données	168
5.2.1.1	Caractères chinois	168
5.2.1.2	Pictogrammes	169
5.2.1.3	Symboles architecturaux	170
5.2.2	Protocole	171
5.2.3	Expérimentations sur la base CaracDB	172
5.2.3.1	Importance de la modélisation de l'étendue	173
5.2.3.2	Apport de la distance	175
5.2.3.3	Validité de la représentation structurelle	175
5.2.4	Limitations de la représentation	177
5.2.4.1	Limites de la tolérance aux distorsions de segmentation	177
5.2.4.2	Limites de la précision de la représentation par PSF	179
5.2.5	Synthèse sur la représentation de symboles par PSF	180
6	Conclusion	183
6.1	Bilan	183
6.2	Perspectives	185
A	Introduction à la logique floue	187
A.1	Concepts généraux	187
A.2	Définition d'un sous-ensemble flou	188
A.3	Caractéristiques d'un sous-ensemble flou	188
A.4	Opérations de base sur les sous-ensembles flous	189
A.5	Normes et conormes triangulaires	190
B	Aspects algorithmiques	193
B.1	Cas de données image (représentation hors-ligne)	193
B.2	Cas de données vectorisées (représentation en-ligne)	195
B.2.1	Cas d'un élément structurant directionnel	195
B.2.2	Cas d'un élément structurant de distance	197
	Bibliographie	199
	Références de l'auteur	209

Table des figures

1.1	Principe de raisonnement par modèle spatial	3
2.1	Schéma général des approches par extraction de descripteurs	9
2.2	Positionnement de primitives par la position du centre de leur boîte englobante	11
2.3	Calcul de descripteurs de positionnement relatif entre primitives de tracé	12
2.4	Modèles relationnels appris pour le positionnement relatif de primitives de tracé	13
2.5	Positionnement relatif de primitives de caractères chinois par vecteurs reliant leurs extrémités	14
2.6	Description de contraintes géométriques entre tracés par <i>Relative Shape Histograms</i>	15
2.7	Représentation d'un caractère chinois par un graphe avec catégorisation des relations spatiales	16
2.8	Catégories de relations spatiales inter-primitives dans des caractères chinois	17
2.9	Catégories de relations spatiales entre primitives de symboles architecturaux	18
2.10	Ambiguïté des relations spatiales entre symboles d'une expression mathématique	20
2.11	Extraction de caractéristiques pour le positionnement relatif d'objets dans une expression mathématique	21
2.12	Structure arborescente et patrons structurels de composition pour les caractères chinois	22
2.13	Caractérisation des espaces inter-mots et inter-lettres dans une ligne de texte manuscrit	23
2.14	Importance de la prise en compte des formes des objets pour la description de relations spatiales entre éléments d'une expression mathématique	24
2.15	Problèmes posés par la représentation des objets par leur boîte englobante	25
2.16	Description de relations directionnelles par histogrammes d'angles	28
2.17	Définition de régions d'intérêt pour une description unifiée d'informations topologiques et de positionnement directionnel pour deux objets d'une image	30
2.18	Schéma général des approches par modèles spatiaux	32
2.19	Positionnement relatif d'objets par définition d'une partition empirique du plan image	36
2.20	Fonction $f(\theta) = \max(0, 1 - \frac{2}{\pi}\theta)$	39
2.21	Élément structurant directionnel flou modélisant la relation <i>être à droite</i>	39
2.22	Paysages flous <i>à droite de R</i> pour différents objets R	40
2.23	Evaluation du positionnement relatif de deux objets par description morphologique directionnelle	42
2.24	Exemple de \mathcal{F} -template modélisant la relation <i>au Nord de R</i>	44
2.25	Calcul de la valeur du \mathcal{F} -template en un point du plan image	44

Table des figures

2.26	Diagramme de Venn issu d'une ontologie de relations spatiales pour l'analyse d'images	48
3.1	Schéma général de l'approche proposée pour l'apprentissage et l'exploitation de méta-modèles spatiaux	53
3.2	Apprentissage naïf d'élément structurant pour la modélisation de relation spatiale	55
3.3	Description de la position d'un point sous différents points de vue par rapport à différentes parties de l'objet de référence	57
3.4	Principe de l'apprentissage du méta-modèle pour le point de vue à droite . . .	60
3.5	Principe de l'exploitation d'un modèle appris selon le point de vue à droite . .	61
3.6	Visualisation des zones pour différents niveaux de quantification des degrés directionnels	62
3.7	Représentation de la fonction de seuillage fréquence-appartenance floue	63
3.8	Exemples d'apprentissage d'histogrammes pour différentes relations spatiales .	64
3.9	Complémentarité des modèles spatiaux issus des points de vue opposés	67
3.10	Complémentarité des quatre points de vue orientés par les directions cardinales	68
3.11	Influence de la forme de la référence sur la description	69
3.12	Intersection des quatre modèles spatiaux directionnels appris pour former le modèle spatial global	71
3.13	Application d'un même méta-modèle spatial à différents objets de référence . .	72
3.14	Élément structurant <i>être proche de</i>	73
3.15	Modèle spatial de distance considéré comme un point de vue additionnel	74
3.16	Intégration de modèles de distance au sein des points de vue directionnels . . .	75
3.17	Ensemble flou trapézoïdal pour la modélisation de la distance	75
3.18	Modèle spatial hybride (direction-distance) appris pour un point de vue	76
3.19	Représentation des deux fonctions de seuillage fréquence-appartenance floue pour l'apprentissage de modèles bipolaires	81
3.20	Exemples d'apprentissage de méta-modèles spatiaux bipolaires	82
3.21	Choix de couleurs pour la représentation des modèles spatiaux bipolaires . . .	84
3.22	Illustration du besoin de modélisation de l'étendue des objets par un exemple d'intersection	86
3.23	Illustration du besoin de modélisation de l'étendue des objets pour distinguer des modèles spatiaux similaires	87
3.24	Exemples d'apprentissage du modèle d'étendue pour différentes configurations .	90
3.25	Limitations de la modélisation de l'étendue par nuage de points	91
3.26	Sélection de points particuliers pour la modélisation de l'étendue	92
3.27	Exemples d'apprentissage du modèle d'étendue pour différentes configurations.	94
3.28	Exemple de l'intérêt de la modélisation de l'étendue par points particuliers . .	95
3.29	Positionnement de primitives distantes	97
3.30	Positionnement de primitives avec concavités et recouvrement	98
3.31	Positionnement de primitives connexes	99
3.32	Positionnement de primitives avec intersection	100
3.33	Positionnement d'objets complexes distants	102
3.34	Positionnement d'objets complexes avec recouvrement	103
3.35	Confusion entre des contextes localement similaires	103

4.1	Exemples de symboles structurés manuscrits : caractères chinois, pictogrammes et symboles architecturaux	110
4.2	Evaluation de la qualité du choix des références de positionnement	115
4.3	Intérêt de la Référence Virtuelle Absolue pour la recherche de références de positionnement	116
4.4	Illustration des relations spatiales modélisées dans un PSF	117
4.5	Exploitation d'un composant de PSF pour un segment de tracé	118
4.6	Fusion de modèles spatiaux construits sur deux références distinctes	119
4.7	Schéma d'exploitation des modèles de positionnement mutuel des références d'un PSF	121
4.8	Exemple d'assignement des segments aux composants d'un PSF	122
4.9	Schéma global de fusion globale des scores de positionnement relatif pour l'exploitation d'un PSF	123
4.10	Visualisation des meilleures hypothèses de références de positionnement pour différents symboles	129
4.11	Segmentation guidée par les modèles spatiaux	130
4.12	Schéma itératif pour l'apprentissage des PSF	135
4.13	Exemples de références déterminées par apprentissage de PSF	140
4.14	Visualisation des composants de PSF appris	141
5.1	Exemples de mots manuscrits hors-ligne extraits de RIMES	161
5.2	Segmentation de mots manuscrits en lettres pour la constitution de la base MOT-OffDB	162
5.3	Jeu de symboles PictoDB	169
5.4	Taux d'erreur en fonction du poids relatif de l'étendue et du positionnement sur la base de validation de CaracDB	173
5.5	Difficultés posées par la base PictoDB	179
5.6	Exemples de classes mal reconnues dans la base ArchiDB	180
A.1	Caractère graduel des degrés d'appartenance à un sous-ensemble flou	188
A.2	Caractéristiques d'un sous-ensemble flou A	189
B.1	Calcul de la dilatation d'un objet représenté par une matrice de pixels	194
B.2	Calcul de la dilatation d'un objet représenté par un polygone avec un élément structurant de direction	196
B.3	Calcul de la dilatation d'un objet représenté par un polygone avec un élément structurant de distance	197

Liste des tableaux

2.1	Caractéristiques de positionnement relatif des primitives de tracé du chiffre <i>sept</i>	12
5.1	Représentation de 16 classes de la base IME-OnDB	147
5.2	Représentation de 16 classes de la base HCC-OnDB	148
5.3	Caractéristiques utilisées dans les expérimentations	151
5.4	Comparaison de méthodes de description des relations spatiales pour les bases IME-OnDB et HCC-OnDB	152
5.5	Significativité statistique à 5% de la comparaison deux à deux des expérimentations du tableau 5.4	153
5.6	Visualisation des modèles de positionnement appris pour la base IME-OnDB	154
5.7	Modèles de positionnement appris pour la base HCC-OnDB	154
5.8	Jeux de caractéristiques avec intégration de la distance	155
5.9	Comparaison des jeux de caractéristiques avec distance pour les deux bases IME-OnDB et HCC-OnDB	155
5.10	Significativité statistique à 5% de la comparaison des expérimentations du tableau 5.9	155
5.11	Modèles de positionnement avec distance appris pour 16 classes de la base IME-OnDB	156
5.12	Modèles de positionnement avec distance appris pour 16 classes de la base HCC-OnDB	157
5.13	Jeux de caractéristiques avec prise en compte de la bipolarité	157
5.14	Comparaison des jeux de caractéristiques bipolaires pour les deux bases IME-OnDB et HCC-OnDB	158
5.15	Modèles de positionnement bipolaires appris pour 16 classes de la base IME-OnDB	158
5.16	Modèles de positionnement bipolaires appris pour 16 classes de la base HCC-OnDB	159
5.17	Jeux de caractéristiques d'étendue	159
5.18	Comparaison des différentes modélisations de l'étendue pour les bases IME-OnDB et HCC-OnDB	160
5.19	Classes de relations spatiales dans la base MATH-OnDB	163
5.20	Résultats de la classification de relations spatiales sur les bases MATH-OnDB et MATH-OnDB avec différents jeux de caractéristiques de positionnement	164
5.21	Visualisation des modèles spatiaux entre paire de lettres de la base MOT-OffDB	166
5.22	Visualisation des modèles spatiaux entre des éléments de la base MATH-OnDB	166
5.23	Jeu de symboles CaracDB	168
5.24	Jeu de symboles ArchiDB	170
5.25	Statistiques des bases de symboles	171

5.26	Taux d'erreur sur la base de CaracDB avec et sans scores d'étendue	173
5.27	Apport de l'étendue pour la réduction de la confusion entre certains symboles .	174
5.28	Taux d'erreur sur la base CaracDB avec et sans modélisation de la distance . .	175
5.29	Apport de la distance pour la précision des cartes de segmentation	176
5.30	Exemples de symboles bien reconnus	176
5.31	Exemples de symboles mal reconnus	177
5.32	Taux d'erreur pour chaque classe de la base PictoDB	178
A.1	Exemples de T-normes et T-conormes duales.	191

Introduction

L'écriture ou la *composition de formes manuscrites* est un moyen de communication universel, ergonomique et extrêmement efficace pour exprimer des idées complexes par l'écriture de mots et de phrases, la saisie d'expressions structurées (par exemple des notations mathématiques), la composition de schémas (organigrammes, plans d'architecte, schémas techniques ou dessins géométriques) ou encore la « prise de note ». La richesse de l'information contenue dans un *document manuscrit* repose sur deux aspects complémentaires. D'abord, les éléments tracés, produits par le parcours de la pointe du stylo au contact de la surface de saisie, véhiculent une grande partie du sens global du document : ce sont les lettres, les symboles mathématiques, les traits qui forment les dessins. Il s'agit des *formes* des objets que l'on perçoit. Ensuite, les relations, non directement visibles, qui régissent le positionnement de ces objets tracés décrivent une information essentielle : la position d'une lettre par rapport à une autre, l'agencement des symboles d'une expression, la structure des traits d'un dessin. Il s'agit du *positionnement spatial* que l'on perçoit indirectement entre les objets. Le sens global de la composition ne peut être interprété qu'en considérant conjointement les formes des objets et leur positionnement relatif.

Les systèmes informatiques actuels qui présentent des interfaces pour la saisie d'information au moyen d'un stylo électronique (*ordinateurs tablettes, smartphones* et autres *tableaux blancs interactifs*) ou directement avec le doigt (dispositifs à écrans et surfaces tactiles) essaient de mettre à profit l'efficacité de ce mode de communication pour simplifier la vie de l'utilisateur. Le développement de ce mode d'interaction entre l'homme et la machine est accompagné de méthodes automatiques toujours plus avancées pour reconnaître le texte, transcrire les expressions, interpréter les schémas ou indexer le contenu des notes saisies « à main levée ».

Les recherches présentées dans ce document s'inscrivent dans le contexte du traitement automatique de *documents manuscrits* et se concentrent sur le problème de la description des relations spatiales entre les objets qui les composent. Nous partons du constat que l'information de positionnement relatif entre les objets manuscrits est souvent décrite d'une

façon réductrice et simplificatrice qui ne l'exploite pas dans toute sa richesse. Notre hypothèse de travail suppose qu'il est possible de concevoir une modélisation plus complète, riche et générique de cette source d'information importante et nécessaire à la compréhension du document manuscrit. Aussi, le premier aspect de nos travaux consiste à proposer une nouvelle modélisation générique et précise pour la description de relations spatiales entre deux objets manuscrits. Le second aspect de nos recherches se focalise sur la description de la structure spatiale de formes manuscrites vues comme une composition d'éléments de tracés agencés selon des relations de positionnement. Par l'application de cette représentation à une tâche de reconnaissance de symboles, nous démontrerons que l'information de positionnement spatial revêt effectivement une importance majeure pour l'interprétation de compositions manuscrites et qu'il est possible de mieux tirer profit de cette source d'information à condition que la modélisation que l'on en fait permette d'en préserver la richesse. Enfin, on pourra conclure que la modélisation proposée pour les relations spatiales peut contribuer efficacement à l'amélioration des méthodes existantes de traitement automatique du document manuscrit en participant à combler le déficit qu'elles accusent dans la qualité de description du positionnement relatif des objets.

L'analyse automatique de documents manuscrits a fait l'objet de très nombreux travaux de recherche qui ont abouti à la conception de quantité d'applications performantes portant sur des problèmes de plus en plus complexes : reconnaissance de chiffres isolés, vérification de montants sur chèques bancaires, tri automatique de courrier postal, reconnaissance de l'écriture sur interface de saisie de texte libre, transcription automatisée d'expressions mathématiques, composition assistée de schémas techniques complexes, indexation du contenu de prises de notes. . . Dans ces différents contextes, l'information sur le positionnement relatif des objets est largement utilisée, implicitement ou explicitement. Pourtant, il n'existe pas de description généralisée pour exprimer « *quelle est la relation de positionnement entre les objets A et B ?* ». Cette question est difficile à résoudre d'une façon générique pour plusieurs raisons. D'abord, les objets peuvent être de natures diverses et avoir un niveau de complexité graphique très différent entre les situations (du simple segment au schéma complexe, de la lettre à la phrase, du symbole à l'expression mathématique. . .). Ensuite, les relations de positionnement elles-mêmes sont des notions abstraites, assez difficiles à définir (même pour l'homme), floues par nature et source d'ambiguïtés. L'imprécision et l'ambiguïté potentielle des relations de positionnement sont d'autant plus sensibles que dans le cas qui nous intéresse les objets eux-mêmes sont soumis aux variations et aux imprécisions inhérentes à leur nature manuscrite. Enfin, la forme exacte des objets à positionner doit être prise en compte précisément dans la description des relations spatiales, car elle affecte profondément la perception que l'on a de leur positionnement relatif (nous mettrons en évidence cette propriété dans la partie 2.1).

En pratique, le problème est plutôt abordé en relevant des mesures sur la position relative des objets. Ces mesures, ou *descripteurs de positionnement* sont des indices choisis au cas par cas pour caractériser au mieux les relations de positionnement relatif en fonction des besoins de l'application visée. Ces descripteurs reposent souvent sur une représentation simplifiée des formes des objets et ils n'offrent qu'une description réductrice de la relation spatiale, optimisée pour le problème à résoudre. Ils ne constituent pas une façon de modéliser réellement une relation spatiale de façon générique c'est-à-dire reproductible sur des objets

quelle que soit leur nature, leur niveau de complexité ou la particularité de leur forme.

Dans cette thèse, nous poursuivons l'objectif plus ambitieux de définir une modélisation générique des relations spatiales entre objets manuscrits, dans le but d'aider à la reconnaissance automatique de documents.

Des études de psychologie conduites par Logan & Sadler [LS99] suggèrent que l'appréhension par le cerveau humain de relations de positionnement passe par l'application de structures appelées *modèles spatiaux* (*spatial templates*) dans l'espace de perception. Un *modèle spatial* est un modèle centré sur un objet perçu (la *référence* de la relation de positionnement) qui décrit dans l'image des zones de préférences de positionnement selon la relation spatiale qu'il définit.

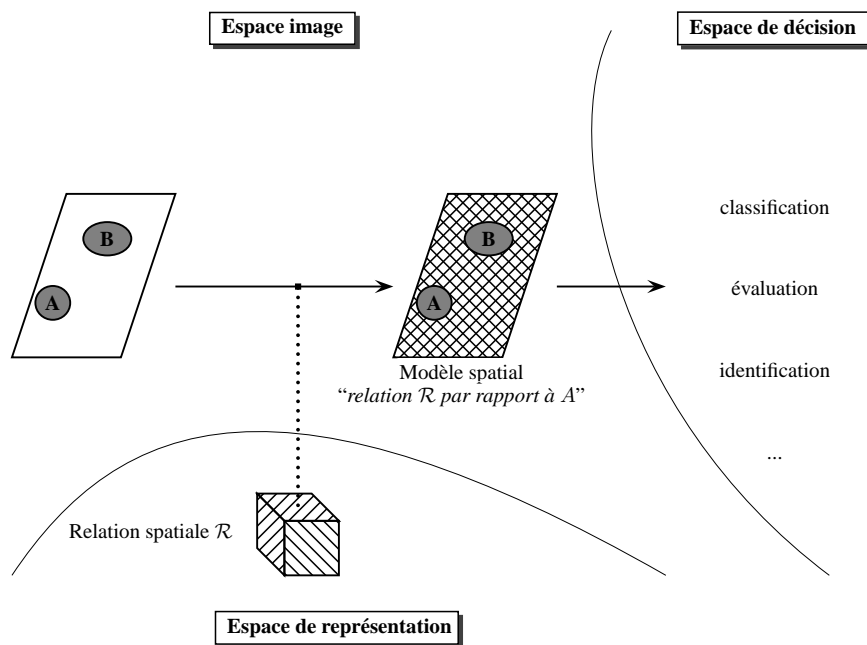


FIGURE 1.1: Une relation spatiale \mathcal{R} peut être appliquée au sein même de l'image et centrée sur un objet qui joue le rôle de référence de positionnement. Il en résulte un *modèle spatial* qui modélise la proposition « être dans la relation \mathcal{R} par rapport à A » (où A est l'objet de l'image pris pour référence). La comparaison de la position de B dans le modèle spatial permet de prendre une décision (classification ou identification de relation spatiale, évaluation de la situation par rapport à la relation spatiale...).

La figure 1.1 illustre très schématiquement ce principe de raisonnement spatial. Une image contenant deux objets A et B est perçue. A partir d'une abstraction de relation spatiale \mathcal{R} , définie dans un certain espace de représentation, il est possible d'appliquer, sur l'image perçue, un *modèle spatial* décrivant cette relation *par rapport* à l'objet A . Ce modèle spatial décrit dans l'image des zones de positionnement plus ou moins favorables par rapport à A en fonction de la relation spatiale \mathcal{R} . L'analyse de la position de B par rapport au modèle spatial permet de prendre une décision sur le positionnement des objets. Plusieurs avantages sont offerts par cette représentation. Tout d'abord, le raisonnement est conduit

intégralement dans l'espace de l'image jusqu'à la prise de décision. Par conséquent, celle-ci peut se faire en ayant connaissance de la situation exacte, sans simplification ni approximation des formes des objets ni de leur positionnement. Ensuite, le modèle spatial constitue une description unifiée permettant de répondre à deux types de problèmes importants pour le raisonnement spatial :

problème d'évaluation : étant donnés deux objets A et B , leur positionnement relatif vérifie-t-il la relation spatiale \mathcal{R} ?

problème de localisation : étant donné un objet A , où doit se situer un objet B positionné relativement à A selon la relation spatiale \mathcal{R} ?

Le problème d'évaluation est souvent la raison d'être des *descripteurs de positionnement* utilisés classiquement pour décrire les relations spatiales. Il s'agit en général d'attribuer un score de correspondance entre la situation perçue, décrite par un ensemble de mesures relevées dans l'image, et un modèle de la relation spatiale \mathcal{R} construit dans le même espace de description. Le problème de localisation est en revanche très largement délaissé par les méthodes de représentation du positionnement relatif. Le fait de pouvoir y répondre confère pourtant aux *modèles spatiaux* une capacité de *prédiction* qui peut être précieuse pour la « lecture » automatique d'expressions composées.

Quelques approches pour l'interprétation d'images mettent en application la théorie cognitive des modèles spatiaux pour la description de relations spatiales *linguistiques*. Il s'agit de cas où le concept de positionnement \mathcal{R} peut être décrit simplement par des mots : par exemple lorsque \mathcal{R} représente la relation *être à droite de*, ou *être proche de*. Pour le traitement des relations spatiales manuscrites, les relations linguistiques et définies empiriquement sont insuffisantes, car la plupart des relations de positionnement entre les objets manuscrits ne peuvent pas être simplement décrites. Prenons par exemple la relation spatiale qui décrit le positionnement relatif d'un exposant par rapport au symbole auquel il est attaché dans une notation mathématique. Ce concept de positionnement ne peut pas être défini simplement par une description telle que *être en haut à droite* du symbole de référence. Il est en réalité beaucoup plus riche et renferme de nombreux aspects : notion de distance (proximité du symbole de référence), d'échelle relative (l'exposant est plus petit que le symbole de référence), prise en compte de la forme du symbole de référence (en haut à droite par rapport à quelle partie du symbole de référence ?) et de l'exposant (quelle partie de l'exposant est en haut, à droite ? une partie est-elle complètement en haut ou complètement à droite ?), tolérance de variation pour chacun de ces paramètres... Tous ces aspects sont présents dans la relation spatiale *exposant* telle que nous la concevons et sont mis à contribution lorsque nous interprétons une notation mathématique. Pourtant, il est très difficile de traduire linguistiquement cette connaissance ou d'en extraire une modélisation réaliste. Cet exemple illustre à la fois la multiplicité des informations à prendre en compte dans une description du positionnement relatif, mais aussi la nécessité de pouvoir *apprendre* automatiquement des modèles à partir de données.

Dans nos travaux, nous exploitons ce même cadre théorique pour définir des relations spatiales \mathcal{R} non linguistiques au moyen de *méta-modèles spatiaux* qui sont appris automatiquement à partir de données. Une relation \mathcal{R} est représentée par un **méta-modèle spatial** qui est utilisable selon le schéma décrit à la figure 1.1. Cette vision originale du problème de

description du positionnement relatif consiste à produire une vraie modélisation d'une relation spatiale dans l'absolu, en tant que telle, et en tenant compte de ses multiples aspects. Le principe de raisonnement par modèles spatiaux permet de garantir une représentation de la relation qui est **générique vis-à-vis des objets**, c'est-à-dire qui n'est pas limitée à des objets de certaines natures ou de certains niveaux de complexité graphique, et **générique vis-à-vis des relations de positionnement** à modéliser, ce qui permet de répondre aux différents besoins de modélisation rencontrés dans l'analyse de documents. Par construction, la représentation s'adapte à la situation sur laquelle elle est appliquée et donc **prend en compte la forme particulière des objets**. En choisissant un espace de représentation adéquat, la représentation des méta-modèles doit permettre de **gérer l'imprécision et la variabilité sur le positionnement relatif** constaté dans les tracés manuscrits. Cette souplesse sera offerte dans nos travaux par le choix d'une représentation des méta-modèles et modèles spatiaux dans le **cadre de la théorie des ensembles flous**. La **possibilité d'apprentissage à partir de données** permet de ne pas limiter les méta-modèles aux seules relations linguistiques ou définies empiriquement. Cela autorise de plus à construire des modèles qui tiennent compte de la variabilité du positionnement des objets manuscrits. La capacité à traiter le problème de localisation est une originalité importante apportée par ce choix de représentation. Dotés de cette **capacité de localisation**, les modèles spatiaux appris permettent d'exhiber la région de l'espace où la relation apprise est vérifiée, par rapport à un objet de référence. Nous pourrions faire usage de cette possibilité de *prédiction* pour *découvrir* des objets positionnés selon les modèles. L'application de ce principe constituera donc une méthode originale pour résoudre le problème de *segmentation* des objets, c'est-à-dire la recherche des frontières entre les objets (dont la définition n'est pas connue a priori dans un document manuscrit à interpréter).

La suite de ce document est organisée en chapitres, dont nous présentons ci-dessous la teneur en quelques phrases.

- Le **chapitre 2** dresse un **état de l'art** des méthodes utilisées pour décrire le positionnement relatif d'objets manuscrits dans les techniques de traitement de document. Ce panorama est également élargi aux méthodes utilisées en analyse d'image pour décrire les relations spatiales entre des objets quelconques. Les principes de la théorie des **modèles spatiaux** et les méthodes exploitant cette représentation sont présentés en détail. Le chapitre est conclu par l'exposé plus précis de notre démarche et du positionnement de nos travaux au regard de l'existant.
- Le **chapitre 3** est dédié à l'introduction du principe, de la formalisation et de la mise en œuvre de notre approche de modélisation des relations spatiales par **méta-modèles spatiaux**. Les méta-modèles spatiaux sont dédiés à la description du positionnement d'un objet par rapport à un objet de référence. Le choix est porté sur un formalisme s'inscrivant dans le cadre de la théorie des ensembles flous et son exploitation fait appel à des opérateurs de morphologie mathématique. La capacité d'apprentissage automatique des modèles à partir de données est mise en évidence. Des résultats qualitatifs sont présentés pour illustrer la large expressivité permise par ces méta-modèles (généricité vis-à-vis des relations spatiales modélisées) et leur capacité à traiter des objets de nature, forme et complexité variables (généricité vis-à-vis des objets à positionner).

- Dans le **chapitre 4**, nous exploitons ces méta-modèles spatiaux au sein de **Patrons Structurés Flous**, une représentation purement spatiale des formes manuscrites structurées. Nous cherchons à exploiter au mieux l’information spatiale pour des formes structurées complexes dans lesquelles il faut identifier des références de positionnement et des éléments à positionner (problème de segmentation). Notre démarche consiste à exploiter finement, grâce à des méta-modèles spatiaux, l’information spatiale disponible dans des symboles manuscrits dans le but de mettre en évidence sa richesse et son importance. Cette application permet non seulement de vérifier le rôle fondamental que joue l’information de positionnement spatial pour l’interprétation de symboles, mais aussi de souligner la contribution apportée par les méta-modèles spatiaux pour la description de cette information. Le **pouvoir de prédiction** des méta-modèles est exploité pour concevoir une **stratégie originale de segmentation des tracés**, où la découverte des segments sera **guidée par leur contexte de positionnement**.
- Le **chapitre 5** est consacré à la présentation des **résultats expérimentaux**. Dans un premier temps, la qualité de description offerte par les méta-modèles spatiaux sera validée en comparaison avec des descriptions classiques de positionnement relatif. Ces tests de validation seront conduits sur une tâche de classification de relations spatiales appliquée à des données manuscrites de différentes natures (gestes d’édition, primitives de caractères chinois, lettres cursives et symboles mathématiques). Ensuite, nous évaluerons la qualité de la représentation offerte par les *Patrons Structurés Flous* pour décrire des formes structurées à l’aide de leurs relations spatiales. Ces expérimentations viseront à confirmer l’importance de l’information spatiale dans la perception de différentes natures de symboles et démontreront l’intérêt d’en préserver toute la richesse pour la performance des méthodes de reconnaissance du manuscrit. L’aptitude des méta-modèles spatiaux à modéliser cette information avec succès pourra être mise en évidence, ce qui soulignera le rôle clé qu’ils peuvent jouer dans l’amélioration de systèmes structurels pour la reconnaissance d’expressions manuscrites.
- Enfin, le **chapitre 6** dresse le bilan et conclut ces travaux. Forts des résultats obtenus qui mettent en évidence l’importance du rôle de l’information spatiale, nous esquissons des perspectives pour renforcer sa contribution aux systèmes d’analyse de documents manuscrits au moyen des méta-modèles spatiaux.

Etat de l'art : description du positionnement relatif

En dépit de l'importance reconnue des informations de positionnement relatif entre les éléments constituant un document pour son analyse et son interprétation, bien peu de travaux se sont concentrés sur l'élaboration d'une technique permettant de les modéliser avec à la fois une large expressivité (pouvoir de décrire différentes catégories de relations spatiales) et une capacité à traiter des objets manuscrits de différentes natures et de niveaux de complexité variables : lettres de l'alphabet, caractères asiatiques, chiffres, mots, phrases, mais aussi symboles, expressions, schémas ou autres formes structurées complexes.

A l'inverse, il existe dans le champ du traitement de documents une abondance de méthodes pour la description et la modélisation du positionnement qui sont dédiées à un type d'objets donné, ou à une catégorie de relations spatiales (relations de distance, directionnelles, ou topologiques). Ce sont pour la plupart des méthodes *ad hoc*, qui sont motivées par un objectif applicatif précis : reconnaissance d'expressions mathématiques, reconnaissance de symboles multi-tracés, reconnaissance de mots manuscrits...

Pour trouver des méthodes plus génériques, il convient de se tourner vers le domaine de l'analyse et l'interprétation d'images au sens large. Au-delà même du problème général de l'interprétation de scènes naturelles, diverses disciplines faisant appel à des représentations de type image nécessitent une analyse poussée des relations spatiales entre des objets. On peut citer notamment l'interprétation d'images médicales, l'analyse d'images satellite, les systèmes d'information géographique ou encore le problème de navigation autonome en robotique.

Dans cette partie, nous allons présenter un éventail des différentes méthodes utilisées pour décrire le positionnement d'objets. Nous proposons pour cela de faire une distinction originale des méthodes en deux grandes catégories.

La première catégorie regroupe les approches qui consistent à extraire des descripteurs et à traiter ces informations spatiales dans un espace de représentation distinct de l'espace de perception (l'image). Ces descripteurs, souvent multidimensionnels, peuvent être utilisés

pour qualifier, catégoriser, identifier ou comparer des relations spatiales. Ils peuvent également répondre à des besoins de classification ou être intégrés à un processus de prise de décision à plus haut niveau. Ce qui caractérise ces approches est l'extraction d'une information spatiale qui est réduite au seul besoin de l'application visée et qui est optimisée pour la tâche voulue (par exemple la discrimination entre différentes situations de positionnement relatif). De plus, ces méthodes reposent souvent sur une représentation simplificatrice des formes des objets, en les assimilant par exemple à des rectangles, des segments ou même à un seul point, en fonction de l'objectif visé. En résumé, ces approches n'ont pas vocation à offrir une modélisation générique des relations spatiales, mais plutôt à extraire des indices sur le positionnement relatif des objets pour contribuer à une interprétation globale.

La seconde catégorie regroupe des approches qui permettent d'appréhender les relations spatiales au sein même de l'espace image. Cette démarche, qui s'appuie sur des études sur la perception humaine des relations spatiales et le raisonnement dans l'espace, se veut par conséquent plus fidèle au processus cognitif d'appréhension des relations spatiales. Elle permet ainsi de conduire un raisonnement spatial exclusivement dans l'espace de l'image, de sorte que l'utilisation ultime de l'information spatiale se fait en ayant connaissance de tout le contexte perceptif : les formes exactes (non simplifiées) des objets mis en relation et leur positionnement relatif précis.

Nous mettrons en évidence les avantages qu'offrent ces approches dites *spatiales* vis-à-vis des objectifs énoncés en introduction de cette thèse. Nos travaux s'inscriront dans cette démarche spatiale pour concevoir une méthode générique de modélisation du positionnement relatif d'objets manuscrits (voir chapitre 3).

2.1 Représentation par descripteurs

Cette catégorie d'approches est la plus largement utilisée pour la description du positionnement d'objets. La démarche consiste à extraire, à partir d'objets définis dans le repère de l'image, un ou plusieurs descripteurs de leur positionnement spatial, qui peuvent ensuite être exploités pour résoudre différents problèmes. L'espace de représentation étant déconnecté de l'espace image, ces descripteurs sont traités comme des *caractéristiques* de la relation décrite et peuvent être traités de la même façon que toute autre nature de descripteurs (par exemple des caractéristiques de formes, d'intensité de niveaux de gris, de vitesse de tracé dans le cas de données *en-ligne*...). De ce fait, les possibilités d'apprentissage automatique de modèles ou de classifieurs sont très vastes.

Les méthodes qui procèdent par extraction de descripteurs n'ont pas pour objectif de modéliser de façon générique toute relation spatiale entre des objets quelconques dans le but d'être exploité dans le cadre d'un raisonnement spatial au sein de l'espace de perception. Plus simplement, elles visent à définir des modèles pour *vérifier* la valeur d'une ou plusieurs mesures sur le positionnement spatial d'objets ou le respect de contraintes de positionnement entre des objets. Les descripteurs extraits sont choisis en fonction de l'objectif visé par l'application. Ils décrivent certains indices sur le positionnement entre les objets, qui sont sélectionnés en fonction des besoins : mesures d'angles, de distances, de degrés de recouvrement... Les formes des objets eux-mêmes sont souvent approximées pour simplifier la représentation de leur positionnement relatif.

La figure 2.1 illustre la démarche générale des approches par descripteurs. *L'espace de*

2.1. Représentation par descripteurs

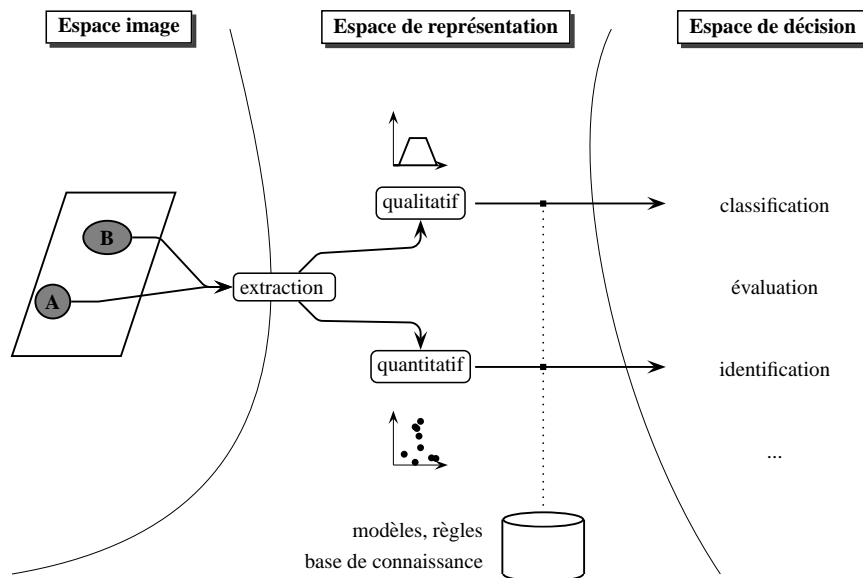


FIGURE 2.1: Schéma général des approches de description du positionnement relatif par extraction de descripteurs.

représentation, dans lequel sont décrits les objets ainsi que les modèles, est distinct de *l'espace image*, où sont définis les objets d'intérêt. Les descripteurs extraits peuvent être de nature quantitative et sont alors susceptibles d'être comparés entre eux ou bien confrontés à un modèle prédéfini (qui peut être issu d'une connaissance experte ou appris à partir de données représentatives). Alternativement, des descripteurs de nature qualitative peuvent être extraits, qui visent alors à qualifier la nature de la relation spatiale, souvent avec un degré de confiance associé. Ils peuvent également nourrir un processus de prise de décision de plus haut niveau. *L'espace de décision* est un terme générique qui peut prendre différentes formes en fonction de l'objectif recherché par l'application. Par exemple, si le but est de classifier la relation spatiale entre des objets de l'image (c'est-à-dire d'étiqueter la relation spatiale), l'espace de décision est l'ensemble des relations spatiales connues (l'ensemble des étiquettes disponibles). La décision peut aussi consister à évaluer le positionnement d'objets par rapport à une relation prédéfinie (la décision consiste alors par exemple en un degré d'adéquation entre 0 et 1). Enfin, la prise de décision fait parfois intervenir d'autres natures d'informations et de connaissances (par exemple relatives à la forme des objets) pour d'autres applications (interprétation de scène, reconnaissance de formes composées...).

Comme nous l'avons évoqué dans l'introduction générale, la description du positionnement au sein de compositions manuscrites peut intervenir à différents niveaux de complexité graphique. Dans ce qui suit, nous faisons une distinction entre les méthodes de positionnement selon qu'elles sont dédiées aux relations spatiales entre *primitives de tracé* ou entre *objets multi-tracés*. Comme nous le verrons, ces différences de niveau induisent des catégories de relations modélisées très différentes, ainsi que des représentations différentes des objets manipulés.

Nous présentons donc d'abord les techniques opérant au niveau des primitives graphiques

simples, avant d'aborder le niveau des relations entre symboles (par exemple au sein d'une expression). Enfin, nous concluons cette section par une incursion dans le domaine de l'analyse d'images, où des descripteurs plus génériques ont été proposés pour la représentation du positionnement d'objets de formes quelconques.

2.1.1 Positionnement de primitives manuscrites

Dans les approches dites *structurelles* pour l'analyse de tracés manuscrits, la description repose sur une décomposition de ceux-ci en éléments indivisibles (appelés par la suite *primitives structurelles*, ou plus simplement *primitives*). La définition des primitives varie d'une méthode à l'autre : sous-tracé (assimilable à un segment de droite, ou délimité par des points singuliers), composante connexe dans un signal hors-ligne, tracé délimité par un poser et un lever de stylo dans un signal en-ligne... Le point commun entre ces approches est la nécessité de décrire, conjointement aux formes de ces primitives, l'information sur leur positionnement au sein d'une forme globale. Ici, il est réellement tiré parti de la nature particulière des objets, qui sont des primitives de tracés et non pas des objets quelconques. Les primitives de tracés sont des formes linéiques par nature (elles n'ont pas d'épaisseur). Elles peuvent donc être approximées par des éléments graphiques tels que des lignes polygonales, des segments, des arcs de cercles... Les relations spatiales qui existent entre ces objets font intervenir des informations de direction, de distance, mais aussi de topologie (intersection, adjacence) ou encore de géométrie (parallélisme, perpendicularité, tangence). Nous présentons d'abord les approches quantitatives pour le positionnement de primitives de tracés, puis les approches qualitatives.

2.1.1.1 Approches quantitatives

La majorité des techniques de positionnement par descripteurs quantitatifs décrivent les positions des primitives en termes de directions, d'angles et de distances. Dans la plupart des travaux adoptant ce type d'approche, l'information de positionnement est mesurée de façon relative entre les différentes primitives, par exemple en mesurant simplement des vecteurs de déplacement entre leurs extrémités.

Dans leurs travaux, Marukatat & Artières [MA04] présentent deux stratégies pour intégrer de l'information sur le positionnement relatif de primitives au sein d'un système de reconnaissance basé sur des Modèles de Markov Cachés (MMC). Pour la reconnaissance d'un caractère, la probabilité a posteriori d'appartenir à une classe donnée selon le modèle appris est décomposée en un produit de deux probabilités, l'une liée à la forme des tracés primitifs, l'autre liée à leur positionnement spatial.

Deux stratégies sont comparées pour le calcul de cette probabilité de positionnement. La première repose sur une mesure du positionnement de chaque primitive par rapport à une référence absolue, définie comme étant la *boîte englobante* normalisée de tout le tracé. La notion de boîte englobante désigne le plus petit rectangle aux côtés parallèles aux axes du repère de l'espace image qui contient tout le tracé. Cette notion est souvent utilisée pour représenter simplement la position et les dimensions d'un objet. Nous mettrons en évidence au paragraphe 2.1.2.3 les inconvénients des représentations de relations spatiales reposant sur les boîtes englobantes des objets.

2.1. Représentation par descripteurs

Dans le cas des travaux de Marukatat & Artières [MA04], la position du centre de la boîte englobante de chaque primitive est modélisée dans le référentiel (défini par la boîte englobante du tracé global) par une loi de probabilité gaussienne dont les paramètres sont estimés sur une base d'apprentissage. Les probabilités de positionnement sont considérées comme indépendantes pour les différentes primitives et peuvent être simplement combinées pour constituer le score global de positionnement. La figure 2.2 représente un symbole (le chiffre *sept*) décomposé en trois primitives notées s_1 , s_2 , et s_3 dont les centres des boîtes englobantes sont mis en évidence.

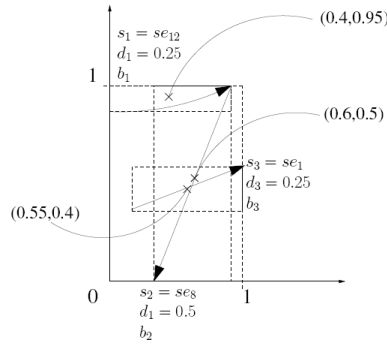


FIGURE 2.2: Dans les travaux Marukatat & Artières [MA04], la position des trois primitives constituant le chiffre *sept* est décrite par la position du centre de leur boîte englobante dans le repère défini par la boîte englobante globale normalisée. Les coordonnées des centres des boîtes englobantes de chaque primitive sont représentées ici.

Le fait d'assimiler un tracé à sa boîte englobante, ou à un seul point (ici son centre), est une pratique couramment utilisée pour simplifier la description du positionnement relatif de tracés. En général, ce choix revient à faire une simplification très grossière qui peut conduire à une description contre-intuitive, comme nous allons le montrer au paragraphe 2.1.2.3. Dans le cas présent, il convient toutefois de remarquer que chaque primitive est décrite par trois informations qui la déterminent assez précisément : il s'agit de son type (qui décrit sa courbure et sa direction), de sa longueur et des coordonnées de sa boîte englobante. Cette bonne description limite dans ce cas les effets négatifs de l'approximation.

La seconde stratégie proposée par les auteurs consiste à décrire le positionnement *relatif* des primitives, c'est-à-dire les unes par rapport aux autres. La description relative peut se faire tout simplement en modélisant les vecteurs reliant les extrémités des primitives. Dans les stratégies à base de MMC, les liens entre primitives sont souvent décrits par une transition dans le modèle stochastique, qui représente le *saut* éventuel fait par le stylo entre le tracé de deux primitives. Marukatat & Artières choisissent plutôt de décrire ces liens à l'aide de six attributs de positions relatives des boîtes englobantes des tracés. Les six mesures proposées correspondent à la proportion de la boîte englobante d'un tracé qui se trouve *au-dessus*, *aligné verticalement*, *en-dessous*, et à *gauche*, *aligné horizontalement* et à *droite* par rapport à la boîte englobante de l'autre tracé. Un seuil fixé à 80% des dimensions de la boîte englobante de référence détermine la frontière entre les zones considérées comme *alignées* ou non. La figure 2.3, extraite de [Mar04], illustre le calcul des trois descripteurs

de positionnement relatif vertical pour un tracé b_s par rapport à un autre tracé b_r , où les tracés sont représentés par leurs boîtes englobantes.

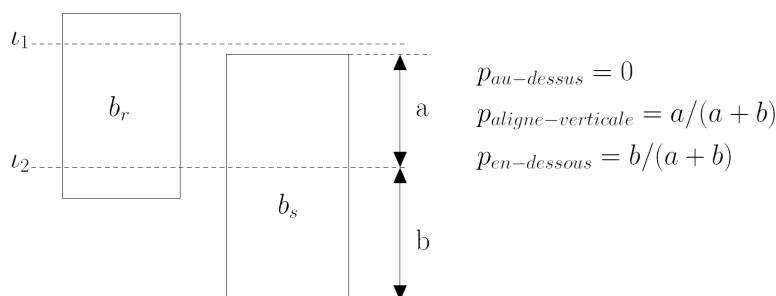


FIGURE 2.3: Calcul des trois caractéristiques d'alignement vertical pour décrire le positionnement relatif d'une primitive de tracé b_s par rapport à une autre primitive b_r . Les tracés sont représentés par leur boîte englobante (d'après [Mar04]).

Les caractéristiques de positionnement relatif calculées entre les primitives du chiffre *sept* présenté à la figure 2.2 sont regroupées dans le tableau 2.1, tiré de la thèse de Marukatat [Mar04].

TABLE 2.1: Valeurs des caractéristiques de positionnement relatif des primitives de tracé du chiffre *sept* de la figure 2.3. Chaque ligne présente les descripteurs extraits pour le positionnement d'une primitive cible relativement à une primitive de référence. Les trois premières colonnes regroupent les valeurs des caractéristiques d'alignement vertical, et les trois suivantes celles d'alignement horizontal. Les valeurs sont reportées d'après [Mar04].

référence	argument	$p_{au-dessus}$	p_{al-v}	$p_{en-dessous}$	p_{gauche}	p_{al-h}	p_{droite}
s_1	s_2	0	0,08	0,92	0	0,1	0,9
s_1	s_3	0	0	1	0	0,2	0,8
s_2	s_3	0	1	0	0,45	0,2	0,35

Les résultats comparatifs obtenus (voir les détails dans [Mar04]) montrent que si la description absolue du positionnement est la plus stable pour des symboles simples (lettres ou chiffres) et conduit aux meilleures performances, c'est en revanche la description relative du positionnement qui est la plus efficace lorsque les symboles décrits sont plus complexes (c'est particulièrement notable sur une base de caractères coréens, qui sont constitués d'un grand nombre de primitives). Une seconde remarque intéressante est que les descriptions absolue et relative du positionnement ne sont pas redondantes et que c'est la combinaison des deux qui permet d'obtenir les meilleurs résultats, quel que soit le niveau de complexité des symboles.

Dans des travaux connexes, Sicard & Artières [SA09] s'intéressent spécifiquement à l'apprentissage de modèles relationnels entre paires d'observations d'une séquence modélisée par

2.1. Représentation par descripteurs

un Modèle de Markov Caché (alors que la majorité des approches à base de MMC supposent l'indépendance entre les observations). Les paires d'observations sont reliées dans ce modèle par des probabilités d'émission de caractéristiques différentielles (dites *delta features*). Ces probabilités sont complètement intégrées au modèle et un paramètre permet de contrôler leur importance par rapport aux probabilités d'émission des observations, ce qui permet de passer d'un modèle purement local (sans prise en compte des relations spatiales) à un modèle purement relationnel (où seules les relations spatiales sont prises en compte). Les résultats tendent à prouver l'importance des caractéristiques relationnelles pour une tâche de reconnaissance de chiffres manuscrits en-ligne. La figure 2.4, extraite de la référence [SA09], représente les distributions apprises sur les caractéristiques différentielles entre les paires de primitives du chiffre *sept*. Pour chaque paire, la distribution est représentée par un vecteur (le vecteur moyen de déplacement) et une ellipse centrée sur l'extrémité du vecteur (représentant la dispersion ou la covariance). Chaque ligne du tableau de la figure 2.4 représente le positionnement des trois primitives par rapport à l'une prise comme référence. Par exemple, dans la troisième case de la première ligne, on peut lire que la troisième primitive est positionnée en dessous, et légèrement à droite par rapport à la première, avec une grande variabilité horizontale et une faible variabilité verticale.

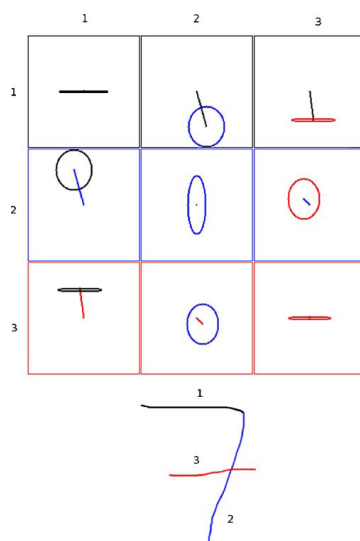


FIGURE 2.4: Visualisation des modèles relationnels appris pour le positionnement relatif des primitives de tracé du chiffre *sept*. Chaque ligne du tableau présente les modèles appris pour les trois primitives par rapport à une même référence. (extrait de [SA09], voir aussi [SAP06]).

D'autres approches basées sur des modèles probabilistes graphiques, tels que des réseaux Bayésiens, des champs de Markov, ou encore des champs aléatoires conditionnels ont émergé, adoptant le même type de description pour le positionnement relatif des éléments graphiques. On peut citer les travaux de Cho *et al.* [CK04] pour la reconnaissance de chiffres, ceux de Zeng *et al.* [ZL08] et Kim *et al.* [KK03] pour la reconnaissance de caractères chinois, ou encore de Kang *et al.* [KK04] pour des caractères coréens. Dans tous ces contextes, les relations sont décrites en termes de distances et de positions relatives dans le repère,

à la manière de la technique présentée plus haut. Il n'est pas porté d'intérêt aux relations topologiques entre primitives de tracé, ni à leurs contraintes géométriques.

Toujours pour la reconnaissance de caractères asiatiques, Shin *et al.* [Shi02] proposent de décrire les relations inter-primitives par les vecteurs relatifs entre leurs extrémités. Le problème de la reconnaissance de caractères chinois est posé comme un problème de mise en correspondance entre les primitives extraites d'un caractère à reconnaître et celles d'un modèle prédéfini. Dans l'algorithme de programmation dynamique utilisé pour rechercher ce meilleur alignement, les auteurs intègrent un coût lié au positionnement relatif des primitives. La description choisie considère quatre vecteurs de déplacements joignant deux à deux les extrémités des primitives. Le coût associé à une relation spatiale tient compte de la différence d'angle et de longueur entre les vecteurs extraits et ceux du modèle. Les quatre vecteurs de positionnement relatif entre primitives sont illustrés dans la figure 2.5, où les primitives sont représentées par des lignes pointillées. Le calcul du coût de mise en correspondance avec un vecteur de positionnement relatif du modèle est également illustré pour les deux vecteurs d_{ss} (qui lient les points de départ des deux primitives). La valeur de l'angle θ et de la distance e déterminent le coût de l'appariement des deux descripteurs de la relation spatiale.

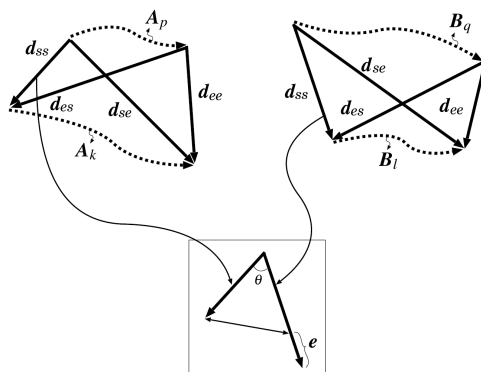


FIGURE 2.5: Le positionnement relatif de deux primitives de caractères chinois A_p et A_k est décrit par quatre vecteurs entre leurs extrémités (d_{ss} , d_{se} , d_{ee} et d_{es}). Le coût d'appariement entre le descripteur d_{ss} d'un symbole à reconnaître et le descripteur d_{ss} d'un prototype structurel fait intervenir la différence d'angle (θ) et de longueur (e) entre les vecteurs. La figure est tirée de [Shi02].

C'est dans les applications de reconnaissance de schématique et de symboles techniques que les relations topologiques et géométriques sont considérées avec un plus grand intérêt. Dans [PR07], Pu & Ramani exploitent un descripteur quantitatif dénommé *Relative Shape Histogram*. Les auteurs se basent sur des histogrammes de distances pour identifier des contraintes géométriques entre éléments de schémas. Dans cette application et à la différence du problème de reconnaissance de caractères, il est mis l'accent sur la définition de descripteurs invariants en rotation et à l'échelle, mais permettant de caractériser l'existence de certaines contraintes géométriques entre les objets. Etant données deux primitives de

2.1. Représentation par descripteurs

tracé, les longueurs de segments liant entre eux des points de ces deux objets sont discrétisées et comptabilisées dans un histogramme dont la forme finale est caractéristique de leur relation géométrique (ainsi que de leur forme). On obtient ainsi une silhouette d'histogramme caractéristique pour un segment tangent à un cercle, une autre pour deux segments parallèles. . . Des exemples de descripteurs extraits sur des primitives dans différentes configurations sont illustrées par la figure 2.6. Pour chaque situation, les deux primitives sont représentées en rouge à gauche de leur histogramme associé. On peut vérifier la stabilité de la silhouette de l'histogramme face à des situations de tracés déformés et subissant diverses rotations. Une fois extraits, ces descripteurs peuvent ensuite être comparés entre eux, ou comparés à des modèles, par exemple pour nourrir par la suite un solveur de contraintes géométriques.

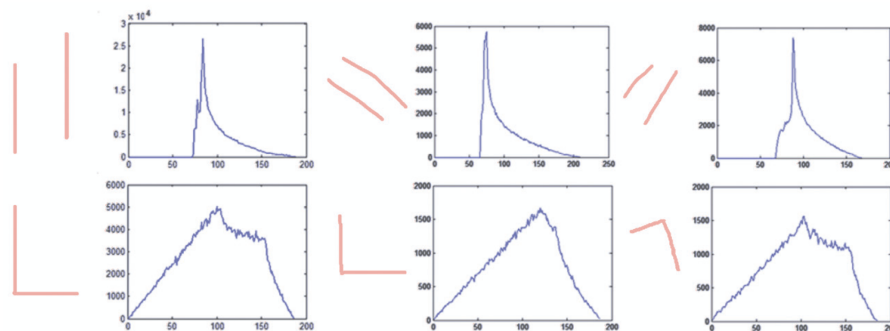


FIGURE 2.6: Description de contraintes géométriques entre tracés par *Relative Shape Histograms*. Les deux situations de parallélisme et de perpendicularité entre des segments manuscrits conduisent à des formes différentes de leur histogramme. Pour une situation donnée, le descripteur est relativement stable face à des variations de forme des tracés et insensible à leur rotation. Les exemples sont tirés de [PR07].

2.1.1.2 Approches qualitatives

Parmi les travaux visant à la reconnaissance de formes manuscrites structurées, il est fréquent de recourir à une représentation des relations entre primitives par des descripteurs qualitatifs. Il s'agit en général de catégoriser la relation spatiale relative de deux primitives parmi une liste de types prédéfinis. Les différents types peuvent se distinguer par des notions topologiques (*à l'intérieur, adjacent, intersectant, connecté*), des contraintes géométriques (*parallèle, perpendiculaire*), des informations de directions (*à droite, au-dessus*) ou de distance (*proche, lointain*). Un degré de confiance est généralement associé à la catégorisation d'une relation inter-primitives, ce qui permet de tenir compte de l'incertitude possible sur cette catégorisation. L'ambiguïté entre les catégories est fréquente et il peut également être souhaitable de pouvoir attribuer plusieurs catégories à la relation entre deux primitives.

Pour la reconnaissance de caractères chinois, plusieurs méthodes proposées procèdent à une catégorisation des relations spatiales entre paires de primitives. Ces méthodes s'appuient souvent sur une représentation du caractère par un graphe dans lequel les sommets

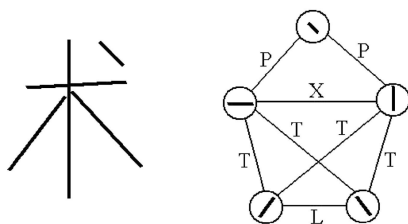


FIGURE 2.7: Représentation d'un caractère chinois (à gauche) par un graphe (à droite). Les sommets du graphe correspondent aux primitives du tracé du caractère et ses arcs décrivent leurs relations spatiales. Celles-ci sont catégorisées en 5 types de relations. Image reprise de [LJN04].

décrivent les primitives de la structure et les arcs décrivent les relations entre ces primitives. Un exemple simple de ce type de représentation est illustré à la figure 2.7, tirée de la référence [LJN04]. Cette illustration représente un caractère chinois constitué de 5 primitives (à gauche) et un graphe qui décrit sa structure (à droite). Chaque sommet du graphe correspond à une des primitives du tracé et ses arcs décrivent les relations spatiales entre primitives. Dans ce cas, les catégories de relations spatiales sont : positionnement distant (P), intersection (X), adjacence (T) ou coïncidence entre extrémités (L).

Dans les travaux de Liu, Kim *et al.* [LKK01], les relations spatiales entre primitives de tracé sont catégorisées comme appartenant à l'une des douze catégories présentées par la figure 2.8 (les primitives sont réduites à de simples segments). Les relations R_1 à R_8 sont toutes des variantes de la relation topologique d'*adjacence*, mais se distinguent par le sens de tracé et la position du point de contact entre les deux primitives, car ce sont des caractéristiques importantes dans l'écriture de caractères chinois. Pour permettre une stabilité dans la définition des catégories, des seuils de tolérances sont définis, par exemple pour décider si deux primitives sont adjacentes (catégories R_1 à R_8) ou s'intersectent (R_{12}). Les primitives sont assimilées à leur point central si elles sont de petite dimension et suffisamment éloignées (relation R_{11}).

Les approches à base de graphes de relations spatiales abondent dans les méthodes de reconnaissance de caractères asiatiques. Dans les travaux de Liu, Cham *et al.* [LCC96], par exemple, les relations entre segments sont décrites par trois attributs qualitatifs : la position verticale relative de centres (qui peut être *au-dessus*, *en-dessous*, ou *indéterminée*), leur position relative horizontale (*à gauche*, *à droite* ou *indéterminée*), et un attribut d'intersection qui peut être activé ou non. Certaines approches se distinguent par leur gestion de l'imprécision, en mettant en œuvre une description floue des attributs, comme c'est le cas dans les graphes relationnels attribués flous (FARG) présentés par Chan & Cheung [CC92] ou encore par Zheng, Ding *et al.* [ZDW97]. La modélisation floue des relations spatiales permet de gérer l'imprécision sur les tracés et donc l'incertitude qui en résulte sur leur catégorisation, en associant aux attributs qualitatifs des degrés de confiance. Par exemple, dans [CC92], les relations spatiales sont décrites au moyen de trois attributs décrivant leur positionnement relatif vertical, horizontal, et leur type de jointure. Le positionnement relatif vertical est décrit par plusieurs sous-ensembles flous non exclusifs : *au-dessus*, *alignés*, *en-dessous*, et une relation spatiale est donc décrite par rapport à ces trois typicalités, avec des degrés

2.1. Représentation par descripteurs

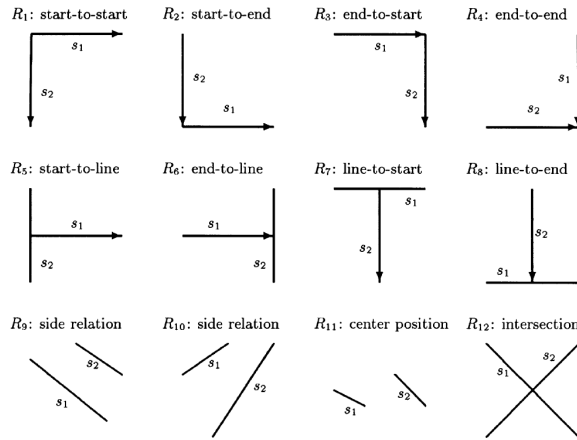


FIGURE 2.8: Illustration des 12 catégories de relations spatiales entre les primitives de tracés composant les caractères chinois selon [LKK01].

d'adéquation divers. De la même façon, le type de jointure décrit peut être : *intersection* avec une confiance de 0,7, *adjacence* avec une confiance de 0,2 et *parallèle* avec un degré de 0.

La démarche suivie par Lee, Kara *et al.* [LBKS07] pour la modélisation de symboles multitracés par graphes relationnels à attributs est très similaire. Les primitives sont catégorisées en deux types : arcs de cercle et segments. Les relations spatiales entre deux primitives sont catégorisées en fonction du nombre de leurs intersections, qui peut varier de 0 (si les primitives ne s'intersectent pas) à 2 (si les primitives présentent deux points d'intersection, par exemple dans le cas d'un segment qui est la corde d'un arc). Le descripteur est affiné par des mesures quantitatives de l'angle entre les primitives lorsqu'elles s'intersectent. Les mesures des positions des intersections sur les primitives sont aussi intégrées au descripteur.

La représentation structurelle adoptée par Xu, Sun *et al.* dans [XZB⁺04] est aussi basée sur des graphes, dénommés graphes de relations spatiales, où les noeuds et les arêtes ont la même interprétation que dans l'approche de Lee, Kara *et al.*. Dans ces travaux, les relations spatiales entre deux primitives sont classées dans une catégorie parmi : *connexion d'extrémités*, *tangence*, *intersection*, *parallélisme*, *concentricité*. Il est à noter que dans ce cas, la catégorie attribuée à une relation entre deux primitives n'est pas évaluée par un coût, mais qu'elle peut être multiple en présence d'ambiguïtés. Pour les relations de parallélisme, de tangence et de concentricité, une catégorie de positionnement directionnel est combinée à la catégorisation pour éliminer les ambiguïtés persistantes. Il s'agit notamment de permettre la distinction entre des cas de symboles symétriques par exemple.

Les travaux de Mas, Sanchez *et al.* [MSL05] sont représentatifs des approches utilisant les descripteurs qualitatifs pour la description et la reconnaissance automatique de symboles structurés. Une grammaire d'adjacence est exploitée pour décrire les symboles à partir d'un ensemble de primitives et de relations entre elles. Un procédé de pré-traitement consistant

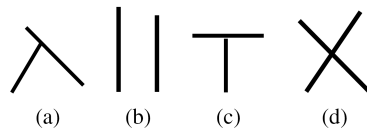


FIGURE 2.9: Illustration des catégories de relations spatiales entre les primitives de symboles architecturaux, définies dans [MSLL07]. Situations d'adjacence (a), de parallélisme (b), de perpendicularité (c) et d'intersection (d).

en une approximation polygonale du tracé permet de définir un ensemble d'éléments qui sont catégorisés comme segments ou arcs et qui constituent les primitives de la représentation structurelle. Pour toutes les paires de primitives extraites prises deux à deux, des contraintes géométriques sont examinées et évaluées avec un degré d'insatisfaction entre 0 et 1 : *parallélisme*, *perpendicularité*, *incidence*, *adjacence*, *intersection*. Ces degrés sont ensuite intégrés au processus de reconnaissance sous la forme de coûts structurels attachés aux règles de production de la grammaire. Dans une version plus aboutie de ces travaux, les auteurs constatent l'existence de différentes natures d'ambiguïtés possibles sous ce formalisme. Tout d'abord, certains symboles graphiquement différents peuvent être décrits au moyen des mêmes règles de production et des mêmes contraintes géométriques. Ensuite, deux symboles topologiquement identiques et décrits par les mêmes règles de production sont indifférenciables par cette description, même lorsque leurs proportions géométriques sont très différentes (par exemple un carré et un rectangle très allongé). Afin de remédier à ces insuffisances, il est proposé d'enrichir la description de contraintes spatiales en y intégrant des mesures quantitatives d'angle, de dimensions relatives et de distance entre les primitives [MLSL06].

Cette solution revient en fait à construire des descripteurs hybrides qualitatifs-quantitatifs. On retrouve cette même idée dans d'autres travaux relatifs à la description de relations entre primitives dans des symboles. Ainsi, Li, Qiu *et al.* [LSQW09] s'intéressent à la définition d'un descripteur pour les relations entre segments de droite. L'application présentée est identique à celle de Mas, Sanchez *et al.* [MSL05] et concerne les mêmes données. Dans ce cas, les auteurs proposent un modèle qualitatif à quatre types de relations définies par la satisfaction ou non de deux critères : parallélisme et intersection. Le type de relation est moins précis que dans [MSL05], puisque par exemple les catégories *incidence* et *intersection* sont vues comme deux cas de la catégorie *intersection et non parallélisme*, et la *perpendicularité* est soit vue comme un cas particulier d'*intersection*, soit n'est pas prise en compte et devient une relation *sans intersection, ni parallélisme*. Cependant, les auteurs enrichissent leur descripteur d'un plus grand nombre d'attributs quantitatifs. Comme dans la méthode précédente, les caractéristiques quantitatives intégrées au descripteur sont des mesures de rapport des longueurs, d'angles et de distances tenant compte des points milieux des segments ou du point d'intersection lorsqu'il existe.

Etant donné le grand nombre de mesures quantitatives intégrées au descripteur de Li, Qiu *et al.*, ainsi que, dans une moindre mesure, dans ceux de Mas, Sanchez *et al.*, on peut

2.1. Représentation par descripteurs

légitimement se poser la question de sa nature : peut-on encore le voir comme un descripteur qualitatif? La réponse à cette question tient dans la prépondérance de la catégorie attribuée à une relation (information qualitative) par rapport aux autres attributs du descripteur (information quantitative) : c'est en fonction de la catégorie d'un descripteur que certaines mesures lui sont intégrées plutôt que d'autres. Toutes les relations spatiales ne sont donc pas représentées dans le même espace et l'information quantitative ne peut être utilisée que pour comparer des relations dans le même espace, c'est-à-dire des relations identifiées comme appartenant à la même catégorie. On peut donc considérer que ces descripteurs sont de nature qualitative avant tout, même s'ils sont hybridés avec d'importantes informations quantitatives.

2.1.1.3 Synthèse

Pour le positionnement relatif de primitives de tracés, on a pu relever différents types de descripteurs visant à distinguer différentes relations spatiales d'intérêt, en fonction des applications. Pour les besoins de reconnaissance de caractères, les relations inter-primitives modélisées combinent souvent informations de positionnement directionnel (à gauche, à droite, ...), de distance et de topologie (intersections, adjacences). C'est dans la reconnaissance de caractères complexes, tels que les caractères chinois, que ces modèles sont les plus évolués afin de distinguer un plus grand nombre de situations entre des paires de primitives. Pour les applications d'interprétation de schémas, les modèles s'intéressent davantage aux contraintes géométriques entre les primitives de tracés.

En ce qui concerne la représentation des objets, elle est souvent assez précise et fidèle au tracé brut grâce aux approximations par des segments (approximation polygonale) ou des arcs de cercle qui se prêtent bien à la représentation d'objets linéiques. De nombreuses méthodes quantitatives utilisent directement les points des tracés pour extraire les descripteurs (par exemple via les vecteurs entre les points extrémités des primitives). Quelques approches se basent au contraire sur une vision plus grossière des tracés, au moyen de leur boîte englobante. Nous reviendrons au paragraphe 2.1.2.3 sur les problèmes posés par la représentation d'objets au moyen de leur boîte englobante.

Les exemples de descripteurs vus jusqu'ici illustrent la multiplicité des besoins de description : même en se limitant aux relations entre des primitives de tracé de forme simple, on relève une large variété dans la nature des relations spatiales décrites. On peut aussi relever que malgré cette simplicité de forme des objets, leurs relations spatiales ne sont pas faciles à décrire et les situations d'ambiguïté sont fréquentes.

2.1.2 Positionnement d'objets manuscrits complexes

La description du positionnement relatif n'intervient pas uniquement entre primitives constituantes d'un objet manuscrit, mais également entre objets manuscrits complexes eux-mêmes. Ce problème est même très général et se pose dans un grand nombre de contextes applicatifs. On peut notamment citer l'exemple très parlant des applications visant à la transcription automatique d'expressions mathématiques, qui sont des expressions constituées de nombreux symboles dont le positionnement relatif capture une grande partie de l'information sémantique. Dans ce contexte particulier, les relations spatiales jouent un rôle considérable pour l'interprétation globale, puisque certaines relations sémantiques, comme

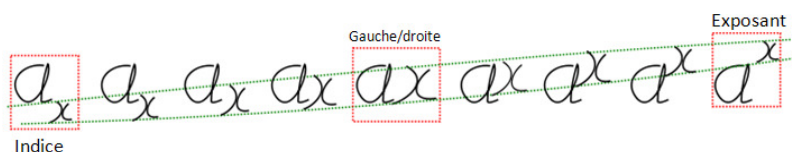


FIGURE 2.10: Illustration de l'ambiguïté des relations de positionnement spatial entre symboles d'une expression mathématique manuscrite (image reprise de [Awa10]).

celle entre un symbole et son exposant (ou son indice), ne sont représentées que par le positionnement relatif des objets.

Plus généralement, la modélisation du positionnement d'objets manuscrits est un enjeu crucial pour l'interprétation d'expression multi-symboles : textes manuscrits en contexte (mots, phrases), expressions mathématiques, représentations de molécules chimiques, schémas divers (organigrammes, schémas techniques, plans architecturaux...). Ce qui distingue la description du positionnement relatif d'objets au sein d'expressions par rapport au positionnement de primitives de tracé au sein d'un objet est d'une part la représentation adoptée pour les objets et d'autre part les types de relations spatiales à décrire. Dans la plupart des cas, les objets sont disjoints (ou du moins leurs tracés ne s'intersectent pas) et les relations modélisées concernent un positionnement relatif en termes de direction et de distance entre des objets approximés par leurs boîtes englobantes. Comme pour les descriptions inter-primitives, on distingue pour le positionnement relatif d'objets les approches quantitatives et les approches qualitatives.

2.1.2.1 Approches quantitatives

Le premier exemple d'applications faisant appel à des descriptions quantitatives entre des objets manuscrits est l'analyse d'expressions mathématiques. C'est dans ce domaine que les questions de description des relations spatiales ont été le plus étudiées [AUAS09] [AMVG10] [ZBC02]. Il est communément admis que de par leur structure bidimensionnelle, les expressions mathématiques sont difficiles à analyser et que de nombreuses ambiguïtés peuvent apparaître dans le positionnement relatif des éléments. La figure 2.10, tirée de la thèse d'Awal [Awa10], illustre ce phénomène avec un exemple simple. Lorsque la position du symbole « x » par rapport au symbole « 2 » varie progressivement entre la position *indice* « idéale » et la position *exposant* « idéale », de nombreuses configurations ambiguës se présentent. Ceci illustre la difficulté d'interpréter la relation entre les deux symboles dans les situations réelles où le caractère manuscrit des objets saisis induit fréquemment une imprécision et une variabilité dans leur positionnement.

Bien qu'ils soient appliqués à des symboles imprimés, les travaux de Aly, Uchida *et al.* [AUAS09] sont représentatifs des approches couramment utilisées pour la classification des relations spatiales entre symboles mathématiques. Leur démarche est d'abord d'attribuer aux symboles un type en fonction de leur étendue verticale par rapport à une ligne de base imaginaire, afin de distinguer les symboles *réguliers* (« a », « c », « e », « e », « x »...), *ascendants* (« b », « d », « k », « > »...), *descendants* (« g », « j »...) et *ascendants-descendants* (« f », « Σ »...). La hauteur des boîtes englobantes des symboles est normalisée en leur ajoutant des éléments ascendants ou descendants virtuels en fonction de leur type.

2.1. Représentation par descripteurs

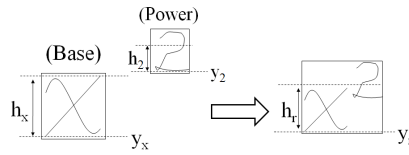


FIGURE 2.11: Extraction de caractéristiques de positionnement relatif entre des objets d’une expression mathématique. La hauteur relative et l’ordonnée du centre des boîtes englobantes des objets sont utilisées pour modéliser la relation spatiale d’exposant. L’image est tirée de [AMVG10].

La relation spatiale entre deux symboles est ensuite décrite par le rapport des hauteurs ainsi que par la distance verticale des centres des boîtes englobantes. Ces deux caractéristiques simples suffisent à discriminer très précisément les types de relations entre symboles dans des expressions mathématiques imprimées, en utilisant un moteur de reconnaissance dédié pour chaque type de symbole.

Dans le cas de tracés manuscrits, une démarche similaire est adoptée par Awal, Mouchère *et al.* [AMVG10]. Une information de hauteur normalisée par le type de symbole est extraite pour chaque symbole mis en jeu dans la relation, ainsi que la différence de la position verticale de leur centre. Des modèles gaussiens dédiés à chaque type de relation permettent de donner un coût probabiliste aux hypothèses de relations spatiales au cours de l’analyse structurelle. Un exemple de mesure des caractéristiques de positionnement extraites selon cette méthode est illustré par la figure 2.11.

Les caractères chinois peuvent être vus comme un assemblage de tracés primitifs agencés entre eux selon des relations de positionnement et de topologie, comme nous l’avons expliqué précédemment. Le nombre de ces primitives peut être assez grand, dépassant fréquemment la dizaine. En réalité, les caractères chinois complexes sont composés de plusieurs symboles de niveau intermédiaire, appelés *radicaux*, qui sont eux-mêmes constitués de primitives de tracés. Cette structuration est un élément fondamental de la langue chinoise. Tout caractère chinois peut être décrit comme un arbre de composition de ses radicaux. La figure 2.12(a) représente un caractère chinois (imprimé) constitué de trois radicaux. Alors qu’il existe des milliers de caractères chinois, le nombre de radicaux est beaucoup plus réduit (de l’ordre de 300). Le niveau de complexité des caractères chinois varie d’un seul radical, pour les plus simples, jusqu’à une dizaine de radicaux, pour les plus complexes. Dès lors, l’analyse d’un caractère chinois peut être assimilée à l’analyse d’une « expression » multi-symboles, dans laquelle les radicaux sont assimilés aux symboles. L’agencement spatial des radicaux doit être décrit pour reconnaître « l’expression » que constitue le caractère complet. Cette analogie est d’autant plus solide que des caractères chinois différents peuvent être composés du même ensemble de symboles organisés différemment, comme c’est le cas pour les expressions mathématiques par exemple. Il faut cependant noter qu’à la différence des expressions mathématiques et malgré sa nature fortement structurée, un caractère chinois a tendance à être tracé comme un tout par les scripteurs. De nombreuses méthodes de reconnaissance se sont néanmoins intéressées à la description de leur structure interne, ce qui a donné lieu à un besoin de description des relations spatiales inter-radicaux [LJN04].

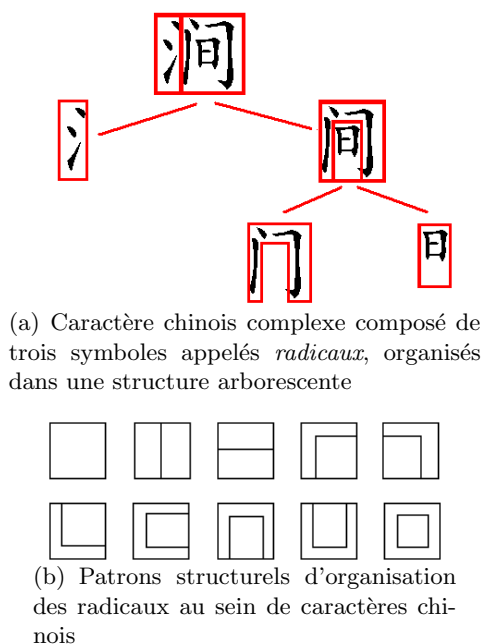


FIGURE 2.12: Les caractères chinois ont une structure en arbre dont les feuilles sont des symboles élémentaires appelés *radicaux* (a). Les radicaux sont organisés entre eux selon 10 patrons de composition (b).

Les relations spatiales entre les radicaux au sein d'un caractère sont très codifiées et doivent respecter l'un des patrons de composition de structure répertoriés. Ces modèles sont au nombre d'une dizaine, et décrivent comment deux radicaux peuvent s'agencer dans un caractère complexe, qui doit lui-même globalement s'inscrire dans un carré. Le nombre exact de patrons structurels varie légèrement selon les sources, mais la figure 2.12(b) reprend ceux les plus fréquemment cités (notamment par Dai, Liu *et al.* [DLX07] et Wang, Fan *et al.* [WF01]). Ils peuvent être appliqués récursivement pour décrire les caractères comportant plus de deux radicaux (comme c'est le cas dans l'exemple de la figure 2.12(a)).

De nombreuses méthodes de reconnaissance exploitent cette représentation hiérarchique pour la reconnaissance de caractères chinois [ML08][OYSS07]. La description des relations spatiales entre radicaux s'appuie alors systématiquement sur la boîte englobante des radicaux. Par exemple, pour Ota, Yamamoto *et al.* [OYSS07], les quatre coordonnées x_{min} , x_{max} , y_{min} et y_{max} de la boîte englobante d'un radical représentent sa position dans le carré unité. La position relativement à un autre radical est décrite par les différences entre les coordonnées des deux boîtes englobantes et un modèle gaussien en quatre dimensions est construit pour chaque configuration.

Un autre exemple d'intégration d'information spatiale pour l'interprétation de l'écrit est l'utilisation de modèles géométriques pour la segmentation et la reconnaissance de lignes de texte. Appliqué à des textes manuscrits en anglais par Xue, Govindaraju *et al.* [XG02] comme à des langues asiatiques par Yin, Wang *et al.* [YWL10], Zhu, Zhou *et al.* [ZZLN10] ou encore Zhou, Yu *et al.* [ZYL⁺07], le principe est le même : il s'agit d'utiliser des caractères

2.1. Représentation par descripteurs

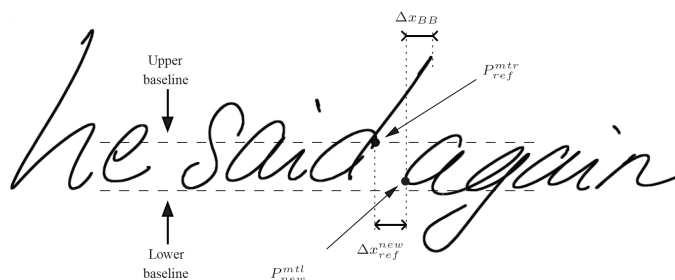


FIGURE 2.13: Difficulté de la caractérisation des espaces inter-mots dans une ligne de texte manuscrit. Des caractéristiques de distance horizontale peuvent être extraites de façon plus stable grâce à la ligne de base. Image extraite de [QBA09].

téristiques géométriques pour aider à identifier des frontières entre les caractères ou entre les mots qui sont écrits sur une ligne, de gauche à droite. Des hypothèses de segments sont formulées et évaluées par des modèles statistiques à partir de caractéristiques telles que les dimensions des boîtes englobantes des segments ou la distance horizontale entre les segments consécutifs [ZZLN10]. Finalement, la confrontation des propriétés géométriques d'une hypothèse de segmentation avec les modèles fournit un coût probabiliste qui participe à l'évaluation de cette hypothèse et ainsi à la reconnaissance globale du texte.

La figure 2.13, proposée par Quiniou, Bouteruche *et al.* [QBA09], présente quelques caractéristiques pouvant être utilisées pour la segmentation de texte manuscrit en mots. Il s'agit d'identifier les frontières entre les mots en distinguant les espaces entre mots des espaces entre lettres d'un même mot. On remarque qu'une déformation courante de l'écriture manuscrite telle que l'inclinaison suffit à rendre cette tâche difficile. L'extraction d'une zone de *ligne de base*, qui permet de donner un contexte géométrique plus stable au mot par rapport aux caractères pris un par un, est alors précieuse.

Dans ces applications dédiées à l'analyse du texte, le positionnement est décrit principalement par des informations de distance. Le positionnement est considéré presque exclusivement dans une direction (de la gauche vers la droite), même si une information de positionnement vertical est parfois intégrée, notamment à travers le positionnement des éléments de tracés par rapport à la ligne de base. Enfin, les éléments de tracés (les lettres et autres segments de mots ou de phrase) sont généralement assimilés à leur boîte englobante.

2.1.2.2 Approches qualitatives

Lorsqu'il s'agit de positionner relativement des objets manuscrits distants les uns des autres, les approches quantitatives sont largement dominantes. En effet, comme nous l'avons vu à la section 2.1.1.2, les approches qualitatives étaient utilisées majoritairement pour décrire des relations topologiques ou des contraintes géométriques. Ce besoin, qui était important pour les relations spatiales entre primitives de tracés qui sont proches et présentent beaucoup de recouvrements, est beaucoup moins présent pour les relations entre objets composés. On trouve néanmoins quelques descripteurs qualitatifs pour le positionnement relatif de symboles dans le contexte de la reconnaissance d'expressions mathématiques.

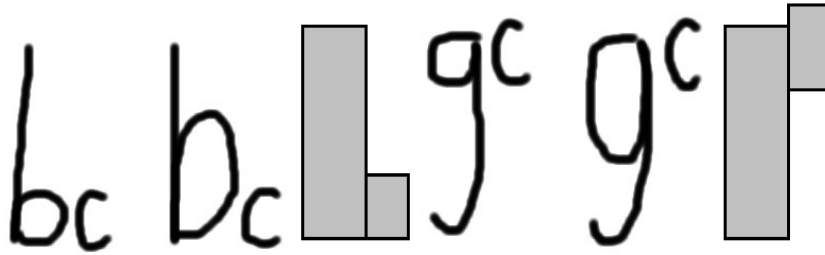


FIGURE 2.14: La prise en compte de la forme des symboles est importante pour la description de leur relation spatiale. A gauche, les deux exemples de paires de symboles « b » et « c » ont les mêmes boîtes englobantes mais sont reliés par des opérateurs différents (relation *produit* et *indice*). Même chose à droite avec deux de couples « g » et « c » liés par les relations de *produit* et d'*exposant*. Image reprise de [GK08].

La première approche, présentée par Genoe & Kechadi [GK08], fait usage de fonctions floues appliquées au décalage vertical entre les objets. Les mesures de décalage sont effectuées, non pas par rapport aux centres des boîtes englobantes, mais par rapport à la zone de la ligne de base de chaque symbole. En effet, les auteurs relèvent que la boîte englobante complète d'un symbole est une information trop grossière qui peut conduire à des décisions contre-intuitives. Par exemple, à la figure 2.14, les paires de symboles « b » et « c » ont les mêmes boîtes englobantes, mais on perçoit des relations sémantiques différentes dans les deux cas (respectivement *produit* et *indice*). En appliquant les règles floues déterminées, chaque type de relation envisagée se voit attribuer un score de satisfaction, qui est intégré au processus général d'analyse structurelle.

Dans le deuxième exemple d'approche qualitative, introduit dans leurs travaux par Fitzgerald, Geiselbrechtinger *et al.* [FGK07], les relations spatiales entre symboles mathématiques sont décrites au moyen de trois propriétés : la position relative verticale, la position relative horizontale et la taille relative. Les positions relatives sont déterminées par rapport à des points particuliers qui sont définis spécifiquement pour chaque relation envisagée. Des sous-ensembles flous de formes sigmoïdes sont conçus empiriquement et appliqués pour déterminer les valeurs acceptables des trois attributs pour chaque relation prédéfinie : *exposant*, *indice*, *au-dessus*, *en-dessous*...

2.1.2.3 Synthèse

Les informations utilisées pour décrire les relations spatiales entre symboles constituant une expression sont essentiellement les informations de positionnement relatif et d'échelle relative. Le positionnement est décrit le plus souvent au moyen d'un ou plusieurs vecteurs de déplacement entre des points particuliers des symboles : centres ou coins de boîtes englobantes, ligne de base. La représentation des objets revient presque toujours à les résumer par leur boîte englobante, voire par un seul point (centre de gravité, centre de boîte englobante). Cette représentation grossière a été critiquée dans plusieurs travaux pour son imprécision et même le caractère contre-intuitif des descriptions auxquelles elle peut conduire. Parmi les travaux cités précédemment, nous avons déjà pu percevoir des problèmes dûs à la re-

2.1. Représentation par descripteurs

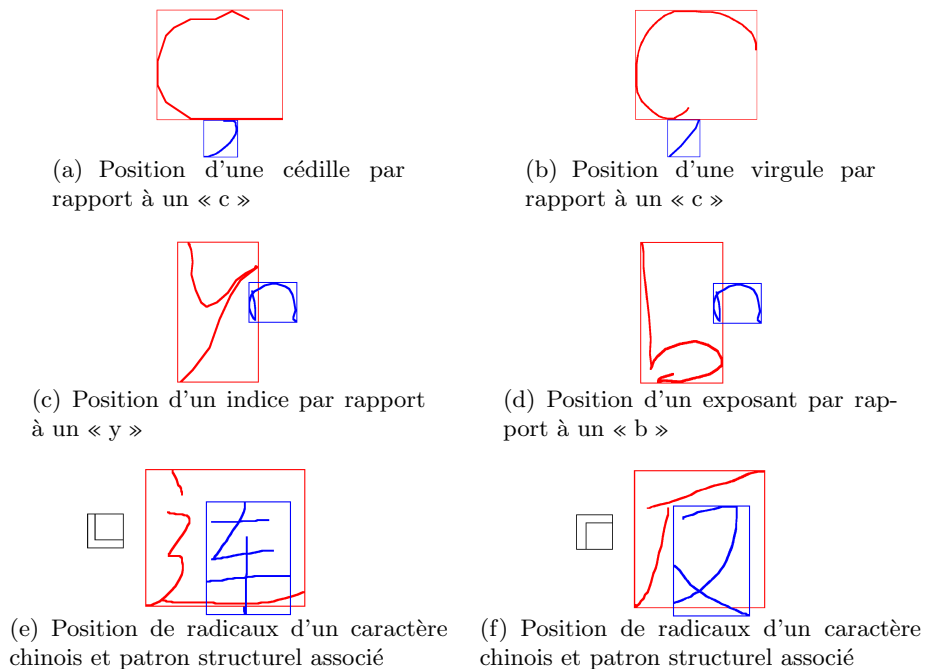


FIGURE 2.15: Des objets positionnés selon des relations différentes peuvent avoir la même description par boîte englobante. La prise en compte plus précise de la forme des objets est nécessaire pour distinguer les cas (a) et (b) (signe diacritique et ponctuation dans du texte), (c) et (d) (relations mathématiques d'indice et d'exposant), et (e) et (f) (structuration de différents caractères chinois).

présentation par boîte englobante des objets, notamment à la figure 2.13 (détection des espaces inter-mots dans du texte manuscrit) ou à la figure 2.14 (importance de la prise en compte de la ligne de base dans le positionnement relatif de symboles mathématiques). La représentation d'objets par leur boîte englobante est largement utilisée, y compris lorsque les boîtes englobantes des symboles à positionner ne sont pas disjointes et se recouvrent largement.

La figure 2.15 représente plusieurs exemples de situations réelles de positionnement relatif entre des objets manuscrits de différentes natures (lettres de l'alphabet latin, symboles d'expressions mathématiques, caractères chinois). La boîte englobante de chaque objet est représentée. Sur chaque ligne, deux paires d'objets sont représentées. On constate que bien qu'elles aient les mêmes boîtes englobantes, elles correspondent en réalité à des exemples de relations spatiales différentes entre les objets. Ces relations spatiales sont pourtant bien perceptibles en considérant le tracé réel des objets. Les exemples (a)-(d) mettent en évidence l'importance de prendre en compte les formes des tracés pour le positionnement relatif d'objets même dans le cas où leurs boîtes englobantes ne se recouvrent pas (objets (a)(b)) ou très peu (objets (c)(d)). L'exemple (e) met en évidence la description contre-intuitive que peut fournir l'utilisation des boîtes englobantes dans le cas où leur recouvrement est important. La position de la boîte englobante bleue par rapport à la rouge laisse penser que l'objet bleu est plutôt en-dessous du rouge, ce qui est le contraire de la relation perçue

en considérant les tracés, l'objet bleu étant nettement au-dessus et à droite du rouge. Ces situations de chevauchement sont très courantes dans les expressions manuscrites, dans le tracé de caractères chinois, mais aussi dans les expressions mathématiques (par exemple avec le symbole de racine carrée) ou dans l'écriture de texte (par exemple lorsqu'il est incliné comme illustré à la figure 2.13)...

Les choix de simplification des objets couramment retenus pour extraire un descripteur de leur positionnement relatif nous semblent incompatibles avec l'idée de concevoir des modèles génériques et non simplificateurs de relation spatiale. S'il est communément admis que deux objets peuvent être assimilés à des points lorsqu'ils sont suffisamment éloignés, cela n'est en pratique pas du tout le cas dans le contexte de symboles manuscrits voisins dans une expression composée. De plus, la représentation par boîte englobante n'est acceptable que si l'on admet que la forme réelle des objets n'importe pas dans la perception de leur positionnement spatial relatif. La figure 2.15 a permis de prouver au contraire la nécessité de prendre en compte les formes des objets dans la perception des relations spatiales. Les particularités de formes, la présence de concavités, de boucles, la grande variation de ces formes en fonction des styles d'écriture individuels et du contexte de saisie sont le propre des tracés manuscrits, et il est indispensable d'en considérer la richesse pour décrire avec finesse le positionnement spatial de ces objets. Dans nos travaux, nous veillerons particulièrement à conserver toute la connaissance sur le tracé exact des objets dans la description de leur positionnement relatif.

2.1.3 Positionnement générique d'objets dans des images

Le panorama sur la description du positionnement relatif de primitives et d'objets manuscrits présenté jusqu'ici laisse apparaître une longue collection de techniques dédiées à des problèmes applicatifs spécifiques. Les besoins en termes d'expressivité (les différentes natures de relations spatiales modélisées) sont naturellement adaptés à l'application, mais nous avons relevé également que la représentation même des objets varie en fonction de leur complexité graphique. Les tracés primitifs sont généralement représentés assez fidèlement par des segments, quand les objets complexes sont représentés plus grossièrement par une boîte englobante ou par un point synthétique. Comme nous venons de le mettre en évidence, ces représentations ne sont pas satisfaisantes pour décrire précisément les relations spatiales entre des objets qui sont caractérisés par des formes souvent concaves, dont le positionnement relatif peut être assez complexe et présenter des recouvrements de boîte englobante.

En analyse d'images, la question du positionnement relatif entre éléments a fait l'objet de nombreux travaux. Dans ce contexte, les objets sont souvent définis comme un ensemble de points de l'image. Cette définition est bien adaptée aux objets surfaciques et est homogène quel que soit le niveau de complexité graphique des objets. Elle permet aussi de mieux formaliser les méthodes de description des relations spatiales, tout en portant souvent plus d'attention à prendre en compte précisément les formes des objets. Les descripteurs de positionnement utilisés en analyse d'images ont l'avantage d'être plus génériques vis-à-vis des types d'objets qu'ils peuvent traiter. Leur objectif reste cependant de décrire le positionnement spatial d'objets sous certains aspects qui sont déterminés en fonction du besoin applicatif.

2.1. Représentation par descripteurs

Dans la suite, nous énumérons quelques familles de descripteurs quantitatifs et qualitatifs pour la représentation de l'information spatiale entre des objets dans des images. Nous nous intéressons plus particulièrement aux méthodes qui délaissent les approximations des formes d'objets au moyen de leurs boîtes englobantes pour plutôt prendre en compte l'ensemble de leurs points.

2.1.3.1 Approches quantitatives

Des descripteurs quantitatifs sont très souvent utilisés pour la description des relations de directions ou de distances entre les objets. La thèse de Dehak [Deh02] ainsi que l'état de l'art de Bloch & Ralescu [BR03] dressent des bons panoramas des descripteurs utilisés pour les relations directionnelles *floues* entre des objets. Le recours à une modélisation floue pour la représentation des relations spatiales a été proposé par Freeman [Fre75] et a été depuis largement adopté, tant cette théorie semble bien adaptée pour la description de ces concepts de positionnement qui sont intrinsèquement imprécis. En effet, il semble naturel de considérer qu'on ne peut pas tracer une frontière nette entre les zones de l'espace qui sont *à droite* d'un objet et celles qui ne le sont pas, mais qu'il existe toute une gamme de nuances entre des zones *parfaitement* à droite, *plutôt* à droite, *un peu* à droite et *pas du tout* à droite. Dans le cas de l'analyse d'images, il est aussi fréquent de considérer que les objets eux-mêmes sont flous, car leur définition peut n'être connue qu'avec une part d'incertitude.

S'il existe en analyse d'images des méthodes qui consistent à résumer les objets à leurs centres ou leurs boîtes englobantes, nous nous intéressons ici plutôt à une famille de descripteurs qui préservent plus fidèlement les formes des objets : il s'agit des descripteurs par histogrammes.

Au sein de cette famille, une première méthode est la description par histogramme d'angles. Dans cette approche, les directions des vecteurs reliant chaque point d'un objet avec chaque point de l'autre objet mis en relation sont comptabilisées dans un histogramme, qui est une fonction d'une variable d'angle dans $[0; 2\pi]$. Parce qu'il est construit à partir de la totalité des points de chacun des objets, ce descripteur prend indirectement en compte la forme de ceux-ci. Miyajima & Ralescu [MR94] proposent de considérer cette fonction comme un sous-ensemble flou, qui peut être comparé à des sous-ensembles flous de relations spatiales prédéfinies (par exemple *à droite de*). La méthode de comparaison utilisée est le calcul du degré de *compatibilité* entre les deux sous-ensembles flous, en exploitant le principe d'extension formulé par Zadeh [Zad78].

Les quatre illustrations de la figure 2.16 résument ce principe d'utilisation des histogrammes d'angles. A partir de deux objets (figure (a)), un histogramme d'angles est extrait (b) puis normalisé. L'histogramme comptabilise la position angulaire de tous les points de B par rapport à tous les points de A. Une fois normalisé, cet histogramme H_{AB} est assimilé à un ensemble flou, qui peut être comparé à un modèle prédéfini d'une relation directionnelle telle que *à droite de* (dont une définition est donnée par l'illustration (c) de la figure 2.16). La comparaison de H_{AB} avec le modèle *à droite de* se fait par le calcul de leur compatibilité. Pour deux ensembles flous F et G , leur compatibilité est un ensemble flou $cp(F, G)$ défini à partir du principe d'extension [Zad78] :

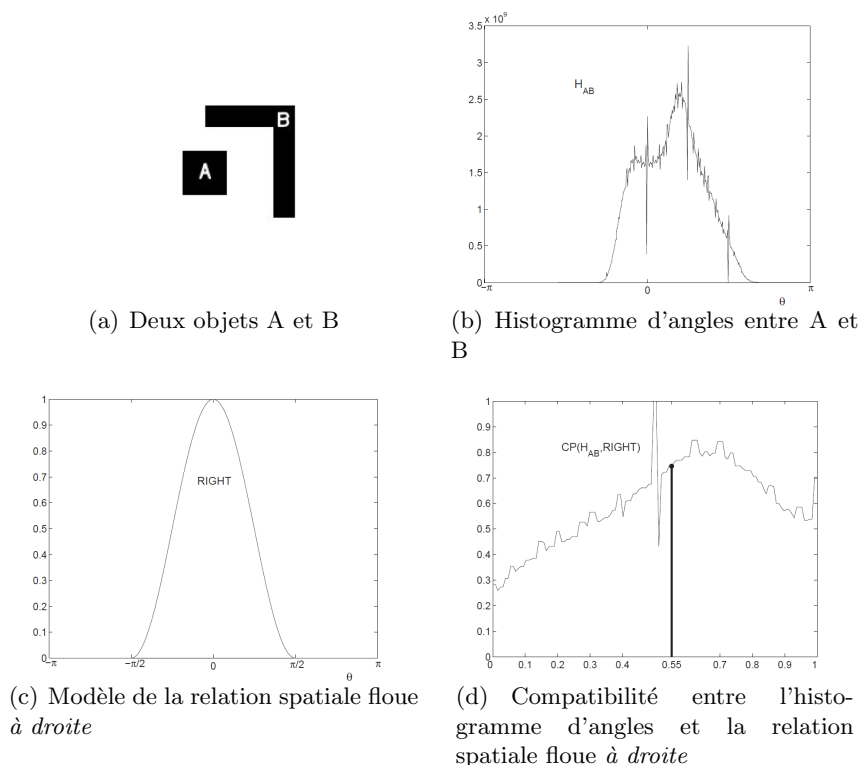


FIGURE 2.16: A partir de deux objets A et B de l'image (a), un histogramme d'angles peut être construit (b). En comparant cet histogramme avec un modèle de relation directionnelle prédéfinie (par exemple à droite (c)), on peut calculer la fonction de compatibilité (d). Le centre de gravité de l'ensemble flou résultat est un score d'adéquation de la position relative de A et B avec le modèle (c). Ici ce score indique que la relation spatiale entre A et B correspond à la relation à droite avec un degré de 0,55. (Illustrations reprises de [Deh02]).

$$\mu_{cp(F,G)}(x) = \begin{cases} 0 & \text{si } \mu_F^{-1}(x) = \emptyset \\ \sup\{\mu_G(y), x = \mu_F(y)\} & \text{sinon.} \end{cases} \quad (2.1)$$

L'ensemble flou de compatibilité entre H_{AB} et le modèle *droite* est représenté par l'illustration (d). Le calcul du centre de gravité de l'ensemble flou de compatibilité fournit finalement un score qui décrit l'adéquation entre la position des deux objets A et B et le modèle de relation directionnelle à droite. Dans le cas des objets de la figure 2.16(d), ce calcul conduit à un score de correspondance de 0,55 entre le positionnement de A et B et la relation à droite (autrement dit, l'objet B est considéré à droite de A avec un degré de satisfaction de 0,55).

Wang & Keller [WK99] ont proposé une autre utilisation des histogrammes d'angles pour la description de relations directionnelles. Leur idée est d'entraîner des réseaux de neurones pour associer à un histogramme d'angles quatre degrés de satisfaction vis-à-vis des quatre directions *nord*, *sud*, *est* et *ouest*. L'apprentissage des réseaux de neurones se

2.1. Représentation par descripteurs

fait à partir d'exemples dont les degrés d'adéquation aux quatre relations directionnelles ont été estimés par des opérateurs humains. L'idée est donc d'apprendre à généraliser une estimation intuitive des degrés de relations directionnelles à partir d'histogrammes d'angles.

Une généralisation des histogrammes d'angles a été proposée par Matsakis & Wendling [MW99]. Les *histogrammes de forces* (ou \mathcal{F} -histogrammes) intègrent une information de distance qui permet de définir une force d'interaction dont l'interprétation est contrôlée par un paramètre r . Lorsque r vaut 2, le descripteur représente la force gravitationnelle exercée par les points d'un objet sur l'autre. Les couples de points ont alors un poids proportionnel au carré de la distance de l'un à l'autre dans la construction de l'histogramme. Un histogramme d'angles est un cas particulier d'histogramme de forces, obtenu lorsque le paramètre r vaut 0 (les couples de points ont alors tous le même poids dans l'histogramme). Les auteurs valident qualitativement leurs descripteurs en les comparant à des histogrammes d'angles classiques. Trois niveaux d'utilisation des \mathcal{F} -histogrammes sont répertoriés par Matsakis [Mat02]. Le premier est la comparaison directe de \mathcal{F} -histogrammes extraits, par exemple dans le but d'identifier des paires d'objets ou des scènes similaires. Un deuxième niveau d'utilisation, dit intermédiaire, consiste à utiliser le \mathcal{F} -histogramme pour vérifier ou attribuer un score à certaines relations de positionnement prédéfinies (telles que *à droite*), à la manière de ce que nous avons détaillé à la figure 2.16 pour les histogrammes d'angles. Le niveau supérieur correspond à l'utilisation des \mathcal{F} -histogrammes pour produire une description linguistique de la relation entre les deux objets [BMK04], d'une façon similaire à ce que nous avons évoqué précédemment pour les histogrammes d'angles.

Les R-histogrammes sont une autre extension des histogrammes d'angles, présentés par Wang & Makedon [WM03]. Ils incorporent au descripteur une information supplémentaire sur la relation topologique entre les objets. Les auteurs proposent de créer un histogramme dans un espace à deux dimensions, les vecteurs entre paires de points des deux objets étant distingués par leur direction (comme dans un histogrammes d'angles) ainsi que par leur *distance qualifiée*. Une distance qualifiée est un descripteur combinant une distance euclidienne entre les deux points et un qualificatif qui décrit si l'un des deux points appartient à l'autre objet. Le but est ainsi de renforcer la capacité du descripteur à distinguer les cas de recouvrement entre les objets (information topologique).

Récemment, Santosh, Wendling *et al.* [SWL10a] ont introduit un nouveau descripteur inspiré des histogrammes d'angles, mais permettant également de distinguer des relations topologiques entre des objets. L'idée est de s'appuyer sur la région caractérisée de manière unique par la disposition des boîtes englobantes des deux objets mis en relation. En fonction des relations topologiques entre les deux objets, cette région peut être un rectangle, être réduite à un segment ou à un seul point. La figure 2.17 illustre la forme de la région R pour différentes configurations. L'information de positionnement directionnel entre les objets est décrite en construisant un histogramme d'angles à chaque sommet de la région R . Cet histogramme est binaire, il décrit donc simplement les directions dans lesquelles les objets A et B sont visibles depuis les sommets de R . Le descripteur est appliqué comme nouvelle signature décrivant la forme de symboles graphiques structurés pour leur indexation et leur reconnaissance dans des documents imprimés [SWL10b][SLW11].

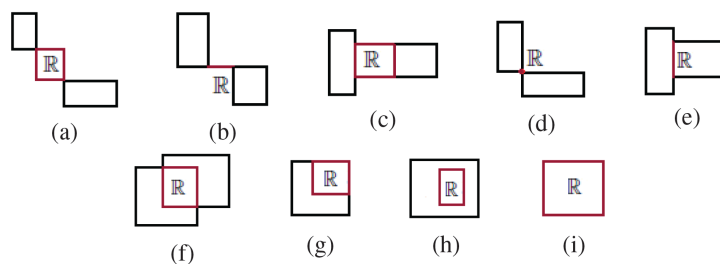


FIGURE 2.17: La forme et le nombre de sommets de la zone R définie à partir des boîtes englobantes des deux objets à positionner permet de caractériser l'information sur la relation topologique entre les objets. Des descripteurs sont extraits de chaque sommet de R et sont combinés pour construire un descripteur global combinant l'information de topologie et de positionnement directionnel. L'image est extraite de [SWL10b].

2.1.3.2 Approches qualitatives

Les descripteurs qualitatifs coïncident la plupart du temps avec la représentation de relations topologiques entre les objets. La description des relations topologiques s'appuie sur des notions ensemblistes (telles que l'inclusion, l'intersection) qui peuvent être directement appliquées aux objets vus comme des ensembles de points du plan. Des extensions floues de ces opérations ensemblistes peuvent aussi naturellement être mises en œuvre dans le cas d'objets vus comme des ensembles flous.

Plusieurs modèles de représentation de relations topologiques existent, qui répertorient toutes les configurations possibles de contact pour une paire d'objets. Egenhofer propose par exemple de décrire un objet par trois ensembles : son intérieur, sa frontière, et son extérieur [EH91]. La relation topologique entre deux objets A et B peut alors être représentée par un vecteur à neuf attributs binaires, où chaque attribut décrit l'existence ou non de points dans l'intersection des trois parties de deux objets (*intérieur de A et intérieur de B* , *intérieur de A et frontière de B* , ..., *extérieur de A et extérieur de B*). Ces informations caractérisent toutes les situations topologiquement différentes telles que l'adjacence, l'inclusion, le recouvrement...

Un autre modèle populaire est le *Region Connection Calculus* (RCC) qui consiste à catégoriser en huit types les relations topologiques possibles entre deux objets vus comme des ensembles de points et de conduire des raisonnements logiques sur la base de ces relations (on parle alors de *raisonnement spatial qualitatif*) [Ren02].

2.1.4 Synthèse des approches par descripteur

La première partie de cet état de l'art a exposé de nombreuses méthodes visant à décrire le positionnement d'objets dans des espaces de représentation dédiés, au moyen de descripteurs extraits de l'espace image. Les méthodes rentrant dans cette catégorie sont multiples, utilisent des descripteurs qualitatifs ou quantitatifs et décrivent des relations spatiales variées en utilisant des informations de topologie, de directions et de distances.

2.2. Représentations spatiales

Ce passage en revue a permis de constater que la plupart des méthodes de positionnement s'intéressent à l'agencement *relatif* des éléments entre eux. Pour l'essentiel, les descripteurs présentés sont destinés à décrire le positionnement d'un objet A par rapport à un objet B , plutôt que par rapport à un repère absolu. La description d'une *scène* complexe, mettant en jeu un grand nombre d'objets se fait donc la plupart du temps en considérant les relations des objets pris deux à deux. Il est bien naturel que les descriptions relatives soient largement dominantes en particulier dans le contexte d'objets manuscrits. Ces objets sont en effet par nature positionnés les uns par rapport aux autres, car ils sont tracés *en contexte* : la position d'un nouvel objet d'une expression manuscrite est affectée par les objets tracés préalablement.

Nous avons d'abord relevé que dans le cadre des objets manuscrits, les descripteurs de relations spatiales sont généralement dédiés à une application précise, un niveau de complexité donné des objets (simples primitives ou symboles complexes) et une catégorie de relations spatiales. Il n'existe pas de méthode générale, susceptible d'être exploitée avec des données de différentes natures, de différents niveaux de complexités et avec un pouvoir d'expression assez général pour couvrir tous les besoins. De plus, nous avons pointé du doigt le manque général de prise en compte des formes des objets dans les descripteurs de leur positionnement relatif, ce qui est paradoxal pour l'analyse du manuscrit où les formes des objets ont des propriétés spécifiques qui ne doivent pas être négligées (avec en particulier la présence de concavités).

Une incursion parmi les descripteurs de relations exploités pour l'analyse d'images en général a permis de mettre en évidence une série de méthodes qui respectent plus précisément les formes des objets au moyen d'histogrammes construits à partir de tous leurs points. Pourtant, comme tout descripteur, un histogramme d'angles ou un histogramme de forces consiste à résumer, compresser l'information de positionnement relatif des objets. Différentes utilisations peuvent être faites de ces descripteurs (comparaison à des modèles, interprétation linguistique de scène...), mais ils n'ont pas vocation à définir des modèles de classes de relations spatiales au sens que nous avons présenté en introduction : une représentation abstraite, applicable à tout type d'objet, prenant en compte tous les aspects important de la relation (directions, distance, mais aussi échelle relative, formes des objets). De plus, les approches par descripteurs sont incompatibles avec le besoin de *prédiction* que nous avons présenté, car ils ne sont pas capables de résoudre le problème de *localisation*. Dans la seconde partie de cet état de l'art, nous introduisons les principes des modèles spatiaux et présentons quelques méthodes qui s'en inspirent pour la description de relation de positionnement.

2.2 Représentations spatiales

Si on considère la question des représentations de relations spatiales sous l'angle du raisonnement humain, il apparaît clairement que l'extraction de descripteurs depuis l'espace de perception (l'image) vers un espace de représentation distinct n'est pas conforme au processus cognitif mis en œuvre dans le cerveau.

Dans la vie de tous les jours, si l'on cherche à vérifier l'affirmation « A est à droite de B », A et B étant des objets que nous percevons, notre réflexe n'est pas de mesurer des angles entre points de A et de B pour les comparer avec un modèle intuitif de la relation

être à droite dans cet espace de représentation angulaire. Plusieurs études de psychologie expérimentale montrent que le procédé cognitif mis en œuvre pour l'appréhension par le cerveau humain des relations spatiales est tout autre : il s'agit de raisonner à partir de modèles directement interprétables dans l'espace de perception et définis par rapport à l'un des deux objets qui prend le rôle de référence pour la relation analysée. Ces modèles sont qualifiés de modèles *spatiaux*, car ils sont définis dans l'espace perceptif (l'image). Dans l'exemple qui nous intéresse, un modèle de la relation à droite tel que nous concevons intuitivement cette relation est appliqué à l'objet *A*, puis la position de *B* est comparée avec ce modèle. La figure 2.18 schématise le principe de raisonnement à l'aide de tels *modèles spatiaux*. La représentation des relations spatiales se faisant au moyen de modèles appliqués dans le plan image, il n'y a pas d'espace intermédiaire de représentation du positionnement relatif des objets. La prise de décision se fait alors directement depuis l'espace de l'image enrichi par le modèle spatial appliqué comme un « calque ».

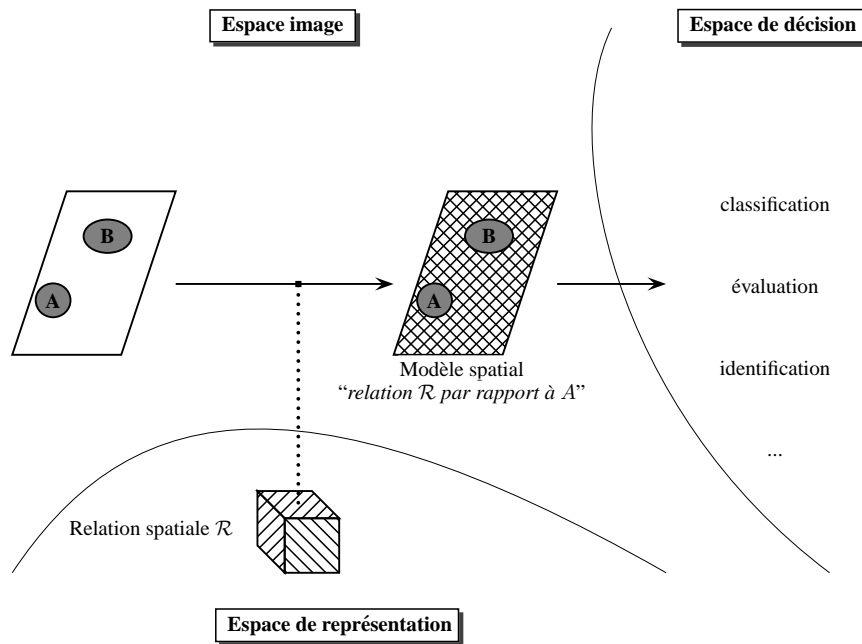


FIGURE 2.18: Schéma général des approches par modèles spatiaux. A partir d'une représentation d'une relation spatiale \mathcal{R} , il est possible de décrire directement dans le plan image un *modèle spatial* adapté à un objet de référence : il décrit la relation \mathcal{R} par rapport à *A*. L'espace image muni de ce modèle spatial peut directement servir à évaluer, identifier, classifier la relation spatiale.

Dans cette partie, nous présentons d'abord les travaux justifiant la théorie des modèles spatiaux pour l'appréhension par l'être humain des relations spatiales, avant de lister les approches qui utilisent des modèles spatiaux pour l'analyse automatique d'images ou de documents.

2.2. Représentations spatiales

2.2.1 Théorie des modèles spatiaux pour l’appréhension de relations spatiales

Nous exposons ici d’abord les principes de la théorie des modèles spatiaux tels qu’ils ont été exposés dans les travaux de cognition spatiale et de psychologie expérimentale. Dans les paragraphes suivants, nous mettons en avant plusieurs particularités des modèles spatiaux qui les distinguent par rapport aux méthodes de représentation des relations spatiales à l’aide de descripteurs évoquées à la section 2.1.

2.2.1.1 Principes

L’appréhension par l’homme des relations spatiales au moyen de *modèles spatiaux* est une théorie couramment admise par les chercheurs en psychologie et cognition : elle a été étudiée notamment par Gapp [Gap95a], Hayward & Tarr [HT95], Logan & Sadler [LS99], Kelleher, Costello *et al.* [KC05] ou encore par Kojima & Kusumi [KK05]. Dans leurs travaux, Logan & Sadler [LS99] introduisent en détail les principes de ces modèles spatiaux, qu’ils nomment en anglais *spatial templates*. Ils s’intéressent dans leur étude aux relations spatiales mettant en jeu deux objets dans l’espace (en 2D ou 3D) et pouvant être décrites à l’aide d’un terme linguistique (par exemple *à droite, au dessus...*). Selon eux, l’appréhension d’une relation spatiale fait d’abord intervenir un cadre de référence. C’est le repère (orthonormé, en 2D ou 3D) qui permet d’orienter les directions, de fixer une échelle et une origine, qui est attachée au centre de l’objet de référence. Le modèle spatial de la relation est ensuite développé dans l’espace muni de ce repère. Un modèle spatial est décrit comme suit par les auteurs :

A spatial template is a representation that is centered on the reference object and aligned with the reference frame [...]. It is a 2D or 3D field representing the degree to which objects appearing in each point in space are acceptable examples of the relation in question.

Il est supposé que le modèle spatial décrit trois grands types de régions d’acceptabilité : les zones de bonne correspondance avec la relation, les zones de moins bonne correspondance mais qui sont néanmoins acceptables et les zones inacceptables. Il est suggéré qu’il n’y a pas de frontière nette entre les zones *bonne* et *acceptable*, mais que la transition entre les deux se fait graduellement. Une fois que le modèle spatial est appliqué au cadre et à l’objet de référence, la position occupée par le second objet mis en relation peut être évaluée. En fonction de la zone qu’il occupe dans le modèle spatial (zone bonne, acceptable, inacceptable), l’acceptabilité de la relation peut être mesurée par un score pouvant varier de 0 (position inacceptable au vu du modèle spatial) à 1 (position parfaite au vu du modèle spatial).

Ce modèle théorique a été validé empiriquement dans les travaux de Logan & Sadler et a servi de cadre de travail à de nombreuses expériences sur la représentation des relations spatiales comme dans les travaux de Gapp [Gap95b], de Carlson & Van Deman [CVD04] ou de Kojima & Kusumi [KK05]. Les relations spatiales étudiées dans ce cadre par les différents chercheurs sont toujours liées à des termes linguistiques simples. Cependant, Logan & Sadler notent également que le cadre de raisonnement autorise aussi la combinaison de modèles spatiaux pour la représentation de relations plus complexes (comme par exemple *en haut à droite*).

2.2.1.2 Expressivité

Le processus de développement des modèles spatiaux selon Logan & Sadler limite a priori leur application aux seules relations spatiales *binaires* et *relatives*. En effet, il s'agit toujours d'étudier la position d'un objet (l'argument) *par rapport* à un autre objet (la référence). Comme nous l'avons constaté précédemment, la très grande majorité des relations spatiales pour l'analyse de tracés manuscrits peuvent être décrites par des relations de nature binaire et relative. De plus, les modèles spatiaux présentent le grand avantage d'être homogènes entre eux : puisqu'ils sont définis dans le même espace (l'espace image) quelle que soit la relation spatiale relative qu'ils représentent (*à droite de X, proche de X...*), il est aisé de les combiner entre eux pour former des modèles spatiaux plus précis. Il est à noter que des modèles spatiaux peuvent très bien être combinés même s'ils sont attachés à des objets de référence différents (mais définis dans la même image), ce qui permet d'enrichir encore le pouvoir d'expression de cette modélisation. Cet usage a notamment été validé par des expérimentations conduites par Gapp auprès d'un panel d'utilisateurs [Gap95b]. On note finalement que les relations spatiales décrites par les modèles spatiaux appartiennent essentiellement aux catégories des relations *métriques* (de direction et de distance), plutôt que topologiques. En effet, les notions de distances et de directions sont bien adaptées à une représentation dans le plan image, alors qu'il est plus difficile d'imaginer une représentation équivalente pour des relations telles que l'*adjacence* ou l'*intersection* d'objets. Logan remarque par exemple que les relations qui impliquent un contact entre les objets (par exemple la relation décrite par le terme « *on* » en anglais) ne se prêtent pas à la représentation par modèle spatial.

2.2.1.3 Gradualité, imprécision et formalisation floue

La description donnée par Logan & Sadler des modèles spatiaux insiste sur le caractère graduel des transitions entre les zones d'acceptabilité qu'ils définissent. En fait, cette caractéristique est motivée par le fait que les concepts relationnels qu'ils décrivent tels que *à droite de X* sont eux-mêmes intrinsèquement imprécis, dans le sens où ils ne peuvent pas être définis de façon satisfaisante par un ensemble strict de points de l'image. Au contraire, il apparaît bien plus conforme à la définition intuitive de la relation *à droite* d'être définie graduellement, sans frontière nette entre les points considérés comme *à droite de X* et ceux qui ne le sont pas du tout. Dès lors, il est naturel de chercher à formaliser les modèles spatiaux dans le cadre de la théorie des sous-ensembles flous introduite par Zadeh [Zad65]. Dans ce cadre, qui généralise la théorie des ensembles classique, il est possible de définir des ensembles non stricts, auxquels les éléments appartiennent avec un degré exprimé entre 0 et 1. Un élément dont le degré d'appartenance à un ensemble vaut 1 est considéré comme très représentatif de l'ensemble, tandis qu'un élément de degré 0 est considéré comme pas du tout représentatif. Tous les degrés entre 0 et 1 permettent d'exprimer des nuances et de considérer les points qui appartiennent fortement, modérément et peu à l'ensemble. Une annexe à la fin de ce rapport présente les principaux concepts et opérateurs de raisonnement logique flou (voir annexe A).

Il a été suggéré de longue date de décrire et modéliser les relations spatiales entre objets au moyen de fonctions floues (voir par exemple les travaux de Freeman [Fre75]). Puisque le raisonnement spatial humain fait appel dans la vie de tous les jours à de nombreux concepts

2.2. Représentations spatiales

intrinsèquement flous (à *gauche*, *loin*, *au Nord-Est...*), il semble conforme à l'intuition de formaliser ces concepts dans ce cadre théorique. Il est à noter que la question d'aborder le raisonnement spatial sous un formalisme permettant la gestion de l'imprécis est toujours un sujet de recherche très actif, comme en témoignent les nombreux travaux et ouvrages récents qui y sont consacrés [JPPS10][MS02][HAB08].

2.2.1.4 Usages : vérification et localisation

Au moment de concevoir des expérimentations pour valider la théorie des modèles spatiaux auprès d'un panel de sujets, Logan relève que ces modèles peuvent être mis en évidence au travers de plusieurs tâches [Log94]. Les tâches demandées aux sujets de l'expérimentation peuvent consister en l'évaluation de la satisfaction d'une relation entre deux objets (*est-ce que X est à droite de Y?*) ou en la recherche d'un objet satisfaisant une relation par rapport à une référence (*qu'y a-t-il à gauche de X?*). Il est intéressant de lier ces tâches à des problèmes d'analyse d'images pour démontrer l'intérêt des modèles spatiaux par rapport aux méthodes à base de descripteurs de positionnement.

La première tâche citée, que l'on qualifie de tâche de *vérification*, est classiquement facile à adresser au moyen de descripteurs de relation spatiale qui peuvent être classifiés, ou comparés à des modèles idéaux de chaque relation construits dans le même espace de représentation. Il s'agit là de la vocation même de cette première famille de représentations. Pour une description par modèle spatial, cette tâche est simplement remplie en comparant la position de l'objet argument avec les zones d'acceptabilité définies par le modèle spatial développé sur l'objet de référence.

La deuxième tâche, dite tâche de *localisation*, est impossible à résoudre directement au moyen de descripteurs. Il faudrait, pour y répondre, extraire un descripteur de la relation spatiale entre l'objet de référence et *tout autre objet possible de la scène*, puis procéder pour chacun à une *vérification* de la relation. Cela est en pratique souvent impossible (ou au moins très coûteux), car les objets peuvent être nombreux, et surtout leur segmentation peut ne pas être connue : c'est alors chaque *hypothèse de segmentation* qui doit être considérée dans l'extraction. Au contraire, le modèle spatial est un outil permettant directement de résoudre ce problème, puisque, par définition, il indique dans quelle(s) zone(s) de l'image il faut chercher l'objet. Nous verrons plus tard que cet avantage des modèles spatiaux peut être bénéfique pour l'analyse de formes structurées. En nous appuyant sur le concept de modèle spatial, nous exploiterons leur capacité de *localisation* pour « prédire » et ainsi découvrir a priori les frontières probables entre les segments, en fonction d'un modèle de positionnement appris et déployé sur l'objet de référence.

2.2.2 Modèles spatiaux simples pour l'analyse du manuscrit

Il existe une catégorie d'approches pour la description du positionnement relatif d'éléments manuscrits qui dérivent des concepts de modèles spatiaux introduits ci-dessus. Il s'agit des stratégies qui appliquent une partition du plan autour d'un objet de référence, y construisant des zones de prédominance pour chaque relation spatiale possible. En effet, ces méthodes consistent à appliquer directement dans le plan une forme de modèle spatial, centré sur l'objet de référence et définissant des zones d'acceptabilité.

Un bon exemple est tiré des travaux de Zhang, Blostein *et al.* [ZBZ05] sur l'analyse d'expressions mathématiques manuscrites. Chaque symbole est caractérisé par sa boîte englobante et par un point correspondant au centre de gravité normalisé selon le *type* du symbole (standard, ascendant, descendant, ou ascendant-descendant). C'est à partir de la boîte englobante du symbole de référence et de seuils prédéfinis que sont développées des zones correspondant aux relations spatiales modélisées : *indice*, *exposant*, *au-dessus*, *à droite*... La figure 2.19 représente une partition construite autour d'un symbole « a », délimitant sept régions qui correspondent aux sept relations spatiales modélisées. Cette partition est à but discriminant. Elle est utilisée pour attribuer une étiquette à la relation spatiale entre deux objets. Par exemple, un objet est étiqueté comme étant *exposant* de l'objet de référence si son *centre* est localisé dans la zone de même nom. Une extension à une partition floue est présentée dans [ZBZ05] ; elle permet de tolérer plusieurs hypothèses avec des scores d'adéquation associés. Pour faire le lien avec le cadre des modèles spatiaux exposé plus haut, on peut considérer que cette partition résulte de plusieurs modèles spatiaux (un par classe de relation spatiale définie). Ainsi, la zone *exposant* de la partition peut être vue comme la zone où l'acceptabilité du modèle spatial *exposant* est supérieure à l'acceptabilité des modèles spatiaux concurrents.

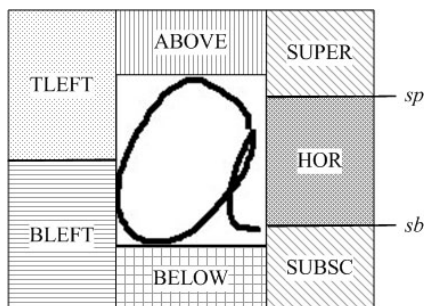


FIGURE 2.19: Partition définissant les classes de relations spatiales possibles autour d'un symbole « a » dans une expression mathématique. *sb* et *sp* sont des seuils définis empiriquement. L'image est tirée de [ZBZ05].

Ce type de description est assez courant pour l'analyse d'expressions mathématiques, en raison de la simplicité de mise en œuvre : on la retrouve notamment dans les travaux de Zanibbi, Blostein *et al.* [ZBC02] et de Kim, Kim *et al.* [KK10]. Un autre exemple est l'utilisation par Macé & Anquetil [MA09] de modèles spatiaux pour le pilotage d'une analyse grammaticale de documents structurés. Le déclenchement de règles d'analyse est dirigé par des zones contextuelles qui sont positionnées aux endroits susceptibles d'accueillir des tracés à analyser. Dans ce cas également, les modèles spatiaux exploités sont des rectangles flous définis empiriquement.

Ces exemples de modèles spatiaux appliqués à l'analyse d'expressions manuscrites sont très rudimentaires. Tout d'abord, ils reposent sur l'assimilation des objets à des boîtes englobantes. Cette critique, que nous avons déjà formulée à l'encontre des modèles à base de descripteurs (voir le paragraphe 2.1.2.3), prend encore davantage d'importance ici, car la représentation dans l'espace image demande une précision accrue. Il suffit pour s'en convaincre

2.2. Représentations spatiales

de revenir à la figure 2.19. En imaginant qu'un objet ait son *centre* situé dans le coin supérieur droit de la boîte englobante du « a », la partition serait incapable de lui attribuer une étiquette de relation spatiale (alors qu'il s'agirait très probablement d'un *exposant*). On peut aussi obtenir des représentations peu intuitives mêmes en cas de légères déformations des symboles : par exemple en imaginant une ligature finale plus longue pour le « a ». En plus des formes des objets, les formes des zones elles-mêmes sont simplistes puisque ce sont des rectangles dont les côtés sont parallèles aux axes du repère de l'image. Cette contrainte très forte peut également conduire à des descriptions peu intuitives. Enfin, un autre défaut de ces modèles est leur caractère empirique, caractérisé par l'utilisation de seuils fixes et sans qu'aucune méthode ne soit fournie pour les déterminer (par exemple par apprentissage automatique sur un jeu de données représentatives). De ce fait, ces modèles spatiaux sont dédiés à des problèmes spécifiques et ils nécessitent une expertise du domaine applicatif visé (par exemple les expressions mathématiques).

Encore une fois, c'est du domaine de l'analyse d'images que viennent les approches les plus riches, qui permettent de répondre à plusieurs de ces limitations.

2.2.3 Modèles spatiaux morphologiques

La mise en œuvre la plus aboutie de modèles spatiaux pour l'analyse des positions relatives d'objets en analyse d'images est sans conteste celle proposée par Bloch, qui fait usage de la théorie de la morphologie mathématique. Bien qu'elle ne s'en revendique pas, cette approche consiste précisément à mettre en application les principes de la théorie des modèles spatiaux formulée dans le domaine cognitif. L'approche morphologique, imaginée par Gader [Gad97], formalisée en détails par Bloch en 1999 [Blo99], a depuis été très étudiée [BR03] et a fait l'objet de nombreuses applications dans différents domaines : analyse d'images médicales [CCB06], analyse d'images satellite [CA07][CA09], navigation autonome en robotique [BS03]... Cette approche présente les avantages de reposer sur des concepts théoriques solidement fondés et de fournir un outillage pour construire des modèles spatiaux sous la forme d'ensembles flous en prenant explicitement en compte la forme des objets. Nous présentons cette méthode dans de plus amples détails car elle constitue la base des travaux présentés dans la suite de cette thèse (chapitre 3).

2.2.3.1 Définition de modèles spatiaux par dilatation morphologique floue

Le principe de cette méthode est de construire la représentation du modèle spatial (qui est appelé dans ces références un *paysage flou*) au moyen d'opérations morphologiques floues appliquées sur l'objet de référence de la relation. La morphologie mathématique est une théorie utilisée en traitement d'images, qui permet de formaliser des transformations non linéaires sur des images à partir de principes ensemblistes [Ser83]. Une extension naturelle de la théorie aux sous-ensembles flous permet de définir des opérations de transformation dont les arguments ou le résultat peuvent être des objets flous. Nous nous intéressons ici à une opération de base de la théorie morphologique appelée *dilatation*. C'est principalement à travers elle que sont définis les modèles spatiaux flous selon les travaux de Bloch.

Une opération de dilatation repose sur l'application d'opérations ensemblistes (des intersections) entre l'objet dilaté et un autre ensemble appelé *élément structurant*. L'élément structurant est défini comme un ensemble de points du plan. Cet ensemble noté v peut être

translaté, et on note v_p son translaté au point p de l'espace \mathcal{S} qui est défini par :

$$\forall q \in \mathcal{S}, q \in v_p \Leftrightarrow (q - p) \in v.$$

La dilatation d'un objet R (défini comme un ensemble classique de points de \mathcal{S}) par l'élément structurant v est aussi un ensemble de \mathcal{S} défini par :

$$\delta_v(R) = \{p \in \mathcal{S} \mid v_p \cap R \neq \emptyset\}. \quad (2.2)$$

Intuitivement, l'élément structurant v est translaté sur les points de l'espace et chaque position p qui lui permet d'intersecter l'objet R appartient au résultat de la dilatation. En particulier, si l'élément structurant est un disque de rayon r centré sur l'origine, le résultat de la dilatation est l'ensemble des points de l'espace qui sont positionnés à une distance de R inférieure à r (cet exemple illustre le nom de *dilatation* puisque l'objet dilaté contient l'objet initial R). Il est important de remarquer que l'objet R initial est explicitement pris en compte dans l'équation de dilatation et que la forme du résultat est fortement liée à la forme de R .

On peut aussi proposer la définition équivalente :

$$p \in \delta_v(R) \Leftrightarrow \exists q \in \mathcal{S}, q \in R \wedge \vec{qp} \in v, \quad (2.3)$$

qui exprime le fait qu'un point appartient à la dilatation de R par v si et seulement s'il existe un point q dans R tel que le vecteur entre q et p appartient à l'élément structurant v . Cette définition est facilement transposable au cas où les ensembles R et v sont définis comme des sous-ensembles flous de \mathcal{S} .

Dans le cas d'un objet R flou, défini par sa fonction d'appartenance μ , et d'un élément structurant flou, désigné par sa fonction d'appartenance ν , le résultat de la dilatation de R par ν est un sous-ensemble flou de \mathcal{S} dont la fonction d'appartenance est définie par :

$$\forall p \in \mathcal{S}, \delta_\nu(\mu)(p) = \sup_{q \in \mathcal{S}} t[\nu(p - q), \mu(q)], \quad (2.4)$$

avec t une t-norme (opérateur implémentant une relation conjonctive floue, voir à l'annexe A). L'expression $t[\nu(p - q), \mu(q)]$ est une transposition floue de l'expression $q \in R \wedge \vec{qp} \in v$ de l'équation (2.3).

Le résultat d'une telle dilatation est considéré comme étant un modèle spatial (dénommé *paysage flou*) construit autour de l'objet de référence R . C'est la définition de l'élément structurant ν qui encapsule la relation spatiale décrite par le modèle spatial résultant de la dilatation. Nous présentons ci-dessous une forme d'élément structurant utilisée dans l'approche morphologique pour construire différents modèles spatiaux directionnels.

2.2.3.2 Exemple : paysages flous directionnels

Des relations directionnelles du type *être dans la direction α par rapport à R* peuvent être décrites par des modèles spatiaux obtenus par dilatation morphologique de l'objet R au moyen d'éléments structurants de la forme :

$$\forall p \in \mathcal{S}, \nu_\alpha(p) = f \left(\arccos \frac{\vec{op} \cdot \vec{u}_\alpha}{\|\vec{op}\|} \right), \quad (2.5)$$

2.2. Représentations spatiales

où o est le centre de l'élément structurant et u_α est le vecteur unitaire de direction α . La figure 2.20 représente la fonction f définie sur $[0; \pi]$ par $f(\theta) = \max(0, 1 - \frac{2}{\pi}\theta)$. C'est une simple fonction linéaire, à valeurs dans $[0; 1]$, qui prend la valeur de 1 lorsque θ vaut 0, qui décroît linéairement quand θ augmente et qui vaut 0 dès que θ dépasse $\pi/2$.

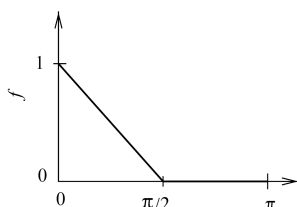


FIGURE 2.20: Représentation de la fonction $f(\theta) = \max(0, 1 - \frac{2}{\pi}\theta)$ utilisée dans la définition des éléments structurants directionnels (équation (2.5)) [BR03].

L'élément structurant directionnel flou défini par l'équation (2.5) peut être représenté par une image 2D en niveau de gris, où l'intensité lumineuse de chaque pixel est proportionnelle au degré d'appartenance du point à l'ensemble flou ν . Par exemple, la figure 2.21 représente un élément structurant directionnel, défini pour la direction $\alpha = 0$ (relation à droite de). Les pixels blancs représentent des points du plan qui appartiennent parfaitement à l'élément structurant flou (ce sont les points p de \mathcal{S} tels que $\nu(p) = 1$). À l'inverse, les pixels noirs représentent des points qui ont une appartenance nulle à l'élément structurant ($\nu(p) = 0$). Cette représentation en niveaux de gris d'ensembles flous définis dans le plan image sera utilisée tout au long de ce document.

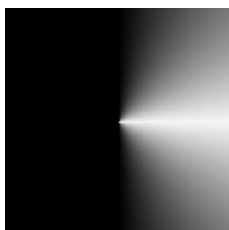


FIGURE 2.21: Élément structurant directionnel flou défini selon l'équation (2.5) avec $\alpha = 0$. Il représente la relation à droite [BR03].

On peut remarquer que l'élément structurant représenté par la figure 2.21 est visuellement conforme à sa définition formelle : il décrit, par rapport à son centre (au centre de l'image), les degrés d'acceptabilité de la relation directionnelle être à droite de. D'après la forme de la fonction f décrite plus haut, tous les points situés dans la moitié gauche de l'image ont un degré nul d'appartenance à l'élément structurant, tandis que tous les points de la partie droite ont un degré positif. Les seuls points qui appartiennent parfaitement à l'élément sont ceux qui sont sur le segment horizontal reliant le centre de l'image avec son bord droit (ce sont les points parfaitement à droite de).

À partir de l'élément structurant ainsi défini, la construction d'un modèle spatial directionnel se fait au moyen de l'opération de dilatation morphologique telle que décrite à

l'équation (2.4). Pour différents objets de référence définis comme des ensembles classiques (non flous) de points du plan, la figure 2.22 représente les modèles spatiaux directionnels obtenus par dilatation avec l'élément structurant représenté à la figure 2.21. La même convention est utilisée pour la représentation en niveau de gris des degrés d'appartenance aux modèles spatiaux flous. A chaque fois, l'objet de référence R est surimposé en rouge, pour plus de clarté par rapport à l'image originale extraite de [BR03].



FIGURE 2.22: Exemples de *paysages flous* directionnels construits sur différents objets de référence, représentés en rouge. Images reproduites d'après [BR03].

Les quatre exemples permettent de mesurer l'impact de la forme de référence sur la forme du modèle spatial résultat. On constate que la notion *être à droite de R* ne signifie pas forcément être complètement à droite de R , mais au moins à droite d'une partie de R . Par conséquent, les zones de \mathcal{S} qui sont à l'intérieur de la cavité du troisième objet, ou la zone encerclée par le quatrième objet, sont considérées comme parfaitement *à droite de R* .

2.2.3.3 Comparaison d'un objet argument avec le modèle spatial

Une fois le paysage flou défini, il est possible d'évaluer avec quel degré un objet cible A respecte la relation décrite par rapport à l'objet de référence R . Dans les travaux de Bloch, une grande attention a été portée aux différentes stratégies pour répondre à ce problème, en particulier afin de pouvoir gérer l'évaluation du positionnement relatif d'objets flous. Dans [Blo99], différents opérateurs sont proposés pour agréger les valeurs d'adéquation des points de A par rapport au paysage flou.

Un opérateur *moyen* consiste à calculer simplement la moyenne des degrés d'adéquation des points de A avec le paysage flou noté μ^R :

$$M^R(A) = \frac{1}{|A|} \sum_{p \in A} \mu^R(p), \quad (2.6)$$

avec $|A|$ qui désigne le cardinal de A . Dans le cas d'un objet A défini par une fonction d'appartenance floue μ_A , l'opérateur de moyenne devient :

$$M^R(A) = \frac{1}{|A|} \sum_{p \in \mathcal{S}} t[\mu^R(p), \mu_A(p)], \quad (2.7)$$

avec t une t-norme, et $|A| = \sum_{p \in \mathcal{S}} \mu_A(p)$ le cardinal flou de A . Cette mesure moyenne donne une évaluation globale de l'adéquation de la position de A avec le paysage flou, en tenant compte de tous ses points.

2.2. Représentations spatiales

D'autres mesures, conformes aux notions de *possibilité* et de *nécessité* utilisées dans la théorie des possibilités pour décrire la correspondance entre deux sous-ensembles flous [DP95], ont aussi été proposées. La nécessité N et la possibilité Π fournissent respectivement une mesure pessimiste et une mesure optimiste de la correspondance entre le sous-ensemble flou A et le paysage flou μ^R , délimitant ainsi un intervalle. La mesure de nécessité est définie par :

$$N^R(A) = \inf_{p \in \mathcal{S}} T[\mu^R(p), 1 - \mu_A(p)], \quad (2.8)$$

avec T une T-conorme (opérateur flou de disjonction). La possibilité est définie par :

$$\Pi^R(A) = \sup_{p \in \mathcal{S}} t[\mu^R(p), \mu_A(p)]. \quad (2.9)$$

Dans le cas où A est un ensemble classique, ces deux mesures correspondent aux valeurs extrêmes de μ^R prises par des points de A .

2.2.3.4 Exemple : évaluation de positions directionnelles

La figure 2.23(a) présente deux objets A et B de formes différentes positionnés par rapport à une même référence R . Les représentations de la figure 2.23 montrent les valeurs d'adéquation du positionnement de A et B par rapport aux quatre relations directionnelles (*être à gauche, à droite, en-dessous, au-dessus de R*). Dans chaque cas, l'intervalle de possibilité/nécessité est représenté, en compagnie de la valeur moyenne. Par exemple, pour l'évaluation de la relation spatiale *B est au-dessus de R*, on lit que la valeur moyenne est environ de 0,5, tandis que la mesure de nécessité vaut 0 et la mesure de possibilité vaut 1. Autrement dit, il existe des points de B qui ne sont pas du tout au-dessus de R et d'autres qui sont parfaitement au-dessus de R . On peut constater que ces trois mesures donnent une indication sur l'étendue de B de part et d'autre des zones d'acceptabilité formées par le paysage flou *au-dessus de R*. Inversement, les trois mesures pour la relation *A est à gauche de R* sont toutes nulles, ce qui décrit que A n'est absolument pas à gauche de R .

2.2.3.5 Expressivité

Dans les exemples présentés, nous avons mis en avant le pouvoir de l'approche morphologique pour construire des modèles spatiaux directionnels. Historiquement, c'est pour décrire ce type de relations spatiales qu'elle a été conçue [Blo99]. Pourtant, par la suite, un pouvoir d'expressivité plus large a été mis en évidence. Nous dressons ici une liste de travaux ultérieurs qui ont montré d'autres utilisations de l'approche.

- Des éléments structurants décrivant des relations de distances ont été présentés (voir par exemple [Blo06]) afin de construire des modèles spatiaux pour des relations du type *proche de X, loin de X*.
- Des éléments structurants hybrides (combinant direction et distances) ont été proposés, notamment avec une forme différente pour la fonction f [CA09].
- Le cas de la relation ternaire *entre*, faisant intervenir deux références (*entre X et Y*) a également été traité au moyen de l'approche morphologique dans [CA09], en s'inspirant de [BCCJ06].

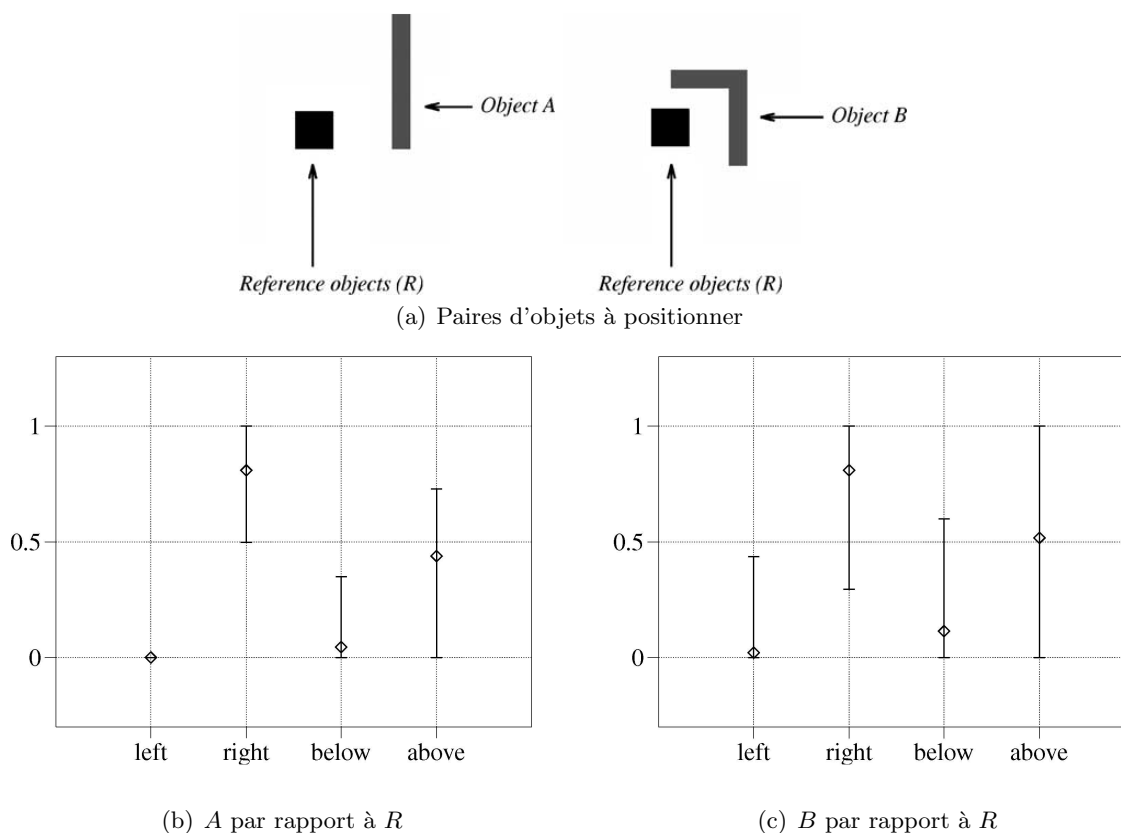


FIGURE 2.23: Des objets A, B et R sont représentés à la figure (a). L'évaluation de quatre relations directionnelles (à gauche de R , à droite de R , en-dessous de R , au-dessus de R) est représentée pour les objets A (image (b)) et B (image (c)). Dans les figures (b,c), pour chacune des quatre directions considérées, les trois mesures de nécessité, possibilité et moyenne sont représentées respectivement par les bornes inférieure et supérieure de l'intervalle ainsi que le losange intermédiaire. Images tirées de [Blo99].

- Dans de nombreux travaux, les modèles spatiaux sont combinés entre eux pour représenter des relations spatiales plus évoluées (au Nord-Est [SK09], proche et à droite de X [HAB08]. ...).
- Des travaux plus anciens, dans le contexte de la reconnaissance de chiffres manuscrits, ont montré que l'utilisation de dilatations morphologiques directionnelles floues peut permettre de détecter la présence de particularités de tracé telles que des boucles ou des concavités [WKG97].
- Atif, Hudelot *et al.* [AHF⁺07] ont suggéré une définition floue pour la relation topologique d'adjacence entre deux objets A et B , qui revient à évaluer la relation (une partie de) A est très proche de B . La relation très proche de peut être représentée par l'approche morphologique. Il convient ensuite d'adapter le mode de calcul de la correspondance entre A et le modèle spatial construit, en considérant par exemple la

2.2. Représentations spatiales

- valeur maximale d'activation du modèle spatial par les points de A .
- Des applications *linguistiques* des paysage flous morphologiques ont été proposées par Wang & Keller [WK99] et par Sledge & Keller [SK09]. Dans le second cas, il s'agit de construire une carte à partir d'une description linguistique de relations spatiales géographiques (*au nord-est, derrière*), faisant parfois intervenir plusieurs références à la fois (*plutôt à gauche de X et au Nord de Y*). La capacité de localisation des paysages flous y est utilisée pour déterminer des régions de positionnement d'objets (bâtiments) à partir de la description linguistique.

Cette liste non exhaustive d'extensions et d'utilisations diverses de l'approche morphologique pour la description de relations spatiales donne un bon aperçu des capacités de la méthode. Conformément aux remarques faites concernant l'expressivité de la théorie cognitive des modèles spatiaux (voir paragraphe 2.2.1.2), l'approche morphologique est plutôt adaptée à la représentation de relations spatiales binaires, relatives, de nature directionnelle ou de distance. Cependant, les possibilités de fusion de modèles dans le cadre de la théorie floue, de prise en compte de plusieurs références, et la formulation de nouvelles interprétations pour des relations topologiques ont été mises en évidence dans les nombreuses extensions de la méthode d'origine.

Notre contribution à la modélisation du positionnement spatial d'objets s'appuiera fortement sur l'approche morphologique. Nous exploiterons dans notre approche ces capacités : couplage d'informations de directions et de distance, possibilité de combiner les modèles spatiaux entre eux et d'appuyer la description sur plusieurs références, ce qui nous permettra d'atteindre une large expressivité dans la modélisation. Par ailleurs, la capacité de localisation sera exploitée pour concevoir des stratégies innovantes de segmentation pour les objets manuscrits (cela fera l'objet d'une des contributions présentées dans le chapitre 4 de ce document).

2.2.4 Modèles spatiaux par l'approche \mathcal{F} -template

Le concept de \mathcal{F} -template, introduit par Matsakis, Wendling *et al.* [MWN10], est un autre moyen pour construire des modèles spatiaux flous dans l'espace 2D ou 3D (voir aussi [WNM06][MNW06]). Il a été introduit assez récemment et est présenté par ses auteurs comme une représentation *duale* des histogrammes de forces (ou \mathcal{F} -histogrammes) que nous avons évoqués au paragraphe 2.1.3.2.

Un \mathcal{F} -template est défini pour une relation spatiale \mathcal{R} et un objet B . Il s'agit d'un sous-ensemble flou de l'espace euclidien qui associe à chaque point p le degré avec lequel il satisfait la relation \mathcal{R} par rapport à la référence B . Cette définition correspond bien à la notion de modèle spatial au sens de Logan & Sadler.

La figure 2.24 montre un objet de référence B et un \mathcal{F} -template construit sur cet objet pour la relation \mathcal{R} *au Nord de*. La convention de représentation d'ensemble flou en niveaux de gris est la même que celle utilisée que dans les illustrations précédentes.

Le principe de définition de ce modèle spatial est de décomposer la relation par rapport à l'objet de référence en une agrégation de relations par rapport à des *sections longitudinales* de l'objet de référence. Les sections longitudinales sont des structures 1D (segments ou union

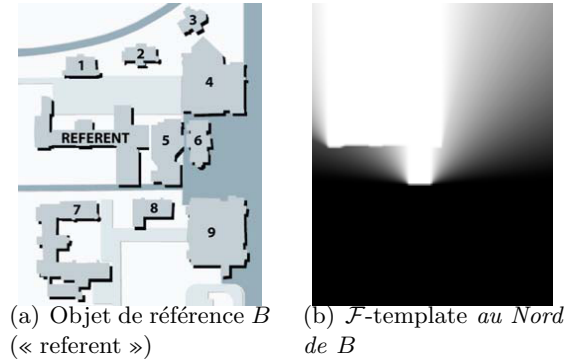


FIGURE 2.24: Exploitation du \mathcal{F} -template décrivant la relation spatiale *au Nord de* sur un objet de référence B (images tirées de [WNM06] et [MNW06]).

de segments) obtenues par intersection de l'objet de référence avec des droites paramétrées par leur direction θ .

Pour le point p de l'espace, la valeur du \mathcal{F} -template associé à la relation \mathcal{R} par rapport à l'objet B est définie formellement par :

$$\mathcal{F}^{\mathcal{R}B}(p) = \sup_{\theta \in [0; \pi]} \mathcal{F}(p, \mathcal{R}, B_p(\theta)), \quad (2.10)$$

où $B_p(\theta)$ désigne la section longitudinale de B par la droite de direction θ passant par p .

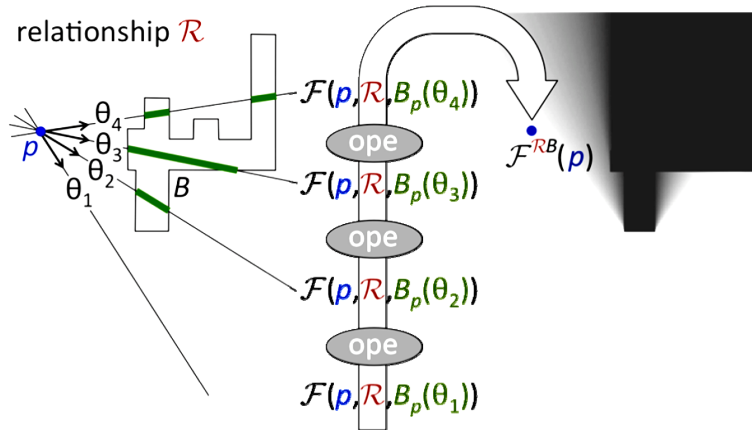


FIGURE 2.25: Calcul de la valeur du \mathcal{F} -template *au Nord de* en un point p , par rapport à un objet B (la convention de représentation de degrés flous en niveau de gris est inversée dans cette illustration). D'après [MWN10].

La figure 2.25 illustre le calcul du \mathcal{F} -template de la figure 2.24 en un point p par rapport au même objet B que précédemment. Il décrit la relation *être au Nord* par rapport à cet objet B . Cette fois-ci, la représentation en niveaux de gris est inversée par rapport à la convention habituelle (le résultat du calcul est donc le « négatif » de la figure précédente,

2.2. Représentations spatiales

avec les points parfaitement au Nord de B qui sont en noir). Les sections $B_p(\theta)$ de l'objet sont illustrées (en vert) pour plusieurs directions θ . Dans le calcul décrit par l'équation (2.10), c'est l'opérateur *sup* qui permet de combiner les valeurs de $\mathcal{F}(p, \mathcal{R}, B_p(\theta))$ pour plusieurs valeurs de θ , mais d'autres opérateurs peuvent s'y substituer [MWN10].

La fonction \mathcal{F} de l'équation (2.10) détermine la manière de calculer le degré d'adéquation (entre 0 et 1) du point p par rapport à $B_p(\theta)$ selon la relation \mathcal{R} . Si \mathcal{R} représente une relation directionnelle (par exemple *être dans la direction* δ), \mathcal{F} doit pouvoir évaluer à quel degré p est dans la direction δ par rapport à $B_p(\theta)$, en tenant compte de l'angle θ considéré et de la position de p par rapport à la section longitudinale $B_p(\theta)$ sur la droite de direction θ . En particulier, dans cet exemple, $\mathcal{F}(p, \mathcal{R}, B_p(\theta))$ vaut 1 si p est parfaitement dans la direction Nord ($\delta = \frac{\pi}{2}$) par rapport à $B_p(\theta)$ (soit quand $\theta = \frac{\pi}{2} + \pi$).

Un \mathcal{F} -template directionnel tel que celui-ci est théoriquement identique au *paysage flou* équivalent construit par approche morphologique. Dans ce cas, l'intérêt d'un \mathcal{F} -template réside en fait dans la plus grande rapidité de son calcul, grâce à l'accumulation des relations sur des sections 1D plutôt que sur tous les points de l'objet de référence. La précision de la représentation est en pratique contrôlée par le nombre de directions θ considérées dans l'équation (2.10).

La mesure de correspondance entre un objet argument A et le \mathcal{F} -template est peu détaillée par les auteurs de la méthode. Dans [CNM06], il est fait mention de la possibilité de calculer cette correspondance de différentes façons, en considérant par exemple le minimum, le maximum ou la moyenne des degrés atteints par les points de A dans le \mathcal{F} -template. Du fait de l'équivalence formelle du \mathcal{F} -template directionnel avec un paysage flou, on peut évidemment appliquer les méthodes d'évaluations développées dans le cadre morphologique (voir au paragraphe 2.2.3.3).

La généralité du formalisme est mise en avant et différentes variantes de fonctions \mathcal{F} sont présentées afin de décrire des relations de distances (*proche, loin*) ou combinant directions et distances [WNM06]. Tout comme dans les \mathcal{F} -histogrammes, il est possible de pondérer dans les \mathcal{F} -templates les points par leur distance à l'objet de référence, ce qui modifie l'interprétation du \mathcal{F} -template en un champ potentiel de forces. Tout comme dans l'approche par morphologie mathématique, la description par \mathcal{F} -template a aussi été pensée pour prendre en compte des objets flous et c'est dans ce contexte que des algorithmes particulièrement performants ont été proposés [MWN10][WNM06]. Des algorithmes ont été proposés pour le calcul de \mathcal{F} -templates aussi bien pour des objets en représentation vectorielle (définis comme des polygones) ou en représentation *raster* (définis sous la forme d'une matrice de pixels). Enfin, comme dans l'approche morphologique, la géométrie des objets mis en relation est parfaitement prise en compte dans la représentation de leurs relations spatiales par \mathcal{F} -templates.

2.2.5 Synthèse des approches par modèles spatiaux

La seconde partie de l'état de l'art a présenté les concepts de modèles spatiaux, issus de théories psychologiques sur l'appréhension par l'homme des relations spatiales. Les modèles spatiaux permettent de raisonner à partir de relations spatiales exprimées directement dans l'espace de l'image, ce qui présente plusieurs avantages, au-delà de l'aspect intuitif que cela apporte. D'abord, l'homogénéité de la représentation des modèles permet de les combiner, de les fusionner indépendamment de la relation spatiale qu'ils décrivent, et même de l'objet

qu'ils ont pour référence. Ensuite, ils offrent un moyen naturel de raisonnement avec des objets et des relations spatiales qui peuvent être définis de façon imprécise, au moyen de sous-ensembles flous dans l'espace image. Pour la tâche d'*évaluation* d'une relation spatiale, ils permettent de développer un raisonnement et de prendre une décision en examinant tout le contexte perceptif dans son ensemble et sans simplification des formes des objets ni de leur positionnement relatif. Enfin, la représentation qu'ils offrent naturellement dans l'espace image permet de résoudre directement le problème de *localisation*.

Nous avons pu voir que quelques travaux offrent des implémentations plus ou moins fidèles de ces concepts pour l'analyse de document et d'image. L'approche par morphologie mathématique floue présente l'avantage de construire des modèles spatiaux conformes à l'intuition et respectant le caractère intrinsèquement flou des relations spatiales décrites. Surtout, elle permet de prendre en compte explicitement la forme des objets dans la description de leurs relations spatiales. En analysant le pouvoir d'expressivité des modèles spatiaux en général, et de ceux construits par l'approche morphologique en particulier, nous avons pu montrer que bon nombre de relations peuvent être exprimées dans ce paradigme (directionnelles, de distance, avec une ou plusieurs références, ainsi que quelques exemples liés à la topologie). Les \mathcal{F} -template sont une alternative au procédé morphologique de construction des modèles spatiaux. Sous certaines conditions, un \mathcal{F} -template est équivalent au paysage flou directionnel défini par dilatation morphologique. Le procédé de construction des \mathcal{F} -templates par agrégation de mesures effectuées sur des sections longitudinales d'objets est adapté à la gestion d'objets surfaciques plutôt que linéiques comme c'est le cas des objets manuscrits que l'on veut traiter. Au contraire, l'approche morphologique n'impose aucune contrainte sur les formes des objets et permet très simplement de traiter des objets linéiques. Elle est aussi plus ancienne et a été beaucoup plus étudiée. Nous avons choisi pour la suite de baser nos travaux sur la modélisation du positionnement relatif d'objets manuscrits sur ce cadre de description par opérateurs morphologiques.

Les approches existantes en analyse d'image qui utilisent les modèles spatiaux sont limitées dans leur potentiel du fait de leur paramétrage toujours statique et prédéfini. En effet, dans tous les cas rencontrés, les modèles spatiaux construits reposent sur des définitions linguistiques simples, implémentées par des *éléments structurants* prédéfinis dans l'approche morphologique, ou des fonctions \mathcal{F} prédéfinies (voir équation (2.10)) pour l'approche \mathcal{F} -template. La définition des éléments structurants et des fonctions \mathcal{F} repose sur des connaissances extérieures traduisant ces relations spatiales linguistiques. Le seul paramétrage prévu pour ces éléments consiste à ajuster la direction principale pour les relations directionnelles, l'influence éventuelle de la distance, ainsi que le « niveau de flou » souhaité pour la représentation (la *tolérance* accordée à une relation décrite linguistiquement)¹.

Il n'existe aucune méthode générale capable de produire des modèles spatiaux pour représenter des relations spatiales non linguistiques, c'est-à-dire non définissables simplement a priori. Dans cette thèse, nous allons chercher à exploiter plus avant le cadre de raisonnement offert par les modèles spatiaux. Notre objectif sera de déterminer une façon de construire des modèles spatiaux non linguistiques, qui décrivent des relations spatiales « découvertes » automatiquement à partir de données.

1. Font exception deux cas cités dans [HAB08] et [AHF⁺07], où un paramètre de l'élément structurant est ajusté automatiquement à partir de données pour détecter des cas atypiques de positionnement relatif.

2.3 Positionnement de nos travaux

2.3.1 Apprentissage de relations spatiales pour l'analyse du manuscrit

La première partie de l'état de l'art a mis en avant les approches pour la description du positionnement relatif d'objets au moyen de descripteurs extraits dans un espace de représentation. Cette technique a l'inconvénient de décorréler la représentation des relations spatiales de l'espace image où évoluent les objets. Le pouvoir d'expressivité de ces représentations est systématiquement restreint par le choix des descripteurs utilisés et les simplifications opérées sur les formes des objets, qui sont guidés par le but applicatif de chaque mise en œuvre. Une fois les descripteurs extraits, les techniques connues pour l'apprentissage et la classification automatique peuvent être utilisées dans le but de les qualifier, de les comparer ou de les catégoriser en plusieurs types de relations spatiales que l'on souhaite pouvoir reconnaître. Pour l'analyse automatique du manuscrit, cette propriété est cruciale car elle permet de définir des modèles qui décrivent le plus souvent des relations spatiales non-linguistiques (qu'il n'est pas possible de décrire au moyen de termes linguistiques usuels). La variabilité du positionnement relatif des objets doit être estimée sur une base de données d'exemples de la relation de positionnement afin d'en construire un modèle fidèle dans l'espace de représentation. Les approches par extraction de descripteurs permettent donc facilement l'apprentissage automatique de modèles de relations spatiales, mais ces modèles sont définis dans des espaces de représentation où l'information sur le positionnement et la forme des objets est simplifiée, compressée et optimisée dans un but précis. Du fait de cette représentation simplificatrice, il n'est en général pas possible de définir une zone de l'espace correspondant aux « bons » positionnements selon un tel modèle. Cela rend impossible pour ces approches de *localiser* des objets.

Dans la seconde partie, nous avons présenté les modèles spatiaux comme une alternative aux approches par extraction de descripteurs. Les avantages qu'ils offrent sont multiples, et l'implémentation qui est rendue possible par les opérateurs morphologiques est aboutie, conforme à l'intuition et capable d'exprimer une large palette des relations spatiales différentes. Elle permet aussi de prendre explicitement en considération la forme réelle des objets à positionner, ce qui est un critère important pour les objets manuscrits dont la forme est sujette à de grandes variations. Par définition, les modèles spatiaux se prêtent parfaitement à la conduite d'un raisonnement complètement réalisé dans l'espace de l'image, ce qui permet une prise de décision qui évite toute simplification des formes des objets et considère leur positionnement relatif dans toute sa complexité. Enfin, ils constituent par nature un outil permettant de *localiser* un objet à partir d'un autre objet connu.

Néanmoins, les approches par modèles spatiaux souffrent d'un déficit de généralisation et d'expressivité : ils ne sont utilisés en l'état que pour décrire des relations spatiales relativement simples et prédéfinies empiriquement. Il n'existe pas de méthode pour construire des modèles spatiaux représentant des relations spatiales plus complexes, non-linguistiques, comme cela est pourtant nécessaire pour traiter des objets manuscrits.

2.3.2 Vers l'apprentissage de méta-modèles spatiaux

La contribution centrale de cette thèse est la présentation d'une stratégie générale pour permettre l'apprentissage de *méta-modèles spatiaux*. Un méta-modèle spatial est une formalisation abstraite d'une relation spatiale non-linguistique, qui peut être construite automatiquement à partir de données. Par construction, un méta-modèle spatial permet de capturer la variabilité dans le positionnement de paires d'objets reliées par une même relation spatiale non-linguistique, ce qui n'est possible jusqu'ici qu'en utilisant l'approche par extraction de descripteurs. Un méta-modèle spatial peut être développé sur un objet de référence pour produire un modèle spatial, à la manière des modèles spatiaux linguistiques présentés plus haut. De cette façon, on bénéficie des avantages des modèles spatiaux pour le raisonnement spatial directement dans l'espace image. La formalisation que nous proposons exploite le cadre de l'approche morphologique décrite précédemment, ce qui permettra de tenir compte de la forme exacte des objets mis en relation, et ceci quelle que soit leur complexité graphique (primitive structurale ou tracé composé). Pour situer conceptuellement l'innovation que constitue la notion de méta-modèle spatial par rapport aux notions existantes, il peut être intéressant de considérer le schéma présenté à la figure 2.26, issu des travaux de Hudelot, Atif *et al.* [HAB08].

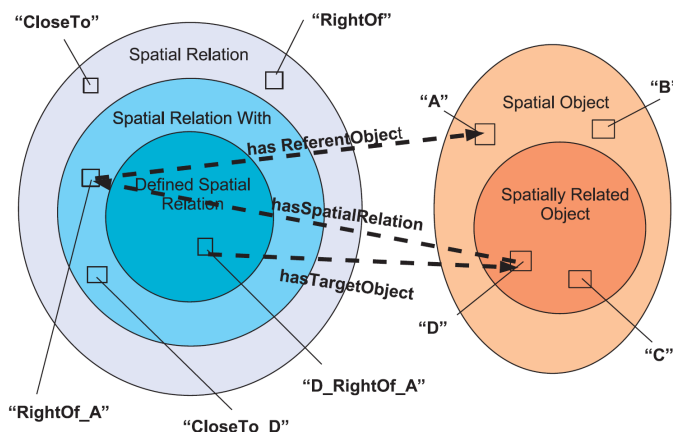


FIGURE 2.26: Diagramme de Venn issu d'une ontologie de relations spatiales pour l'analyse d'images présentée dans [HAB08].

Il s'agit d'un *diagramme de Venn* qui représente les concepts, utiles à l'interprétation d'image, identifiés dans l'ontologie proposée par les auteurs. Dans cette ontologie, il a été fait le choix de représenter les relations spatiales en tant que concepts à part entière (et non pas en tant que relations entre des objets). Trois concepts représentent, à différents niveaux, les relations spatiales entre objets : *Spatial Relation*, *Spatial Relation With* et *Defined Spatial relation*. Le concept de plus haut niveau (*Spatial Relation*) fait référence à une relation spatiale entre des objets, dans l'absolu. Des exemples de *Spatial Relation* sont notamment *CloseTo*, *RightOf*, qui sont des relations spatiales linguistiques. Concrètement, ces *Spatial Relation* absolues sont représentées par le choix d'un élément structurant (dans l'approche morphologique) ou d'une fonction \mathcal{F} (dans l'approche \mathcal{F} -template). Le concept

2.3. Positionnement de nos travaux

de *Spatial Relation With* correspond à une *Spatial Relation* qui a été appliquée sur un objet de référence (ce qui est conforme au procédé cognitif décrit par Logan *et al.*). Sur le schéma, un lien *has reference object* représente la relation entre le concept *Spatial relation With* et le concept *Spatial Object* qui désigne son objet de référence. Des exemples de concepts *Spatial relation With* sont donnés : *à droite de A, proche de D...* Concrètement, ces concepts sont des modèles spatiaux, représentés par des *paysages flous* (approche morphologique), ou par des \mathcal{F} -templates. Enfin, le concept central de *Defined Spatial Relation* décrit une relation spatiale complètement définie (avec sa référence et son argument) et peut donc être évaluée (par exemple : *D est à droite de A*).

Il est intéressant de remarquer que les deux premiers concepts cités font écho aux notions de vocabulaire récemment distinguées par Matsakis & Wendling [MWN10], sous les noms respectifs de *relationship* et *relationship to a reference*. Ces auteurs distinguent encore le terme de *relative position*, qui évoque plutôt un descripteur de relation spatiale (par exemple un \mathcal{F} -histogramme).

Selon l'ontologie schématisée par la figure 2.26, la notion de méta-modèle spatial que nous introduisons est assimilable au concept le plus général : *Spatial Relation*. Il s'agit en effet d'une abstraction de relation spatiale dans l'absolu, qui pourra être instanciée sur un objet de référence puis évaluée pour un objet cible. La nouveauté conceptuelle est donc le pouvoir d'apprendre automatiquement cette catégorie de relations spatiales à partir de données et sans information a priori.

Dans le chapitre 3, nous définissons et formalisons les méta-modèles spatiaux. Nous montrons comment la combinaison de connaissance sur les directions et les distances permet de définir automatiquement, à partir de données représentatives, une description abstraite des relations spatiales et comment celle-ci peut être exploitée sous la forme de modèles spatiaux qui s'adaptent aux formes des objets de référence. Nous mettrons l'accent sur la capacité de la modélisation à considérer des relations spatiales entre des objets de nature et de complexité variable, tout en permettant d'exprimer une large palette de situations de positionnement spatial d'un objet relativement à un autre.

Le chapitre 4 présentera une exploitation aboutie de l'information spatiale au sein de formes manuscrites structurées, qui mettent en jeu de multiples relations de positionnement. Dans le cadre du traitement de symboles manuscrits structurés, nous chercherons à construire des modèles de représentation basés uniquement sur l'information du positionnement relatif de leurs composants, décrite au moyen de méta-modèles spatiaux. Les enjeux d'identification de références de positionnement et de segmentation des tracés seront à considérer. La démarche poursuit le double objectif de confirmer l'importance de l'information spatiale pour la perception de symboles et de mettre en évidence la qualité de la description fournie par les méta-modèles spatiaux. Nous ferons notamment usage de la propriété originale de *localisation* afin de guider la segmentation des tracés par ces modèles de positionnement appris.

Méta-modèles spatiaux pour l'apprentissage de relations spatiales

Dans ce chapitre, nous présentons la formalisation des méta-modèles spatiaux que nous proposons pour la représentation de relations spatiales non-linguistiques entre éléments manuscrits. Le premier objectif poursuivi est d'offrir une large expressivité vis-à-vis des catégories de relations spatiales, à la lumière des besoins que nous avons rencontrés dans l'état de l'art (chapitre 2). Une généralité vis-à-vis des objets pris en charge est également souhaitable, car elle permet l'application des mêmes modèles sur des objets de nature et de niveaux de complexité divers. Au regard du problème posé par les approximations d'objets au moyen de leur boîte englobante, on s'attachera aussi à prendre en compte plus précisément les formes des objets positionnés, car c'est la seule façon d'aboutir à une modélisation fine et conforme à l'intuition de leur positionnement relatif. Ces objectifs seront adressés via le principe d'exploitation de modèles spatiaux, qui seront de plus formalisés dans une représentation floue, ce qui permettra une description réaliste des concepts de relations spatiales tout en prenant en compte l'imprécision et la variation sur le positionnement relatif des objets manuscrits. Enfin, la possibilité d'apprentissage que nous apportons permettra de construire ces méta-modèles spatiaux pour décrire des relations non linguistiques qui bénéficient de tous les avantages déjà cités de raisonnement par modèles spatiaux. L'un de ces avantages, qui réside dans la capacité à résoudre le problème de *localisation* d'objets dans l'espace, confère aux méta-modèles spatiaux un pouvoir de *prédiction* tout à fait original exprimant *dans quelle zone de l'image peut se trouver un objet B par rapport à un objet A selon la relation spatiale apprise*.

Ce chapitre est organisé en sept sections. L'approche est d'abord introduite dans les grandes lignes (section 3.1). Ensuite, nous présentons la définition formelle des méta-modèles et les détails de leur mise en œuvre qui exploite la description par opérateurs de morphologie mathématique (section 3.2). L'information spatiale de distance et son intégration au formalisme sont ensuite traités distinctement dans une section dédiée (section 3.3). La quatrième partie développe une extension des méta-modèles à une modélisation floue *bipolaire*

de l'information spatiale, pour distinguer plus explicitement les zones de « bon » et de « mauvais » positionnement dans l'espace (section 3.4). Une dernière contribution consiste à ajouter aux méta-modèles une notion sur *l'étendue* des objets, dont nous verrons qu'elle permet d'augmenter le pouvoir d'expressivité des méta-modèles spatiaux à la gestion de certaines relations de positionnement topologique (section 3.5). Enfin, le chapitre sera conclu par une présentation de différents résultats qualitatifs donnant un aperçu de leurs capacités de modélisation et de leur généralité (section 3.6).

3.1 Principes

Nous avons mis en lumière le besoin de construire automatiquement, à partir de données, des modélisations des relations spatiales afin de capturer la variabilité du positionnement relatif des objets. Ce besoin est d'autant plus prégnant que nous nous intéressons à des relations entre objets manuscrits, qui présentent des relations spatiales complexes et difficiles à décrire empiriquement. Nous avons présenté en introduction l'exemple du positionnement relatif d'un exposant par rapport à un symbole dans une expression mathématique, mais beaucoup d'autres exemples de relations non-linguistiques peuvent être cités : positionnement relatif de deux primitives d'un caractère chinois, position de la barre d'un « t », position d'un symbole au sein d'un schéma complexe... Pour définir une méthode générique de représentation applicable à tous ces contextes, aucun a priori ne devrait être fait sur les relations spatiales à modéliser entre les objets. Il est donc souhaitable d'éviter le recours à des directions privilégiées de description, de limiter la représentation à certaines situations topologiques entre les objets (par exemple limiter la représentation aux objets qui ne s'intersectent pas, ou dont les boîtes englobantes ne se recouvrent pas), de faire reposer la description sur la définition de seuils de distance minimaux... Afin de garantir la large expressivité, aucune gestion particulière de cas spécifiques ne doit être requise. Que les objets soient distants, proches ou lointains, avec ou sans recouvrements de leurs boîtes englobantes, qu'ils présentent ou non des concavités, qu'ils s'intersectent, se touchent ou bien se confondent, le même procédé de modélisation doit être capable de gérer ces situations. Par ailleurs, la forme des objets doit être exploitée dans toute sa complexité pour définir finement le positionnement spatial. Le choix que nous faisons d'exploiter le principe des modèles spatiaux pour représenter les relations spatiales permet de se prémunir contre toute limitation concernant la nature et la forme des objets, puisqu'il n'est pas nécessaire d'extraire des caractéristiques de positionnement. La généralité de la représentation par rapport aux relations spatiales rencontrées dans le traitement du manuscrit sera notamment assurée par une bonne couverture de toutes les directions de l'espace. Dans un premier temps, nous présentons le principe général de notre approche pour l'apprentissage de méta-modèles spatiaux.

3.1.1 Schéma général et interprétation

Les modèles spatiaux présentés dans la seconde partie de l'état de l'art (voir partie 2.2) ont une interprétation simple, dérivée du terme linguistique qu'ils décrivent. Par exemple, dans le cas de la représentation de la relation décrite par le terme linguistique *à droite de*, le modèle spatial appliqué à un objet de référence *A* pouvait être simplement interprété

3.1. Principes

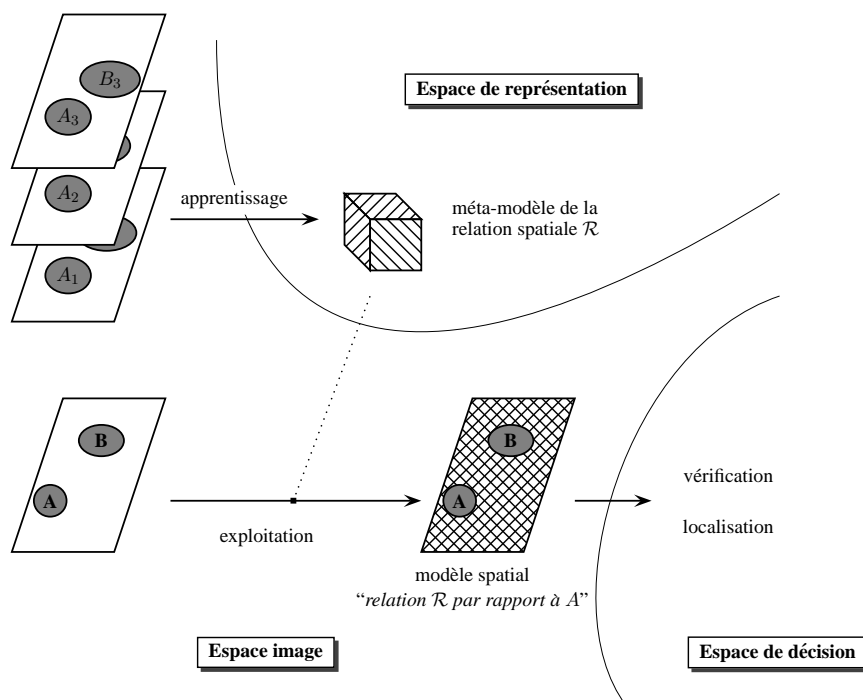


FIGURE 3.1: Schéma général de l'approche proposée pour l'apprentissage et l'exploitation de méta-modèles spatiaux. A partir d'exemples de paires d'objets positionnés selon une relation spatiale \mathcal{R} , un *méta-modèle* de \mathcal{R} peut être défini. Son exploitation se fait selon le principe du raisonnement par modèles spatiaux : étant donnée une image comportant un objet A de référence, le méta-modèle peut être développé dans l'espace image. Il en résulte un modèle spatial décrivant la relation \mathcal{R} par rapport à l'objet A . Ce modèle spatial peut servir directement à la prise de décision, aussi bien par *évaluation* d'un positionnement relatif que par *localisation* d'objet.

comme l'ensemble flou, défini dans le plan, qui associe à tout point son degré d'adéquation par rapport à la proposition logique *être à droite de A*. La généralisation que nous proposons consiste à apprendre, à partir d'exemples représentatifs d'une relation spatiale \mathcal{R} entre deux objets, une abstraction de cette relation (en général non-linguistique) que nous appelons *méta-modèle de relation spatiale*. La figure 3.1 résume schématiquement ce principe d'apprentissage et d'exploitation du méta-modèle spatial. Il s'agit en quelque sorte d'essayer de reproduire la capacité d'apprentissage dont est doté le cerveau humain : à partir de plusieurs exemples d'une relation spatiale entre des paires d'objets différents (chacun ayant sa forme singulière), il est possible de construire une représentation abstraite d'une relation spatiale. Ce qui rend cette représentation abstraite intéressante est qu'elle n'est pas définitivement liée aux objets d'apprentissage mais qu'elle peut se généraliser à d'autres objets en adaptant la représentation de la relation à leurs particularités.

Un méta-modèle spatial appris représente donc un concept spatial *être dans la relation \mathcal{R} par rapport à*, dans l'absolu. Le modèle spatial construit par application de cette abstraction sur un objet de référence A est un sous-ensemble flou du plan, qui associe à chaque point

du plan un degré pouvant être interprété comme le degré d'adéquation de ce point à la proposition logique *être dans la relation \mathcal{R} par rapport à A* . Il est construit en fonction de la forme précise de cet objet A , qui est différent des objets ayant servi l'apprentissage du méta-modèle. Comme pour le cas de modèles spatiaux linguistiques, la définition floue du modèle spatial de la relation \mathcal{R} offre une représentation souple des zones d'acceptabilité de la relation et des nuances entre bonnes et mauvaises régions du modèle. Comme tout modèle spatial, celui-ci peut être combiné avec d'autres modèles spatiaux et peut répondre aux tâches de vérification et de localisation d'objets.

3.1.2 Morphologie mathématique et apprentissage

Comme précisé précédemment, nous avons choisi de recourir à la morphologie mathématique floue pour la description des relations spatiales. Dans la partie 2.2.3, nous avons montré que la description de relations spatiales par morphologie mathématique utilise l'opération de *dilatation* morphologique entre l'objet qui est la référence du modèle spatial et un ensemble appelé *élément structurant*. C'est l'élément structurant qui paramètre la forme de la relation spatiale décrite par rapport à la référence, et nous avons évoqué différentes formes d'éléments structurants pouvant être utilisés pour la description de relations telles que *être à droite de* ou *être proche de*. Une stratégie pour apprendre automatiquement à générer des modèles spatiaux pour des relations non-linguistiques peut donc passer par la modification des éléments structurants impliqués dans les opérations de dilatation.

Quelques travaux ont exploré l'idée de paramétrer automatiquement des éléments structurants à partir de données. Les paramètres appris concernent les formes des fonctions floues appliquées sur les mesures d'angles ou de distances dans [AHF⁺07] et [CCB06]. Dans le contexte de l'interprétation d'images acquises par IRM, il est proposé de s'appuyer sur les informations de configuration spatiale pour détecter des cas pathologiques (en l'occurrence la présence de tumeurs) dans lesquels une déviation sensible peut être détectée par rapport à l'organisation d'un modèle générique sain. Dans ces travaux, des éléments structurants différents sont distingués pour l'interprétation des cas pathologiques par rapport au cas sains. Les paramètres qui varient entre les modèles sont les paramètres des fonctions floues qui définissent la tolérance de la relation spatiale (de direction ou de distance) décrite. A partir de cas étiquetés comme sains ou non, les moyennes et variances des mesures extrêmes d'angles et de distances entre les objets positionnés sont utilisées pour adapter les niveaux de tolérances des deux modèles. En revanche, les directions des éléments structurants ne sont pas elles-mêmes modifiées. Il s'agit en fait dans ces travaux d'adapter une connaissance générique à la gestion de cas atypiques.

Le problème que nous avons à résoudre est plus ouvert et plus général. Puisque l'apprentissage du modèle de relation spatiale doit être complètement automatique, il n'est pas possible de définir un modèle générique de la relation avant de l'affiner par exemple en apprenant les paramètres des fonctions floues qui le définissent. Il faut donc découvrir à partir des exemples les directions et distances d'intérêt pour la modélisation de la relation.

Une stratégie naïve pour apprendre un élément structurant modélisant une relation spatiale pourrait être de le construire à partir d'une direction (et éventuellement d'une distance)

3.1. Principes

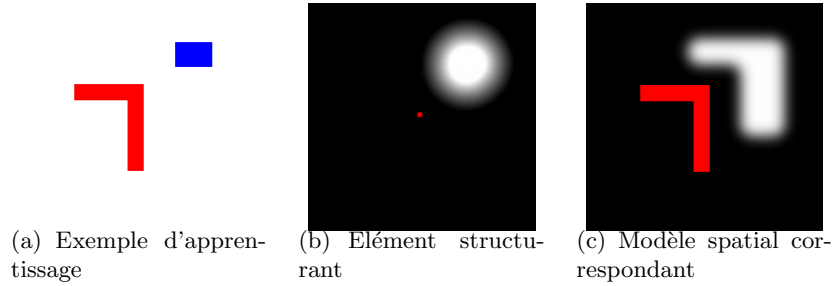


FIGURE 3.2: Exemple d'apprentissage unique (a) pour le vecteur moyen \vec{v} , élément structurant défini par l'équation (3.1) (b) et modèle spatial obtenu (c) pour l'un des objets de référence de la figure 2.22.

moyenne, associée à une variance estimée sur les exemples d'apprentissage. Supposons, pour simplifier, qu'un vecteur \vec{v} soit le résultat de l'apprentissage d'une direction et d'une distance moyennes estimées sur un ensemble d'exemples pour la relation spatiale \mathcal{R} . \vec{v} peut avoir été calculé simplement comme la moyenne des vecteurs entre les centres des boîtes englobantes des objets, à partir de leurs histogrammes d'angles. . . Un élément structurant défini par ce vecteur pourrait être le suivant :

$$\nu_v(p) = h(\|\vec{op} - \vec{v}\|), \quad (3.1)$$

avec h une fonction décroissante dans $[0; 1]$, telle que $h(0) = 1$, et o le centre de l'élément structurant. Cet élément structurant est un sous-ensemble flou du plan qui vaut 1 au point p tel que $\vec{op} = \vec{v}$. Il peut être interprété comme un représentant de la relation spatiale *être translaté de \vec{v} par rapport à*. La figure 3.2(a) montre un exemple unique d'apprentissage à partir duquel un vecteur de déplacement \vec{v} peut être estimé. L'objet pris pour référence est l'un des objets de la figure 2.22, représenté en rouge. La figure 3.2(b) représente visuellement l'élément structurant obtenu à partir ce vecteur moyen. Le point rouge représente le centre de l'élément structurant.

Le modèle spatial obtenu par dilatation morphologique d'un objet par cet élément structurant est illustré à la figure 3.2(c), en utilisant le même objet de référence que dans l'exemple d'apprentissage. On peut constater que le modèle spatial obtenu n'est pas satisfaisant. Au lieu de livrer une description d'une relation spatiale comme souhaité, il décrit en pratique une transformation de l'objet de référence par une translation floue (définie par le vecteur \vec{v} et la fonction h de l'équation (3.1)). L'interprétation qui peut être donnée au sous-ensemble flou obtenu est *être le point translaté de \vec{v} d'un point de R* , où R désigne l'objet de référence.

L'exemple trivial présenté au-dessus fournit une mauvaise description de la relation spatiale souhaitée pour deux raisons qui se cumulent. La première est que l'on ne peut pas simplement résumer la relation spatiale entre deux objets au seul moyen d'un vecteur de déplacement moyen (quelle que soit la manière dont il est estimé). En effet, cette représentation consiste implicitement à simplifier les objets sous la forme de simples points. A moins que les objets ne soient suffisamment distants l'un de l'autre, il n'est pas satisfaisant de les approximer par un point (par exemple leur centre de gravité ou le centre de

leur boîte englobante). La deuxième raison qui rend ce modèle spatial peu satisfaisant est liée à la première et tient à la définition de l'opération de dilatation morphologique. Par définition (voir l'équation (2.3)), un point du plan a pour degré d'appartenance au modèle spatial le maximum de son degré d'appartenance à l'élément structurant lorsque le centre de celui-ci parcourt l'objet de référence. Dans le cas où l'élément structurant est très *précis* (c'est-à-dire qu'il définit une zone d'acceptabilité très réduite dans le plan), le résultat de la dilatation a donc une forme très similaire à celle de l'objet de référence, ce qui ne constitue pas en général une définition acceptable d'un modèle spatial pour la relation que l'on cherche à modéliser.

3.1.3 Combinaison de points de vue directionnels

Le problème pointé du doigt dans le paragraphe précédent montre que la dilatation morphologique par un élément structurant trop précis produit des résultats inappropriés pour la description de modèles spatiaux. Pour parvenir à construire des modèles spatiaux non-linguistiques précis, il est possible de combiner plusieurs modèles spatiaux moins précis, qui portent chacun une partie de l'information sur la relation spatiale. Par définition, et conformément à l'intuition, les modèles spatiaux peuvent se combiner dans le plan image pour fournir une description plus précise d'une relation spatiale complexe, comme l'avait déjà constaté Logan (voir 2.2.2). De plus, la représentation au moyen de sous-ensembles flous permise par la morphologie mathématique floue est parfaitement adaptée pour formaliser ces combinaisons, tout en préservant la modélisation de l'imprécision sur le résultat. Ce principe de combinaison de modèles spatiaux partiels sera au cœur de notre représentation et nous l'exploiterons à plusieurs niveaux pour construire des modèles de positionnement aussi précis que possible.

La richesse de la description permise par la combinaison selon différentes directions provient du fait que la forme de l'objet de référence est alors mieux prise en compte dans toute sa complexité. Pour un point de l'espace donné, l'analyse de sa position par rapport à une référence à la forme complexe sous plusieurs angles de directions (ou selon différents *points de vue*) repose sur différentes portions du tracé de la référence. En effet, la description de la position d'un point est toujours faite en cherchant la partie de la référence qui *supporte* au mieux la relation directionnelle du point de vue considéré. Différents points de vue conduisent à décrire la position d'un même point par rapport à différentes parties de la référence.

La figure 3.3 résume ce principe par la représentation d'un objet de référence présentant une concavité. Les positions de deux points p_1 et p_2 du plan sont décrites par rapport à ce même objet, selon quatre points de vue *en haut*, *en bas*, *à gauche* et *à droite*. Pour chaque point de vue, la portion de la référence qui supporte la description directionnelle est mise en évidence en rouge. Le score d'adéquation μ de la position de p_1 et p_2 par rapport à chaque point de vue est indiqué. On constate que dans le cas de p_1 , qui est positionné à l'intérieur de la concavité, la portion de la référence utile diffère selon chaque point de vue (a-d). Le point peut être ainsi décrit comme à la fois parfaitement *en haut*, *à droite* et *en bas* (d'une partie) ($\mu = 1$ (a,b,d)) et plutôt *à gauche* ($\mu = 0,7$ (c)) de la référence. Le point p_2 n'est pas parfaitement *en haut* ($\mu = 0,8$ (e)) ni *en bas* ($\mu = 0,7$ (f)) d'une portion de la référence, car il est situé en dehors de la concavité. Il est en revanche parfaitement *à droite*

3.1. Principes

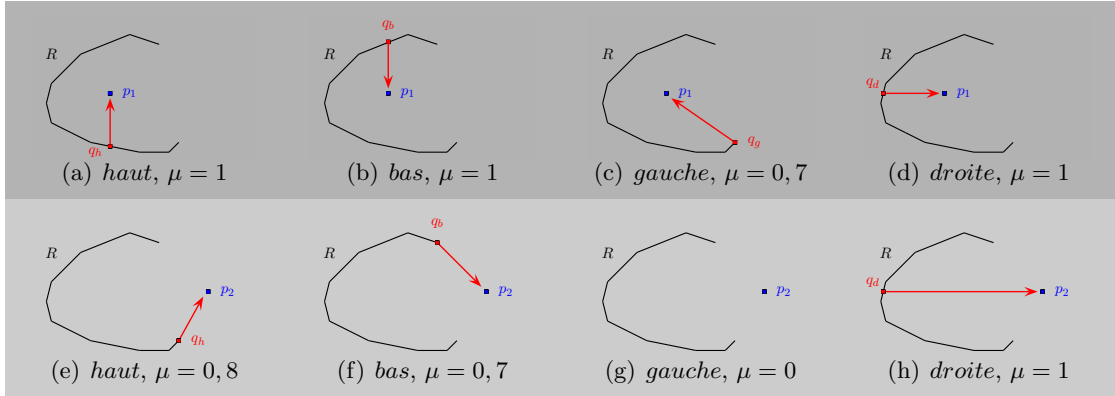


FIGURE 3.3: Représentation, pour un objet de référence R , des différentes portions qui servent de support pour décrire la position de points p_1 (a-d) et p_2 (e-h) selon quatre points de vue directionnels : *en haut*, *en bas*, *à gauche* et *à droite*. Pour chaque situation, le meilleur point q de la référence est noté en rouge et le vecteur qp associé est représenté par une flèche rouge. Le score d'adéquation μ de la position de p par rapport à R selon chaque point de vue est indiqué. p_1 est à la fois parfaitement en haut, en bas et à droite de R . p_2 est parfaitement à droite de R .

d'une partie de R ($\mu = 1$ (h)) et pas du tout *à gauche* de R ($\mu = 0$ (g)). Ce simple exemple confirme que les différents points de vue apportent alors des informations complémentaires qui permettent de préciser plus finement la position du point d'intérêt grâce à l'appui sur des portions différentes de l'objet de référence.

Dans la recherche de généralité des relations exprimables par notre approche, il est important que tout positionnement relatif d'objets puisse être décrit, quelles que soient les directions mises en jeu, sans qu'aucune direction ne soit privilégiée ni négligée. L'idée que nous poursuivons consiste à choisir un ensemble de directions primaires (prédéfinies et linguistiques), qui sont autant de *points de vue* selon lesquels on va analyser la relation spatiale. Chaque point de vue considéré pourra apporter une information différente et complémentaire sur le positionnement relatif des objets. Pour chaque point de vue, nous allons décrire *dans quelle mesure* la relation spatiale \mathcal{R} est en accord avec le point de vue. Par exemple, selon la relation directionnelle linguistique *être à droite*, on va décrire *dans quelle mesure* la relation \mathcal{R} autorise-t-elle un objet à *être à droite* de l'objet de référence. En fonction de la relation \mathcal{R} modélisée, la réponse à cette question pourra varier entre *pas du tout à droite* et *complètement à droite*. La description sera répétée selon plusieurs points de vue, de façon à couvrir toutes les possibilités de positionnement relatif dans le plan. Finalement, la construction du modèle spatial de la relation \mathcal{R} combinerà l'ensemble des points de vue : un point du plan sera considéré conforme au modèle spatial si son positionnement est conforme à la description de \mathcal{R} par rapport à l'objet de référence selon tous les points de vue.

En prolongeant l'analogie avec le raisonnement spatial humain et linguistique, on peut remarquer que la stratégie proposée va fournir des descriptions interprétables et intuitives.

En effet, en reprenant l'exemple des deux objets de la figure 3.2(a), on peut admettre qu'une description intuitive de leur relation spatiale est *plutôt en haut et à droite*. Cette description linguistique repose sur une décomposition de la relation spatiale en deux *points de vue* indépendants, sous laquelle elle est analysée : *en haut* et *à droite*. La relation est évaluée comme étant *plutôt* conforme à chacun des points de vue. Finalement, la description globale exprime que la localisation de l'objet cible est à l'intersection des zones de l'espace qui satisfont à la fois *être plutôt à droite* et *être plutôt en haut* de la référence. Dans ce cas, la précision de la description provient effectivement de la combinaison de plusieurs points de vue. Pour les méta-modèles de positionnement, la précision de la description proviendra de la combinaison de point de vues qui auront fait l'objet d'un apprentissage, chacun apportant des indices renforçant la précision de la description.

La qualité de description offerte par des méta-modèles spatiaux sera vérifiable *visuellement*. En effet, par définition, ils permettent de définir une description directement dans l'espace perceptif. On pourra donc représenter par une image un méta-modèle spatial appliqué à un objet de référence, ce qui fournira un indice visuel de la qualité de description et de la précision des modèles.

3.2 Formalisation et mise en œuvre

Dans cette partie, nous présentons la formalisation choisie pour nos méta-modèles spatiaux ainsi que la mise en œuvre de leur apprentissage à partir de données. Comme nous l'avons introduit plus haut, un méta-modèle spatial est défini par plusieurs *points de vue* partiels sur la relation spatiale modélisée. Ces points de vue sont liés à des directions prédéfinies, représentées par des éléments structurants de direction. Formellement, le méta-modèle de la relation \mathcal{R} est représenté par :

$$\mathcal{M}^{\mathcal{R}} = \langle (\nu_{\alpha_i}, H_{\alpha_i}^{\mathcal{R}}) \rangle_{i=1..d}, \quad (3.2)$$

où ν_{α_i} représente l'élément structurant de direction α_i , $H_{\alpha_i}^{\mathcal{R}}$ est une fonction apprise, qui détermine *dans quelle mesure* la relation portée par ν_{α_i} s'applique aux points de l'espace selon \mathcal{R} et d est le nombre de points de vue considérés dans le modèle. Dans la partie suivante, nous présentons d'abord la définition et l'interprétation données aux fonctions $H_{\alpha_i}^{\mathcal{R}}$, ainsi que leur apprentissage. Par la suite, nous décrirons comment sont choisis les points de vue pour permettre la modélisation générique des relations spatiales. Enfin, nous nous intéresserons à la stratégie de combinaison des points de vue pour la génération d'un modèle spatial global décrivant la relation spatiale apprise.

3.2.1 Apprentissage de points de vue

3.2.1.1 Formalisation

Un point de vue est lié dans notre modélisation à une direction orientée par un angle α , exprimé en radians. Dans l'état de l'art, nous avons vu qu'une définition d'un élément structurant de direction peut être donnée par (voir Bloch [Blo99]) :

$$\nu_{\alpha}(p) = \max \left(0, 1 - \frac{2}{\pi} \arccos \frac{\vec{op} \cdot \vec{u}_{\alpha}}{\|\vec{op}\|} \right), \forall p \in \mathcal{S}, \quad (3.3)$$

3.2. Formalisation et mise en œuvre

avec \mathcal{S} qui désigne l'espace image. Cet élément structurant décrit la relation *être dans la direction* α . Les points situés exactement dans la direction α par rapport à son centre o ont une appartenance parfaite à l'élément structurant, tandis que les points p qui sont dans le demi-plan opposé à la direction α n'y appartiennent pas du tout.

Le but de chaque point de vue directionnel dans notre description est de fournir une information sur la relation spatiale \mathcal{R} , en exprimant dans quelle mesure \mathcal{R} est compatible ou non avec la relation directionnelle décrite par ν_α . Pour cela, nous considérons qu'un modèle spatial issu d'un point de vue sur une relation est le résultat de la composition fonctionnelle de la dilatation morphologique d'un objet de référence R par l'élément structurant directionnel original ν_α avec une fonction intermédiaire $H_\alpha^\mathcal{R}$:

$$\mu_\alpha^\mathcal{R}(R) = (H_\alpha^\mathcal{R} \circ \mu_\alpha(R)). \quad (3.4)$$

On rappelle que $\mu_\alpha(R)$ est le modèle spatial *classique* représentant la relation linguistique *être dans la direction* α par rapport à R . Il est défini par dilatation morphologique de l'objet R au moyen de l'élément structurant ν_α (cette opération a été présentée en détails au paragraphe 2.2.3, voir l'équation (2.3)). $H_\alpha^\mathcal{R}$ est une fonction de $[0; 1]$ dans $[0; 1]$, qui peut être vue comme un *filtre* flou appliqué sur le modèle spatial directionnel $\mu_\alpha(R)$. Le résultat $\mu_\alpha^\mathcal{R}(R)$ est, comme $\mu_\alpha(R)$, un sous-ensemble flou du plan, qui associe à tout point de \mathcal{S} un degré d'appartenance entre 0 et 1.

La forme de la fonction de transformation $H_\alpha^\mathcal{R}$ influe directement sur la forme et l'interprétation du modèle spatial $\mu_\alpha^\mathcal{R}(R)$. Cette forme est a priori quelconque et peut varier grandement en fonction de la relation \mathcal{R} . Quelques exemples de formes de fonctions H et de leur interprétation associée sont donnés au paragraphe 3.2.1.3. Le plus souvent, la forme de H définit un modèle spatial qui représente une relation spatiale impossible ou difficile à exprimer linguistiquement.

3.2.1.2 Apprentissage

L'apprentissage d'un point de vue directionnel consiste à déterminer la forme de sa fonction de composition H (voir l'équation (3.4)). Comme la forme de la fonction H est inconnue a priori, nous proposons de l'approximer par un histogramme. Cela permet de ne pas faire d'hypothèses et donc de ne pas contraindre sa forme.

Considérons un ensemble de paires d'objets d'apprentissage $(R_i, A_i)_{i=1..N}$, qui sont positionnés relativement selon une relation spatiale \mathcal{R} dont nous souhaitons construire un méta-modèle. Dans chaque paire, l'objet R_i constitue la référence de positionnement et l'objet A_i l'objet argument (ou objet cible). Pour un point de vue déterminé par la direction α , nous cherchons à décrire *dans quelle mesure les objets A_i sont dans la direction α par rapport à leur référence R_i* . Pour cela, nous considérons la distribution des degrés d'adéquation des points des objets A_i avec les modèles spatiaux directionnels *être dans la direction* α par rapport à R_i . Ces modèles spatiaux $\mu_\alpha^{R_i}$ sont obtenus par dilatation morphologique de l'objet R_i par l'élément structurant flou de direction α : ν_α . C'est cette distribution que nous cherchons à approximer par un histogramme H . La figure 3.4 illustre ce principe d'apprentissage.

L'illustration (a) montre des exemples d'objets de référence R_i (en rouge) et d'objets cible A_i (en bleu) qui sont positionnés selon une relation spatiale \mathcal{R} que l'on veut modéliser.

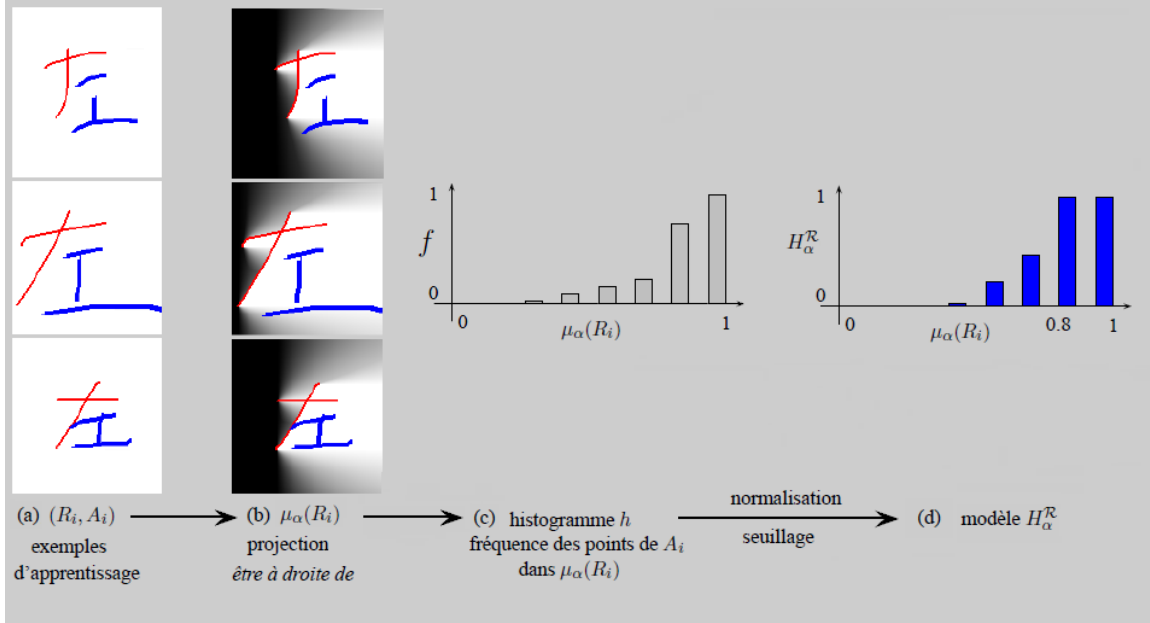


FIGURE 3.4: Apprentissage à partir d'exemples (a) de la fonction H^R normalisée (d) modélisant la relation \mathcal{R} selon le point de vue à droite de. Les objets d'apprentissage sont d'abord décrits selon la relation *être à droite* (b) et les degrés atteints dans ces modèles spatiaux par les points des objets cibles A_i (en bleu) sont comptabilisés dans un histogramme de fréquences h (c). Cet histogramme est normalisé et seuillé pour former le modèle flou H^R (d). Ce modèle décrit les degrés admissibles des objets cibles pour la relation modélisée, selon le point de vue à droite. Il exprime notamment que les points *très à droite* de R_i sont parfaitement admis par le modèle : $H_\alpha^R(x) = 1$ pour $x \geq 0,8$.

Le modèle spatial *être à droite de* est développé sur la référence de chaque exemple d'apprentissage (b). L'histogramme h représenté en (c) comptabilise la distribution des points des objets A_i d'apprentissage (en bleu) par rapport à leur degré d'adéquation avec la relation *être à droite de* R_i dans les modèles spatiaux $\mu_\alpha(R_i)$. Cet histogramme h mesure donc la fréquence de points des objets d'apprentissage atteignant les différents niveaux de gris dans les images (b). Il doit ensuite être normalisé et seuillé pour définir le modèle H^R (d) sous la forme d'un ensemble flou. H^R associe aux différentes valeurs de $\mu_\alpha(R_i)$ (degrés d'*être à droite*) leur degré de représentativité de la relation spatiale apprise.

La figure 3.5 rappelle le principe d'exploitation de ce modèle H^R , formalisé précédemment par l'équation (3.4). Pour un objet de référence R , la description *être à droite* ($\mu_\alpha(R)$) (image (a)) est composée avec le modèle appris H^R (image (b)) pour former un modèle spatial transformé μ_α^R (image (c)). Ce modèle spatial décrit les zones admises pour l'objet cible par rapport à la référence selon la relation \mathcal{R} apprise en considérant le point de vue à droite.

Nous présentons ci-dessous plus formellement les différentes étapes du procédé d'apprentissage du modèle H^R exposées à la figure 3.4.

3.2. Formalisation et mise en œuvre

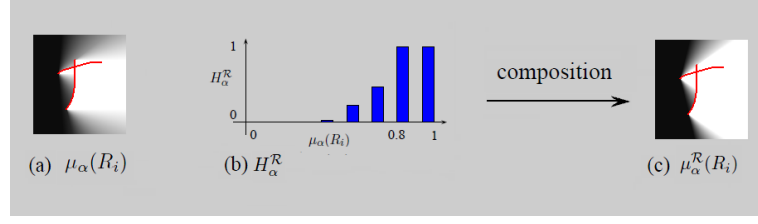


FIGURE 3.5: Exploitation du modèle $H^{\mathcal{R}}$ sur un nouvel objet R . Le modèle (b) peut être composé avec la description être à droite (a) pour former un modèle spatial appris (c) qui décrit les degrés admissibles pour l'objet cible attendu d'après la relation \mathcal{R} , selon le point de vue à droite.

Définition de l'histogramme de fréquences La distribution des valeurs d'adéquation des objets d'apprentissage avec les modèles spatiaux $\mu_{\alpha}^{R_i}$ est approximée par un histogramme noté $h_{\alpha}^{\mathcal{R}}$. Pour $k = 0..K$, on note $i_k = k/K \in [0; 1]$ et on dénomme $(I_k)_{k=0..K+1}$ les ensembles de points définis par $I_0 = \{i_0 = 0\}$, $I_1 =]i_0; i_1[$, $I_k = [i_{k-1}; i_k[$ ($k=2..K$), $I_{K+1} = \{i_K = 1\}$ (les I_k forment une partition de $[0; 1]$). Chaque section de l'histogramme $(h_{\alpha}^{\mathcal{R}})_k$ comptabilise le nombre de points des exemples d'apprentissage qui ont un degré d'appartenance à $\mu_{\alpha}^{R_i}$ dans l'ensemble I_k :

$$(h_{\alpha}^{\mathcal{R}})_k = \sum_{i=1..N} |S_{i_k}| \quad \text{avec} \quad S_{i_k} = \{p \in A_i, \mu_{\alpha}^{R_i}(p) \in I_k\}. \quad (3.5)$$

Cet histogramme de fréquences doit ensuite être normalisé et transformé pour former la fonction H à valeurs dans $[0; 1]$. En premier lieu, la normalisation est essentielle et doit garantir que la valeur 1 est atteinte par l'histogramme. Ainsi l'ensemble flou appris (et donc le modèle de la relation spatiale) sera un ensemble flou *normal*, c'est-à-dire admettant certains points comme lui étant parfaitement compatibles. Ensuite, l'interprétation de la fonction H que l'on cherche à déterminer n'est pas exactement celle que l'on peut faire de l'histogramme de fréquences h . La fonction H recherchée doit représenter *les degrés de positionnement admissibles pour les points de A par rapport à R selon la direction α* . Il est donc nécessaire de définir la notion d'*admissibilité* par rapport aux fréquences mesurées par h .

Normalisation L'histogramme h comptabilise tous les points des objets d'apprentissage pour la relation \mathcal{R} selon le point de vue considéré. Le nombre de points dans l'histogramme dépend de la taille des objets A_i et de l'échantillonnage des objets (le nombre de points considérés par unité de surface ou de longueur). Le nombre de points peut donc être très variable entre deux objets.

Soit h_{max} la fréquence maximale mesurée dans l'histogramme h . h_{max} représente le nombre maximal de points des objets d'apprentissage qui tombent dans le même intervalle de degrés lorsque l'on considère la relation être dans la direction α par rapport à R_i . Il faut alors remarquer que les ensembles de degrés (I_k) , ne couvrent pas des portions du plan équivalentes. En effet, les ensembles extrêmes I_0 et I_{K+1} couvrent des sections de l'image plus grandes que les autres intervalles. La figure 3.6 illustre les zones du plan couvertes par les ensembles $I_{k,k=0..K+1}$ avec $K = 2$ et $K = 4$, en prenant pour exemple le point de vue

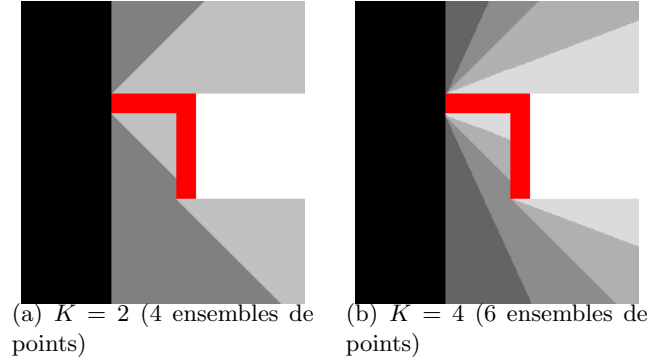


FIGURE 3.6: Représentation des zones liées aux intervalles I_k de l'histogramme pour la relation *être à droite de R*. Un niveau de gris est affecté aux zones associées à un même ensemble I_k .

être à droite de en référence l'objet de la figure 2.22. Dans les deux cas, la zone la plus sombre contient les points du plan dont les degrés d'appartenance au modèle spatial *être à droite de R* sont dans I_0 (c'est-à-dire nuls). Inversement, la zone blanche contient les points parfaitement à droite de R , dont les degrés sont dans I_{K+1} (c'est-à-dire égaux à 1). En comparant les deux figures, construites avec quatre et six régions pour autant d'intervalles, on remarque que les régions extrêmes correspondant à I_0 et I_{K+1} ne changent pas, tandis que les zones des autres ensembles $I_{k,k=1..K}$ diminuent de taille lorsque K augmente.

Ces figures illustrent le phénomène d'accumulation de points aux degrés extrêmes. Compte tenu de la définition de l'élément structurant de la relation *être à droite de*, en moyenne une grande partie du plan image a un degré nul ou parfait d'appartenance à la relation. En pratique, lorsque l'on comptabilise les points qui appartiennent à chacune de ces zones, on constate une accumulation de points dans les tranches extrêmes de l'histogramme h . La normalisation de cet histogramme, qui consiste à le ramener à une fonction à valeurs dans $[0; 1]$, ne peut donc pas se faire simplement en le divisant par sa fréquence maximale h_{max} . Il faut en fait distinguer deux cas. Dans le cas général, on considère que $h_{max} = \max_{k=1..K} (h_k)$; h_{max} est donc le nombre maximal d'occurrences de points dans un intervalle non-extrême de l'histogramme. On normalise ensuite tout l'histogramme à l'aide de h_{max} :

$$\hat{h}_k = \min \left(\frac{h_k}{h_{max}}, 1 \right), k = 0..K + 1. \quad (3.6)$$

Un cas particulier est cependant à considérer. Il se produit lorsque l'essentiel des occurrences comptabilisées par h se concentrent sur l'une des extrémités (ou sur les deux), c'est-à-dire lorsque $h_0 + h_{K+1} > \beta * \sum_{i=0}^{K+1} h_i$ (dans notre implémentation, nous fixons le seuil β à 0,9). Dans ce cas (qui se produit par exemple dès que tous les objets A_i sont dans la zone *pas du tout à droite de R*), tout l'histogramme est normalisé par $h_{max} = \max(h_0, h_{K+1})$.

Seuillage Une fois l'histogramme normalisé, \hat{h} doit être transformé pour transposer la notion de fréquence en terme d'appartenance à un modèle flou. Nous proposons d'appliquer une fonction de seuillage sur les fréquences relatives de \hat{h} qui est déterminée par deux

3.2. Formalisation et mise en œuvre

paramètres a et b . Cette fonction g , croissante sur $[0; 1]$ et telle $g(0) = 0$ et $g(1) = 1$ est définie simplement par :

$$g(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } x \geq b. \end{cases} \quad (3.7)$$

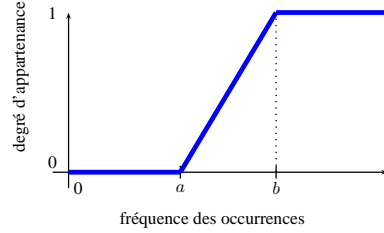


FIGURE 3.7: La fonction g permet de seuiller les fréquences trop rares (inférieures au seuil a) dans la transformation des fréquences en degrés d'adéquation au modèle. Le seuil b fixe la fréquence de représentativité parfaite du modèle.

L'utilisation de fonctions de formes sigmoïdes ou autres *S-fonctions* est courante pour la transformation de fréquences en degrés d'appartenance floue (voir par exemple [KKGK02]). L'idée est de contrôler le lien entre les fréquences mesurées et le degré de conformité à l'ensemble flou résultant. Dans le cas de la fonction g de l'équation (3.7), ce contrôle dépend des deux paramètres a et b . Ici, a détermine la fréquence minimale à partir de laquelle un élément x de $[0; 1]$ sera considéré comme représentatif de la relation. Autrement dit, une zone de l'image liée à I_k telle que $\hat{h}_k < a$ dans l'histogramme normalisé est considérée comme *pas du tout acceptable* pour la relation spatiale ($H_k^R = 0$). Inversement, b détermine la fréquence minimale au-delà de laquelle un élément x est considéré comme ayant un degré d'être à droite parfaitement représentatif de la relation ($H_k^R = 1$).

Un seuil a strictement supérieur à 0 signifie que certains points des exemples d'apprentissage vont être considérés comme des anomalies qui ne doivent pas être prises en compte par le modèle, car trop rares. Un seuil b strictement inférieur à 1 signifie que plusieurs zones du plan (associées à des degrés d'être à droite différents) peuvent être considérées simultanément comme parfaitement en accord avec le modèle. Si $b = 1$, en revanche, seule la zone qui comporte le plus d'occurrences sera considérée comme parfaitement en accord avec la relation apprise.

L'histogramme H est donc finalement déduit de l'histogramme de fréquences normalisées par application de g :

$$H_k = g(\hat{h}_k), \quad k = 0..K + 1. \quad (3.8)$$

3.2.1.3 Exemples

La figure 3.8 présente six exemples de relation spatiales conduisant à des formes d'histogrammes différents.

Pour chacun des six exemples, un seul exemple d'apprentissage est utilisé (représenté par l'image de la première colonne à gauche). La relation directionnelle à droite de est appliquée et un histogramme de fréquences normalisées est construit pour ce point de vue

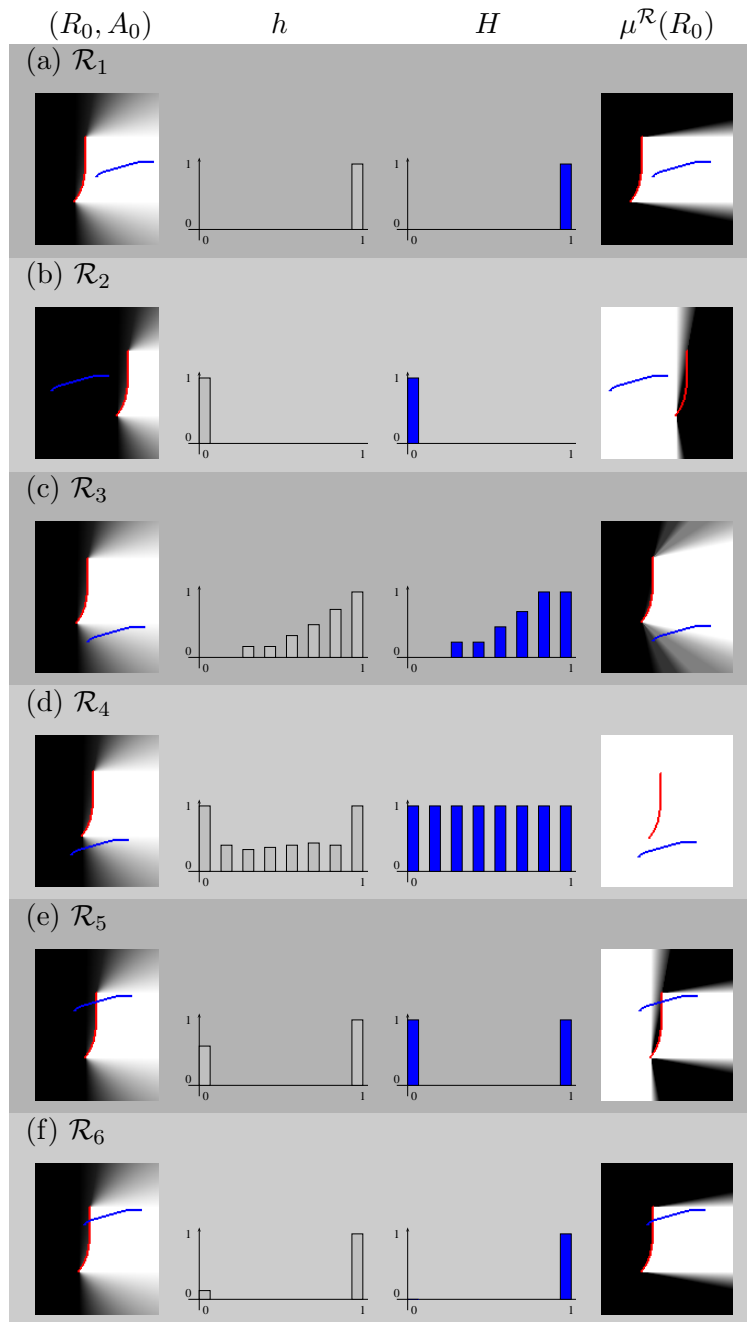


FIGURE 3.8: Exemples d'apprentissage d'histogrammes pour différentes relations spatiales. Pour chaque cas, un exemple d'apprentissage (R_0, A_0) est présenté avec le modèle spatial à droite de. L'histogramme de fréquences brut h est représenté en gris (fréquences de points d'apprentissage en fonction des niveaux de gris dans l'image de la première colonne). L'histogramme normalisé H est représenté en bleu ; il constitue le modèle appris. La dernière colonne illustre l'application de ce modèle sur le même objet de référence.

3.2. Formalisation et mise en œuvre

(deuxième colonne). Il s'agit de l'histogramme noté jusqu'ici par h , normalisé par sa valeur maximale. Il représente directement les fréquences mesurées sur les points d'apprentissage. La troisième colonne représente l'histogramme H qui est issu de la normalisation et du seuillage de h par la fonction g (voir (3.8)). Enfin, la dernière colonne représente la même paire d'objets que la première colonne, avec le modèle spatial obtenu par combinaison du modèle spatial directionnel avec la fonction apprise H . L'interprétation du modèle spatial représenté dans cette dernière colonne diffère de la première colonne : ce modèle représente la relation *être à droite de R dans la même mesure que les objets d'apprentissage selon \mathcal{R}* .

Le premier exemple (a) représente un cas où l'objet à positionner est parfaitement à droite de la référence. Les points comptabilisés par l'histogramme ont tous un degré d'appartenance au modèle spatial *à droite de* qui vaut 1. L'histogramme H appris décrit donc le fait que, selon la relation spatiale \mathcal{R}_1 , l'objet cible doit être *strictement à droite* de la référence. Le modèle spatial obtenu représente bien cette description, puisque la zone d'admissibilité (en blanc) est plus réduite et stricte que celle de la relation linguistique *à droite de*.

Le second exemple est l'exact opposé du premier. Selon \mathcal{R}_2 , l'objet cible n'est pas du tout *à droite de* la référence. Le modèle spatial résultat représente cela en excluant toute la partie du plan à droite de la référence et en considérant comme parfaitement acceptable tout ce qui n'est pas du tout à droite. En revanche, aucune information n'est prise en compte sur la position de l'objet cible parfaitement à gauche ou non de la référence.

Selon \mathcal{R}_3 , l'objet cible est positionné plutôt en bas et plutôt à droite de la référence. Les degrés pris par les points d'apprentissage sont plus nuancés, quelques-uns étant peu à droite quand d'autres, plus nombreux, sont parfaitement à droite. L'histogramme est seuillé et normalisé par rapport à sa valeur maximale en excluant les valeurs extrêmes pour former H . Le modèle spatial résultat accorde une acceptabilité différente aux points de l'objet cible en fonction de la fréquence qui a été mesurée dans chaque zone. Les points parfaitement à droite de la référence sont considérés comme parfaitement en accord avec \mathcal{R}_3 . Les points légèrement à droite de la référence, qui étaient moins nombreux en apprentissage, sont considérés comme moins bons mais néanmoins acceptables par le modèle.

Dans l'exemple \mathcal{R}_4 , les points de l'objet cible couvrent toutes les nuances de degrés par rapport à la relation *être à droite*. Après normalisation, et grâce au paramètre b de la fonction de seuillage, avec $b < 1$ (voir (3.7)), tous les degrés sont considérés comme acceptables pour \mathcal{R}_4 . Il en résulte un modèle spatial complètement permissif, qui n'exclut de la zone d'acceptabilité aucune partie du plan. Cela signifie que le point de vue directionnel *à droite* n'apporte pas d'information sur la relation spatiale \mathcal{R}_4 .

Le cinquième exemple met en jeu deux objets qui s'intersectent. L'objet cible s'étend donc de part et d'autre de la référence (à gauche et à droite). L'histogramme apprend donc que les points peuvent être *pas du tout à droite* ou *parfaitement à droite* (mais qu'aucune nuance entre les deux n'est permise). Le modèle spatial résultat présente donc deux zones distinctes d'acceptabilité pour la relation \mathcal{R}_5 .

Enfin, la relation \mathcal{R}_6 est entachée d'ambiguïté topologique, puisque les objets s'intersectent mais seule une très petite proportion de l'objet cible est à gauche de la référence, quand tout le reste est à droite. L'histogramme de fréquences illustre ce déséquilibre. Ici, c'est l'influence du paramètre a de la fonction de seuillage g qui est mis en évidence (voir (3.7)). Dans ce cas, la proportion de points pas du tout à droite de la référence est considé-

rée comme négligeable (car inférieure à a). Le modèle considère donc que l'objet cible doit être strictement à droite de la référence et le résultat est identique à la relation spatiale \mathcal{R}_1 . Si la proportion de points pas du tout à droite était plus importante (entre a et b), le modèle appris comporterait deux zones (comme pour la relation \mathcal{R}_5), mais avec une différence d'acceptabilité (la zone à gauche étant moins acceptable que la zone à droite). Bien entendu, les modèles sont en réalité appris à partir d'un nombre important d'exemples et l'ambiguïté sur la relation en est donc significativement réduite.

3.2.2 Sélection des points de vue directionnels

Nous avons exposé l'intérêt de combiner différents points de vue sur une relation spatiale pour enrichir la précision de sa description et nous venons de montrer comment l'information sur une relation spatiale peut être prise en compte au niveau d'un point de vue grâce à l'apprentissage d'une fonction de composition. Nous nous intéressons maintenant à la question du choix des points de vue. Sous quels angles faut-il analyser une relation spatiale pour en offrir la meilleure description possible ?

L'idée de notre modélisation est de représenter l'information sur la relation spatiale au niveau de chaque point de vue en limitant les zones directionnelles acceptables (en éliminant les zones inacceptables). C'est l'intersection de plusieurs points de vue qui doit apporter la précision finale du modèle. Il faut donc choisir un éventail de directions (points de vue) permettant de décrire des relations riches par intersection. La question de la complémentarité des directions choisies est donc essentielle. Nous proposons ici d'abord une stratégie générique qui fournit des jeux de points de vue utilisables pour la description de relations spatiales en général. Des stratégies alternatives sont ensuite présentées pour répondre à des problèmes plus spécifiques comme par exemple l'identification automatique de relations spatiales.

3.2.2.1 Stratégie générique

L'objectif d'une sélection générique de points de vue est de permettre la description de relations spatiales aussi variées que possibles, sans limiter ni favoriser certaines directions de l'espace. C'est bien sûr une répartition régulière des directions sur les angles entre 0 et 2π qui permet la meilleure couverture de toutes les directions du plan.

Choix des directions Il est important de bien couvrir toutes les directions de $[0; 2\pi]$ et non seulement les directions du demi-espace orienté par $[0; \pi]$ car les points de vue opposés apportent une information *complémentaire* et non pas redondante. En d'autres termes, on n'a pas accès à la même information sur la relation spatiale analysée selon le point de vue de direction α et le point de vue opposé, de direction $\alpha + \pi$. Pour s'en convaincre, on peut considérer les exemples de la figure 3.9. Dans le premier exemple, un objet cible est situé complètement à droite de son objet de référence 3.9(a). Cette relation spatiale est analysée sous les points de vue opposés à *droite de* et à *gauche de*. Dans chaque cas, le modèle spatial appris par la méthode exposée précédemment est appliqué sur l'objet de référence. Le premier décrit que l'objet cible est *complètement à droite de* la référence 3.9(b). Le second décrit que l'objet cible *n'est pas du tout à gauche* de la référence 3.9(c). Bien que partiellement redondants, les deux modèles spatiaux ne sont pas identiques et aucun des

3.2. Formalisation et mise en œuvre

deux n'est inclus dans l'autre. Dans le second exemple de la figure 3.9, une paire d'objets qui s'intersectent est considérée 3.9(d). Les deux points de vue considérés sont *en haut* (e) et *en bas* (f). On voit que les informations décrites selon les deux points de vue sont très complémentaires car la forme des deux modèles spatiaux est très différente. Cela est dû au fait que, par définition de l'opération de dilatation morphologique, ce ne sont pas les mêmes points de la référence qui permettent de dire qu'un point p du plan est *en haut* ou *en bas* de la référence R . En particulier, les points de l'objet cible sont à la fois *en haut* de la partie inférieure de R et *en bas* de la partie supérieure de R .

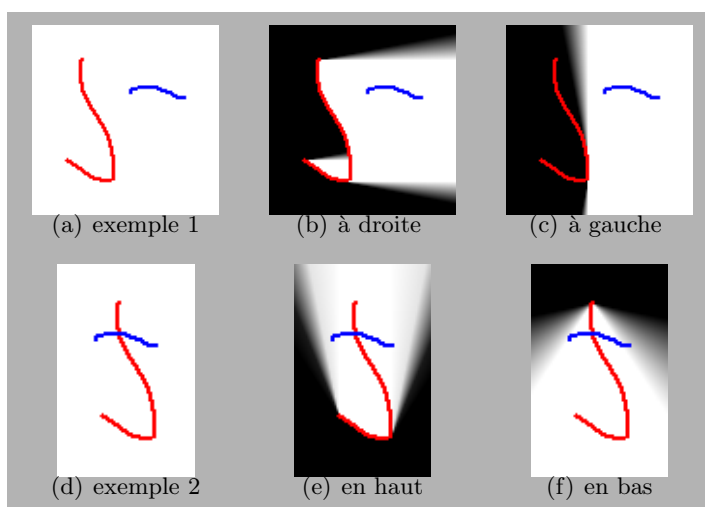


FIGURE 3.9: Complémentarité des points de vue opposés. Pour l'exemple 1 (a), les modèles spatiaux appris selon les points de vue *à droite* (b) et *à gauche* (c) comportent une part d'information complémentaire, car ce ne sont pas les mêmes portions de la référence qui servent de support à la description selon ces deux directions. Il en va de même pour l'exemple 2 (d) avec les points de vue *en haut* (e) et *en bas* (f).

Nombre de directions Si le nombre de points de vue participant à un modèle n'est pas contraint dans la définition donnée au paragraphe 3.2.1.1, il est souhaitable en pratique de le limiter à un nombre réduit (de l'ordre d'une dizaine au maximum). En effet, la complexité du modèle et surtout le coût de son exploitation augmentent linéairement avec le nombre de points de vues. Un compromis est donc nécessaire entre la précision permise par la multiplicité des points de vue d'une part, la simplicité et la performance d'exploitation du modèle d'autre part. En pratique, quatre ou huit axes directionnels semblent constituer de bons compromis.

Quatre axes cardinaux Finalement, le choix le plus naturel consiste à considérer quatre axes orientés vers les quatre points cardinaux *Nord*, *Sud*, *Est* et *Ouest*, qui sont aussi dénommés par les termes linguistiques *haut*, *bas*, *gauche* et *droite*. La figure 3.10(a) représente ces quatre directions. Dans ce cas, on a donc $\alpha_0 = 0$, $\alpha_1 = \pi/2$, $\alpha_2 = \pi$ et $\alpha_3 = 3\pi/2$.

La complémentarité de ces quatre points de vue est confirmée par la forme des quatre éléments structurants directionnels correspondants. Pour rappel, un élément structurant

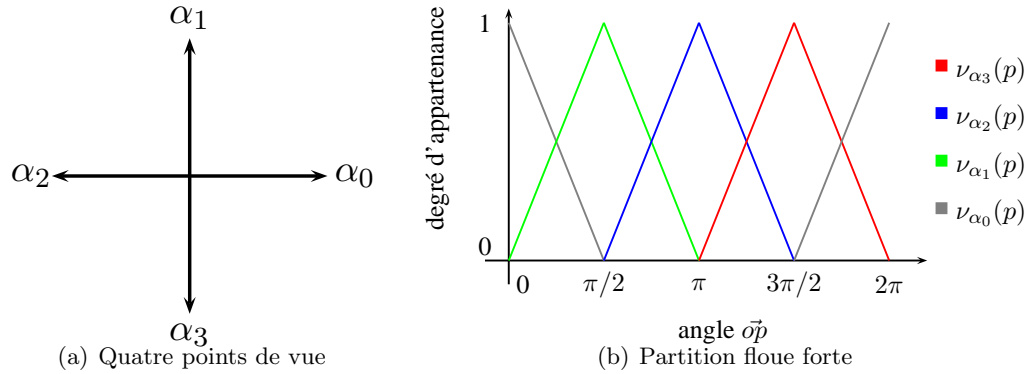


FIGURE 3.10: (a) Quatre points de vues orientés selon les directions cardinales. (b) Les quatre éléments structurants associés forment une partition floue forte dans le plan lorsque l'objet de référence est réduit à un point

directionnel orienté par une direction α est défini par un sous-ensemble flou dans le plan, dont le support est le demi-plan orienté par α (voir équation (3.3)). En imaginant le cas limite où l'objet de référence est réduit à un point unique, les quatre dilatations obtenues par les quatre éléments structurants forment une *partition floue forte* du plan. Une *partition floue forte* est définie par des ensembles flous $(\mu_i)_i$ du même domaine \mathcal{X} tels que (voir pour plus de détails [BMM03]) :

$$\forall x \in \mathcal{X}, \sum_i \mu_i(x) = 1.$$

Cette propriété garantit que tout point du plan peut être complètement décrit par son appartenance à un ou plusieurs éléments structurants directionnels, même dans le cas où la référence est réduite à un point. La figure 3.10(b) permet de s'en rendre compte simplement, en traçant les fonctions d'appartenance aux éléments structurants en fonction de l'angle $\vec{\sigma}_p$ (o étant le centre des éléments structurants, c'est-à-dire le point référence, et p étant n'importe quel point du plan). Le choix des quatre points de vue cardinaux garantit donc une description satisfaisante des positions dans toutes les directions malgré un petit nombre de points de vue.

3.2.2.2 Stratégies alternatives

Des stratégies alternatives sont imaginables pour la sélection de points de vue différents. Une première idée peut être d'opérer une rotation des quatre directions cardinales de façon à ce que l'un des points de vue corresponde à la direction principale de la relation spatiale que l'on cherche à décrire. On note d'abord que cette notion de direction principale, ou moyenne, entre deux objets, n'a de sens que lorsque les objets sont suffisamment distants (et donc assimilables à des points), ce qui limite la généralité de cette opération. Surtout, le fait d'orienter un des points de vue dans une direction « favorable » pour la relation spatiale \mathcal{R} n'apporte aucune garantie de meilleure description, car nous avons montré que tout point (et toute direction) de l'espace est déjà suffisamment « couvert » par les éléments structurants cardinaux. En revanche, l'augmentation du nombre de points de vue et leur

3.2. Formalisation et mise en œuvre

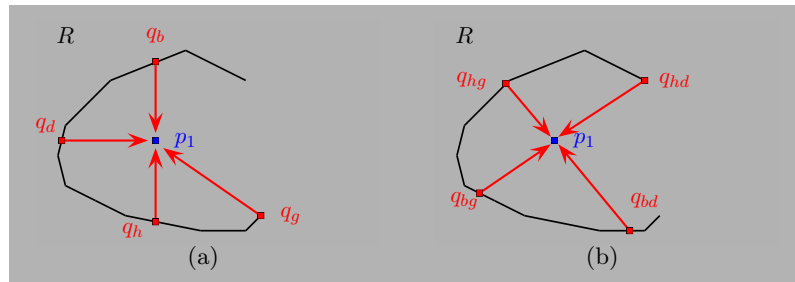


FIGURE 3.11: Différents jeux de quatre points de vue orientés différemment peuvent conduire à une description différente de la même relation. Dans le cas (a), le point p est parfaitement à droite, en haut et en bas de R , mais pas complètement à gauche. Dans le cas (b), le point est vu parfaitement depuis les quatre points de vue.

spécialisation peut bénéficier à la description en fonction de la forme de l'objet de référence. La figure 3.11 illustre cette influence de la forme de l'objet de référence sur la qualité de la description offerte par un jeu de points de vue.

Dans le premier cas de figure (image (a)), les quatre points de vue choisis permettent de décrire que le point p est à l'intérieur d'une concavité de la référence R , mais que cette concavité n'est pas fermée puisque p n'est pas parfaitement à gauche de R . Dans le second cas de figure (image (b)), le point est vu comme parfaitement positionné par rapport aux quatre points de vue. Il a dans ce cas la même description que s'il était positionné dans une boucle fermée de R .

Le choix des quatre points de vue cardinaux permet d'assurer que tout point du plan est « vu » de la référence selon au moins un des points de vue. Il ne permet en revanche pas de garantir la meilleure description quelles que soient la forme de la référence et la nature de la relation spatiale.

Une deuxième technique de choix de points de vue est guidée par un but plus spécifique. Lorsque le but de l'application est de faire une *reconnaissance de relation spatiale*, c'est-à-dire lorsqu'il faut déterminer si une paire d'objets est positionnée selon la relation \mathcal{R}_1 ou \mathcal{R}_2 , la sélection des points de vue peut être faite en cherchant à faciliter l'identification et la discrimination des relations spatiales. Nous présenterons dans la partie expérimentale un exemple de sélection de points de vue adaptée pour une tâche de reconnaissance de relations spatiales (voir le paragraphe 5.1.4.1). Une telle stratégie de sélection de directions peut être mise en place en deux temps. D'abord, à partir d'une base d'exemples étiquetés $(A_i, R_i, \mathcal{R}_i)$ (paires d'objets positionnés avec $\mathcal{R}_i = \mathcal{R}_1$ ou \mathcal{R}_2), il est possible d'extraire les degrés moyens atteints par les objets A_i par rapport aux R_i selon un grand nombre de points de vue $(\alpha_i)_{i=0..N}$ (avec par exemple $N = 16, 32$ ou plus). Ces degrés moyens par direction peuvent être considérés comme des *caractéristiques* de positionnement relatif et un algorithme de sélection de caractéristiques peut être utilisé pour ne conserver que les n directions les plus discriminantes entre les classes \mathcal{R}_1 et \mathcal{R}_2 , avec par exemple $n = 4$. Dans un second temps, les n directions sélectionnées peuvent constituer les points de vue pour l'apprentissage de méta-modèles de positionnement des deux relations spatiales.

Quel que soit le jeu de points de vue finalement choisi, il est préférable qu'il soit commun à tous les méta-modèles de positionnement construits dans un même contexte applicatif, pour des évidentes raisons de performance. En effet, la majeure partie de l'effort de calcul d'un modèle spatial pour un point de vue réside dans le calcul de la dilatation morphologique, qui peut alors être mis en commun entre tous les méta-modèles spatiaux appliqués à un même objet.

3.2.3 Conjonction des points de vue

Conformément à l'exposé de la partie 3.2.1.2, une fonction de composition H est apprise pour chaque point de vue décrivant une relation spatiale \mathcal{R} . En ayant sélectionné un ensemble de points de vue qui assure une description précise des relations, le méta-modèle spatial est donc défini par ses éléments structurants directionnels et les fonctions H associées (voir (3.2)). Nous nous concentrons ici sur la méthode de fusion de l'information portée par chaque point de vue $(\nu_\alpha, H_\alpha^{\mathcal{R}})$ pour offrir une description globale de la relation \mathcal{R} . Chaque point de vue de direction α permet de construire dans le plan image un modèle spatial sous la forme d'un ensemble flou qui représente *être dans la direction α par rapport à R conformément à \mathcal{R}* . Une zone est considérée comme acceptable par ce modèle si elle admet un positionnement conforme aux points d'apprentissage de \mathcal{R} quand on les analyse selon la direction α . Nous proposons que la fusion de l'information portée par les éléments structurants soit opérée directement entre les modèles spatiaux résultants dans l'espace image. Ce choix est directement applicable car les modèles spatiaux sont homogènes dans leur représentation (sous-ensembles flous du plan). La fusion est mise en œuvre par un opérateur de conjonction floue (t-norme). Le modèle spatial global de la relation \mathcal{R} est ainsi défini pour un objet de référence R :

$$\mu_{\mathcal{M}}^{\mathcal{R}}(R)(p) = t[\mu_{\alpha_i}^{\mathcal{R}}(R)(p)]_{i=1..d}, \forall p \in \mathcal{S}, \quad (3.9)$$

avec d le nombre de points de vue, t l'opérateur de t-norme floue et $\mu_{\alpha_i}^{\mathcal{R}}(R)$ le modèle spatial issu du point de vue α_i , construit par composition de la dilatation morphologique de R par l'élément structurant ν_{α_i} et de la fonction $H_{\alpha_i}^{\mathcal{R}}$. L'interprétation de cette opération est très intuitive : il s'agit de considérer un point du plan comme positionné conformément à \mathcal{R} par rapport à R s'il est bien positionné selon tous les points de vue. Par exemple, si on adopte les quatre points de vue cardinaux, un point est dans une zone acceptable d'après $\mathcal{M}^{\mathcal{R}}$ s'il est à la fois *autant à droite* et *autant à gauche* et *autant en haut* et *autant en bas* de R que les points des exemples d'apprentissage de la relation spatiale \mathcal{R} .

L'opérateur mathématique retenu comme t-norme peut être choisi parmi les t-normes floues classiques (voir l'annexe A). Le choix d'une fusion conjonctive est cohérent avec l'interprétation proposée pour l'apprentissage des éléments structurants et la forme des fonctions histogrammes définies pour $H_\alpha^{\mathcal{R}}$. Par exemple, il a été montré dans la figure 3.8 que lorsqu'aucune information ne peut être décrite selon un point de vue, sa fonction de composition est définie comme complètement permissive ($H_{\alpha_i}^{\mathcal{R}} = 1$ et donc $\mu_{\alpha_i}^{\mathcal{R}} = \mathcal{S}$) et ce point de vue prend ainsi un rôle neutre vis-à-vis de l'intersection globale.

Enfin, si nous avons décidé que les modèles spatiaux $\mu_{\alpha_i}^{\mathcal{R}}$ aient au moins une valeur modale (ce sont des ensembles flous *normaux*), aucune garantie de cette sorte ne peut être donnée sur le résultat de la fusion globale $\mu_{\mathcal{M}}^{\mathcal{R}}(R)$. En effet, en fonction de la forme de l'objet

3.3. Prise en compte de la distance

de référence R , une relation \mathcal{R} peut tout simplement ne pas être satisfiable. Par le second exemple de la figure 3.9, nous avons montré qu’une description valide de positionnement pour un objet intersectant une référence verticale peut être à la fois *en-dessous* et *au-dessus* de R . La même description appliquée à une référence parfaitement horizontale donne une intersection vide, aucun point du plan ne pouvant satisfaire ces deux contraintes à la fois.

Exemples La figure 3.12 illustre les modèles spatiaux obtenus pour chaque point de vue cardinal (a, b, c, d), suite à l’apprentissage à partir d’exemples comme nous l’avons illustré sur un point de vue à la figure 3.4. On retrouve une partie redondante entre les deux modèles spatiaux des points de vue *haut* (a) et *bas* (b). Le premier décrit que A n’est pas au-dessus de R , le second que A est en-dessous de R . Cependant, ils se complètent bien et leur intersection est bénéfique pour la précision de la description. Le même constat peut être fait entre les modèles spatiaux des points de vue *à gauche* et *à droite*. Finalement, le modèle spatial global, obtenu par intersection floue, est présenté par la figure 3.12(e). Il décrit les zones du plan qui satisfont un positionnement conforme au modèle de \mathcal{R} par rapport à R .

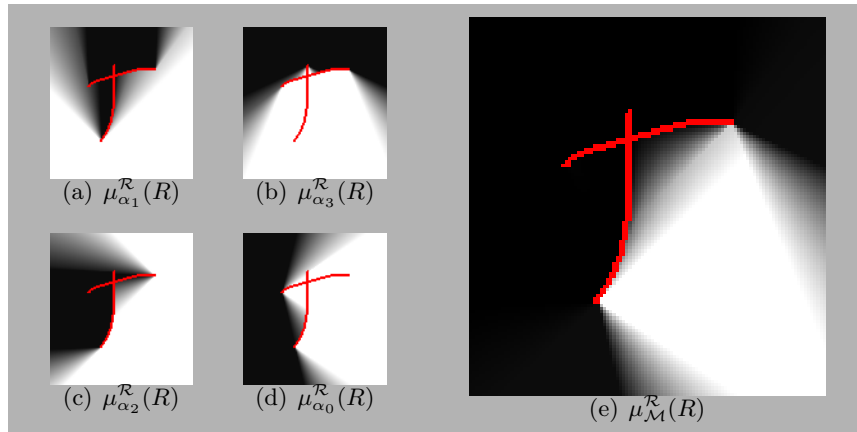


FIGURE 3.12: Modèles spatiaux pour les quatre points de vue cardinaux (a,b,c,d) et intersection globale (e) par rapport à un objet R (en rouge).

La figure 3.13 présente différents modèles spatiaux obtenus par application du même méta-modèle spatial sur différents objets de référence. Les scores d’adéquation moyens pour des objets cibles associés à ces références sont également représentés. Ils évaluent à quel degré la relation entre les objets de chaque paire est conforme au modèle spatial global de la relation \mathcal{R} apprise.

3.3 Prise en compte de la distance

Jusqu’ici, la modélisation que nous avons formulée ne repose que sur des informations de direction entre les objets et c’est par combinaison de points de vue directionnels que des modèles spatiaux satisfaisants peuvent être établis pour représenter des relations spatiales

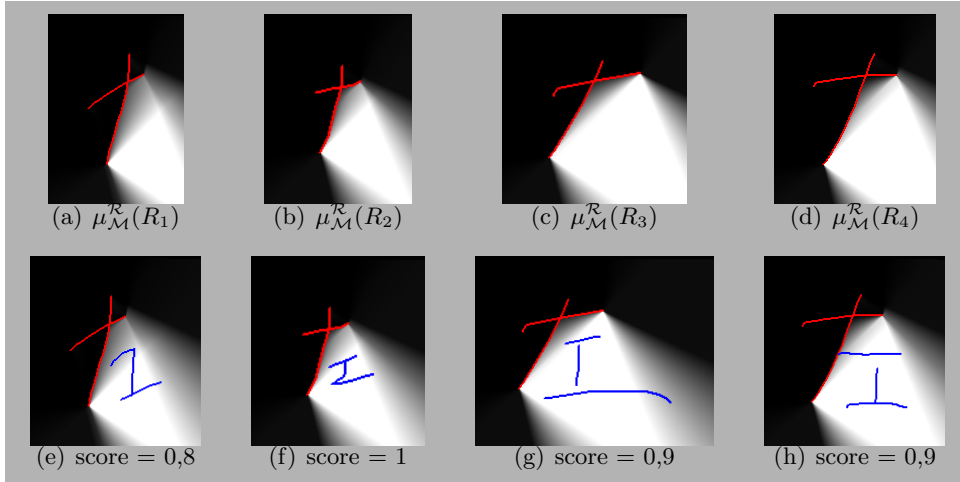


FIGURE 3.13: Application d'un même méta-modèle spatial à différents objets de référence. Les modèles spatiaux adaptés aux formes des références sont présentés par les figures (a) à (d). Les représentations des objets arguments associés sont données par les figures (e) à (h), ainsi que les scores d'évaluation de la relation spatiale pour chaque cas (calculés par la moyenne d'adéquation des points des objets).

non-linguistiques. Cependant, il semble naturel de chercher à incorporer à cette modélisation une information sur la distance entre les objets, de manière à améliorer la qualité de la description offerte. Dans cette section, nous considérons deux stratégies pour incorporer cette information de distance. La première méthode consiste simplement à analyser la distance entre les objets comme un point de vue supplémentaire sur la relation spatiale, ce qui permet une prise en compte globale de la distance. La seconde méthode embarque l'information sur la distance selon chaque point de vue, au moyen d'éléments structurants hybrides (de direction et distance combinées), ce qui offre une description de la distance direction par direction.

3.3.1 Stratégie globale

La technique la plus directe pour intégrer l'information de distance dans notre modélisation est de considérer la distance comme un point de vue additionnel de nos méta-modèles. Cette vision est rendue possible par l'homogénéité de la représentation des points de vue sous la forme de modèles spatiaux dans l'espace image. Le point de vue additionnel de distance repose sur une description par un élément structurant de distance défini comme suit :

$$\nu_{\tau}(p) = \beta_{\tau}(\|\vec{op}\|) \quad \text{avec } \beta_{\tau}(t) = \max(0, 1 - t/\tau), \forall p \in \mathcal{S}. \quad (3.10)$$

o est le centre de l'élément structurant et β est une fonction décroissante de \mathbb{R}^+ dans $[0; 1]$, telle que $\beta(0) = 1$ et $\beta(t) = 0$ dès que $t > \tau$. Ici, nous choisissons pour β une simple fonction linéaire. Un tel élément structurant décrit en fait la relation spatiale de distance *être proche de*. La figure 3.14 représente un élément structurant flou de distance défini de la sorte.

3.3. Prise en compte de la distance

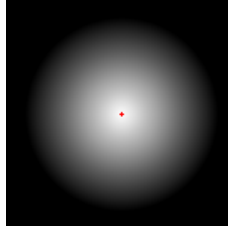


FIGURE 3.14: Élément structurant *être proche de*.

Un modèle spatial de distance μ_{dist} est construit par dilatation d'un objet R de référence par cet élément structurant. C'est un ensemble flou défini dans le plan \mathcal{S} et qui représente la relation *être proche de* R . De façon à adapter le sens de *proche* par rapport aux dimensions de l'objet de référence, le paramètre τ est déterminé, à un facteur multiplicateur près, d'après la longueur de la diagonale de la boîte englobante de R .

Apprentissage Suivant le même principe que celui appliqué pour les points de vue directionnels, une fonction de composition $H_\tau^{\mathcal{R}}$ est associée à un élément structurant ν_τ pour constituer un modèle de point de vue de distance. C'est une fonction déterminée à partir de l'histogramme des fréquences des degrés de *être proche de* atteints par les points des objets d'apprentissage. L'interprétation que l'on peut donner au modèle spatial issu du point de vue $\mu_\tau^{\mathcal{R}}(R)$ est qu'il décrit la relation *être proche de* R dans la même mesure que les points d'apprentissage selon \mathcal{R} .

La figure 3.15(a) représente un modèle spatial de distance, issu d'un point de vue de distance appris, appliqué par dilatation et composition sur un objet R . Cet ensemble flou $\mu_{dist}^{\mathcal{R}}$ représente les zones de l'image qui sont acceptables d'après la relation \mathcal{R} en termes de distance à l'objet R . La distance étant décrite globalement, indépendamment des directions, la zone d'acceptabilité décrit un « halo » dans toutes les directions autour de l'objet de référence.

Exploitation L'exploitation du méta-modèle global implique un point de vue additionnel de distance. On a désormais :

$$\mathcal{M}^{\mathcal{R}} = \langle (\nu_{\alpha_0}, H_{\alpha_0}^{\mathcal{R}}), \dots, (\nu_{\alpha_{d-1}}, H_{\alpha_{d-1}}^{\mathcal{R}}), (\nu_{dist}, H_{dist}^{\mathcal{R}}) \rangle, \quad (3.11)$$

avec d le nombre de points de vue directionnels. Le résultat de la fusion par intersection floue des points de vue est déterminé exactement comme précédemment (voir l'équation (3.9)), le modèle spatial de distance jouant le même rôle que les modèles spatiaux directionnels. La figure 3.15(b) représente le modèle spatial global, obtenu par fusion de quatre points de vue directionnels et d'un point de vue additionnel de distance. Ce modèle spatial peut être comparé à celui présenté par la figure 3.12(e), ce qui permet de mettre en valeur l'intérêt de l'introduction d'un modèle de distance dans la description. Le résultat est en effet plus conforme à la relation spatiale modélisée, car l'étendue de la zone d'acceptabilité dans la direction *en bas à droite* est beaucoup mieux contenue. Les points trop proches de l'objet de référence sont également exclus de la zone d'acceptabilité.

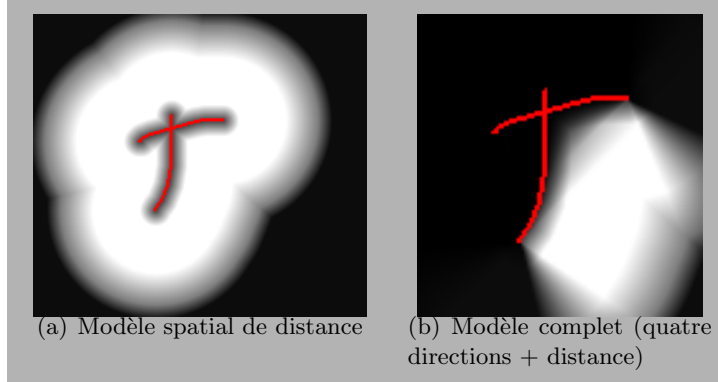


FIGURE 3.15: Un modèle spatial de distance peut être vu comme un point de vue additionnel sur la relation (a). Le résultat de l'application du modèle complet, obtenu par intersection des cinq points de vue, est présenté (b).

3.3.2 Stratégie directionnelle

Une autre méthode pour la prise en compte de la distance dans la modélisation est d'intégrer sa description direction par direction. On peut supposer qu'une description plus précise de l'information de distance peut être construite en utilisant un modèle différent pour chaque point de vue de direction. L'idée est donc ici de construire un modèle spatial hybride couplant modèle de direction et de distance, qui décrira les deux informations pour chaque point de vue. Par rapport à la définition que nous avons donnée au paragraphe 3.2.2, nous proposons ici de construire un ensemble flou décrivant les distances acceptables pour la relation \mathcal{R} et pour chaque section de l'histogramme H . La figure 3.16 illustre par un schéma l'intégration de modèles de distance individuels pour chaque direction modélisée par l'une des sections de l'histogramme H associé à un point de vue.

Apprentissage Plus précisément, si $(H_\alpha^{\mathcal{R}})_k$ représente la k -ième section de l'histogramme H (fonction de composition pour le point de vue de direction α), le modèle de distance associé, noté $(\delta_\alpha^{\mathcal{R}})_k$ doit être construit en fonction de la distribution des degrés de distances atteints par les points de S_{i_k} . Cette distribution est notée $(\Delta_\alpha)_k$:

$$(\Delta_\alpha)_k = \cup_i \{ \mu_{dist}(R_i)(p), p \in S_{i_k} \} \quad S_{i_k} = \{ p \in S, \mu_\alpha^{R_i}(p) \in I_k \}, \quad (3.12)$$

avec i l'indice d'une paire (R_i, A_i) d'objets d'apprentissage, $\mu_{dist}(R_i)$ un modèle spatial de distance construit comme dans la section précédente, décrivant la relation spatiale *proche* de R_i . I_k est défini comme dans la partie 3.2.1.2. $(\Delta_\alpha)_k$ est approximé par un ensemble flou $(\delta_\alpha)_k$ défini par une fonction d'appartenance de forme trapézoïdale, de paramètres a, b, c et d (voir aussi sa représentation à la figure 3.17) :

3.3. Prise en compte de la distance

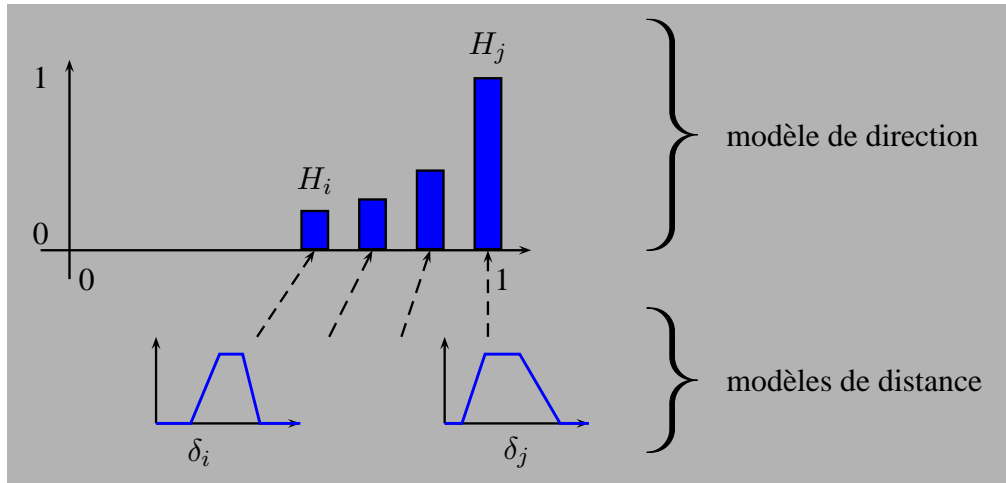


FIGURE 3.16: Schéma d'intégration de modèles de distance direction par direction au sein d'un point de vue. A chaque direction quantifiée dans le modèle H du point de vue est associé un modèle de distance dédié.

$$\delta(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x \leq b \\ 1 & \text{si } x \geq b < x \leq c \\ 1 - \frac{x-c}{d-c} & \text{si } c < x \leq d \\ 0 & \text{si } x > d. \end{cases} \quad (3.13)$$

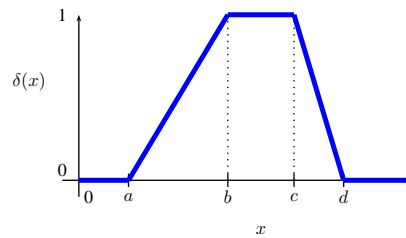


FIGURE 3.17: Définition de l'ensemble flou trapézoïdal $(\delta_\alpha)_k(x)$ modélisant la distance apprise pour les points dont le degré directionnel est dans la section H_k de l'histogramme.

Nous choisissons pour a et d respectivement les valeurs minimale et maximale de $(\Delta_\alpha)_k$, de sorte que :

$$x > \inf(\Delta_\alpha)_k \text{ et } x < \sup(\Delta_\alpha)_k \Rightarrow (\delta_\alpha)_k(x) > 0,$$

avec b et c qui désignent respectivement les valeurs du premier et troisième quartiles de $(\Delta_\alpha)_k$, de sorte que la moitié des éléments de $(\Delta_\alpha)_k$ aient une appartenance de 1 à l'ensemble flou.

L'ensemble des $(\delta_\alpha)_{k,k=1..K+1}$ constitue une famille d'ensembles flous trapézoïdaux, qui peuvent être intégrés à la définition de la fonction de composition apprise hybride pour le point de vue de direction α :

$$\mu_{\alpha,d}^{\mathcal{R}}(R)(p) = t\left[\underbrace{(H_{\alpha_k}^{\mathcal{R}} \circ \mu_\alpha(R))(p)}_{\text{direction}}, \underbrace{(\delta_{\alpha_k} \circ \mu_{\tau_R}(R))(p)}_{\text{distance}}\right] \quad \forall p \in S \text{ t.q. } \mu_\alpha(R)(p) \in I_k, \quad (3.14)$$

avec t un opérateur de t-norme floue, $H_{\alpha_k}^{\mathcal{R}} \circ \mu_\alpha(R)$ est le modèle spatial directionnel selon le point de vue α (comme défini dans l'équation (3.3)), et $\delta_{\alpha_k} \circ \mu_{\tau_R}(R)$ est le modèle de distance associé. $\mu_{\tau_R}(R)$ est défini comme décrit à l'équation ((3.10)).

La figure 3.18(a) illustre un exemple de modèle spatial hybride obtenu par rapport à un objet R , pour le point de vue à droite de . Cette illustration peut être comparée au modèle spatial de la figure 3.12(d), qui représente le même point de vue de la relation spatiale \mathcal{R} , autour du même objet de référence R , mais sans considération de distance. L'apport de la distance apparaît nettement, puisque le modèle spatial de la figure 3.18(a) ne présente plus une zone d'acceptabilité de taille infinie vers la droite, ce qui est plus conforme à \mathcal{R} .

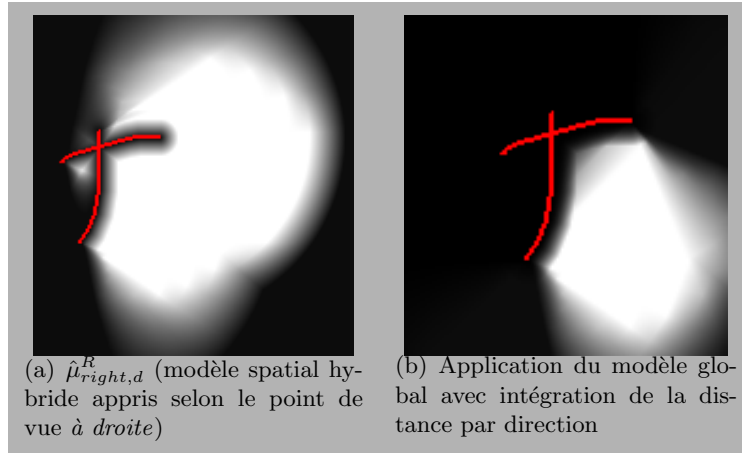


FIGURE 3.18: Modèle spatial hybride (de direction et distance) associé au point de vue *droite* pour un objet R (a) et modèle spatial global résultant de l'application du méta-modèle, à partir de quatre points de vue hybrides (b).

Exploitation L'exploitation d'un méta-modèle spatial qui intègre des modèles de distances au sein de chaque point de vue directionnel est tout à fait identique à la procédure adoptée pour un méta-modèle purement directionnel. Les modèles spatiaux hybrides construits pour chaque point de vue sont fusionnés comme précédemment, par intersection floue (voir l'équation (3.9)). La figure 3.18(b) représente le modèle spatial résultat obtenu à partir de points de vue qui intègrent les modèles de distances. Il peut être comparé à ceux représentés par les figures 3.12(e) (sans modèle de distance) et 3.15(b) (avec un modèle global de distance).

3.4 Prise en compte de la bipolarité de l'information spatiale

Dans les modèles spatiaux construits jusqu'ici, comme dans ceux présentés dans l'état de l'art, l'information sur la qualité de positionnement des points du plan par rapport à l'objet de référence est décrite par un ensemble flou, dont la fonction d'appartenance reflète les zones plus ou moins acceptables. Pourtant, plusieurs études ont montré que l'information spatiale peut être gérée sous une représentation *bipolaire*, c'est-à-dire en distinguant les informations positives (par exemple des zones de positionnement favorables) et les informations négatives (par exemple les zones de positionnement interdites, impossibles) [Blo11][Blo09]. L'idée de représentation bipolaire de connaissance imprécise est utilisée notamment en base de données pour décrire les données ou pour exprimer toute la complexité des requêtes utilisateurs [DTZB10][MDT09][DP02]. Dans ce contexte, il est par exemple utile de distinguer l'information positive (qui indique une *préférence* formulée dans la requête, un *souhait*, une situation favorable) de l'information négative (qui exprime une *contrainte*, une *interdiction*, ou une situation indésirable).

En raisonnement spatial, beaucoup de concepts manipulés sont intrinsèquement bipolaires, comme par exemple les concepts directionnels opposés (*gauche* et *droite*). Les modèles spatiaux, quant à eux, sont des cartes de préférences de positionnement mais on peut considérer qu'ils intègrent des préférences positives (par exemple des zones de positionnement favorables, constatées par l'expérience) et des préférences négatives (zones interdites, impossibles). Dans sa formulation initiale des modèles spatiaux, Logan précisait déjà qu'ils définissaient des zones (plus ou moins) favorables et des zones inacceptables [LS99]. Cette dualité de la représentation n'est pas complètement reproduite dans les modèles spatiaux tels que nous les avons présentés jusqu'ici. Le procédé d'apprentissage proposé pour les méta-modèles ne permet pas de distinguer finement les zones de positionnement qui sont *peu* représentatives de la relation de celles qui ne le sont *pas du tout*. Il paraît pourtant souhaitable de pouvoir pénaliser plus fortement un objet qui est localisé dans une zone explicitement *interdite* du modèle spatial, par rapport à un objet dans une zone *mauvaise* (c'est-à-dire dans laquelle peu de points d'apprentissage ont été constatés). Nous allons dans cette partie développer une extension des méta-modèles spatiaux afin de permettre l'application de modèles spatiaux bipolaires qui font une distinction explicite entre ces notions de *préférences* et de *contraintes* de positionnement.

Dans ses travaux les plus récents, Bloch a développé un cadre théorique complet pour l'intégration de représentation bipolaire au sein de la morphologie mathématique, notamment dans le but de construire des représentations bipolaires de relations spatiales dans des images (sous la forme de *modèles spatiaux bipolaires*) [Blo11][Blo09]. Dans cette partie, nous introduisons d'abord quelques définitions concernant nos choix de représentation de l'information bipolaire, qui sont reprises des travaux de Bloch, puis nous présentons brièvement sa solution pour la construction de descriptions bipolaires floues des relations spatiales. Ensuite, nous précisons notre choix de représentation afin de construire une extension bipolaire des méta-modèles spatiaux décrits précédemment.

3.4.1 Définitions

Un ensemble flou bipolaire est défini dans l'espace \mathcal{S} par une paire de fonctions (v, ω) telles que :

$$\forall x \in \mathcal{S}, v(x) + \omega(x) \leq 1.$$

La fonction v associe à x un degré flou d'appartenance entre 0 et 1; c'est l'information positive sur x . La fonction ω associe à x un degré flou de non-appartenance entre 0 et 1; c'est l'information négative sur x . Cette représentation permet de représenter à la fois la bipolarité de l'information et son caractère flou, afin de gérer l'imprécision. La différence $1 - v(x) - \omega(x)$ représente un degré de *neutralité*, d'*indifférence* ou d'*indétermination* pour le point x . On remarque aussi qu'un ensemble flou classique du plan est un cas particulier d'ensemble flou bipolaire, avec $\mu = (v, 1 - v)$. La contrainte $v(x) + \omega(x) \leq 1$ permet de garantir une cohérence dans les préférences positives et négatives exprimées. D'autres représentations sont possibles, qui relâchent ou au contraire renforcent cette contrainte, ce qui autorise un plus ou moins grande indépendance entre les deux composantes. Plus de détails sur ces définitions alternatives et leurs interprétations sont donnés par De Tré, Zadrozny *et al.* [DTZB10].

Des extensions des opérateurs de logique floue ont été proposés pour les ensembles bipolaires. Nous présentons ici certains d'entre eux. Tout d'abord, en suivant la notation adoptée dans [Blo11], on note \mathcal{L} l'ensemble des paires de nombres (a, b) de $[0; 1]$ tels que $a + b \leq 1$. Un ordre partiel \preceq peut être défini sur \mathcal{L} :

$$\forall (a_1, b_1), (a_2, b_2) \in \mathcal{L}, (a_1, b_1) \preceq (a_2, b_2) \text{ ssi } a_1 \leq a_2 \text{ et } b_1 \geq b_2.$$

Il s'agit d'un ordre partiel composante par composante, qui correspond à un ordre de Pareto. Le treillis complet (\mathcal{L}, \preceq) associé admet pour plus grand élément $(1, 0)$ et pour plus petit élément $(0, 1)$.

Un opérateur de négation bipolaire, défini sur \mathcal{L} , est un opérateur décroissant N tel que $N((0, 1)) = (1, 0)$ et $N((1, 0)) = (0, 1)$. Un exemple d'opérateur de négation peut être défini par $N((a, b)) = (b, a)$.

Un opérateur de t-norme bipolaire, défini de $\mathcal{L} \times \mathcal{L}$ dans \mathcal{L} , est un opérateur \top , croissant sur ses deux arguments, qui a pour élément neutre $(1, 0)$ et pour élément absorbant $(0, 1)$. Une t-norme bipolaire \top peut être définie à partir d'une t-norme et d'une t-conorme floues classiques, notées t et T (par exemple avec $t = \min$ et $T = \max$) :

$$\top((a_1, b_1), (a_2, b_2)) = (t(a_1, a_2), T(b_1, b_2)).$$

On parle alors pour \top de t-norme bipolaire *t-représentable*. Une t-conorme bipolaire \perp peut être définie de façon duale :

$$\perp((a_1, b_1), (a_2, b_2)) = (T(a_1, a_2), t(b_1, b_2)).$$

3.4.2 Morphologie bipolaire floue

Considérant l'espace \mathcal{B} des ensembles flous bipolaires définis dans \mathcal{S} , on peut le munir d'un ordre partiel de Pareto défini par :

$$\forall \mu_1 = (v_1, \omega_1), \mu_2 = (v_2, \omega_2) \in \mathcal{B}, \mu_1 \preceq \mu_2 \text{ ssi } \forall x \in \mathcal{S}, v_1(x) \leq v_2(x) \text{ et } \omega_1(x) \geq \omega_2(x).$$

3.4. Prise en compte de la bipolarité de l'information spatiale

La structure algébrique de treillis complet de (\mathcal{B}, \preceq) permet de développer des opérateurs morphologiques sur des ensembles flous bipolaires. En particulier, l'opération de dilatation d'un objet de l'espace par un élément structurant peut être définie lorsque tous deux sont des ensembles flous bipolaires de \mathcal{S} . Les formalisations de ces opérations et leurs propriétés sont présentées en détails dans la référence [Blo11].

Un objet est décrit par un ensemble flou bipolaire dans l'espace et des éléments structurants bipolaires flous sont définis. Un exemple d'élément structurant bipolaire associe une fonction floue de direction *être à droite de* (pour la partie positive) avec une fonction floue de direction opposée *être à gauche de*. La dilatation bipolaire conduit à la définition d'un ensemble flou bipolaire, qui décrit la relation *être à droite de* au moyen de deux composantes. Sous conditions de choix de certains opérateurs bipolaires, il est démontré que chacune des composantes de l'ensemble flou bipolaire résultant peut être interprété comme le résultat d'une opération morphologique simple des composantes positives (respectivement négatives) de l'objet initial et de l'élément structurant (ou de leur complément ensembliste) [Blo11].

Différentes relations d'ordre sont proposées alternativement à \preceq , ce qui peut permettre de donner un poids plus ou moins important à l'une des deux composantes par rapport à l'autre, et ainsi d'affecter les résultats des opérations morphologiques. Par exemple, une relation d'ordre *lexicographique* qui consiste à comparer deux éléments de \mathcal{L} d'abord selon leurs composantes négatives, puis selon leurs composantes positives si égalité, peut être définie. Ce faisant, on privilégie la prise en compte des contraintes (informations négatives) par rapport aux préférences (informations positives).

Discussion sur la définition des objets La mise en œuvre bipolaire de la description spatiale proposée dans les travaux de Bloch a l'avantage de reposer sur un cadre théorique solidement établi, dans lequel toute l'information est représentée par des ensembles bipolaires : les objets, les éléments structurants et les modèles spatiaux obtenus. Cela est rendu possible grâce à l'extension de la morphologie mathématique au cadre bipolaire. En pratique, cette formalisation peut être adaptée au cas d'images, dont les objets sont définis avec une part d'imprécision qui peut être décrite par des ensembles bipolaires. Par exemple, dans une image obtenue par IRM d'une coupe de cerveau, il existe une part d'imprécision dans les objets identifiés. Un *segment* défini comme une zone homogène de l'image peut correspondre à une structure identifiée (thalamus, cortex cérébral, ventricule...) mais comporter une certaine imprécision sur ses frontières. Ce segment, couramment décrit en pratique par une fonction floue d'appartenance des points de l'image, peut aussi être modélisé par un ensemble flou bipolaire décrivant les zones qui lui appartiennent plus ou moins favorablement et des zones qui en sont plus ou moins exclues.

Dans le cas qui nous intéresse, les objets sont des tracés manuscrits. En particulier, lorsqu'ils sont représentés par un signal en-ligne, ils ont des propriétés particulières qu'il est intéressant de conserver pour développer des implémentations performantes des opérations morphologiques. Ils ont en effet une nature *linéique* (sans « épaisseur ») qui se prête à une représentation par approximation polygonale. Dès lors, il est peu naturel et peu efficace d'adopter une représentation bipolaire des objets, ce qui nécessite de les considérer comme des ensembles (surfaiques) de l'image et de s'intéresser à leur complément ensembliste. De la même façon que nous nous sommes restreints à considérer les objets manuscrits comme

des objets simples (et non pas flous), nous souhaitons conserver cette représentation pour la construction de modèles spatiaux bipolaires, car elle autorise une moindre complexité de mise en œuvre (nous donnons des détails sur ces aspects algorithmiques dans l'annexe B). Dans notre réflexion, nous allons donc nous intéresser uniquement à construire des modèles spatiaux bipolaires à partir d'objets dont la simplicité de représentation est conservée.

3.4.3 Méta-modèles spatiaux bipolaires

L'approche que nous proposons pour la représentation de l'information binaire floue est assez différente de celle introduite par Bloch. En effet, nous nous intéressons spécifiquement à construire un résultat (le modèle spatial) binaire à partir d'opérations morphologiques floues classiques développées sur des objets simples (non flous et non binaires). Nous choisissons de décrire chaque point de vue sur une relation spatiale par deux fonctions de composition, l'une décrivant l'information positive (de *préférence* de positionnement) et l'autre l'information négative (de *contraintes* de positionnement) sur la relation modélisée. Ainsi un méta-modèle de la relation spatiale \mathcal{R} est décrit par :

$$\mathcal{M}^{\mathcal{R}} = \langle (\nu_{\alpha_i}, H_{\alpha_i}^{\mathcal{R}}, G_{\alpha_i}^{\mathcal{R}}) \rangle_{i=0..d-1}, \quad (3.15)$$

où ν_{α_i} est l'élément structurant directionnel, $H_{\alpha_i}^{\mathcal{R}}$ et $G_{\alpha_i}^{\mathcal{R}}$ sont deux fonctions de composition décrivant respectivement l'information *positive* et *négative* pour le point de vue orienté par α_i .

3.4.3.1 Apprentissage des fonctions des points de vue binaires

La fonction H de composition positive est apprise comme nous l'avons décrit plus haut (voir au paragraphe 3.2.1) et a la même interprétation. Par combinaison avec la dilatation morphologique au moyen de l'élément structurant directionnel, elle permet de représenter les zones du plan qui sont considérées comme acceptables (ou souhaitables) pour le positionnement d'un objet par rapport à un autre selon le modèle. Cette acceptabilité est apprise à partir des fréquences constatées de présence des points des objets au cours de l'apprentissage. Comme précédemment, nous définissons le modèle spatial de *préférence positive* par composition de la dilatation de R (l'objet de référence) par l'élément structurant linguistique de direction α avec la fonction représentée par l'histogramme H :

$$v_{\alpha}^{\mathcal{R}}(R) = (H_{\alpha}^{\mathcal{R}} \circ \mu_{\alpha}(R)). \quad (3.16)$$

Un modèle dual représente l'information de *préférence négative* de positionnement pour la relation spatiale. Son rôle est donc d'exhiber les zones de l'espace où un objet *ne doit pas* se situer d'après la relation apprise. La construction de ce modèle spatial négatif va se faire également à partir des fréquences constatées en apprentissage, l'objectif étant d'identifier des zones de l'espace où il est *impossible* que les objets se situent d'après la relation apprise. A partir de la base d'exemples d'apprentissage de la relation spatiale à modéliser, on considérera que les régions jamais occupées par ces exemples seront de fait à exclure explicitement du modèle spatial (et donc à considérer comme des zones *interdites*). Le modèle spatial de *préférence négative* est construit par application de la fonction de composition négative G sur la dilatation de R par le même élément structurant directionnel :

3.4. Prise en compte de la bipolarité de l'information spatiale

$$\omega_\alpha^{\mathcal{R}}(R) = (G_\alpha^{\mathcal{R}} \circ \mu_\alpha(R)). \quad (3.17)$$

Apprentissage La fonction G est construite à partir de l'histogramme de fréquences normalisé \hat{h} , construit sur la base d'exemples d'apprentissage (comme expliqué à la section 3.2.1.2). Si l'histogramme H est construit en appliquant une fonction de seuillage g à \hat{h} , G est construit en lui appliquant une fonction f décroissante, telle que $f(0) = 1$ et $f(1) = 0$. Nous choisissons de définir simplement f en fonction du paramètre a qui définissait le seuil minimal d'activation pour g (voir équation (3.7)) :

$$f(x) = \begin{cases} 1 - \frac{x}{a} & \text{si } x \leq a \\ 0 & \text{sinon.} \end{cases} \quad (3.18)$$

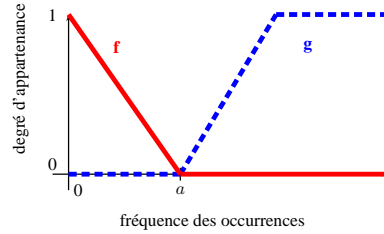


FIGURE 3.19: Représentation de la fonction f (en rouge) de seuillage pour l'apprentissage du modèle *néгатif*. La fonction g pour le seuillage de la fonction *positive* est figurée en pointillés bleus (voir sa définition à la figure 3.7). Les gammes de degrés aux fréquences rares (inférieures au seuil a) seront considérées comme des zones de préférence négative dans le modèle bipolaire.

Les deux fonctions f et g sont liées par le paramètre a , ce qui permet de garantir que $\forall x \in [0; 1], g(x) + f(x) \leq 1$. La figure 3.19 représente la fonction f par rapport à la fonction g . Comme précisé plus haut, a détermine le seuil de fréquence minimal pour qu'une valeur soit considérée comme représentative de la relation (au sens des préférences positives exprimées par H). En deçà de a , la fréquence est considérée comme une *anomalie*. Elle n'est donc pas prise en compte dans le modèle positif ($g(x) = 0$) mais est prise en compte dans le modèle de préférence négative ($f(x) > 0$). Si une valeur x n'est strictement jamais constatée sur les points d'apprentissage, elle est alors jugée complètement *interdite* par le modèle négatif ($f(x) = 1$).

Exemple L'intérêt de la représentation bipolaire peut être mis en évidence au moyen de l'exemple repris de la figure 3.8. Nous avons construit pour les deux relations \mathcal{R}_1 et \mathcal{R}_6 le même modèle de positionnement, alors que les deux paires d'objets ont des positionnements relatifs différents (voir exemples (a) et (f) de la figure 3.8). Dans la figure 3.20, les modèles spatiaux positifs et négatifs sont représentés, pour les mêmes objets et les deux relations \mathcal{R}_1 et \mathcal{R}_6 .

Les images (a) et (g) représentent les paires d'objets d'apprentissage et les histogrammes de fréquences associées pour le point de vue à *droite de* sont représentés par (b) et (h). Les

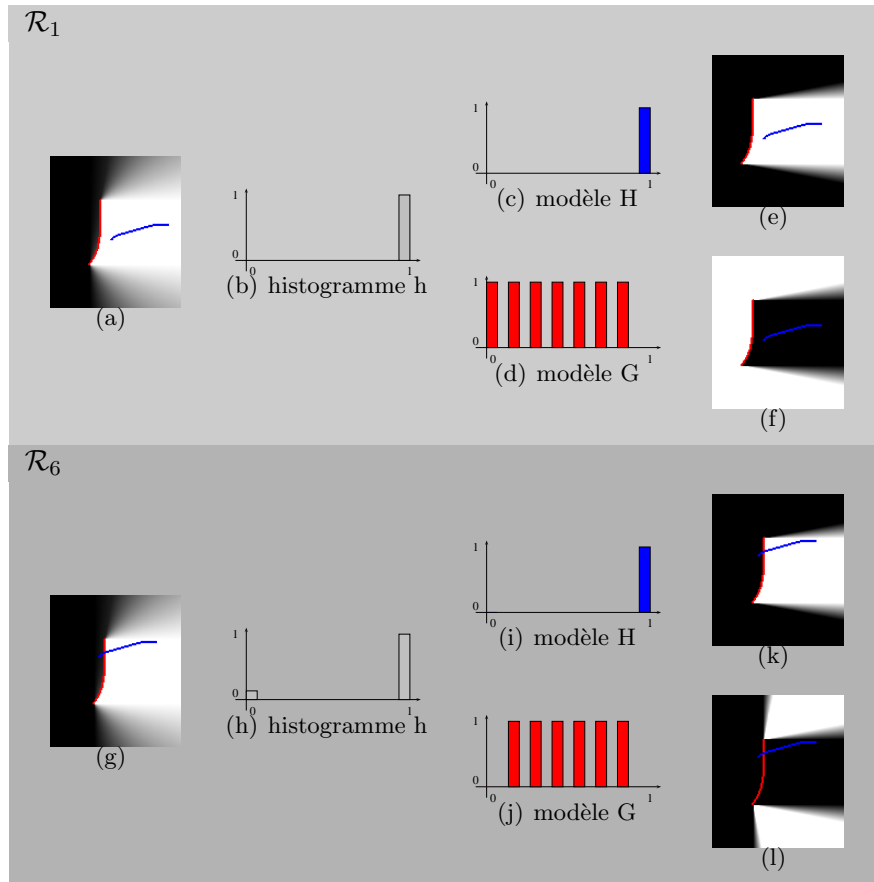


FIGURE 3.20: Exemples d'apprentissage de méta-modèles spatiaux bipolaires pour deux relations spatiales indifférenciables selon leur composante positive (relations spatiales (a) et (f) de la figure 3.8).

histogrammes de préférences positives H (en bleu) et négatives G (en rouge) sont aussi représentés pour les deux relations (c) et (d) pour \mathcal{R}_1 , (i) et (j) pour \mathcal{R}_6 . Enfin, la dernière colonne montre les modèles spatiaux positifs et négatifs correspondant (e) et (f) pour \mathcal{R}_1 , et (k) et (l) pour \mathcal{R}_6 . Les modèles positifs ont la même interprétation qu'auparavant : les zones blanches des images (e) et (k) décrivent les zones de préférence positive, c'est-à-dire les zones acceptables. Les modèles spatiaux négatifs sont aussi représentés en niveaux de gris, les zones *interdites* ou *inacceptables* correspondant aux zones les plus claires dans les images (f) et (l). Les modèles spatiaux positifs restent inchangés par rapport à la figure 3.8 et sont toujours identiques pour les deux relations. Par contre, on peut noter que les modèles des deux relations se distinguent désormais par leur composante négative. En effet, la composante négative de \mathcal{R}_1 exclut du modèle spatial toute zone du plan qui n'est pas strictement à droite de R (zone blanche de l'image (f)). Dans ce cas, le modèle négatif est égal au complémentaire du modèle positif. Dans le cas de \mathcal{R}_6 (image (l)), la zone à gauche de l'objet n'est pas considérée comme *interdite* par le modèle négatif. En apprentissage, certains points étaient situés à gauche de l'objet de référence. Leur nombre était trop faible pour être considéré dans le modèle positif, mais suffisant pour que la zone ne soit pas

3.4. Prise en compte de la bipolarité de l'information spatiale

complètement interdite par le modèle négatif. Ici, leur fréquence dans \hat{h} est égale au seuil a , ce qui revient finalement à considérer que la zone à gauche de R est globalement neutre selon ce point de vue (elle n'est ni une zone de préférence positive, ni une zone de préférence négative).

Distance et bipolarité Dans l'exposé et dans l'exemple ci-dessus, nous n'avons pas évoqué le rôle de la distance (dont nous avons pourtant proposé deux modèles d'intégration) vis-à-vis des modèles spatiaux bipolaires. En fait, de même que nous avons construit nos méta-modèles sur les informations de direction en premier lieu, puis de distance en complément, nous utilisons la distance comme une information d'appoint dans le modèle bipolaire. L'information de distance est plus sensible et plus variable que l'information de direction, notamment à cause du problème de mise à l'échelle. Les descriptions de distance que nous avons proposées tiennent compte de l'échelle des objets en adaptant la notion de proximité par rapport à une mesure grossière de dimension de la référence (liée à la longueur de la diagonale de sa boîte englobante). En pratique, la direction s'avère être une information beaucoup plus stable que la distance normalisée. Si l'intégration de distance est faite suivant la stratégie globale, il est aisé de reproduire sur le point de vue de distance le même raisonnement bipolaire que pour les points de vue directionnels. Etant donnée la description trop générale que cela offre, il est cependant probable que dans la plupart des cas, cette modélisation ne permette pas d'éliminer des zones du plan par « interdiction » dans le modèle négatif (les composantes négatives risquent alors d'être totalement permissives et donc peu informatives). Dans le mode d'intégration directionnel, plus fin, on risque davantage d'être confronté à l'instabilité de la distance. Dans ce cas, nous utilisons l'information de distance uniquement dans la partie positive des modèles, sans la faire intervenir dans la partie négative. Le modèle spatial positif est construit selon le procédé vu plus haut (partie 3.3.2). L'information sur la distance n'est pas utilisée pour renforcer la précision de la composante négative des modèles, mais uniquement pour affiner la description des préférences positives de positionnement.

3.4.3.2 Construction du modèle spatial bipolaire global

Selon un point de vue α_i déterminé et étant donné un objet de référence R dans l'espace \mathcal{S} , un modèle spatial bipolaire est un ensemble flou bipolaire, défini de \mathcal{S} dans \mathcal{L} , qui est issu des deux modèles spatiaux flous $v_{\alpha_i}^{\mathcal{R}}$ et $\omega_{\alpha_i}^{\mathcal{R}}$ (composante positive et négative respectivement) :

$$\mu_{\alpha_i}^{\mathcal{R}}(R) = (v_{\alpha_i}^{\mathcal{R}}(R), \omega_{\alpha_i}^{\mathcal{R}}(R)). \quad (3.19)$$

Grâce à la définition proposée pour les fonctions f et g , on a :

$$\forall p \in \mathcal{S} \quad v_{\alpha_i}^{\mathcal{R}}(R)(p) + \omega_{\alpha_i}^{\mathcal{R}}(R)(p) \leq 1,$$

ce qui vérifie que $\mu_{\alpha_i}^{\mathcal{R}}(R)$ est bien un ensemble flou bipolaire de \mathcal{S} . La fusion des modèles spatiaux bipolaires obtenus pour chaque point de vue se fait par une opération de conjonction, mise en œuvre au moyen d'une t-norme bipolaire \top :

$$\mu_{\mathcal{M}}^{\mathcal{R}}(R)(p) = \top[\mu_{\alpha_i}^{\mathcal{R}}(R)(p)]_{i=1..d}, \forall p \in \mathcal{S}. \quad (3.20)$$

Représentation des modèles spatiaux bipolaires Deux images en niveaux de gris sont nécessaires pour représenter un tel modèle spatial bipolaire, puisque les deux composantes ne peuvent être représentées sur la même échelle de gris. Cependant, pour alléger la représentation, nous proposons d'utiliser dans la suite une échelle de couleurs dans les images illustrant des ensembles flous bipolaires. La figure 3.21 montre la gamme de couleurs que nous utiliserons pour représenter les valeurs de \mathcal{L} (ensemble des degrés bipolaires $\mathcal{L} = \{ \langle a; b \rangle \}$ de $[0; 1]^2$ tels que $a + b \leq 1$).

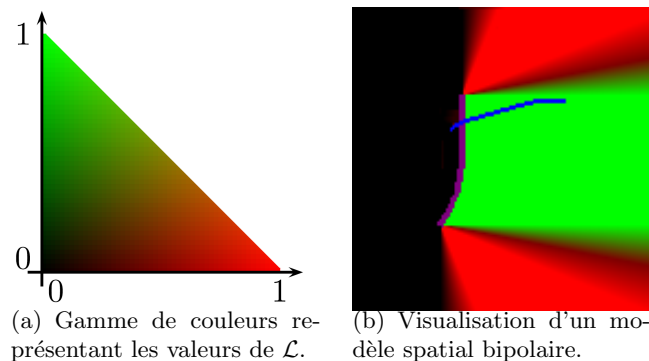


FIGURE 3.21: Gamme de couleurs utilisées pour représenter les valeurs de \mathcal{L} (a) et exemple de représentation d'un modèle spatial bipolaire en utilisant cette gamme de couleurs (b) (ce modèle est le résultat de la fusion d'un modèle spatial bipolaire dont les deux composantes sont représentées à la figure 3.20 (k) et (l)).

La composante verte est utilisée pour représenter le degré d'appartenance à l'ensemble bipolaire et la gamme de rouge pour le degré de non-appartenance. La figure 3.21(b), qui présente un exemple de visualisation de modèle spatial bipolaire, illustre cette représentation. Il s'agit du modèle spatial bipolaire représenté dans la figure 3.20 au moyen de ses deux composantes positive (image (k)) et négative (image (l)), qui sont ici fusionnées dans une représentation unique grâce à cette gamme de couleurs. L'objet de référence, qui était par convention représenté en rouge jusqu'ici, est maintenant représenté en violet.

3.5 Intégration d'un modèle d'étendue

Les méta-modèles spatiaux que nous avons définis jusqu'ici permettent de modéliser le positionnement relatif d'objets en tenant compte de la forme de la référence et en s'appuyant sur des directions et des distances considérées comme *admisses* suite à un apprentissage à partir de données. L'extension de la méthode à la formalisation de modèles spatiaux bipolaires permet en outre de distinguer des zones favorables ou défavorables de positionnement par l'expression de préférences positives ou négatives.

En revanche, la description ne permet pas de modéliser *comment* l'objet cible occupe l'espace. Une relation de positionnement spatial n'est pas uniquement caractérisée par la définition des zones admissibles pour l'objet cible (étant donné un objet de référence), mais également par la manière dont il couvre l'espace. Indépendamment des zones où il *peut* être

3.5. Intégration d'un modèle d'étendue

positionné, un objet a également une *portée* ou une *étendue* par rapport à la référence qui détermine en partie la relation spatiale.

Dans cette section, nous proposons d'ajouter au méta-modèle une modélisation de *l'étendue* de l'objet à positionner par rapport à la référence, qui apporte une information complémentaire à la modélisation de son positionnement. Tandis que le modèle de positionnement décrit les zones de l'image où l'objet cible *peut* être localisé, le modèle d'étendue décrit *quelle doit être l'étendue de l'objet cible*. L'étendue est décrite comme une mesure de la *portée* ou de *l'amplitude des degrés* atteints par les points de l'objet dans les modèles spatiaux linguistiques (*à droite de, à gauche de...*) construits par rapport à la référence. La description de l'étendue se fait donc dans les espaces de représentation déjà exploités pour construire les points de vue du méta-modèle spatial.

Un objet à positionner par rapport à une référence aura donc un bon degré d'étendue par rapport au modèle si son amplitude dans l'espace des degrés directionnels est conforme au modèle, selon chaque point de vue. Ce score pourra être combiné au score de positionnement de l'objet cible dans un score global d'*évaluation* décrivant conjointement l'adéquation de positionnement et d'étendue par rapport à la relation spatiale apprise.

Par définition, l'étendue sort du cadre de la représentation par modèle spatial, puisque l'objet cible doit être défini pour en mesurer l'étendue et lui associer un score. La mesure d'adéquation de l'étendue par rapport au modèle ne pourra donc servir qu'à valider la relation spatiale d'un objet par rapport à un objet de référence. Elle ne pourra être utilisée que dans le cadre d'une tâche de *vérification* de relation spatiale et non pour la *localisation* d'objets.

Dans la suite, nous introduisons d'abord les objectifs de la modélisation de l'étendue en illustrant les besoins par deux exemples. Nous montrons qu'un modèle d'étendue permet de distinguer des cas de relations spatiales indifférenciables par les simples modèles spatiaux (par le positionnement pur). Un apport important est notamment la possibilité de vérification de propriété topologique telle que l'intersection des objets. Par la suite, nous présenterons deux modélisations possibles pour prendre en compte cette notion d'étendue d'objets, qui seront utilisables alternativement en fonction de la nature des objets manipulés (représentation en-ligne ou hors-ligne).

3.5.1 Objectif

Les modèles spatiaux flous résultant de l'application de méta-modèles permettent de décrire précisément le positionnement possible d'un objet par rapport à un objet de référence selon la relation spatiale non-linguistique apprise. Cette description ne suffit pourtant pas à garantir qu'un objet *bien positionné* dans le modèle spatial est un *bon représentant* de la relation spatiale. Il faut en effet pour cela vérifier non seulement que l'objet est bien positionné, mais aussi qu'il a une occupation de l'espace similaire aux objets d'apprentissage.

Ce besoin est mis en évidence sur l'exemple du modèle spatial représenté par la figure 3.22. Il s'agit du modèle décrivant la relation apprise entre deux objets s'intersectant (une référence verticale et un objet cible horizontal). L'illustration représente uniquement la composante positive du modèle spatial lié au point de vue *à droite* (nous avons détaillé l'apprentissage de ce modèle par la figure 3.8, voir l'exemple (e)).

L'analyse de ce modèle spatial montre que deux zones sont considérées comme acceptables : la zone parfaitement à droite de la référence et le demi-plan à gauche de la référence.

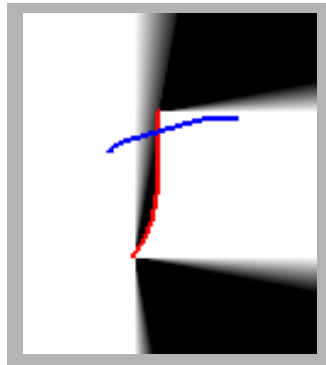


FIGURE 3.22: Illustration du besoin de modélisation de l'étendue des objets pour décrire une situation d'intersection. Le modèle spatial appris à partir d'objets s'intersectant, selon le point de vue à *droite*, considère comme admissible des objets situés complètement à gauche ou à droite de la référence. (Ce modèle spatial a été défini en détail à la figure 3.8(e)).

D'après le modèle, un objet peut être donc correctement positionné :

- s'il est complètement situé dans la zone parfaitement à droite de R ;
- s'il est complètement situé dans la zone à gauche de R ;
- s'il est partagé entre ces deux zones, une partie de l'objet étant à droite et l'autre partie à gauche de R .

Ainsi, même si les objets d'apprentissage avaient tous la propriété de s'étendre de part et d'autre de l'objet de référence, cette information a été perdue dans le modèle spatial. Celui-ci ne décrit en effet que les zones *possibles* de positionnement, mais ne peut exprimer le fait que l'objet cible doit intersecter la référence. Un modèle d'étendue, en revanche, pourra modéliser cette caractéristique en décrivant le fait qu'un objet n'est conforme à la relation spatiale que si certains de ses points sont *parfaitement à droite de R* tandis que d'autres ne sont *pas du tout à droite de R* . Lors de la vérification d'une relation spatiale, un objet qui serait positionné complètement à droite de la référence (et qui aurait donc un bon score de positionnement) serait pénalisé parce qu'il ne s'étend pas dans la zone à gauche de la référence. En supposant que l'objet cible est *connexe*, la combinaison du modèle de positionnement (représenté par la figure 3.22) avec ce modèle d'étendue revient à décrire que l'objet cible doit *intersecter* l'objet de référence.

Le second exemple illustrant le besoin de l'introduction d'une modélisation de l'étendue des objets cibles est livré par la figure 3.23. Deux modèles spatiaux identiques sont représentés et développés sur le même objet de référence, avec des objets cibles (imaginaires) différents. On peut imaginer que le premier modèle spatial (image (a)) résulte de l'apprentissage d'une relation \mathcal{R}_a à partir d'objets cibles étendus verticalement et dont la position est très stable par rapport à leur référence. Le second modèle (image (b)) peut avoir été appris à partir d'objets moins étendus verticalement, mais dont la position verticale est très variable par rapport à la référence (relation spatiale \mathcal{R}_b). Les deux cas ont abouti à la définition du même modèle spatial, qui considère comme également acceptable le positionnement des deux objets représentés.

Pourtant, il est souhaitable de pouvoir différencier ces deux relations spatiales. \mathcal{R}_a est caractérisée par le fait que l'objet cible doit couvrir la zone d'acceptabilité verticalement

3.5. Intégration d'un modèle d'étendue

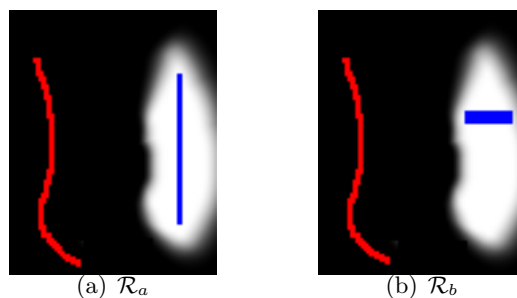


FIGURE 3.23: Mise en évidence du besoin de modélisation de l'étendue des objets pour différencier des relations dont le modèle spatial est similaire. Deux relations spatiales distinctes peuvent aboutir au même modèle spatial malgré une différence importante dans la façon dont l'objet cible couvre l'espace : avec une grande étendue verticale (a) ou non (b).

dans toute son étendue, tandis que pour \mathcal{R}_b l'objet cible doit la couvrir horizontalement mais aussi avoir une faible étendue verticale. Pour \mathcal{R}_a , un modèle d'étendue (selon le point de vue *en haut*, par exemple) consistera à décrire que l'objet doit avoir à la fois des points nettement *en haut* et d'autres faiblement *en haut* par rapport à la référence. Pour \mathcal{R}_b , on pourra décrire que l'objet cible *ne doit pas* avoir à la fois des points fortement *en haut* et d'autres faiblement *en haut* par rapport à la référence.

En fait, étant donnée la possible variabilité de positionnement (par exemple la variabilité verticale des objets pour la relation \mathcal{R}_b), l'étendue sera décrite relativement au degré médian des points des objets par rapport à la relation *être en haut de R*. C'est en effet *l'amplitude* des degrés qui doit être décrite par le modèle d'étendue, plutôt que le niveau moyen (*dans quelle mesure les points de l'objet cible sont en haut de R*) qui est déjà décrit par le modèle de positionnement. En ce sens, la description de l'étendue fournit un indice sur l'aspect intrinsèque de l'objet cible. Cette étendue est néanmoins bien à distinguer d'une simple mesure de *dimensions* dans l'espace euclidien, car elle est décrite dans un espace de représentation qui est défini en fonction d'une direction et qui tient compte de la forme de la référence. Par exemple, selon le point de vue *à droite*, les objets des deux images de la figure 3.23 ont une étendue identique et réduite à zéro puisque tous leurs points sont *parfaitement à droite* de *R*.

3.5.2 Etendue d'un objet vu comme un nuage de points

Dans le cas le plus général, un objet à positionner par rapport à un objet de référence est représenté par un ensemble de points. Cela est valable pour la représentation en-ligne (ensemble de points échantillonnés) ou hors-ligne (ensemble de pixels) du signal manuscrit, et quelle que soit la complexité graphique de l'objet (fait en un seul tracé comme en plusieurs). L'étendue de l'objet peut alors se ramener à la variance ou l'amplitude du nuage de ses points projeté dans les espaces définis par les points de vue directionnels considérés pour la relation. La mesure de l'étendue moyenne des objets d'apprentissage est faite pour chaque point de vue. En notant $(R_i, A_i)_{i=1..N}$ les N paires d'objets d'apprentissage (représentatives de la relation spatiale \mathcal{R}), on cherche à construire un modèle du profil moyen de chaque

objet dans l'espace des degrés atteints par ses points selon le point de vue considéré (par exemple *être dans la direction* α). Pour un objet A_i , on note $E_{\alpha,i}$ la distribution des degrés atteints par ses points dans le modèle spatial $\mu_\alpha(R_i)$ et $c_{\alpha,i}$ sa valeur centrale, c'est-à-dire :

$$c_{\alpha,i} = \frac{\inf\{E_{\alpha,i}\} + \sup\{E_{\alpha,i}\}}{2}. \quad (3.21)$$

$c_{\alpha,i}$ est donc le milieu de l'étendue des degrés occupés par A_i . Cette valeur est utilisée pour centrer les étendues des différents objets d'apprentissage et ainsi construire un sous-ensemble flou décrivant un modèle de profil des objets A_i par rapport à cette valeur centrale. Un histogramme $(E_{\alpha,i}^c)_k$ peut être construit pour chaque objet, en s'appuyant sur les ensembles I_k définissant une partition de $[0; 1]$ (ce sont les ensembles qui ont servi à construire H et G , voir leur définition au paragraphe 3.2.1.2) :

$$(E_{\alpha,i}^c)_k = |S_{k_m - k_c}| \quad \text{avec } S_k = \{p \in A_i, \mu_\alpha(R_i)(p) \in I_k\}, \quad (3.22)$$

avec k_m l'indice médian de l'histogramme et k_c l'indice k tel que $c_{\alpha,i} \in I_k$. Ainsi, $(E_{\alpha,i}^c)$ est l'histogramme centré sur la valeur c de la distribution des degrés atteints par un objet A_i . Les fréquences mesurées par $(E_{\alpha,i}^c)$ sont normalisées par rapport à la fréquence maximale. Les profils de tous les objets A_i d'apprentissage ainsi centrés peuvent être accumulés dans un histogramme (E_α^c) , qui représente donc le profil moyen d'étendue des objets A_i autour de leur valeur centrale c (pour le point de vue orienté par α). L'histogramme est également normalisé par rapport à sa fréquence maximale.

Exploitation par mesure de similarité d'histogrammes Le calcul de l'adéquation d'étendue entre un objet cible et le modèle d'étendue, chacun étant décrit par son histogramme centré et normalisé, repose sur le calcul d'un indice de similarité entre les deux histogrammes. De très nombreux indices de similarité d'histogrammes existent dans la littérature et une bonne synthèse en a été faite par Cha [Cha07]. Nous avons retenu un indice qui procède à l'intersection des deux histogrammes E et F de longueur identique K , dénommé indice de similarité de Ruzicka, qui considère les deux histogrammes comme deux ensembles flous :

$$s = \frac{\sum_{k=0}^K \min(E_k, F_k)}{\sum_{k=0}^K \max(E_k, F_k)}. \quad (3.23)$$

Le score s_{α_i} mesure la similarité entre les distributions de degrés d'un objet par rapport à un modèle appris, pour un point de vue donné (α_i) sur la relation spatiale. La mesure globale d'adéquation de l'étendue d'un objet par rapport à tout le modèle peut se faire en combinant les scores de similarité obtenus pour chaque point de vue par une moyenne géométrique :

$$\sigma_{\mathcal{M}}^{\mathcal{R}}(R)(A) = \prod_{i=1}^d s_{\alpha_i}^{\frac{1}{d}}. \quad (3.24)$$

Le score $\sigma^{\mathcal{R}}(R)(A)$ fournit une mesure de score d'étendue de l'objet A par rapport à R relativement au modèle spatial \mathcal{M} . Ce score peut être utilisé en combinaison avec le score de positionnement pur $\mu^{\mathcal{R}}(R)(A)$ (défini précédemment) pour répondre à une tâche d'évaluation de relation spatiale.

3.5. Intégration d'un modèle d'étendue

Exemples de modèles La figure 3.24 représente plusieurs modèles d'étendue construits à partir de deux exemples, pour différentes relations spatiales (relations \mathcal{R}_1 à \mathcal{R}_5). Dans chaque cas, deux couleurs distinguent les points des deux objets d'apprentissage dans l'histogramme, ce qui permet de constater la similarité de l'étendue des objets par rapport au point de vue considéré pour la relation spatiale (*être à droite de*).

Les modèles d'étendues pour les relations \mathcal{R}_1 et \mathcal{R}_2 décrivent tout simplement que dans chaque cas, tous les points des objets ont exactement le même degré d'adéquation à la relation *être à droite de*. En effet, dans le premier cas, tous les points sont parfaitement à droite de la référence et dans le second cas, tous les points sont dans la zone *pas du tout à droite* de la référence.

Les cas des relations \mathcal{R}_3 et \mathcal{R}_4 sont très différents, puisque cette fois les points des objets sont ventilés sur une large gamme de degrés selon la direction *à droite de R*. En fait, pour ces deux relations, le modèle d'étendue décrit que un objet doit couvrir quasi uniformément toutes les nuances de degrés entre 0 et 1 vis-à-vis de cette relation de position directionnelle.

Finalement, le cas de la relation \mathcal{R}_5 met en évidence l'intérêt de la modélisation de l'étendue pour le cas de la modélisation d'une intersection entre les objets. Dans ce cas, un objet *doit s'étendre de part et d'autre de la référence* pour être conforme au modèle d'étendue représenté par cet histogramme, c'est-à-dire avoir certains points *parfaitement à droite* (degré de 1) et d'autres points *pas du tout à droite* de R (degré de 0).

Discussion Le modèle d'étendue présenté ci-dessus est efficace pour distinguer les situations de la figure 3.24 entre elles. Par exemple, la distinction entre les relations \mathcal{R}_1 , \mathcal{R}_3 et \mathcal{R}_5 est facilitée par les modèles d'étendue selon le point de vue *à droite de* qui sont tous très différents. Tout comme pour la richesse de description du positionnement, la combinaison des points de vue bénéficie aussi largement à la modélisation de l'étendue car elle augmente les possibilités de distinguer des cas de relations spatiales. Par ailleurs, d'un simple point de vue pratique, le calcul de l'adéquation d'un objet par rapport à un modèle d'étendue peut se faire directement, à la suite du calcul de son bon positionnement. Il suffit pour cela de mémoriser la distribution des degrés atteints par les points de l'objet à positionner par rapport aux relations directionnelles selon chaque point de vue, de la centrer et de la comparer au modèle par la mesure donnée à l'équation (3.23). Puisqu'il s'appuie sur les degrés obtenus par application des opérateurs directionnels morphologiques, le modèle d'étendue bénéficie aussi des avantages offerts par cette représentation et prend donc en compte la forme de l'objet de référence.

Néanmoins, les modèles d'étendue souffrent d'une imprécision due au fait de considérer les objets comme des nuages de points qui sont projetés indépendamment selon chaque point de vue. Par exemple, la figure 3.25 représente deux situations où les objets à positionner ont des étendues qui semblent très différentes (le premier est un segment vertical, le second un segment horizontal). Pourtant, lorsque l'on compare les modèles d'étendue appris pour chacune des situations, on constate qu'ils sont au contraire très similaires. En effet, les nuages de points projetés sur chaque point de vue dans les deux cas ont des profils sensiblement équivalents et couvrent la même gamme de degrés selon les directions *être en haut*, *être à droite*...

Dans la seconde stratégie de représentation de l'étendue des objets, nous identifions des points particuliers de l'objet dont la position par rapport au nuage de points projeté est

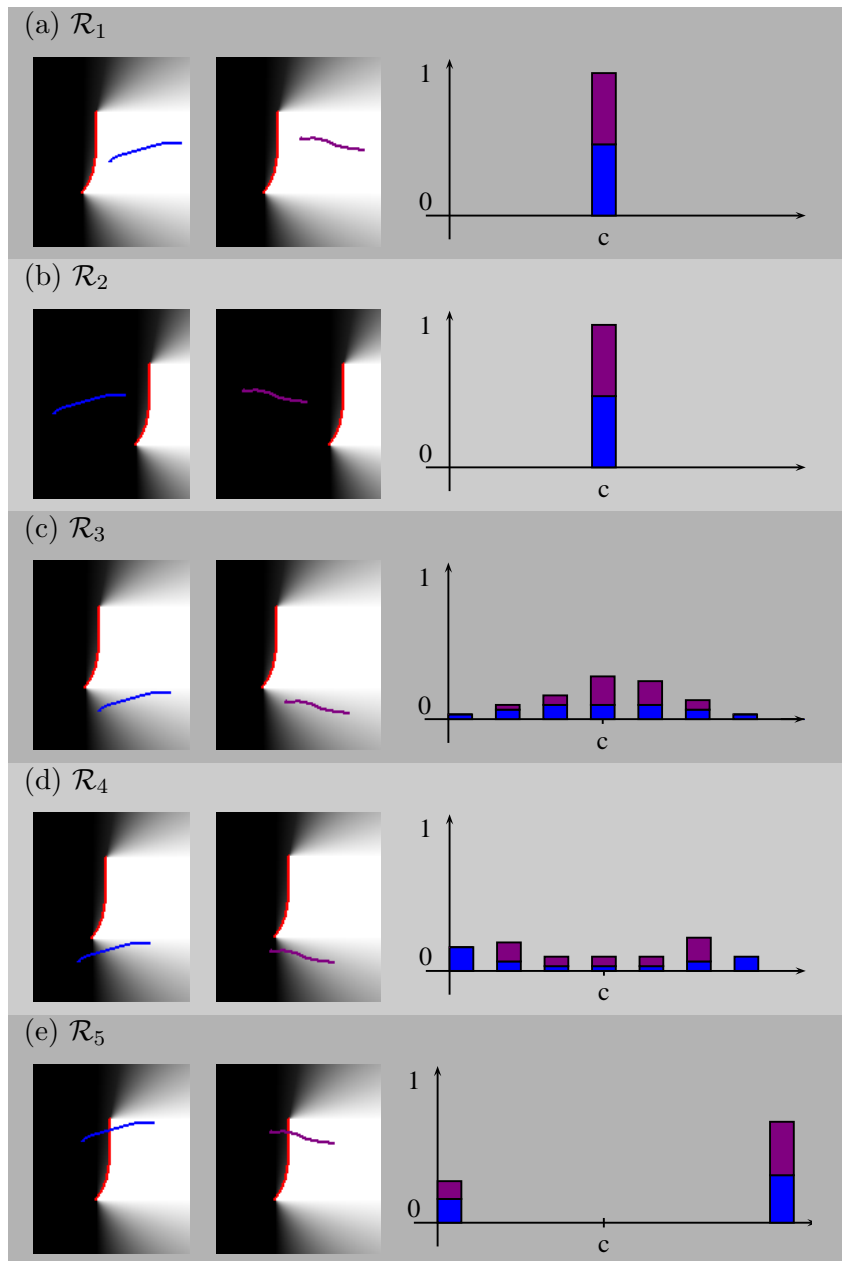


FIGURE 3.24: Visualisation du modèle d'étendue appris à partir de deux exemples d'une même relation spatiale, pour 5 relations différentes. A chaque fois, le modèle d'étendue est un histogramme qui comptabilise la distribution des points des objets dans l'espace des degrés selon le point de vue à droite.

analysée pour tous les points de vue à la fois, ce qui permet une plus grande précision du modèle d'étendue.

3.5. Intégration d'un modèle d'étendue

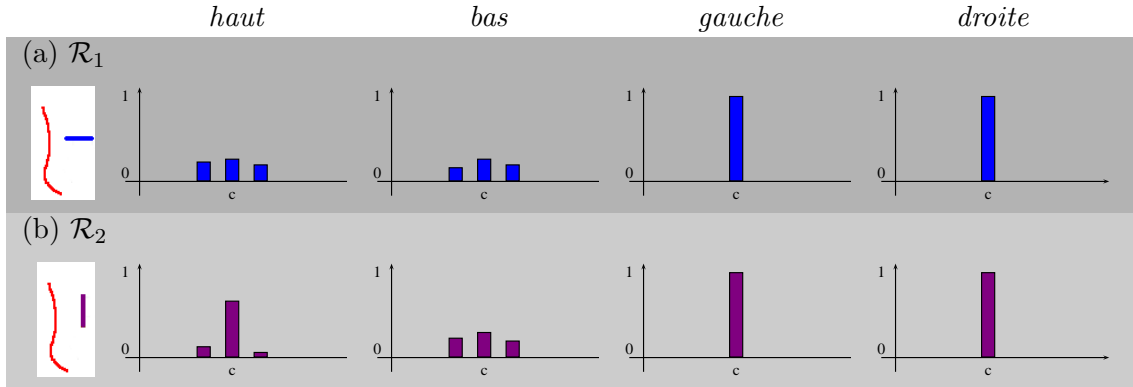


FIGURE 3.25: La projection indépendante des objets vus comme des nuages de points peut conduire à des modèles d'étendue identiques même pour des situations très différentes : cas d'un objet horizontal (a) et d'un objet vertical (b).

3.5.3 Etendue d'un objet par considération de ses points singuliers

Dans le cas où les objets à positionner sont des primitives de tracé relativement simples, la position de certains de leurs points (par exemple leurs extrémités) les caractérise assez précisément. En particulier lorsque les tracés sont rectilignes ou en arc de cercle, la connaissance de leurs extrémités peut fournir une description satisfaisante de leur étendue. L'avantage de se concentrer sur des points particuliers est que l'objet n'est plus vu comme un nuage de points projeté indépendamment selon les points de vue. Au contraire, les points particuliers sont décrits un à un, selon tous les points de vue à la fois. Pour être en bonne adéquation avec le modèle d'étendue, un même point de l'objet à positionner devra donc être bien positionné selon tous les modèles partiels d'étendue à la fois (c'est-à-dire selon tous les points de vue sur la relation). Cette différence majeure par rapport à la version précédente permettra notamment de distinguer les cas de l'objet horizontal et vertical de la figure 3.25.

Points singuliers et projection Pour des objets en-ligne, les levés et posés de stylo déterminent les points de début et de fin de tracés qui constituent des points d'intérêt de l'objet manuscrit. Nous nous intéressons ici uniquement au cas d'objets en-ligne, car il est particulièrement aisé de définir des points d'intérêt pour ces objets. Cependant, d'autres stratégies d'extraction de points pourraient être étudiées pour le cas d'objets hors-ligne (par exemple en appliquant une méthode de détection de points d'intérêt à base de gradients directionnels).

Le principe proposé consiste à modéliser la position moyenne des points particuliers des objets au sein du nuage de points selon chaque point de vue considéré pour la relation. Comme précédemment, la description de l'étendue est normalisée par rapport à la valeur centrale c de la distribution des degrés de l'objet (voir l'équation (3.22)).

Pour un objet simple, constitué d'un seul tracé comportant p points $S = s_1..s_p$, nous définissons trois points particuliers s_a , s_b et s_m . s_a et s_b correspondent aux extrémités du tracé, tandis que s_m désigne le point de mi-parcours du tracé. La correspondance entre s_a , s_b et les deux extrémités s_1 et s_p se fait selon des critères géométriques. s_a est défini

comme l'extrémité la plus proche du coin haut gauche de la boîte englobante du tracé et s_b est défini comme l'autre extrémité du tracé. En cas d'ambiguïté, c'est-à-dire lorsque les deux extrémités sont presque équidistantes du coin haut gauche, c'est l'extrémité la plus à gauche qui est considérée pour s_a . Le point de mi-parcours du tracé est le point qui partage au mieux la longueur du tracé en deux parties de longueur égale. s_m est donc défini comme le premier point $s_i, i = 1..p$ tel que :

$$\sum_{k=1}^i \|s_k s_{k+1}\| > \frac{1}{2} * \sum_{k=1}^{p-1} \|s_k s_{k+1}\|.$$

La figure 3.26 présente deux exemples de tracés simples en-ligne avec les points particuliers s_a, s_b et s_m déterminés pour chaque cas.

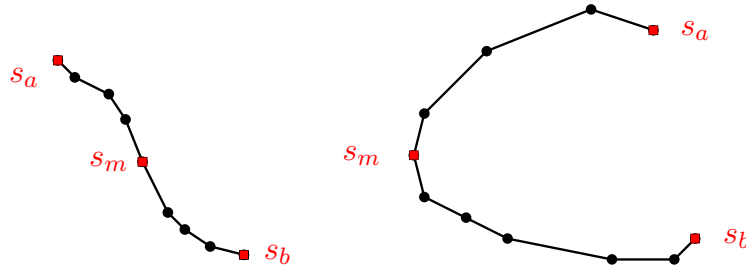


FIGURE 3.26: Exemples de points particuliers extraits de deux tracés simples.

Pour un point de vue, trois modèles sont construits, chacun modélisant la position attendue pour l'un des points d'intérêt dans la distribution des degrés. Cette position moyenne est à nouveau décrite par un histogramme, centré sur la valeur c définie précédemment. Ainsi, la position du point particulier s_a relativement au centre de la distribution c des degrés de l'objet selon le point de vue α est considérée pour chaque exemple d'apprentissage i .

Pour une paire d'objets (R_i, A_i) , considérée selon le point de vue α , $(E_{\alpha,i}^c)$ représente la distribution des degrés atteints par les points de A , centrée par rapport à c . Parmi ces points, le point particulier s_a atteint un degré exprimé par $\mu_{\alpha}(R_i)(s_a)$. Le modèle de position des points s_a est construit en comptabilisant à partir des exemples d'apprentissage le nombre de fois où le degré $\mu_{\alpha}(R_i)(s_a)$ tombe dans chaque section de l'historgramme $(E_{\alpha,i}^c)$. Il en résulte un histogramme $(h_a)_k$ qui est normalisé par sa fréquence maximale. Cet histogramme est ensuite utilisé comme un modèle flou de la position du point s_a par rapport au point de normalisation c selon le point de vue porté par α .

Exemples La figure 3.27 présente quelques exemples des histogrammes appris pour positionner les trois points particuliers s_a, s_m et s_b selon le point de vue être à droite de. Chaque ligne du tableau présente un cas de figure différent, illustré dans la première colonne par une paire d'objets d'apprentissage. La référence est en rouge et l'objet argument en bleu, avec ses points particuliers mis en évidence par des couleurs : orange pour s_a , vert pour s_m et marron pour s_b . Les trois modèles d'étendue associés sont représentés par leurs histogrammes h_a, h_m et h_b dans les trois colonnes de droite.

3.5. Intégration d'un modèle d'étendue

Dans les deux premiers cas, tous les points de l'objet argument ont le même degré (ils sont tous parfaitement à droite de R pour la relation \mathcal{R}_1 , et pas du tout à droite dans le cas de \mathcal{R}_2). Les modèles d'étendue imposent donc que les points particuliers s_a , s_b et s_m aient le même degré que le point central de la distribution des degrés (c).

Pour les cas \mathcal{R}_3 et \mathcal{R}_4 , les modèles sont plus informatifs. Ils décrivent que le point s_a doit avoir un degré *d'être à droite* inférieur au centre de la distribution c , tandis que s_b doit être *plus à droite* que c . Le point de mi-parcours doit quant à lui coïncider avec le degré central de la distribution.

Dans le dernier cas, les positions de points extrémités sont comparables à celles observées pour les relations \mathcal{R}_3 et \mathcal{R}_4 . Le modèle de position pour le point s_m , en revanche, a la particularité d'être bimodal. Cela est dû au fait que, parmi les exemples d'apprentissage, les points des objets A_i s'étendent entre les degrés 0 et 1 (puisqu'ils s'étendent de part et d'autre de l'objet de référence R_i associé). Dans ce cas ; le centre de la distribution vaut 0,5 pour tous les objets d'apprentissage. Le point de mi-parcours, en revanche, peut prendre les valeurs de 0 (si plus de la moitié du tracé de l'objet est dans la partie sombre du paysage flou) ou de 1 (si plus de la moitié du tracé est dans la partie claire du paysage). Le modèle appris tolère donc les deux possibilités pour le degré du point de mi-parcours par rapport à c .

Exploitation Le calcul d'un score d'adéquation entre ce type de modèles d'étendue et une paire d'objets met en œuvre une fusion point par point, considérant tous les points de vues pris en compte dans le modèle. La force de cette représentation est en effet de considérer la position d'un même point s_a selon tous les points de vue à la fois. Un score de bon positionnement peut donc être calculé pour chaque point particulier. Par exemple pour le point s_a , son score de bon positionnement σ_a par rapport au modèle appris pour une relation \mathcal{R} est défini par :

$$\sigma_a^{\mathcal{R}} = t[h_a(\mu_a^c)_{\alpha_i}]_{i=1..d} \quad \text{avec} \quad \mu_a^c = \mu_{\alpha_i}(R)(s_a) - c, \quad (3.25)$$

où d est le nombre de points de vue considérés dans le modèle et l'indice i représente chaque point de vue. Le modèle de position h_a est appliqué comme une fonction floue sur μ_a^c , qui représente le degré de s_a dans le paysage flou $\mu_{\alpha_i}(R)$ recentré par rapport à c . L'opérateur de t-norme utilisé dans la combinaison traduit le fait que s_a doit être bien positionné par rapport au modèle d'étendue h_a selon chaque point de vue α_i . On a donc une parfaite homogénéité avec l'évaluation du positionnement d'un point de l'espace dans un modèle spatial, qui met en œuvre la conjonction floue de ses degrés de positionnement par rapport à tous les points de vue du méta-modèle (voir l'équation (3.9)).

Un calcul similaire peut être fait pour chaque point particulier. Un score global d'évaluation de l'étendue d'un objet par rapport à un modèle peut être défini en combinant les scores des différents points particuliers :

$$\sigma_{\mathcal{M}}^{\mathcal{R}}(R)(A) = (\sigma_a^{\mathcal{R}})^{\frac{1}{3}} * (\sigma_m^{\mathcal{R}})^{\frac{1}{3}} * (\sigma_b^{\mathcal{R}})^{\frac{1}{3}}. \quad (3.26)$$

Tout comme pour la mesure d'étendue par nuages de points (voir au paragraphe précédent), le score $\sigma_{\mathcal{M}}^{\mathcal{R}}(R)(A)$ peut être combiné à un score de positionnement pour une tâche d'évaluation du positionnement relatif de deux objets par rapport à un modèle appris.

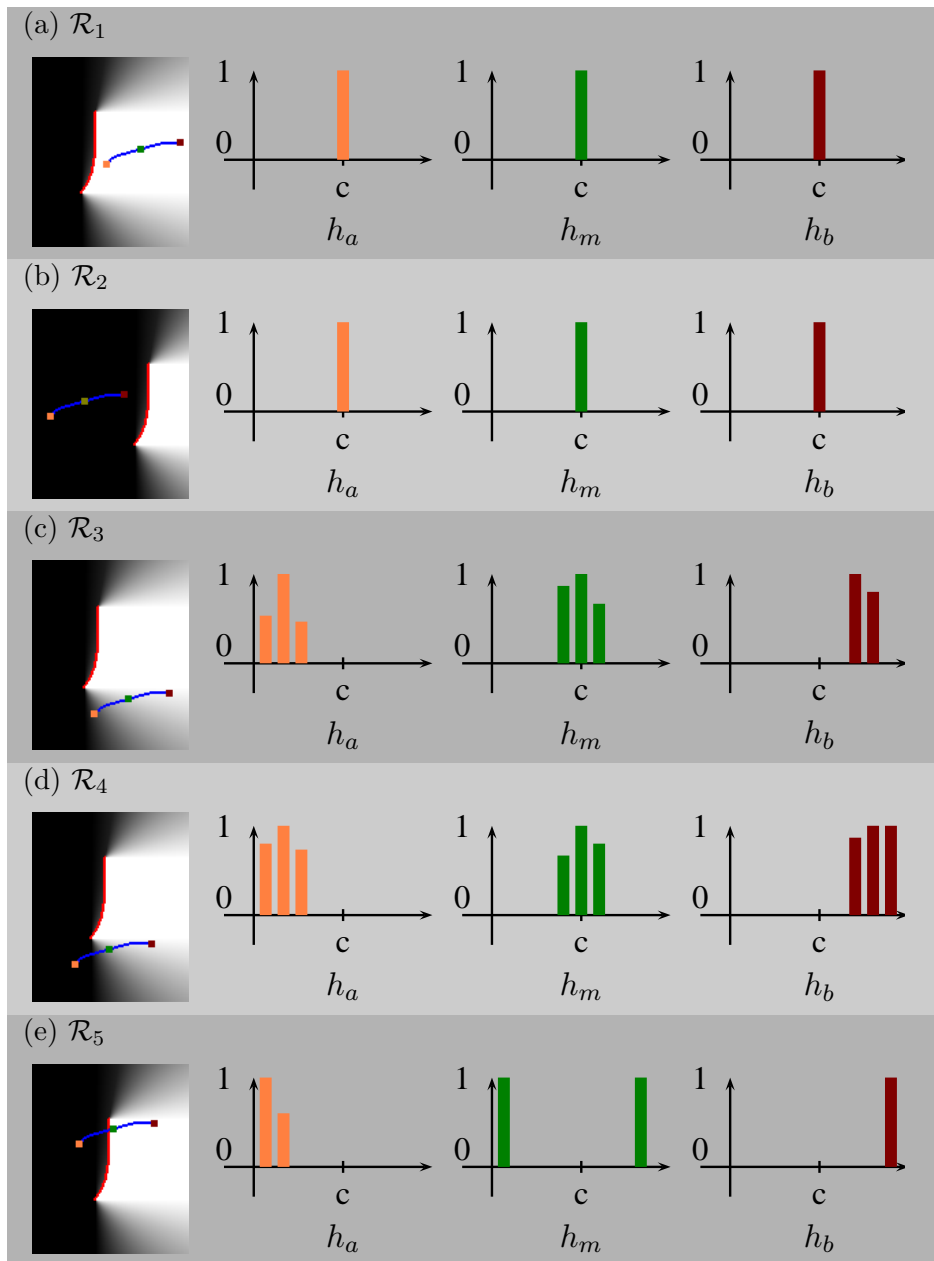


FIGURE 3.27: Exemples d'apprentissage du modèle d'étendue pour différentes configurations.

Discussion Comme nous l'avons précisé, le fait de considérer les points à travers tous les points de vue à la fois, plutôt que de projeter indépendamment les objets comme des nuages de points selon chaque point de vue, permet une plus grande précision de la description de l'étendue. Nous illustrons cet avantage au moyen de la figure 3.28, où les deux exemples de la figure 3.25 sont repris. Pour chacun des deux cas, les trois modèles d'étendue liés aux trois points particuliers sont représentés pour les quatre points de vue (un point de vue par colonne). La comparaison des modèles obtenus pour les points s_a entre le cas \mathcal{R}_1 (objet

3.5. Intégration d'un modèle d'étendue

horizontal) et le cas \mathcal{R}_2 (objet vertical) montre qu'ils diffèrent grandement selon le point de vue *bas*. Il est donc possible de distinguer aisément les deux cas de figure au moyen de ces modèles d'étendue plus élaborés, grâce à l'information plus précise qu'ils contiennent sur l'occupation de l'espace par l'objet argument.

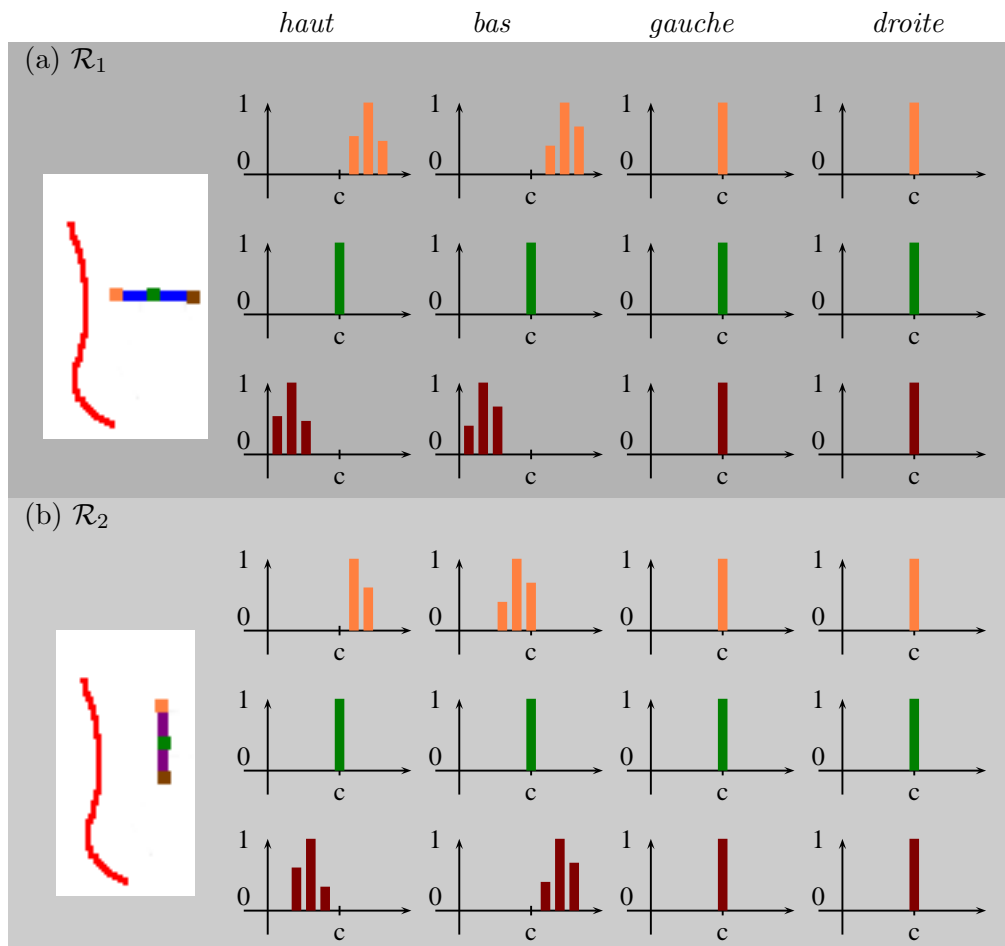


FIGURE 3.28: La modélisation de l'étendue par prise en compte des points particuliers des objets permet de distinguer le cas d'un objet horizontal (a) et d'un objet vertical (b), selon la direction *en bas* (deuxième colonne d'histogrammes).

3.5.4 Synthèse

Deux modèles différents ont été présentés pour décrire l'étendue de l'objet argument en complément du modèle de positionnement pur exposé dans les sections précédentes. Par définition, ces modèles ne peuvent être utilisés que pour la tâche d'évaluation de positionnement relatif de deux objets. Le principe de ces modèles est de vérifier le positionnement simultané de l'objet cible dans différentes régions de l'espace, les régions étant décrites par les points de vue directionnels par rapport à l'objet de référence. Ces modèles permettent de favoriser les objets dont l'étendue, ou la façon de couvrir l'espace autour de la référence,

est similaire à celle constatée sur les objets d'apprentissage. Leur apport se traduit par la capacité à distinguer des relations spatiales qui offrent des zones d'acceptabilité similaires avec des objets à l'étendue différente, mais aussi de modéliser une contrainte de nature topologique dans le cas de d'une situation d'intersection (positionnement de l'objet de part et d'autre de la référence).

Le premier modèle, dit par projection de nuage de points, pourra être utilisé notamment dans le cas où les objets à positionner sont de forme complexe, comportent des tracés multiples, ou lorsqu'il est difficile d'en extraire des points particuliers (par exemple dans le cas d'objets hors-ligne). Le second modèle, qui décrit plus précisément l'étendue des objets, pourra être utilisé lorsque les objets sont de simples tracés manuscrits en-ligne (notamment s'ils sont assimilables à des segments ou arcs de cercle).

3.6 Résultats qualitatifs

Nous avons présenté dans ce chapitre une formalisation nouvelle pour modéliser les relations spatiales entre des tracés manuscrits, en insistant sur les objectifs de généralité de la représentation par rapport à la nature et à la complexité des objets traités, mais aussi à la catégorie des relations spatiales exprimées pour répondre aux besoins de description des relations spatiales qui ont été recensés au cours de l'état de l'art (chapitre 2). Nous avons pour cela distingué les besoins de positionnement entre primitives simples de tracés et les besoins de positionnement entre des objets plus complexes, par exemple au sein d'expressions ou de schémas structurés.

Dans cette partie, nous allons démontrer qualitativement la capacité des méta-modèles spatiaux à décrire des situations de positionnement entre des objets de différentes natures, de complexité variable. Cette présentation est organisée en distinguant ici aussi les besoins de description de relations entre tracés primitifs simples et entre tracés plus complexes. Une validation quantitative de l'intérêt des méta-modèles spatiaux pour la description du positionnement relatif de différents types d'objets sera présentée en détail dans la partie expérimentale (voir le chapitre 5).

Pour avoir un aperçu des techniques algorithmiques et des considérations de complexité pour le calcul des dilatations morphologiques, le lecteur est invité à se reporter à l'annexe B. Cette annexe met en évidence la simplicité de mise en œuvre des méta-modèles spatiaux aussi bien sur des données hors-ligne (où les objets sont représentés par un signal de type image), que sur des données en-ligne (où les objets sont décrits par des séquences ordonnées de points échantillonnés).

3.6.1 Modélisation de relations entre primitives simples

Nous présentons ici quelques exemples de situations de relations entre primitives de tracé simples en illustrant à chaque fois par plusieurs exemples les capacités de descriptions offertes par les méta-modèles. Nous justifierons sur les cas de figure pertinents les bénéfices apportés par les différentes variantes des méta-modèles spatiaux : introduction de modèles de distance, modélisation bipolaire, intégration de modèles d'étendue.

3.6. Résultats qualitatifs

3.6.1.1 Primitives de tracé distantes

Nous nous intéressons d'abord au cas général du positionnement relatif de primitives distantes (pas de recouvrement de boîtes englobantes ni de connexion des tracés). Ce cas de figure correspond au besoin le plus fréquent et on peut trouver des exemples de ces relations dans de nombreuses natures de données. Le tableau 3.29 donne quelques exemples de ces relations issues de l'écriture du français ((a),(b)) et du chinois ((c),(d)). La première colonne présente les deux objets mis en relation (la référence est en rouge, l'argument en bleu) et la seconde colonne représente leurs boîtes englobantes. Un méta-modèle spatial appris est développé dans la troisième colonne sur la référence; il ne comporte pas d'information de distance. Le méta-modèle représenté dans la quatrième colonne comporte quant à lui une information de distance (selon la stratégie *directionnelle*, voir la section 3.3.2).

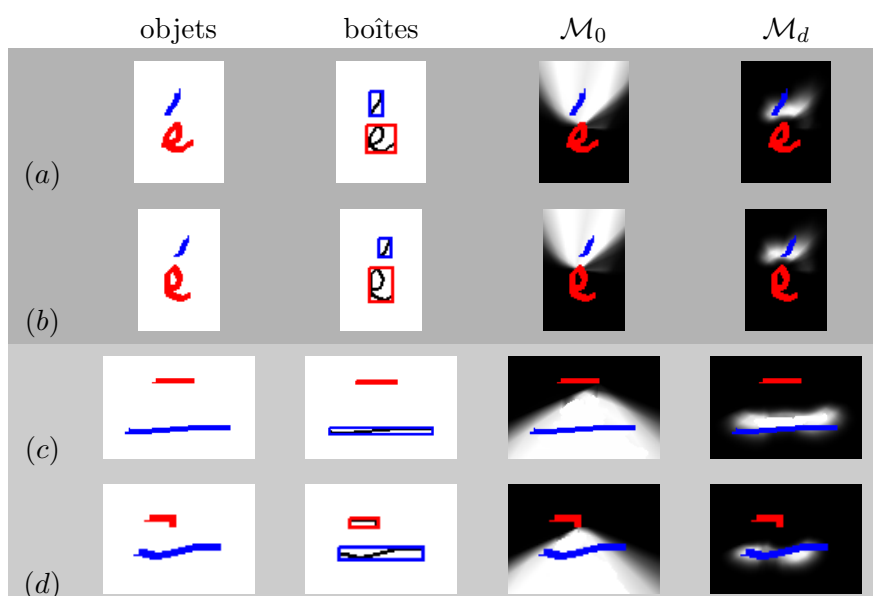


FIGURE 3.29: Modèles spatiaux appris pour des primitives distantes (\mathcal{M}_0 sans modèle de distance et \mathcal{M}_d avec modèles de distance *directionnels*). Deux relations spatiales sont présentées, l'une issue de l'écriture du français ((a) et (b)) et l'autre de l'écriture de caractères chinois ((c) et (d)). Deux paires d'objets différents illustrent chaque relation.

Pour chacune des deux relations illustrées dans la figure 3.29, les modèles sont appliqués à deux paires d'objets différents, ce qui permet d'apprécier la capacité des modèles à bien s'adapter à la forme de la référence et de vérifier la bonne correspondance de l'objet argument avec le modèle spatial. La différence visuelle entre les modèles \mathcal{M}_0 (sans distance) et \mathcal{M}_d (avec distance) fournit un indice de l'intérêt de prendre en considération la distance lorsque l'on décrit ce type de relations spatiales. Les modèles sans distance sont en effet trop permissifs car ils s'étendent à l'infini dans la direction autorisée. A l'inverse, les modèles limités en distance sont capables de beaucoup mieux décrire une zone réaliste du positionnement de l'objet bleu. Les deux situations illustrées ici mettent en jeu des primitives de tracé simple assez différentes : dans le premier cas il s'agit d'un tracé cursif de la lettre « e », tandis que dans le second cas il s'agit d'un simple segment de droite horizontal. On

constate que, dans les deux cas, le comportement des modèles spatiaux appris est conforme à la description intuitive de la relation. Les directions mises en jeu dans ces deux relations sont le haut et le bas respectivement. Les modèles ayant été construits à partir des quatre points de vue cardinaux (stratégie générique), la direction principale de la relation n'a pas d'influence sur la richesse de description. On aurait donc des résultats similaires pour des situations de primitives disposées selon d'autres directions par rapport à la référence.

3.6.1.2 Concavités et recouvrement de boîtes englobantes

L'un des intérêts majeurs de la prise en compte des formes des tracés dans la modélisation que nous avons présentée est la bonne gestion des cas de concavités et des recouvrements entre les boîtes englobantes des objets. Dans ce cas, l'approximation des tracés par leur boîtes englobantes est à proscrire pour les raisons déjà présentées au paragraphe 2.1.2.3. Comme le montrent les exemples du tableau 3.30, les modèles-spatiaux s'accommodent bien de ces situations. Il s'agit de deux relations issues respectivement de l'écriture du chinois ((a) et (b)) et du tracé de schémas électriques ((c) et (d)). Dans les deux cas, l'objet argument est situé plus ou moins nettement dans une concavité de l'objet de référence de sorte que les boîtes englobantes des objets se recouvrent. Dans le premier cas ((a) et (b)), l'objet argument est pourtant perçu comme nettement à gauche de la référence. Dans le second cas ((c) et (d)), l'objet est encerclé par la référence (la concavité est fermée, c'est une boucle complète).

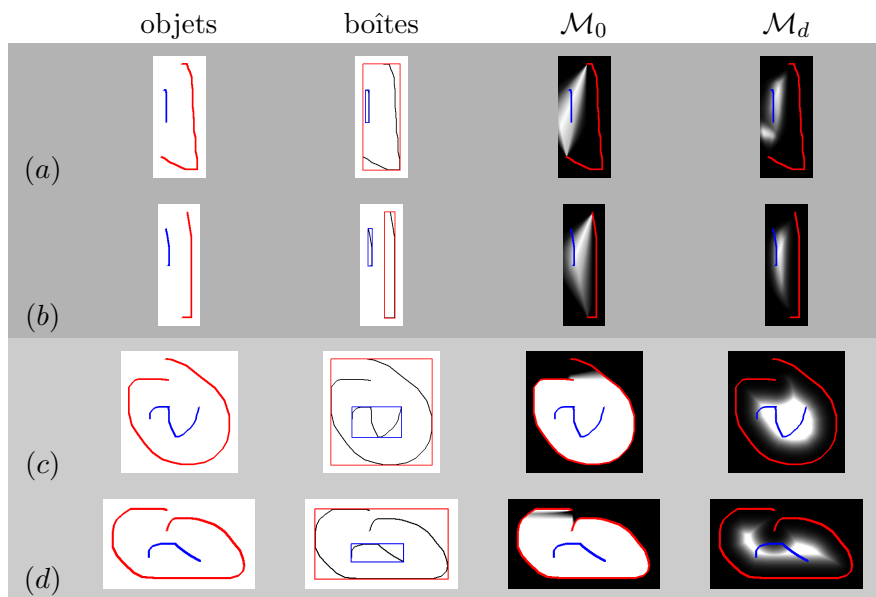


FIGURE 3.30: Modèles spatiaux appris pour des primitives présentant concavités et recouvrements de leurs boîtes englobantes (\mathcal{M}_0 sans modèle de distance et \mathcal{M}_d avec modèles de distance *directionnels*). Deux relations spatiales sont présentées, l'une issue de l'écriture du chinois ((a) et (b)) et l'autre de symboles électriques ((c) et (d)). Deux paires d'objets différents illustrent chaque relation.

Ces illustrations permettent de vérifier la bonne prise en compte des concavités de la

3.6. Résultats qualitatifs

forme de référence dans la description. L'apport de la distance est là aussi manifeste. Le cas de la relation d'encerclement ((c) et (d)) illustre un pouvoir singulier des modèles spatiaux. Dans ce cas particulier, l'apprentissage permet de distinguer la zone à l'intérieur de R de la zone à l'extérieur de R . Il s'agit en fait d'un modèle de la relation à l'intérieur de, qui peut s'appliquer quelle que soit la forme de la référence pourvu qu'elle présente une boucle. Le méta-modèle appris est aisément interprétable à partir de sa décomposition en points de vue : il décrit simplement le fait que l'objet bleu doit être à la fois parfaitement à droite, à gauche, en haut et en bas de la référence.

3.6.1.3 Primitives adjacentes

Le cas de figure où les primitives mises en relation sont connectées (relations d'adjacence) est également très courant. Nous prenons comme exemple une relation issue de l'écriture du français (la cédille du « c ») et une autre issue de caractères chinois. Dans la figure 3.31, une paire d'objets illustre chacune de ces relations. Les modèles spatiaux représentés sont le modèle avec distance directionnelle \mathcal{M}_d ainsi que le modèle spatial bipolaire \mathcal{M}_d^* , avec la même distance.

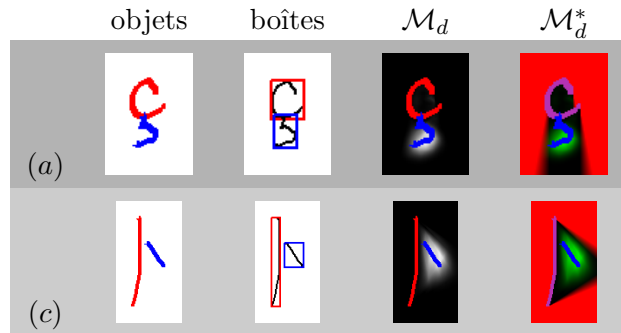


FIGURE 3.31: Modèles spatiaux appris pour des primitives connexes ou adjacentes (\mathcal{M}_d avec modèles de distance *directionnels* et \mathcal{M}_d^* avec bipolarité). Deux relations spatiales sont présentées, l'une issue de l'écriture du français (a) et l'autre du chinois (b).

Ces deux exemples mettent bien en évidence l'apport de la modélisation bipolaire dans le cas de la description de situations d'adjacence. Dans le cas de la relation entre le « c » et la cédille (a), la relation spatiale est caractérisée par le fait que la cédille doit être située *en-dessous* de la lettre. C'est bien ce qui est modélisé par le modèle spatial \mathcal{M}_d , qui ne fait cependant pas de distinction entre les zones à l'intérieur et celles situées sur les côtés du « c ». Le modèle bipolaire \mathcal{M}_d^* , en revanche, a la capacité de distinguer la zone de préférence positive (en vert, c'est le positionnement attendu de la cédille) et les zones de préférences négatives (en rouge, sur les côtés, ce sont les zones absolument interdites pour la cédille). La zone en noir décrit une zone neutre, qui n'exprime pas de préférences positives ou négatives sur la position de la cédille. L'intérieur de la lettre « c » est ainsi une zone neutre ; elle n'est pas interdite car il est assez fréquent de constater que la cédille intersecte le tracé du bas du « c » et déborde ainsi dans cette zone. Dans le cas de la situation (b), la bipolarité permet aussi une finesse de description accrue. Ainsi, la zone à gauche de la référence verticale est formellement exclue du modèle, tandis que la zone de préférence positive exprime une

position proche de l'objet de référence. La zone neutre autorise une certaine souplesse à la zone de préférence, mais uniquement dans la zone à droite de la référence. Autrement dit, il est préférable pour un objet à positionner de s'étendre à droite de la référence plutôt que de l'intersecter en débordant vers la gauche.

La représentation du positionnement par modèles spatiaux ne permet pas de décrire la relation d'adjacence indépendamment de la rotation. On peut cependant noter sur ces exemples que d'une part, les modèles spatiaux appris permettent de gérer correctement les cas de primitives connexes, d'autre part, il est possible d'envisager la modélisation de ce type de relations ou de contraintes géométriques a posteriori de l'évaluation de la position d'un objet par modèle spatial. L'idée que nous avons exploitée pour exprimer l'étendue des objets est un exemple de ces possibilités (voir à la section 3.5).

3.6.1.4 Objets avec intersection

Le cas de primitives dont le tracé s'intersecte est un exemple particulier de relation de positionnement. Comme pour la relation d'adjacence, il n'est pas possible de construire un modèle spatial qui modélise directement une relation d'intersection. Cependant, il est possible d'évaluer la relation d'intersection a posteriori de l'examen de la position de l'objet argument par rapport à la référence. Cette mesure a posteriori a été définie sous la forme de modèles d'étendue (voir à la section 3.5). La modélisation d'une intersection par notre formalisme est donc avant tout un positionnement exprimé sous la forme de directions et de distance par rapport à une référence (c'est le rôle du modèle spatial) et ensuite par une contrainte sur l'étendue de l'objet argument qui doit s'étendre de part et d'autre de la référence (c'est le rôle du modèle d'étendue).

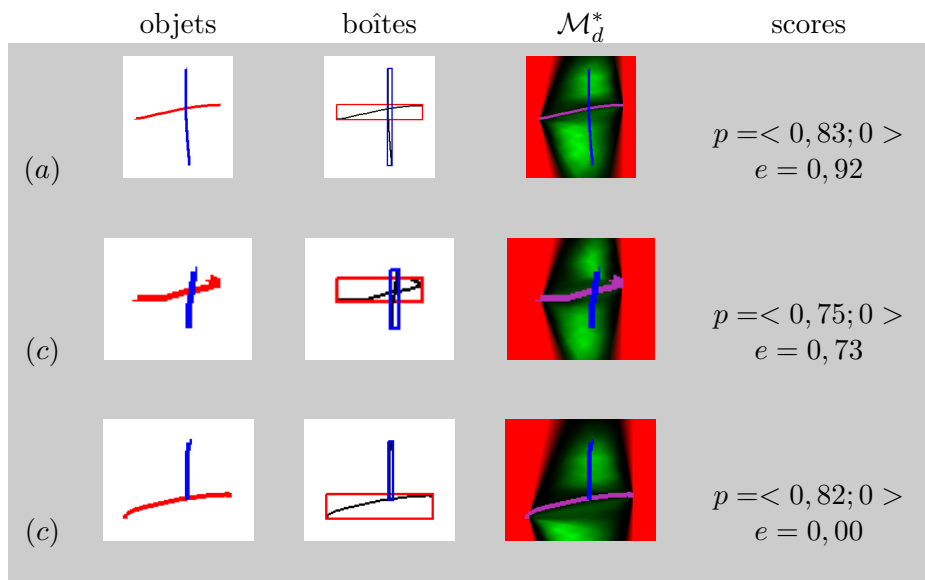


FIGURE 3.32: Modèle spatial bipolaire appris pour des primitives présentant une intersection (\mathcal{M}_d^*). Les scores de positionnement (p) et d'étendue (e) mesurés pour trois exemples d'objets sont indiqués. La relation apprise est issue d'une expression mathématique, il s'agit des deux tracés de l'opérateur « + ».

3.6. Résultats qualitatifs

La figure 3.32 représente une situation d'intersection entre des objets extraits d'expressions mathématiques (il s'agit du tracé de l'opérateur « + »). Les deux premiers exemples sont des tracés du symbole complet, qui présentent une intersection. Dans le troisième exemple, le tracé du symbole est incomplet et il n'y a pas d'intersection. Les modèles spatiaux bipolaires avec distance sont représentés et les scores de positionnement et d'étendue des objets sont donnés dans la dernière colonne. Le rôle de la modélisation de l'étendue est ici mis en avant (elle est décrite par la représentation par points particuliers, selon la méthode de la section 3.5.3). Les deux objets arguments de (a) et (b) ont un bon score de positionnement car ils sont situés dans la zone de préférence positive du modèle bipolaire. Ils ont également un bon score d'étendue car ils s'étendent chacun de part et d'autre du tracé de référence, ce qui est bien la situation attendue par le modèle d'étendue. L'objet argument de la paire (c) a également un bon score de positionnement, car il est aussi situé dans une zone de préférence positive. En revanche, il est nettement pénalisé par un score d'étendue nul, car il ne s'étend que d'un seul côté de la référence. La paire d'objets (c) est donc jugée comme ne respectant pas la relation spatiale décrite conjointement par le modèle de positionnement et le modèle d'étendue.

3.6.2 Modélisation de relations entre objets manuscrits complexes

Nous poursuivons le passage en revue des capacités d'expression des méta-modèles spatiaux en considérant le positionnement relatif d'objets plus complexes, souvent constitués eux-mêmes de plusieurs tracés. Nous avons vu dans le chapitre 2 que les descripteurs de positionnement les plus utilisés pour ces situations reposent sur l'approximation des objets au moyen de leurs boîtes englobantes. Nous distinguons à nouveau ici les relations spatiales entre des objets distants (voir la figure 3.33) et les relations entre objets qui présentent un recouvrement de leur boîtes englobantes (voir la figure 3.34).

Les relations spatiales entre objets distants sont illustrées à la figure 3.33 par des relations issues d'expressions mathématiques. Il s'agit la relation de positionnement entre une barre de fraction horizontale et son dénominateur ((a) et (b)) et de la relation de positionnement entre une expression et son exposant ((c) et (d)).

Dans le premier cas, la forme de l'objet de référence est très stable (une barre de fraction horizontale), tandis que l'objet argument peut prendre des formes et des niveaux de complexité très divers (d'un simple chiffre à une expression complexe). Le résultat de l'apprentissage est représenté par deux modèles spatiaux bipolaires sans distance (\mathcal{M}^*) et avec distance directionnelle (\mathcal{M}_d^*). Les modèles spatiaux appris décrivent des zones de préférence positives au milieu et en dessous de la barre de fraction. Des zones neutres s'étendent sur les côtés, toujours en-dessous de la référence. Enfin, les zones au-dessus et nettement à droite ou à gauche de la référence sont complètement exclues.

Dans le second cas, les formes et complexités des deux objets sont variables, comme le montrent les deux exemples (c) (forme multi-tracés en référence, chiffre isolé en argument) et (d) (lettre isolée en référence et expression composée en argument). Le même méta-modèle permet de bien appréhender ces différents cas de figures, comme en attestent les représentations des modèles spatiaux. L'intérêt de l'intégration de la distance dans les modèles semble assez discutable, en raison de la grande variabilité dans les dimensions des objets référence et arguments entre les cas de figure.

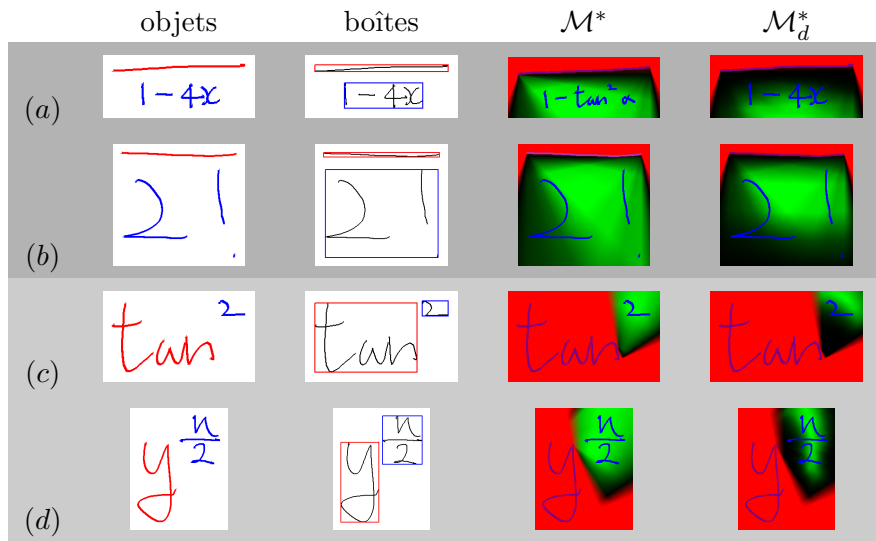


FIGURE 3.33: Modèles spatiaux bipolaires (sans distance \mathcal{M}^* et avec distance \mathcal{M}_d^*) appris pour des objets complexes distants. Les deux relations spatiales présentées sont issues d'expressions mathématiques : position du dénominateur par rapport à la barre de fraction ((a),(b)) et position de l'exposant par rapport à une expression ((c),(d)).

La figure 3.34 décrit deux relations spatiales entre des objets complexes qui présentent un important recouvrement de leurs boîtes englobantes. Il s'agit de la relation *racine carrée* issue d'expressions mathématiques et d'une relation de structure de caractères chinois correspondant au modèle de structure *haut-droite* qui a été présenté à la figure 2.12).

Deux exemples de paires d'objets sont donnés pour chaque relation. Les modèles spatiaux représentés sont les versions bipolaires sans distance (\mathcal{M}^*) et avec distance *directionnelle* (\mathcal{M}_d^*). Ils permettent de vérifier le bon comportement de l'apprentissage et de l'exploitation des méta-modèles dans cette situation où les objets mis en relation sont plus complexes (multi-tracés) et variables (par exemple l'expression sous la racine peut être un symbole chiffre ou une expression composée). L'apport de l'intégration de distance semble bénéfique car elle permet de mieux affiner les zones de préférences positives. Par exemple, dans le cas de la racine carrée (a), l'information de distance permet de mieux restreindre la zone de préférences positives à la *ligne de base* de l'expression, plutôt qu'à toute la zone située sous la racine.

3.6.3 Problème de localité du contexte de positionnement

Une caractéristique importante des méta-modèles spatiaux est la prise en compte de la forme du tracé de l'objet de référence pour décrire les relations de positionnement. Cette propriété constitue un avantage pour les différentes situations que nous avons présentées jusqu'ici, autant pour la description du positionnement d'objets de tracés simples que d'objets plus complexes. Elle permet également d'assurer une modélisation fidèle à l'intuition de relations d'enchevêtrement des objets ou de recouvrement de leurs boîtes englobantes.

Toutefois, une conséquence de la prise en compte de la forme de la référence peut être

3.6. Résultats qualitatifs

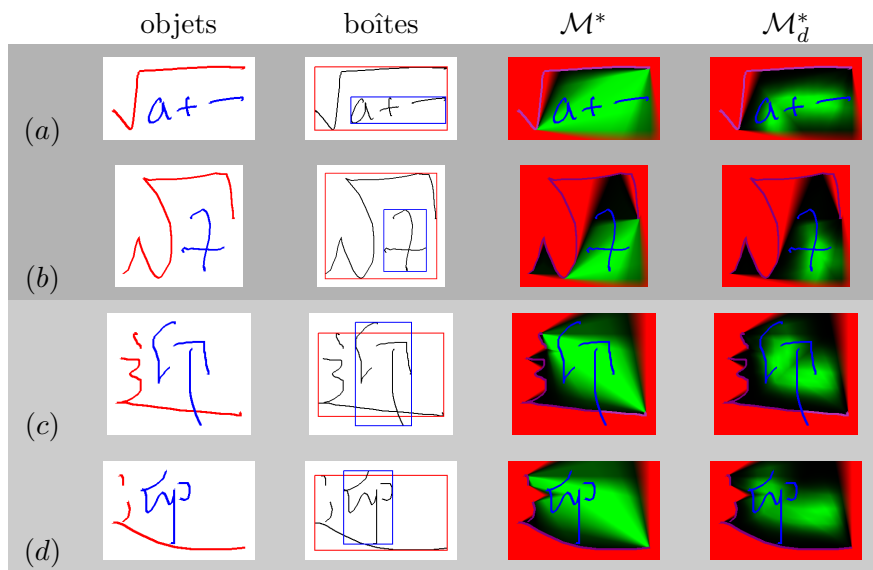


FIGURE 3.34: Modèles spatiaux bipolaires (sans distance \mathcal{M}^* et avec distance \mathcal{M}_d^*) appris pour des objets complexes qui présentent un recouvrement de leur boîtes englobantes. Les deux relations spatiales présentées, sont issues d'expressions mathématiques (racine carrée) ((a),(b)) et de l'écriture du chinois ((c),(d)).

une impossibilité à distinguer des relations de positionnement différentes par rapport à une référence qui présente des *contextes localement similaires*. Ce phénomène peut notamment se produire dans le cas d'objets arguments situés dans une concavité d'un objet de référence qui en comporte plusieurs. Par exemple, la figure 3.35 présente deux cas de figure où le modèle de positionnement appris à partir d'exemples au positionnement stable est incapable de restituer une zone unique de positionnement acceptable pour deux relations spatiales.

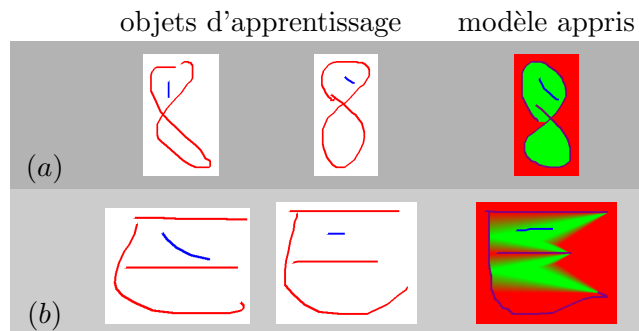


FIGURE 3.35: Dans le cas où la référence comporte plusieurs concavités localement similaires, le modèle spatial appris ne peut pas les différencier. On peut le vérifier sur les deux boucles d'un « 8 » (a) : les objets d'apprentissage sont situés dans la boucle supérieure, mais le modèle appris admet les deux boucles comme des zones acceptables. Sur l'exemple du « E » (b), on constate le même phénomène avec les concavités ouvertes vers la droite.

Dans le cas de la première relation (a), un tracé est positionné dans la boucle supérieure d'un « 8 » pour les exemples d'apprentissage. Le modèle spatial représenté, pourtant, admet de façon équivalente les deux boucles d'un « 8 » comme étant des zones de positionnement acceptables selon cette relation. Le modèle n'est en effet défini que par rapport à la forme locale de la référence et toute zone *encerclée* par le tracé de la référence est vue comme admissible. Il s'agit d'un modèle pour une relation topologique qui peut être exprimée par *l'objet est à l'intérieur de la référence*. Le cas du « E » (b) illustre le même phénomène, avec cette fois un contexte local de concavité qui se répète dans la référence. En l'état, le modèle spatial appris ne peut pas distinguer les concavités supérieure et inférieure l'une de l'autre.

Cette incapacité du modèle à différencier des contextes locaux similaires peut être un avantage, dans le cas où l'on veut modéliser la relation topologique *à l'intérieur de* quelle que soit la forme de l'objet de référence : un carré, un cercle, un « 8 »... Cependant, cette limitation de la description doit aussi pouvoir être contournée pour permettre l'apprentissage de relations telles que *dans la concavité supérieure du « E »*, par exemple. En pratique, des situations de confusions apparaîtront lors de la description de relations entre une référence complexe (comportant des contextes locaux similaires) et un objet argument positionné dans une concavité de la référence. Pour faire face à ce type de problème, la solution que nous adopterons consistera à décomposer davantage la référence en plusieurs éléments, plus simples et offrant une meilleure capacité de description. Par exemple, dans le cas (b), le positionnement de l'objet bleu peut être simplement décrit par la combinaison de deux méta-modèles spatiaux : l'un exprimant la relation *au-dessus de la barre horizontale au milieu du « E »*, l'autre la relation *à droite de la barre verticale du « E »*.

La combinaison de description de positionnement spatial depuis plusieurs références est rendue possible par les modèles spatiaux, de la même façon que nous avons déjà combiné plusieurs *points de vue* directionnels. Un gain en précision du positionnement relatif peut être apporté par la combinaison de modèles supportés par plusieurs références simples, plutôt que par un modèle unique sur une référence complexe. Nous ferons usage de ce principe de combinaison des références au chapitre 4 pour décrire le positionnement relatif de primitives au sein de symboles structurés.

3.7 Synthèse

Dans ce chapitre, nous avons introduit une nouvelle formulation pour représenter des relations spatiales entre des objets manuscrits. Le principe de raisonnement par modèles spatiaux a été exploité plus avant pour définir des *méta-modèles* qui peuvent être appris automatiquement à partir de données. Par leur exploitation sous la forme de modèles spatiaux, ils permettent de profiter de tous les avantages offerts par le raisonnement spatial au sein de l'espace image que nous avons mis en évidence dans le chapitre 3. Ils sont donc capables de prendre en compte parfaitement les formes des objets. Ils sont dotés de la capacité de *prédire* le positionnement d'un objet par rapport à une référence et ils offrent un cadre naturel et intuitif pour considérer les situations sans simplifier ou approximer ni les formes ni le positionnement relatif des objets.

Les choix que nous avons présentés tout au long de ce chapitre de formalisation ont été guidés par l'objectif de généralité vis-à-vis des besoins de description de positionnement

3.7. Synthèse

dans le cadre du traitement automatique de documents manuscrits. Il a notamment été porté une grande attention à garantir la même expressivité quelles que soient les directions mises en jeu dans les relations. L'information de distance a également fait l'objet d'une attention particulière et deux variantes ont été proposées pour l'intégrer dans les méta-modèles. Pour pousser plus loin le cadre de raisonnement en gérant l'imprécision sur le positionnement relatif des objets, il a été proposé d'étendre la modélisation spatiale par ensembles flous à une modélisation floue *bipolaire*, où les préférences positives et négatives sur le positionnement des objets sont exprimées sur deux échelles distinctes. Enfin, nous avons proposé de coupler à la mesure du *positionnement* une mesure de *l'étendue* des objets dans l'espace des degrés directionnels exprimés relativement à la référence. Un modèle d'étendue apporte un complément au modèle de positionnement, qui permet de décrire des contraintes sur *l'occupation de l'espace* par l'objet cible (indépendamment de son bon ou de son mauvais positionnement).

Des résultats qualitatifs ont été présentés afin de démontrer la capacité de la formalisation à traiter des données de nature et complexité variable. Un large spectre de relations spatiales a été couvert par ces exemples, démontrant la bonne généralité du formalisme pour des situations d'objets distants, connectés, s'intersectant... La bonne capacité d'application des méta-modèles en des modèles spatiaux adaptés aux formes des objets de référence a aussi pu être constatée. Dans le chapitre 5, des expérimentations seront présentées afin de comparer la qualité de description des relations spatiales offertes par les méta-modèles spatiaux en comparaison avec d'autres techniques de description de positionnement par des descripteurs. Nous montrerons qu'ils permettent de faciliter la reconnaissance automatique de relations de positionnement avec des objets de différentes natures (gestes d'édition, primitives de caractères chinois, lettres manuscrites, symboles mathématiques).

Dans le chapitre 4, nous allons développer plus en profondeur notre démarche de focalisation des efforts de modélisation sur les aspects spatiaux des compositions manuscrites. Nous allons exploiter la richesse de description permise par les méta-modèles en les mettant au service de la représentation d'objets composés, où se posent les problèmes de segmentation et d'identification d'éléments de référence pour le positionnement. Dans le contexte du traitement de symboles manuscrits structurés, nous mettrons à profit la précision et la généralité des méta-modèles pour souligner l'importance de la modélisation du positionnement spatial. Cette utilisation des méta-modèles illustrera également la contribution qu'ils peuvent apporter pour l'analyse d'expressions manuscrites.

Représentation de formes structurées par *Patrons Structurels Flous*

Nous avons introduit dans le chapitre 3 le concept de méta-modèle spatial, qui constitue une modélisation nouvelle pour les relations spatiales entre des *objets manuscrits*, en visant à une large généralité d'expression des relations et une applicabilité à des natures d'objets variables. Ces méta-modèles spatiaux ont été développés avec la volonté de capturer au mieux la richesse du positionnement relatif de deux objets manuscrits.

La seconde contribution de nos travaux, présentée dans ce chapitre, consiste à prolonger cette démarche à un niveau plus avancé : nous cherchons à représenter au mieux les aspects spatiaux qui caractérisent des compositions manuscrites structurées. Nous introduisons la notion de *Patron Structurel Flou*, une représentation structurelle qui combine plusieurs méta-modèles de positionnement pour former un *patron spatial* des formes manuscrites composites. L'objectif poursuivi est double. Il s'agit tout d'abord de mettre en exergue le rôle majeur joué par l'information de structure et de positionnement relatif afin de souligner l'intérêt qui doit lui être consacré dans les méthodes de reconnaissance. Ensuite, nous souhaitons démontrer l'avantage qu'offrent les méta-modèles spatiaux pour décrire et modéliser cette information avec une plus grande précision et un spectre d'exploitation plus large que les descripteurs traditionnellement utilisés. Contrairement à ces derniers, les méta-modèles spatiaux sont à même de mettre en lumière le rôle fondamental de l'aspect spatial car ils ne procèdent pas à des approximations simplificatrices des formes des objets ni de leurs positionnements. De plus, à travers leur capacité unique de *prédiction*, ils permettent d'accorder la priorité à l'information de positionnement pour guider la tâche de *segmentation* dans le but d'identifier les composants de la structure.

Dans la première partie de ce chapitre, nous présentons les enjeux et la démarche adoptée pour représenter des symboles manuscrits structurés. Nous introduisons ensuite les Patrons Structurels Flous, une représentation originale pour ces symboles qui se concentre sur leurs aspects spatiaux, en s'appuyant sur des méta-modèles. Le chapitre est conclu par une validation qualitative de la représentation, qui confirme la pertinence de la démarche et atteste

donc de l'importance effective de l'information spatiale dans ce contexte. Cette validation confirmera également la précision des méta-modèles de positionnement et mettra en évidence le potentiel de leur contribution à l'analyse du manuscrit en général.

4.1 Démarche proposée

Pour la représentation de symboles manuscrits isolés en-ligne, de nombreuses méthodes existantes décrivent la structure de symboles manuscrits, en particulier au moyen de graphes ou à l'aide de règles grammaticales [Wen03]. Dans ces approches, les relations spatiales entre les primitives structurelles sont modélisées explicitement, le plus souvent avec des descripteurs tels que ceux rencontrés dans la première partie de l'état de l'art. Ces approches accordent pourtant souvent une importance prioritaire aux formes des tracés des primitives plutôt qu'à l'information spatiale, ne serait-ce que parce que l'étape de segmentation est guidée par la recherche de formes élémentaires (segments, arcs de cercle).

Cet ordre de priorité nous semble discutable et signale selon nous une sous-exploitation de la richesse d'information que comporte pourtant l'aspect spatial pour ces symboles. Le principe que nous défendons dans ce chapitre est au contraire de représenter les symboles exclusivement par la modélisation fine des relations spatiales entre les éléments qui les constituent. Notre démarche consiste volontairement à se priver de l'information de la forme même des éléments qui constituent le symbole afin de mettre en évidence toute la force de l'information spatiale et ainsi de convaincre de l'intérêt de la modéliser avec une grande précision.

4.1.1 Stabilité et importance de l'information spatiale

Le choix de concentrer l'effort de modélisation sur l'information de structure et les relations spatiales entre les éléments est motivé par l'analyse des propriétés de différentes catégories de symboles que nous avons étudiées : les caractères chinois, les dessins de pictogrammes et les symboles architecturaux. Des exemples des trois types de symboles sont donnés dans la figure 4.1.

Caractères chinois Nous considérons un petit nombre de caractères chinois, qui sont assimilés dans notre étude à une catégorie de symboles (voir les éléments (a) à (f) de la figure 4.1). Ce sont des symboles très structurés, composés théoriquement à partir de tracés naturels appartenant à un jeu de formes primitives prédéfini. En pratique, le grand nombre de tracés par caractère et la structure bidimensionnelle complexe incitent les scripteurs à lier entre eux plusieurs de ces tracés théoriquement distincts. Les différences de styles de tracé conduisent à plusieurs variantes pour le même caractère. On constate cependant une forte stabilité de certaines caractéristiques : les éléments sont tracés plutôt du haut vers le bas et de gauche à droite, de sorte que les distorsions de segmentation sont assez reproductibles entre scripteurs. Les formes individuelles de certains tracés peuvent varier beaucoup, notamment pour les tracés les plus petits. Les plus grands tracés, souvent verticaux ou horizontaux, sont relativement stables.

Pictogrammes Ce type de symboles correspond à de petits dessins visuellement simples, représentant des pictogrammes utilisés par exemple pour signaler un danger ou donner une information (une flamme, une voiture, ...). Des exemples de ce type de symboles

4.1. Démarche proposée

sont présentés par la figure 4.1 (éléments (g) à (j)). Ces symboles sont choisis pour être communément reconnus par tout le monde, mais il n'existe paradoxalement pas de façon standard de les dessiner. On observe dans ce cas de grandes variations dans toutes les caractéristiques de tracés en fonction des scripteurs. Il y a notamment une très grande variété dans le nombre de tracés (de 1 à 8 tracés pour le même symbole parfois). Certaines parties des dessins peuvent être « pleines », on peut trouver des pointillés, et l'ordre des tracés est très inconstant.

Symboles architecturaux Ce troisième type de symboles présente des caractéristiques différentes des deux précédents. Il s'agit ici de symboles constitués uniquement de segments et d'arcs de cercle (voir les éléments (k) à (n) de la figure 4.1). Les symboles sont composés de l'assemblage de polygones qui se chevauchent et s'intersectent souvent. Dans ce cas, il n'existe pas non plus de méthode de dessin standard et on observe de grandes variations dans les stratégies de tracés d'un scripteur à l'autre. Par exemple, un rectangle peut être fait d'un seul tracé naturel (sans lever de stylo), comme de deux, trois ou quatre tracés, voire davantage car il n'est pas rare de relever des cas de rectangles imparfaits qui ont été retouchés plusieurs fois.

Au regard de ces trois différentes catégories de symboles, il apparaît que l'information de structure et de positionnement relatif entre les éléments est plus stable et moins sujette à distorsions que peut l'être la forme même de ces éléments. Ce constat que nous faisons ici dans le contexte du traitement du manuscrit est l'un des fondements des méthodes de *raisonnement spatial* qui s'appliquent à de nombreux domaines, comme ceux de l'interprétation d'images médicales ou d'images satellites. Par exemple, dans l'analyse d'images médicales, il est établi que les relations spatiales jouent un rôle crucial car elles sont moins sujettes à variabilité que les propriétés intrinsèques des éléments de la structure (forme, luminosité, ou texture par exemple) [AHF⁺07]. La représentation que nous proposons pour ces symboles repose sur ce même constat transposé aux objets manuscrits.

4.1.2 Représentation de symboles par *Patron Structurel Flou* (PSF)

Un Patron Structurel Flou (ou PSF, en abrégé) est un modèle de symbole manuscrit qui le décrit comme un assemblage d'entités positionnées les unes par rapport aux autres par des méta-modèles spatiaux flous. Il constitue une modélisation purement structurelle et spatiale du symbole : il n'est imposé aucune contrainte sur la forme des tracés. En particulier, les tracés ne sont pas classifiés en différentes catégories prédéfinies de formes et aucun modèle statistique de leur forme n'est utilisé.

Ce choix serait certainement pénalisant dans l'optique de construire une méthode de *reconnaissance* de symboles, même si on peut constater que la forme de plusieurs tracés semble peu significative pour certains symboles (par exemple les petits tracés dans le caractère chinois de l'exemple (b), ou le point dans le pictogramme de l'exemple (j), voir la figure 4.1).

Il permet en revanche de bien mettre en évidence le pouvoir intrinsèque de représentation que porte l'information de structure et de positionnement relatif des éléments du symbole. La méthode est intégralement formalisée dans un contexte de raisonnement bipolaire flou et un score bipolaire pourra être attribué à la correspondance entre un symbole et un PSF. Un algorithme d'apprentissage sera proposé pour construire des PSF automatiquement à



FIGURE 4.1: Exemples de symboles structurés manuscrits : caractères chinois (a-f), pictogrammes (g-j), symboles architecturaux (k-n).

4.1. Démarche proposée

partir de données. La qualité de la représentation offerte par les PSF ainsi construits pourra ensuite être évaluée en mesurant leur capacité à répondre à une tâche de reconnaissance de symboles par une démarche de classification très simple : la classe attribuée à un symbole à reconnaître sera celle dont le prototype PSF représente au mieux ce symbole.

Par rapport à la description de relations spatiales entre deux objets, la modélisation de structures comporte des enjeux supplémentaires : il faut découvrir quels sont les éléments stables de la structure, en particulier pour définir des objets de référence. Grâce à la capacité de prédiction spatiale dont disposent les méta-modèles spatiaux, nous proposons une stratégie originale pour résoudre le problème de segmentation qui consiste à identifier les différents éléments de la structure.

4.1.3 Segmentation guidée par les modèles spatiaux

La capacité des méta-modèles spatiaux à répondre conjointement aux problèmes d'évaluation de relation spatiale et de localisation d'objets permet d'envisager des approches originales pour résoudre le problème de segmentation des tracés.

Le problème de segmentation se pose dans toutes les situations d'analyse structurelle d'expressions. Dans notre contexte, la segmentation consiste à identifier des primitives structurelles parmi les tracés qui constituent un symbole manuscrit. Dans le traitement d'objets *en-ligne*, des tracés *naturels* sont directement extraits du signal capturé par le dispositif de saisie. Un tracé *naturel* correspond à la portion de tracé réalisée d'une seule traite, sans lever de stylet. Comme nous l'avons montré dans les divers exemples illustrant les trois types de symboles, la composition à la main de symboles complexes peut donner lieu à de multiples variantes dans l'ordre et le nombre de leurs tracés naturels. Pour faire face à cette variabilité, la tâche de segmentation doit procéder au regroupement de tracés ou bien à la division d'un tracé en plusieurs éléments, afin de se ramener à une description canonique de la structure.

Le problème de la recherche de segmentation pose un défi bien connu, énoncé sous le nom de « paradoxe de Sayre » : *la segmentation est nécessaire pour la reconnaissance des entités, mais la reconnaissance est nécessaire pour la bonne segmentation des entités* [Say73]. En pratique, ce paradoxe est contourné en formulant un certain nombre d'hypothèses de segmentation. On peut par exemple identifier des points de tracé qui semblent correspondre à des points de rupture possible à partir de l'analyse de la forme locale du tracé (point de forte courbure, point de rebroussement, point multiple...). Des modèles de positionnement relatif peuvent être utilisés pour évaluer le positionnement des segments extraits, rejeter de mauvais candidats, ou participer à la recherche de la meilleure hypothèse globale. Les problèmes de segmentation et d'analyse des relations spatiales sont fortement interdépendants.

Nous proposons de s'appuyer directement sur le positionnement relatif des tracés, plutôt que sur leur forme locale, pour proposer des hypothèses de segmentation. Dans l'approche par Patron Structurel Flou, nous mettons en place une stratégie de segmentation guidée par le positionnement relatif des éléments. Cette stratégie est rendue possible par la *capacité de prédiction* des méta-modèles spatiaux : étant donné un objet de référence, le modèle définit dans quelle zone du plan doit se situer un élément. Ces zones de préférence, qui modélisent une position attendue pour un objet de la structure, peuvent servir à guider la recherche de points de segmentation. Un tel procédé de segmentation a l'avantage d'être guidé par des

modèles qui ont été appris à partir de données et qui s'adaptent à la forme des tracés de la référence.

4.2 *Patrons Structurels Flous* pour la représentation de symboles

L'objectif poursuivi par notre démarche est d'évaluer dans quelle mesure la représentation de symboles manuscrits au seul moyen de l'information sur leur structure interne décrite par des modèles de positionnement est porteuse de sens. Cette évaluation pourra être effectuée en cherchant à reconnaître automatiquement un symbole par la mesure de son score d'adéquation avec un PSF (qui constitue un modèle de caractère). Ce score ne repose que sur l'adéquation conjointe des éléments de la structure du symbole à reconnaître avec les méta-modèles spatiaux qui décrivent leur positionnement attendu ainsi que leur étendue. Le cadre théorique de la logique floue sera utilisé pour fusionner les scores d'adéquation provenant des différents éléments du modèle et pour conduire le mécanisme de prise de décision. Plus précisément, les scores bipolaires flous de positionnement et d'étendue issus des méta-modèles spatiaux seront intégrés dans un raisonnement flou bipolaire complet. Nous formalisons ce raisonnement dans la partie 4.2.2, puis présentons le procédé d'exploitation de la représentation par PSF pour le calcul d'un score de correspondance (partie 4.2.3). Nous proposons ensuite un mécanisme d'apprentissage supervisé permettant de construire automatiquement à partir de données les PSF, sans aucune connaissance a priori sur la segmentation interne des symboles en primitives structurelles (partie 4.2.4).

Avant tout cela, nous nous intéressons au problème du choix du référentiel dans lequel sont décrits les modèles spatiaux de positionnement qui composent un PSF.

4.2.1 Définition du référentiel de positionnement

Tel que nous l'avons défini de façon informelle, un PSF est un modèle structurel reposant sur un ensemble de méta-modèles spatiaux qui décrivent l'organisation relative des composants de la structure. Nous avons introduit en détail dans le chapitre 3 les principes de construction et d'exploitation des méta-modèles spatiaux qui permettent de modéliser le positionnement relatif d'un objet par rapport à un autre. Pour l'apprentissage comme pour l'exploitation d'un méta-modèle spatial, nous avons jusqu'ici toujours considéré que la définition de l'objet de référence était connue et qu'elle n'était pas remise en cause. Nous nous intéressons ici à la définition du référentiel de description du positionnement des éléments de la structure des symboles. Différents choix sont à faire pour définir sur quel(s) objet(s) de référence les méta-modèles spatiaux doivent être construits et appliqués au sein du PSF.

4.2.1.1 Référentiel concret

Dans le cas des PSF, qui ont vocation à représenter la structure de n'importe quel type de symbole, il faut définir une stratégie pour déterminer dans quel référentiel va être décrit le positionnement des éléments de la structure. La qualité de description permise par les méta-modèles spatiaux provient en premier lieu de leur adaptation aux formes des tracés manuscrits. Comme nous l'avons déjà mis en évidence, cette propriété permet de ne pas sacrifier la forme des objets dans la description de leur positionnement relatif et de prendre

4.2. *Patrons Structurels Flous* pour la représentation de symboles

en compte les particularités du tracé manuscrit. De plus, un modèle de positionnement sera d'autant plus précis et informatif que la forme de la référence est riche et singulière. Pour ces raisons, il apparaît que le choix du référentiel de positionnement au sein d'un PSF doit nécessairement se porter vers des portions du tracé réel saisi par le scripteur. Les objets définissant la référence de positionnement doivent être des éléments du symbole manuscrit : on parle de référence *concrète*, par opposition à un référentiel *virtuel* qui pourrait être défini par des éléments tels que la boîte englobante du tracé ou encore des axes définissant un repère sur le centre de gravité du symbole.

Nous avons montré en fin de chapitre 3 que les méta-modèles spatiaux permettent de décrire un large spectre de catégories de relations spatiales. En particulier, le positionnement relatif de tracés simples peut être précisément décrit, y compris dans des situations d'intersection ou d'adjacence des deux objets, lorsqu'ils sont distants ou encore quand l'un est entouré par l'autre... Cette capacité permet donc de considérer que les références sont partie prenante du symbole, et ce quelle que soit la complexité de la structure ou l'intrication de ses éléments manuscrits.

Enfin, contrairement à certains descripteurs spatiaux qui ne peuvent être exploités que sur des tracés rectilignes ou en arc de cercle, la capacité des méta-modèles spatiaux à s'adapter à toute forme d'objet permet de ne pas imposer de contrainte de forme, de taille ou encore de complexité graphique dans le choix des références. Cela autorise donc une parfaite généralité dans les possibilités de symboles décrits.

4.2.1.2 Choix d'un référentiel global

La question du nombre et de la sélection des éléments manuscrits de référence, ainsi que du nombre de modèles spatiaux, se pose également. Il s'agit de décider si le positionnement des éléments de structure doit être décrit pour tous par rapport à une référence unique (globale), ou bien par rapport à tous les autres éléments, ou encore par rapport à certains éléments sélectionnés (par exemple les plus proches, les plus grands, les plus stables...).

Il convient de limiter la complexité de la représentation tout en garantissant une bonne description du positionnement de tous les éléments de la structure. La solution la plus complète, qui consiste à décrire le positionnement de tous les éléments par rapport à tous les autres, est écartée d'emblée car elle conduit à une complexité déraisonnable pour l'exploitation des PSF et surtout à une représentation inadaptée de la structure. En effet, le nombre d'éléments pouvant se chiffrer à une dizaine (voire plus), l'explosion combinatoire engendrée par la recherche du meilleur appariement entre une structure à reconnaître et un PSF devient difficile à gérer. Surtout, la description complète de toutes les relations spatiales entre les éléments pris deux à deux n'est en général pas pertinente. En effet, il existe souvent des éléments prégnants, qui exercent un rôle dominant pour la perception de la structure. Les autres éléments sont alors positionnés avant tout par rapport à ces éléments prégnants, plutôt que par rapport à tous les autres.

Le caractère chinois (e) de la figure 4.1 comporte deux éléments prégnants particulièrement évidents dans sa structure : un long tracé vertical qui sépare le caractère en deux moitiés, gauche et droite, et un long tracé horizontal qui le coupe en deux dans sa hauteur. Ces deux éléments définissent en quelque sorte un repère de positionnement pour les autres éléments de la structure.

Dans notre approche, nous cherchons à identifier ces éléments prégnants et les considérons comme les références globales pour le positionnement des autres éléments. Le choix intuitif d'appuyer la description des relations spatiales sur les éléments prégnants rejoint l'une des règles formulée dans le modèle psychologique de Clark [CPM73] pour la perception des relations spatiales. Cette règle est citée par Freeman dans ses travaux [Fre75]. Elle stipule simplement que « *si un objet dans une image est perceptuellement proéminent, il sera utilisé comme référence pour toute relation de positionnement avec un autre objet* ». Sous l'angle du problème de reconnaissance de formes, les objets saillants d'une structure offrent l'avantage d'être faciles à identifier parmi un ensemble de tracés manuscrits car ils sont généralement de grande dimension par rapport aux autres éléments. De plus, ils présentent des caractéristiques relativement stables. Par définition, leur forme, leur orientation et leur position jouent un rôle fortement structurant pour la perception du symbole. Il est donc naturel de constater une certaine stabilité de ces éléments entre différentes variantes manuscrites d'un même symbole.

4.2.1.3 Optimisation du pouvoir descriptif du référentiel

Puisque le champ d'application visé concerne des symboles de nature bidimensionnelle, il est nécessaire que le choix du référentiel de positionnement puisse couvrir efficacement le positionnement d'éléments dans les deux dimensions. Pour cela, nous proposons de faire reposer l'expression des modèles de positionnement spatial sur un couple de références, l'une notée r_x et dite *horizontale*, l'autre notée r_y et dite *verticale*. Chaque référence est une portion de tracé manuscrit, avec r_x (respectivement r_y) qui est une portion de tracé supposément horizontale (respectivement verticale). Les notions d'horizontalité et de verticalité dans la recherche des références r_x et r_y sont souples et n'imposent aucune restriction. Elles se traduisent simplement par l'intégration d'un critère d'horizontalité dans la recherche de la meilleure référence r_x et de verticalité dans la recherche de r_y . En pratique, si un symbole n'est composé que de tracés horizontaux, la référence r_y sera fatalement une portion de tracé horizontale. Néanmoins, dans le cas général, la recherche de l'orthogonalité des références permet de garantir une bonne description du positionnement de tout élément. En effet, la position d'un élément donné pourra ainsi être décrite conjointement par deux méta-modèles spatiaux liés à deux références offrant un angle de vue différent sur l'élément.

Pour toute portion S extraite du tracé d'un symbole, on peut calculer un score exprimant la qualité de la description offerte par ce tracé s'il est utilisé comme référence *horizontale* ou *verticale* pour le positionnement des autres éléments du symbole. L'idée est simplement d'exprimer que S est une bonne hypothèse de référence *horizontale* si tout autre point de tracé du symbole peut être décrit comme étant situé soit au-dessus, soit en-dessous de S . Les notions *en-dessous* et *au-dessus* sont décrites à l'aide d'opérateurs morphologiques flous de directions classiques (selon la représentation admise précédemment pour les descriptions de relations directionnelles par rapport à une référence, voir par exemple à la section 2.2.3.2).

Ainsi, le score du choix de S comme référence horizontale dépend de la mesure suivante :

$$\sum_{p \in \mathcal{S} \setminus S} \max(\mu_{haut}(S)(p), (\mu_{bas}(S)(p)), \quad (4.1)$$

où \mathcal{S} est l'ensemble des points du tracé. Le tracé S est vu comme l'ensemble de ses points ; $\mu_{haut}(S)$ et $\mu_{bas}(S)$ sont les paysages flous directionnels définis par les éléments structurants

4.2. *Patrons Structurels Flous* pour la représentation de symboles

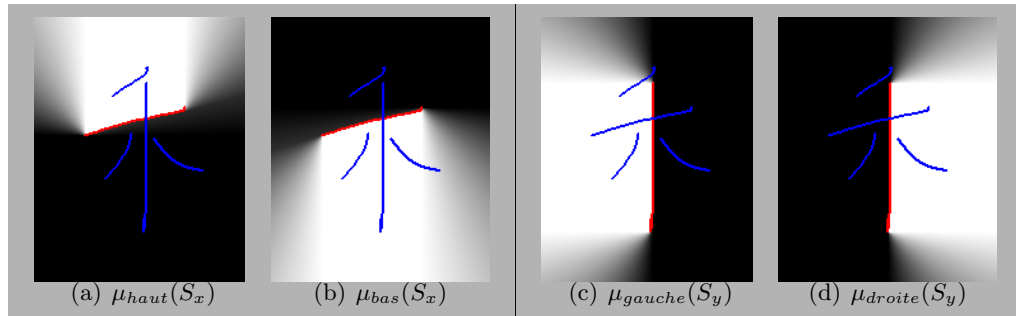


FIGURE 4.2: Evaluation de la qualité d'une portion de tracé S_x prise comme référence horizontale (a,b) et S_y comme référence verticale (c,d)

de direction *haut* et *bas* appliqués sur la référence S . Une définition similaire reposant sur les directions à *gauche* et à *droite* est utilisée pour estimer la qualité de description offerte par le choix de S comme référence *verticale*. La fonction de score ainsi définie permet d'optimiser la qualité de la description ultérieure du positionnement, en sélectionnant les références permettant de décrire facilement le positionnement de tout autre point du tracé. La figure 4.2 représente les deux modélisations directionnelles morphologiques *haut* et *bas* pour un choix de référence horizontale S_x dans le premier caractère de l'exemple (e) de la figure 4.1. De même, les directions à *gauche* et à *droite* sont appliquées sur un tracé S_y pris comme hypothèse de référence verticale. Ces deux éléments de tracés constituent les meilleures références au sens du score de l'équation 4.1 car ils permettent de « couvrir » presque parfaitement les autres tracés par ces quatre directions.

Les deux références sont sélectionnées indépendamment l'une de l'autre, ce qui peut conduire au choix de la même portion de tracé pour les deux références. Cela se produirait par exemple dans le cas du symbole « % », pour lequel il est bien naturel de sélectionner le segment diagonal comme unique référence pour décrire le positionnement des deux petits cercles.

4.2.1.4 Positionnement absolu du référentiel par *Référence Virtuelle Absolue* (RVA)

Comme expliqué au paragraphe 4.2.1.1, il est crucial pour la qualité et la finesse de la description offerte par les méta-modèles spatiaux de choisir des objets manuscrits comme références de positionnement plutôt que des éléments virtuels. Cependant, le choix de références manuscrites pose le problème de la segmentation et de l'extraction des bonnes portions de tracé parmi le dessin du symbole complet. Cette tâche est difficile et ne peut pas être résolue de façon définitive a priori : il sera donc nécessaire d'extraire plusieurs hypothèses de références horizontales et verticales à partir d'un tracé à reconnaître. Il est indispensable d'extraire les « bonnes » références pour exploiter correctement un PSF, puisque tous les modèles de positionnement relatif reposent sur elles. Afin de faciliter la recherche des tracés de référence, nous proposons d'introduire dans la modélisation une part de positionnement par rapport à un référentiel absolu et virtuel (défini par des objets non concrets) qui offre



FIGURE 4.3: La figure (a) représente trois candidats (en rouge) pour constituer la référence horizontale de positionnement. Les figures (b),(c) et (d) démontre que ces références peuvent être distinguées en fonction de la position du bord gauche de la boîte englobante par rapport à elles-mêmes.

l'avantage d'être directement disponible : la boîte englobante du symbole. Le principe n'est pas de construire des modèles en s'appuyant sur ce référentiel virtuel, pour les raisons déjà évoquées au paragraphe 4.2.1.1, mais tout au contraire de décrire comment les quatre segments définissant la boîte englobante du tracé du symbole sont positionnés par rapport aux hypothèses de référence r_x et r_y . En d'autres mots, il s'agit de positionner des éléments du référentiel absolu par rapport aux références manuscrites choisies, plutôt que le contraire. De ce fait, on conserve le pouvoir d'expressivité des modèles spatiaux qui sont toujours développés sur des références concrètes, tout en permettant de stabiliser et d'identifier plus facilement ces références grâce aux éléments virtuels.

La figure 4.3 illustre sur un simple exemple l'intérêt de positionner les bords de la boîte englobante pour aider au choix d'une bonne référence de tracé. L'image (a) présente un caractère dans lequel trois éléments de tracés apparaissent comme de bons candidats pour constituer la référence horizontale (ils sont représentés en rouge). Les images (b),(c) et (d) montrent que la description de la position du bord gauche de la boîte englobante est très différente par rapport à ces trois références. Le bord gauche est *en bas à gauche* de la première hypothèse de référence (image (a)), *à gauche* de la seconde hypothèse (image (b)) et *en haut à gauche* de la troisième (image (c)). La description de la position des bords de la boîte englobante dans le modèle permettra donc de faciliter la recherche de ces bonnes références.

Dans la suite, nous désignons les quatre segments définissant les cotés de la boîte englobante du tracé sous le nom de Référence Virtuelle Absolue (ou RVA). Il est à noter que ces éléments virtuels seront décrits par des méta-modèles spatiaux en tout point comparables à ceux utilisés pour décrire le positionnement des composants du symbole par rapport aux références et seront intégrés dans la fusion d'informations de façon homogène.

4.2.2 Formalisation des PSF

Dans cette partie, nous présentons une définition des PSF et formalisons le calcul de leur correspondance avec un symbole tracé. Nous proposons de définir un PSF par un quadruplet $\langle a_y^x, a_x^y, (\mathcal{M}_i)_{i=1..n}, (\mathcal{N}_i)_{i=1..4} \rangle$, dont les éléments sont présentés en détails ci-dessous.

- a_y^x est un méta-modèle spatial décrivant le positionnement de la référence r_x par rapport à la référence r_y .
- Réciproquement, a_x^y est un méta-modèle spatial décrivant le positionnement de la référence r_y par rapport à la référence r_x .
- Chaque \mathcal{M}_i est un couple de deux méta-modèles spatiaux décrivant le positionnement d'un élément de la structure par rapport aux deux références r_x et r_y . Par conséquent,

4.2. Patrons Structurels Flous pour la représentation de symboles

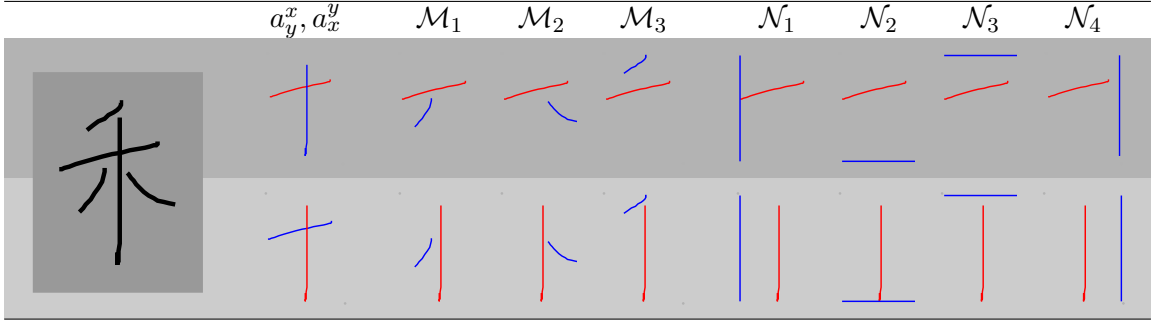


FIGURE 4.4: Illustration des différentes relations spatiales entre paires d'objets modélisées dans le PSF du symbole représenté à gauche. Dans chaque situation, la référence est en rouge et l'objet à positionner en bleu. Les images de la colonne a_y^x, a_x^y représentent le positionnement mutuel des références r_x et r_y . Les colonnes \mathcal{M}_i montrent le positionnement des trois composants du PSF par rapport à r_x (première ligne) et à r_y (seconde ligne). Les quatre dernières colonnes représentent le positionnement des éléments de RVA par rapport à r_x (première ligne) et à r_y (seconde ligne).

on peut noter $\mathcal{M}_i = (a_x^i, a_y^i)$ où a_x^i (respectivement a_y^i) est le méta-modèle spatial de positionnement du composant i par rapport à la référence r_x (respectivement r_y). n désigne le nombre d'éléments considérés dans la représentation structurelle, c'est-à-dire le nombre de *composants* dans le symbole, en plus des deux références r_x et r_y . (\mathcal{M}) peut éventuellement être vide, si l'union des deux références couvre tout le symbole ($n = 0$).

- Chaque \mathcal{N}_i décrit le positionnement d'un élément de Référence Virtuelle Absolue (l'un des bords de la boîte englobante du symbole) par rapport aux deux références r_x et r_y . On peut noter $\mathcal{N}_i = (b_x^i, b_y^i)$, avec b_x et b_y les méta-modèles de positionnement relatifs aux références r_x et r_y respectivement. Les éléments de référence virtuelle absolue sont toujours au nombre de quatre.

Il convient de souligner le fait qu'un PSF n'est composé que de modèles de positionnement relatif, sans aucune description de formes. En particulier, les primitives ne sont pas étiquetées par un type tel que *segment* ou *arc de cercle* comme c'est souvent le cas dans les approches structurelles habituelles. La notion de *Patron Structurel Flou* repose bien sur un assemblage d'éléments à la forme quelconque, mais disposés les uns par rapport aux autres selon des modèles flous qui s'adaptent aux formes des tracés des références. Il faut également noter que tous les méta-modèles spatiaux intégrés dans le PSF sont assortis d'un modèle de l'étendue des objets, défini selon la méthode exposée dans la section 3.5.

La figure 4.4 résume les différents composants qui constituent un PSF pour un exemple de caractère chinois. Chaque image illustre l'une des relations spatiales modélisées dans le PSF. A chaque fois, l'objet en rouge est la référence pour la description du positionnement de l'objet bleu. La première ligne concerne les descriptions de positionnement relatives à r_x et la seconde à r_y . La première colonne représente les relations spatiales entre les deux références. Les trois colonnes suivantes concernent les composants du modèle, tandis que les quatre dernières concernent les éléments de Référence Virtuelle Absolue.

Avant de présenter la formalisation du calcul de l'adéquation d'une forme à reconnaître par rapport à un PSF, nous présentons le détail du calcul du score d'adéquation d'un segment de tracé donné avec un composant individuel du PSF.

4.2.2.1 Correspondance d'un segment avec un composant du PSF

Pour un segment S_i donné, on peut déterminer un score bipolaire de son positionnement et un score de l'adéquation de son étendue avec un composant \mathcal{M}_j du PSF. Le schéma de la figure 4.5 présente le mode de calcul de ces deux scores.

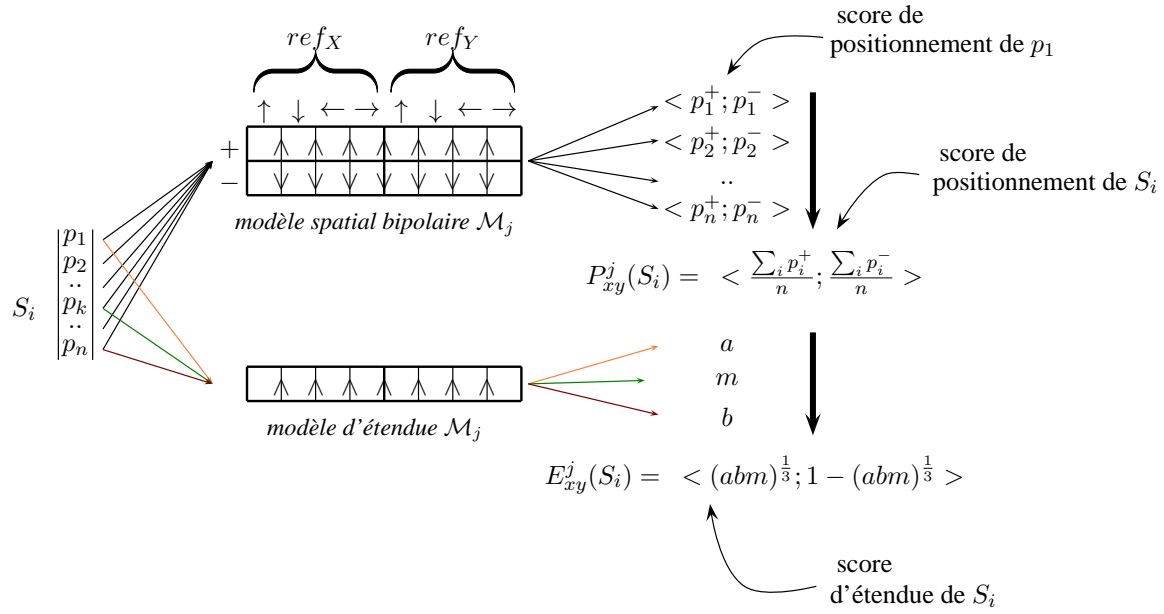


FIGURE 4.5: Schéma d'exploitation du composant j d'un PSF pour un segment S_i (modèle spatial bipolaire et modèle associé d'étendue par points particuliers). Le score de positionnement global est la moyenne arithmétique des scores bipolaires des points de S_i . Le score global d'étendue est une moyenne arithmétique des scores d'étendue de chaque point particulier (a,m,b).

Le score $P_{xy}^j(S_i)$ mesure l'adéquation du positionnement des points de S_i par rapport au modèle \mathcal{M}_j développé sur les deux références r_x et r_y . Un composant du PSF est défini par deux méta-modèles spatiaux, chacun relatif à l'une des deux références $\mathcal{M}_j = (a_x^j, a_y^j)$. Pour un point p_i du segment S_i , son adéquation au modèle de positionnement doit donc prendre en compte son positionnement par rapport aux deux références, selon quatre points de vue directionnels pour chacune d'entre elles. On rappelle que selon l'équation (3.20), l'adéquation du positionnement d'un point p par rapport à un modèle bipolaire a^j appliqué à une référence R est exprimé par :

$$\mu^{a^j}(R)(p) = \top[\mu_{\alpha_i}(R)(p)]_{i=1..4}, \quad (4.2)$$

avec \top une t-norme bipolaire et α_i qui désigne l'un des quatre points de vue considérés dans le méta-modèle a^j .

4.2. Patrons Structurels Flous pour la représentation de symboles

Dans le cas d'un point p à positionner conjointement par rapport à deux modèles distincts a_x^j et a_y^j , chacun étant développé sur une référence différente, on peut calculer l'adéquation de son positionnement par :

$$\mu^{\mathcal{M}_j}(r_x, r_y)(p) = \top(\mu^{a_x^j}(r_x)(p), \mu^{a_y^j}(r_y)(p)). \quad (4.3)$$

L'utilisation d'un opérateur de conjonction signifie que, pour être bien positionné par rapport à la description du composant \mathcal{M}_j , le point p doit être bien positionné vis-à-vis des deux modèles et des deux références à la fois.

La figure 4.6 représente le calcul de l'intersection de deux méta-modèles spatiaux définis sur les références r_x et r_y du caractère pris en exemple à la figure 4.4. Les deux premiers modèles décrivent la position du premier composant du caractère par rapport à r_x et r_y respectivement, tandis que le troisième modèle illustre le résultat de leur fusion selon l'équation (4.3). Cette figure illustre pour cet exemple l'intérêt de décrire le positionnement par rapport aux deux références simultanément : le modèle complet est en effet bien plus précis que les deux autres pris séparément.

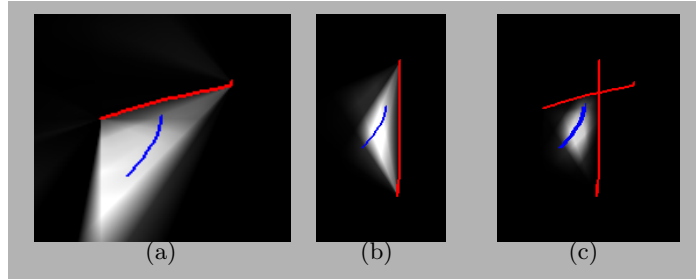


FIGURE 4.6: Méta-modèle pour le composant \mathcal{M}_1 du caractère de la figure 4.4 par rapport à sa référence r_x (a), par rapport à r_y (b) et visualisation du modèle défini par leur intersection selon l'équation (4.3)(c).

Par distributivité de la t-norme, on peut en fait considérer le score global $\mu^{\mathcal{M}_j}(r_x, r_y)(p)$ (équation (4.3)) comme le résultat d'application d'un modèle à huit points de vue, dont les quatre premiers sont relatifs à la référence r_x et les quatre autres sont relatifs à r_y . Ainsi, dans le schéma de la figure 4.5, on exprime l'adéquation du positionnement d'un point p_i par rapport au modèle comme le résultat de la conjonction bipolaire de huit mesures d'adéquation (selon les points de vue *haut* par rapport à ref_x , ..., *droite* par rapport à ref_x , *haut* par rapport à ref_y , ..., *droite* par rapport à ref_y). Le score bipolaire $\mu_j^{\mathcal{M}}(r_x, r_y)(p)$ y est noté pour simplifier $\langle p^+; p^- \rangle$. Il exprime, par un degré flou bipolaire, dans quelle mesure le point p est correctement positionné par rapport aux modèles de positionnement considérant quatre points de vue depuis les deux références à la fois.

Le score d'adéquation de positionnement pour le segment complet est calculé en moyennant l'adéquation bipolaire par rapport au modèle point à point. Cette stratégie de fusion correspond à la mise en œuvre bipolaire de la mesure moyenne d'adéquation formulée à la section 2.2.3.3 (voir l'équation (2.6)). L'opérateur de moyenne arithmétique donne un poids identique à chaque point de S_i , sans pénaliser lourdement le score global en cas d'existence de points mal positionnés. On note le score global $P_{xy}^j(S_i)$.

En parallèle, un score d'adéquation de l'étendue de S_i peut être calculé relativement au modèle. On adopte ici la stratégie d'étendue par points particuliers, qui avait été présentée à la section 3.5.3. Pour chaque point particulier identifié dans le segment S_i (points a , b et m), on peut évaluer l'adéquation de sa position relativement au modèle d'étendue pour chacun des deux modèles. Comme pour les scores de positionnement, on peut agréger les huit scores d'adéquation puisqu'il s'agit du même point de tracé qui est analysé sous huit points de vue (quatre points de vue directionnels depuis chacune des deux références). L'étendue n'est pas décrite par un modèle bipolaire et le résultat du calcul pour un point est donc un score flou simple. La fusion entre les scores d'étendue pour chaque point est faite par une *moyenne géométrique* qui détermine un score global $E_{xy}^j(S_i)$. Cet opérateur (plutôt qu'une moyenne arithmétique comme pour le score de positionnement) traduit le besoin d'avoir conjointement chacun des points particuliers bien positionné selon le modèle d'étendue. Le score global est donc fortement pénalisé si l'un des points particuliers a un mauvais score individuel. Le résultat du calcul est un score flou classique qui est ramené à sa généralisation bipolaire ($\langle u; v \rangle$ avec $v = 1 - u$) pour son intégration dans le mécanisme de fusion globale.

Le schéma présenté ici pour le calcul de la correspondance entre une portion de tracé S_i et un composant du PSF \mathcal{M}_j est reproduit presque à l'identique pour le calcul des scores de la RVA. La seule différence est que les modèles \mathcal{N}_j sont appliqués sur des éléments graphiques virtuels B_i (qui sont les quatre segments qui définissent les bords de la boîte englobante du symbole) et non pas sur des portions de tracés réels.

Pour le cas des tracés de référence r_x et r_y , le schéma est légèrement différent, puisqu'ils ne sont positionnés que l'un par rapport à l'autre (une seule référence de positionnement). Les deux schémas de la figure 4.7 présentent l'exploitation des modèles de positionnement de chaque référence par rapport à l'autre.

4.2.2.2 Fusion globale

Nous exposons ici comment est calculé le score global de correspondance entre un symbole à reconnaître \mathcal{S} et un modèle de PSF noté \mathcal{T} . Considérons le cas général d'un symbole segmenté en deux références r_x , r_y ainsi que k éléments $\{S_i\}_{i=1..k}$ et d'un PSF \mathcal{T} comportant n composants $\{\mathcal{M}_i\}_{i=1..n}$. On désigne par σ la fonction qui assigne les segments S aux composants de \mathcal{T} .

$$\sigma : \{S_1..S_k\} \rightarrow \{\mathcal{M}_1..M_n\} \quad (4.4)$$

Pour le calcul de l'adéquation de \mathcal{S} et \mathcal{T} , la fonction σ n'est pas obligatoirement surjective sur $\{\mathcal{M}_1..M_n\}$. Autrement dit, on peut tolérer que certains composants du modèle \mathcal{T} ne soient mis en correspondance avec aucun élément de tracé de \mathcal{S} . On parle alors d'assignement *non couvrant*. Certains éléments de tracés peuvent également n'être associés à aucun composant de \mathcal{T} par σ : on parle alors d'assignement *incomplet*.

On définit σ' l'application pseudo-inverse de σ définie comme suit de $\{\mathcal{M}_1..M_n\}$ dans $\{S_1..S_k\} \cup \{\emptyset\}$.

$$\sigma'(\mathcal{M}) = \begin{cases} \{S \in \{S_1..S_k\} | \sigma(S) = \mathcal{M}\} & \text{si } \exists S \in \{S_1..S_k\} | \sigma(S) = \mathcal{M} \\ \emptyset & \text{sinon.} \end{cases} \quad (4.5)$$

4.2. Patrons Structurels Flous pour la représentation de symboles

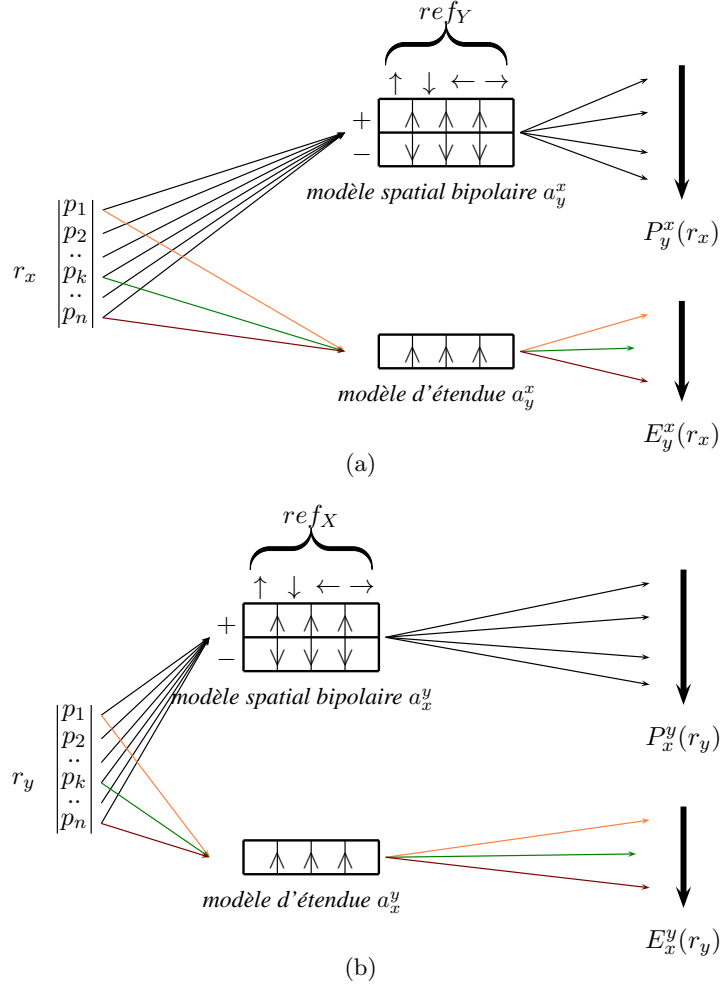
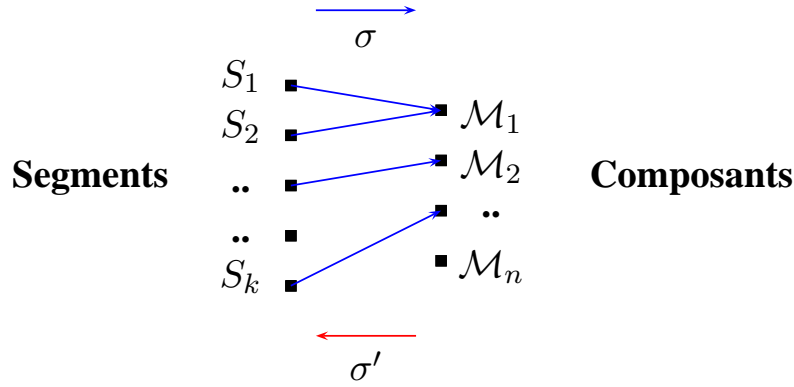


FIGURE 4.7: Schéma d'exploitation du modèle de position de la référence horizontale par rapport à la référence verticale (a) et de la référence verticale par rapport à la référence horizontale (b).

La figure 4.8 représente un exemple d'assignement σ incomplet (tous les segments ne sont pas affectés à un composant du modèle) et non couvrant (tout composant n'a pas de segment affecté). Les segments S_i non assignés sont ceux tels que $\sigma(i) = \emptyset$, tandis que les composants \mathcal{M}_j non couverts par σ sont tels que $\sigma'(\mathcal{M}_j) = \emptyset$.

Pour le calcul de l'adéquation globale d'un symbole avec un PSF, les différentes mesures bipolaires floues de positionnement et d'étendue doivent être combinées dans un mécanisme de fusion. Le système de fusion proposé permet d'imiter un raisonnement logique interprétable pour la combinaison de ces différentes sources d'informations. Dans notre cas applicatif, la construction du mécanisme de fusion consiste à traduire en opérateurs et modèles logiques flous le raisonnement exprimé comme suit.

Le calcul d'un score global d'adéquation pour le symbole $\mathcal{S} = r_x, r_y, \{S_i\}_{i=1..k}$ par rapport au PSF \mathcal{T} , compte tenu de la fonction d'assignement σ , doit mettre en jeu :


 FIGURE 4.8: Exemple d'assignement σ incomplet et non couvrant.

- la bonne adéquation de positionnement et d'étendue relative des références r_x et r_y par rapport aux modèles a_y^x et a_x^y de \mathcal{T} ,
- la bonne adéquation de positionnement et d'étendue des RVA (relativement à r_x, r_y) avec les modèles $\{\mathcal{N}_j\}_{j=1..4}$,
- la bonne adéquation de positionnement et d'étendue des éléments S_i (relativement à r_x, r_y) avec les modèles du composant $\{\mathcal{M}_j\}_{j=1..n}$ auxquels ils sont assignés par σ ,
- une pénalisation pour les éléments de tracé non assignés ($S \in \{S_1..S_k\} | \sigma(i) = \emptyset$),
- une pénalisation pour les composants de \mathcal{T} non couverts ($\mathcal{M} \in \{\mathcal{M}_1..M_n\} | \sigma'(\mathcal{M}) = \emptyset$).

Schéma général de fusion Le schéma de la figure 4.9 représente le mécanisme de fusion combinant tous les éléments mis en œuvre pour déterminer le score global de correspondance entre un symbole \mathcal{S} muni de ses deux références r_x, r_y et un PSF \mathcal{T} , en fonction de l'assignement σ . Nous détaillons dans les paragraphes suivants les différents opérateurs de combinaison intervenant dans ce procédé de fusion.

Score individuel des éléments de la structure Pour chaque élément de la structure (chaque ligne de la figure 4.9), un score s_i est calculé en combinant les scores de positionnement et d'étendue obtenus selon le calcul schématisé par la figure 4.5. La combinaison est réalisée par un opérateur de moyenne géométrique pondérée. L'opération représentée dans la figure 4.9 par le schéma :

$$P \wedge E \rightarrow s$$

se traduit donc par le calcul suivant :

$$s = \langle s^+; s^- \rangle = \left\langle (p^+)^\alpha (e^+)^\beta; 1 - (1 - p^-)^\alpha (1 - e^-)^\beta \right\rangle, \quad (4.6)$$

avec $P = \langle p^+; p^- \rangle$ et $E = \langle e^+; e^- \rangle$. α et β sont les poids déterminant l'importance relative du positionnement et de l'étendue dans le calcul du score s . Ils vérifient $\alpha + \beta = 1$. Le résultat s est un score flou bipolaire ($s \in \mathcal{L}$, ensemble des nombres bipolaires défini au paragraphe 3.4.1).

4.2. Patrons Structurels Flous pour la représentation de symboles

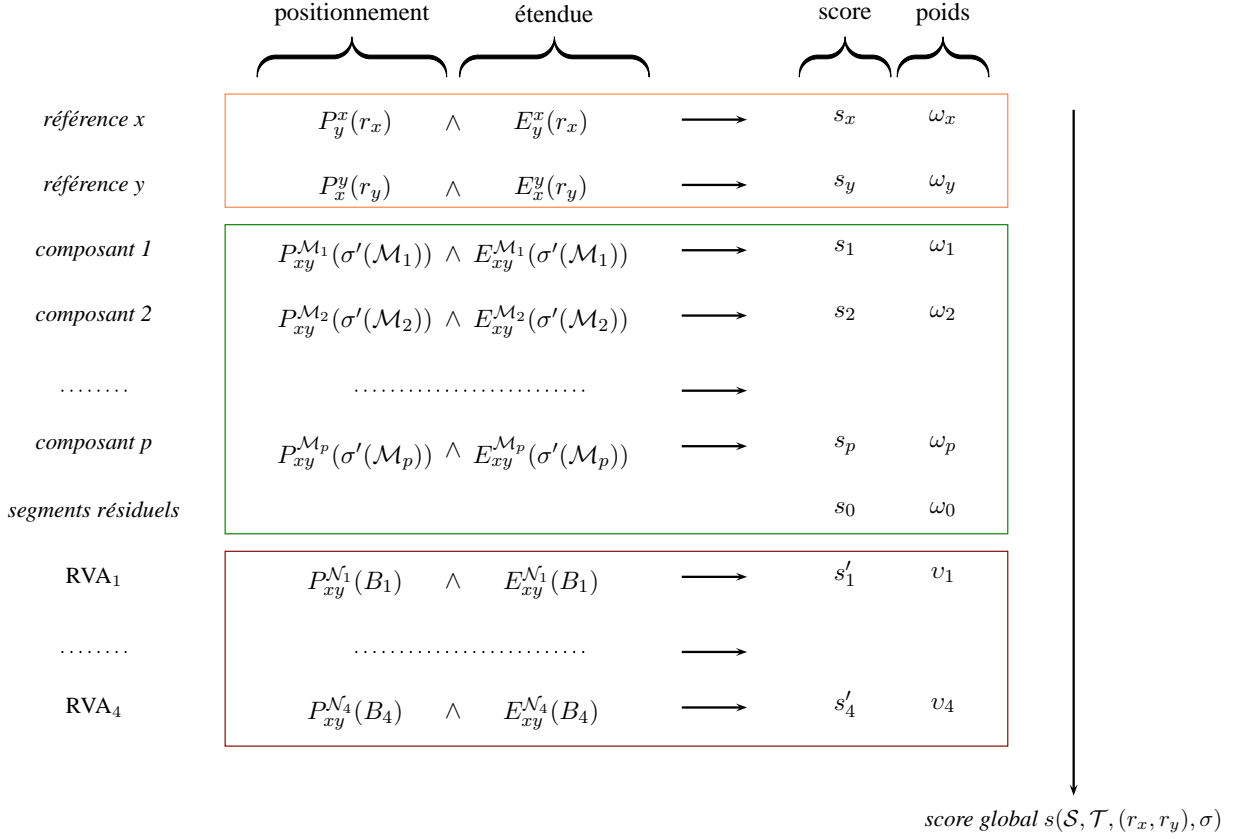


FIGURE 4.9: Schéma de fusion globale pour le calcul du score d'adéquation entre un symbole \mathcal{S} et un PSF \mathcal{T} . Chaque ligne combine un score de positionnement et un score d'étendue pour constituer un score complet pour l'un des éléments de la structure. Le score global d'adéquation est obtenu par combinaison pondérée des scores des éléments de structure.

Nous présentons ici une simplification de notation par l'introduction de deux opérateurs. Dans ses travaux, Xu [Xu07] définit des opérateurs d'agrégation de scores flous *intuitionnistes*, qui sont formellement équivalents à des scores bipolaires flous. Nous nous en inspirons pour définir l'opérateur de moyenne géométrique bipolaire. Notons \otimes l'opérateur de t-norme bipolaire construit à partir de la t-norme floue *produit* et de sa t-conorme duale \top , de sorte que, pour tout a, b dans \mathcal{L} , on a :

$$\begin{aligned} a \otimes b &= \langle t(a^+, b^+); \top(a^-, b^-) \rangle \\ &= \langle a^+ * b^+; 1 - (1 - a^-)(1 - b^-) \rangle. \end{aligned}$$

Nous définissons aussi l'opérateur \bullet , de $\mathbb{R} \times \mathcal{L}$ dans \mathcal{L} tel que

$$\lambda \bullet a = \langle (a^+)^{\lambda}; 1 - (1 - a^-)^{\lambda} \rangle, \lambda > 0. \quad (4.7)$$

Cet opérateur vérifie que :

$$\forall \lambda > 0, a \in \mathcal{L}, \quad \lambda \bullet a \in \mathcal{L}. \quad (4.8)$$

A l'aide des deux opérateurs \otimes et \bullet , on peut exprimer l'expression de l'équation (4.6) par la notation suivante :

$$s = (\alpha \bullet P) \otimes (\beta \bullet E). \quad (4.9)$$

Pour tous les éléments de la structure, on définit ainsi un score s_i à partir de la fusion selon cette même équation de combinaison (4.9). Ce schéma de fusion est homogène quel que soient les éléments de structure concernés : références r_x et r_y (premier bloc du schéma 4.9), composants \mathcal{M}_i du PSF (second bloc) et Référence Virtuelle Absolue (troisième bloc).

Deux cas particuliers peuvent toutefois se présenter. Les *segments résiduels* éventuels sont les segments du tracé qui n'ont été associés à aucun modèle, pour lesquels il est donc impossible de calculer un score de positionnement et d'étendue. Ces tracés non assignés doivent être pénalisés pour le score global de la correspondance, car ils traduisent la présence de tracés non attendus par le modèle. Le score s_0 qui leur est attribué est une constante de pénalisation $s_0 = c_0 \in \mathcal{L}$.

Enfin, il peut se présenter le cas de composants du modèle auquel aucun segment n'est assigné (ce sont les composants non couverts, \mathcal{M}_j tels que $\sigma'(\mathcal{M}_j) = \emptyset$). On leur attribue alors un score s_j défini par une autre constante de pénalisation c_1 . c_0 et c_1 sont deux hyperparamètres du système.

Calcul du score global Le calcul de la fusion globale se fait au moyen des opérateurs introduits précédemment selon l'équation suivante :

$$\begin{aligned} s(\mathcal{S}, \mathcal{T}, r_x, r_y, \sigma) &= (\omega_x \bullet s_x) \otimes (\omega_y \bullet s_y) \\ &\quad \otimes (\omega_1 \bullet s_1) \otimes \dots \otimes (\omega_p \bullet s_p) \\ &\quad \otimes (\omega_0 \bullet s_0) \\ &\quad \otimes (v_1 \bullet s'_1) \otimes \dots \otimes (v_4 \bullet s'_4). \end{aligned}$$

Le calcul du score global de correspondance entre \mathcal{S} et \mathcal{T} fait donc bien intervenir tous les éléments de la structure, la contribution de chacun étant pondérée par un poids ω_i, v_i . Ces poids doivent respecter la contrainte $\omega_x + \omega_y + \omega_1 + \dots + \omega_p + \omega_0 + v_1 + \dots + v_4 = 1$.

L'idée générale du schéma de fusion est de pondérer uniformément tous les éléments de la structure, puisqu'a priori nous ne souhaitons pas considérer certains éléments comme étant plus importants que d'autres. Nous notons pour la suite w le poids donné à un élément de structure « normal ».

Une première exception à l'équilibre des poids concerne les éléments de Référence Virtuelle Absolue. On peut considérer que leur poids relatif doit être amoindri car il ne s'agit pas d'éléments réels du tracé, mais d'éléments de référence virtuelle. On corrige donc leur poids par un facteur ρ : $v_i = \rho * w, 0 \leq \rho \leq 1$.

Par ailleurs, les scores s_x et s_y issus des deux éléments r_x et r_y ont la particularité d'être naturellement moins importants que les autres scores s_i . Cela est dû au fait que le calcul de positionnement et d'étendue de ces objets ne provient de la fusion que d'un seul modèle spatial à quatre points de vue, tandis que les scores de positionnement et d'étendue des autres objets sont issus de la fusion de deux méta-modèles spatiaux à quatre points de vue chacun. Les schémas de la figure 4.7 illustrent cette particularité par comparaison avec

4.2. Patrons Structurels Flous pour la représentation de symboles

le schéma général 4.5. L'opérateur de t-norme utilisé pour combiner les scores des quatre points de vue est un produit direct (il ne s'agit pas d'une moyenne *pondérée*), qui donne donc un « poids » naturel plus important en fonction du nombre de facteurs dans le produit. Afin de corriger le déséquilibre naturel entre s_x , s_y et les autres scores s_i , il convient de donner un poids double à ces deux scores dans la fusion : $\omega_x = \omega_y = 2 * w$.

Enfin, le poids donné au score de pénalisation pour les segments résiduels s_0 doit être adapté en fonction de l'existence ou non de ces résidus ainsi que de leur importance. Plus la proportion de tracé non-assigné est importante, plus le poids associé à s_0 doit être important. On détermine le poids associé au score de pénalisation s_0 de la façon suivante :

$$\omega_0 = \delta * w \text{ avec } \delta = \begin{cases} 0 & \text{si } S^* = \emptyset \\ \max\left(1, p \frac{\sum_{S \in S^*} |S|}{\sum_{S \in \mathcal{S}} |S|}\right) & \text{sinon,} \end{cases} \quad (4.10)$$

où S^* est l'ensemble des tracés résiduels ($S^* = \{S_i, \sigma(S_i) = \emptyset\}$) et $|S|$ est le nombre de points d'un élément de tracé. De cette façon, le poids donné au score de pénalisation est au moins égal au poids w dès lors que des tracés résiduels existent. Si tous les tracés sont assignés aux composants, le poids est nul et s_0 n'intervient pas dans le calcul du score global. Enfin, si la proportion des tracés résiduels est suffisamment importante (supérieure à $\frac{1}{p}$, soit la proportion moyenne de points dans un composant de \mathcal{S}), le poids ω_0 est augmenté proportionnellement.

Les scores des autres composants de la structure ont un poids $\omega_i = w$, y compris pour le cas des composants qui ne se sont vus assigner aucun élément de tracé par σ (leur score est alors défini par la constante de pénalisation c_1).

Finalement, les poids sont complètement déterminés par la contrainte qui impose que leur somme doit valoir 1. On a donc :

$$\begin{aligned} & \omega_x + \omega_y + \omega_1 + \dots + \omega_p + \omega_0 + v_1 + \dots + v_4 \\ &= 2w + 2w + pw + \delta w + 4pw \\ &= 1. \end{aligned}$$

On peut donc déterminer w pour toute configuration du PSF et de σ par le calcul suivant :

$$w = \frac{1}{4 + p + \delta + 4\rho}. \quad (4.11)$$

4.2.2.3 Comparaison de scores bipolaires

Les scores de correspondance entre un symbole \mathcal{S} et des prototypes PSF doivent pouvoir être comparés, notamment pour pouvoir décider quel prototype PSF représente le mieux le symbole. Pour une tâche de reconnaissance de symboles, la classe retenue pour \mathcal{S} est celle dont le prototype \mathcal{T}^* maximise le score d'adéquation $s(\mathcal{S}, \mathcal{T})$. Puisque les scores d'adéquation sont des scores flous bipolaires, la comparaison des scores nécessite la définition d'un ordre total sur \mathcal{L} (ensemble des scores flous bipolaires). Plusieurs définitions sont possibles, comme le tri total lexicographique évoqué à la section 3.4.2. Nous choisissons d'utiliser plutôt le tri défini dans [MDT09], également formulé par Xu dans le contexte de la théorie floue *intuitionniste* d'Atanassov [Xu07] :

$$\forall x, y \in \mathcal{L}, x \prec_a y \quad \text{ssi} \quad \begin{cases} x^+ - x^- < y^+ - y^- \\ \text{ou} \quad x^+ - x^- = y^+ - y^- \quad \text{et} \quad x^+ < y^+, \end{cases} \quad (4.12)$$

L'ordre total \prec_a permet ainsi de comparer des scores bipolaires et de les trier. Cet ordre est plus utilisable en pratique car plus stable qu'un tri lexicographique qui est hypersensible aux différences sur l'une des deux composantes (positive ou négative).

Formellement, on peut écrire que la décision de reconnaissance pour un symbole \mathcal{S} est la classe portée par \mathcal{T}^* , avec :

$$\mathcal{T}^* = \max_{i=1..C} s(\mathcal{S}, \mathcal{T}_i). \quad (4.13)$$

C représente le nombre de prototypes de PSF (nombre de classes) et l'opérateur *max* est construit à partir de l'ordre total \prec_a . Le score $s(\mathcal{S}, \mathcal{T}_i)$ est défini par :

$$\begin{aligned} s(\mathcal{S}, \mathcal{T}_i) &= s(\mathcal{S}, \mathcal{T}_i, (r_x, r_y)^*, \sigma^*) \\ &= \max_{(r_x, r_y) \in R_x \times R_y} \left(\max_{\sigma \in \Omega} s(\mathcal{S}, \mathcal{T}, r_x, r_y, \sigma) \right), \end{aligned}$$

avec R_x et R_y les ensembles d'objets de référence possibles pour r_x et r_y et Ω l'ensemble des assignements σ autorisés pour un couple de références déterminé. Pour le meilleur score obtenu, l'assignement σ^* encode la meilleure segmentation du symbole guidée par le PSF associé.

4.2.3 Exploitation des PSF

Nous avons formalisé le calcul de la correspondance entre un symbole préalablement segmenté, dont les références sont déterminées et pour un assignement donné de ses segments avec les composants d'un PSF. Nous présentons maintenant le mécanisme d'exploitation de cette formalisation pour calculer le score d'adéquation d'un symbole avec un PSF, en tenant compte de problématiques telles que la recherche de références, la segmentation des tracés, et la détermination du meilleur assignement.

Le score d'adéquation retenu pour un symbole par rapport à un prototype PSF est le meilleur score obtenu en considérant toutes les hypothèses de références extraites. Nous présentons d'abord la méthode pour extraire et trier des hypothèses de références pertinentes pour décrire le positionnement relatif des éléments du symbole.

Etant donné un choix de références, il faut ensuite proposer une segmentation du tracé. Dans notre méthode, cette segmentation s'appuie sur les méta-modèles spatiaux compris dans le PSF. Nous en présentons la mise en œuvre dans un second temps.

Enfin, pour un choix de références et pour une segmentation proposée, il faut chercher le meilleur assignement des segments vers les composants. La description de la recherche de cette fonction d'assignement fait l'objet de la troisième partie.

4.2.3.1 Hypothèses de références

Considérons un symbole \mathcal{S} que l'on cherche à reconnaître, représenté par son signal en ligne. \mathcal{S} est constitué de N *tracés naturels* : $\mathcal{S} = t_1..t_N$, où chaque tracé est une suite de

4.2. *Patrons Structurels Flous* pour la représentation de symboles

points $t_i = p_i^1, p_i^2, \dots, p_i^{k_i}$. L'extraction d'hypothèses de références consiste à trouver dans \mathcal{S} des portions de tracés qui forment de « bonnes » références pour décrire le positionnement des éléments de sa structure. Pour cela, nous détectons d'abord un certain nombre de points d'intérêt du tracé, puis nous trions les portions de tracé délimitées par ces points en fonction du critère exprimé à l'équation (4.1).

Extraction de points d'intérêt A partir du tracé de \mathcal{S} , nous cherchons à extraire certains points particuliers du tracé : ce sont des points supposés délimiter des portions de tracés constituant de « bonnes » références. La recherche des points particuliers est faite indépendamment pour la recherche de références horizontales d'une part, verticales d'autre part. Les points de début et de fin de chaque tracé naturel constituent un premier ensemble de points d'intérêt (il s'agit des points p_i^1 et $p_i^{k_i}$). Nous extrayons aussi les points du symbole d'abscisse et d'ordonnée minimales et maximales. Enfin, nous appliquons sur chaque tracé naturel t_i un algorithme d'approximation polygonale (directement appliqué d'après [PV93]). L'approximation polygonale à n segments de droite du tracé t_i détermine $n - 1$ sommets internes au tracé qui définissent des points d'intérêt de t_i . Le nombre de segments de droite dans l'approximation est déterminé par un seuil relatif sur l'erreur d'approximation du tracé. n est donc augmenté progressivement jusqu'à ce que la réduction de l'erreur par rapport à l'approximation à $n - 1$ segments soit inférieure au seuil fixé. En moyenne le nombre n retenu varie entre 1 (lorsque t_i est suffisamment rectiligne) et 5 environ pour les tracés les plus complexes constatés. Cette méthode de détection de points singuliers s'est avérée plus stable par rapport à des essais conduits avec des méthodes basées sur des variations locales de formes (obtenus par exemple par la méthode de Rosenfeld [RJ73]).

L'ensemble de tous les points d'intérêt extraits pour la recherche de références horizontales et verticales sont notés $\{q_i^x\}$ et $\{q_i^y\}$, respectivement. Si plusieurs points q_i extraits sont identiques ou trop proches, ils sont fusionnés en un seul point d'intérêt moyen.

Constitution d'hypothèses de références Les combinaisons de points d'intérêt pris deux à deux définissent des portions de tracés qui sont des hypothèses de références horizontales ou verticales. Pour limiter le nombre d'hypothèses, toutes les combinaisons de points de $\{q_i\}$ ne sont pas autorisées. Si le point d'intérêt q_a correspond au point $p_{i_a}^{j_a}$ (point d'indice j_a sur le tracé t_{i_a}) et q_b au point $p_{i_b}^{j_b}$ (point d'indice j_b sur un autre tracé t_{j_b}), on définit la référence associée par le plus court chemin reliant les deux points *indépendamment de l'ordre des tracés*. Certaines hypothèses de référence sont alors exclues si le chemin qui les relie emprunte des *sauts* trop importants entre les tracés. Si $i_a = i_b$, les deux points proviennent du même tracé naturel t_{i_a} et la paire (q_a, q_b) définit une portion de t_{i_a} qui est considérée comme une hypothèse de référence acceptable. Sinon, on vérifie que la somme des distances entre les extrémités des tracés constituant le meilleur chemin de q_a à q_b est inférieure à un seuil D_{max} . Enfin, on interdit à une portion de tracé de référence de commencer sur une fin de tracé naturel ou de terminer sur un début de tracé naturel. Toutes les paires de points qui respectent ces quatre conditions définissent les hypothèses de références admissibles.

Tri et sélection des hypothèses Les hypothèses de références $\{q_a^x, q_b^x\}$ sont triées selon la mesure de qualité définie à l'équation (4.1). Cette mesure favorise les portions de tracés

qui couvrent le reste du symbole dans toute sa dimension horizontale ou verticale. Les deux listes d'hypothèses de références horizontales L_x et verticales L_y sont triées selon ce critère et limitées à un nombre maximal d'hypothèses.

Exemples Le tableau de la figure 4.10 présente quelques exemples de références *horizontales* et *verticales* extraites pour différents exemples de symboles manuscrits et triées de gauche à droite par score décroissant. Les portions de tracés références sont représentées en rouge. Les exemples mis en évidence par un encadré illustrent la capacité de la méthode à isoler des références en fragmentant ou en connectant plusieurs tracés naturels.

4.2.3.2 Segmentation guidée par les modèles spatiaux

Etant donné un couple de tracés de références $(r_x, r_y) \in L_x \times L_y$, on note $\mathcal{S}|_{xy}$ le *reste* du tracé du symbole \mathcal{S} . $\mathcal{S}|_{xy}$ contient tous les éléments de tracé de \mathcal{S} qui ne sont pas couverts par l'une des deux références r_x et r_y .

Supposons que l'on cherche à calculer la valeur d'adéquation entre \mathcal{S} muni des deux références r_x et r_y et un PSF \mathcal{T} comportant n composants. Pour pouvoir assigner les éléments de $\mathcal{S}|_{xy}$ aux différents composants de \mathcal{T} , il est d'abord nécessaire de déterminer une *segmentation* possible des tracés de $\mathcal{S}|_{xy}$. En effet, à ce stade $\mathcal{S}|_{xy}$ contient des tracés naturels (tels qu'ils ont été saisis par le scripteur), ainsi que d'éventuels résidus de tracés résultant de l'extraction des références r_x et r_y . L'objectif de cette étape est d'identifier dans $\mathcal{S}|_{xy}$ des points de segmentation potentiels. A cause de la variabilité des habitudes d'écriture, un même tracé naturel doit parfois être segmenté car il correspond à la saisie de plusieurs éléments de la structure connectés entre eux. Il faut donc déterminer en quels points de $\mathcal{S}|_{xy}$ les tracés peuvent être segmentés. Classiquement, la recherche de points de segmentation passe par la détection de points d'intérêt dans la forme des tracés. Ce type de méthode repose sur une analyse locale des formes, sans prise en compte du contexte de leur positionnement. Dans notre approche, au contraire, nous exploitons la capacité prédictive des méta-modèles de positionnement pour formuler ces hypothèses de segmentation : on parle alors de segmentation guidée par le contexte spatial des tracés.

Etant donnée la sélection au préalable d'un couple de références (r_x, r_y) , il est possible d'appliquer les méta-modèles spatiaux des différents composants du PSF sur ces références. Ils définissent n modèles spatiaux bipolaires $\mu^{\mathcal{M}_i}(\{r_x, r_y\})$, $i = 1..n$ qui peuvent être évalués en tout point de l'espace du tracé. Pour rappel, en un point p du plan, $\mu^{\mathcal{M}_i}(\{r_x, r_y\})(p)$ exprime le degré (bipolaire) d'adéquation de p avec la relation spatiale \mathcal{M}_i . Ce score est élevé si p est positionné par rapport à (r_x, r_y) conformément au i ème composant dans le modèle de PSF.

En supposant que $\mathcal{S}|_{xy}$ corresponde bien au tracé du symbole représenté par le PSF, il doit être possible de localiser des éléments de tracés dans chacune des zones de « bon » positionnement définies par les méta-modèles appliqués aux références. En s'appuyant sur ce principe, la recherche des frontières entre les différents éléments à segmenter dans $\mathcal{S}|_{xy}$ peut se concentrer sur les zones du tracé où l'on constate un changement de *modèle dominant*. On peut calculer l'adéquation de chaque point p du tracé de $\mathcal{S}|_{xy}$ avec chaque modèle \mathcal{M}_i . Le modèle j avec lequel p a le meilleur degré d'adéquation est le modèle dominant au point p , on note alors $\mathcal{M}^*(p) = \mathcal{M}_j$.

4.2. Patrons Structurels Flous pour la représentation de symboles

hypothèses de références <i>horizontales</i>	hypothèses de références <i>verticales</i>

FIGURE 4.10: Visualisation des 5 meilleures hypothèses de références horizontales et verticales pour différents exemples de symboles. Les tracés références sont représentés en rouge. Les exemples encadrés illustrent différents cas de références constituées par segmentation ou connexion des tracés originaux.

Si en parcourant les points d'un tracé de $\mathcal{S}|_{xy}$ on observe un changement du modèle dominant, on peut faire la supposition que le tracé t est constitué en réalité de plusieurs éléments de la structure qu'il faut segmenter. Le point estimé de changement du modèle dominant est alors considéré comme un point de segmentation potentiel.

Soit $t = p^1..p^k$ un tracé de $\mathcal{S}|_{xy}$, on considère que le point p^i de t est un point de segmentation potentiel si les conditions suivantes sont vérifiées :

$$\begin{cases} \mathcal{M}^*(p^{i-\delta}) = \mathcal{M}^*(p^{i-\dots}) = \mathcal{M}^*(p^{i-1}) = \mathcal{M}_a \\ \mathcal{M}^*(p^{i+1}) = \mathcal{M}^*(p^{i+\dots}) = \mathcal{M}^*(p^{i+\delta}) = \mathcal{M}_b \\ a \neq b, \end{cases} \quad (4.14)$$

avec $\delta \in \{2, 3, \dots\}$. Le paramètre δ permet de contrôler la stabilité du changement de modèle dominant (sa valeur minimale est de 2). Selon cette stratégie, un point de segmentation potentiel est envisagé dès qu'un changement durable de modèle spatial dominant est détecté. Les points de début et de fin de tracé naturel (p^1, p^k) sont également considérés comme des points de segmentation potentiels.

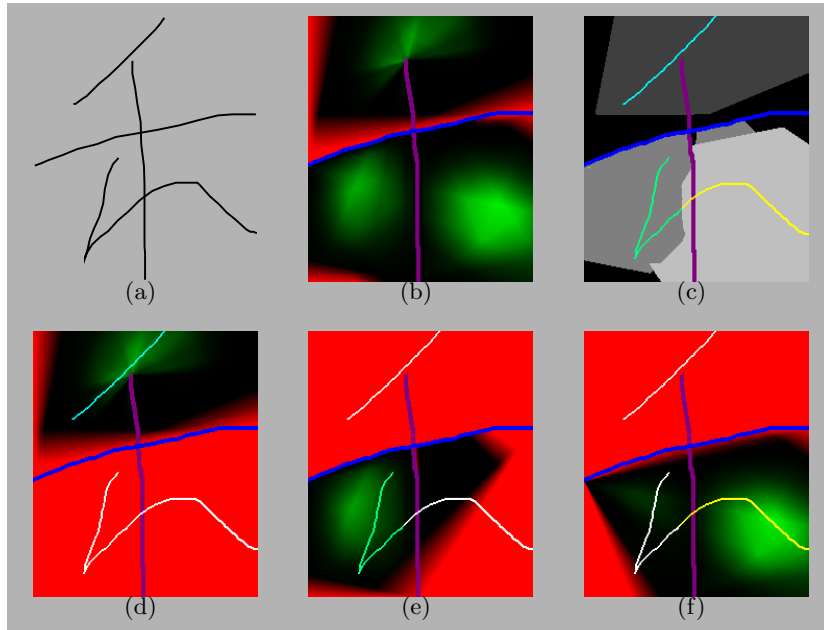


FIGURE 4.11: Exemple de tracé à segmenter (a) et application des 3 modèles spatiaux du PSF correspondant sur une paire de références (surimposées en violet) (b). Les 3 modèles spatiaux sont représentés isolément (d, e, f). La figure (c) représente une carte des modèles dominants, en trois niveaux de gris différents pour les trois modèles. Les éléments de tracé à segmenter sont colorés selon le modèle dominant en chaque point.

La figure 4.11 décrit le principe de segmentation des tracés guidés par les modèles spatiaux sur un exemple de caractère chinois à reconnaître (image (a)). Le PSF correspondant comporte trois composants, qui sont représentés par trois méta-modèles de positionnement spatial. La figure (b) représente l'union des trois modèles spatiaux obtenus par application de ces méta-modèles sur une paire de références. Le reste du caractère est composé de

4.2. Patrons Structurels Flous pour la représentation de symboles

deux tracés manuscrits. La représentation de l'image (c) permet de mettre en évidence en différents niveaux de gris les zones de dominance des trois modèles. Les portions de tracés sont colorées en fonction du modèle dominant en chacun de leur point. Les points de segmentation obtenus sont dans ce cas les extrémités des tracés et le point de changement de couleur du tracé (c).

L'avantage de cette stratégie de segmentation est la prise en compte du contexte de positionnement des tracés. Les hypothèses de segmentation sont formulées en fonction de la forme des références et de modèles spatiaux, appris à partir de données, qui traduisent la variabilité du positionnement des différents éléments de la structure. En pratique, cela se traduit par une extraction de points de segmentation différents en fonction du PSF qui est appliqué sur un même symbole à reconnaître. Le principe de segmentation revient en effet à laisser chaque modèle spatial d'un PSF appris sélectionner les éléments de tracés qui semblent lui revenir, d'après leur positionnement. L'information utilisée pour segmenter les tracés repose donc sur une vision globale du symbole et de sa structure, plutôt que sur une information locale relative à la forme des tracés.

4.2.3.3 Recherche de l'assignement optimal des segments

Les points de segmentation potentiels qui ont été définis au paragraphe précédent par l'application des modèles spatiaux définissent une suite de N éléments indivisibles $\mathcal{S}|_{xy} = s_1 \dots s_N$. La mise en correspondance de ces éléments avec les composants du PSF impose la résolution d'un problème d'optimisation combinatoire. Il s'agit de trouver σ^* , le meilleur assignement possible des N segments vers les n composants (le terme d'assignement est utilisé abusivement, car comme nous l'avons présenté à la section 4.2.2.2, la fonction σ peut être non couvrante et incomplète). Les assignements sont évalués par la fonction d'adéquation s dont le calcul a été schématisé à la figure 4.9.

Nous proposons de résoudre le problème combinatoire par un algorithme bien connu de *séparation et évaluation* (aussi souvent appelé en anglais : *branch and bound*). Le principe de l'algorithme est de construire l'arbre des assignements possibles progressivement, en éliminant dynamiquement certaines branches dont on peut affirmer qu'elles ne conduisent pas à la meilleure solution. Le calcul d'une fonction d'évaluation (fonction de *bounding*) permet d'estimer une borne supérieure sur le score qu'il est possible d'atteindre dans chaque branche inexplorée. Si le résultat de cette estimation optimiste de l'exploration d'une branche est moins bon que le score d'une solution réelle déjà trouvée, l'exploration de toute la branche peut être abandonnée. La phase de séparation (ou *branching*) consiste à subdiviser le problème d'assignement en sous-problèmes, c'est-à-dire à compléter un assignement partiel avec une nouvelle association (tracé, composant).

Fonction de *bounding* La fonction de score que l'on cherche à optimiser est le résultat global $s(\mathcal{S}, \mathcal{T}, r_x, r_x, \sigma_i)$ déterminé par le mécanisme de fusion complet (voir la figure 4.9).

La fonction de *bounding* doit vérifier les conditions suivantes :

$$\begin{cases} g(H_i) \succeq_a s(\mathcal{S}, \mathcal{T}, (r_x, r_x), H_i^*) & \text{pour tout nœud } H_i \text{ de l'arbre} \\ g(H_i) = s(\mathcal{S}, \mathcal{T}, (r_x, r_x), H_i) & \text{pour toute feuille } H_i \text{ de l'arbre} \\ g(H_i) \preceq_a g(H_j) \text{ si } H_j \text{ est le nœud père de } H_i, \end{cases} \quad (4.15)$$

avec H_i^* le meilleur assignement possible obtenu dans le sous-arbre de H_i . La fonction g estime donc une borne supérieure du score s qu'il est possible d'atteindre en explorant la branche portée par un assignement partiel H_i . Supposons qu'à un moment donné de l'exploration de l'arbre, on a déjà trouvé une solution admissible σ_1 dont le score vaut $s_1 = s(\mathcal{S}, \mathcal{T}, (r_x, r_x), \sigma_1)$. Si pour un nœud H_i dont le sous-arbre n'a pas encore été exploré, on a $g(H_i) < s_1$, alors on sait que l'exploration du sous-arbre de H_i peut être évitée.

La fonction de *bounding* doit constituer un bon compromis entre la précision de l'estimation qu'elle fournit et le coût de son calcul. Si elle est trop grossière, elle sera moins efficace et ne permettra pas d'élaguer de grandes parties de l'arbre de recherche. A l'inverse, si elle est très précise, son calcul risque de générer une complexité élevée.

Le problème est donc de construire, à partir d'un assignement partiel H_i , un majorant simple et aussi proche que possible du score du meilleur assignement complet H_i^* au sens de la fonction de fusion globale présentée au paragraphe 4.2.2.2. On se sert du fait que l'opérateur de combinaison \otimes est monotone vis-à-vis de l'ordre total \prec_a défini par (4.12), c'est-à-dire :

$$\forall a, b, c \in \mathcal{L}, a \prec_a b \Rightarrow a \otimes c \prec_a b \otimes c. \quad (4.16)$$

De la même façon, on peut exploiter la propriété suivante pour l'opérateur \bullet :

$$\forall a, b \in \mathcal{L}, \lambda \in \mathbb{R}, a \prec_a b \Rightarrow \lambda \bullet a \prec_a \lambda \bullet b. \quad (4.17)$$

Ces propriétés garantissent que la recherche d'un majorant global pour $s(\mathcal{S}, \mathcal{T}, (r_x, r_x), H_i^*)$ peut se limiter à majorer chacun de ses termes dans le calcul de la fusion globale (décrit au paragraphe 4.2.2.2). Pour un assignement partiel $H_i = \{i_1, \dots, i_m, -1, \dots, -1\}$, les m premiers éléments de tracés sont déjà assignés à des composants du PSF. Les $N - m$ éléments restants à assigner sont attribués au composant pour lequel leur score d'adéquation de positionnement est le meilleur. Toutefois, si ce score est inférieur à la constante de pénalisation c_0 pour un tracé résiduel, le tracé n'est pas assigné. Le score de fusion globale est majoré en ne tenant pas compte des pénalisations pour les éléments non assignés, ni pour les modèles non couverts (le poids de ces éléments dans la fusion est reversé vers le composant ayant le meilleur score s_i). Enfin, tous les scores d'étendue E sont majorés par leur valeur maximale théorique, c'est-à-dire le score bipolaire $\langle 1; 0 \rangle$. Ceci retarde l'évaluation des scores E uniquement aux feuilles de l'arbre, lorsque l'assignement H_i est complètement défini. La raison de ce choix est que les scores E ne sont pas additifs : cela signifie que l'on ne peut pas donner un bon estimateur de $E(\{S_i, S_j\})$ à partir de $E(\{S_i\})$. En revanche, les scores effectifs des positionnements et des étendues des références comme des RVA sont utilisés dans tous les calculs de $g(H_i)$, puisqu'ils ne dépendent pas de l'assignement σ cherché. Finalement, pour l'évaluation d'une feuille de l'arbre, la fonction g est équivalente à la mise en œuvre de la fusion globale.

Recherche d'une solution initiale La performance de l'algorithme dépend en partie de sa capacité à définir une bonne solution initiale σ_0 . Dans le cas particulier de la reconnaissance de symboles par PSF, l'algorithme est déroulé successivement sur différents choix de références r_x, r_y et sur différents prototypes de PSF (pour la recherche du meilleur score global de correspondance). Le score de la meilleure solution trouvée pour tous les choix de références et tous les PSF est connu et l'algorithme se contente donc d'explorer les branches de l'arbre potentiellement meilleures que cette meilleure solution. De ce fait, on dispose à

4.2. *Patrons Structurels Flous* pour la représentation de symboles

toutes les exécutions de l'algorithme (sauf la première) d'un score faisable de bonne qualité. Par ailleurs, une partie du calcul de g repose sur les informations des scores des références qui sont déjà connus avant le déroulement de l'algorithme. Ces éléments participent donc à l'élagage des branches de l'arbre. En particulier un score de références très faible peut conduire rapidement à élaguer de grandes parties de l'arbre, voire tout l'arbre (c'est alors une *rejet implicite* de ce choix de références).

4.2.3.4 Synthèse

La stratégie de recherche de la meilleure segmentation d'un symbole par rapport à un PSF permet de mettre en œuvre la formalisation du calcul présenté à la section 4.2.2. Les points essentiels de cette mise en œuvre résident dans la méthode originale consistant à guider les hypothèses de segmentation des tracés à partir des modèles de leur positionnement qui sont développés sur les hypothèses de références choisies. Cette segmentation se fait donc en fonction du modèle appris et en respectant la forme particulière des références du symbole. Le guidage de la segmentation s'appuie aussi sur une vision globale de la structure du symbole, par opposition aux méthodes de détection de points singuliers qui sont plus locales. Il en découle une capacité à détecter des points de segmentation indétectables par l'analyse de forme du tracé. La fonction d'assignement des segments permet d'associer différents segments au même composant d'un PSF, ce qui autorise par exemple la gestion de tracés en pointillés. Par ailleurs, une souplesse est introduite afin de tolérer le calcul de scores de correspondance même en l'absence de certains composants ou en présence d'éléments surnuméraires.

4.2.4 Apprentissage des PSF

Nous avons présenté précédemment les principes de définition, la formalisation et l'exploitation des PSF pour le calcul du score de correspondance avec un symbole. Dans cette dernière section, nous nous intéressons au problème de leur apprentissage à partir de données. Pour une classe de symboles à modéliser, disposant d'un certain nombre d'exemples manuscrits de la classe, l'apprentissage du PSF consiste à :

- identifier et stabiliser un choix de « bonnes » références r_x et r_y ,
- définir le nombre de composants dans le PSF,
- apprendre tous les méta-modèles de positionnement spatial et d'étendue : positionnement mutuel des références, positionnement des composants par rapport aux deux références et positionnement des RVA par rapport aux deux références.

Nous introduisons dans cette partie un mécanisme d'apprentissage permettant de découvrir automatiquement tous ces éléments à partir des exemples d'apprentissage. L'apprentissage est effectué classe par classe et le modèle PSF résultant est donc intrinsèquement lié à la classe qu'il représente. Découvrir automatiquement les références et les éléments de la structure du PSF impose de procéder par un mécanisme d'optimisation itératif, car le problème à résoudre est un problème d'optimisation combinatoire difficile.

Le principe de l'algorithme itératif proposé s'inspire des méta-heuristiques utilisées pour les problèmes d'optimisation complexes : algorithmes évolutionnistes, recuit simulé... L'idée commune de ces techniques, qui est reprise dans notre stratégie d'apprentissage, consiste à chercher itérativement une *bonne* solution au problème d'optimisation à partir d'une

solution courante, tout en introduisant une certaine dose de hasard dans la sélection de la solution de l'étape suivante, afin d'éviter de tomber dans des optima locaux. Nous proposons de s'appuyer sur le schéma suivant (illustré graphiquement par la figure 4.12).

1. A chaque itération i , on dispose d'un modèle de PSF courant : \mathcal{T}_i .
2. Les exemples d'apprentissage sont *alignés* au mieux avec \mathcal{T}_i : par recherche de leur meilleures références et du meilleur assignement de leurs tracés aux composants de \mathcal{T}_i , comme présenté au paragraphe 4.2.3.3. Des *inversions* sont aléatoirement imposées sur certains exemples pour sélectionner la meilleure alternative de référence ou d'assignement. Ces alignements constituent alors des exemples segmentés du symbole à apprendre, c'est-à-dire dans lesquels les tracés sont « étiquetés » comme référence (r_x, r_y) ou composants (s_j) du modèle.
3. Une *sélection* aléatoire « élimine » des exemples avec une probabilité de 5%.
4. Les méta-modèles spatiaux et modèles d'étendue associés sont appris à partir de ces exemples segmentés. Ils constituent un nouveau PSF \mathcal{T}'_i .
5. \mathcal{T}'_i est appliqué sur les exemples d'une base de validation (par le procédé d'exploitation du paragraphe 4.2.3.3), afin de calculer pour chacun un score d'adéquation.
6. Une fonction h évalue la qualité globale du modèle par rapport à la base de validation en fonction des scores d'adéquation des exemples. C'est cette fonction h que l'on cherche à optimiser.
7. En fonction du score atteint selon cette fonction h et d'une variable aléatoire, on choisit de conserver ou non \mathcal{T}'_i à l'itération suivante, selon le principe du recuit simulé.

Le principe de ce schéma est de construire un modèle de PSF de plus en plus représentatif des données. Un bon modèle de PSF permet théoriquement de bien étiqueter les données d'apprentissage à l'étape d'alignement. Cela doit conduire à apprendre des méta-modèles spatiaux et d'étendue de meilleure qualité et donc à un modèle de PSF représentatif pour les données de la base de validation. Suite à un nombre prédéfini d'itérations, le PSF qui a permis d'atteindre le meilleur score selon h est conservé comme modèle de la classe de symboles.

Ce schéma est décliné en plusieurs variantes selon des étapes distinctes de l'apprentissage que nous détaillons ci-après. Tout d'abord, une étape d'initialisation est nécessaire pour construire un premier modèle de PSF. Ce premier modèle est partiel : nous ne nous intéressons dans l'initialisation qu'à identifier des références stables et identifiables dans chaque symbole de l'ensemble d'apprentissage et à apprendre le positionnement des RVA vis-à-vis de ces références. Dans un second temps, le modèle est enrichi par la modélisation du positionnement mutuel des deux références. Enfin, la troisième étape détermine le nombre de composants du PSF ainsi que leurs modèles de positionnement. Chaque étape est initialisée avec le meilleur PSF obtenu à l'étape précédente.

4.2.4.1 Initialisation

L'objectif de l'étape d'initialisation est de chercher à stabiliser le choix de références pour décrire le positionnement dans les symboles à modéliser. On dispose de N_A exemples d'apprentissage du symbole, notés $\{A_i\}_i$, et de N_V exemples de validation, notés $\{V_i\}_i$. Pour chacun des exemples des deux bases, on peut extraire une liste d'hypothèses de références

4.2. Patrons Structuels Flous pour la représentation de symboles

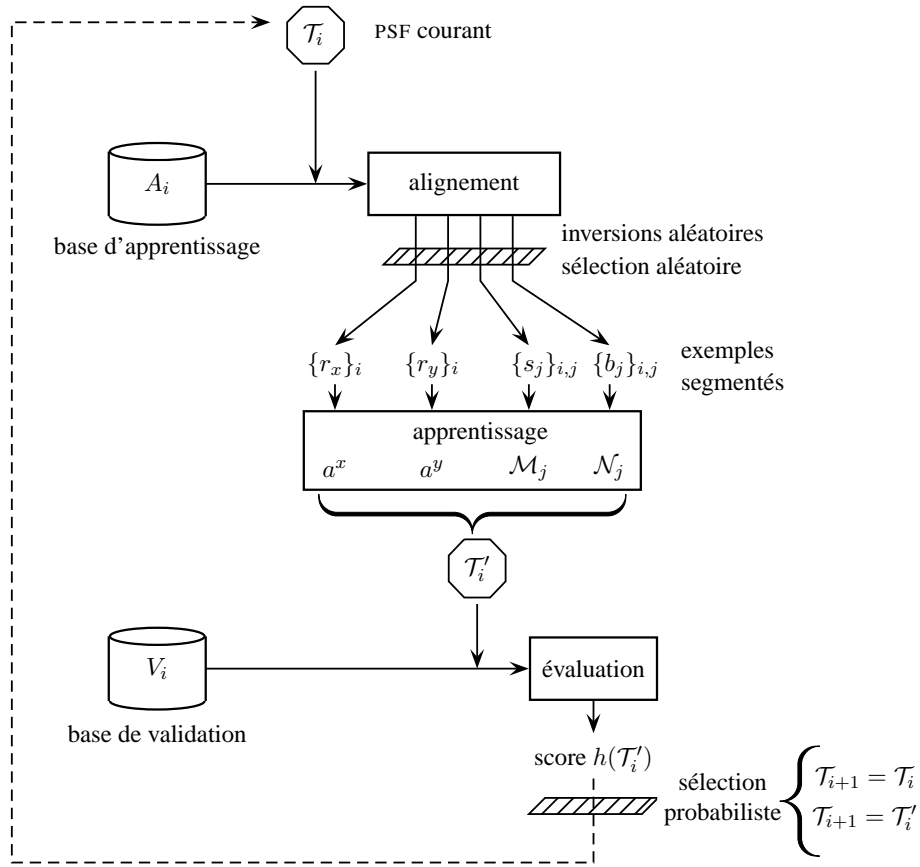


FIGURE 4.12: Schéma itératif pour l'apprentissage des PSF.

horizontales et verticales (d'après le procédé d'extraction décrit au paragraphe 4.2.3.1). Ces listes sont ordonnées d'après le critère de l'équation (4.1). Les hypothèses r_x et r_y d'un même symbole sont combinées pour former toutes les hypothèses de paires de références et on note $r_j^{A,i}$ la j -ième hypothèse de paire de références pour le i -ème exemple d'apprentissage (ensemble A). Ces paires sont ordonnées par le score combiné des deux éléments, de sorte que la paire $r_1^{A,i}$ est celle pour laquelle le score moyen de r_x et r_y est le plus grand.

Pour un nombre d'essais d'initialisation $nbInit$, on tire au hasard l'une des paires de références de l'un des objets d'apprentissage qui servira de *germe* pour la suite de l'apprentissage (la probabilité de tirer une paire de références est proportionnelle à son score, les références supposées meilleures sont ainsi plus probablement utilisées comme germe). A partir de ce germe $r_j^{A,i}$, on apprend un premier PSF $\mathcal{T}_{i,j}^0$ rudimentaire : il modélise simplement le positionnement des éléments de RVA par rapport à la paire de références $r_j^{A,i}$. Pour chaque bord $\{B_k\}_{k=1..4}$ de la boîte englobante du tracé du symbole A_i , on construit une paire de modèles spatiaux $\mathcal{N}_k = (b_x^k, b_y^k)$ décrivant son positionnement relativement à chacune des deux références de $r_j^{A,i}$.

Ce premier modèle de PSF $\mathcal{T}_{i,j}^0$ est utilisé pour sélectionner les paires de références *équivalentes* parmi les autres exemples d'apprentissage $A_j, j \neq i$. L'équivalence des paires de référé-

rences d'un symbole à l'autre repose sur les modèles de positionnement \mathcal{N}_k du PSF partiel appris. Le référentiel virtuel absolu joue ici un rôle crucial car il permet d'apparier facilement des hypothèses de références extraites de symboles différents. Comparativement aux autres informations de positionnement relatif dans un PSF classique (positionnement mutuel des références, positionnement des composants), le positionnement (et l'étendue) des RVA offre plus de stabilité, car les éléments B_k sont moins sujets à distorsions que les objets manuscrits et surtout car ils sont identifiables directement et sans ambiguïté (il n'y a pas besoin de les segmenter). Pour chaque exemple $A_{k,k \neq i}$, on cherche à maximiser :

$$s^0(A_k, \mathcal{T}_{i,j}^0, r^{A,k}), \quad (4.18)$$

où s^0 est la fonction d'adéquation d'un symbole avec un PSF partiel et $r^{A,k}$ dénote une paire de références de l'exemple. Ce score ne dépend que des modèles de positionnement des RVA et sa maximisation revient à choisir la paire d'hypothèses $r^{A,k}$ qui est *équivalente* à la paire $r_j^{A,i}$ (celle qui a servi de germe à l'apprentissage de $\mathcal{T}_{i,j}^0$). Les exemples d'apprentissage ainsi appariés avec le modèle partiel de PSF fournissent des paires de références supposément équivalentes. Un nouveau PSF plus représentatif de la base d'apprentissage peut alors être construit, cette fois avec les N_A exemples d'apprentissage. Sa définition reste limitée au positionnement des RVA par rapport aux références sélectionnées. On note ce nouveau PSF $\mathcal{T}_{i,j}^{0'}$.

Dans une seconde phase, ce PSF est évalué sur la base de validation, au moyen d'une fonction $h_0 : \mathcal{T}, \{V_i\}_i \rightarrow \mathbb{R}$.

Cette fonction h_0 , que l'on cherche à maximiser, doit exprimer la qualité de description offerte par un modèle \mathcal{T} , estimée sur la base de validation. Des scores d'adéquation $s^0(V_k, \mathcal{T}^{0'}, r^{V,k})$ élevés traduisent une bonne description des exemples de validation par le modèle. Il est pourtant nécessaire d'introduire un terme de régularisation dans la définition de h_0 , afin de contraindre les modèles de positionnement du PSF à une certaine précision. Sans cette contrainte, les modèles peuvent s'avérer très souples et très peu précis, voire totalement permissifs (admettant toute paire d'objets comme un bon représentant de la relation spatiale apprise). Ce comportement est à éviter afin d'aboutir à des modèles de PSF précis, c'est-à-dire qui n'admettent comme bons représentants que des exemples du symbole appris. Dans la fonction h_0 , nous chercherons donc à la fois à maximiser le niveau des scores s^0 , mais aussi à maximiser l'écart entre le score obtenu pour la meilleure paire de références par rapport au score obtenu sur la meilleure paire alternative de références. L'idée est que si le PSF est suffisamment précis (ce qui est souhaitable), il ne doit admettre qu'une seule paire d'hypothèses de références avec un bon score d'adéquation. On note :

$$g_1^k = s_0(V_k, \mathcal{T}^{0'}, r_1^{V,k}) = \langle (g_1^k)^+; (g_1^k)^- \rangle \quad (4.19)$$

le score bipolaire de l'exemple de validation V_k avec sa meilleure paire de références $r_1^{V,k}$ par rapport au modèle de PSF $\mathcal{T}^{0'}$. Le score obtenu avec la meilleure paire de références alternative $r_2^{V,k}$ est noté :

$$g_2^k = s_0(V_k, \mathcal{T}^{0'}, r_2^{V,k}) = \langle (g_2^k)^+; (g_2^k)^- \rangle. \quad (4.20)$$

La fonction h_0 d'évaluation d'un PSF par rapport à la base de validation V est définie par :

4.2. Patrons Structurels Flous pour la représentation de symboles

$$h_0(\mathcal{T}^{0'}, \{V_k\}_{k=1..N_V}) = \frac{1}{N_V} \left(\frac{\sum_k (g_1^k)^+}{\sum_k (g_1^k)^+ + \sum_k (g_2^k)^+} \right) \left(\frac{\epsilon + \sum_k (g_2^k)^-}{\epsilon + \sum_k (g_1^k)^- + \sum_k (g_2^k)^-} \right), \quad (4.21)$$

où $\epsilon > 0$, par exemple $\epsilon = 10^{-3}$.

Le processus présenté ci-dessus est répété *nbInit* fois, en tirant à chaque fois un germe différent, ce qui conduit à autant de modèles de PSF. On retient à l'issue de cette étape le modèle \mathcal{T}^0 qui maximise le score calculé par h_0 . Ce modèle repose donc sur le meilleur choix de références au sens de h_0 en considérant simplement l'adéquation des références avec les éléments de RVA. Ce choix pourra être remis en cause dans les étapes suivantes, mais il permet d'offrir une bonne initialisation pour la seconde étape, où plus de paramètres seront pris en compte dans les modèles de PSF.

4.2.4.2 Itérations pour la recherche de références

Cette deuxième étape vise à renforcer le choix de bonnes paires de références en intégrant dans l'apprentissage le *positionnement mutuel* des références. L'algorithme est cette fois-ci itératif : il vise à maximiser une fonction h_1 en mettant à jour un modèle de PSF que l'on cherche à améliorer progressivement (selon le principe illustré par le schéma de la figure 4.12). L'initialisation de l'algorithme itératif se fait en prenant le meilleur modèle \mathcal{T}^0 issu de l'étape précédente. Il constitue le premier modèle courant \mathcal{T}_0 . Les étapes suivantes de l'algorithme sont alors répétées pour chaque itération.

1. A partir du modèle courant \mathcal{T}_i , on cherche la meilleure paire de références pour chaque exemple de la base d'apprentissage A . Des *inversions* sont provoquées aléatoirement pour choisir la meilleure paire alternative (le tirage est pondéré par l'inverse de la différence de score entre les deux hypothèses de références).
2. Un filtre de sélection élimine des exemples avec une probabilité de 5%.
3. A l'aide des paires de références retenues, on apprend un nouveau PSF \mathcal{T}'_i partiel, modélisant le positionnement des RVA par rapport aux références ainsi que le positionnement mutuel des références.
4. Le modèle \mathcal{T}'_i est évalué sur la base de validation, au moyen d'une fonction h_1 similaire à la fonction h_0 (définie à l'équation (4.21)). La différence est que le score d'adéquation d'un exemple avec le modèle est calculé en tenant compte du positionnement mutuel des références et de leurs scores d'étendue, en plus du positionnement et de l'étendue des RVA.
5. En fonction du score issu de h , on décide de conserver ou non \mathcal{T}'_i pour l'itération suivante. Ce choix se fait en fonction de l'écart entre le score $h_1(\mathcal{T}'_i)$ et le meilleur score $h_1(\mathcal{T}^*)$ obtenu au cours des itérations ainsi qu'en fonction d'une variable de température T_i qui décroît au fur et à mesure des itérations selon une loi de probabilité exponentielle, d'après le principe de recuit simulé. On a donc pour l'itération $i + 1$:

$$\begin{cases} \mathcal{T}_{i+1} = \mathcal{T}'_i & \text{si } h_1(\mathcal{T}'_i) > h_1(\mathcal{T}^*) \\ \mathcal{T}_{i+1} = \mathcal{T}'_i & \text{si } X < e^{(h_1(\mathcal{T}^*) - h_1(\mathcal{T}'_i))/T_i} \\ \mathcal{T}_{i+1} = \mathcal{T}_i & \text{sinon,} \end{cases} \quad (4.22)$$

avec X un réel tiré aléatoirement entre 0 et 1. Ce schéma permet d'explorer l'espace de solutions en tolérant la considération de PSF moins bons que la meilleure hypothèse trouvée, avec une probabilité qui diminue au cours des itérations. La variable de température T_i est définie par $T_i = T_0 e^{\lambda i}$, avec T_0 la température initiale et λ la vitesse de refroidissement (par exemple $\lambda = 0.99$).

Dans le cas où le PSF n'est pas modifié ($\mathcal{T}_{i+1} = \mathcal{T}_i$), l'itération $i+1$ commence dans les mêmes conditions que l'itération i . Ce sont alors les *inversions* et la *sélection* aléatoires mises en place au début de l'itération (étape 1 et 2) qui permettent de modifier les données d'apprentissage à l'itération $i+1$ et de mener ainsi à la construction d'une solution \mathcal{T}'_{i+1} différente.

Cette étape de l'apprentissage s'arrête au terme d'un nombre prédéfini d'itérations. Le meilleur modèle de PSF \mathcal{T}^* obtenu au cours des itérations est retenu pour l'étape finale qui a pour but d'apprendre et de stabiliser le PSF complet.

4.2.4.3 Itérations pour l'apprentissage du PSF complet

A ce stade, les deux étapes précédentes ont permis de stabiliser le choix des références. Il est donc possible de compléter le PSF en y intégrant les autres composants de tracé. L'algorithme utilisé respecte le même principe de déroulement et est initialisé par le meilleur modèle construit à l'étape précédente. En utilisant ce modèle, on peut sélectionner les meilleures références pour chaque objet d'apprentissage. Le reste des tracés des objets forment des éléments en nombre variable en fonction des exemples et de leur segmentation naturelle. Nous choisissons d'initialiser le nombre de composants avec le nombre le plus fréquent de tracés naturels constatés dans les exemples d'apprentissage. Ces éléments sont appariés sur tous les exemples d'apprentissage, puis un modèle de leur positionnement est appris pour chacun. Ces modèles spatiaux sont intégrés au PSF qui constitue donc une première version d'un modèle de symbole complet.

Le même algorithme de recuit simulé est déroulé itérativement, après avoir réinitialisé la variable de température à sa valeur T_0 . Pour la i -ème itération, les étapes de l'algorithme sont les suivantes.

1. A partir du modèle courant (complet) \mathcal{T}_i , on calcule le score d'adéquation de chaque exemple d'apprentissage, pour le meilleur choix de références et d'assignement σ . Des *inversions* forcent la sélection de références ou d'assignement alternatifs. Pour un exemple donné, la fonction σ retenue fournit une bonne segmentation du symbole. Un segment peut donc être étiqueté par le composant du PSF auquel il est associé selon l'assignement σ .
2. Une *sélection* élimine aléatoirement des exemples d'apprentissage avec une probabilité de 5%.
3. Les éléments des symboles d'apprentissage retenus sont utilisés pour apprendre un nouveau PSF complet : \mathcal{T}'_i . En utilisant l'étiquetage retenu pour les segments, on dispose d'éléments pour apprendre le modèle de positionnement de chaque composant par rapport aux références.
4. Le modèle de \mathcal{T}'_i est évalué par une fonction h_2 sur la base de validation. Cette fonction tient compte du score complet d'évaluation pour chaque exemple.

4.3. Validation qualitative

5. Le principe de recuit simulé est utilisé pour décider ou non de conserver \mathcal{T}_i' à l'étape suivante (de façon exactement similaire à celui décrit précédemment).

Finalement, à l'issue d'un nombre prédéfini d'étapes, on retient le modèle de PSF qui a permis d'obtenir la valeur maximale de la fonction h_2 sur la base de validation. Il s'agit du modèle qui représente au mieux la classe de symboles. Il pourra servir notamment de prototype de la classe de symboles pour l'exploitation en reconnaissance.

4.3 Validation qualitative

Nous avons introduit le formalisme original des PSF pour représenter des symboles manuscrits à partir de l'information sur leur structure et sur le positionnement relatif de leurs éléments internes. Un algorithme d'apprentissage a été mis au point, ce qui permet de construire ces représentations structurelles automatiquement à partir de données. Puisque les PSF sont une composition de méta-modèles spatiaux, il est possible de les visualiser par des images représentant ces modèles spatiaux appliqués sur les bonnes références. Nous utilisons ici cette propriété pour évaluer qualitativement le résultat de l'apprentissage de PSF pour différentes natures de symboles et différentes situations de segmentation. Les images qui illustrent cette partie représentent des PSF appris automatiquement selon l'algorithme décrit à la section 4.2.4 et développés sur des tracés par le procédé d'exploitation décrit à la section 4.2.3.

Dans le chapitre dédié aux expérimentations (chapitre 5), nous reviendrons sur les capacités des PSF en évaluant leur capacité à reconnaître ces symboles. Le but de la présente partie est plutôt de montrer quelques exemples du résultat de l'apprentissage de PSF pour différents types de symboles.

Intéressons nous d'abord aux sélections des références de positionnement dans la description des symboles. Puisque la formalisation des PSF repose avant tout sur la définition d'une paire de références qui servent de support à la description du positionnement, il est important de vérifier si l'algorithme d'apprentissage est effectivement capable de faire émerger une paire de références stable et satisfaisante dans tous les cas. Nous rappelons que selon le principe exposé à la section 4.2.1, nous avons choisi de favoriser la recherche d'éléments de références perceptivement saillants (objets de grande dimension, couvrant tout le symbole selon l'une de ses directions). Le choix automatique des références par l'algorithme d'apprentissage peut donc être vérifié qualitativement sur quelques exemples.

La figure 4.13 contient différents symboles où sont mis en évidence les deux éléments de tracés automatiquement déterminés pour constituer les références de positionnement. Les éléments de tracés en couleur (en bleu pour l'élément de référence *horizontale* et en violet pour l'élément de référence *verticale*) sont les meilleures références au sens du modèle de PSF appris pour la classe du symbole.

Les différents exemples de la figure sont des caractères chinois (a,b,c) et des pictogrammes (d,e,f). Leur complexité est variable et la présence d'éléments structurels prégnants est plus ou moins évidente selon les cas. Les symboles (b),(c) et (d), qui sont les plus complexes graphiquement, sont structurés autour de longs tracés rectilignes horizontaux et verticaux, qu'il paraît intuitif de considérer comme références de positionnement. C'est particulièrement évident pour le cas (c), où les deux tracés qui s'intersectent au centre

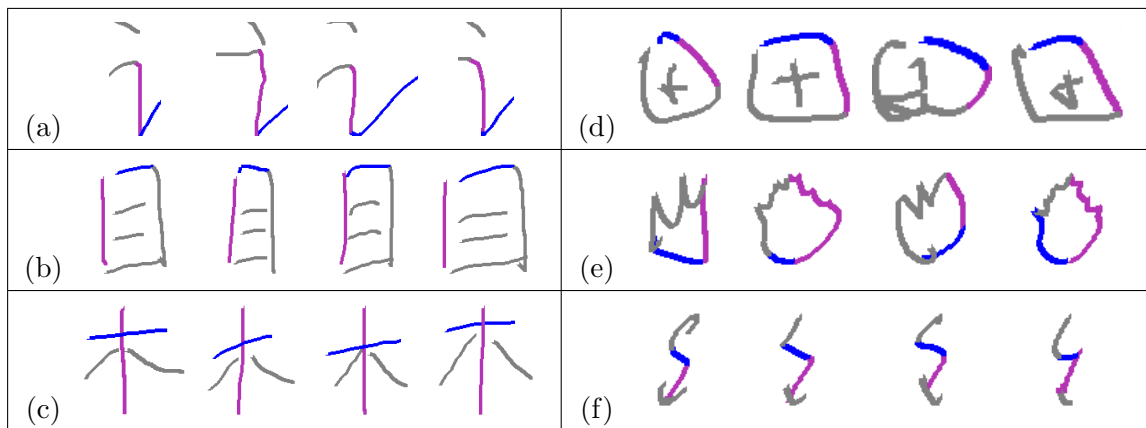


FIGURE 4.13: Exemples de références établies pour des symboles : caractères chinois (a,b,c) et pictogrammes (d,e,f). L'élément de référence horizontale est représenté en bleu, l'élément de référence verticale en violet.

du caractère sont naturellement sélectionnés par l'algorithme. Les exemples (b) et (c) démontrent qu'en cas de plusieurs possibilités de références (plusieurs segments horizontaux et verticaux sont de bons candidats), l'algorithme est capable de choisir l'une d'entre elles et de la retrouver correctement dans chaque symbole. Les cas (a),(e) et (f) correspondent à des situations où il est moins évident de choisir des portions de tracé comme références (pas d'éléments particulièrement prégnants). Le procédé d'apprentissage parvient néanmoins à converger vers une solution à chaque fois avec une certaine stabilité malgré les variantes de segmentation et de distorsions de tracés.

Le second aspect des PSF pouvant être évalué qualitativement concerne les méta-modèles spatiaux qui ont été appris pour modéliser les positions des éléments de la structure des symboles. Ils peuvent être visualisés un à un, en les développant sur les références de positionnement déterminées automatiquement. La figure 4.14 illustre ainsi plusieurs PSF appris sous la forme de leurs modèles spatiaux. Les modèles spatiaux représentés sont à mettre en correspondance avec les composantes formelles du PSF tel qu'il a été défini à la section 4.2.2 (voir en particulier la figure 4.4).

Cette figure expose quatre PSF modélisant des symboles chinois et des symboles architecturaux et développés chacun sur deux exemples. Les modèles spatiaux qui constituent le PSF sont représentés un à un par des images distinctes.

Sur la première ligne, on a par exemple un symbole chinois \mathcal{S} dont les références sont mises en évidence, ainsi que les modèles spatiaux correspondant à ses éléments de structure selon le PSF. Le premier élément visualisé résulte de la combinaison (union bipolaire) des modèles de positionnement des quatre éléments de la RVA par rapport aux deux références. Les deux éléments suivants sont les modèles de positionnement relatif et mutuel des références r_x et r_y . Enfin, quatre modèles spatiaux décrivent la position de quatre composants (\mathcal{M}_i) de structure par rapport aux références. Sur chaque image, la partie de tracé colorée en jaune correspond à la partie affectée au composant selon la meilleure correspondance trouvée. Le même PSF est développé sur un autre représentant de la même classe de symboles sur la seconde ligne. Ce second symbole met en évidence le bon comportement de la

4.3. Validation qualitative

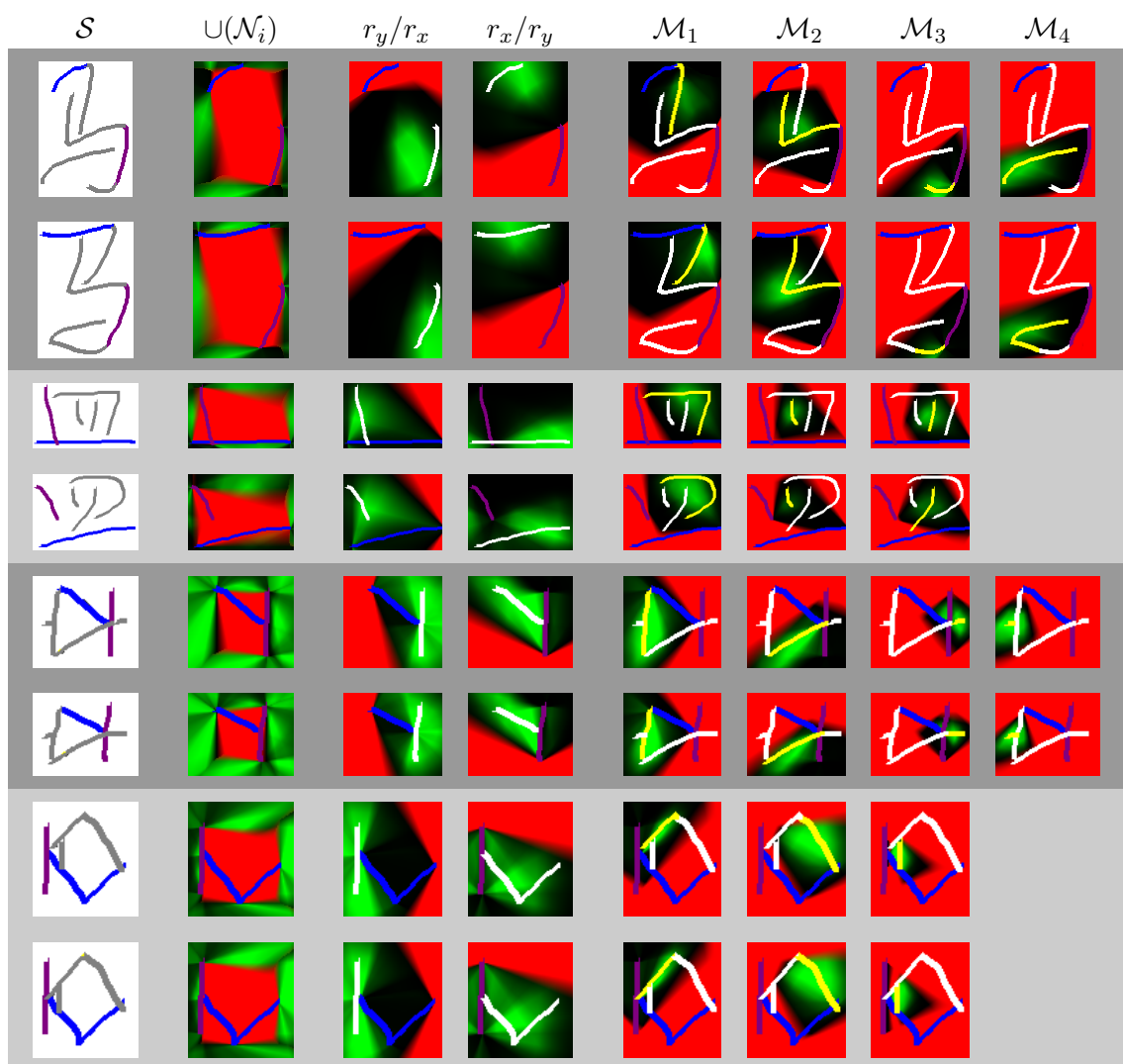


FIGURE 4.14: Les modèles spatiaux constituant quatre PSF sont présentés dans cette figure. Chaque ligne représente, pour un symbole (\mathcal{S}), tous les modèles spatiaux composant le PSF associé. On trouve parmi eux les modèles de RVA (position de la boîte englobante par rapport aux références) : ils sont ici fusionnés dans une seule image $\cup(\mathcal{N}_i)$. Les autres modèles décrivent le positionnement relatif des références (r_y/r_x et r_x/r_y) et les modèles pour les autres composants du PSF : \mathcal{M}_i . Chaque PSF est exploité sur deux tracés différents. Les modèles spatiaux sont bipolaires avec intégration de la distance *directionnelle*.

stratégie de segmentation guidée par les modèles spatiaux, qui permet de segmenter le tracé pour en affecter différentes portions à deux composants (voir les composants \mathcal{M}_3 et \mathcal{M}_4).

Le second PSF présenté dans cette figure comporte trois composants. Son application à deux tracés différents permet d'apprécier la bonne adaptation des modèles spatiaux face à des distorsions importantes dans les formes des tracés individuels.

Le troisième et le quatrième PSF illustrés représentent des symboles architecturaux com-

prenant quatre et trois composants. Les modèles spatiaux représentés confirment la bonne description de cet autre type de symboles offerte par les PSF appris.

4.4 Synthèse

Nous avons introduit dans cette partie le concept de Patron Structurel Flou, une représentation structurelle pour les symboles manuscrits. Cette représentation est construite par un assemblage de méta-modèles spatiaux, qui permettent de décrire la richesse de l'information de structure et de positionnement relatif des éléments.

Vérification de l'importance de l'information spatiale La possibilité de définir des modèles de structure réalistes pour des symboles manuscrits au seul moyen de l'information de structure et de positionnement relatif de leurs éléments atteste de l'importance de cette nature d'informations dans la perception de ces symboles. Une réflexion approfondie a été menée pour établir un référentiel de positionnement cohérent avec le principe de définition des PSF. Des références de positionnement sont issues de tracés réels et permettent d'assurer une bonne qualité de description du positionnement de tous les autres éléments de la structure. De plus, ces références se positionnent d'elles-mêmes par rapport aux éléments de la boîte englobante des symboles, ce qui permet une plus grande stabilité de leur extraction et facilite notamment l'apprentissage automatique des modèles.

La formalisation du schéma de fusion global par un raisonnement bipolaire flou permet une prise en compte homogène des différents éléments de la structure, en combinant pour chacun une description de son positionnement et de son étendue par rapport aux modèles. Le schéma d'exploitation présenté permet de calculer des correspondances entre des tracés et des PSF, en tolérant toutes variations dans l'ordre et le nombre des tracés. La stratégie de segmentation utilisée s'appuie sur la modélisation apprise du positionnement relatif des objets pour formuler des hypothèses de segmentation adaptées à l'hypothèse de PSF considéré, en tenant compte du contexte global et des formes particulières des références. Enfin, un algorithme d'apprentissage permet de définir des PSF automatiquement à partir de données, sans nécessiter de connaissance sur la segmentation interne des symboles d'apprentissage.

Validation de l'intérêt applicatif des méta-modèles spatiaux Cette application au contexte des symboles manuscrits a été l'occasion d'exploiter en profondeur les avantages des méta-modèles spatiaux et de la théorie sous-jacente du raisonnement spatial au sein de l'espace image. Le choix de combiner deux références distinctes pour décrire le positionnement relatif des objets correspond à une exploitation plus avancée du principe de combinaison de modèles spatiaux identifié par Logan. Nous avons montré qu'une bonne stratégie de sélection des références permet en outre de maximiser l'intérêt de cette combinaison, et avons abouti à faire reposer la description sur deux références aussi orthogonales que possible. La stratégie de segmentation proposée dans cette approche consiste à exploiter directement la capacité de prédiction offerte par les méta-modèles spatiaux appris. La mise en œuvre des PSF témoigne donc non seulement de la qualité qu'offrent les méta-modèles spatiaux pour la description du positionnement et de l'étendue entre des objets qui présentent des relations spatiales non-linguistiques, mais aussi de leur fonctionnalité de prédiction qui peut

4.4. Synthèse

être avantageusement utilisée pour résoudre des problèmes d'analyse structurelle de tracés manuscrits.

Une rapide étude qualitative de la modélisation a permis de confirmer la capacité de représentation offerte par les PSF ainsi que la convergence de l'algorithme d'apprentissage vers des représentations structurelles visuellement satisfaisantes pour quelques exemples. La bonne adaptation aux formes permise par les modèles spatiaux ainsi que la stratégie de segmentation guidée par ces modèles est visiblement à même de modéliser les variantes de tracés constatées dans les différentes bases de symboles. Dans le chapitre 5, il sera présenté une expérimentation visant à éprouver quantitativement la démarche sur une tâche de reconnaissance de symboles structurés, dans le but de sonder dans quelle mesure l'information spatiale décrite dans un PSF est suffisante ou non pour reconnaître automatiquement différentes natures de symboles. Les résultats permettront d'appuyer nos conclusions sur l'intérêt de l'information spatiale ainsi que la qualité de sa modélisation par les méta-modèles spatiaux, mais aussi de mettre en évidence les limites de cette représentation.

Expérimentations

Dans ce chapitre, nous décrivons les expérimentations qui ont été conduites pour valider les différentes contributions présentées dans cette thèse et discutons des résultats obtenus.

La première partie se concentre sur les expérimentations menées pour démontrer la qualité de la représentation du positionnement relatif d'objets manuscrits fournie par les méta-modèles spatiaux introduits au chapitre 3. La description par méta-modèles est comparée à des descripteurs classiques de positionnement relatif, sur une tâche de classification de relations spatiales appliquée à des données de différentes natures.

La seconde partie expérimentale s'intéresse au problème de reconnaissance de symboles au moyen de la représentation structurale présentée au chapitre 4. L'objectif de cette validation est de démontrer la capacité de la représentation par PSF à reconnaître des symboles de différentes natures et ainsi de souligner à la fois la valeur intrinsèque de l'information spatiale et la contribution des méta-modèles spatiaux pour la représentation de cette information au sein d'un système structurel de reconnaissance de tracés manuscrits.

5.1 Classification de relations spatiales entre objets manuscrits

Dans cette première partie, nous expérimentons la performance intrinsèque des méta-modèles de positionnement que nous avons introduits dans le chapitre 3 pour décrire des relations spatiales entre des objets manuscrits. Ces expérimentations sont faites en dehors de tout système structurel d'analyse de tracés et sans tenir compte de problèmes de segmentation. Elles portent uniquement sur la tâche d'évaluation du positionnement relatif de deux objets manuscrits, l'un étant étiqueté comme la référence, l'autre comme l'argument de la relation spatiale. Dans les tests présentés ici, le but est de classifier une relation spatiale parmi un ensemble de classes prédéfinies, au moyen de différents jeux de caractéristiques et d'un moteur de reconnaissance statistique standard.

Nous présentons d'abord en détail les différents jeux de données utilisés pour nos expérimentations. Ensuite, nous exposons le protocole expérimental général, avant de présenter et d'analyser les résultats des différents tests.

5.1.1 Jeux de données

Nous avons sélectionné plusieurs jeux de données afin d'éprouver la capacité des modèles de positionnement introduits à traiter différentes catégories de relations spatiales et différentes natures d'objets. Ces jeux couvrent une large palette de besoins de positionnement relatif qui se posent pour l'analyse de documents manuscrits. Nous introduisons ici les deux premiers jeux de données sélectionnés, sur lesquels seront évaluées toutes les variantes des méta-modèles spatiaux et qui serviront de base pour la comparaison avec d'autres techniques de description du positionnement relatif. Dans le paragraphe 5.1.4, nous élargirons les expérimentations à d'autres natures de données pour mettre en avant la bonne généralité des méta-modèles.

5.1.1.1 Accents et gestes d'édition en-ligne

La base IME-OnDB est une base de gestes manuscrits en-ligne. Celle-ci comporte 18 classes de gestes graphiques saisis par 15 scripteurs, chacun ayant tracé entre 15 et 20 exemplaires de chaque geste. Elle contient un total de 5525 enregistrements, un enregistrement étant une paire d'objets manuscrits collectés dans le contexte d'une application de saisie et d'édition de texte sur un dispositif de type assistant personnel (PDA). Les gestes correspondent à de purs gestes d'édition graphique (suppression d'un caractère, espace, retour chariot...) ou bien à la saisie de signes diacritiques (accent, cédille, apostrophe) ou encore de ponctuation (virgule, point). Tous les gestes ont été saisis *en contexte*, c'est-à-dire par rapport à un caractère de référence associé (une lettre de l'alphabet). Le tableau 5.1 présente 16 classes parmi ces gestes graphiques, avec à chaque fois la lettre de référence en rouge et le geste en bleu. La base de données est disponible publiquement sur le site de l'équipe IMADOC¹.

















Nature des objets Dans ce jeu de données, les références sont constituées d'un seul tracé qui correspond à une lettre de l'alphabet latin. Les objets arguments (les gestes d'édition ou les signes diacritiques) sont également constitués d'un seul tracé simple. Puisque plusieurs de ces gestes ont une forme très semblable (par exemple la *virgule* et l'*apostrophe*), la modélisation de leur positionnement relativement à la référence est nécessaire pour une bonne reconnaissance. Il s'agit donc d'un jeu de données approprié pour une évaluation et une comparaison de caractéristiques de positionnement relatif entre les objets. Pourtant, la description de forme des objets argument est également nécessaire, puisque des classes comme *accent aigu*, *accent grave*, *accent circonflexe*, *tréma* et *suppression d'accent* reposent vraisemblablement sur des relations spatiales très similaires.

Variabilité des références Une particularité de cette base de données est que pour une classe donnée de gestes à reconnaître (et donc pour une classe de relation spatiale),

1. <http://www.irisa.fr/imadoc/web/>

5.1. Classification de relations spatiales entre objets manuscrits

TABLE 5.1: Représentation de 16 classes de la base IME-OnDB.

 accent aigu	 accent grave	 accent circonflexe	 tréma
 suppression d'accent	 cédille	 suppression de cédille	 changement de casse
 suppression du caractère	 suppression du dernier caractère	 retour chariot	 insertion
 espace	 apostrophe	 virgule	 point

les références peuvent varier de façon importante. En effet, différentes lettres de l'alphabet apparaissent comme objet de référence pour chaque classe. Ainsi, les exemples de la classe *accent circonflexe* admettent pour références les lettres a, e, i, o, u ; ceux de la classe *apostrophe* admettent les lettres c, d, j, l, m, n, s, t ; tandis que la classe *cédille* n'admet pour référence que la lettre c . La variabilité des formes des références est donc accrue pour certaines classes et le besoin de prise en compte de ces formes pour décrire les relations spatiales est d'autant plus important dans ce jeu de données.

Nature des relations spatiales Les natures de relations spatiales présentes dans la base sont assez variées. La plupart des relations mettent en jeu des objets distants, déconnectés, mais certaines concernent des objets adjacents (par exemple la relation *cédille*), qui s'intersectent (*suppression de caractère*) ou sont inclus l'un dans l'autre (*changement de casse*). Les positions relatives couvrent toutes les directions du plan et nous ferons le choix de considérer les quatre directions principales du plan pour décrire ce vocabulaire de relations spatiales dans notre modélisation (cela correspond donc au choix général décrit à la section 3.2.2). Les objets sont tout de même principalement à droite de leur référence (*suppression du dernier caractère*, *retour chariot*, *insertion*, *espace*, *point*, *virgule*, *apostrophe*), au-dessus (*accent aigu*, *accent grave*, *accent circonflexe*, *tréma*, *suppression d'accent*, *apostrophe*), et en-dessous (*cédille*, *suppression de cédille*).

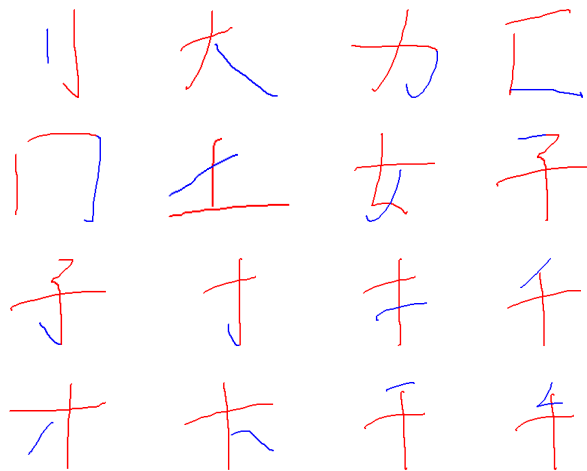
5.1.1.2 Primitives de caractères chinois

Les caractères chinois sont des formes structurées, qui peuvent être décrites comme une composition de tracés primitifs organisés dans un espace de forme carrée, en deux dimensions. Ils peuvent être constitués d'un nombre important de ces primitives (jusqu'à

plus d'une vingtaine) et offrent donc un vocabulaire de relations spatiales inter-primitives très riche.

A partir d'une base de données de caractères chinois en-ligne saisis au sein du *Laboratoire Franco Chinois d'Informatique, d'Automatique et de Mathématiques Appliquées* (LIAMA) de Pékin, nous avons constitué un jeu de relations inter-primitives extraites de caractères. La base originale complète contient 60 exemples pour les 6763 caractères définis par le standard GB2312 pour le chinois simplifié. 60 scripteurs ont participé à la collecte, chacun écrivant une seule fois chaque caractère. Davantage de détails sur cette base sont donnés dans les travaux de Ma & Liu sur la reconnaissance du chinois [ML09]. Pour notre expérimentation, nous avons défini 17 classes de paires d'objets extraits de caractères simples. Pour chaque classe, nous isolons environ 120 enregistrements (2 par scripteur) de chaque classe. Un enregistrement est une paire d'objets, l'un étiqueté comme objet de référence, l'autre comme argument de la relation. Dans la suite, nous faisons référence à ce jeu de données sous le nom de HCC-OnDB. Le tableau 5.2 représente des exemples de 16 classes de cette base de données, toujours avec les objets de référence en rouge et les objets argument en bleu.

TABLE 5.2: Représentation de 16 classes de la base HCC-OnDB.



Nature des objets Les objets de référence sont faits de deux tracés primitifs, tandis que les arguments sont toujours faits d'un seul tracé, de forme simple. Le tableau 5.2 illustre les classes retenues dans notre jeu de données (les références sont représentées en rouge, les arguments en bleu). Contrairement à la base IME-OnDB, les objets de référence des enregistrements pour une même classe sont tous des représentants du même objet. Seules les différences entre les styles des scripteurs et la variabilité naturelle des tracés manuscrits interviennent dans ce jeu de données.

Nature des relations spatiales Les relations spatiales décrites sont de multiples natures : les objets sont tantôt distants, tantôt adjacents (certains sont issus d'un même tracé

5.1. Classification de relations spatiales entre objets manuscrits

naturel ayant été segmenté) et certains présentent des intersections. Les positions relatives des arguments par rapport aux références couvrent les quatre directions cardinales (haut, bas, gauche, droite). Les boîtes englobantes des tracés de référence et de l’argument se recouvrent largement en général.

5.1.2 Protocole expérimental

Chaque base d’exemples est constituée de paires d’objets segmentés (une référence et un argument), positionnés dans le même référentiel et dont la relation spatiale est à reconnaître. Tous les tests menés dans cette section consistent à reconnaître automatiquement la classe d’un nouvel exemple à partir de caractéristiques extraites décrivant le positionnement relatif des objets (ainsi que des informations sur la forme de l’objet argument). Un classifieur de type *Support Vector Machine* (SVM) est utilisé comme moteur de reconnaissance pour attribuer une classe à un exemple décrit à partir des caractéristiques extraites. La comparaison de la qualité offerte par différents espaces de caractéristiques sera faite en fonction des performances de reconnaissance obtenues en sortie du système.

Nous évaluerons les différentes contributions apportées par les variantes des méta-modèles au moyen des deux bases IME-OnDB et HCC-OnDB en mettant en œuvre un processus de validation croisée omni-scripteur. Ces contributions seront également évaluées en comparaison avec des descripteurs de positionnement relatifs classiquement utilisés pour le positionnement relatif d’objets.

5.1.2.1 Processus de validation omni-scripteur

Les bases IME-OnDB et HCC-OnDB contiennent l’information sur l’identité des scripteurs ayant réalisé les saisies, ce qui permet de mettre en place un processus de validation dit omni-scripteur. Dans ce cadre, nous distinguons les scripteurs de test des autres scripteurs. Les données sur lesquelles sont évaluées les qualités de description des différentes méthodes proviennent de scripteurs différents de ceux ayant produit les données qui permettent d’entraîner les modèles. Ce cadre de test correspond au cadre le plus général et permet d’éprouver le pouvoir de généralisation d’un espace de représentation sans sous-estimer la variabilité inter-scripteur, comme c’est le risque lorsque les données sont divisées aléatoirement en données d’apprentissage et de test.

Pour chacune des deux bases, on adoptera la méthode de validation croisée omni-scripteur suivante. La base B globale est découpée en 15 groupes $B = \cup_i B_i, i=1..15$ en prenant soin de regrouper toutes les données d’un même scripteur dans le même groupe. Dans le cas de IME-OnDB, chaque groupe contient exactement les données d’un scripteur, tandis que dans le cas de HCC-OnDB, chaque groupe contient les données de 4 scripteurs (parmi les 60 scripteurs de la base). Le nombre d’exemples dans chaque groupe est environ de 370 pour la base IME-OnDB et de 70 pour la base HCC-OnDB. Pour une expérimentation (c’est-à-dire pour un espace de représentation à évaluer), nous procéderons à 15 cycles d’apprentissage du classifieur, en laissant à chaque fois de côté un groupe B_i différent pour l’apprentissage et en l’utilisant pour évaluer le classifieur appris sur les 14 autres groupes. C’est le score de reconnaissance moyen calculé sur les 15 expérimentations qui sera présenté dans chaque cas comme résumé de la qualité d’une méthode de description.

5.1.2.2 Apprentissage et optimisation du classifieur

Le classifieur retenu pour évaluer la qualité de la description offerte par chaque espace de représentation est un SVM muni d'un noyau gaussien. Les deux paramètres du SVM C (variable de coût des erreurs en apprentissage) et γ (paramètre de la fonction gaussienne) sont optimisés par une recherche en grille faite indépendamment pour chaque apprentissage. L'optimisation consiste à chercher la meilleure valeur conjointe de ces deux paramètres au moyen d'une validation croisée à 10 groupes constitués aléatoirement parmi la base d'apprentissage. La valeur du paramètre de coût, C , est cherchée parmi la gamme de valeurs sur l'échelle logarithmique $10^{-1}..10^4$, et la valeur du paramètre γ est optimisée dans la gamme $10^{-4}..10^1$. Le critère optimisé au cours de cette recherche en grille est le taux de reconnaissance sur la base de validation (un dixième de la base d'apprentissage). Les paramètres retenus pour l'apprentissage du classifieur final sont donc ceux qui maximisent le taux moyen de reconnaissance sur la validation croisée à 10 groupes.

5.1.2.3 Significativité des comparaisons

Le but de ces expérimentations étant de démontrer la supériorité de certaines méthodes de description par rapport à d'autres pour la tâche de classification automatique d'exemples de relations spatiales, il est nécessaire d'établir un protocole permettant de mesurer la significativité des comparaisons effectuées entre les différentes expérimentations.

La significativité statistique de la comparaison de différents scores de reconnaissance obtenus par deux méthodes de description est renforcée lorsque le nombre de jeux de données comparés augmente. Nous utiliserons cette propriété pour asseoir au mieux la significativité de nos comparaisons. Dans notre protocole expérimental, les 30 taux de reconnaissance obtenus pour chaque méthode de description (15 tests omni-scripteur pour chaque base de données) peuvent être utilisés pour renforcer la validité des comparaisons.

Plutôt que de chercher à mesurer l'écart type entre les taux de reconnaissance obtenus par deux méthodes sur les 30 jeux de données, nous proposons de procéder à un test statistique de type *test t* (ou de Student). Une simple mesure de l'écart-type serait peu informative dans notre contexte omni-scripteur car certaines expérimentations sont intrinsèquement plus difficiles que d'autres. En effet, des styles de scripteurs particuliers peuvent rendre spécialement difficile la reconnaissance de certains groupes de données B_i . L'écart type des taux de reconnaissance obtenus par une méthode de description sur les 30 jeux de données sera inévitablement affecté par ces disparités.

Le principe du *test t* est de vérifier l'hypothèse de la supériorité de la moyenne d'une loi de probabilité par rapport à une autre. Une statistique intermédiaire peut être calculée, tenant compte de l'écart de performance entre les deux méthodes comparées sur chaque jeu de données, du nombre de jeux de données et de l'écart type sur l'écart de performance entre les deux méthodes. Cette statistique peut ensuite être comparée à la loi de Student tabulée, pour décider, avec un niveau de confiance prédéfini, si on peut affirmer que la méthode A est significativement meilleure que la méthode B . Le nombre de mesures (c'est-à-dire le nombre de jeux de données) intervenant dans la comparaison détermine le nombre de degrés de libertés à considérer pour la loi de Student à appliquer.

5.1. Classification de relations spatiales entre objets manuscrits

TABLE 5.3: Caractéristiques utilisées dans les expérimentations.

type	désignation	nombre	description
forme	f	9	caractéristiques géométriques de forme
échelle	e	3	ratio boîtes englobantes
position	b	9	caractéristiques boîtes englobantes et centres
position	h	18	caractéristiques issues de l’histogramme d’angles
position	m_1	k	adéquation selon k directions, sans distance
position	m_2	n	n méta-modèles (basiques)

5.1.3 Expérimentations

Nous présentons ici les résultats obtenus pour la tâche de classification d’exemples de relations spatiales au moyen de différents jeux de caractéristiques de description, selon les modalités décrites dans la partie précédente.

Nous cherchons avant tout à valider la qualité de description offerte par les méta-modèles dont la présentation a fait l’objet de la partie centrale de cette thèse (chapitre 3). Tout d’abord, nous comparons ces modèles avec différents descripteurs supposés plus simples du positionnement spatial d’objets. Par la suite, nous nous concentrons sur les apports de chacune des variantes introduites dans le chapitre consacré : apport de l’intégration de l’information de distance, apport de la modélisation spatiale bipolaire et apport de la mesure sur l’étendue des objets. Toutes ces expérimentations sont conduites sur les deux jeux de données IME-OnDB et HCC-OnDB, selon le protocole d’expérimentation et de validation statistique exposé précédemment.

5.1.3.1 Apport des méta-modèles de positionnement

Dans un premier temps, nous réalisons une comparaison de la version basique des méta-modèles de positionnement présentés au chapitre 3 avec différents descripteurs de positionnement plus ou moins évolués. Les différents jeux de caractéristiques que nous utilisons sont résumés par le tableau 5.3.

Le jeu de caractéristiques f est constitué de neuf caractéristiques simples qui décrivent la forme de l’objet argument. Il sera intégré à toutes les expérimentations pour les deux bases de données IME-OnDB et HCC-OnDB. De même, le jeu de caractéristiques e sera toujours intégré. Il correspond à la mesure de la proportion relative des boîtes englobantes des deux objets (l’argument par rapport à la référence), ainsi qu’au ratio hauteur sur largeur de l’objet argument.

Les quatre jeux suivants sont des jeux de caractéristiques de positionnement relatif et c’est sur leur comparaison que repose l’intérêt de cette expérimentation.

Le jeu b est constitué de neuf mesures de positionnement relatif des boîtes englobantes des deux objets : décalages en x et en y des quatre bords des boîtes, distance et direction du vecteur reliant les centres des boîtes. . . Toutes ces mesures sont normalisées par la longueur de la diagonale de la boîte englobante de la référence.

Le jeu h est tout simplement le résultat direct du descripteur classique calculé par la

TABLE 5.4: Comparaison des taux de reconnaissance obtenus avec différents jeux de caractéristiques de positionnement pour les deux bases IME-OnDB et HCC-OnDB.

Exp.	<i>carac. communes</i>	<i>carac. positionnement</i>	% (IME-OnDB)	% (HCC-OnDB)
e_0	f, e	\emptyset	55.35	66.30
e_1	f, e	b	96.18	91.03
e_2	f, e	h	96.38	93.09
e_3	f, e	m_1	96.47	93.09
e_4	f, e	m_2	96.72	94.12

méthode d’histogramme d’angles. Tous les points échantillonnés des deux objets sont pris en compte deux à deux et les angles des vecteurs ainsi formés sont comptabilisés dans un histogramme. Dans notre cas, les angles sont quantifiés en 18 sections. Une fois l’histogramme normalisé, il fournit ainsi 18 caractéristiques de positionnement relatif.

Le jeu m_1 est calculé par application directe de k éléments structurants directionnels, selon la méthode proposée par Bloch et déjà éprouvée dans le contexte d’objets manuscrits par Bouteruche *et al.* [BMA06]. L’adéquation moyenne de l’objet argument par rapport au modèle spatial construit par l’un de ces éléments structurants fournit une caractéristique de positionnement. Dans cette expérimentation, on prend $k = 4$ (les éléments structurants retenus sont orientés par les quatre directions cardinales).

Enfin, le jeu m_2 est constitué par les mesures d’adéquation de l’objet argument par rapport à n modèles appris selon la version basique des méta-modèles définis au chapitre 3 (sans intégration de distance, ni prise en compte de la bipolarité, ni modèle d’étendue). A partir de la base d’apprentissage (la même que celle qui permet l’apprentissage du classifieur), nous construisons un méta-modèle de positionnement pour chaque classe de relation spatiale à reconnaître, selon la méthode d’apprentissage introduite dans nos travaux. En exploitation, ces modèles sont appliqués un à un sur l’objet de référence de l’exemple à reconnaître et sont évalués avec l’objet argument, fournissant ainsi une caractéristique par modèle (soit une par classe : 18 pour IME-OnDB et 17 pour HCC-OnDB). Les modèles sont construits en considérant les quatre points de vue directionnels par défaut (quatre directions cardinales), et tous les histogrammes sont définis avec huit sections.

Résultats Le tableau 5.4 présente le taux de reconnaissance obtenu avec les différents jeux de caractéristiques, sur les deux bases de données. Chaque taux de reconnaissance apparaissant dans ce tableau est la moyenne des quinze tests opérés lors de la validation croisée omni-scripteur.

Le tableau 5.5 associé représente le résultat de la validité statistique des comparaisons des ensembles de caractéristiques pris deux à deux.

Les résultats de ces tests statistiques confirment que la hiérarchie qui apparaît entre les méthodes du tableau 5.4 est significative. Autrement dit, pour ces deux jeux de données, les modèles spatiaux introduits dans nos travaux surpassent toutes les autres descriptions considérées pour le positionnement relatif des objets. En analysant plus en détail ces résultats, on note que les descripteurs par histogrammes d’angles sont supérieurs aux descripteurs qui reposent sur les boîtes englobantes. Cette supériorité peut s’expliquer par la prise en

5.1. Classification de relations spatiales entre objets manuscrits

TABLE 5.5: Significativité statistique à 5% de la comparaison deux à deux des expérimentations du tableau 5.4 (le test statistique tient compte des deux bases de données). Un symbole « ✓ » à l'intersection de la ligne e_i et de la colonne e_j du tableau exprime que la description e_i est significativement supérieure à la description e_j (à un niveau de test de 5%).

	e_0				
e_1	✓	e_1			
e_2	✓	✓	e_2		
e_3	✓	✓	✓	e_3	
e_4	✓	✓	✓	✓	e_4

compte précise de la forme réelle des tracés dans e_2 par rapport à l'approximation grossière des objets qui est faite dans l'expérience e_1 . Ensuite, l'expérimentation e_3 démontre la bonne qualité de description offerte par l'espace de représentation par opérateurs morphologiques directionnels. Comme nous l'avions évoqué, l'application directe de quatre éléments structurants de direction fournit un jeu de caractéristiques très performant. Ici on vérifie que ce jeu surpasse les deux descripteurs e_1 et e_2 . Enfin, l'expérience e_4 témoigne de la valeur ajoutée de notre modélisation et confirme donc l'intérêt de l'apprentissage de modèles de positionnement pour la qualité de description des relations spatiales entre les objets.

La base de données HCC-OnDB permet de mettre en évidence des différences plus importantes entre les qualités de description, notamment parce que les classes de positionnement qu'elle comporte présentent plus de situations d'intersection entre les tracés. Ces cas de figures sont particulièrement mal gérés par les descripteurs basés sur les boîtes englobantes, qui sont nettement surpassés par les techniques plus fidèles aux tracés réels.

Visualisation des modèles spatiaux Le tableau 5.6 illustre les modèles spatiaux appris pour chaque classe de la base IME-OnDB.

Pour chaque classe, le modèle est développé sur un objet manuscrit pris comme exemple de référence. Ces exemples permettent de valider visuellement la bonne qualité de la description fournie par les modèles par rapport à ces exemples. La première ligne du tableau montre que les modèles peuvent présenter des différences assez fines pour décrire le positionnement des différents types d'accents (il s'agit bien plus que d'une simple description du type *être en haut* par rapport à la référence). Les représentations des classes *apostrophe* ou *virgule* témoignent du bon comportement des modèles spatiaux face à ces classes qui reposent sur différentes lettres de référence. Les cas d'intersection de la référence avec l'argument ainsi que d'inclusion de l'argument inclus dans une concavité de la référence sont également bien modélisés.

Le tableau 5.7 illustre les modèles spatiaux appris pour chaque classe de la base HCC-OnDB. Ces visualisations confortent qualitativement le bon comportement des modèles par rapport à ces objets d'une autre nature, où les références sont constituées de plusieurs tracés primitifs.

TABLE 5.6: Visualisation de 16 modèles de positionnement appris pour les 16 classes de la base de données IME-OnDB présentées dans le tableau 5.1.

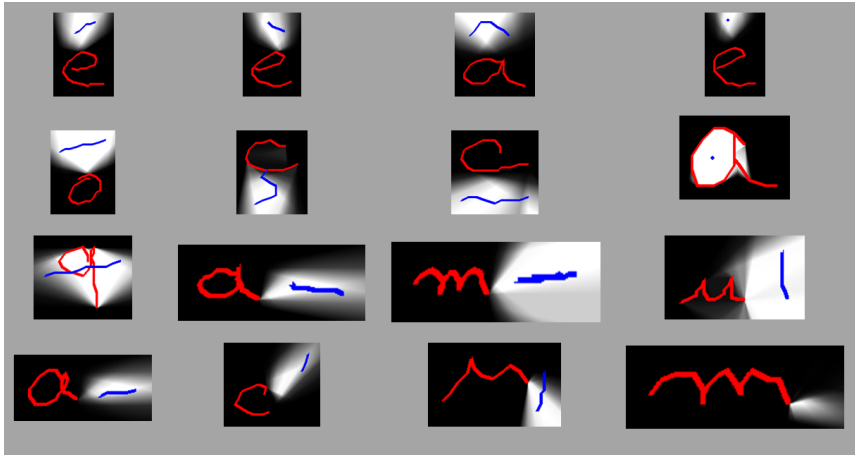
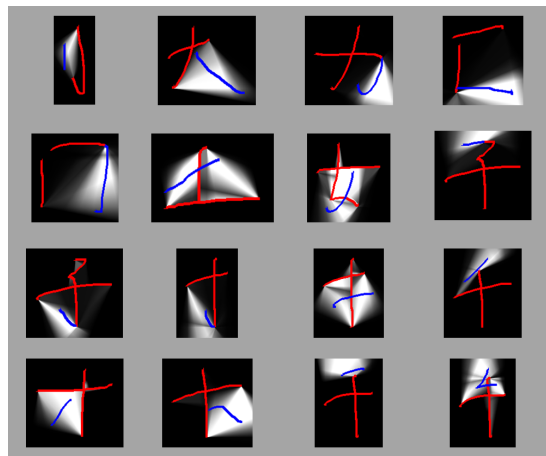


TABLE 5.7: Visualisation de 16 modèles de positionnement appris pour les 16 classes de la base de données HCC-OnDB présentées dans le tableau 5.2.



5.1.3.2 Apport de la distance

Nous avons étudié avec un intérêt particulier dans le chapitre 3 le problème de l'intégration de la distance dans les méta-modèles de positionnement. La seconde expérimentation vise à confirmer l'intérêt de cette intégration tout en comparant les deux méthodes que nous avons proposées : intégration globale de la distance (on parlera de *distance globale*), prise comme un point de vue supplémentaire sur la relation spatiale, ou intégration de modèles de distance au sein même de chaque point de vue directionnel (on parlera de *distance embarquée*).

Le tableau 5.8 résume les nouveaux jeux de caractéristiques introduits pour mesurer

5.1. Classification de relations spatiales entre objets manuscrits

TABLE 5.8: Jeux de caractéristiques avec intégration de la distance.

type	désignation	nombre	description
position	m_1d_g	$k + 1$	adéquation selon $k +$ distance globale
position	m_2d_g	n	n modèles, distance globale
position	m_2d_e	n	n modèles, distance embarquée

TABLE 5.9: Comparaison des taux de reconnaissance obtenus avec différents jeux de caractéristiques incorporant l'information sur la distance pour les deux bases IME-OnDB et HCC-OnDB.

Exp.	<i>carac. communes</i>	<i>carac. positionnement</i>	% (IME-OnDB)	% (HCC-OnDB)
e_3	f, e	m_1	96,47	93,09
e_{31}	f, e	m_1d_g	96,67	94,12
e_4	f, e	m_2	96,72	94,12
e_{41}	f, e	m_2d_g	96,72	94,44
e_{42}	f, e	m_2d_e	96,79	94,85

TABLE 5.10: Significativité statistique à 5% de la comparaison deux à deux des expérimentations du tableau 5.9.

	e_3				
e_{31}	✓	e_{31}			
e_4	✓	-	e_4		
e_{41}	✓	-	-	e_{41}	
e_{42}	✓	✓	✓	-	e_{42}

l'apport de l'information de distance sur la tâche de reconnaissance. Le jeu m_1d_g est similaire au jeu m_1 défini précédemment, mais intègre en plus un point de vue de distance (une caractéristique additionnelle). m_2d_g et m_2d_e sont les deux jeux de caractéristiques issues de méta-modèles spatiaux mettant en œuvre les deux stratégies proposées pour l'intégration de l'information de distance. Comme précédemment, un méta-modèle est appris pour chaque classe de relation spatiale. Pour un exemple à reconnaître, on extrait son degré d'appartenance à chacun des modèles.

Résultats Le tableau 5.9 résume les résultats obtenus en prenant en compte l'information de distance au moyen des trois jeux de caractéristiques nouvellement introduits (lignes grisées) et le tableau associé 5.10 donne les résultats des tests de significativité relatifs à ces comparaisons.

On peut constater que, d'une façon générale, l'apport de la distance semble être plus bénéfique pour la base de données HCC-OnDB que pour la base IME-OnDB. Une analyse

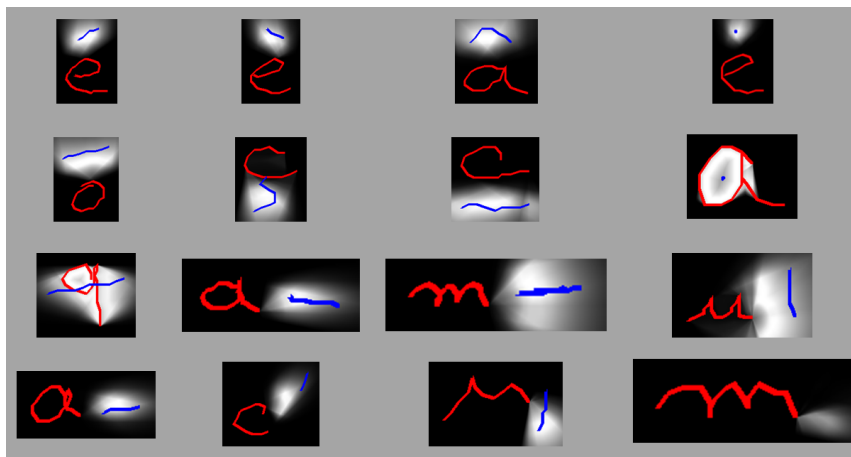
fine des erreurs sur la base IME-OnDB permet de voir que la plupart des erreurs de confusion concernent des classes de gestes au positionnement similaire, mais à la forme distincte. Il y a par exemple beaucoup de confusions entre les classes *accent aigu* et *accent circonflexe*, qui ne peuvent pas être différenciées par l'information supplémentaire de distance.

Les tests de significativité montrent que l'apport de la distance comme caractéristique supplémentaire dans le jeu de caractéristiques morphologiques directes est significativement bénéfique (expérimentation e_{31} par rapport à e_3).

La comparaison des deux méthodes d'intégration dans les méta-modèles spatiaux penche en faveur de la méthode par intégration fine de la distance (c'est à dire au sein de chaque point de vue directionnel, plutôt que globalement). En effet, même si la comparaison directe des deux méthodes ne permet pas de conclure à une différence significative (e_{42} vs e_{41}), l'apport de la distance embarquée dans le jeu e_{42} améliore significativement la description par rapport aux modèles sans distance (jeu e_4). En fait, si on limite le test de significativité aux seules mesures effectuées sur la base HCC-OnDB (soit la base de données pour laquelle l'information de distance a le plus d'importance), on peut démontrer que l'intégration de distance embarquée est effectivement supérieure à l'intégration globale (différence significative à 5% sur les 15 jeux de données omni-scripteur).

Visualisation des modèles spatiaux Le tableau 5.11 illustre les modèles spatiaux appris pour chaque classe de la base IME-OnDB, avec une intégration de la distance direction par direction (distance embarquée).

TABLE 5.11: Visualisation de 16 modèles de positionnement avec distance appris pour 16 classes de la base de données IME-OnDB (telles que définies dans le tableau 5.1).

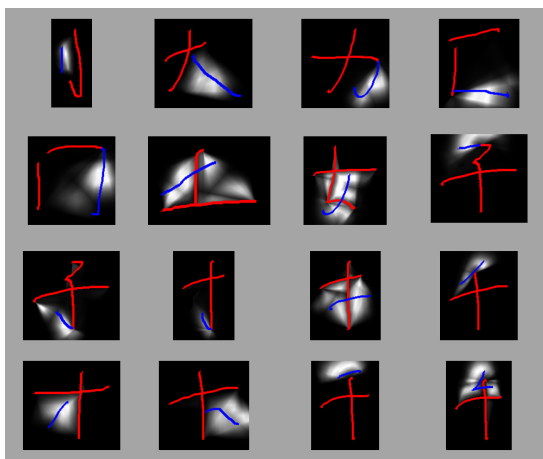


Par comparaison avec le tableau 5.6, on peut remarquer un gain en précision permis par l'intégration de la distance dans les modèles. Des nuances de niveaux de gris supplémentaires apparaissent. Certaines zones, qui semblaient s'étendre à l'infini sont désormais mieux délimitées (voir par exemple la classe *accent circonflexe*). En général, les formes des zones d'acceptabilité s'affinent, parfois jusqu'à refléter fidèlement la forme des objets arguments (par exemple pour les classes *accent grave* et *accent aigu*).

5.1. Classification de relations spatiales entre objets manuscrits

La tableau 5.12 présente les modèles spatiaux pour les classes de la base HCC-OnDB.

TABLE 5.12: Visualisation de 16 modèles de positionnement avec distance appris pour 16 classes de la base de données HCC-OnDB (telles que définies dans le tableau 5.2).



5.1.3.3 Apport de la bipolarité

Le tableau 5.13 présente les jeux de caractéristiques avec prise en compte de la bipolarité de l'information spatiale au sein de méta-modèles appris. Les jeux m_2n et m_2nd_e correspondent à l'exploitation des modèles bipolaires respectivement sans distance et avec modèles de distances embarqués. Les jeux de caractéristiques m_2n et m_2nd_e comportent deux fois plus de caractéristiques que leurs équivalents non bipolaires, puisque pour chaque méta-modèle de positionnement on conserve les deux composantes du score bipolaire d'évaluation.

Résultats Les résultats de classification obtenus avec ces jeux de caractéristiques sont présentés dans le tableau 5.14, qui indique également à titre de rappel les résultats obtenus par les modèles équivalents sans prise en compte de la bipolarité.

Ces résultats mettent en évidence une légère amélioration permise par la prise en compte de la bipolarité dans les méta-modèles spatiaux. Cet effet est reproduit sur les quatre comparaisons et un test de significativité prenant en compte les quatre comparaisons (donc avec 60 tests comparatifs élémentaires) permet de conclure à la significativité de l'apport de la bipolarité à un niveau de confiance de 5%.

TABLE 5.13: Jeux de caractéristiques avec prise en compte de la bipolarité.

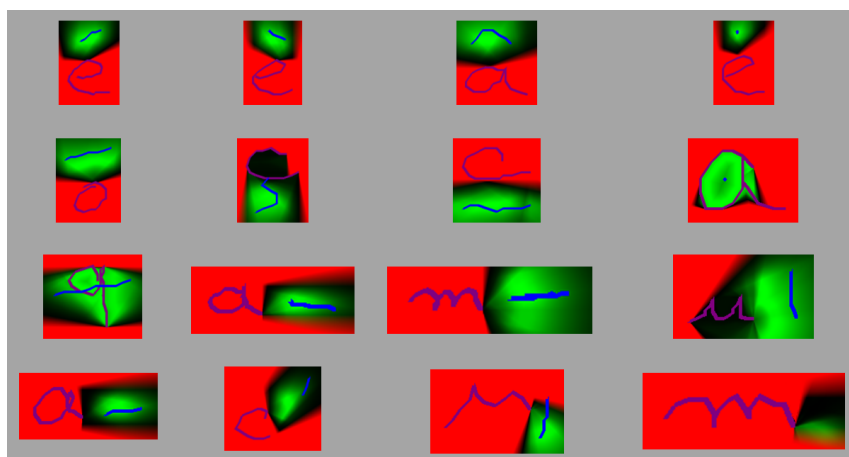
type	désignation	nombre	description
position	m_2n	$2n$	n modèles bipolaires, sans distance
position	m_2nd_e	$2n$	n modèles bipolaires, distance embarquée

TABLE 5.14: Comparaison des taux de reconnaissance obtenus avec prise en compte de la bipolarité de l'information spatiale pour les deux bases IME-OnDB et HCC-OnDB.

Exp.	carac. communes	carac. positionnement	% (IME-OnDB)	% (HCC-OnDB)
e_4	f, e	m_2	96,72	94,12
e_{42}	f, e	$m_2 d_e$	96,79	94,85
e_5	f, e	$m_2 n$	96,86	94,54
e_{52}	f, e	$m_2 n d_e$	97,10	95,67

Visualisation des modèles spatiaux Le tableau 5.15 illustre les modèles spatiaux bipolaires appris pour chaque classe de la base IME-OnDB.

TABLE 5.15: Visualisation de 16 modèles bipolaires de positionnement appris pour 16 classes de la base de données IME-OnDB (telles que définies dans le tableau 5.1).



On peut percevoir les nuances apportées par la distinction entre les zones de préférence négatives et positives grâce au code de couleurs que nous avons présenté au paragraphe 3.4.3.2. Par exemple, le modèle spatial bipolaire de la classe *cédille* permet de distinguer une zone positive (située en dessous de la lettre « c »), une zone négative (au dessus et sur les cotés du « c ») et une zone neutre (à l'intérieur du « c »). Cette description traduit le fait que dans certains exemples le tracé de la cédille déborde à l'intérieur de la concavité du « c ».

Le tableau 5.16 présente les modèles spatiaux bipolaires pour la base de données HCC-OnDB.

5.1.3.4 Apport de la modélisation de l'étendue

Dans cette partie, nous cherchons à mesurer l'apport de la modélisation de l'étendue de l'objet argument dans la description du positionnement, en écho à la section 3.5. Le tableau 5.17 présente les deux jeux de caractéristiques introduits, définis à partir des deux méthodes

5.1. Classification de relations spatiales entre objets manuscrits

TABLE 5.16: Visualisation de 16 modèles bipolaires de positionnement appris pour 16 classes de la base de données HCC-OnDB (telles que définies dans le tableau 5.2).

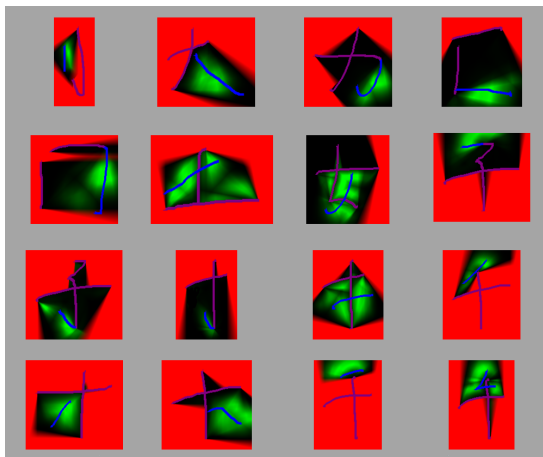


TABLE 5.17: Jeux de caractéristiques d'étendue

type	désignation	nombre	description
étendue	ext_g	12	étendue par nuage de points
étendue	ext_{abm}	12	étendue par points singuliers

de description de l'étendue proposées dans notre étude : étendue par *nuage de points* (voir section 3.5.2) ou étendue par *points singuliers* (voir section 3.5.3).

Les caractéristiques extraites ne sont pas des scores d'adéquation issus des modèles d'étendue (comme ce qui est présenté dans la section 3.5), mais une description directe de l'étendue selon la représentation choisie. Par exemple, dans le cas de la représentation de l'étendue par projection de nuage de points, les douze caractéristiques sont les trois mesures (minimale, médiane et maximale) de la distribution des points projetés selon chacun des quatre points de vue (voir à la section 3.5.2). Pour la seconde représentation, les douze caractéristiques correspondent aux abscisses des valeurs modales des histogrammes appris pour chacun des trois points particuliers selon les quatre points de vue (voir à la section 3.5.3).

Résultats Les résultats présentés dans le tableau 5.18 résument les taux de reconnaissance atteints en intégrant les caractéristiques d'étendue avec deux différents jeux de caractéristiques de positionnement : le simple jeu de caractéristiques à partir de mesures sur les boîtes englobantes (b) et le meilleur jeu de caractéristiques obtenu sur les expérimentations précédentes, fourni par les méta-modèles spatiaux bipolaires avec *distance embarquée* (m_2nd_e). Pour chaque jeu de positionnement, on procède au test de reconnaissance sans lui intégrer d'information d'étendue, puis en le combinant avec chaque représentation d'étendue ext_g et ext_{abm} . Le jeu de caractéristiques de formes f n'est pas utilisé dans ces expérimentations afin de mieux mettre en évidence l'apport des modèles d'étendue pour ces données.

TABLE 5.18: Comparaison des différentes modélisations de l'étendue pour les deux bases IME-OnDB et HCC-OnDB.

Exp.	carac. étendue	carac. positionnement	% (IME-OnDB)	% (HCC-OnDB)
e_{10}	\emptyset	b	81,83	90,41
e_{11}	ext_g	b	89,59	94,43
e_{12}	ext_{abm}	b	93,52	94,64
e_{520}	\emptyset	m_2nd_e	86,75	92,99
e_{521}	ext_g	m_2nd_e	89,92	94,95
e_{522}	ext_{abm}	m_2nd_e	94,55	95,46

Les résultats obtenus en intégrant des caractéristiques sur l'étendue des objets par les deux méthodes ext_g et ext_{abm} permettent de vérifier l'intérêt de cette modélisation, sur les deux bases et avec les deux méthodes de positionnement.

En effet, la supériorité de e_{11} et e_{12} par rapport à e_{10} apparaît très nettement, de même que celle de e_{511} et e_{512} par rapport à e_{510} . La comparaison entre les deux méthodes d'intégration de l'étendue est favorable à la méthode par points particuliers plutôt que la méthode par nuage de points sur la base IME-OnDB, et ce quel que soit le jeu de caractéristiques de positionnement associé (boîtes englobantes ou méta-modèles spatiaux). Les deux méthodes ont des résultats plus comparables sur la base de données HCC-OnDB. Prises globalement en considérant les deux bases, toutes les différences de performances sont cependant significatives.

Ces tests démontrent que la modélisation de l'étendue apporte bel et bien des indices supplémentaires utiles à la reconnaissance des relations spatiales et que, comme nous l'avions pressenti, la description par considération des points singuliers des objets est plus performante que la description par projection de l'objet vu comme un nuage de points.

5.1.4 Généralisation à d'autres jeux de données

Afin d'éprouver la capacité des méta-modèles spatiaux à s'adapter à d'autres natures de données et de relations spatiales entre des objets manuscrits, nous élargissons dans cette partie nos expérimentations à d'autres jeux de données. Nous présentons deux extensions du cadre expérimental d'une part à des données extraites de mots manuscrits hors-ligne et d'autre part à des symboles mathématiques extraits d'expressions en-ligne. Les propriétés de chaque jeu de données sont présentées dans deux paragraphes distincts. Les résultats obtenus pour la classification de leurs relations spatiales sont exposés et commentés dans un troisième temps.

5.1.4.1 Lettres en contexte de mots manuscrits hors-ligne

La reconnaissance de texte manuscrit est un des problèmes les plus étudiés dans le domaine de l'analyse de documents. Les besoins sont considérables et couvrent aussi bien le traitement de document saisi en-ligne (par exemple grâce aux méthodes embarquées de saisie de texte manuscrit à l'aide d'un stylo électronique sur une surface sensitive) que le

5.1. Classification de relations spatiales entre objets manuscrits

traitement de documents hors-ligne (composés sur un support papier avant d’être numérisés sous la forme d’une image).

Nous avons extrait un jeu de données à partir des mots manuscrits saisis dans le cadre de la campagne RIMES (pour *Reconnaissance et Indexation de données Manuscrites et de fac similÉS*) [GCBG09]. Il s’agit de courriers manuscrits réalisés sur papier avant d’être numérisés, puis annotés et retranscrits manuellement. Une base de mots a été extraite de ce lot de documents afin de constituer une référence pour l’évaluation et la comparaison de systèmes de reconnaissance de mots manuscrits hors-ligne dans le cadre de compétitions [GEA09]. La figure 5.1 présente quelques exemples de mots contenus dans cette base, sous la forme d’images en niveaux de gris.




FIGURE 5.1: Trois exemples de mots contenus dans la base RIMES pour la reconnaissance automatique de mots manuscrits hors-ligne.

Des paires de lettres consécutives ont été extraites des 7798 mots du jeu de validation de la base RIMES. Puisque la vérité terrain n’est disponible qu’au niveau du mot complet et non pas au niveau des lettres, nous nous sommes appuyés sur le système de reconnaissance de Guichard *et al.* [GTC10]. Connaissant l’étiquette d’un mot, ce système est capable, grâce à l’utilisation de Modèles de Markov Cachés, de produire la segmentation la plus probable d’un mot en lettres et ainsi d’offrir un ensemble d’imageries de lettres segmentées et étiquetées. Un classifieur statistique de type *Support Vector Machine* (SVM) dont les scores de sortie sont probabilisés est ensuite utilisé. Il permet de vérifier la bonne segmentation des lettres, grâce à un seuil de rejet sur le score de reconnaissance des imageries. Les exemples conservés constituent ainsi une base de paires de lettres, dont la segmentation est relativement fiable et qui sont correctement étiquetées. Nous proposons de comparer différentes descriptions des relations spatiales sur cette base de paires de lettres. Au total, environ 13000 paires de lettres sont conservées, constituant le jeu de données dénommé MOT-OffDB dans la suite.

Dans la figure 5.2, un mot est représenté (a) avec le résultat de sa segmentation fournie par le moteur de reconnaissance (b). Les vignettes (c-l) représentent les paires de lettres extraites qui constituent des exemples de différentes classes de relations spatiales (avec les objets référence représentés en rouge et les arguments en bleu).

Classes de relations spatiales On définit une classe de positionnement spatial par la relation qui lie deux lettres successives d’un même mot en fonction de leur catégorisation en quatre grands types de lettres. Les types de lettres dépendent de leur étendue verticale par rapport à la ligne de base du mot. On distingue ainsi les lettres *ascendantes* (b,d,h,k,l,t,majuscules), *descendantes* (g,j,p,q,y,z), *médianes* (a,c,e,i,m,n,o,r,s,u,v,w,x), et *ascendantes-descendantes* (f). Les lettres supportant un accent ou un tréma sont assimilées à la catégorie des *ascendantes*, tout comme les majuscules, tandis que les « c » portant une cédille sont assimilés à des *descendantes*. On retient en fait onze classes de relations spatiales, correspondant aux onze catégories de couples qui apparaissent effectivement dans la base d’apprentissage.

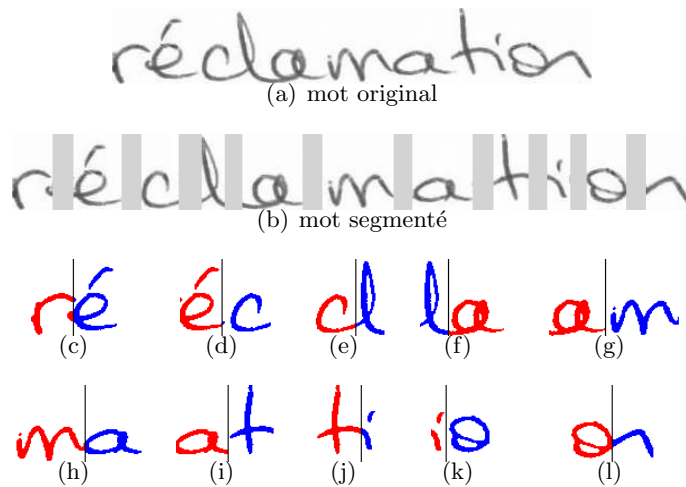


FIGURE 5.2: Mot extrait de la base RIMES (a) avec sa segmentation en lettres fournie par le reconnaisseur par Modèles de Markov Cachés (b), et les paires de lettres extraites comme représentants de différentes classes de relations spatiales (c-l).

Adaptation du choix des points de vue Les couples d'objets ont une particularité dans leur positionnement : par définition, à cause de la segmentation en colonnes fournie par les MMC, on a toujours une frontière verticale qui sépare l'objet de référence (à gauche) et l'objet argument (à droite). Par conséquent, l'analyse de la relation spatiale entre les objets selon tout point de vue orienté vers le demi plan gauche de l'espace est inutile. Cela suggère pour nos modèles de redistribuer les points de vue parmi différentes directions orientées dans le demi plan droit, en éliminant le point de vue *gauche* de notre description. Pour nos expérimentations sur cette base, nous choisirons dans nos modèles de distribuer cinq points de vue orientés par les cinq directions suivantes : $\{-\frac{\pi}{2}, -\frac{\pi}{4}, 0, \frac{\pi}{4}, \frac{\pi}{2}\}$. On peut voir cette adaptation des directions des points de vue comme une application du principe de stratégie dynamique de sélection des points de vue évoquée à la section 3.2.2.2.

Nature des objets Dans le cas de ce jeu de données, les objets sont représentés par des images binarisées. La référence et l'argument sont donc vus comme des matrices de pixels, ce qui implique un coût d'exécution plus important de l'algorithme de dilatation morphologique (voir à l'annexe B). Dans ces données, les objets ne sont pas parfaitement linéiques. Nous appliquons donc un filtre d'extraction des contours sur l'objet de référence pour réduire le coût de calcul sans altérer la représentation, puisque seuls les points de contour de la référence jouent un rôle dans le calcul de la dilatation lorsque les objets sont distants, comme c'est le cas ici.

Variabilité des objets Pour une même classe de positionnement, les objets référence et argument peuvent représenter différentes lettres. L'adaptation face à une grande variabilité de formes est donc cruciale pour ce jeu de données (toutes les lettres ascendantes et les majuscules sont par exemple regroupées dans un modèle unique).

5.1. Classification de relations spatiales entre objets manuscrits

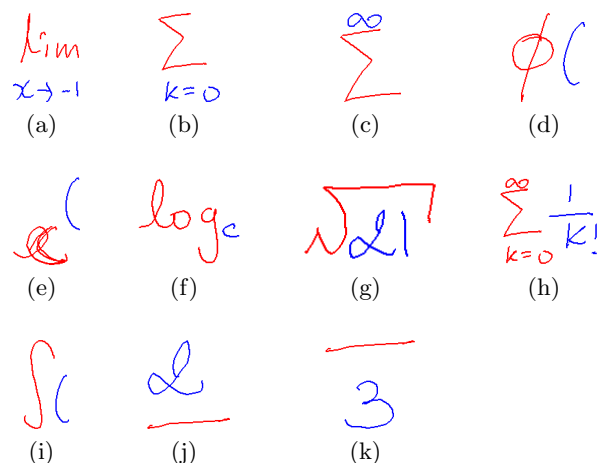
5.1.4.2 Symboles mathématiques en-ligne

Les expressions mathématiques manuscrites constituent un cadre applicatif où les problèmes de positionnement relatif entre les objets sont très sensibles. Comme nous l'avons constaté dans l'état de l'art, c'est dans ce champ d'application que la plupart des travaux dédiés au problème de la description du positionnement relatif d'objets manuscrits ont été conduits. Les relations spatiales qui interviennent dans les expressions mathématiques sont riches car elles exploitent réellement les deux dimensions du plan. Elles présentent souvent des ambiguïtés et mettent en relation des objets à la complexité graphique et aux dimensions très variables (par exemple un chiffre simple en exposant d'un symbole ou une expression complexe sous une racine carrée).

Nous avons extrait un jeu de données à partir de la base de données constituée très récemment pour la compétition de reconnaissance d'expressions mathématiques *CROHME* (*Competition on Recognition of Online Handwritten Mathematical Expressions*)² [MVGG+11]. Nous exploitons les 296 expressions (données saisies en-ligne) composant la base *Part-I* fournie par les organisateurs. La segmentation de chaque expression est connue et accompagnée de la vérité terrain, ce qui permet de construire l'arbre sémantique de l'expression complète. De cet arbre, nous extrayons des paires d'objets mis en relation selon l'une des classes de relations spatiales identifiées. Ces classes sont : *à droite*, *en indice*, *en exposant*, *en numérateur*, *en dénominateur*, *sous une somme*, *au dessus d'une somme*, *à droite d'une somme*, *à droite d'une intégrale*, *sous une racine carrée*, *sous une expression*. Les 10126 paires extraites constituent le jeu de données dénommé MATH-OnDB dans la suite.

Le tableau 5.19 illustre par un exemple chacune des classes de relations spatiales contenues dans cette base.

TABLE 5.19: Exemples d'objets positionnés selon les 11 classes de relations spatiales définies dans la base MATH-OnDB.



2. Les résultats de cette compétition ont été présentés dans le cadre de la conférence ICDAR, en septembre 2011 [MVGG+11]

TABLE 5.20: Résultats de la classification de relations spatiales sur les bases MATH-OnDB et MOT-OffDB avec différents jeux de caractéristiques de positionnement.

	carac. positionnement	nombre	% (MATH-OnDB)	% (MOT-OffDB)
e_0	b	12	97,90	80,76
e_{10}	m_1	7	97,95	80,35
e_{11}	m_1d_g	8	97,95	81,92
e_{20}	m_2	13	98,13	81,88
e_{21}	m_2d_g	13	97,95	81,93
e_{22}	m_2d_e	13	97,98	82,46
e_{30}	m_2n	23	98,16	81,84
e_{31}	m_2nd_g	23	98,11	82,51
e_{32}	m_2nd_e	23	98,10	83,46

Nature des objets Les objets sont des objets manuscrits en-ligne, composés d'un seul ou de plusieurs tracés. Les références sont très variables dans leur forme. Par exemple, la relation spatiale *exposant* peut s'appliquer sur un nombre, une lettre (qui peut elle-même être médiane, ascendante, descendante), une majuscule, mais aussi une expression composée (par exemple une racine carrée, une fraction. . .). Il en résulte un besoin d'élasticité important dans les modèles de positionnement, ainsi qu'une certaine robustesse à la distance et à l'échelle.

Nature des relations spatiales Les objets sont toujours disjoints (aucune intersection des tracés n'est présente), mais leurs boîtes englobantes peuvent se recouvrir partiellement (c'est souvent le cas pour les relations *exposant* ou *indice*) ou totalement (pour la relation *sous une racine carrée*). Toutes les directions du plan sont à considérer; nous conservons donc les quatre points de vue cardinaux classiques pour la définition des méta-modèles.

5.1.4.3 Résultats

Le tableau 5.20 présente les résultats obtenus pour la tâche de reconnaissance automatique des relations spatiales pour les deux bases de données MATH-OnDB et MOT-OffDB avec différents jeux de caractéristiques de positionnement relatif. Le protocole expérimental est plus simple que le protocole omni-scripteur présenté précédemment. Les taux de reconnaissance présentés ici sont mesurés sur un jeu de test unique constitué au hasard à partir des bases de données, sans tenir compte de l'identité des scripteurs. Les significativités des différences entre les méthodes ne sont pas calculées. Ces expérimentations ont pour but de démontrer la reproductibilité des résultats en vérifiant si la description du positionnement relatif par méta-modèles spatiaux est plus performante qu'une description simple par mesures de boîtes englobantes sur ces autres types de données. Les mêmes notations que dans la partie précédente sont utilisées pour les jeux de caractéristiques.

Base MATH-OnDB Bien que les taux de reconnaissance des relations spatiales dans la base MATH-OnDB soient déjà élevés avec la description par caractéristiques extraites des

5.1. Classification de relations spatiales entre objets manuscrits

boîtes englobantes (e_0), on note un léger gain de performance permis par les meilleurs méta-modèles spatiaux (e_{20} ou e_{30}). Le bon comportement de la description par boîtes englobantes s'explique par le fait qu'il n'y a dans la base de données qu'une seule relation spatiale où les boîtes des objets se recouvrent largement (il s'agit de la relation racine carrée). Ce cas, généralement mal géré par cette technique, est donc ici facile à identifier.

L'intégration des modèles de distance dans le cas de la base MATH-OnDB n'est pas bénéfique pour la description. Cela s'explique par la grande variabilité des formes et la complexité des objets positionnés selon les mêmes relations spatiales. La relation *exposant*, par exemple, peut faire intervenir des objets aussi différents qu'un simple symbole isolé (par exemple un chiffre) ou le mot-clé « *log* », de dimensions beaucoup plus grandes (voir l'exemple (f) du tableau 5.19). Non seulement la distance autorisée pour une même relation est très variable (en fonction de la complexité graphique de l'objet argument), mais cette distance est aussi difficile à normaliser (en fonction des dimensions de l'objet référence). Pour permettre une modélisation satisfaisante de la distance dans un cas réel, il faudrait pouvoir estimer *en contexte* un paramètre d'échelle, par exemple lié à la dimension moyenne des symboles saisis dans une expression complète. Ce paramètre permettrait alors de mieux normaliser les mesures de distance.

La modélisation bipolaire (e_{30}) ne semble pas apporter de gain significatif pour cette base de données.

Base MOT-OffDB Sur la base de données de relations spatiales entre paires de lettres issues de mots manuscrits, on obtient une hiérarchie des jeux de caractéristiques similaire à celle établie dans la section précédente sur les gestes d'édition ou les primitives du chinois. La description par boîtes englobantes (e_0) est dominée par les méta-modèles spatiaux simples (e_{20}) et bipolaires (e_{30}). On remarque cette fois-ci une réelle importance de l'information de distance dont l'intégration bénéficie nettement aux deux types de méta-modèles (e_{21} , e_{22} et e_{31} , e_{32}). La technique d'intégration par distance *embarquée* surpasse sensiblement la technique globale dans les deux cas. Enfin, la bipolarité apporte ici un réel gain notamment en combinaison avec l'intégration de la distance (e_{31} et e_{32}). Cela s'explique par le rôle de la modélisation de la distance dans les méta-modèles bipolaires : elle permet de raffiner la description de l'information positive et renforce la distinction entre zones de préférence positives et zones de préférence négatives (ces dernières n'étant définies que sur des critères de direction).

Visualisation Le tableau 5.21 représente quelques exemples des modèles spatiaux appris entre des paires de lettres consécutives extraites des mots (base MOT-OffDB).

Les modèles représentés sont les modèles bipolaires avec intégration de la distance *embarquée*. Ils correspondent aux meilleurs modèles d'après les résultats rapportés dans le tableau 5.20 pour cette base (jeu de caractéristiques m_2nd_e).

Le tableau 5.22 représente par un exemple chaque classe de positionnement relatif définie dans la base d'expressions mathématiques MATH-OnDB (voir le tableau 5.19). Les modèles représentés sont les modèles bipolaires sans information de distance.

TABLE 5.21: Visualisation des modèles spatiaux bipolaires (avec distance *embarquée*) appris sur la base MOT-OffDB : relation *médiane/médiane* (a,b,c), *médiane/ascendante* (d,e,f,g,h), *ascendante/médiane* (i,j), *ascendante/ascendante* (k,l), *descendante/médiane* (m), *descendante/ascendante* (n), *descendante/descendante* (o) et *ascendante-descendante/médiane* (p).

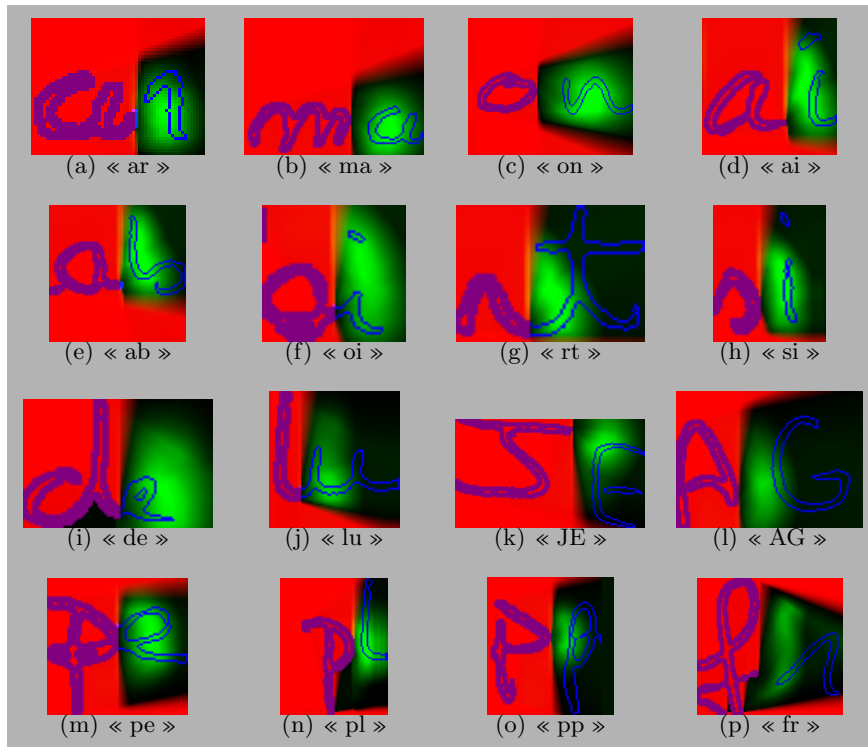
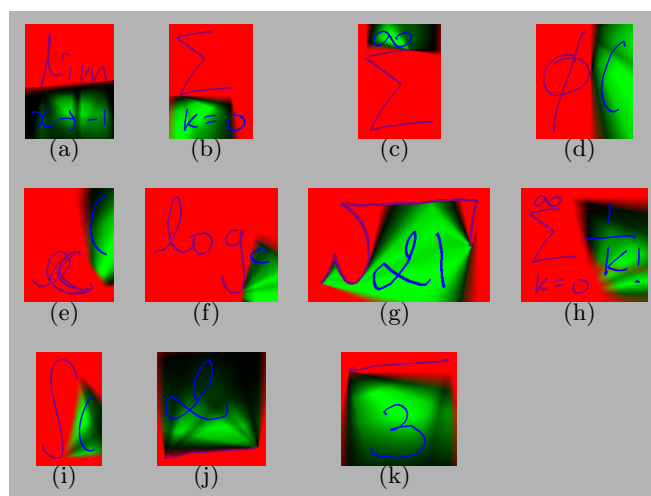


TABLE 5.22: Visualisation des modèles spatiaux bipolaires (sans distance) appris sur la base MATH-OnDB.



5.2. Evaluation des PSF pour la représentation de symboles structurés

5.1.5 Synthèse

Dans cette première partie expérimentale, nous avons démontré la qualité et la précision de description des positionnements relatifs d'objets manuscrits au moyen des méta-modèles spatiaux introduits dans cette thèse. Les expérimentations conduites sur des données de différentes natures (commandes gestuelles, écriture de mots, primitives de caractères chinois et symboles mathématiques) mettent en évidence la supériorité des méta-modèles spatiaux par rapport à l'utilisation de descripteurs de positionnement plus classiques. Les objectifs d'une large expressivité des types de relations et d'une grande généralité face à des types d'objets différents sont donc bien validés par cette phase expérimentale.

Ces résultats valident d'une part le bien fondé de l'approche : la prise en compte précise des formes des objets contribue à la bonne modélisation de leur positionnement relatif. Ils permettent d'autre part de vérifier que les méta-modèles spatiaux offrent une mise en œuvre efficace de ce principe. Les différentes variantes des modèles introduites dans le chapitre 3 permettent toutes de contribuer à la performance des modèles vis-à-vis de certains aspects de la description. L'intégration de l'information de distance offre une précision accrue sur les bases de gestes d'édition, de primitives du chinois et de paires de lettres. La représentation bipolaire élargit également le pouvoir de description en distinguant les zones de préférences positives, négatives et neutres du modèle spatial. Cela se traduit par un gain de précision notable sur les bases de gestes, de paires de lettres et de primitives du chinois. Enfin, le modèle d'étendue associé à un méta-modèle de positionnement fournit un outil complémentaire pour la vérification a posteriori de contraintes de positionnement qui sont bénéfiques à la discrimination entre les relations spatiales (sur les bases de gestes d'édition et de primitives du chinois).

5.2 Evaluation des PSF pour la représentation de symboles structurés

Cette seconde partie expérimentale vise à mesurer quantitativement la finesse et la pertinence de la description de formes structurées au moyen des Patrons Structurels Flous que nous avons introduits au chapitre 4. On va pour cela évaluer leur capacité à reconnaître des symboles manuscrits, en mettant en œuvre les algorithmes d'apprentissage et d'exploitation que nous avons présentés aux sections 4.2.4 et 4.2.3. La tâche évaluée consiste à reconnaître la classe d'un symbole donné parmi un ensemble de classes modélisées chacune par un prototype de PSF. Dans l'absolu, répondre à ce problème de reconnaissance nécessite d'apprendre à *discriminer* correctement les exemples des différentes classes. Dans notre contexte particulier, en revanche, l'objectif visé n'est pas d'obtenir la meilleure performance de reconnaissance possible, mais plutôt de se servir de la tâche de reconnaissance pour vérifier la qualité de la représentation offerte par les PSF en s'intéressant à la question : *dans quelle mesure les PSF permettent-ils de reconnaître des symboles ?*

Les PSF offrent une description *intrinsèque* à chacune des classes et ils ne sont pas entraînés pour les discriminer. Plus simplement, ils visent à décrire au mieux les symboles en se concentrant sur leur structuration interne et le positionnement relatif de leurs éléments. La reconnaissance de symboles selon la démarche proposée ici attribue à un symbole la classe du PSF qui le décrit au mieux d'après le score bipolaire flou calculé selon la formalisation

TABLE 5.23: Les 50 classes retenues dans le jeu de symboles CaracDB.

présentée dans la section [4.2.2.2](#).

Dans une première partie, nous décrivons les différents jeux de données utilisés pour cette validation expérimentale. Nous présentons ensuite le protocole mis au point pour procéder à l'apprentissage des PSF, à la définition des hyper-paramètres et à la tâche de reconnaissance elle-même. Enfin, la troisième partie présente et discute les résultats obtenus en reprenant différentes variantes des méta-modèles spatiaux déjà comparées dans la première partie expérimentale. Les limitations de la représentation sont également mises en évidence et analysées.

5.2.1 Jeux de données

La validation de la démarche de reconnaissance de symboles manuscrits par Patrons Structurels Flous est conduite sur plusieurs bases de données comprenant des symboles de différentes natures et aux caractéristiques propres. Les trois bases de données de symboles que nous avons retenues sont décrites ici. Elles correspondent aux trois types de symboles que nous avons évoqué dans la section [4.1](#) (voir en particulier les exemples de la figure [4.1](#)) : il s'agit de caractères chinois, de pictogrammes et de symboles architecturaux.

5.2.1.1 Caractères chinois

Le premier jeu de symboles considéré est un extrait d'une base de caractères chinois isolés et en-ligne, qui a été collectée au sein du laboratoire *Laboratoire Franco Chinois d'Informatique, d'Automatique et de Mathématiques Appliquées* (LIAMA). C'est la même base qui avait déjà fourni le jeu de données HCC-OnDB dans la première partie expérimentale (voir la section [5.1.1.2](#)).

La base originale contient 60 exemples pour chacun des 6763 caractères du dictionnaire choisi (un caractère par classe et par scripteur) [ML09]. Nous avons sélectionné 50 classes de caractères pour constituer le jeu de « symboles » dénoté par la suite CaracDB. Le tableau [5.23](#) illustre un exemple de chaque classe retenue dans ce jeu de données.

5.2. Evaluation des PSF pour la représentation de symboles structurés

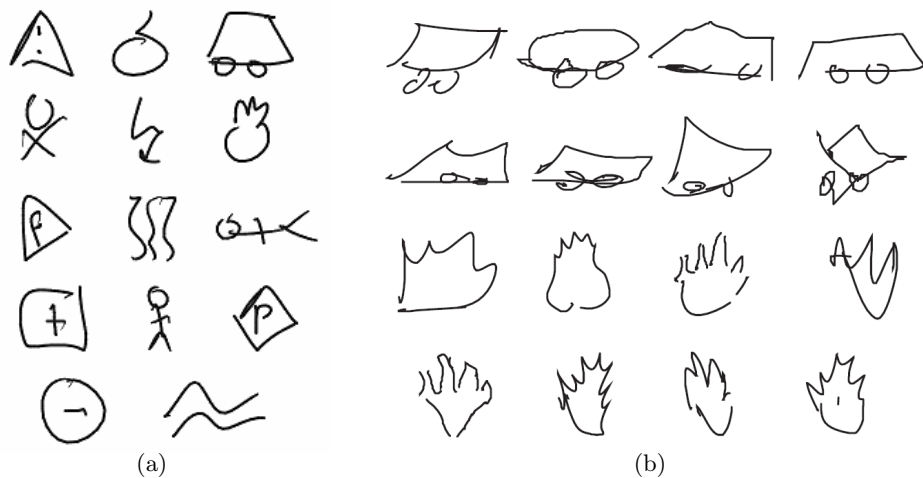


FIGURE 5.3: Les 14 classes du jeu de symboles PictoDB constitué à partir de la base NicIcon sont présentées en (a) (tiré de [NWV08]). La figure (b) illustre la variabilité constatée des tracés pour les deux classes *car* et *fire* (tiré de [WNvGV09]).

Les caractères sélectionnés sont des caractères simples et correspondent le plus souvent à des *radicaux* du chinois. Pour rappel, les radicaux sont les éléments intermédiaires de structure des caractères complexes : ils sont composés de plusieurs tracés élémentaires et leur combinaison permet de former les caractères les plus complexes (voir la figure 2.12 de l'état de l'art). Le nombre de tracés naturels constituant chaque symbole est assez réduit dans cette base de données, variant d'un seul tracé pour les caractères les plus simples à moins d'une dizaine pour les plus complexes. Une certaine variabilité de segmentation est présente mais les variantes sont reproductibles d'un scripteur à l'autre. Il s'agit dans la grande majorité des cas de la connexion de plusieurs tracés théoriquement disjoints. Ce sont en général les mêmes éléments de tracés qui se retrouvent connectés d'un exemple à l'autre, à cause de l'ordre de séquence qui est induit par les règles de l'écriture du chinois.

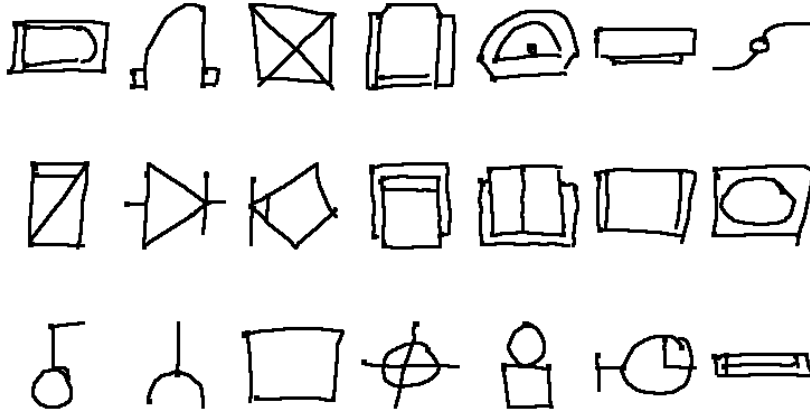
Cette base de données servira de base d'évaluation principale pour nos expérimentations. Ses caractéristiques telles que le niveau de complexité des symboles, le nombre de classes et la variabilité de segmentation permettront d'évaluer plusieurs propriétés des PSF.

5.2.1.2 Pictogrammes

La seconde source de données est la base NicIcon, publiée par Niels, Willems *et al.* [NWV08], qui contient des gestes multi-tracés saisis à la fois en-ligne et hors-ligne. Les *gestes* définis sont au nombre de 14 et sont en réalité des symboles choisis pour « présenter une ressemblance visuelle avec les objets qu'ils représentent ou correspondre à des symboles associés connus de tous » [NWV08]. La figure 5.3(a) décrit les 14 symboles retenus.

Il s'agit donc par définition de pictogrammes, qui ont pour but d'être aisément reconnaissables, mais pour lesquels il n'existe pas de méthode de dessin standard. En conséquence, la base de données NicIcon, qui contient des données saisies par 32 scripteurs (à raison de 700 exemples par scripteur), est caractérisée par une grande instabilité dans la segmentation

TABLE 5.24: Les 21 classes du jeu de symboles ArchiDB.



des symboles. La figure 5.3(b) illustre la variabilité constatée dans les formes des tracés et leur segmentation pour les classes *car* et *fire*. La complexité graphique des symboles est assez limitée : les auteurs de la base rapportent un nombre moyen de tracés élémentaires par symboles variant selon les classes entre 2,05 et 4,25 [WNvGV09].

Pour nos besoins expérimentaux, nous utilisons les données de 10 scripteurs en conservant 55 exemples par classe et par scripteur (symboles en-ligne). Le jeu de données ainsi constitué est dénoté dans la suite PictoDB.

Cette seconde base de données permettra d'éprouver la capacité de la démarche à construire des modèles structurels spatiaux dans le contexte de symboles dont la segmentation est très variable et qui présentent de grandes distorsions de tracés. Ce type de données constitue en fait un véritable défi pour l'évaluation de méthodes structurelles de reconnaissance de formes.

5.2.1.3 Symboles architecturaux

Le troisième jeu de données provient d'une base de symboles graphiques collectés au *Centre de Visió per Computador* (CVC) de Barcelone. Il s'agit de symboles en-ligne, tracés dans le contexte de plans architecturaux. Ils correspondent au jeu de données expérimental utilisé dans les travaux de Mas, Sanchez *et al.* [MSLL07] pour la reconnaissance de symboles par grammaire d'adjacence. La base que nous étudions contient des données correspondant à 21 classes de symboles saisis par 9 scripteurs à raison d'une dizaine de symboles par scripteur et par classe. Ce jeu de symboles est noté ArchiDB pour la suite. Le tableau 5.24 illustre par un exemple chaque classe de symboles de cette base.

Ces symboles reposent sur des tracés élémentaires rectilignes ou en arcs de cercle (ils sont d'ailleurs classifiés selon ces deux types dans les travaux de Mas, Sanchez *et al.*). De la même façon que dans la base PictoDB, il n'existe pas de dessin standard pour ces types de symboles, ce qui fait que les exemples de la base ArchiDB présentent une grande variabilité dans la segmentation des tracés. Les complexités des classes peuvent être assez importantes, des exemples de certaines classes allant jusqu'à contenir une vingtaine de tracés.

5.2. Evaluation des PSF pour la représentation de symboles structurés

Ce troisième jeu de données nous permettra de confronter la représentation PSF à des symboles d'une plus grande complexité et d'éprouver le principe de positionnement des éléments de la structure par rapport à un référentiel de positionnement unique.

5.2.2 Protocole

Toutes les expérimentations sur les jeux de données sont abordés dans un cadre omniscritteur, ce qui est rendu possible par la connaissance de l'identité des scripteurs dans les trois jeux de symboles constitués.

L'algorithme d'apprentissage est donc exécuté sur les données saisies par des scripteurs distincts de ceux qui permettront de mesurer les taux de reconnaissance. Plus précisément, chaque jeu de données est décomposé en trois bases : une base d'apprentissage, une base de validation et une base de test. Pour chaque classe, les exemples de la base d'apprentissage et de validation sont utilisés par l'algorithme d'apprentissage (selon le schéma itératif représenté à la figure 4.12). La base d'apprentissage sert à entraîner le PSF, tandis que la base de validation permet d'en évaluer la qualité au cours du procédé itératif. La base de validation est également utilisée pour optimiser certains hyper-paramètres des modèles. La base de test, enfin, sert à évaluer sur un jeu indépendant la capacité de généralisation des PSF pour la reconnaissance de symboles tracés par des scripteurs différents.

Le tableau 5.25 résume pour les trois jeux de données le nombre de scripteurs et d'exemples d'apprentissage, de validation et de test.

TABLE 5.25: Statistiques (nombre de scripteurs et d'exemples) des ensembles d'apprentissage, validation et test pour les trois jeux de données. Les nombres d'exemples présentés sont cumulés sur toutes les classes.

jeu	classes	apprentissage		validation		test	
		script.	ex.	script.	ex.	script.	ex.
CaracDB	50	25	1250	15	728	20	943
PictoDB	14	4	280	4	280	2	700
ArchiDB	21	3	735	2	525	2	630

Paramètres des méta-modèles spatiaux Les paramètres des méta-modèles spatiaux composant les PSF sont les paramètres génériques déjà utilisés dans la première partie expérimentale. Les point de vue choisis sont donc les quatre directions cardinales et tous les histogrammes sont construits avec huit sections. Les modèles que nous intégrons dans cette partie pour la modélisation de la distance et de l'étendue sont ceux qui se sont révélés les plus performants dans la première partie expérimentale. Ainsi le modèle retenu pour la distance est le modèle de distance *embarquée*. Pour l'étendue, nous utiliserons la méthode de représentation par les points singuliers.

Paramètres de l'apprentissage des PSF Le procédé itératif d'apprentissage automatique des PSF requiert la définition de plusieurs paramètres : la variable de température initiale du recuit simulé, la vitesse de refroidissement, le nombre de *germes* d'initialisation et le nombre d'itérations fixé pour les étapes de convergence. Idéalement, le nombre de germes

et d'itérations doit pouvoir être assez grand. Toutefois, pour des raisons pratiques de limitation du temps nécessaire à l'apprentissage, nous avons du limiter le nombre d'itérations à un maximum de 50 (25 étapes de recherche des références et 25 étapes de recherche de la structure complète). Le nombre de germes utilisé est fixé à 50. Etant donné le faible nombre d'étapes itératives, nous avons choisi pour le recuit simulé une initialisation de la température à un niveau assez faible de 0,15. Cela signifie que même en début d'apprentissage, il n'est pas permis d'explorer des parties de l'espace de recherche trop peu prometteuses par rapport à la meilleure solution déjà obtenue. De même, la vitesse de refroidissement est assez élevée afin d'imposer une convergence assez rapide. L'importance de la bonne qualité de la solution initiale pour l'algorithme en est donc renforcée. Nous utilisons donc un nombre important de germes, ce qui permet d'augmenter la probabilité d'initialiser le recuit simulé avec une bonne solution initiale.

Paramètres d'exploitation des PSF Parmi les hyper-paramètres nécessaires à l'exploitation des PSF, nous avons défini empiriquement les constantes c_0 et c_1 . Ces valeurs déterminent respectivement le score d'un tracé non attribué et le score d'un composant non assigné dans l'évaluation d'un assignement entre un ensemble de tracés et un PSF. La valeur de c_0 est fixée arbitrairement à une valeur bipolaire de $\langle 0, 1; 0, 9 \rangle$. Il s'agit d'un *mauvais* score bipolaire, c'est-à-dire à une valeur qui est sensiblement inférieure au score bipolaire moyen constaté dans la correspondance d'un tracé avec son composant associé dans un PSF. Le score c_1 est quand à lui adapté dans chaque PSF en fonction de statistiques mesurées au cours de l'apprentissage du PSF. Si, au cours de l'apprentissage, un composant était toujours pourvu d'un tracé, son score c_1 est choisi comme très pénalisant ($c_1 = \langle 0, 1; 0, 9 \rangle$). Si en revanche on a constaté en apprentissage un certain nombre d'exemples pour lesquels le composant n'était pas pourvu, on lui attribue alors un score c_1 plus élevé en proportion, de façon à tolérer des assignements non couvrants sur ce composant.

Le poids attribué aux éléments de structure de la RVA dans la fusion globale est également déterminé arbitrairement pour tous les jeux de données. Il est fixé à $\rho = 0, 5$, ce qui signifie qu'un élément de RVA a un poids inférieur de moitié aux autres éléments de la structure. Quelques tests expérimentaux ont montré que ce poids a une influence assez faible sur les performances de reconnaissance, mais qu'il doit être sensiblement supérieur à 0.

Le paramètre de pondération entre les scores de positionnement et d'étendue des éléments de structure fait l'objet d'une optimisation sur laquelle nous reviendrons au paragraphe 5.2.3.1.

5.2.3 Expérimentations sur la base CaracDB

Comme nous l'avons précisé, c'est principalement à l'aune de la base de symboles CaracDB que nous allons valider la capacité de la représentation par PSF à conduire la tâche de reconnaissance de symboles. Outre l'évaluation de la qualité de la représentation structurelle, nous profitons de ce nouveau cadre expérimental pour confirmer les contributions de l'intégration de la distance et de l'apport de l'étendue aux méta-modèles spatiaux.

5.2. Evaluation des PSF pour la représentation de symboles structurés

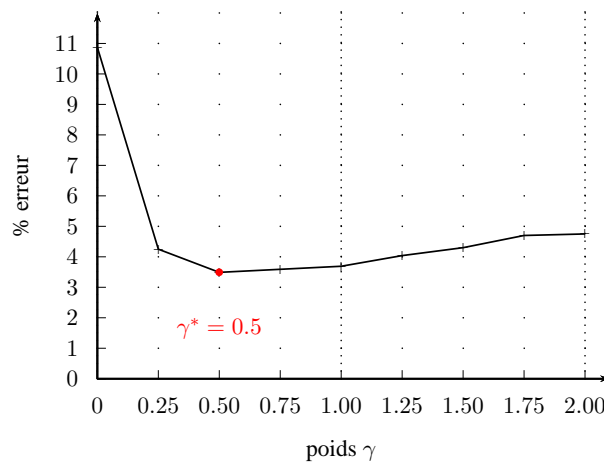


FIGURE 5.4: Evolution du taux d'erreur mesuré sur la base de validation de CaracDB en fonction du poids γ des scores d'étendue par rapport aux scores de positionnement.

5.2.3.1 Importance de la modélisation de l'étendue

Dans un premier temps, nous présentons les expérimentations permettant d'optimiser le poids donné aux scores d'étendue par rapport aux scores de positionnement issus des méta-modèles spatiaux. En référence au paragraphe 4.2.2.2, nous rappelons que le score bipolaire d'adéquation d'un tracé avec un composant structurel auquel il est assigné est exprimé par la combinaison de son score de positionnement, pondéré par un facteur α et de son score d'étendue, pondéré par β , avec $\alpha + \beta = 1$. Pour fixer ces paramètres, nous cherchons la valeur optimale du rapport $\gamma = \beta/\alpha$ sur la base de validation. Suite à l'apprentissage des modèles de PSF pour toutes les classes de la base CaracDB à l'aide des bases d'apprentissage et de validation, nous appliquons la tâche de reconnaissance à la base de validation. La courbe présentée à la figure 5.4 exprime le taux d'erreur mesuré sur pour différentes valeurs du rapport β/α entre 0 et 2.

Ce test permet de vérifier tout d'abord l'importance de la modélisation de l'étendue pour la reconnaissance de ces caractères. Si le poids de l'étendue est nul, le taux d'erreur augmente de façon très importante. La valeur retenue pour γ suite à cette optimisation est $\gamma^* = 0,5$. On attribuera donc au score d'étendue un poids deux fois inférieur au poids des scores de positionnement dans la fusion globale.

Sur la base de test, les résultats de reconnaissance mesurés sont en phase avec les résultats de la base de validation. Le tableau 5.26 présente les taux d'erreurs obtenus sur la base de test, d'abord sans intégration des scores d'étendue ($\gamma = 0$), puis avec pondération des scores d'étendue par le facteur γ^* .

TABLE 5.26: Taux d'erreur sur la base de CaracDB avec et sans utilisation des scores d'étendue.

base	% erreur ($\gamma = 0$)	% erreur ($\gamma = \gamma^*$)
CaracDB	12,62	4,03

TABLE 5.27: Les trois classes de symboles prises en exemple dans ce tableau illustrent la réduction des confusions permise par l'introduction de l'étendue dans la fusion. Pour chaque cas (a,b,c), la première colonne représente une classe de symbole A et la seconde colonne une classe concurrente B. Les deux dernières colonnes présentent le taux d'erreur pour la classe A lorsque le score d'étendue n'est pas pris en compte et lorsqu'il est intégré avec un poids γ^* .

	symbole A	symbole B	% erreur classe A ($\gamma = 0$)	% erreur classe A ($\gamma = \gamma^*$)
(a)	刀	力	31,6	5,26
(b)	寸	寺	21,05	10,53
(c)	大	尢	42,11	15,79

Ce résultat confirme l'apport de l'étendue pour la modélisation des relations spatiales, déjà mis en évidence au paragraphe 5.1.3.4. En analysant finement les résultats, on remarque que certaines classes sont particulièrement affectées par l'apport des modèles d'étendue. Le tableau 5.27 présente trois cas particuliers de classes qui bénéficient nettement de l'introduction de l'étendue dans la fusion.

Les deux cas (a) et (b) du tableau illustrent chacun des paires de classes de symboles qui sont mal reconnues en l'absence de modèle d'étendue car elles ne se distinguent que par la présence ou non d'une intersection entre des éléments. L'apport du score d'étendue est alors très nettement visible sur la réduction du taux d'erreur (dernière colonne). La notion d'étendue a été initialement introduite comme un remède contre les faiblesses des descriptions de positionnement pur notamment dans ces situations d'intersections de tracés (voir au paragraphe 3.5).

Le cas (c) ne correspond pas à des classes qui sont confondues à cause de la présence d'intersection. Il s'agit plutôt de la *forme* d'un des éléments de la structure qui diffère entre les deux symboles A et B. Le gain apporté par l'étendue sur cette classe montre que celle-ci permet également d'aider à distinguer ces deux caractères par la description de *comment un objet occupe l'espace*.

5.2. Evaluation des PSF pour la représentation de symboles structurés

5.2.3.2 Apport de la distance

La reconnaissance de symboles par PSF permet de reproduire un des tests déjà effectués dans la première partie expérimentale afin de vérifier l’apport de la distance dans les méta-modèles spatiaux. Nous comparons donc ici simplement la performance de reconnaissance des PSF sur la base de test dans le cas où leurs méta-modèles spatiaux ne comportent pas d’information de distance et dans le cas où ils intègrent une modélisation de la distance *embarquée*. Cette modélisation s’est avérée supérieure à l’intégration d’une distance globale, lors des tests de classification de relations spatiales.

Le tableau 5.28 présente ces deux taux d’erreurs mesurés sur la base de test, en intégrant les scores d’étendue pondérés d’après le résultat de l’optimisation présentée précédemment.

TABLE 5.28: Taux d’erreur sur la base de test de CaracDB avec et sans modélisation de la distance.

% erreur sans distance	% erreur avec distance
5,41	4,03

Le résultat confirme dans ce contexte de reconnaissance de symboles l’apport significatif de l’intégration des modèles de distance au sein des méta-modèles de positionnement qui constituent les PSF. Le gain en reconnaissance est ventilé assez uniformément sur les classes de la base de données et il est délicat de mettre en avant des situations précises qui bénéficient de façon significative de l’apport de la distance. Toutefois, le gain de la qualité de description offert par la considération de la distance dans les modèles spatiaux se traduit de façon perceptible sur les représentations des *cartes de segmentation* des symboles (qui permettent de visualiser les zones de dominance des modèles spatiaux associés aux composants des modèles). Ces cartes servent à conduire la segmentation des tracés et jouent par conséquent un rôle majeur dans le procédé de reconnaissance.

Le tableau 5.29 illustre des cartes de segmentations obtenues avec des modèles spatiaux purement directionnels (première colonne) et qui intègrent une information de distance (deuxième colonne).

Il apparaît nettement sur ces deux exemples que l’information de distance apporte une précision supplémentaire dans les zones de dominance des modèles. Elle renforce donc la précision de l’étape de segmentation des tracés par les méta-modèles spatiaux.

5.2.3.3 Validité de la représentation structurelle

Le taux de reconnaissance avec la meilleure configuration (modèles de distance *embarquée* et poids de l’étendue optimisé) permet d’atteindre une performance de bonne reconnaissance de près de 96% (voir le tableau 5.29). Cela signifie que les PSF sont bien adaptés à la description des symboles de la base CaracDB (les radicaux de caractères chinois) et sont capables de les reconnaître efficacement parmi 50 classes.

Ce résultat confirme tout d’abord le rôle fondamental joué par l’information spatiale dans ces symboles. Aucun élément de forme n’a en effet été modélisé, ce qui n’empêche pas d’atteindre une performance acceptable sur cette base. Le procédé d’apprentissage des PSF a été bien capable de construire des représentations adaptées des caractères, qui sont de plus suffisamment précises pour réaliser la tâche de reconnaissance. La stratégie de

TABLE 5.29: Cartes de segmentation de symboles décrits par des méta-modèles sans distance (a) et (c) ou avec distance (b) et (d). Les cartes représentent par différents niveaux de gris les zones de dominance des modèles spatiaux attachés aux composants structurels du PSF. Les tracés en couleur représentent les références et les tracés en blanc sont les éléments à assigner aux composants.

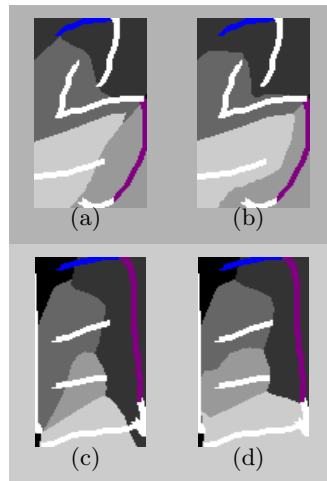


TABLE 5.30: Exemples correctement reconnus pour quatre classes de symboles. Le symbole en gras représente à chaque fois un prototype standard de la classe.



choix d'un référentiel de positionnement s'avère efficace sur cette base de données, puisqu'il est toujours possible d'extraire des tracés répondant aux critères que nous avons mis en avant : tracés plus ou moins rectilignes, verticaux ou horizontaux, de grande étendue par rapport au symbole. Le procédé d'exploitation a également montré sa capacité à surmonter correctement les variantes de segmentations qui existent dans la base.

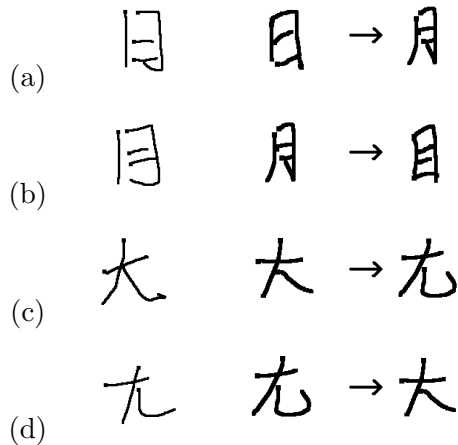
Le tableau 5.30 représente ainsi quelques exemples de symboles correctement reconnus malgré d'importantes variations de formes et différentes variantes de segmentation.

Le tableau 5.31 illustre quant à lui quelques cas d'erreurs de classification, qui mettent en évidence deux types d'erreurs à distinguer.

Les cas d'erreurs (a) et (b) sont dûs à la modélisation insuffisante de certaines relations spatiales. En effet, les caractères confondus doivent pouvoir être distingués par la description des relations spatiales entre certains éléments de leur structure. Cela montre une première

5.2. Evaluation des PSF pour la représentation de symboles structurés

TABLE 5.31: Exemples de caractères mal reconnus. La première colonne représente le symbole mal reconnu, la seconde colonne présente la classe réelle du symbole et la troisième la classe attribuée par erreur.



limite de la représentation et en particulier du choix que nous avons fait de décrire toutes relations de positionnement par rapport à un référentiel unique.

Les deux cas (c) et (d) mettent plutôt en évidence la limite intrinsèque de la description purement spatiale. En effet, la confusion entre ces symboles provient de l'absence de description de la forme. Ces cas particulièrement ambigus défient la représentation par PSF. Malgré le gain en précision apporté par la mesure de l'étendue des objets, que nous avons montré au tableau 5.27, cette mesure ne permet pas de distinguer ces cas difficiles.















5.2.4 Limitations de la représentation

Nous avons déjà constaté des limitations de la représentation par PSF pour la reconnaissance de symboles sur la base CaracDB. Pour éprouver plus en profondeur les limites de la représentation purement spatiale des symboles, nous expérimentons dans cette partie les PSF face aux bases de données PictoDB et ArchiDB. La base PictoDB va d'abord nous permettre de sonder les limites de la tolérance des PSF face à des symboles qui présentent de plus grandes distorsions de forme et de segmentation de leurs tracés. Ensuite, une expérimentation sur la base ArchiDB testera la description sur des symboles graphiquement plus complexes.

5.2.4.1 Limites de la tolérance aux distorsions de segmentation

La base de symboles PictoDB constitue un défi pour construire une représentation structurelle des caractères. La reconnaissance de ces symboles par PSF permet toutefois d'atteindre une performance de 83% de bonne reconnaissance. Le tableau 5.32 donne plus en détail les scores de reconnaissance atteints pour chacune des 14 classes de symboles de la base PictoDB.

TABLE 5.32: Taux d'erreur pour chaque classe de la base PictoDB.

classe	% erreur	classe	% erreur	classe	% erreur
	34		32		2
	4		2		18
	0		66		2
	8		2		12
	54		2	Total	17

Les mauvais taux de reconnaissance atteints pour certaines classes sont à mettre au compte de la grande difficulté pour l'algorithme d'apprentissage à converger vers des modèles stables de PSF. Cette difficulté provient en premier lieu de la grande distorsion des tracés et au caractère très cursif de certains exemples, la stratégie d'extraction de référence étant peu adaptée à la gestion de tracés courbes. Par exemple, en présence de cercles, les références extraites sont assez instables car elles résultent d'une procédé d'approximation polygonale qui est peu pertinent sur ces formes. Cela explique notamment le mauvais taux de reconnaissance pour la classe représentée en bas à gauche du tableau 5.32. La classe la plus mal reconnue (au centre du tableau) cumule deux difficultés. D'abord, la cursivité des tracés rend difficile l'extraction de références stables. Ensuite, elle a la particularité d'avoir un nombre d'éléments de structure variable, très difficile à déterminer de façon stable par l'algorithme d'apprentissage.

La figure 5.5 illustre les difficultés que posent certaines classes pour construire des modèles stables de leur structure.

Pour les symboles de la première ligne, le caractère très cursif de leur écriture pose un défi pour la méthode d'extraction des références qui est peu adaptée notamment à la gestion des tracés circulaires. Les symboles de la seconde ligne ont la particularité très marquée d'être fortement variables en fonction des styles d'écriture. Cette classe de symboles pourrait être définie par : *trois tracés non rectilignes, descendants, plus ou moins parallèles*. On comprend aisément en regardant ces exemples que la méthode d'apprentissage des PSF, qui a été conçue de façon à pouvoir détecter par exemple la présence de *crochets* à l'extrémité de tracés dans des caractères chinois, est mise à mal par ce type de formes instables. L'algorithme échoue en fait à déterminer des références et des éléments de structure de façon stable. Le résultat est donc un PSF mal défini, qui reconnaît mal les tracés de cette classe et qui perturbe la reconnaissance des autres classes.

En dépit de quelques classes très mal modélisées, la représentation par PSF permet glo-

5.2. Evaluation des PSF pour la représentation de symboles structurés

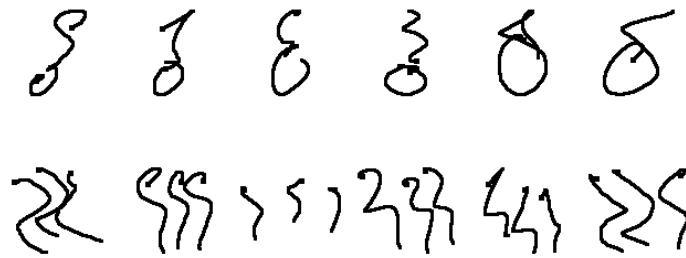


FIGURE 5.5: Illustration des difficultés posées par deux classes de symboles de la base PictoDB.

bablement de maintenir un niveau de reconnaissance de 83%, ce qui montre une certaine capacité à gérer des cas de segmentation beaucoup plus difficiles que dans la base CaracDB. Lorsque l'algorithme d'apprentissage parvient à apprendre des représentations suffisamment stables, leur exploitation permet de procéder à la reconnaissance avec une précision raisonnable. La limite qui apparaît clairement sur cette base de données vient avant tout de la stratégie d'extraction de points particuliers pour générer les hypothèses de références, qui s'applique mal à plusieurs de ces classes.

5.2.4.2 Limites de la précision de la représentation par PSF

La base de données ArchiDB comporte 21 classes, dont certaines correspondent à des symboles nettement plus complexes que les deux bases précédentes, puisque certains exemples sont constitués de près d'une vingtaine de tracés. Ces symboles posent alors un défi à la représentation par PSF. D'abord, la complexité combinatoire engendrée est nettement augmentée car elle croît de façon exponentielle avec le nombre de tracés élémentaires constituant les symboles. Il semble qu'un vingtaine de tracés soit la limite maximale acceptable pour les algorithmes développés. Ensuite, cette complexité permet de pointer du doigt les limites de la précision de la description spatiale. Un choix de conception des PSF a été de sélectionner un référentiel unique par rapport auquel tous les autres éléments de tracés sont positionnés. Lorsque de nombreux éléments de structure constituent les symboles, il apparaît que cette description est insuffisante et qu'il est nécessaire de décrire le positionnement relatif des éléments entre eux et pas seulement par rapport à une référence centralisée. C'est d'ailleurs le choix que font Mas, Sanchez *et al.* [MSLL07] pour reconnaître ces mêmes classes de symboles.

Une expérimentation sur cette base de 21 classes donne un taux de bonne reconnaissance de l'ordre de 70% seulement. La figure 5.6 donne trois exemples de classes qui sont particulièrement mal reconnues par l'approche PSF.

Ce qui caractérise ces symboles est qu'il ne présentent pas d'éléments de structure saillants constituant des références naturelles pour décrire le positionnement de tous les éléments de la structure. Ils sont constitués d'un nombre assez important d'éléments qu'il est difficile de positionner précisément par rapport à un référentiel unique. L'analyse des PSF appris pour ces classes montre que la convergence est bien réussie, mais que les modèles spatiaux sont peu précis et qu'en particulier la carte de segmentation est peu efficace pour



FIGURE 5.6: Exemples de classes mal reconnues dans la base ArchiDB.

segmenter certains tracés trop éloignés des éléments de référence. On atteint donc avec ces exemples la limite de la capacité de description de symboles par les PSF tels qu'ils ont été définis.

5.2.5 Synthèse sur la représentation de symboles par PSF

Les expérimentations conduites dans cette deuxième partie ont permis de mettre en évidence en premier lieu la bonne capacité des PSF à représenter de façon précise des caractères chinois simples. Les résultats de reconnaissance obtenus sur la base CaracDB valident la démarche purement spatiale que nous avons adoptée : l'information sur la structure interne de ces symboles est intrinsèquement assez riche pour pouvoir les reconnaître de façon efficace. Du même coup, il est confirmé que les méta-modèles spatiaux offrent une description fidèle de cette information. De plus, leur pouvoir de prédiction exploité ici pour segmenter les tracés en fonction de leur positionnement relativement à des références est efficace et permet de faire face à diverses situations de segmentation. Les apports obtenus par l'introduction dans les modèles de l'information de distance et des scores d'étendue permettent en outre de consolider les conclusions de la première partie expérimentale.

Plusieurs limitations ont également été mises en évidence dans ces expérimentations, en particulier à l'aide des deux bases PictoDB et ArchiDB. Tout d'abord, par définition, les PSF n'ont pas été entraînés et optimisés pour maximiser leur capacité à reconnaître des symboles. Une meilleure performance pourrait donc être espérée non pas en modifiant leur définition, mais par exemple en optimisant les poids de leurs éléments de structure dans la fusion dans un cadre discriminant.

Ensuite, la première limitation réelle des PSF a été constatée dans le cadre de symboles à la segmentation très chaotique et en particulier lorsque les tracés sont très cursifs. L'extraction de références stables au cours de l'apprentissage est alors difficile et la stratégie d'extraction de références s'avère inadaptée à traiter les tracés circulaires. Cette limitation pourrait donc être contournée en redéfinissant une stratégie plus générale pour extraire des références stables y compris en l'absence de portions rectilignes de tracés.

La seconde limitation constatée provient du choix de positionner tous les éléments par rapport à un référentiel de positionnement unique. Ce choix limite la capacité de description à des symboles à la complexité assez réduite. Dépasser cette limitation en expressivité demanderait de modéliser des relations spatiales supplémentaires entre les éléments de tracés et non pas seulement par rapport aux références globales. Un coût combinatoire plus important serait alors engendré par une telle représentation et une adaptation des algorithmes serait nécessaire (il faudrait en particulier améliorer la fonction de *bounding* utilisée dans la recherche du meilleur assignement).

5.2. Evaluation des PSF pour la représentation de symboles structurés

Il est intéressant de noter que ces limitations ne proviennent pas du choix de représentation purement spatiale des symboles. Ce n'est donc pas en premier lieu la restriction à l'information spatiale qui limite l'approche de représentation de symboles par PSF, mais plutôt les choix que nous avons fait pour simplifier cette information (en particulier le choix d'un référentiel unique de positionnement).

Conclusion

6.1 Bilan

Dans ces travaux, nous nous sommes intéressés au problème de la description de l'information spatiale entre objets manuscrits pour aider à l'interprétation des documents de cette nature. Faisant le constat d'un certain manque dans la littérature, nous avons proposé une modélisation pour les relations spatiales qui soit transversale aux besoins rencontrés dans l'analyse de documents manuscrits, c'est-à-dire avec l'objectif de pouvoir s'appliquer à des situations de positionnement variées, entre objets de différentes natures et différents niveaux de complexité. L'idée conductrice de nos recherches est que les relations spatiales renferment une importance majeure quant au sens à interpréter et qu'elles ne sont probablement pas exploitées dans toute leur richesse par les techniques de description existantes.

Les méta-modèles spatiaux que nous avons introduit reposent sur les fondements de la perception cognitive de relations spatiales et offrent une représentation intuitive pour juger du positionnement relatif de deux objets. L'intérêt majeur de cette approche est de permettre la conduite d'un raisonnement spatial au sein même de l'espace de perception ce qui autorise une prise de décision tenant compte de la situation de positionnement sans simplification ni approximation des formes ou des relations entre les objets concernés. Cette caractéristique se traduit aussi par la capacité de la représentation à prédire, étant donné un objet de référence, quelles zones du plan sont acceptables pour la relation apprise.

La formalisation des méta-modèles spatiaux a été guidée par les objectifs de généralité et d'expressivité de la modélisation. La sélection de directions de raisonnement et l'incorporation de distance dans les modèles ont été abordés dans cette optique. L'expressivité et la précision de la méthode de description reposent sur deux aspects fondamentaux. D'abord, l'analyse de la relation est conduite en considérant la combinaison de plusieurs points de vue qui sont choisis afin d'offrir une bonne couverture des positionnements relatifs possibles. Ensuite, l'apprentissage automatique de la description selon chaque point de vue est essentielle

et permet d'appréhender des relations spatiales complexes dans toute leur richesse. L'originalité de la prise en compte du caractère bipolaire de l'information spatiale permet en outre de modéliser finement des nuances d'acceptabilité et d'inacceptabilité de positionnement. Enfin, la considération d'un score d'étendue de l'objet cible, décrivant comment l'objet occupe l'espace, s'est avérée capable d'apporter à la modélisation une capacité supplémentaire à distinguer des situations qui ne peuvent pas l'être au seul regard du positionnement strict des objets. La possibilité d'apprendre automatiquement les méta-modèles spatiaux disposant de ces qualités est essentielle pour appréhender les relations spatiales non-linguistiques qui se présentent entre les objets manuscrits. Elle constitue aussi, sur le plan théorique, une réelle nouveauté conceptuelle au regard des méthodes existantes s'inspirant du raisonnement par modèles spatiaux.

Les résultats obtenus par exploitation de cette modélisation ont permis de vérifier, tant qualitativement que quantitativement, que la précision de la description du positionnement relatif offerte surpasse les familles de méthodes classiquement utilisées pour cela, sur différents jeux de données présentant diverses natures de relations spatiales.

La deuxième partie de nos travaux a permis d'approfondir la démarche de mise en évidence du rôle de l'information spatiale pour l'interprétation du manuscrit. Si l'information spatiale joue effectivement un tel rôle, et puisque nous avons défini une modélisation permettant de la décrire plus fidèlement que les méthodes existantes, dans quelle mesure n'est-il pas possible d'interpréter le sens de certaines structures manuscrites en se limitant à la description de cette connaissance ?

Nous avons défini une représentation structurelle nouvelle pour décrire les symboles manuscrits, à partir de la combinaison de méta-modèles spatiaux modélisant le positionnement relatif des éléments de la structure. Les enjeux supplémentaires par rapport à la description du positionnement de deux objets prédéfinis concernent la détermination d'objets de référence et la recherche de la segmentation des éléments de la structure. Des stratégies cohérentes avec le mode de description des relations spatiales ont été définies pour résoudre ces deux problèmes. La recherche d'un référentiel de positionnement est faite avec l'objectif d'optimiser le pouvoir descriptif offert par ce référentiel au regard du reste de la structure du symbole. La segmentation des éléments de tracés est quant à elle guidée par les modèles spatiaux appris qui décrivent leur positionnement attendu par rapport au référentiel. La propriété de combinaison des modèles spatiaux est aussi exploitée pour construire des relations de positionnement plus précises, exprimées par rapport à une paire de références. Cette seconde partie de nos travaux permet donc également de mettre en évidence les contributions que peuvent apporter les méta-modèles spatiaux aux systèmes d'analyse d'expressions manuscrites.

La qualité de la représentation structurelle ainsi définie a été évaluée en procédant à une tâche de reconnaissance automatique de différents types de symboles. Les résultats obtenus sur une base de symboles chinois prouvent que dans ce contexte, cette représentation purement spatiale des symboles est suffisamment précise pour permettre leur bonne reconnaissance. Cela confirme à la fois la richesse de l'information de positionnement relatif que renferment ce type de formes et la qualité des méta-modèles spatiaux à bien représenter cette information. La bonne tolérance face à différentes situations de segmentation conforte par ailleurs la stratégie de segmentation guidée par les modèles de positionnement. D'autres

6.2. Perspectives

expérimentations sur des symboles posant des défis par la plus grande variabilité de leur segmentation ou par la plus grande complexité de leur composition ont mis en évidence les limites de notre représentation. Les limites identifiées ne sont paradoxalement pas imputables en premier lieu à l'absence de description des formes individuelles des tracés qui constituent les symboles. Au contraire, ce sont encore certaines caractéristiques spatiales qui pourraient être intégrées à la description pour en enrichir la précision (en particulier le positionnement relatif de tous les éléments de structure relativement les uns aux autres).

Il serait, pour conclure, inapproprié de se risquer à formuler une limite absolue à la richesse comportée par l'information spatiale d'une expression manuscrite : nous n'avons fait dans cette thèse qu'en sonder la profondeur au moyen de ces nouveaux outils que sont les méta-modèles spatiaux. D'autre part, chercher à quantifier la plus-value d'une description des formes des éléments de tracés tient de la gageure. En effet, l'importance de l'information de forme des primitives structurelles semble s'amenuiser à mesure que ces primitives sont définies comme des portions de tracés plus petites, plus élémentaires, aux formes plus simples. A l'inverse, l'information spatiale revêt alors une part d'importance croissante. Un élément de réponse qui semble émerger est la limite pratique posée par la complexité combinatoire engendrée lorsque le modèle de structure s'enrichit d'un grand nombre de relations spatiales.

6.2 Perspectives

Le recul sur les travaux présentés dans ce document nourrit trois catégories de perspectives à considérer pour poursuivre ces recherches. Les premières sont des perspectives de fond sur la modélisation du positionnement spatial : nous évoquons pour cela quelques pistes pour renforcer le pouvoir d'expressivité des méta-modèles spatiaux. Les secondes se rapportent à la représentation de structures par Patrons Structurels Flous : nous présentons quelques avancées possibles pour améliorer leur pouvoir de description. Les troisièmes concernent les contributions possibles de la représentation de l'information spatiale par méta-modèles de positionnement au service de l'interprétation de documents manuscrits structurés : nous esquissons aussi quelques perspectives dans cette direction.

Dans la définition des méta-modèles spatiaux, il pourrait être bénéfique d'adopter une stratégie plus intelligente de sélection des directions sous lesquelles la relation spatiale entre deux objets est examinée (choix des points de vue). Cette sélection pourrait ainsi être guidée par une connaissance sur la forme de l'objet de référence : il est plus ou moins informatif d'examiner le positionnement *à droite* d'un objet selon qu'il s'agit d'un objet étiré verticalement ou horizontalement par exemple. Cette adaptation du choix des directions pourrait permettre de définir également des modèles de relations dont le référentiel est lié à l'échelle ou à l'orientation de l'objet de référence, par exemple dans le but de créer des modèles invariants en rotation.

Une seconde piste qui reste partiellement inexplorée concerne le potentiel de description offert par les mesures a posteriori des points d'un objet argument par rapport à un modèle spatial. Nous avons fait usage de mesures a posteriori pour décrire l'étendue des objets, ce qui s'est avéré bénéfique pour discerner plusieurs cas de positionnement spatial. Il est

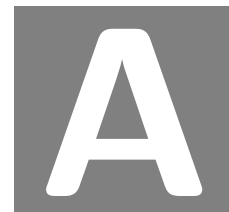
évident que d'autres mesures complémentaires de cette nature seraient capables de renforcer le pouvoir d'expressivité des modèles. Par exemple, l'analyse de la distribution des degrés de distance des points d'une forme linéique par rapport à une autre peut permettre d'évaluer une situation de parallélisme.

La seconde partie de nos travaux a présenté une utilisation possible des méta-modèles spatiaux pour l'analyse de symboles manuscrits. Cette démarche n'est absolument pas une fin en soit : son rôle a été de confirmer les suppositions initiales concernant l'importance de l'information spatiale ainsi que de démontrer la capacité de l'outil que constituent les méta-modèles pour décrire au mieux cette information.

Des possibilités d'amélioration de la description offerte par les Patrons Structurels Flous peuvent être envisagées comme une prolongation de ce travail. Il serait intéressant de considérer des techniques plus robustes pour mieux extraire des hypothèses de références sur des tracés curvilignes et dont la segmentation est instable. De plus, une réflexion pourra être conduite afin d'adapter la nature du référentiel de positionnement de façon à améliorer le pouvoir de description face à des symboles de plus grande complexité : sur combien de références faut-il appuyer la description ? Quels sont les éléments de tracés dont le positionnement relatif est significatif ?

La perspective de travail naturelle appelée par nos conclusions est de mettre à profit les méta-modèles spatiaux pour renforcer la qualité des relations spatiales décrites au sein de systèmes structurels d'analyse du manuscrit. Il s'agit là de la troisième catégorie de perspectives. Plus généralement que pour la reconnaissance de symboles structurés, c'est pour l'interprétation d'expressions complexes que les méta-modèles spatiaux sont susceptibles d'avoir la plus grande contribution : notations mathématiques, schémas techniques... Dans ces applications aussi, l'information de positionnement spatial joue un rôle majeur qui n'est pas exploité à fond. La qualité des méta-modèles spatiaux ainsi que leur capacité de prédiction peut être efficacement mise au service des stratégies automatiques d'analyse structurelle pour l'interprétation de ce type de données.

Dans le cadre de périphériques à interfaces stylo, la capacité de prédiction peut aussi être mise au service de l'utilisateur pour lui suggérer des zones de tracés potentielles en fonction de ce qu'il a déjà saisi. Les systèmes reposant sur une analyse incrémentale des tracés de l'utilisateur pour la composition de documents complexes (comme dans les travaux de Macé [MA09]) pourraient en particulier bénéficier de cette capacité.



Introduction à la logique floue

La théorie de la logique floue a été formalisée pour la première fois par Zadeh en 1965 [Zad65]. Elle est particulièrement bien adaptée au traitement de données *imprécises*, qui composent souvent le monde réel. C'est pour cette raison qu'elle est aujourd'hui utilisée dans des domaines très variés tels que l'informatique, mais aussi la robotique, la médecine, l'environnement... La logique floue s'appuie sur la théorie des sous-ensembles flous, qui est une extension de la théorie des ensembles classiques avec une prise en compte d'ensembles définis de manière qualitative.

Dans cette section, nous rappelons les concepts fondamentaux associés à la théorie de la logique floue. Nous commençons par une présentation « intuitive » des sous-ensembles flous, puis réalisons une présentation plus formelle. Cette introduction est inspirée des travaux de Bouchon-Meunier [BMM03]. Le lecteur intéressé pourra se référer à cet ouvrage pour des informations plus complètes.

A.1 Concepts généraux

La théorie des sous-ensembles flous généralise le concept de sous-ensemble « classique » en introduisant un *caractère graduel* à la notion d'appartenance d'un élément à un sous-ensemble. Ce caractère graduel et son intérêt sont illustrés par un exemple sur la figure A.1 : il s'agit de l'appartenance au concept de « jeunesse » en fonction de l'âge. La figure A.1(a) présente une fonction dite non-graduelle, définissant alors que l'appartenance à ce concept s'arrête à un âge bien défini (ici 35 ans). Cette définition paraît peu adaptée à une modélisation réelle du concept de jeunesse. Au contraire, la figure A.1(b) présente une fonction graduelle définissant ce même concept. Elle modélise le fait que l'appartenance au concept de jeunesse décroît avec l'âge, ce qui paraît plus naturel.

Le caractère graduel permet de modéliser des connaissances qui sont imprécises, soit de par leur aspect qualitatif (langage naturel), soit de par le biais de leur obtention (instrument de mesure). Il découle alors de la notion de sous-ensembles flous un ensemble de concepts

"Jusqu'à quel âge peut-on dire que quelqu'un est jeune ?"

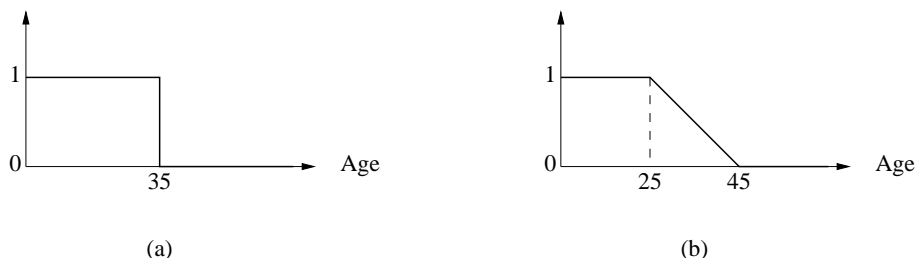


FIGURE A.1: Caractère non graduel (a) et caractère graduel (b) du degré d'appartenance.

permettant une meilleure approche de la réalité, l'introduction de transitions graduelles entre différentes catégories, l'apport d'une notion d'appartenance d'un élément à plusieurs classes (on parle de multi-appartenance) ou encore la fusion de ces notions graduelles pour prendre des décisions les plus robustes possibles.

A.2 Définition d'un sous-ensemble flou

Sous-ensemble flou Soit E un ensemble dénombrable ou non, appelé *ensemble de référence*. Un *sous-ensemble flou* A de E est défini par une fonction d'appartenance de μ_A à valeurs dans $[0; 1]$ qui associe à chaque élément x de E un degré d'appartenance $\mu_A(x)$ indiquant le degré d'appartenance de x à A .

On peut noter, pour le cas particulier où la fonction d'appartenance μ_A prend ses valeurs dans $\{0; 1\}$, que le sous-ensemble flou A est un sous-ensemble classique.

Il existe différents types de fonctions d'appartenance que l'on choisit généralement régulières et convexes pour leurs bonnes propriétés (la figure A.1(b) présente un exemple de fonction d'appartenance).

A.3 Caractéristiques d'un sous-ensemble flou

Dans cette section, nous définissons les caractéristiques de base permettant la représentation d'un sous-ensemble flou noté A sur un ensemble de référence noté E . Ces caractéristiques sont généralement fortement liées à la forme de la fonction d'appartenance μ_A de A . L'ensemble de ces caractéristiques sont illustrées sur la figure A.2.

Support On appelle *support* de A , noté $Supp(A)$, l'ensemble des éléments de E pour lesquels le degré d'appartenance à A n'est pas nul :

$$Supp(A) = \{x \in E | \mu_A(x) > 0\}.$$

Noyau On appelle *noyau* de A , noté $Noy(A)$, l'ensemble des éléments de E pour lesquels le degré d'appartenance à A est égal à 1 :

$$Noy(A) = \{x \in E | \mu_A(x) = 1\}.$$

A.4. Opérations de base sur les sous-ensembles flous

Hauteur On appelle *hauteur* de A , noté $h(A)$, la plus grande valeur prise par la fonction d'appartenance associée à A :

$$h(A) = \sup_{x \in E} \mu_A(x).$$

A est dit *normalisé* si $h(A) = 1$.

α -coupes On appelle α -*coupe* de A , noté A_α , l'ensemble des éléments de E pour lesquels le degré d'appartenance à A est au moins égal à α :

$$A_\alpha = \{x \in E \mid \mu_A(x) \geq \alpha\}.$$

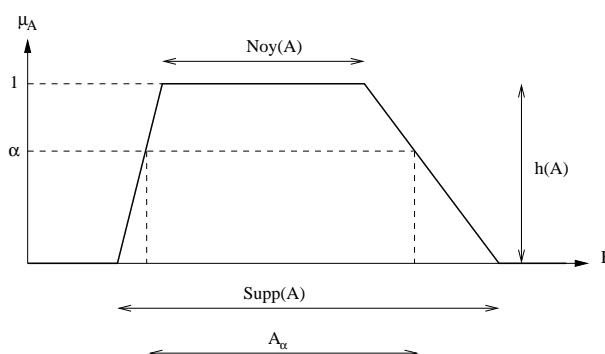


FIGURE A.2: Caractéristiques d'un sous-ensemble flou A .

A.4 Opérations de base sur les sous-ensembles flous

Le concept de sous-ensembles flous étant une généralisation du concept de sous-ensembles classiques, les opérations ensemblistes classiques ont fait l'objet d'extensions aux sous-ensembles flous.

Soient A et B deux sous-ensembles flous d'un même ensemble de référence E , on généralise alors les opérations ensemblistes sur A et B de la manière suivante :

Égalité A et B sont dits *égaux* si leurs fonctions d'appartenance prennent la même valeur pour tout élément de E :

$$A = B : \forall x \in E, \mu_A(x) = \mu_B(x).$$

Inclusion On dit que A est *inclus* dans B si tout élément x de E appartient à B avec un degré au moins aussi grand que celui de son appartenance à A :

$$A \subseteq B : \forall x \in E, \mu_A(x) \leq \mu_B(x).$$

Intersection L'*intersection* de deux sous-ensembles flous A et B de E est un sous-ensemble flou composé des éléments de E auxquels on attribue le plus petit des deux degrés d'appartenance à A et à B :

$$C = A \cap B : \forall x \in E, \mu_C(x) = \min(\mu_A(x), \mu_B(x)).$$

Union L'union de deux sous-ensembles flous A et B de E est un sous-ensemble flou composé des éléments de E auxquels on attribue le plus grand des deux degrés d'appartenance à A et à B :

$$C = A \cup B : \forall x \in E, \mu_C(x) = \max(\mu_A(x), \mu_B(x)).$$

Complément On définit le *complément* A^C d'un sous-ensemble flou A de E de telle sorte qu'un élément x de E appartient d'autant plus à A^C qu'il appartient peu à A :

$$A^C : \forall x \in E, \mu_{A^C}(x) = 1 - \mu_A(x).$$

A.5 Normes et conormes triangulaires

La généralisation de l'union et de l'intersection ensembliste classique aux sous-ensembles flous n'est en fait pas définie de manière unique. Zadeh a proposé, en premier, d'utiliser les opérateurs *min* pour l'intersection et *max* pour l'union de sous-ensembles flous, comme nous l'avons défini dans la section précédente.

Il existe cependant une multitude d'opérateurs que l'on peut utiliser pour mettre en œuvre les opérations d'union et d'intersection. Ces opérateurs sont regroupés en deux familles :

- les *normes triangulaires*, notées *T-normes*, qui définissent les opérateurs d'*intersection* ou de *conjonction*,
- les *conormes triangulaires*, notées *T-conormes*, qui définissent les opérateurs d'*union* ou de *disjonction*.

T-norme $T : [0; 1] \times [0; 1] \rightarrow [0; 1]$ est une fonction de type T-norme si et seulement si pour chaque élément x, y et z de $[0; 1]$, elle vérifie les propriétés suivantes :

- | | |
|---|--|
| (T_1) $T(x, 1) = x$ | <i>existence d'un élément neutre 1</i> |
| (T_2) $x \leq y, u \leq v \Rightarrow T(x, u) \leq T(y, v)$ | <i>isotonie</i> |
| (T_3) $T(x, y) = T(y, x)$ | <i>commutativité</i> |
| (T_4) $T(x, T(y, z)) = T(T(x, y), z)$ | <i>associativité.</i> |

T-conorme $S : [0; 1] \times [0; 1] \rightarrow [0; 1]$ est une fonction de type T-conorme si et seulement si pour chaque élément x, y et z de $[0; 1]$, elle vérifie la propriété :

- | | |
|-----------------------|--|
| (S_1) $S(x, 0) = x$ | <i>existence d'un élément neutre 0</i> |
|-----------------------|--|

ainsi que les propriétés (T_2) , (T_3) et (T_4) .

Il est possible de définir tout un ensemble de mises en œuvres différentes de ces deux familles d'opérateurs. Elles ont alors des propriétés différentes qui vont induire le raisonnement qui va en découler. En fait, le choix d'une T-norme et d'une T-conorme est très lié et on choisit souvent d'associer les fonctions *duales* pour la conduite d'un raisonnement logique flou.

Dualité Une T-norme T et une T-conorme S sont duales si et seulement si :

$$1 - T(x, y) = S(1 - x, 1 - y)$$

$$1 - S(x, y) = T(1 - x, 1 - y).$$

A.5. Normes et conormes triangulaires

La dualité permet de conserver les lois de De Morgan classiques. Le tableau A.1 présente les principales T-normes et T-conormes duales, couramment utilisées dans les systèmes de contrôle flou.

T-norme	T-conorme	nom
$\min(x, y)$	$\max(x, y)$	Zadeh
$x.y$	$x + y - x.y$	probabiliste
$\max(x + y - 1, 0)$	$\min(x + y, 1)$	Lukasiewicz

TABLE A.1: Exemples de T-normes et T-conormes duales.

Aspects algorithmiques pour le calcul des dilatations morphologiques

L'exploitation des méta-modèles de positionnement repose sur la description de relations spatiales linguistiques qui sont construites par dilatation morphologique. C'est cette opération morphologique qui concentre l'essentiel de la complexité de calcul, puisque l'exploitation des modèles *appris* se fait ensuite par la simple application des fonctions histogrammes sur les degrés mesurés dans chaque point de vue (voir l'équation (3.4)).

Le calcul de dilatation morphologique se fait, pour un objet R et un élément structurant ν , selon l'équation suivante :

$$\forall p \in \mathcal{S}, \delta_\nu(\mu)(p) = \sup_{q \in R} \nu(p - q), \quad (\text{B.1})$$

avec t une t -norme, p un point de l'image \mathcal{S} et q un point de R . Ce calcul doit être répété pour chaque point de l'image dans laquelle on veut évaluer une relation de positionnement. La complexité du calcul engendré dépend donc de la définition de l'objet R . Dans le contexte applicatif envisagé (le traitement de données manuscrites), on peut considérer que les objets sont toujours des formes *linéiques* (sans épaisseur), à la différence des objets généralement considérés en analyse d'images. Nous présentons ici les aspects algorithmiques et une analyse de complexité pour le calcul de l'opération de dilatation (pour un seul point p de l'image) selon que l'objet manuscrit de référence R est représenté par un signal *hors-ligne* (représenté par un signal de type image) ou *en-ligne* (représenté par une séquence de points échantillonnés).

B.1 Cas de données image (représentation hors-ligne)

Dans le cas de données hors-ligne, la représentation de l'objet R est une matrice de pixels blancs et noirs (il est possible de binariser l'image source pour se ramener à cette représentation). D'après l'équation (B.1), le calcul de la dilatation d'un objet R par un

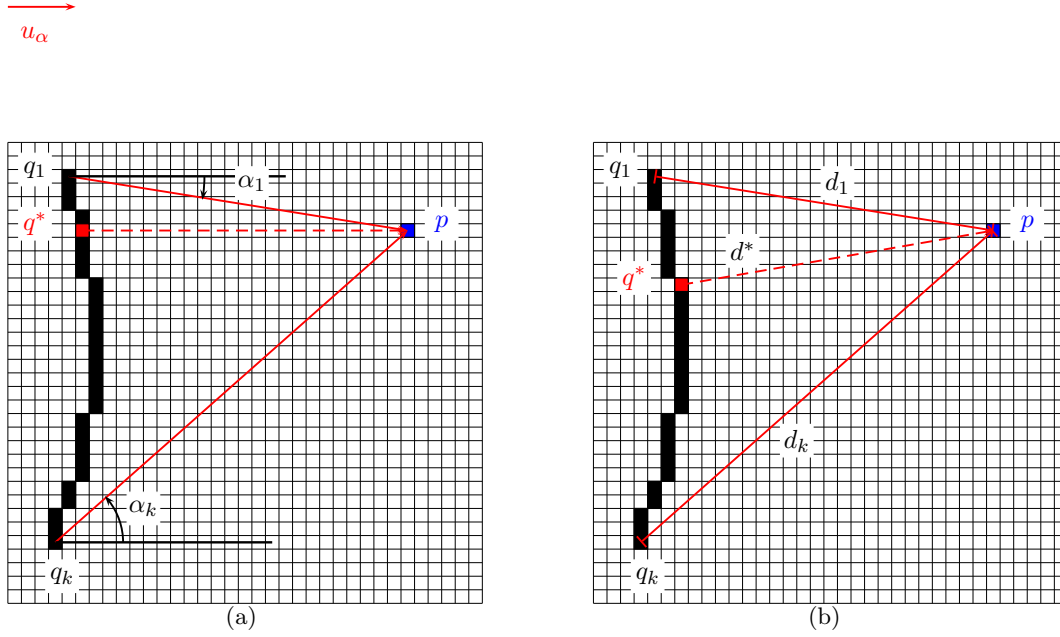


FIGURE B.1: Calcul de la dilatation d'un objet représenté par une matrice de pixels. Dans le cas d'une dilatation par un élément structurant directionnel, le calcul revient à chercher le point q^* de R qui minimise l'angle $\widehat{qp, u_\alpha}$ (a). Pour un élément de distance, il faut déterminer le point qui minimise la distance q^*p (b).

élément structurant nécessite le parcours de tous les points de R . Pour un point p du plan, il faut chercher le point q^* dans R qui permet de maximiser :

$$\nu(p - q). \quad (\text{B.2})$$

Finalement, l'appartenance de p à la dilatation de R par ν est égale à son appartenance à l'élément structurant centré sur q^* .

Le déroulement de cet algorithme de calcul dans le cas où R est représenté par une matrice de pixels autorise simplement à une optimisation par un changement de la représentation. En général, il est intéressant de dresser la liste des points de R qui sont beaucoup moins nombreux que N (nombre de pixels de l'image). C'est particulièrement vrai dans la mesure où les objets n'ont pas d'épaisseur, contrairement à des objets image en général. La discrétisation de l'espace donne également de l'intérêt au fait de tabuler les calculs de ν pour les différents vecteurs $\vec{q}\vec{p}$.

Le déroulement algorithmique est donc très simple. On peut le résumer par un schéma tel que ceux de la figure B.1. Pour les deux cas représentés (éléments structurant de direction (a) ou élément structurant de distance (b)), le déroulement est le même et consiste à chercher le point q^* de la référence qui optimise le degré de $\nu(p - q)$. Dans le premier cas (image (a)), la fonction ν modélise la similarité entre l'angle \widehat{qp} et u_α , tandis que dans le second cas (image (b)), ν modélise une relation de proximité entre q et p .

La complexité du calcul de l'appartenance d'un point au paysage en $\mathcal{O}(N)$, avec N le nombre de pixels de l'image, est réduite à $\mathcal{O}(N_R)$, avec N_R le nombre de points de R . Le

B.2. Cas de données vectorisées (représentation en-ligne)

calcul d'un score de correspondance entre un objet A et un modèle spatial construit par dilatation sur une référence R a donc une complexité en $\mathcal{O}(N_R N_A)$, avec N_A le nombre de points de A .

Dans ses travaux, Bloch propose une optimisation pour calculer plus rapidement une approximation du paysage flou directionnel [Blo99]. Etant donné que nos objets ont un nombre de points beaucoup plus petit que dans le cas d'analyse d'images en général, nous ne mettons pas en place ce type de calcul d'approximation.

B.2 Cas de données vectorisées (représentation en-ligne)

Dans le cas où les données sont issues d'un procédé d'acquisition dit *en-ligne*, le tracé manuscrit est représenté par une suite de coordonnées, échantillonnées à intervalles de temps réguliers, le long de la trajectoire effectuée par le pointeur (stylet électronique, doigt) sur la surface de saisie (tablette graphique, écran sensitif, surface tactile). La représentation du tracé est donc une liste de coordonnées de points, qui constitue une approximation polygonale de la trajectoire réelle. Il s'agit d'une représentation vectorielle qui offre des avantages pour la performance de l'algorithme de dilatation morphologique. Certains points de l'échantillonnage dénotent des ruptures dans le tracé. Ils correspondent à des points où le contact entre le pointeur et la surface a été interrompu (par exemple un levé de stylo). Ces points délimitent des portions continues de tracés, que l'on appellera ici tracés naturels (ils correspondent aux portions de tracé saisies entre un posé et un levé de stylo).

Le tracé S a pour représentation en-ligne la suite de points de l'espace \mathcal{S} qui sont regroupés en t tracés naturels $s_{i,i=1..t}$:

$$S = s_1..s_t, t \geq 0,$$

avec $s_i = p_1^i..p_{n_i}^i$ un tracé primitif constitué de n_i points de \mathcal{S} . Formellement, le tracé S est assimilé à l'ensemble de points du plan défini par :

$$S = \cup_{i=1}^t \cup_{j=1}^{n_i-1} [p_j^i; p_{j+1}^i].$$

L'objet est décrit comme un ensemble de segments qui relie deux points successifs au sein d'un même tracé naturel. Cette représentation d'objet par un ensemble de segments offre naturellement une description de formes *linéiques*, sans épaisseur, avantageuse pour la calcul de la dilatation.

B.2.1 Cas d'un élément structurant directionnel

La figure B.2 montre le calcul de la dilatation d'un tracé $q_1q_2q_3q_4$, en deux points p du plan. Ici, l'élément structurant appliqué est de direction α et le résultat de la dilatation en un point p passe par la recherche du point q^* de la référence tel que \vec{qp} soit aussi proche que possible du vecteur de direction α . La figure montre les deux cas de figure qui peuvent se présenter, selon qu'il existe un point q de la référence R tel que \vec{qp} est de direction α (c'est le cas de la figure (b)) ou qu'il n'en existe pas (cas (a)).

Pour la situation (a), il suffit de calculer la déviation entre les angles \vec{qp} et α pour tout sommet q de l'objet de référence. On se rend compte que la déviation minimale α_1

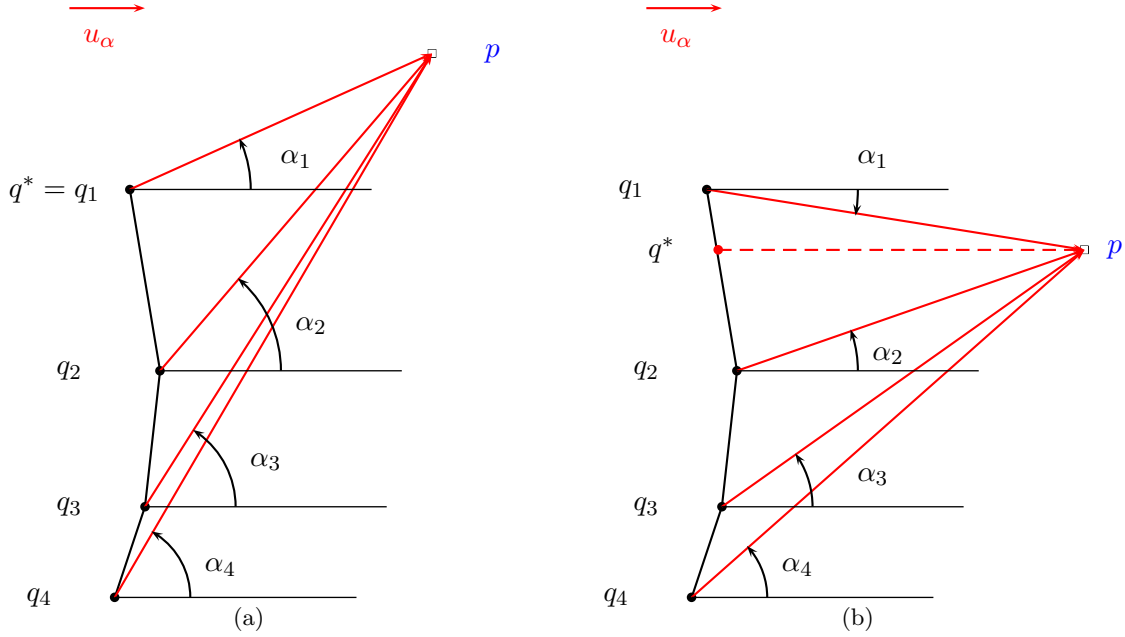


FIGURE B.2: Calcul de la dilatation d'un objet représenté par un polygone avec un élément structurant de direction. La situation (a) présente le cas où il n'existe pas de point q tel que $\vec{q}\vec{p}$ est colinéaire à u_α ; on a alors q^* qui est confondu avec l'un des points q_i . La situation (b) illustre le cas où q^* est un point intermédiaire sur un segment $q_i q_{i+1}$.

est atteinte au point q_1 . Le calcul d'appartenance de p au modèle spatial directionnel ne dépend que de la déviation angulaire α_1 .

Dans la situation (b), on montre que l'interpolation de segments entre les points de l'objet de référence est utile pour calculer rapidement l'angle de déviation minimal pour le point p . En effet, dans ce cas, on peut détecter facilement qu'il existe sur le segment $[q_1 q_2]$ un point q^* tel que $\vec{q^*}\vec{p}$ est colinéaire à \vec{u}_α . Ce cas de figure se produit lorsque les conditions suivantes sont réunies :

$$\begin{cases} \vec{q_1}\vec{p} \cdot \vec{u}_\alpha > 0 \\ \vec{q_2}\vec{p} \cdot \vec{u}_\alpha > 0 \\ \vec{q_1}\vec{p} \cdot \vec{u}_{\alpha'} < 0 \text{ et } \vec{q_2}\vec{p} \cdot \vec{u}_{\alpha'} > 0 \text{ avec } \alpha' = \alpha + \frac{\pi}{2} \\ \text{ou } \vec{q_1}\vec{p} \cdot \vec{u}_{\alpha'} > 0 \text{ et } \vec{q_2}\vec{p} \cdot \vec{u}_{\alpha'} < 0 \end{cases} \quad (\text{B.3})$$

Le calcul exact de q^* n'est pas nécessaire, puisque son existence sur $[q_1 q_2]$ suffit à garantir que la déviation minimale pour p est nulle (donc p est parfaitement en direction α par rapport à la référence) et que l'appartenance de p au modèle spatial directionnel est donc de 1. Dans toutes les situations, la complexité engendrée par le calcul est linéaire avec le nombre de points dans R .

B.2. Cas de données vectorisées (représentation en-ligne)

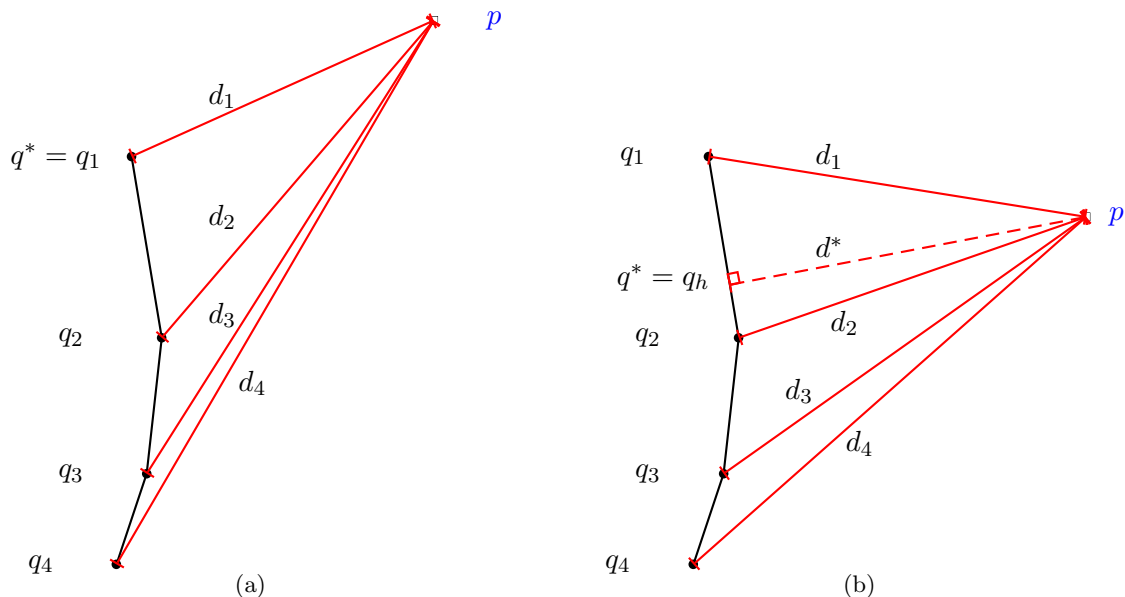


FIGURE B.3: Calcul de la dilatation d'un objet représenté par un polygone avec un élément structurant de distance. La situation (a) présente le cas où q^* est confondu avec l'un des points q_i . La situation (b) illustre le cas où q^* est un point intermédiaire sur un segment q_iq_{i+1} .

B.2.2 Cas d'un élément structurant de distance

Pour le cas du calcul d'une dilatation avec un élément structurant de distance (*être proche de*), il faut chercher le point q^* de la référence qui est le plus proche de p . Ici aussi, deux situations sont présentées dans la figure B.3. Dans la première, le point de la référence qui est le plus proche de p coïncide avec un des points q_i de l'échantillonnage (il s'agit de q_1) et par conséquent la distance minimale vaut d_1 (image (a)). Le calcul de l'appartenance de p au modèle spatial *être proche de* R ne dépend que de d_1 . Le second cas montre que, comme pour le calcul avec un élément structurant de direction, le segment entre deux points consécutifs $[q_1q_2]$ doit être pris en compte dans la recherche de q^* . Pour chaque portion de l'objet de référence, délimitée par les points q_i et q_{i+1} , il faut calculer le point projeté de p sur la droite (q_iq_{i+1}) . Si ce point projeté appartient au segment $[q_iq_{i+1}]$ (et donc à la référence), alors il constitue un candidat pour q^* . Dans l'exemple de la figure (b), le point q_h , projeté de p sur la droite q_1q_2 appartient bien au segment et constitue le point de la référence le plus proche de p ($q^* = q_h$).

Synthèse La représentation des tracés sous la forme d'ensemble de segments peut être exploitée efficacement pour le calcul de dilatation morphologique. Tous les points d'un même segment de la référence sont considérés d'une seule traite et la complexité du calcul de la dilatation en un point p est en $\mathcal{O}(t_R)$, avec t_R le nombre de segments dans le polygone qui représente la référence. En général, le nombre de points t_R est très inférieur au nombre de

pixels N_R dans un signal hors-ligne.

Des optimisations peuvent en pratique réduire le temps moyen de calcul : il est par exemple bénéfique de fusionner les suites de segments colinéaires en un seul segment pour réduire le nombre de points q_i . On peut aussi chercher à diminuer le nombre de segments dans la représentation en réalisant une approximation plus grossière des tracés (par un rééchantillonnage spatial, par exemple). Ici aussi, on peut avoir intérêt à tabuler les angles car l'espace d'acquisition est en réalité souvent lui-même discret.

Bibliographie

- [AHF⁺07] J. Atif, C. Hudelot, G. Fouquier, I. Bloch, and E. Angelini. From generic knowledge to specific reasoning for medical image interpretation using graph-based representations. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 224–229, 2007.
- [AMVG10] A. Awal, H. Mouchere, and C. Viard-Gaudin. Improving online handwritten mathematical expressions recognition with contextual modeling. In *Proceedings of the 12th International Conference on Frontiers in Handwriting Recognition*, pages 427–432, 2010.
- [AUAS09] W. Aly, S. Uchida, Fujiyoshi A., and M. Suzuki. Statistical classification of spatial relationships among mathematical symbols. In *Proceedings of the 10th International Conference on Document Analysis and Recognition*, pages 1350–1354, 2009.
- [Awa10] A. Awal. *Reconnaissance de structures bidimensionnelles : Application aux expressions mathématiques manuscrites en-ligne*. PhD thesis, Université de Nantes, 2010.
- [BCCJ06] I. Bloch, O. Colliot, and R.M. Cesar Jr. On the ternary spatial relation between. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 36(2) :312–327, 2006.
- [Blo99] I. Bloch. Fuzzy relative position between objects in image processing : a morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7) :657–664, 1999.
- [Blo06] I. Bloch. Fuzzy Spatial Relationships for Model-Based Pattern Recognition in Images and Spatial Reasoning Under Imprecision. *Lecture Notes In Computer Science*, 2955 :164, 2006.
- [Blo09] I. Bloch. Bipolar Fuzzy Mathematical Morphology for Spatial Reasoning. In *Proceedings of the 9th International Symposium on Mathematical Morphology and Its Application to Signal and Image Processing*, page 34. Springer, 2009.
- [Blo11] I. Bloch. Lattices of fuzzy sets and bipolar fuzzy sets, and mathematical morphology. *Information Sciences*, 181 :2002–2015, 2011.
- [BMA06] F. Bouteruche, S. Macé, and E. Anquetil. Fuzzy relative positioning for on-line handwritten stroke analysis. In *Proceedings of the 10th International Workshop on Frontiers in Handwriting Recognition*, pages 391–396, 2006.

- [BMK04] R. Bondugula, P. Matsakis, and J.M. Keller. Force histograms and neural networks for human-based spatial relationship generalization. In *Proceedings of Int. Conf. on Neural Networks and Computational Intelligence*, 2004.
- [BMM03] B. Bouchon-Meunier and C. Marsala, editors. *Logique Floue, Principes, Aide à la Décision*. Hermès-Lavoisier, 2003.
- [BR03] I. Bloch and A. Ralescu. Directional relative position between objects in image processing : a comparison between fuzzy approaches. *Pattern Recognition*, 36(7) :1563–1582, 2003.
- [BS03] I. Bloch and A. Saffloti. *Intelligent Systems for Information Processing : From Representation to Applications*, chapter On the Representation of Fuzzy Spatial Relations in Robot Maps. JAI Press, 2003.
- [CA07] R.C. Cinbis and S. Aksoy. Relative position-based spatial relationships using mathematical morphology. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 97–100. IEEE Computer Society, 2007.
- [CA09] R.C. Cinbis and S. Aksoy. Image mining using directional spatial constraints. *IEEE Geoscience and Remote Sensing Letters*, Special Issue on Image Information Mining : Pursuing Automation of Geospatial Intelligence for Environment and Security, 2009.
- [CC92] K.P. Chan and Y.S. Cheung. Fuzzy-attribute graph with application to chinese character recognition. *Systems, Man and Cybernetics, IEEE Transactions on*, 22(1) :153–160, 1992.
- [CCB06] O. Colliot, O. Camara, and I. Bloch. Integration of fuzzy spatial relations in deformable models—application to brain mri segmentation. *Pattern Recognition*, 39(8) :1401–1414, 2006.
- [Cha07] S.H. Cha. Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(2) :1, 2007.
- [CK04] S.J. Cho and J.H. Kim. Bayesian network modeling of strokes and their relationships for on-line handwriting recognition. *Pattern Recognition*, 37(2) :253–264, 2004.
- [CNM06] S. Coros, J.B. Ni, and P. Matsakis. Object localization based on directional information case of 2d vector data. In *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pages 163–170. ACM, 2006.
- [CPM73] H. H. Clark, Carpenter P.A., and Just M.A. *Visual information processing*, chapter On the meeting of semantics and perception, pages 311–381. New York Academic Press, 1973.
- [CVD04] L.A. Carlson and S.R. Van Deman. The space in spatial language. *Journal of Memory and Language*, 51(3) :418–436, 2004.

Bibliographie

- [Deh02] R. Dehak. *Inférence Quantitative des Relations Spatiales Directionnelles*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 2002.
- [DLX07] R. Dai, C.-L. Liu, and B. Xiao. Chinese character recognition : history, status and prospects. *Frontiers of Computer Science in China*, 1(2) :126–136, 2007.
- [DP95] D. Dubois and H. Prade. Tolerant fuzzy pattern matching : An introduction. *Fuzziness in Database Management Systems*, pages 42–58, 1995.
- [DP02] D. Dubois and H. Prade. Bipolarity in flexible querying. *Lecture Notes On Artificial Intelligence*, pages 174–182, 2002.
- [DTZB10] G. De Tré, S. Zadrozny, and A.J. Bronselaer. Handling bipolarity in elementary queries to possibilistic databases. *Fuzzy Systems, IEEE Transactions on*, 18(3) :599–612, 2010.
- [EH91] M.J. Egenhofer and J. Herring. Categorizing binary topological relations between regions, lines, and points in geographic databases. Technical report, University of Maine, 1991.
- [FGK07] JA Fitzgerald, F. Geiselbrechtner, and T. Kechadi. Mathpad : A Fuzzy Logic-Based Recognition System for Handwritten Mathematics. In *Proceedings of the 9th International Conference on Document Analysis and Recognition*, volume 2, 2007.
- [Fre75] J. Freeman. The modelling of spatial relations. *Computer Graphics and Image Processing*, 4(2) :156–171, 1975.
- [Gad97] P.D. Gader. Fuzzy spatial relations based on fuzzy morphology. In *Proc. Sixth IEEE International Conference on Fuzzy Systems*, volume 2, pages 1179–1183 vol.2, 1997.
- [Gap95a] K.P. Gapp. Basic meanings of spatial relations : Computation and evaluation in 3d space. In *Proceedings of the National Conference on Artificial Intelligence*, pages 1393–1393. JOHN WILEY & SONS LTD, 1995.
- [Gap95b] K.P. Gapp. An empirically validated model for computing spatial relations. *KI-95 : Advances in Artificial Intelligence*, pages 245–256, 1995.
- [GCBG09] E. Grosicki, M. Carre, J.M. Brodin, and E. Geoffrois. Results of the rimes evaluation campaign for handwritten mail processing. In *Proceedings of the 10th International Conference on Document Analysis and Recognition*, pages 941–945. IEEE, 2009.
- [GEA09] E. Grosicki and H. El Abed. Icdar 2009 handwriting recognition competition. In *Proceedings of the 10th International Conference on Document Analysis and Recognition*, pages 1398–1402. IEEE, july 2009.
- [GK08] R. Genoe and T. Kechadi. On the recognition of online handwritten mathematics using feature-based fuzzy rules and relationship precedence. In *Proceedings of IEEE International Conference on Fuzzy Systems*, pages 1641–1646. IEEE, 2008.

- [GTC10] L. Guichard, A.H. Toselli, and B. Coüasnon. Handwritten word verification by svm-based hypotheses re-scoring and multiple thresholds rejection. In *2010 12th International Conference on Frontiers in Handwriting Recognition*, pages 57–62. IEEE, 2010.
- [HAB08] C. Hudelot, J. Atif, and I. Bloch. Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, 159 :1929–1951, 2008.
- [HT95] W.G. Hayward and M.J. Tarr. Spatial language and spatial representation. *Cognition*, 55(1) :39–84, 1995.
- [JPPS10] R. Jeansoulin, O. Papini, H. Prade, and S. Schockaert, editors. *Methods for Handling Imperfect Spatial Information*, volume 256 of *Studies in Fuzziness and Soft Computing*. Springer Verlag, 2010.
- [KC05] J. Kelleher and F. Costello. Cognitive representations of projective prepositions. In *Proceedings of the Second ACL-Sigsem Workshop of The Linguistic Dimensions of Prepositions and their Use in Computational Linguistic Formalisms and Applications*, 2005.
- [KGK02] L.T. Kóczy, T.D. Gedeon, and J.A. Kóczy. Fuzzy tolerance relations and relational maps applied to information retrieval. *Fuzzy sets and systems*, 126(1) :49–61, 2002.
- [KK03] I.J. Kim and J.H. Kim. Statistical Character Structure Modeling and Its Application to Handwritten Chinese Character Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1422–1436, 2003.
- [KK04] K.W. Kang and J.H. Kim. Utilization of hierarchical, stochastic relationship modeling for hangul character recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9) :1185–1196, 2004.
- [KK05] T. Kojima and T. Kusumi. How do we judge the acceptability of spatial terms? In *Proceedings of the 2005 International Conference on Cyberworlds*, pages 406–412. IEEE Computer Society, 2005.
- [KK10] D. H. Kim and J.H. Kim. Top-down search with bottom-up evidence for recognizing handwritten mathematical expressions. In *Proceedings of the 12th International Conference on Frontiers in Handwriting Recognition*, pages 507–512, 2010.
- [LBKS07] W.S. Lee, L. Burak Kara, and T.F. Stahovich. An efficient graph-based recognizer for hand-drawn symbols. *Computers & Graphics*, 31(4) :554–567, 2007.
- [LCC96] J. Liu, WK Cham, and M.M.Y. Chang. Stroke order and stroke number free on-line chinese character recognition using attributed relational graph matching. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume 3, pages 259–263. IEEE, 1996.
- [LJN04] C.L. Liu, S. Jaeger, and M. Nakagawa. Online recognition of chinese characters : the state-of-the-art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2) :198–213, 2004.

Bibliographie

- [LKK01] C.L. Liu, I.J. Kim, and J.H. Kim. Model-based stroke extraction and matching for handwritten Chinese character recognition. *Pattern Recognition*, 34(12) :2339–2352, 2001.
- [Log94] G.D. Logan. Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology : Human Perception and Performance*, 20(5) :1015, 1994.
- [LS99] G.D. Logan and D.D. Sadler. *Language and Space*, chapter A Computational Analysis of the Apprehension of Spatial Relations. The MIT Press, 1999.
- [LSQW09] T. Li, B. Shu, X. Qiu, and Z. Wang. A complete descriptor of line-segment-pair for symbol recognition. In *Proceedings of the International Conference on Computer Graphics*, pages 89–95. ACM, 2009.
- [MA04] S. Marukatat and T. Artières. Handling spatial information in on-line handwriting recognition. In *Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition*, pages 14–19, 2004.
- [MA09] S. Macé and E. Anquetil. Eager interpretation of on-line hand-drawn structured documents : The DALI methodology. *Pattern Recognition*, 42 :3202–3214, 2009.
- [Mar04] S. Marukatat. *Une approche générique pour la reconnaissance de signaux écrits en ligne*. PhD thesis, Université Pierre et Marice Curie (Paris 6), 2004.
- [Mat02] P. Matsakis. Understanding the spatial organization of image regions by means of force histograms : a guided tour. In P. Matsakis and L. M. Sztandera, editors, *Applying Soft Computing in Defining Spatial Relations*, volume 106 of *Studies in Fuzziness and Soft Computing*, pages 99–122. Physica-Verlag, 2002.
- [MDT09] T. Matthé and G. De Tré. Bipolar query satisfaction using satisfaction and dissatisfaction degrees. In *Proceedings of the 2009 ACM symposium on Applied Computing*, pages 1699–1703. ACM, 2009.
- [ML08] L.L. Ma, , and C.L. Liu. A new radical-based approach to online handwritten chinese character recognition. In *Proceedings of the 19th International Conference on Pattern Recognition*, 2008.
- [ML09] L.L. Ma and C.-L. Liu. On-line handwritten chinese character recognition based on nested segmentation of radicals. In *Proceedings of the Chinese Conference on Pattern Recognition*, 2009.
- [MLSL06] J. Mas, B. Lamiroy, G. Sánchez, and J. Lladós. Automatic learning of symbol descriptions avoiding topological ambiguities. In *Proceedings of the 3rd Eurographics Workshop on Sketch-Based Interfaces and Modeling*, 2006.
- [MNW06] P. Matsakis, J.B. Ni, and X. Wang. Object localization based on directional information : Case of 2D raster data. In *Proceedings of the 18th International Conference on Pattern Recognition*, volume 2, pages 142–146. IEEE, 2006.

- [MR94] K. Miyajima and A. Ralescu. Spatial organization in 2d images. In *Proceedings of the Third IEEE Conference on Fuzzy Systems*, pages 100–105, 1994.
- [MS02] P. Matsakis and L.M. Sztandera, editors. *Applying soft computing in defining spatial relations*, volume 106 of *Studies in Fuzziness and Soft Computing*. Springer, 2002.
- [MSL05] J. Mas, G. Sanchez, and J. Lladós. An adjacency grammar to recognize symbols and gestures in a digital pen framework. In Jorge Marques, Nicolas Perez de la Blanca, and Pedro Pina, editors, *Pattern Recognition and Image Analysis*, volume 3523 of *Lecture Notes in Computer Science*, pages 25–72. Springer Berlin / Heidelberg, 2005.
- [MSLL07] J. Mas, G. Sanchez, J. Lladós, and B. Lamiroy. An incremental on-line parsing algorithm for recognizing sketching diagrams. In *Proceedings of the 9th International Conference on Document Analysis and Recognition*, volume 1, pages 452–456, 2007.
- [MVG⁺11] H. Mouchère, C. Viard-Gaudin, U. Garain, D.H. Kim, and J.H. Kim. CROHME2011 : Competition on Recognition of Online Handwritten Mathematical Expressions. In *Proceedings of the 11th International Conference on Document Analysis and Recognition*, pages 1497–1500, 2011.
- [MW99] P. Matsakis and L. Wendling. A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7) :634–643, 1999.
- [MWN10] P. Matsakis, L. Wendling, and J. Ni. A general approach to the fuzzy modeling of spatial relationships. In Robert Jeansoulin, Odile Papini, Henri Prade, and Steven Schockaert, editors, *Methods for Handling Imperfect Spatial Information*, volume 256 of *Studies in Fuzziness and Soft Computing*, pages 49–74. Springer Berlin / Heidelberg, 2010.
- [NWV08] R. Niels, D. Willems, and L. Vuurpijl. The nicicon database of handwritten icons. In *Proceedings of the 11th International Conference on the Frontiers of Handwriting Recognition*, pages 296–301, August 2008.
- [OYSS07] I. Ota, R. Yamamoto, S. Sako, and S. Sagayama. On-line handwritten kanji recognition based on inter-stroke grammar. In *Proceedings of the 9th International Conference on Document Analysis and Recognition*, volume 2, pages 1188–1192, 2007.
- [PR07] J. Pu and K. Ramani. Implicit geometric constraint detection in freehand sketches using relative shape histogram. In *Proceedings of the 4th Eurographics workshop on Sketch-based interfaces and modeling*, SBIM '07, pages 107–113, New York, NY, USA, 2007. ACM.
- [PV93] J.-C. Perez and E. Vidal. Optimum polygonal approximation of digitized curves. *Pattern Recognition Letters*, 15 :743–750, 1993.

Bibliographie

- [QBA09] S. Quiniou, F. Bouteruche, and E. Anquetil. Word extraction associated with a confidence index for on-line handwritten sentence recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(5) :945–966, 2009.
- [Ren02] J. Renz. *Qualitative spatial reasoning with topological information*, volume 2293. Springer Verlag, 2002.
- [RJ73] A. Rosenfeld and E. Johnston. Angle detection on digital curves. *Computers, IEEE Transactions on*, 100(9) :875–878, 1973.
- [SA09] R. Sicard and T. Artières. Modelling sequences using pairwise relational features. *Pattern Recognition*, 42(9) :1922 – 1931, 2009.
- [SAP06] R. Sicard, T. Artières, and E. Petit. Modeling on-line handwriting using pairwise relational features. In *Proceedings of the 10th International Workshop on Frontiers in Handwriting Recognition*, pages 267–274, 2006.
- [Say73] K.M. Sayre. Machine recognition of handwritten words : A project report. *Pattern Recognition*, 5(3) :213–228, 1973.
- [Ser83] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA, 1983.
- [Shi02] J. Shin. Optimal stroke-correspondence search method for on-line character recognition. *Pattern recognition letters*, 23(5) :601–608, 2002.
- [SK09] I.J. Sledge and J.M. Keller. Mapping natural language to imagery : Placing objects intelligently. In *Proceedings of the 2009 IEEE International Conference on Fuzzy Systems*, pages 518–523. IEEE, 2009.
- [SLW11] K.C. Santosh, B. Lamiroy, and L. Wendling. Symbol Recognition using Spatial Relations. *Pattern Recognition Letters*, 2011.
- [SWL10a] K. C. Santosh, L. Wendling, and B. Lamiroy. Unified pairwise spatial relations : An application to graphical symbol retrieval. In Jan Van Leeuwen, editor, *Graphics Recognition. Achievements, Challenges, and Evolution*. Springer, 2010.
- [SWL10b] K. C. Santosh, L. Wendling, and B. Lamiroy. Using spatial relations for graphical symbol description. In *Proceedings of the 20th International Conference on Pattern Recognition*, Istanbul Turquie, 2010.
- [Wen03] L. Wenyin. On-line graphics recognition : State-of-the-art. In *Proceedings of the 5th IAPR International Workshop on Graphics Recognition, 2003*. Cite-seer, 2003.
- [WF01] A.-B. Wang and K.-C. Fan. Optical recognition of handwritten Chinese characters by hierarchical radical matching method. *Pattern Recognition*, 34(1) :15–35, 2001.
- [WK99] X. Wang and J.M. Keller. Human-based spatial relationship generalization through neural/fuzzy approaches. *Fuzzy Sets and Systems*, 101(1) :5–20, 1999.

- [WKG97] X. Wang, J.M. Keller, and P. Gader. Using spatial relationships as features in object recognition. In *Proc. Annual Meeting of the North American Fuzzy Information Processing Society*, pages 160–165, 1997.
- [WM03] Y. Wang and F. Makedon. R-histogram : quantitative representation of spatial relations for similarity-based image retrieval. In *Proceedings of the 11th ACM international conference on Multimedia*, pages 323–326. ACM, 2003.
- [WNM06] X. Wang, J.B. Ni, and P. Matsakis. Fuzzy object localization based on directional (and distance) information. In *Proceedings of the 2006 IEEE International Conference on Fuzzy Systems*, pages 256–263. IEEE, 2006.
- [WNvGV09] D. Willems, R. Niels, M. van Gerven, and L. Vuurpijl. Iconic and multi-stroke gesture recognition. *Pattern Recognition*, 42(12) :3303–3312, 2009.
- [XG02] H. Xue and V. Govindaraju. Incorporating contextual character geometry in word recognition. In *Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition*, pages 123–127. IEEE, 2002.
- [Xu07] Z. Xu. Intuitionistic fuzzy aggregation operators. *Fuzzy Systems, IEEE Transactions on*, 15(6) :1179–1187, 2007.
- [XZB⁺04] X. Xiaogang, S. Zhengxing, P. Binbin, J. Xiangyu, and L. Wenyin. An online composite graphics recognition approach based on matching of spatial relation graphs. *International Journal on Document Analysis and Recognition*, 7(1) :44–55, 2004.
- [YWL10] F. Yin, Q.F. Wang, and C.L. Liu. Integrating geometric context for text alignment of handwritten chinese documents. In *12th International Conference on Frontiers in Handwriting Recognition*, pages 7–12. IEEE, 2010.
- [Zad65] L. Zadeh. Fuzzy Sets. *Information and Control*, 8 :338–353, 1965.
- [Zad78] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *International Journal on Fuzzy Sets and Systems*, 1(1) :3–28, 1978.
- [ZBC02] R. Zanibbi, D. Blostein, and J.R. Cordy. Recognizing mathematical expressions using tree transformation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(11) :1455–1467, 2002.
- [ZBZ05] L. Zhang, D. Blostein, and R. Zanibbi. Using fuzzy logic to analyze superscript and subscript relations in handwritten mathematical expressions. In *Proceedings of the 8th International Conference on Document Analysis and Recognition*, pages 972–976, 2005.
- [ZDW97] J. Zheng, X. Ding, and Y. Wu. Recognizing on-line handwritten chinese character via farg matching. *Proceedings of the 4th International Conference on Document Analysis and Recognition*, pages 621–624, 1997.
- [ZL08] J. Zeng and Z.Q. Liu. Markov random field-based statistical character structure modeling for handwritten Chinese character recognition. *IEEE transactions on pattern analysis and machine intelligence*, 30(5) :767–780, 2008.

Bibliographie

- [ZYL⁺07] X. Zhou, J.-L. Yu, C.-L. Liu, T. Nagasaki, and K. Marukawa. On-line handwritten japanese character string recognition incorporating geometric context. In *Proceedings of the 9th International Conference on Document Analysis and Recognition*, pages 48–52, 2007.
- [ZZLN10] B. Zhu, X.D. Zhou, C.L. Liu, and M. Nakagawa. A robust model for on-line handwritten japanese text recognition. *International journal on document analysis and recognition*, 13(2) :121–131, 2010.

Références de l'auteur

Journal international

- [DA12] A. Delaye and E. Anquetil. Learning of fuzzy spatial relations between hand-written patterns, in *International Journal on Data Mining, Modelling and Management, Special issue on Pattern Recognition*, to appear, 2012.

Conférence internationale

- [DA11] A. Delaye and E. Anquetil. Fuzzy Relative Positioning Templates for Symbol Recognition, in *Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR'11)*, pages 1220–1224, 2011.
- [LDA11] Pei-Yu Li, A. Delaye and E. Anquetil. Evaluation of Continuous Marking Menus for Learning Cursive Pen-based Commands, in *Proceedings of the 15th conference of the International Graphonomics Society (IGS'11)*, pages 217–220, 2011.
- [DSA11] A. Delaye, R. Sekkal and E. Anquetil. Continuous Marking Menus for Learning Cursive Pen-based Gestures, in *Proceedings of the International Conference on Intelligent User Interfaces (IUI'11)*, pages 319–322, 2011.
- [DA10a] A. Delaye and E. Anquetil. Learning Spatial Relationships in hand-drawn Patterns using fuzzy Mathematical Morphology, in *Proceedings of the International Conference on Soft Computing for Pattern Recognition (SoCPar'10)*, pages 162–167, 2010.
- [MDL10] L.-L. Ma, A. Delaye and C.-L. Liu. Special Radical Detection by Statistical Classification for Online Handwritten Chinese Character Recognition, in *Proceedings of the International Conference on Frontiers in Handwriting Recognition (ICFHR'10)*, pages 501–506, 2010.

- [DAM09] A. Delaye, E. Anquetil and S. Macé. Explicit fuzzy modeling of shapes and positioning for handwritten Chinese character recognition, in *Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR'09)*, pages 1125–1129, 2009.
- [DMA09b] A. Delaye, S. Macé, and E. Anquetil. Modeling Relative Positioning of Handwritten Patterns, in *Proceedings of the 14th conference of the International Graphonomics Society, (IGS'09)*, 2009.
- [DMA08a] A. Delaye, S. Macé, and E. Anquetil. Hybrid statistical-structural recognition of Chinese characters with Fuzzy Inference System, in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR'08)*, 2008.

Conférence nationale

- [DA10b] A. Delaye et E. Anquetil. Modélisation du positionnement relatif de tracés manuscrits par morphologie mathématique floue, dans *Actes du 17ème Congrès Francophone de Reconnaissance des Formes et d'Intelligence Artificielle (RFIA'10)*, 2010.
- [DA10c] A. Delaye et E. Anquetil. Apprentissage de relations spatiales floues pour la reconnaissance de tracés manuscrits dans *Actes du 5ème l'atelier Représentation et Raisonnement sur le Temps et l'Espace (RTE'10)*, pages 9–12, 2010.
- [DM10] A. Delaye et H. Mouchère. Vers une approche générique pour la reconnaissance de formes manuscrites structurées : application aux équations mathématiques et aux caractères chinois, dans *Actes du 11ème Colloque International Francophone sur l'Écrit et le Document (CIFED'10)*, pages 297–312, 2010.
- [DMA08b] A. Delaye, S. Macé, et E. Anquetil. Reconnaissance hybride statistique-structurale de sinogrammes par Système d'Inférence Floue, dans *Actes du 10ème Colloque International Francophone sur l'Écrit et le Document (CIFED'08)*, pages 49–54, 2008. **Prix du meilleur article CIFED'08.**

L'essor des interfaces homme-machine permettant la saisie d'informations à l'aide d'un stylo électronique est accompagné par le développement de méthodes automatiques pour interpréter des données de plus en plus riches et complexes : texte manuscrit, mais aussi expressions mathématiques, schémas, prise de notes libre... Pour interpréter efficacement ces documents manuscrits, il est nécessaire de considérer conjointement les formes des objets qui les constituent et leur positionnement spatial. Nos recherches se concentrent sur la modélisation du positionnement spatial entre des objets manuscrits, en partant du constat qu'il n'est pas exploité dans toute sa richesse par les méthodes actuelles.

Nous introduisons le concept de méta-modèle spatial, une modélisation générique pour décrire des relations spatiales entre des objets de nature, complexité et formes variables. Ces modèles, qui peuvent être appris à partir de données, offrent une richesse et une précision inédite car ils autorisent la conduite d'un raisonnement spatial directement dans l'espace image. L'appui sur le cadre de la théorie des sous-ensembles flous et de la morphologie mathématique permet la gestion de l'imprécision et offre une description des relations spatiales conforme à l'intuition. Un méta-modèle est doté d'un pouvoir de prédiction qui permet de décrire la relation spatiale modélisée au sein de l'image, par rapport à un objet de référence. Cette capacité rend possible la visualisation des modèles et fournit un outil pour segmenter les tracés en fonction de leur contexte. En exploitant ces modèles, nous proposons une représentation pour des objets manuscrits à la structure complexe. Cette représentation repose uniquement sur la modélisation de leurs informations spatiales, afin de démontrer l'importance de ces informations pour l'interprétation d'objets manuscrits structurés. La segmentation des tracés en primitives structurelles est guidée par les modèles de positionnement, via leur capacité de prédiction.

Les résultats expérimentaux, portant sur des objets de complexité et de natures diverses (caractères chinois, gestes d'édition, symboles mathématiques, lettres), confirment la bonne qualité de description du positionnement offerte par les méta-modèles. Les tests de reconnaissance de symboles par l'exploitation de leur information spatiale attestent d'une part de l'importance de cette information et valident d'autre part la capacité des méta-modèles à la représenter avec une grande précision. Ces résultats témoignent donc de la richesse de l'information spatiale et du potentiel des méta-modèles spatiaux pour l'amélioration des techniques de traitement du document manuscrit.

The rise of pen-enabled interfaces is supported by the development of automatic methods for interpretation of more and more rich and complex input data: handwritten text, mathematical equations, sketches, free handwritten notes... For efficiently recognizing these handwritten documents, one has to consider jointly the shapes of their components and the relative positioning between them. Our research focuses on the modeling of relative positioning between handwritten objects, considering that the potential of this part of the information is not fully exploited in the current methods.

We introduce spatial meta-templates, a generic modeling for describing spatial relations between objects of diverse nature, complexity, and shape. These models can be trained from data and provide richer and more accurate descriptions because they authorize to reason about spatial information directly in the image space. Relying on fuzzy sets theory and mathematical morphology allows dealing with imprecision and offers intuitive description of spatial relations. A meta-template is endowed with a prediction capacity: it provides the description of modeled spatial relations with respect to a reference object in the image, as a spatial template. This enables to conduct segmentation of objects depending on their spatial context. By exploiting these models, we present a new representation for structured handwritten objects. It relies only on modeling of the spatial information so as to demonstrate the importance of spatial information for interpretation of these objects. The segmentation of handwritten strokes into structural primitives is driven by positioning models, making use of their prediction ability.

Experimental results, obtained with objects of diverse nature and complexity (Chinese characters, editing gestures, mathematical symbols, letters), validate the quality of positioning description offered by our models. The performance on the task of recognizing symbols with a spatial-based representation further attests the importance of this information and confirms the ability of meta-templates to model it properly and accurately. These results both show the richness of spatial information and give an insight on the potential of meta-templates for improving methods for handwritten document interpretation.