



HAL
open science

Extraction d'éléments curvilignes guidée par des mécanismes attentionnels pour des images de télédétection: approche par fusion de données

Gilles Cotteret

► **To cite this version:**

Gilles Cotteret. Extraction d'éléments curvilignes guidée par des mécanismes attentionnels pour des images de télédétection: approche par fusion de données. Autre [cs.OH]. Université Paris Sud - Paris XI, 2005. Français. NNT: . tel-00163679

HAL Id: tel-00163679

<https://theses.hal.science/tel-00163679>

Submitted on 19 Jul 2007

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.

GILLES COTTERET

**EXTRACTION D'ÉLÉMENTS CURVILIGNES
GUIDÉE PAR DES MÉCANISMES ATTENTIONNELS
POUR DES IMAGES DE TÉLÉDÉTECTION**
Approche par fusion de données

Thèse de doctorat présentée en cotutelle
à la Faculté des études supérieures de l'Université Laval, Québec
dans le cadre du programme de doctorat en Sciences Géomatiques
pour l'obtention du grade de Philosophiæ Doctor (Ph.D.)

Défendue le 9 décembre 2005 à Orsay
devant le jury composé de :

Dr. Geoffrey Edwards, Université Laval, Canada, Professeur, *directeur au Canada*
Dr. Frank Ferrie, McGill, University, Canada, Professeur, *rapporteur*
Dr. Robert Jeansoulin, LSIS/CNRS, Marseille, Directeur de Recherche, *rapporteur*
Dr. Gérard Ligozat, Université Paris-Sud XI, Professeur, *directeur en France*
Dr. Bernard Moulin, Université Laval, Canada, Professeur, *examineur*
Dr. Philippe Tarroux, École Normale Supérieure, Paris, Professeur, *examineur*

FACULTÉ DE FORESTERIE ET DE GÉOMATIQUE
UNIVERSITÉ LAVAL
QUÉBEC
et

UNIVERSITÉ DE PARIS-SUD XI
LABORATOIRE D'INFORMATIQUE POUR LA MÉCANIQUE ET LES
SCIENCES DE L'INGÉNIEUR, ORSAY, FRANCE
pour l'obtention du grade de Docteur de l'Université de Paris-Sud XI
(spécialité informatique)

Décembre 2005

GILLES COTTERET

**EXTRACTION D'ÉLÉMENTS CURVILIGNES
GUIDÉE PAR DES MÉCANISMES
ATTENTIONNELS POUR DES IMAGES DE
TÉLÉDÉTECTION**

Approche par fusion de données

Thèse de doctorat présentée en cotutelle
à la Faculté des études supérieures de l'Université Laval, Québec
dans le cadre du programme de doctorat en Sciences Géomatiques
pour l'obtention du grade de Philosophiæ Doctor (Ph.D.)

FACULTÉ DE FORESTERIE ET DE GÉOMATIQUE
UNIVERSITÉ LAVAL
QUÉBEC

et

UNIVERSITÉ DE PARIS-SUD XI
LABORATOIRE D'INFORMATIQUE POUR LA MÉCANIQUE ET LES
SCIENCES DE L'INGÉNIEUR, ORSAY, FRANCE
pour l'obtention du grade de Docteur de l'Université de Paris-Sud XI
(spécialité informatique)

Décembre 2005

Résumé

L'extraction d'éléments curvilignes d'images de télédétection, surtout proches de la limite de résolution ou lorsqu'elles sont bruitées, représente toujours un défi important pour les algorithmes informatiques, alors que pour les interprètes humains c'est une tâche immédiate. Dans ce travail une méthode novatrice est présentée pour guider l'extraction d'éléments curvilignes (routes, cours d'eau, *etc*) d'images de télédétection. Cette méthode a pour but de mettre à jour les systèmes d'informations géographiques (SIG) grâce à un modèle inédit (ELECA) qui comporte trois parties. Le modèle ELECA tire avantage de deux idées principales. Les données utilisées ne sont pas seulement les images de télédétection mais également les données des SIG qui ne sont pas nécessairement à jour. Le modèle s'appuie également sur des progrès récents des sciences psychocognitives en imitant partiellement le mouvement des yeux durant une recherche visuelle.

Ainsi les trois parties du modèle ELECA sont (1) une méthode de recherche de l'information locale qui utilise un œil virtuel pour ne se focaliser qu'aux lieux d'information probable, évitant ainsi un traitement global de l'image ; (2) une méthode simple et rapide d'extraction de l'information locale par une adaptation astucieuse d'une analyse en composantes connexes ; et (3) une méthode originale de fusion des informations locales qui s'appuie sur des travaux de reconstruction de l'espace global par raisonnement spatial qualitatif.

Le modèle ELECA évite de cette manière certains pièges dans lesquels tombent souvent les méthodes courantes/classiques. En particulier, la technique proposée peut être mise en œuvre pour des images partiellement masquées ou de basse résolution pour lesquelles seulement des interprètes humains peuvent traiter l'image présentement. La méthode a été aussi élaborée pour être rapide et efficace afin de pouvoir réaliser des mises à jour de SIG très rapides lorsque nécessaire.

La dernière partie de ce travail est consacrée à l'élaboration d'une solution informatique supportant le modèle ELECA. Une architecture logicielle souple et adaptative est

mise en avant, permettant l'intégration de développements ultérieurs pour le modèle. Enfin un volet algorithmique montre concrètement que les solutions du modèle ELECA sont implantables en termes informatiques.

Abstract

The extraction of curvilinear elements from remote sensing images, especially when noisy or near the limit of resolution constitutes a significant challenge for data-processing algorithms. In this work a method is presented for linear feature extraction in remote sensing (RS) images. An original model (ELECA) is introduced allowing out of date geographical information system (GIS) data to be updated through the use of a visual search method that mimics human eye movements.

The ELECA model is composed of three parts : (1) a visual search module using virtual gaze to avoid processing the entire image; (2) a simple and fast method for local information extraction by a clever adaptation of connected-component labeling; and (3) an original method for the fusion of local information to construct a global representation at the scale of the image based on qualitative spatial reasoning techniques.

The ELECA model avoids several problems characteristic of current methods. In particular, the proposed technique can be applied to low resolution or partially occluded images for which currently only human interpreters can successfully process the image. The technique is also designed to be very fast and efficient when a quick GIS update is needed.

The last part of this project is devoted to the design of software which supports the ELECA model. The proposed software architecture is adaptive and allows the integration of future model developments. Finally it is shown how the ELECA model could be implemented.

Resumen

La extracción de elementos curvilíneos de las imágenes de teledetección, sobre todo en proximidad del límite de resolución o en presencia de ruido, representa siempre un reto importante para los algoritmos informáticos, mientras que para los intérpretes humanos es una tarea inmediata. Este trabajo presenta un método nuevo para guiar la extracción de elementos curvilíneos (carreteras, ríos, etc.) de imágenes de teledetección. El objetivo del método consiste en actualizar un sistema de información geográfica (SIG) gracias a un modelo inédito (ELECA) constituido de tres partes. El modelo ELECA desarrolla dos ideas principales. El modelo puede utilizar las imágenes de teledetección y los datos del SIG que están desactualizados. El modelo se basa asimismo en resultados recientes de la investigación en ciencias psico-cognitivas, dado que imita parcialmente el movimiento de los ojos durante una búsqueda visual.

Las tres partes del modelo ELECA son : (1) un método de búsqueda de la información local que utiliza un ojo virtual par focalizarse únicamente donde se encuentra la información probable, evitando un tratamiento global de la imagen ; (2) un método simple y rápido de extracción de la información local, que utiliza una adaptación astuciosa de un análisis en componentes conexos ; y (3) un método original de fusión de informaciones locales que se apoya sobre trabajos de reconstrucción del espacio global por razonamiento espacial cualitativo.

El modelo ELECA evita así algunas trampas en las que caen a menudo los métodos clásicos. En particular, la técnica propuesta puede ser utilizada para imágenes parcialmente enmascaradas o de baja resolución, que sólo pueden ser tratadas actualmente por intérpretes humanos. El método ha sido elaborado para ser eficiente en la actualización rápida de un SIG si es necesario.

La última parte de este trabajo presenta una solución informática elaborada para soportar el modelo ELECA. Una arquitectura de programa flexible y adaptable es presentada, permitiendo la integración de desarrollos ulteriores del modelo. Finalmente,

un aspecto algorítmico muestra concretamente que las soluciones del modelo ELECA¹. se pueden implantar en términos informáticos.

¹ELECA : Elementos Lineares Extraídos por Cálculo Atencional

Avant-propos

Écrire l'avant-propos c'est comme regarder en arrière pour la première fois après une longue période toute consacrée à l'élaboration du mémoire : cela donne un peu le vertige tous ces visages croisés durant cette période doctorale . . . mélancolie quand tu nous tiens !

Cette aventure n'aurait pas eu lieu sans que mon directeur, Geoffrey Edwards, m'ait fait confiance sur la foi d'une page web où il était écrit en clignotant que je recherchais une thèse de doctorat ! Merci Geoffrey ! J'ai apprécié énormément que vous me donniez la chance d'explorer des domaines si nouveaux pour moi, curieux comme je le suis, j'ai été gâté.

Je voudrais remercier particulièrement Bernard Moulin pour l'encadrement si rigoureux qu'il a su me donner sans jamais que je ne le sente comme une contrainte. Bernard est tout entier dévoué à ses étudiants et il n'est pas rare de recevoir un courrier de sa part à 22h00 un samedi soir alors que cela fait moins d'une demie heure qu'on lui a envoyé le nôtre ! Merci Bernard de m'avoir "repêché" pendant un moment de doute.

Ce travail ayant été effectué en co-tutelle entre la France et le Québec, j'ai eu la chance d'avoir un second directeur en la personne de Gérard Ligozat. Gérard m'a accueilli au LIMSI lors de ma troisième année de thèse comme si j'y avais passé les deux années précédentes ; je me suis senti tout de suite à l'aise. Merci Gérard pour les discussions sur mon travail, votre recul sur ma thèse m'a aidé énormément à prendre de la hauteur alors que j'avais "le nez dans le guidon". Je tiens à remercier Gérard également pour avoir dû supporter les vicissitudes administratives annuelles d'inscriptions au doctorat alors que j'étais au Québec.

Je tiens à remercier tout spécialement Sylvie Daniel pour la lecture si précise (et si rapide !) qu'elle a bien voulu faire de la première version de mon document. Ses commentaires ont considérablement amélioré le document. Merci beaucoup Sylvie !

Quand on lit l'avant-propos des thèses, l'importance de la famille, et des parents en particulier, revient systématiquement, à croire que ces grands étudiants attardés que sont les doctorants n'ont jamais coupé le cordon ombilical! De mon côté, il est bien coupé, la plupart du temps 6000 km m'ont séparé de ma famille. Pourtant le soutien inconditionnel, constant et aimant de mes parents et de ma sœur Élisabeth a été et est toujours ma force et mon moteur quotidien. Il m'est impossible d'exprimer ma gratitude envers mes parents à la hauteur de tout ce qu'ils m'ont donné. J'ai tout simplement les meilleurs parents que l'on puisse avoir. Merci Papa. Merci Maman. Merci Bilos.

À la fin du doctorat, on se trouve souvent être l'un des "plus vieux" étudiants du laboratoire, néanmoins durant toute cette période, trois comparses ont toujours été là. Je voudrais remercier Gabriela, Mir et Hervé pour l'amitié sincère qu'ils me témoignent, pour avoir refait le monde les nuits de travail au laboratoire, pour la solidarité des étudiants au doctorat, pour avoir partagé ensemble nos joies et nos peines quotidiennes. Durant ma période au CRG, j'ai eu la chance de rencontrer beaucoup de personnes du monde entier avec lesquelles j'ai construit une amitié qui va bien au-delà des relations de travail. Je pense en particulier à mon "vieux pote" David Bachy et à Sylvie, David Duguay, Boussaad, Michel, Exequiel, PE...

J'ai eu l'occasion également de faire le même type de rencontre au LIMSI; je remercie particulièrement Nicolas, Haïfa et Guillaume pour leur si gentil accueil et leur amitié.

J'ai eu également la chance de croiser sur ma route au LIMSI Jamal Atif qui a su me mettre le pied à l'étrier en C++, et Youssef Machrouh qui m'a fait le plaisir de s'intéresser à mon travail quand je suis arrivé et qui m'a conseillé, merci à tous les deux. Je souhaite également remercier Josée Fitzback pour son aide en ingénierie informatique.

La période de rédaction est intense et je veux remercier Karine pour son réconfort durant ce moment. Je voudrais également remercier Louise pour ses commentaires sur le document et Aldo Napoli pour ses conseils de valorisation du travail. J'adresse un remerciement particulier à Véronique Mimeault qui m'a aidé durant les deux dernières années de ce travail.

Vivre dans 9m² pendant quatre ans peut paraître fou, pour moi cela a été une formidable expérience que celle d'habiter aux résidences universitaires de l'Université Laval, pavillon Parent. Le contexte international m'a permis des rencontres toutes tellement enrichissantes. Merci à tous! Merci à mes amis en France pour leur soutien sans faille.

Ce travail n'aurait pas été possible sans le soutien, sous la forme d'une bourse doctorale, du Réseau de Centres d'Excellence en Géomatique, GEOIDE que je remercie.

Je tiens à remercier chaleureusement Messieurs Frank Ferrie, Robert Jeansoulin et Philippe Tarroux de m'avoir fait l'honneur de bien vouloir évaluer ce travail.

Enfin, j'adresse un remerciement tout spécial à Carmen Couture, Annick Jaton et Sylvie Levesque qui ont su m'aider, tout au long de ma scolarité doctorale, à me dépêtrer de situations administratives à première vue inextricables, et ceci toujours avec sourire et gentillesse.

*À mes parents qui m'ont toujours soutenu,
quels que fussent mes choix.
À ceux qui n'ont pas [eu] cette chance dans la vie.*

*Le commencement de toutes les sciences,
c'est l'étonnement de ce que les choses sont ce qu'elles sont.
Aristote, Métaphysique, I, 2.*

Table des matières

Résumé	iii
Abstract	v
Resumen	vii
Avant-propos	ix
Table des matières	xv
Table des figures	xxi
Table des algorithmes	xxv
Liste des tableaux	xxvii
I Cadre général	1
1 Introduction	3
1.1 Contexte	4
1.1.1 Généralités	4
1.1.2 Spécificités	7
1.2 Problématique	8
1.2.1 Projets GEOIDE	8
1.2.2 Doctorat	10
1.3 Démarche de résolution	11
1.4 Objectifs du travail	14
1.5 Méthode de recherche	14
1.6 Articulation générale du document	20
2 État de l’art	23
2.1 Extraction d’éléments curvilignes d’images de télédétection	24
2.1.1 Mise en situation	25

2.1.2	Types d'images	26
2.1.3	Extraction manuelle <i>vs</i> extraction automatique	28
2.1.4	Techniques d'extraction en télédétection	30
2.1.5	Bilan	37
2.2	Approche de Marr	38
2.2.1	Ébauche primitive	40
2.2.2	Ébauche primitive brute	40
2.2.3	Ébauche primitive complète	43
2.2.4	Au-delà de Marr	48
2.2.5	Notions complémentaires : reconnaissance des formes	48
2.3	Mécanismes attentionnels	52
2.3.1	Généralités	52
2.3.2	Psycho-physiologie de l'attention visuelle	53
2.3.3	Modélisation computationnelle de la vision et de l'attention	62
2.4	Multirésolution, recherche et interrogation	74
2.4.1	Découpage simple	75
2.4.2	Systèmes de recherche d'information visuelle	76
2.4.3	Approches par ondelettes	78
2.4.4	Interrogation par objet	81
2.5	Raisonnement spatial qualitatif	82
2.5.1	Qu'est-ce que le raisonnement qualitatif?	83
2.5.2	Allen : du temporel au spatial	84
2.5.3	Orientation	86
2.5.4	Distance	88
2.5.5	Exemple de raisonnement spatial simple : <i>Cataglyphis</i>	89
II Modèle d'extraction intelligente de données		91
3 Principes et objets de l'extraction d'objets linéaires		93
3.1	Données du modèle ELECA	95
3.1.1	Données en entrée	95
3.1.2	Données en sortie	106
3.1.3	Résumé des spécifications des données du modèle	107
3.2	Description d'un exemple didactique	107
4 Guidage du regard virtuel		111
4.1	Contexte général	112
4.1.1	Objectif	112
4.1.2	Idée directrice	113
4.2	Exemple type	114

4.3	Modèle d'exploration de l'image	115
4.3.1	Saillance	116
4.3.2	Comportement du regard virtuel	130
4.4	Conclusion	137
5	Extraction de l'information locale	141
5.1	Contexte d'extraction pour le modèle ELECA	142
5.1.1	Objectif	143
5.1.2	Idée directrice	144
5.2	Exemple type	145
5.3	Modèle simple d'extraction de l'information locale	146
5.3.1	Filtrage du bruit de l'imagette	146
5.3.2	Détection de contours	149
5.3.3	Seuillage et analyse par composantes connexes	150
5.3.4	Extraction "instantanée" de l'orientation	154
5.4	Conclusion	155
6	Fusion des informations locales	159
6.1	Raisonnement spatial qualitatif et appréhension globale	160
6.1.1	Objectif	160
6.1.2	Idée directrice	161
6.2	Exemple	163
6.3	Vers un modèle de relations spatiales	164
6.3.1	Relations spatiales et segments de lignes	164
6.3.2	Scène visuelle et relations spatiales	170
6.4	Synthèse et équivalences	181
6.4.1	Ébauche du modèle CLE	181
6.4.2	Perspectives pour le modèle CLE	183
III	Développement logiciel	185
7	Conception du système	187
7.1	Du modèle ELECA au modèle informatique	188
7.1.1	Synthèse du modèle ELECA	188
7.1.2	Algorithme général tiré du modèle	190
7.2	Portée du logiciel	192
7.3	Conception logicielle	193
7.3.1	Étapes de traitement — Diagramme d'activité	194
7.3.2	Architecture logicielle — Diagramme de classes	211
7.4	Conclusion	220

8 Réalisation logicielle et enseignements tirés	221
8.1 Implantation	222
8.1.1 Couverture de l'image par le comportement saccadique	222
8.1.2 Extraction de l'information locale	232
8.2 Enseignements et bilan	238
8.2.1 Considérations générales	238
8.2.2 Solution adaptée au problème	239
8.2.3 Comparaison avec l'existant	240
8.3 Conclusion	243
IV Bilan	245
9 Contributions	247
9.1 Introduction	248
9.1.1 Le passé	248
9.1.2 Le présent	249
9.1.3 <i>Quid</i> du futur ?	249
9.2 Le modèle ELECA	250
9.2.1 Recherche de l'information dans l'image	251
9.2.2 Extraction de l'information de l'image	251
9.2.3 Fusion des informations locales en une carte globale	252
9.2.4 Modèle ELECA et objectif général de la thèse	253
9.3 Aspects logiciels	254
9.3.1 Conception d'un système original	254
9.3.2 Algorithmes et implantation du système	255
9.4 Considérations générales	255
9.4.1 Aspect transversal de la recherche	255
9.4.2 Aspect traitement d'image de télédétection	256
9.4.3 Stratégies générales	260
9.5 Conclusion	262
10 Conclusions et perspectives	263
10.1 Conclusions	263
10.2 Perspectives	265
Appendice	269
Bibliographie	271
Glossaire	293

A Compléments algorithmiques	297
A.1 Algorithme de Douglas-Peucker	297
A.2 Algorithme de fermeture transitive	300
A.2.1 Fermeture transitive d'un graphe	300
A.2.2 Algorithme de Roy-Warshall	303
B Exemple de spécifications du logiciel	305
Cas de la classe Database	305
C Compléments d'architecture	311
C.1 Extrait du fichier de paramètres XML	311
C.2 Interface de saisie des paramètres sous windows	312
C.3 Complément sur le diagramme d'activité	313
Index	315

Table des figures

1.1	Mouvements des yeux sur une image.	5
1.2	Modèle de contrôle <i>bottom-up</i> de l'attention Visuelle.	6
1.3	Banda Aceh avant et après le tsunami du 26 décembre 2004.	12
1.4	But du projet.	15
1.5	Méthodologie de la recherche doctorale.	17
2.1	Exemple d'images à très haute résolution, Quickbird et Ikonos	27
2.2	Opérations élémentaires en morphologie mathématique.	32
2.3	L'algorithme <i>watershed</i>	33
2.4	Ébauche primitive brute	42
2.5	Théorie de la Gestalt : regroupement par proximité.	44
2.6	Théorie de la Gestalt : regroupement par similitude.	44
2.7	Théorie de la Gestalt : regroupement par bonne continuation.	44
2.8	Théorie de la Gestalt : regroupement par fermeture.	45
2.9	Théorie de la Gestalt : regroupement par taille relative.	45
2.10	Théorie de la Gestalt : regroupement par environnement.	46
2.11	Théorie de la Gestalt : regroupement par symétrie.	46
2.12	Les trois règles de Marr.	47
2.13	Erreur de reconnaissance de caractère.	49
2.14	Principe de l'analyse par traits.	50
2.15	Principe de la vision naturelle.	54
2.16	Mode d'exploration oculaire d'une image.	55
2.17	Mode d'exploration oculaire d'une image en fonction de la tâche.	56
2.18	Modèle type pour le contrôle ascendant de l'attention	66
2.19	Modèle d'attention prenant en compte l'inhibition de retour	70
2.20	Les zooms appliqués par Machrouh sur une scène visuelle	73
2.21	Représentation d'une courbe par quadtree	76
2.22	Tendance et détail d'une courbe.	79
2.23	Exemple de graphe iconique hiérarchique.	81
2.24	Représentation de l'orientation entre points.	87
2.25	Représentation combinée de la distance et de l'orientation.	89
2.26	La fourmi du désert <i>Cataglyphis</i> en alerte.	90

3.1	Exemple d'image LANDSAT TM (nord de la ville de Québec)	97
3.2	Exemple d'image RADARSAT I (vallée de la Matépédia, Québec) . . .	100
3.3	Exemple d'image RSO aéroportée (forêt de Montmorency, Québec) . .	102
3.4	Ligne.	105
3.5	Exemple de données d'entrée raster.	107
3.6	Exemple d'éléments curvilignes.	108
3.7	Exemple de données vectorielles en sortie	109
4.1	Exemple de zone de recherche souhaitée.	114
4.2	Exemple de parcours du regard virtuel.	115
4.3	Illustration d'un problème de géoréférence.	117
4.4	Corridors de recherche.	118
4.5	Les deux modes du modèle	119
4.6	Première extraction.	121
4.7	Différents types de déplacement au sein des corridors de recherche. . . .	122
4.8	Densités linéaires en fonction de la courbe.	123
4.9	Cas possibles de nouveaux éléments.	125
4.10	Nouvel élément issu d'un nouvel élément.	125
4.11	Exemple de subdivision de l'image suite au mode reconnaissance. . . .	126
4.12	Stratégies en mode découverte.	128
4.13	Exemple d'une première saccade.	131
4.14	Deux types de saccades au sein des corridors de recherche	133
4.15	Deux types de saccades hors des corridors de recherche	134
4.16	Zone de saccade	135
4.17	Différents types de zones de saccade	136
4.18	Différence de largeur des corridors.	137
5.1	Exemple de parcours saccadique pour deux objets.	142
5.2	Exemple de vecteurs orientations.	145
5.3	Application du filtre de Lee à une image RSO.	148
5.4	Application du filtre de Robinson.	151
5.5	Image seuillée à 90.	152
5.6	Orientations instantanées.	155
6.1	Fusion des informations locales dans une perspective de carte globale. .	163
6.2	Exploration d'un environnement avec des occlusions.	166
6.3	Les six vues issues de la figure 6.2	167
6.4	Relations d'orientations entre deux dipôles.	168
6.5	Les 14 relations de Schlieder selon le formalisme de Moratz <i>et al.</i>	169
6.6	Dix relations supplémentaires.	169
6.7	Progression de l'appréhension de la connaissance dans le modèle LE. . .	172

6.8	Progression de l’appréhension de la connaissance dans le modèle CLE.	172
6.9	Transposition du formalisme de Moratz <i>et al.</i> à CLE.	173
6.10	Cas de points communs à deux dipôles dans le modèle CLE.	174
6.11	Deux voisinages en fonction de l’orientation	176
6.12	Voisinage (en gris) d’une courbe, cas général.	177
6.13	Reflète du parcours de recherche dans la liste des découvertes.	179
6.14	Deux voisinages possibles pour un même morceau de Courbe.	180
6.15	Exemple de courbe solution.	182
7.1	Vue d’ensemble du modèle ELECA.	189
7.2	<i>Use case</i> général	194
7.3	Diagramme d’activité général	194
7.4	Diagramme d’activité présentant les différents modules de VISION	195
7.5	Module de chargement	197
7.6	Module M1 : initialisation	199
7.7	Modules M2 –Reconnaissance grossière– et M3 –Reconnaissance fine.	201
7.8	Module M5 : Découverte — balayage attentionnel.	206
7.9	Modules M6 à M8.	209
7.10	Diagramme de classes général.	212
7.11	Relation entre points et éléments dans la structure de données	214
7.12	Héritages de la classe <code>Point</code>	215
7.13	Héritages de la classe <code>CurvilinearElt</code>	215
7.14	Mécanisme d’agrégation de la classe <code>GroupOfNewDiscoveredPoints</code>	216
7.15	Classe <code>Database</code> : agrégation de <code>DBCurvilinearElt</code>	217
7.16	Classes du système de contrôle de VISION.	218
7.17	Classes des systèmes de détection.	219
8.1	Magnitude, vecteur et “grand cercle” pour le mode reconnaissance.	226
8.2	Technique de recherche autour d’un point.	227
8.3	Comportement de l’œil le long d’un élément de la base de données.	231
8.4	Étiquetage des pixels après le premier passage.	234
A.1	Principe de base de l’algorithme de Douglas-Peucker.	298
A.2	Graphe et sa matrice associée.	301
A.3	Connexion réciproque des pixels.	302
A.4	Fermeture transitive du graphe.	302
B.1	Spécifications pour la classe <code>Database</code> – 1/5.	306
B.2	Spécifications pour la classe <code>Database</code> – 2/5.	307
B.3	Spécifications pour la classe <code>Database</code> – 3/5.	308
B.4	Spécifications pour la classe <code>Database</code> – 4/5.	309
B.5	Spécifications pour la classe <code>Database</code> – 5/5.	310

C.1 Interface	312
C.2 Module M4 : Développement de la liste des priorités en découverte . . .	313

Table des algorithmes

5.1	Algorithme général d'analyse en composantes connexes.	154
7.1	Algorithme général.	191
8.1	Algorithme simplifié de comportement de l'œil en reconnaissance. . . .	223
8.2	Algorithme rechercheAutour simplifié.	228
8.3	Algorithme méthodeDécouverte simplifié.	229
8.4	Algorithme général pour un point de fixation.	233
8.5	Algorithme classique d'analyse par composantes connexes.	233
8.6	Algorithme d'analyse par composantes connexes modifié.	237
A.1	Algorithme de Douglas-Peucker pas à pas.	299
A.2	Algorithme de Roy-Warshall.	303

Liste des tableaux

2.1	Les 13 relations d'Allen.	85
3.1	Exemple d'organisation d'une base de données topographiques	104
3.2	Spécifications des données du modèle ELECA	106
4.1	Cycles de spécifications-solutions pour le guidage du regard virtuel. . .	139
5.1	Cycles de spécifications-solutions pour l'extraction locale.	157
6.1	Points porteurs de LE vis-à-vis de la problématique	171
6.2	Transposition des notions de LE vers CLE	181
8.1	Termes utilisés dans les algorithmes du chapitre 8	225

Première partie

Cadre général

Chapitre 1

Introduction

*Le commencement est la moitié de tout.
Pythagore.*

Ce travail de doctorat se situe dans le domaine de l'extraction des formes. La première section de ce chapitre présente le contexte dans lequel s'inscrit ce travail de doctorat. La problématique que nous avons définie est présentée dans une deuxième section puis suit la démarche de résolution proposée dans cette thèse. Dans la section suivante, les objectifs de la recherche sont définis. La cinquième section est consacrée à l'explication de la méthode de recherche qui a été suivie. Enfin une dernière section donne un aperçu de l'articulation générale du document.

1.1 Contexte

1.1.1 Généralités

Depuis longtemps l'extraction de formes sur des images basse résolution en imagerie digitale a nécessité d'importantes ressources manuelles. De nombreux efforts ont été consacrés pour réduire le temps investi pour ce type de tâche en utilisant des techniques d'extraction semi-automatiques ou automatiques. Les éléments curvilignes, dont les réseaux hydrographiques [Yee87], [DLN94], [Fit99], [Zha00] et [DNR02], les linéaments géologiques [WH89] et les réseaux routiers [FTW81], [MD88], [ZC93], [GJ96], [MZ96] et [BCZC97], ont été particulièrement étudiés. En dépit de ces travaux, l'extraction d'éléments curvilignes sur des images de télédétection reste un problème ouvert et un défi important pour les algorithmes de traitement d'image [Qua04]. C'est particulièrement le cas lorsque les éléments cibles sont proches de la limite de résolution ou que les images sont bruitées.

Les méthodes classiques en télédétection utilisent des approches multispectrales comme des classifications de type supervisé ou non supervisé. Ces méthodes reposent sur des techniques de vision numérique élaborées il y a déjà quelque temps puisqu'elles mêmes reposent sur des éléments de traitement du signal [CC01]. La télédétection, plus récemment, s'est penchée sur les approches logiques permettant de mettre en œuvre des filtres morphologiques (Fitzback [Fit99] en utilise par exemple) développées en traitement d'image [Tou90] et en mathématiques appliquées [VC95]. Enfin quelques travaux de télédétection ont intégré des composantes issues de l'étude de la perception visuelle, en particulier le groupement perceptuel [AM96]. De manière générale, historiquement, la télédétection a utilisé des techniques issues de la vision numérique (et du traitement

d'images) avec un décalage de quelques années.

La vision numérique se tourne maintenant vers des approches s'inspirant de la vision humaine et de sa composante attentionnelle [Mac02], [NI02], [SI04]. En effet, la compréhension du système visuel humain a fait d'énormes progrès depuis le début des années 1980. Suivant les travaux pionniers de David Marr [Mar82], les modèles de l'attention visuelle ont été considérablement raffinés grâce à l'accélération des capacités de traitement informatique. Les modèles récents [Wol94], [BG00], [Mac02], [Itt03], sont souvent basés sur les théories attentionnelles et sur les résultats de la recherche en neuropsychologie.

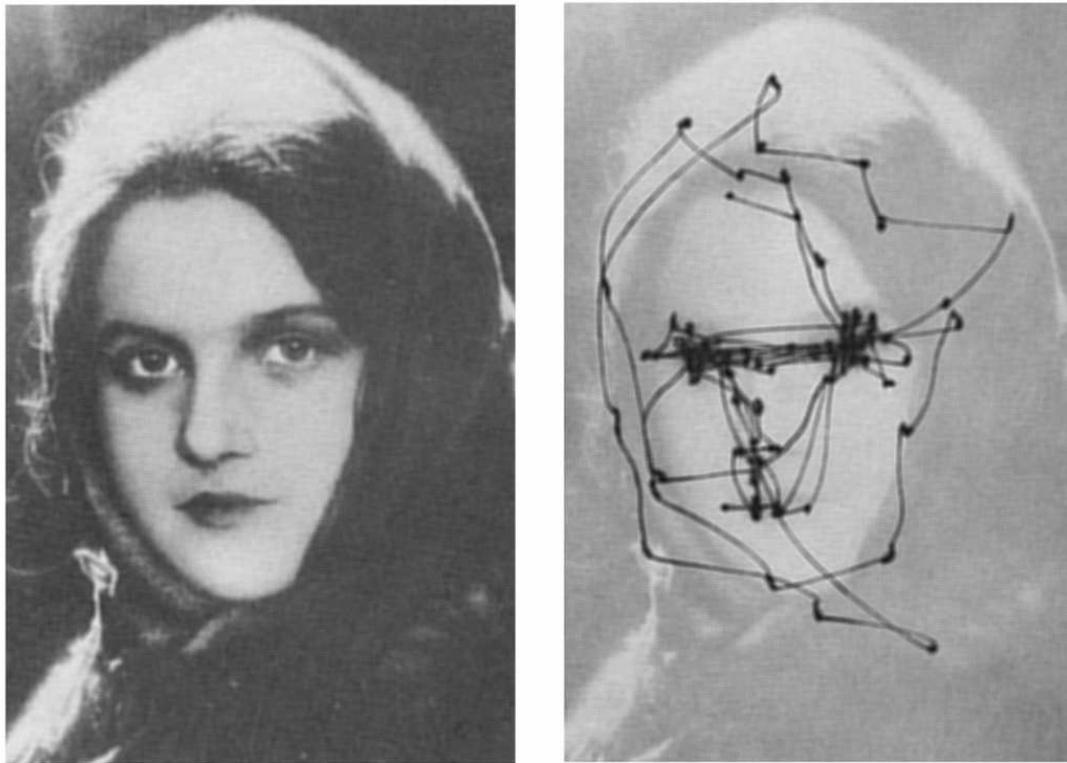


FIG. 1.1 – Mouvements des yeux sur une image. Ces mouvements ne sont pas aléatoires mais répondent à un pattern particulier, d'après [Yar67].

Il a été montré en effet que le mouvement des yeux durant une tâche de recherche

visuelle n'est pas aléatoire, figure 1.1, mais est guidé par des traits particuliers de la scène visuelle [Yar67], [Gre00] et [HB05]. Dans le cas qui nous intéresse, le résultat le plus important est que le regard naturel des sujets humains semble *suivre* les traits saillants à l'intérieur de l'image.

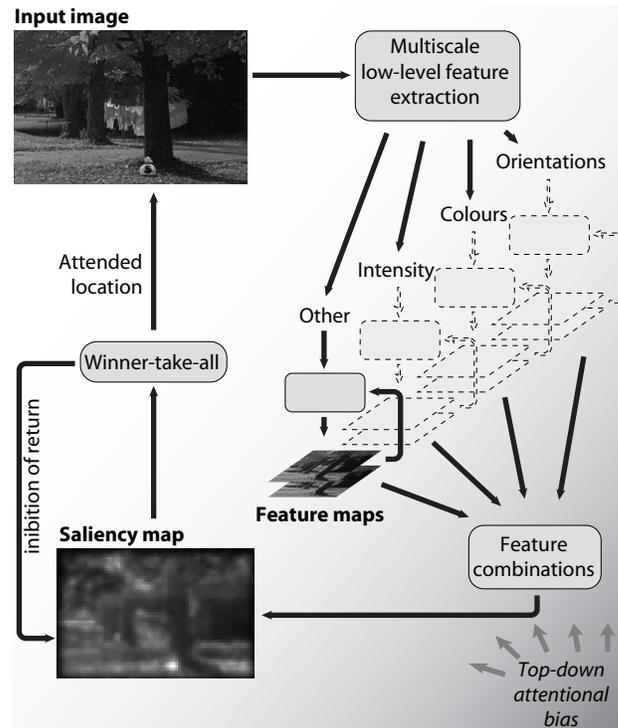


FIG. 1.2 – Diagramme de flux d'un modèle type de contrôle *ascendant* de l'attention visuelle ; version simplifiée de [KU85] et modifiée par [IK01] afin d'intégrer une composante *descendante*. Pour plus de détails voir page 66.

La tâche de recherche visuelle (et donc les modèles qui en ont découlé) a été considérée pendant longtemps comme un processus *ascendant*¹. Un des modèles les plus populaires et des plus repris a été proposé en 1985 par Koch et Ullman [KU85]. La figure 1.2 est une vue d'ensemble du modèle en question. Le modèle repose sur l'hypothèse qu'une carte de saillance encode l'intensité des stimuli pour l'ensemble des positions de la vue présentée au sujet. Cette carte de saillance est produite à partir d'entrées provenant des processus primaires de la vision (*traitement pré-attentionnel des caractéristiques*

¹Ascendant : *bottom-up* dans la littérature anglo-saxonne ; voir 2.2 page 38.

*visuelles*²); elle autorise ainsi une stratégie de contrôle efficace qui permet au “focus attentionnel” de parcourir les éléments présents dans la carte de saillance dans un ordre décroissant d’intensité de saillance. Plus tard le modèle de Koch et Ullman a été modifié par Itti et Koch [IK01] afin d’intégrer une composante *descendante*³ au modèle comme le montre l’élément en bas à droite de la figure 1.2. Toutes les notions abordées dans ce paragraphe sont explicitées plus en détail dans le chapitre 2.

Certains modèles de recherche visuelle sont purement descendants [Itt03]. Ils sont principalement utilisés pour le contrôle de robots. Ces modèles guidés par la tâche⁴ ont été inspirés par les travaux entrepris par Yarbus en 1967 [Yar67]. Yarbus a montré que le pattern de déplacement du regard pour une image donnée change en fonction de la tâche que le sujet a à accomplir – voir figure 2.17, page 56.

1.1.2 Spécificités

L’intérêt de se tourner vers les nouvelles méthodes d’extraction est guidé par les besoins actuels de la géomatique et des systèmes d’information géographique en particulier. Dans le cas de ce doctorat le contexte spécifique est propre à deux grands projets consécutifs du Réseau de Centres d’Excellence du Canada en Géomatique – GEOIDE⁵. Ces deux projets, ENV#19 et MNG#ELD, ont comme intitulés respectifs *Extraction of features from remote-sensed imagery for a search and rescue synthetic vision database*⁶ et *Intelligent data fusion for aircraft navigation and disaster management*⁷. Ils

²Dans la littérature anglophone on parle de *pre-attentive computation of visual features*.

³Descendant : *top-down* dans la littérature anglo-saxonne ; voir 2.2 page 38

⁴*task-dependent models*.

⁵GEOIDE : La GÉOmatique pour des Informations et des Décisions Éclairées – *GEOmatics for Informed DEcisions* – <http://www.geoide.ulaval.ca> .

⁶Extraction d’objets d’images de télédétection pour la création d’une base de données en support à la vision synthétique pour la recherche et le sauvetage.

⁷Fusion intelligente de données pour la navigation aérienne et la gestion des désastres.

regroupent des universités, des partenaires publics et des sociétés privées. Cette thèse de doctorat a été financée par ces projets.

L'un des objectifs principaux de ces projets est de créer une base de données qui doit servir à améliorer le pilotage des hélicoptères lors de missions de sauvetage en utilisant un casque muni d'un écran qui permet d'obtenir une vision synthétique et/ou augmentée. Dans ce contexte la base de données doit être la plus à jour possible.

1.2 Problématique

1.2.1 Projets GEOIDE

Le problème général auquel s'attaquent les projets GEOIDE dans lequel s'inscrit ce doctorat est l'amélioration de la navigation aérienne à vue. En effet, il s'agit du mode principal de navigation pour les pilotes d'hélicoptères de recherche et de sauvetage lorsque ceux-ci sont proches du sol. Ce problème peut se décomposer à la manière des poupées gigognes ; c'est-à-dire en problèmes de plus en plus spécifiques.

La navigation aérienne à vue nécessite, par définition, de voir le terrain survolé. Dans le cas courant de mauvaises conditions climatiques (ce qui est le plus fréquent lors de missions de sauvetage), le terrain réel n'est pas toujours visible ; en particulier dans un contexte canadien ou neige et brouillard peuvent former un rideau impénétrable à la vue. Pour s'affranchir de cela, on utilise un terrain synthétique (issu d'une base de données chargée dans l'hélicoptère) qui est projeté sur un écran du casque porté par le pilote afin que ce dernier puisse "voir" le terrain comme s'il n'y avait pas d'éléments

climatiques pour faire obstacle à la vue. On appelle cette technique vision synthétique⁸ ou vision augmentée⁹.

Deux problèmes découlent de l'utilisation de cette technique de vision augmentée. Le premier problème réside dans la gestion de l'affichage de la base de données de telle façon que le pilote se trouve dans des conditions similaires à l'environnement naturel qui lui est familier. Ce problème fait l'objet de recherches spécifiques dans les projets GEOIDE – Universités York et McGill en particulier. Le second problème est de construire la base de données qui sera affichée sous forme de terrain synthétique.

Dans le contexte de la recherche et du sauvetage, le problème est que *la base de données se doit d'être impérativement à jour* (voir plus loin, la figure 1.3 en particulier) et être construite (ou mise à jour) très *rapidement*. En effet, il faut que cette base soit chargée à bord de l'hélicoptère le plus rapidement possible lors du déclenchement de la mission de recherche et sauvetage. La seule alternative alors est de se tourner vers des *méthodes automatiques d'extraction* à partir d'images de télédétection acquises le plus récemment possible. Le problème d'extraction d'informations d'images de télédétection étant complexe, il a été divisé selon trois axes. Un premier axe concerne les modèles numériques de terrain (aspects abordés à l'University of British Columbia en particulier), un second axe concerne les éléments surfaciques présents dans les images (CRESTech¹⁰ en particulier) et enfin le troisième axe est consacré aux problèmes d'extractions d'objets curvilignes d'images de télédétection ; c'est dans cet axe que se place ce travail doctoral.

⁸*Synthetic vision.*

⁹*Enhanced vision.*

¹⁰Centre for Research in Earth and Space Technology, Ontario

1.2.2 Doctorat

L'extraction d'éléments curvilignes sur des images de télédétection est un problème difficile du point de vue algorithmique; spécialement lorsque les objets cibles sont proches de la limite de résolution ou qu'ils sont en partie masqués — végétation, nuages...

Ainsi, nous faisons la série de constatations que :

1. l'extraction d'éléments curvilignes à partir d'images de télédétection reste une question cruciale car c'est un problème non résolu en informatique d'une manière générale et que la nécessité de méthodes efficaces est toujours plus grande en raison de l'accroissement du nombre de données et des outils basés sur des services géolocalisés;
2. l'extraction d'éléments curvilignes est souvent basée sur des techniques multispectrales (approches de classification globales descendantes), lorsque les images permettent une telle démarche (les méthodes actuelles ne peuvent être appliquées avec succès sur toutes les images);
3. les modèles de recherche visuelle sont nombreux et relativement bien connus mais coûteux, laborieux et ne prennent pas en compte certaines spécificités des données existantes comme des données extérieures fournies par des systèmes autres que la seule imagerie (dans notre cas il s'agit de données issues de bases de données géographiques) limitant les angles d'attaque au problème d'extraction;
4. les mouvements des yeux sont guidés par des caractéristiques précises de la scène visuelle pour une tâche donnée, ce qui évite un traitement global de la scène visuelle et permet une très grande efficacité de recherche.

En effet, les bases de données mises à jour et réalisées rapidement deviennent de

plus en plus cruciales pour les services géolocalisés (*location-based services*) tels que le pilotage d'hélicoptère pour la recherche et le sauvetage qui est le contexte spécifique de ce doctorat. La tristement célèbre figure 1.3 illustre ce propos. On voit comment des éléments (curvilignes en particulier) peuvent disparaître en très peu de temps. Dans cet exemple nous nous situons dans le cas du Tsunami du 26 décembre 2004 en Asie du Sud-Est ; si l'on se place dans une perspective plus canadienne, le dépôt ou la fonte de neige, par exemple, produit le même effet sur le paysage.

Comme nous l'avons vu un peu plus haut, l'extraction d'éléments curvilignes est toujours un défi important pour les algorithmes de traitement d'image alors que pour des sujets humains il s'agit d'une tâche naturelle et immédiate. Cette tâche devient simple et aisée pour des experts en photo-interprétation.

Par conséquent le problème est de savoir s'il est possible de proposer une méthode, de type automatique et générique, qui s'appuie sur des approches inusitées en télédétection comme les mécanismes attentionnels pour solutionner la question de l'extraction des éléments curvilignes d'image de télédétection, en particulier quand ces images ne sont pas traitables par les méthodes conventionnelles.

1.3 Démarche de résolution

Nous proposons d'utiliser une technique d'échantillonnage par saccades inspirée de la recherche visuelle naturelle (humaine) afin de guider un regard virtuel qui permettrait d'identifier les éléments curvilignes sur des images de télédétection [Cot02]. Cette stratégie doit pouvoir accroître considérablement l'efficacité des techniques d'ex-

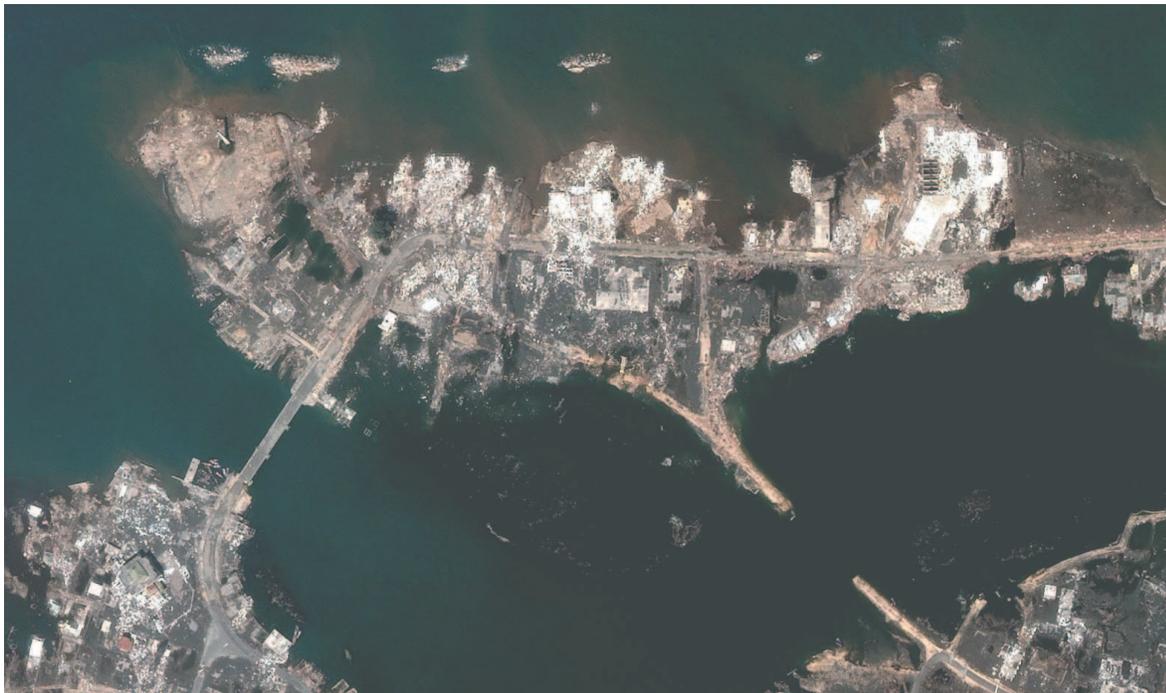


FIG. 1.3 – Banda Aceh avant (23 juin 2004) et deux jours après le tsunami (28 décembre 2004), images QuickBird – Digital Globe.

traction automatique. Elle devrait également être très utile pour mettre à jour les bases de données géographiques (comme celles utilisées par les systèmes d'information géographique – SIG) de façon rapide et automatique.

Nous supposons que les éléments curvilignes (c'est-à-dire les routes, les cours d'eau, les lignes de haute tension, *etc.*) sont déjà présents dans les bases de données géographiques. En revanche ce sont souvent des représentations du terrain qui ne sont plus à jour puisqu'elles reflètent simplement une information exacte au moment où la carte/base de données a été produite. Dans le but de mettre à jour les bases de données géographiques, nous utilisons les anciennes données afin de guider la recherche dans l'image. Cette stratégie nous permet de concentrer la recherche directement aux endroits où la probabilité de trouver de l'information recherchée est élevée. Ceci est réalisé en nous inspirant du comportement de l'attention humaine et de son centre d'intérêt. Il devient ainsi possible de traiter précisément et de mettre à jour les données géographiques en utilisant l'information nouvellement collectée sur l'image.

Nous proposons de parcourir l'image en utilisant une technique d'échantillonnage par saccades [CMEL03] pour ne pas traiter l'entièreté de l'image comme le fait l'œil humain, voir chapitre 2. Cette approche d'échantillonnage par saccades permet de réaliser un réel sous échantillonnage. En effet, nous verrons plus avant que d'autres techniques (basées sur la multi résolution) utilisent également un sous échantillonnage de l'image mais doivent pour ce faire traiter préalablement l'image dans son ensemble. La manière de sous échantillonner l'image que nous proposons apparaît donc originale et comme devant être plus efficace vis-à-vis des techniques connues. Enfin, la figure 1.4 résume les entrées du système que nous proposons ainsi que ses sorties. Les entrées (données) sont l'image de télédétection (à jour) et les données SIG qui (elles) ne sont pas à jour ; les sorties sont les données SIG à jour.

1.4 Objectifs du travail

L'objectif général du travail est, par conséquent, de proposer une méthode d'extraction des éléments curvilignes à partir d'images de télédétection qui puisse permettre la mise à jour de bases de données géographiques, figure 1.4. La base de données doit servir à améliorer le pilotage des hélicoptères¹¹. Les données utilisées sont d'une part les données géographiques déjà existantes et d'autre part des images de télédétection le plus récemment acquises.

Trois objectifs spécifiques doivent être réalisés afin d'atteindre cet objectif général. Ces objectifs correspondent à trois étapes importantes du modèle général développé dans cette thèse et doivent être réalisés les uns après les autres :

1. Mise au point d'une méthode efficace pour guider le regard virtuel sur la scène visuelle afin d'éviter un traitement inutile de l'ensemble de l'image.
2. Mise au point d'une technique qui permette, au sein des zones identifiées par le regard virtuel, d'extraire l'information locale pertinente.
3. Élaboration d'un outil qui rassemble l'ensemble de l'information locale au sein d'une perspective globale et ceci dans le but de mettre à jour la base de données.

1.5 Méthode de recherche

La méthodologie qui a été suivie durant cette recherche doctorale est présentée dans la figure 1.5. Cette recherche s'est inscrite dans les deux projets plus larges abordés plus haut (à savoir GEOIDE ENV#19 et GEOIDE MNG#ELD) ayant (eu) leurs contraintes

¹¹Dans le cadre des projets GEOIDE évoqués plus haut dans ce chapitre.

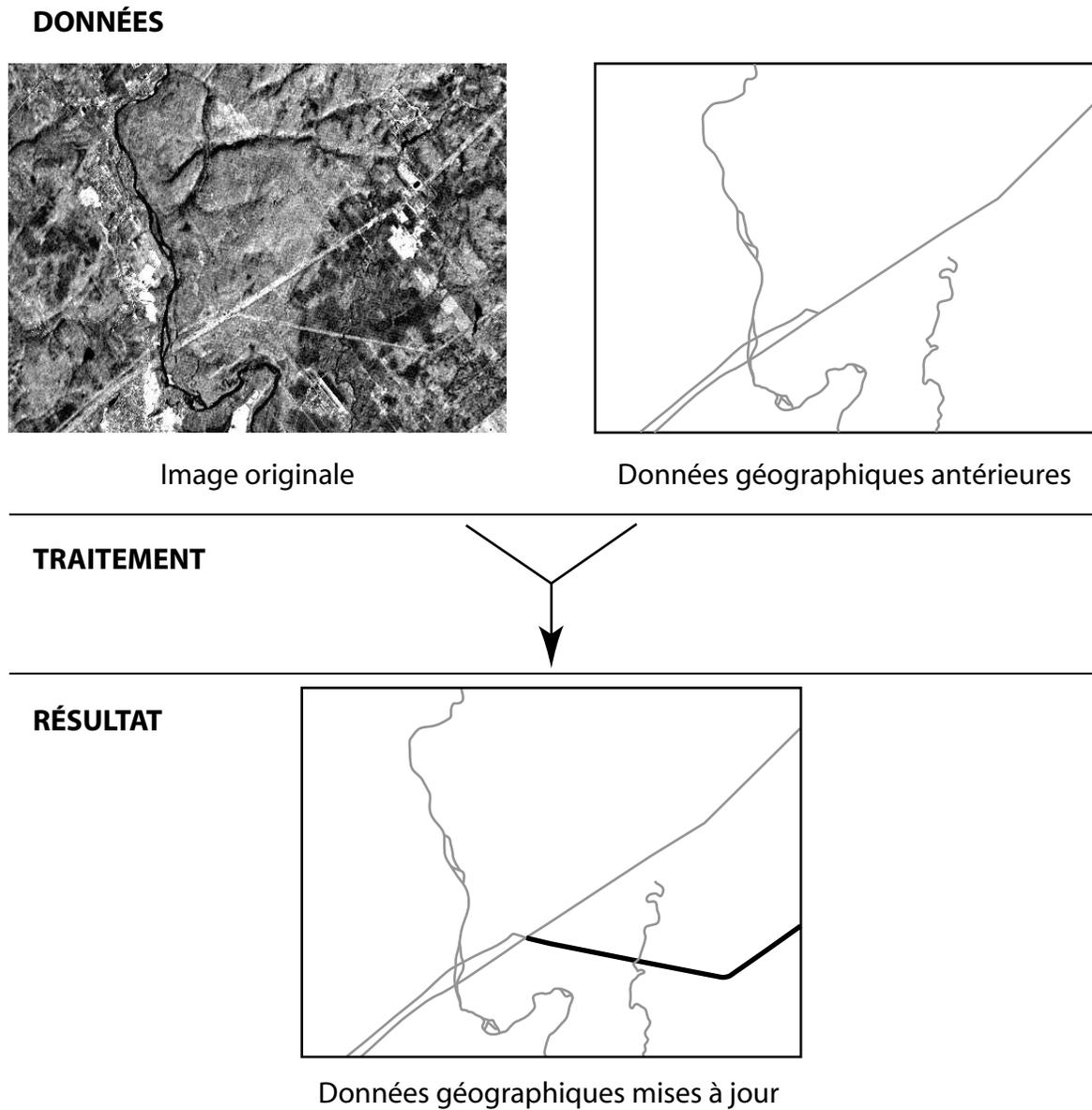


FIG. 1.4 – But du projet.

propres, et qui ont conditionné quelque peu les objectifs du travail comme mentionnés au cours des deux sections précédentes.

La première étape de la recherche a été de circonscrire la problématique abordée au sein de ces projets. En particulier, nous avons appréhendé le domaine d'étude (inconnu jusqu'alors), la problématique des projets généraux (GEOIDE), la problématique du

doctorat à proprement parler au sein de ces deux grands projets et également le type de données avec lesquelles il allait falloir travailler. En outre, des cours de télédétection et de traitement d'image ont aidé ce travail préliminaire d'éclaircissement.

La seconde étape a consisté à effectuer des cycles de recherches bibliographiques afin de préciser le contexte de recherche. Ces cycles ont visé deux aspects intimement liés durant toute cette étape. Il s'est agi d'une part, d'asseoir nos connaissances sur les domaines abordés et, d'autre part, de faire un travail exploratoire de domaines connexes à la géomatique en s'affranchissant quelque peu des limites traditionnelles de sa définition¹². Les domaines explorés ont compris, entre autres, la télédétection¹³, la fusion de données¹⁴, le traitement d'image¹⁵, l'informatique¹⁶, le raisonnement qualitatif spatial, la cognition spatiale¹⁷, la vision et les mécanismes attentionnels. Les aspects plus spécifiques au sujet ont été abordés dans une étape ultérieure, voir plus loin.

La troisième étape méthodologique de la recherche a été conditionnée par l'examen doctoral. L'examen doctoral est une étape académique obligatoire dans le cheminement du doctorat en Sciences Géomatiques. L'examen comporte plusieurs volets, ceux qui nous intéressent ici sont au nombre de trois. D'abord les objectifs de la recherche, qui ont été énoncés plus haut, ont été définis à cette étape. Ensuite le rapport préparé pour l'examen doctoral [Cot02] a permis d'évaluer la faisabilité du projet en déroulant un

¹²Les quatre disciplines de bases de la géomatique sont le droit, la physique, les mathématiques et l'informatique. Ses sept disciplines spécifiques sont -1-la géodésie, -2-la télédétection, -3-la photogrammétrie, -4-la topométrie, -5-la cartographie et les SIRS (Systèmes d'Information à Références Spatiales), -6-l'hydrographie et -7-la gestion et l'aménagement fonciers [Car00].

¹³Également pris le cours de programme gradué "*Télédétection Spectrale*", hiver 2001.

¹⁴Un groupe de discussion en fusion de données a été organisé durant l'automne 2000 et l'hiver 2001 au Centre de Recherche en Géomatique (CRG) [JEC01], animé par le Dr. Robert Jeansoulin – <http://fusion.scg.ulaval.ca/>.

¹⁵Également pris le cours de programme gradué "*Traitement des images*", hiver 2001.

¹⁶Auditeur libre au cours du Professeur Bernard Moulin de "*Projets orientés-objets (conception et gestion)*" à la session d'hiver 2002.

¹⁷Ce doctorat s'est déroulé au sein de la Chaire de recherche du Canada en géomatique cognitive du Dr. Geoffrey Edwards – <http://crccg.crg.ulaval.ca/>.

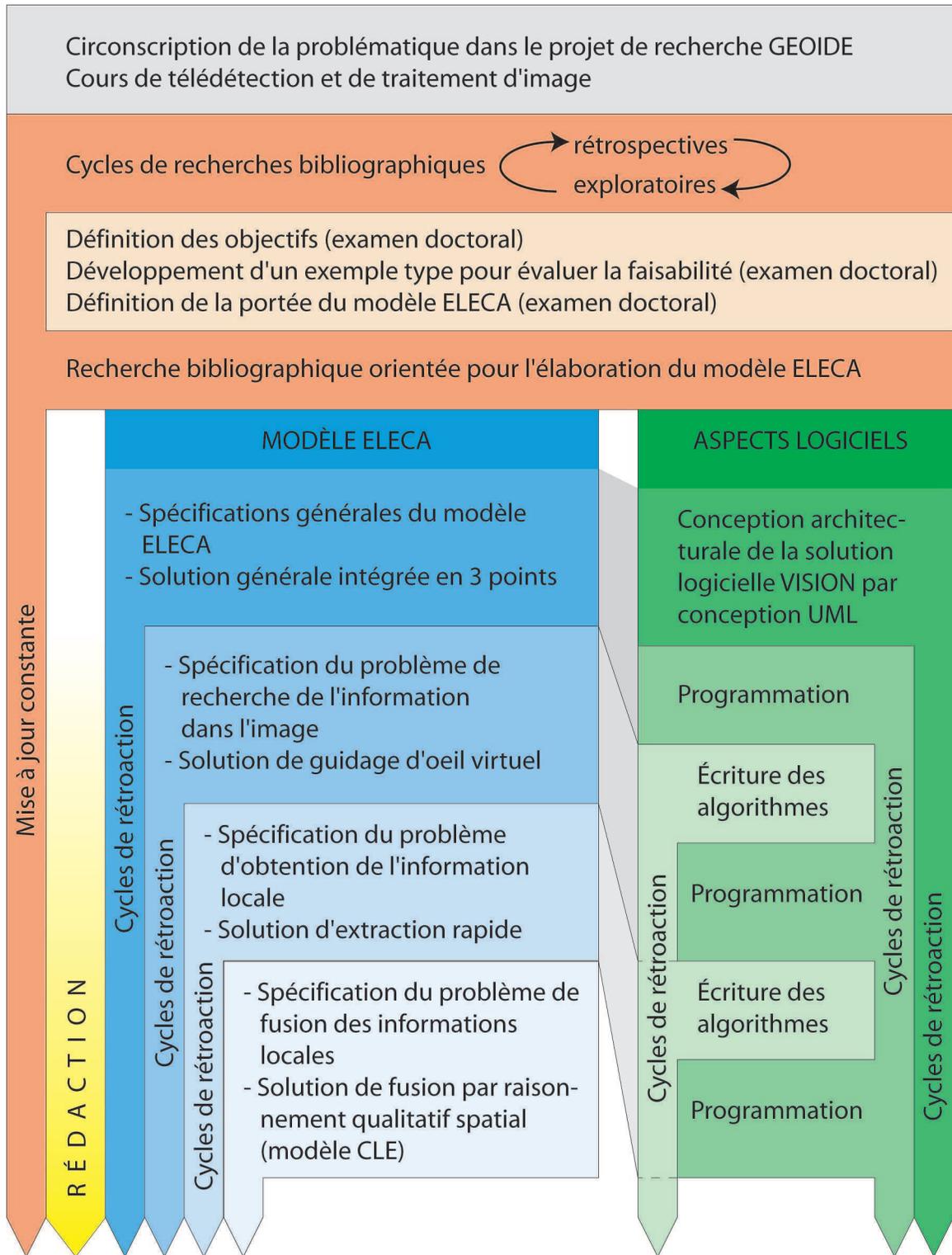


FIG. 1.5 – Méthodologie de la recherche doctorale.

exemple type. Cet exemple montre pas à pas, de manière illustrée, les problèmes que doit résoudre la recherche¹⁸. Enfin, cette étape de l'examen doctoral a permis de définir la portée du modèle envisagé pour répondre à la problématique posée et atteindre les objectifs fixés. Ce modèle, qui est au cœur de ce travail, a été baptisé *ELECA* : *Extraction of Linear Elements by Cognitive Approach* – Éléments Linéaires Extraits par Calcul Attentionnel. Le modèle ELECA a été élaboré par la suite, voir plus bas.

Une fois cette importante étape achevée, une nouvelle recherche bibliographique a couvert des aspects plus précis mais moins nombreux que lors de la seconde étape. En particulier l'accent a été mis sur les mécanismes attentionnels et la reconnaissance des formes. Comme le montre la figure 1.5, la recherche bibliographique s'est poursuivie ensuite durant toute la durée du doctorat.

Pour les étapes ultérieures, qui constituent le cœur de la thèse, nous avons adopté *une approche par cycles successifs de spécifications qui mènent à des solutions*, avec un niveau d'abstraction à chaque fois plus bas. Ceci a été fait en parallèle aussi bien pour le modèle ELECA, en bleu sur la figure 1.5 que pour la solution logicielle proposée, en vert. En permanence, des cycles de rétroaction sur les étapes précédentes ont permis de corriger les imperfections identifiées dans les étapes suivantes à la manière des itérations successives d'un processus UML [Mou02]. Pour ne pas alourdir le texte, nous ne répréciserons plus que chaque étape abordée plus bas comporte des cycles de rétroaction ; ces cycles sont visibles sur la figure 1.5. De plus, une documentation technique a été écrite durant tout le processus de conception et de programmation du logiciel.

La première étape du modèle ELECA a été de spécifier de manière générale le

¹⁸L'utilisation d'un exemple type a été reprise dans la seconde partie de cette thèse afin, dans ce cas ci, de situer le lecteur de manière didactique.

modèle lui-même en termes de données manipulées principalement ; cet aspect sera abordé au chapitre 3. La solution adoptée reprend les trois objectifs énoncés plus haut. La conception de la solution logicielle a débuté en parallèle. Cette conception a été réalisée en utilisant le formalisme UML – chapitre 7. La première phase de la conception a été de modéliser la structure de données.

La seconde étape de réalisation du modèle ELECA a été de spécifier le problème de recherche de l'information dans l'image et d'y apporter une solution originale par guidage d'un œil virtuel sur l'image – chapitre 4. La littérature concernant les domaines liés à cette étape fait l'objet des sections 2.3 et 2.4 du chapitre consacré à l'état de l'art. La conception UML s'est poursuivie durant cette étape puis la programmation a commencé par l'architecture logicielle. Le programme a été baptisé *VISION*. Toujours durant cette étape d'ELECA, et grâce à la solution proposée, l'écriture des algorithmes pour le guidage de l'œil virtuel a commencé – chapitre 8.

La troisième étape d'élaboration du modèle ELECA a porté sur la spécification du problème de l'obtention d'informations locales pertinentes. Une solution d'extraction rapide a été élaborée – chapitre 5. Un survol de la littérature en lien avec ces aspects est présenté dans le chapitre 2, principalement au point 2.2.5. Durant cette étape, du côté logiciel l'écriture des algorithmes pour le guidage de l'œil virtuel a été achevée dans ses grands traits¹⁹.

La quatrième étape méthodologique a été de spécifier le problème de fusion des informations locales. Une ébauche de solution a été formalisée (le modèle CLE) par équivalence avec un modèle de la littérature – chapitre 6. La solution repose sur une approche du raisonnement qualitatif spatial. Le raisonnement qualitatif spatial est abordé

¹⁹Comme toujours, les cycles de rétroaction vont revenir tout au long des étapes suivantes pour affiner ces algorithmes, figure 1.5.

d'un point de vue bibliographique au paragraphe 2.5. Du côté du volet logiciel, au cours de cette dernière étape, les algorithmes en lien avec l'extraction locale ont été écrits (chapitre 8) et la programmation de ces derniers entamée.

Enfin la dernière étape méthodologique a été d'évaluer, d'une part la contribution de ce travail (chapitre 9), d'autre part la portée du logiciel (chapitre 8), et enfin les perspectives ouvertes par cette thèse – chapitre 10.

1.6 Articulation générale du document

Cette thèse est divisée en quatre parties pour une lecture plus facile. De plus nous avons produit ce document en prenant délibérément le parti de présenter de nombreuses figures suivant l'adage qui veut qu'un bon dessin vaille mieux qu'un long discours²⁰. Enfin avant de passer à un survol général de la thèse, mentionnons qu'en plus des figures (FIG.) et des tableaux (TAB.) classiques, nous avons rajouté un nouveau type d'objet de référence, l'algorithme, noté ALG.

La première partie situe le lecteur par rapport aux domaines abordés tant du côté du travail de la thèse proprement dit que du côté de la littérature.

Le chapitre 1 présente le thème de recherche et la problématique générale de cette thèse. Il donne des pistes qui seront explorées ensuite pour pouvoir dégager les objectifs de recherche. Enfin, le chapitre 1 aborde également la méthode de recherche qui a été suivie durant ce travail.

Le chapitre 2 aborde la littérature dans les différents domaines survolés par

²⁰Ou bien : “une image vaut mille mots”, en français du Québec.

ce doctorat. Elle commence par dresser le bilan de ce qui se fait dans le domaine de l'extraction d'objets linéaires en télédétection puis creuse ensuite des domaines plus spécifiques en lien avec le travail.

La deuxième partie est consacrée à l'élaboration d'un modèle qui réponde aux objectifs fixés.

Le chapitre 3 propose au lecteur une revue (1) des types de données qui interviennent dans la thèse, (2) du contexte lié à ces données et (3) des problèmes sous-jacents au contexte. Ce chapitre décrit également un exemple type qui suivra le lecteur afin de rendre la présentation plus didactique.

Le chapitre 4 aborde le problème de la recherche de l'information dans l'image. Il montre quels sont les problèmes et propose un modèle inspiré de la vision humaine pour les solutionner.

Le chapitre 5 est consacré à l'extraction des information locales. Il spécifie les problèmes généraux et les problèmes spécifiques au modèle proposé. Il fournit une réponse algorithmique simple à l'extraction locale rapide d'éléments curvilignes.

Le chapitre 6 s'attaque à la fusion des informations locales. Après avoir fait une lecture particulière de travaux en raisonnement spatial qualitatif, il montre qu'il est possible de concevoir un modèle de fusion qui s'appuie sur les travaux précités en les interprétant de manière originale.

La troisième partie est consacrée à l'élaboration d'une solution logicielle qui repose sur le modèle élaboré dans la partie précédente.

Le chapitre 7 s'attache à montrer qu'il est possible de fournir une solution logicielle au modèle proposé dans les trois chapitres précédents. Dans un premier temps il montre que le modèle est transposable dans le monde informatique,

puis dans un second temps explicite comment le faire et développe une architecture propre.

Le chapitre 8 propose des solutions algorithmiques originales au modèle développé dans les chapitres précédents et directement implantables. Ce chapitre dresse également un bilan des aspects logiciels du travail.

La quatrième partie dresse un bilan général du travail réalisé dans cette thèse.

Le chapitre 9 est une revue des principales contributions de ce travail. Il replace le travail dans le contexte historique de la recherche, puis montre sa portée dans les différents domaines abordés. Enfin il se penche sur l'atteinte des objectifs fixés au début du travail.

Le chapitre 10 tire les conclusions de ce travail en en soulignant ses points forts et ses faiblesses. Il conclue par un bref survol des perspectives de recherche qui découlent de ce travail.

L'appendice contient la liste des références bibliographiques de l'ouvrage, *un court glossaire* et les annexes de la thèse. Les termes explicités dans le glossaire sont signalés par cette *police*.

Bonne lecture.

Chapitre 2

État de l'art

*L'art est fait pour troubler. La science rassure.
Georges Braque, Le Jour et la Nuit.*

Ce chapitre dresse un portrait, d'une part, de ce qui se fait à l'heure actuelle en télédétection appliquée à l'extraction d'éléments curvilignes, et, d'autre part, des recherches dans les différents domaines que ce travail doctoral touche :

Dans un premier temps, la section 2.1 situe le lecteur en télédétection d'objets curvilignes en donnant les enjeux par rapport aux contextes généraux (historique et types de données traitées) puis passe en revue les principales méthodes utilisées. Dans un second temps, sections 2.2 à 2.4, nous abordons les aspects relatifs aux visions numérique et naturelle, en nous attachant sur les transferts et connexions qu'il existe entre les deux. Enfin dans un troisième temps, §2.5, nous abordons le raisonnement spatial qualitatif, en particulier en considérant qu'il peut fournir des outils de reconstruction de l'environnement par navigation pour pouvoir étudier ultérieurement la construction de structures globales de l'image. Nous faisons le lien inédit entre reconstruction d'un l'environnement à partir de vues locales et construction de la compréhension des structures de l'image à partir de vues locales issues des saccades du regard virtuel.

2.1 Extraction d'éléments curvilignes d'images de télédétection

Cette section a pour objet de dresser un état des lieux général en télédétection sur les méthodes actuelles d'extraction des éléments linéaires. Rappelons ici que l'extraction d'éléments curvilignes d'images de télédétection est le contexte spécifique de ce doctorat. Cette section a été guidée en particulier par la lecture de [Qua04]. Dans un premier temps, nous donnons un aperçu historique de l'évolution des techniques de télédétection pour l'extraction d'objets curvilignes d'images. Dans un deuxième temps un survol des types d'images est fait ; puis dans un troisième temps, nous passons en revue les techniques les plus classiques en télédétection pour terminer dans un quatrième temps sur un bilan comparatif des différentes méthodes évoquées.

L'extraction automatique des structures linéaires des images de télédétection est depuis quelques décennies (lancement du premier satellite de télédétection –américain– LANDSAT-1 le 22 juillet 1972¹) le sujet d'intenses recherches. La multiplicité des capteurs et le nombre de plus en plus important de données forcent la réflexion vers des méthodes et techniques tournées vers le traitement le plus automatique possible des images de télédétection. Présentement les techniques s'appuient encore très largement sur l'opérateur (spécialistes de traitement d'image ou experts en interprétation) pour répondre aux exigences de mise à jour des systèmes d'information géographique [Qua04].

¹Par comparaison le premier satellite de télédétection européen a été lancé le 21 février 1986. Il s'agissait de SPOT-1 (Système Probatoire d'Observation de la Terre) dont le projet avait pourtant débuté au début des années 70. Rejeté par l'Europe (Agence Spatiale Européenne), ce sera la France (quelque peu aidée par la Suède et la Belgique) qui prendra ensuite le projet à sa charge, entraînant un gros retard.

2.1.1 Mise en situation

L'extraction de l'information d'image de télédétection (au sens large) remonte aux débuts de la photographie. Comme pour beaucoup de sciences, ce sont des besoins militaires qui ont poussé le développement des premières techniques pour extraire l'information à partir de données acquises par la photographie. Ainsi [WA00] mentionne que dès la moitié du XIX^{ème} siècle le corps des ingénieurs de l'armée française a utilisé des photos aériennes pour la reconnaissance et la cartographie.

Une des premières raisons pour lesquelles les sciences cartographiques se sont tournées vers la reconnaissance de structures curvilignes vient de l'importance qu'occupent les routes dans notre société. Le besoin immédiat qui en découle est que l'information les concernant soit la plus à jour possible [PMS⁺99].

La recherche vers des solutions automatisées remonte aux années 70 [Qua04]. Dans les années 70 et 80, la résolution spatiale des images, comme celles fournies par SPOT ou Landsat Thematic Mapper, ne permettait souvent pas de voir des structures linéaires étroites comme les routes. À ce propos Whang et Zhang [WZ00] ont mis en évidence que le succès de l'extraction de structures linéaires reposait particulièrement à l'époque sur la résolution spatiale du satellite. La nature² de l'objet sur l'image (pour un même objet) change en fonction de la résolution. En effet dans les images à basse résolution les routes apparaissent comme des structures curvilignes (ce sur quoi porte la recherche de cette thèse), alors que dans des images à plus haute résolution, les routes sont des régions homogènes qui satisfont certaines contraintes de taille et de forme. L'accroissement de l'accessibilité à des images à haute résolution spatiale a permis le développement de techniques d'extraction spécialement dédiées à ce type d'image [Qua04]. Par exemple

²Autrement dit son aspect sur l'image.

de nombreuses techniques d'extraction de routes sur des images à haute résolution utilisent des paires de bords (les deux bords de la route) pour détecter les cibles ; ces techniques sont par définition impossibles à utiliser à basse résolution puisque que dans ce cas les éléments apparaissent comme n'ayant pas d'épaisseur.

Évolution actuelle : guidée par les besoins d'automatisation La télédétection, quelle que soit la résolution à laquelle on travaille, a besoin de se doter d'outils d'extraction efficaces. En effet, l'extraordinaire quantité de données collectées (capteurs plus nombreux, plus efficaces, à plus haute résolution avec des taux de transfert des données toujours plus grands), rend le traitement manuel de ces dernières quasi impossible en terme de temps de traitement, mais également en termes pécuniaires. Or le maintien des bases de données à jour est essentiel pour de nombreux secteurs [Gib03]. Historiquement, les éléments curvilignes étaient extraits manuellement grâce à l'interprétation d'experts en utilisant des systèmes stéréoscopiques de vision [Yee87]. La mise à jour automatique de structures curvilignes qui sont partiellement occultées par la végétation, les nuages ou le manque de résolution devient rapidement (très) compliquée. Le travail humain, s'appuyant sur de telles données, devient lourd en termes de temps de traitement, augmentant ainsi de beaucoup le coût lié à la production de données pour un système d'information géographique.

2.1.2 Types d'images

Différents types d'imageries ont été utilisés dans la littérature pour faire l'extraction d'éléments curvilignes, en voici une liste non exhaustive [Qua04] :

- imagerie RADAR ;



FIG. 2.1 – Exemple d'images à haute résolution, Quickbird et Ikonos ; en haut image Quickbird de Banda Aceh acquise le 28 décembre 2004, en bas image Ikonos de la même région, acquise le 29 décembre 2004.

- données LiDAR³ ;
- imagerie thermique de régions agricoles ;
- imagerie multispectrale du visible ou du proche infrarouge (par exemple SPOT et LANDSAT) ;
- imagerie hyperspectrale ;
- imagerie à (très) haute résolution spatiale.

La littérature rapporte la plupart du temps l'extraction d'objets curvilignes menée sur une seule bande en haute résolution spatiale. La dernière génération de satellites (QuickBird, Ikonos) avec des capteurs permettant d'acquérir des images à très haute résolution spatiale offre des résolutions spatiales allant jusqu'au mètre, figure 2.1⁴. Le satellite Ikonos⁵ a une résolution de 1 mètre en panchromatique et de 4 mètres en multispectral ; le satellite Quickbird⁶ est encore plus précis avec des résolutions panchromatique et multispectrale respectivement de 0,61 mètre et de 2,44 mètres. D'une manière générale les méthodes d'extraction sur ces images reposent sur des techniques spectrales qui ne prennent pas en compte les relations spatiales entre les objets dans l'image contrairement à l'approche qui est préconisée dans cette thèse et développée dans les chapitres ultérieurs.

2.1.3 Extraction manuelle *vs* extraction automatique

Le traitement informatique et le traitement par des humains des données de télédétection ont chacun des avantages et des faiblesses. Un interprète humain peut facilement appréhender des images entières et reconnaître les objets ; de son côté l'informa-

³*Sensu stricto* le LiDAR n'est pas vraiment considéré comme un système d'imagerie puisqu'il fournit un nuage de points, mais il est abordé généralement avec les autres techniques permettant l'extraction d'éléments curvilignes [Qua04].

⁴Voir également la figure 1.3, page 12 pour laquelle la résolution est encore plus haute.

⁵Ikonos : <http://www.spaceimaging.com>

⁶Quickbird : <http://www.digitalglobe.com>

tique est performante dans l'optimisation, la délimitation de détails et la répétitivité des tâches [MJMC⁺96]. L'extraction manuelle utilise les capacités cognitives et l'expérience de la personne qui manipule les images mais est très coûteuse en terme de temps de traitement [BSM⁺99]. L'accroissement considérable de la quantité de données à traiter pour des besoins cartographiques oblige à imaginer des techniques d'extraction automatique des éléments de l'image de télédétection [Yee87]. D'après Pigeon *et al.* [PST⁺99], l'utilisation de méthodes automatiques ou semi automatiques peut également réduire de manière significative le temps de formation des opérateurs.

Les sujets humains ont une capacité particulière à regrouper de petits éléments atomiques comme des points ou des lignes en des structures générales de l'image [Gui99]. La littérature rapportait il y a encore peu [GL97] que la mise au point de techniques totalement automatiques était encore très loin de l'objectif. L'humain est capable de retrouver de l'information dans un environnement bruité et de s'adapter à différentes conditions sans qu'aucune information explicite n'ait été fournie. La mise au point de codes informatiques capables de réaliser ce type de comportement est un défi important [Qua04] qui reste toujours à relever.

Il convient de souligner que l'atteinte d'un objectif de mise au point de techniques d'extraction semi automatique est plus réaliste qu'un objectif d'automatisation complète, mais que chercher dans cette dernière direction est très pertinent pour l'amélioration des techniques actuelles. Le but est toujours de minimiser le plus possible l'intervention manuelle pour accélérer le temps de traitement de l'image d'une part et pour réduire le temps en ressources humaines d'autre part.

2.1.4 Techniques d'extraction en télédétection

Le spectre des techniques déployées pour l'extraction d'éléments curvilignes d'images de télédétection est très large. Le niveau d'automatisation de ces techniques varie également considérablement. Certaines nécessitent un nombre d'opérations important pour pouvoir fonctionner, alors que d'autres peuvent, à partir d'un petit nombre d'hypothèses de départ, fonctionner avec une bonne autonomie. La mise à jour de systèmes d'information géographique à partir des données de télédétection est un sujet de recherche habituel. De plus en plus on voit l'émergence de techniques qui essaient de tirer parti de l'information contenue dans les données des systèmes d'information géographique [Qua04]. D'une manière générale le développement de techniques d'extraction d'éléments curvilignes sur des images satellitales est très important. Le problème est toujours ouvert. L'automatisation du processus est encore loin d'être atteinte. Néanmoins, des progrès substantiels ont été faits, en particulier avec l'avènement des images à très haute résolution spatiale – citons par exemple l'amélioration du contrôle des données par des méthodes statistiques [GGPD04] ou l'utilisation plus poussée de méthodes combinées d'appariement et de multi résolution [ZC04].

Il convient de souligner que la plupart des approches actuelles d'extraction d'éléments linéaires sont des combinaisons de différents types de méthodes ; discriminer parmi ces méthodes pour les regrouper en grandes tendances devient de plus en plus difficile. Ainsi, les techniques principales sont passées en revue en insistant sur les plus utilisées. Les quatre premières techniques présentées le sont dans un ordre chronologique vis-à-vis de leur utilisation en télédétection. Ainsi, les techniques reposant sur la morphologie mathématique ont fait l'objet de recherches importantes en télédétection à la fin des années 1980. La transformée de Hough a été étudiée en télédétection principalement au début des années 1990. Avec l'avènement de capteurs à haute résolution (puis à

très haute résolution) spatiale les techniques multi résolution ont pris un essor particulier depuis le milieu des années 1990, et continuent à être très étudiées actuellement. Enfin, les techniques utilisant l'intégration de connaissances sont étudiées en télédétection principalement depuis la fin des années 1990 par l'intermédiaire de systèmes experts le plus souvent. La dernière technique présentée est la classification. La classification est utilisée depuis qu'existent les capteurs multi spectraux et est toujours beaucoup utilisée ; souvent en amont des autres techniques.

2.1.4.1 Morphologie mathématique

Le traitement d'image par morphologie mathématique consiste à mettre en relation les pixels d'une image avec ceux d'un *élément structurant* [CC01]. Par exemple la figure 2.2 présente la *dilatation* et l'*érosion* d'une image par un élément structurant 3×3 fait de 1, il s'agit d'un *filtrage*. L'extraction d'éléments curvilignes comme des routes repose alors sur la différence d'intensité qu'il existe *a priori* entre l'élément curviligne et le fond de l'image. Caloz et Collet [CC01] insistent sur l'importance et l'aspect critique du choix de l'élément structurant car il influence de manière très significative le résultat du filtrage. Ainsi, si l'on utilise un élément structurant comme celui présenté sur la figure 2.2, qui est dit à "voisinage étendu" (car on utilise les huit pixels qui cernent le *pixel central*), l'érosion est forte (il reste peu de pixels) et la dilatation est forte (beaucoup de pixels ont été ajoutés). En revanche si l'on utilise comme élément structurant les pixels du "voisinage restreint" (c'est-à-dire les quatre pixels au-dessus, en dessous, à droite et à gauche du pixel central, donc formant la forme "+"), la dilatation et l'érosion sont plus nuancées. Dans le cas de l'extraction d'éléments linéaires on utilise des éléments structurants directionnels que l'on fait tourner.

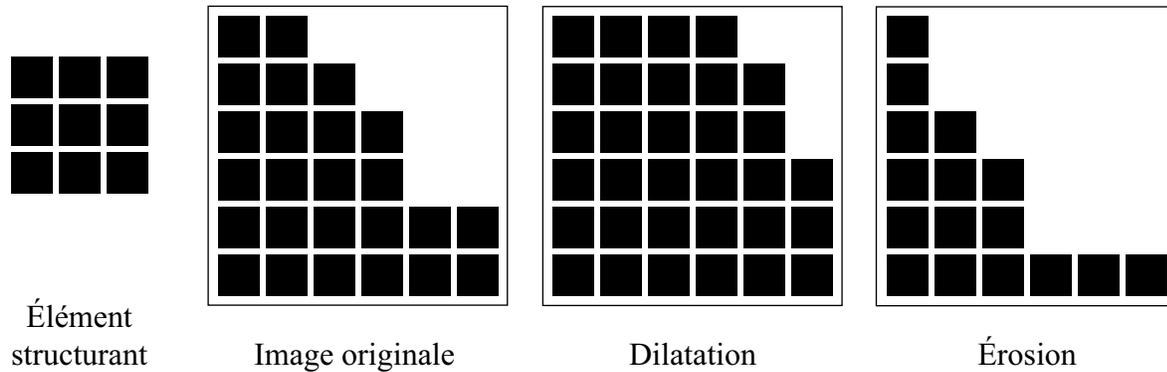


FIG. 2.2 – Opérations élémentaires en morphologie mathématique [Qua04].

L'érosion et la dilatation sont les deux opérations élémentaires de la morphologie mathématique. Les autres opérations sont pour la plupart des opérations de combinaison plus ou moins complexes de ces deux opérateurs. Par exemple, on parle d'*ouverture* pour une érosion suivie d'une dilatation ; dans le cas inverse (dilatation puis érosion), il s'agit d'une *fermeture*.

Une combinaison d'érosions conditionnelles, la squelettisation, est utilisée pour ne garder que les *lignes de force* dans une image [CC01], figure 2.3. Fitzback [Fit99] utilise cette technique pour s'affranchir des points isolés dans l'image. Ansoult *et al.* ont utilisé cette technique combinée à un "élagage" pour mettre en évidence les structures linéaires fermées (figure 2.3), il s'agit de l'algorithme *watershed*⁷ [ASL90]. Les méthodes qui mettent en œuvre des outils de morphologie mathématique, nécessitent pour la plupart un pré-traitement des données coûteux.

⁷ *Watershed* : point tournant, moment critique [Larousse].

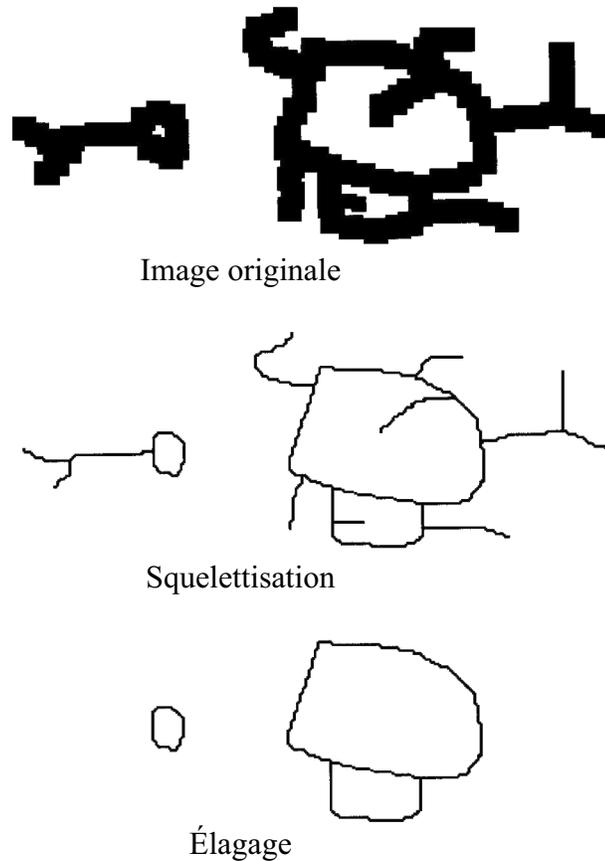


FIG. 2.3 – L'algorithme *watershed* [ASL90], d'après [Qua04]

2.1.4.2 Transformée de Hough

La transformée de Hough est un outil très connu en traitement d'images en général. La transformée de Hough standard permet de passer d'un espace de courbes analytiques à l'espace \mathcal{H} des paramètres de ces courbes. Par exemple si dans l'espace de l'image \mathcal{I} une droite a pour équation $y = ax + b$, alors elle deviendra $\{a, b\}$ dans \mathcal{H} .

Ainsi, toute forme pouvant être paramétrisée devient détectable par la transformée de Hough. S'il s'agit d'objets linéaires, ils correspondront à des points uniques dans \mathcal{H} [Qua04]. Le problème de la transformée de Hough est qu'elle nécessite des images pré-traitées. Il faut en effet que l'image ait d'abord été filtrée pour faire ressortir les

arêtes (de l'image) puis ensuite seuillée en image binaire. La réussite de la détection des éléments curvilignes par la transformée de Hough dépend du traitement initial appliqué à l'image [FC98].

2.1.4.3 Multi-résolution

L'apparence des objets sur les images de télédétection dépend non seulement de leur *signature spectrale* mais aussi de la résolution à laquelle ils sont "vus". L'essor des satellites à très haute résolution spatiale comme QuickBird et Ikonos (voir plus haut) a donné un second souffle aux techniques de multi-résolution dans le domaine de la télédétection⁸. Il faut noter que ces techniques sont utilisées largement en vision numérique [MLT01b].

Sur des images à (très) haute résolution spatiale, les objets curvilignes tels que les routes et les cours d'eau apparaissent comme des *surfaces* allongées avec des signatures spectrales relativement homogènes au sein desdites surfaces. En revanche sur des images à plus basse résolution, les mêmes objets apparaissent comme des *structures curvilignes 1D*.

L'idée (qui peut surprendre à première vue) est de dégrader les images à haute résolution afin de réduire la résolution. Le but est de réduire la résolution jusqu'à ce que les structures linéaires que l'on cherche aient perdu leur épaisseur – elles passent d'un aspect surfacique 2D à un aspect 1D. Les structures linéaires ainsi identifiées sont localisées et servent à guider la recherche sur les images originales à plus haute résolution. La création de ces pyramides d'images est coûteuse en terme de temps de trai-

⁸Ikonos a été lancé en septembre 1999 et Quickbird en octobre 2001

tement. Trois techniques sont principalement utilisées pour créer les images dégradées en télédétection : soit par décimation en ne gardant qu'un pixel sur quatre, soit par un type plus élaboré (et plus coûteux) de décimation qui consiste à ne garder qu'un pixel sur quatre qui est la moyenne des quatre, soit en utilisant une approche par ondelettes. D'autres techniques multi résolution qui ne sont pas seulement propres au contexte de la télédétection font l'objet de la section 2.4.

2.1.4.4 Intégration des connaissances

L'ingénierie des connaissances consiste à formuler dans une forme exploitable des connaissances (d'experts le plus souvent) qui sont communément non exprimées explicitement et/ou sont intuitives [PMS⁺99]. L'intégration⁹ des connaissances sous-entend que les connaissances sont déjà à disposition dans le système pour pouvoir appliquer la méthode. Or il s'avère que l'extraction des connaissances est déjà une étape complexe [Pig01].

L'intégration de connaissances multiformes peut s'avérer une tâche très difficile. Habituellement il s'agit de combiner différents opérateurs locaux sur l'image pour extraire les éléments linéaires en appliquant des contraintes pré-déterminées [Qua04]. Un score est alors attribué aux différentes couches d'informations obtenues à partir des connaissances que le système a synthétisé. Par exemple pour déterminer si un élément curviligne est une route ou un cours d'eau, il est possible de combiner la connaissance que l'on a *a priori* sur leur signatures spectrales¹⁰ et la sinuosité en fonction de la pente¹¹.

⁹En fonction de la littérature on parle soit d'*intégration* des connaissances, soit de *fusion* des connaissances. Il s'agit ici d'un débat de spécialistes et de domaines d'application.

¹⁰Du type : l'intensité des routes est plus forte que celle des cours d'eau.

¹¹Du type : les cours d'eau sont sinueux en plaine et plus rectilignes en zones de relief, à l'inverse des routes.

2.1.4.5 Extraction basée sur la classification

Les satellites utilisant des senseurs multi-spectraux comme Landsat TM, SPOT ou Ikonos permettent de faire des classifications sur les images. En général, la classification est utilisée dans les domaines de l'agriculture, de la foresterie, *etc.*

Pour le domaine de l'extraction d'éléments curvilignes, ces approches sont utiles en tant que pré-traitements de l'image. L'image est segmentée en fonction du couvert, ce qui permet ensuite de mieux diriger l'extraction. [Qua04] mentionne que des techniques de classification ont même été utilisées directement pour l'extraction d'éléments curvilignes sur des données hyper-spectrales, mais que cela était insuffisant ; la classification ne devenant alors qu'une étape de l'extraction.

D'un point de vue général, la classification est un processus le plus souvent supervisé par un expert, qui plus est quand l'objectif final est d'extraire les objets curvilignes. Cette restriction est un frein important à l'automatisation du processus d'extraction. D'après [Qua04], l'intégration de connaissances issues des systèmes d'information géographique devrait permettre de s'affranchir à terme d'une partie de l'intervention humaine.

2.1.4.6 Autres types d'extraction

Appariement de patrons L'appariement de patrons¹² consiste à faire glisser des patrons sur l'image et à calculer un score de similarité par corrélation du patron et de l'image. Les patrons sont en général prédéterminés en termes de formes et d'intensités –

¹²Appariement de patrons : *template matching* – traduction libre.

voir aussi §2.2.5. Habituellement, les patrons dépendent du contexte de l'image, ce qui implique qu'une automatisation totale n'est ici pas possible – sauf si un prétraitement coûteux apportant des informations sur le contexte est appliqué ou bien en utilisant des bases de connaissances.

Autres approches. Enfin, mentionnons quelques approches moins utilisées ou en devenir dans le domaine de l'extraction d'éléments curvilignes d'images de télédétection :

- Programmation dynamique : permet d'augmenter l'optimisation de l'extraction en utilisant une recherche récursive [GL97].
- Permutation de pixels¹³ : s'appuie sur les relations spatiales *et spectrales* entre les pixels [ISA95].
- Importance du contexte : comme mentionné plus haut, en fonction du contexte les objets peuvent avoir des aspects différents [Yee87]. Cela peut permettre soit d'adapter la recherche par rapport à l'objet recherché soit segmenter l'image en fonction du contexte [BSM⁺99].
- Algorithmes génétiques [ISA95].
- Réseaux de *neurones* [LRSO⁺00].

2.1.5 Bilan

La plupart des techniques présentées plus haut fonctionnent dans un (ou quelques) contexte(s) particulier(s) mais pas dans le cas général. En outre, d'après Quackenbush [Qua04], la littérature rapporte souvent des évaluations de techniques en termes qualitatifs (visuelles) mais rarement en termes quantitatifs. Quand cela est tout de même

¹³Permutation de pixel : *pixel swapping* – traduction libre.

fait, la méthodologie d'évaluation est suffisamment vague pour que le lecteur ne puisse apprécier réellement la performance de la méthode.

2.2 Approche de Marr

Dans le domaine de la vision numérique, l'un des ouvrages de référence des plus importants est celui de Marr [Mar82]. Il pose les fondements que doit avoir toute approche de vision numérique pour être cohérente – ce travail concerne la stéréoscopie et la représentation d'objets 3D. Nous verrons que la technique en trois étapes que Marr présente, en partant d'images numériques (donc 2D), est très pertinente pour le travail en deux dimensions également ; en particulier la première étape sur laquelle nous insisterons. De plus, ce livre propose également une approche dont la plupart des travaux ultérieurs ont découlé. Le travail de doctorat présenté ici s'appuie sur des techniques inspirées des visions naturelle et artificielle puisque nous cherchons à tirer parti des stratégies de la vision humaine comme évoqué dans le chapitre 1. Il convient donc de faire une recherche bibliographique sur la vision numérique d'une part, et, sur la vision naturelle d'autre part. La vision naturelle est abordée du point de vue des mécanismes attentionnels dans la section suivante. Dans la présente section, nous cherchons à dégager le cheminement historique et les aspects actuels de la vision numérique qui a été guidée par la vision naturelle. En particulier, nous voyons comment la perception visuelle¹⁴ a été le moteur de l'approche de Marr. Cette manière de nourrir l'approche numérique avec des aspects de vision naturelle s'inscrit dans la droite ligne de l'objectif de la thèse. Ainsi, aussi bien les aspects techniques que la philosophie de

¹⁴Les mécanismes attentionnels, qui sont une partie très importante de la vision naturelle, ne sont abordés qu'après cette section consacrée à l'approche de Marr car ses travaux ont été surtout guidés par la perception visuelle plutôt que par les mécanismes attentionnels. À l'époque des travaux de Marr, les mécanismes attentionnels étaient beaucoup moins étudiés et connus qu'ils ne le sont maintenant.

la démarche de Marr sont pertinents dans le cadre de ce travail.

D'un point de vue méthodologique, Marr explique qu'il est impossible de considérer la vision selon un seul aspect. On ne peut pas, par exemple, considérer le processus de traitement de l'information (c'est-à-dire appréhender ce qui existe à partir d'une image) sans prendre en compte la représentation de cette information dans le cerveau. Marr propose, pour comprendre les systèmes complexes de traitement de l'information, de décomposer le problème en trois niveaux. D'abord il existe un niveau abstrait [Mar76](le quoi et le pourquoi), qui est la maîtrise de la théorie et son application au cas. Ensuite vient le niveau de la représentation et de l'algorithme. Enfin, le dernier niveau est le niveau physique dans lequel on réalise physiquement l'algorithme, réalisation contrainte par le matériel.

Depuis longtemps les chercheurs ont essayé de générer une représentation en trois dimensions à partir d'une image. Il n'est pas prouvé que ce soit l'approche adoptée par l'être humain. Deux approches ont été principalement utilisées pour l'application de techniques en vision numérique [BG90] : l'approche *descendante*¹⁵ et l'approche *ascendante*¹⁶. L'approche descendante est une approche principalement dirigée par les concepts. C'est plutôt la connaissance *a priori* qui va guider le processus. De l'autre côté, l'approche ascendante est guidée par les données. Marr propose dans sa théorie de la vision un découpage en trois étapes. Bien que cette théorie soit ascendante en grande partie, elle présente également une composante de type descendante comme nous allons le voir. Les trois étapes envisagées par Marr sont les suivantes :

- l'ébauche primitive ;
- l'ébauche 2,5D ;

¹⁵Bien que la terminologie *descendante* soit française on retrouve également le terme anglo-saxon *top-down* dans des documents de langue française.

¹⁶Approche ascendante ou *bottom-up*.

- la représentation du modèle 3D.

C'est à partir de cette architecture de base que la plupart des travaux ultérieurs ont été menés. Dans tout traitement visuel on considérera que les données d'entrée¹⁷ sont l'image.

2.2.1 Ébauche primitive

Évidemment dans cette théorie la question sur laquelle nous allons le plus nous pencher est l'ébauche primitive. Le but de l'ébauche primitive est de mettre en exergue les variations d'intensité au sein de l'image et de dégager une organisation générale des caractéristiques de l'image. Cette ébauche primitive peut se scinder en deux parties : *l'ébauche primitive brute* et *l'ébauche primitive complète*. L'ébauche primitive brute est réalisée suivant des approches globales alors que l'ébauche primitive complète s'appuie sur des aspects perceptifs.

2.2.2 Ébauche primitive brute

L'ébauche primitive brute doit réaliser le repérage des variations d'intensité dans l'image. Ces variations sont souvent interprétées comme étant des *arêtes*. Ainsi, de très bons résultats ont été obtenus en utilisant un *filtre* de forme spéciale qui détectait les "vallées de luminance" sur l'image d'un visage [PR85]. Dans ce type d'approche l'échelle des changements¹⁸ présents (et visés) à l'intérieur de l'image est la même (elle

¹⁷*Input.*

¹⁸Par échelle des changements, on entend *variations des fréquences spatiales*.

ne varie pas); le cas des visages est un cas particulier¹⁹. Le problème qui se pose est que les variations d'intensité ne sont pas obligatoirement des arêtes car les échelles de changement à l'intérieur de l'image sont le plus souvent différentes – c'est-à-dire variables d'un endroit à l'autre.

Sans être exhaustif, citons deux approches qui prennent en compte cet aspect multi-échelle : l'approche de Marr et Hildreth [MH80] et l'approche de Watt [Wat87] [Wat88]. Ces deux approches utilisent la même technique de base, c'est-à-dire la combinaison des résultats de filtres de différentes tailles appliqués à l'ensemble de l'image, c'est pourquoi on parle d'*approche globale*. Les résultats de ces filtres fournissent, en fonction de l'échelle, les zones où les variations d'intensité sont les plus importantes.

Dans l'approche proposée par Marr [MH80], on essaye de faire un appariement des résultats obtenus avec des filtres de tailles différentes, voir figures 2.4d et 2.4e. Suivant l'appariement, quatre types de combinaisons sont dégagés : les taches (figure 2.4f), les segments de bordure (figure 2.4g), les barres (figure 2.4h) et les terminaisons – discontinuités.

Dans l'approche de Watt [Wat88], on applique un algorithme qui exploite la sommation des sorties des filtres aux différentes échelles en séparant leurs parties positives et leurs parties négatives. Cet algorithme répond bien au groupement et à la détection, mais ne permet pas d'avoir de positions précises car la structure même de l'algorithme permet une mémorisation de la topologie des éléments au sein de l'image mais pas leur taille ou leur position [Wat87]. Selon Bruce et Green [BG90], la finalité de l'ébauche primitive brute est de faire apparaître la représentation des arêtes.

¹⁹En effet un visage peut se décomposer selon plusieurs échelles; par exemple une petite échelle qui est la forme du visage, une échelle moyenne qui correspond aux traits du visage et à une grande échelle, les plis et rides.

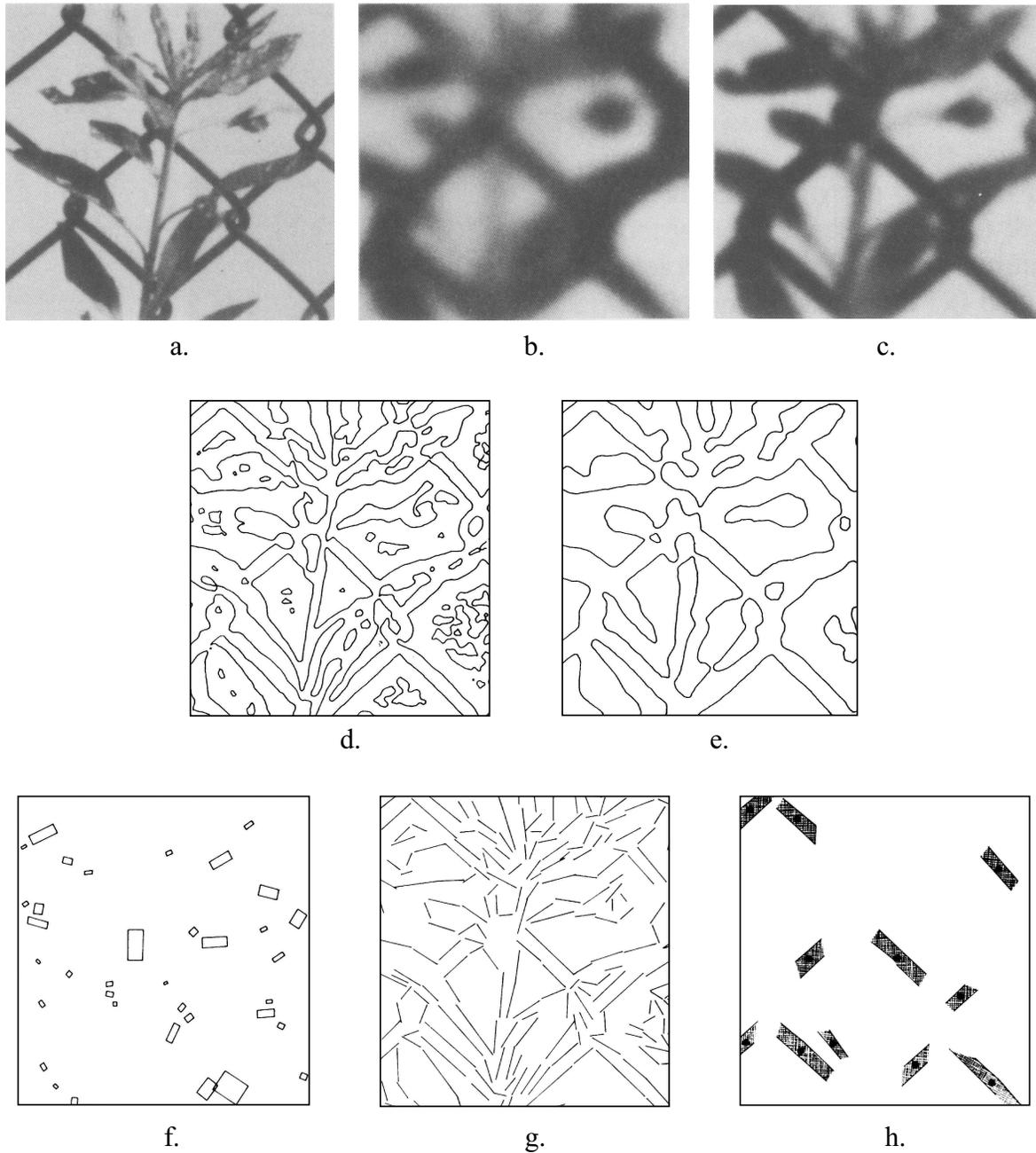


FIG. 2.4 – Ébauche primitive brute d'après [MH80]. a. Image de départ. b. et c. Image de gauche rendue floue à l'aide de filtres gaussiens de deux tailles différentes pour mettre en évidence deux domaines fréquences spatiales différentes. d. et e. sont les passages à zéros du laplacien (dérivées secondes) des images b. et c. La combinaison de d. et e. fournit l'ébauche primitive brute; ici f. les taches, g. les segments de bordure et h. les barres.

2.2.3 Ébauche primitive complète

Lorsque l'ébauche primitive brute est obtenue, on réalise l'ébauche primitive complète qui permet de dégager des structures plus larges (formes et régions) par regroupement [Mar82]. Ces regroupements sont fonction de *l'organisation perceptive*. De nombreux travaux sont effectués dans ce domaine. L'une des théories les plus importantes est le principe de la Gestalt. Les lois gestaltistes de l'organisation qui sont encore considérées à l'heure actuelle sont :

- le principe de proximité, figure 2.5 ;
- le principe de similitude, figure 2.6 ;
- le principe de destin commun, surtout conditionné par le champ du mouvement ;
- le principe de bonne continuation, figure 2.7 ;
- le principe de fermeture, figure 2.8 ;
- le principe d'orientation, d'environnement (figure 2.10), de symétrie (figure 2.11) et de taille relative – figure 2.9.

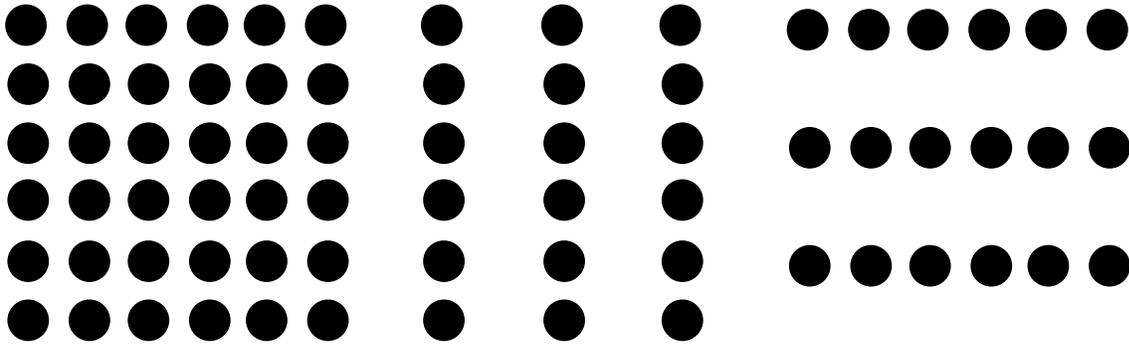


FIG. 2.5 – Théorie de la Gestalt : regroupement par proximité. Suivant la proximité des disques entre eux nous pouvons percevoir des colonnes ou des lignes.

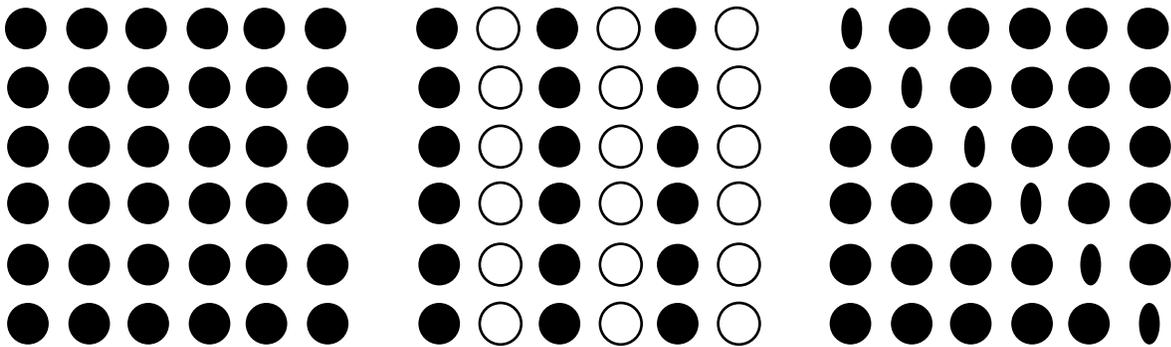


FIG. 2.6 – Théorie de la Gestalt : regroupement par similitude. La similarité nous fait percevoir des colonnes au centre et une diagonale à droite; la similarité l'emporte sur la proximité.



FIG. 2.7 – Théorie de la Gestalt : regroupement par bonne continuation. a. Deux courbes semblent se croiser. b. En réalité il peut s'agir de deux lignes brisées.

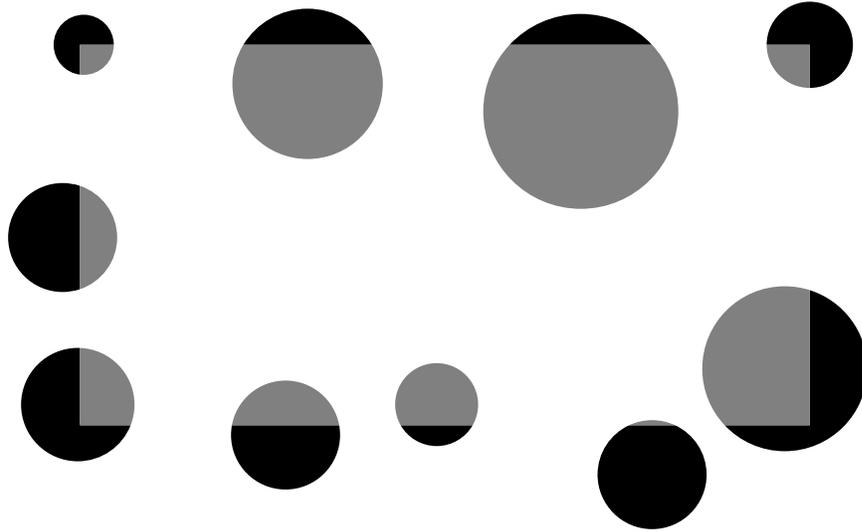


FIG. 2.8 – Théorie de la Gestalt : regroupement par fermeture ; bien qu'il n'existe que des disques bicolores on perçoit un rectangle gris. Certains gestaltistes [BG90] ont suggéré que les sept étoiles principales de la constellation de la Grande Ourse, pouvaient être perçues comme une casserole ou un chariot, grâce aux principes de fermeture et de bonne continuation.

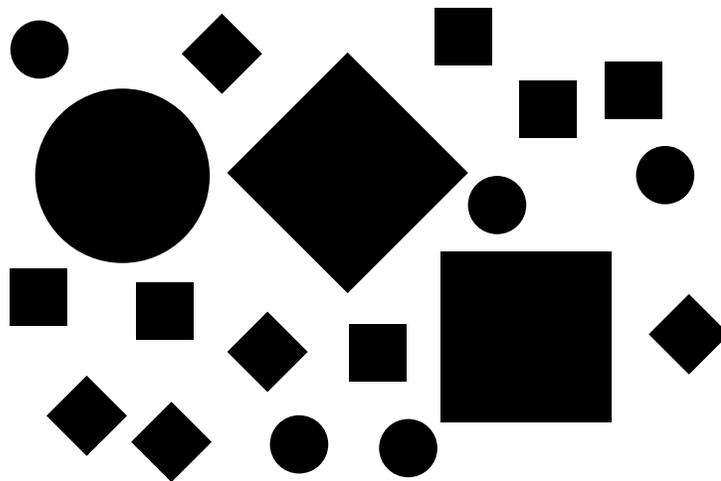


FIG. 2.9 – Théorie de la Gestalt : regroupement par taille relative.

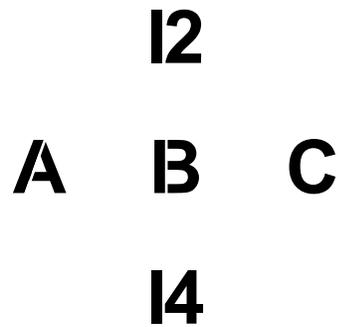


FIG. 2.10 – Théorie de la Gestalt : regroupement par environnement. Suivant que l'on lise de gauche à droite ou de haut en bas, l'élément au centre de la figure est perçu comme "B" ou "13".

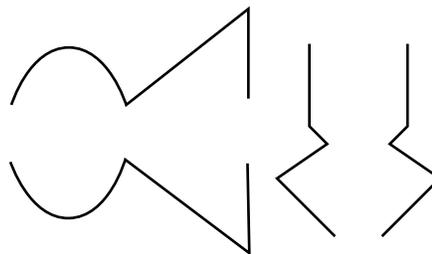


FIG. 2.11 – Théorie de la Gestalt : regroupement par symétrie. Les éléments de la figure sont regroupés deux par deux par symétrie. Ici on voit que la symétrie l'emporte sur la proximité.

Marr propose de faire un regroupement récursif des premiers résultats qui ont été dégagés dans l'ébauche primitive brute. L'organisation obéit alors à trois règles proches de la théorie de la Gestalt, figure 2.12 :

- le *regroupement* est un groupement selon la proximité ;
- le *rassemblement curvilinéaire* est un groupement en une structure curviligne ;
- le *rassemblement thêta* est un groupement selon la tendance d'objets similaires comme présentés sur la figure.

De ce point de vue, le regroupement obéit à de grands concepts généraux et l'approche a une composante descendante. Mais *sensu stricto* on ne se sert pas de connaissances *a priori*. Autrement dit, dans son approche Marr ne "s'attend pas" à trouver un type d'objet dans l'image, mais en revanche il utilise des mécanismes de regroupements qui reflètent l'organisation générale de la scène visuelle. Cette organisation ainsi dégagée permet de guider la perception des objets dans l'image.

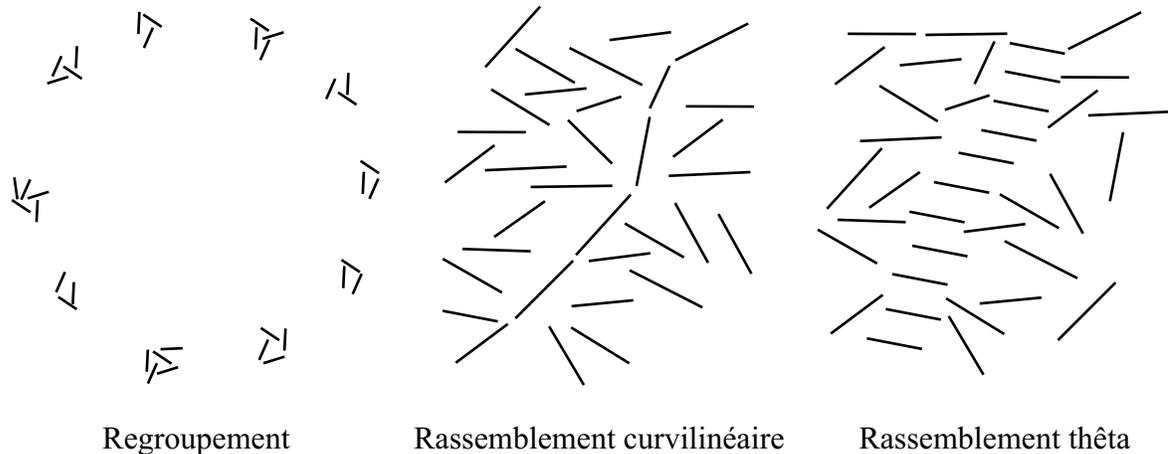


FIG. 2.12 – Les trois règles de Marr.

2.2.4 Au-delà de Marr

Enfin, au-delà de l'approche de Marr, il existe de nouvelles approches qui se situent dans la même philosophie d'utilisation de l'organisation perceptive à des fins d'extraction d'informations d'images²⁰. Par exemple, très récemment, les premiers travaux de Zlatoff *et al.* [ZTB04c] et [ZTB04a] tendent à mêler les connaissances *a priori* concernant un domaine d'étude particulier à la théorie de la Gestalt.

Ces travaux portent sur l'interprétation d'image et en particulier ils cherchent à mettre en relation la **segmentation** de l'image avec des objets sémantiques dans le but de fournir une méthode d'indexation sémantique automatique d'images. Dans un premier temps l'image est segmentée grâce aux connaissances *a priori* du domaine concerné [ZTB04b]. L'exemple pris est le domaine de l'architecture historique (une stèle) : segmentation en couronnement, corps, inscriptions, *etc.* À partir de la segmentation, l'image est décrite topologiquement. Via des techniques empruntées à la théorie de la Gestalt et en s'appuyant sur une **ontologie** appropriée (développement d'objets sémantiques élémentaires), Zlatoff *et al.* proposent une description sémantique multi niveau de l'image. Les auteurs pensent pouvoir ainsi fournir à terme une indexation automatique d'images aériennes.

2.2.5 Notions complémentaires : reconnaissance des formes

La reconnaissance des formes est un aspect très important et incontournable de la vision numérique. Cette notion est abordée rapidement ici afin de montrer quelle est sa

²⁰D'autres approches comme [AM96] par exemple, prenant également en compte le groupement perceptuel, sont abordées plus loin car elles concernent plus spécialement la modélisation de l'**attention visuelle** pour des contextes particuliers et à des fins spécifiques — voir §2.3.3.2, page 63.

portée dans le cadre de ce doctorat. D'une manière générale ces approches se heurtent à la variabilité des objets et des images qui sont visés dans ce travail. En revanche nous verrons que la psychologie apporte des clés intéressantes dans le domaine de la reconnaissance des formes.

2.2.5.1 Reconnaissances particulières

Si l'on voulait placer la reconnaissance des objets dans l'approche de Marr, celle-ci interviendrait suite aux étapes d'ébauches primitives – brute et complète. De nombreuses approches ont été explorées, et elles peuvent donner de bons résultats mais toujours pour des applications précises. Si l'on est intéressé à avoir une solution générale, le problème n'est toujours pas résolu.

Un des domaines où les mécanismes de reconnaissance de formes fonctionnent convenablement est la reconnaissance de caractères. De nombreux algorithmes ont ainsi pu être testés mais hélas ils n'ont pas fait leurs preuves dans d'autres domaines. D'une manière générale ce sont des approches algorithmiques sans lien proche avec une démarche cognitive. Leurs principes de fonctionnement peuvent être très variés.



FIG. 2.13 – Erreur de reconnaissance de caractère. À gauche un A et un R, au milieu le gabarit d'un A. À droite le gabarit du A a tendance à s'apparier plutôt avec le R qu'avec le A.

Une première approche est d'*appairer l'objet à un gabarit*. Mais un gabarit peut s'apparier à plusieurs objets, par exemple un A et un R dans le cas de lettres comme pour la figure 2.13. Une seconde approche est *l'analyse par traits*. Selon le nombre de

traits et leurs caractéristiques (horizontales ou verticales) dans les figures analysées, des démons²¹ (qui correspondent aux différentes formes possibles) vont réagir de manière plus ou moins forte, selon leur correspondance avec l'objet, figure 2.14. Si l'on reste toujours dans le domaine des caractères on voit que la limite de ce système est atteinte lorsque l'on présente des objets qui ne sont pas des caractères et donc pour lesquels il n'existe pas de démon. Une forme quelconque peut alors être prise pour une lettre.

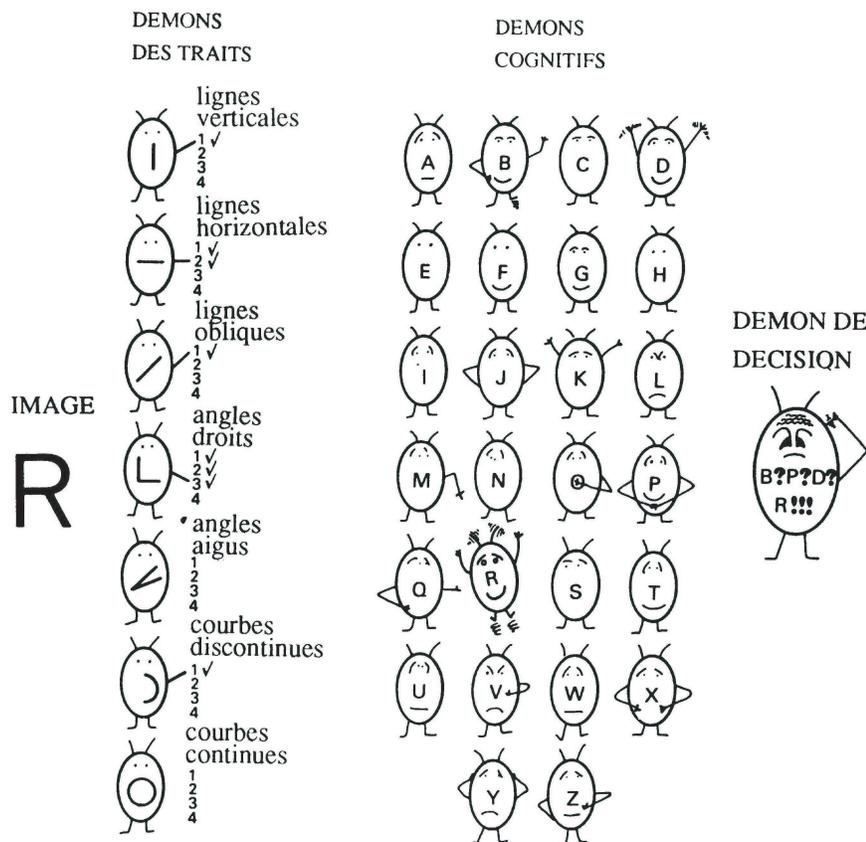


FIG. 2.14 – Principe de l'analyse par traits, d'après [LN72] et [Sel59] adapté par [BG90].

Décrire structurellement les objets peut être un autre mécanisme de reconnaissance. Cette approche diffère nettement de l'analyse par traits car elle décrit les *relations structurelles* qu'il y a entre les éléments qui composent une lettre – contrairement à l'analyse par traits qui a un nombre fini de démons descripteurs. Par exemple un T (T)

²¹Les démons sont de petits programmes ou parties de programme qui travaillent en arrière plan ; c'est-à-dire qu'ils sont actifs en permanence et peuvent réagir dès qu'ils sont sollicités.

est descriptible comme :

- deux lignes ;
- une horizontale et une verticale ;
- la ligne verticale est *sous* la ligne horizontale ;
- la ligne verticale touche la ligne horizontale en son milieu *sans la couper*.

Mais dans cette approche le pré-traitement est très difficile à mettre en œuvre avant de pouvoir faire la reconnaissance à proprement parler [BG90].

2.2.5.2 Approche de Marr et Nishihara

Marr et Nishihara [MN78] ont proposé de découper les objets à différentes échelles en “cônes généralisés”. Ce découpage se fait grâce aux contours. À partir de ce découpage on consulte ce que l’on a en mémoire dans le système. D’autres approches sont des variantes, la différence résidant dans la manière dont se fait le découpage – en “super-quadratiques”, en “géons”. Mais ce type de découpage des objets n’est pas applicable pour de nombreux objets comme les visages et les nuages car ils ne peuvent, par définition, se découper en objets prédéterminés, leurs formes n’étant pas divisibles, ou simplement trop variées pour être gardées en mémoire par le système ; comme dans le cas des objets curvilignes auxquels nous nous attaquons dans ce travail. Cela est dû non seulement à la forme intrinsèque des objets, mais également à leur nature sur l’image : formes, intensités, nature (ici dans le sens de route, cours d’eau, *etc*), qualité de l’image. . .

2.2.5.3 Perception des formes

Enfin pour compléter cette section sur la reconnaissance des formes, mentionnons que des recherches en psychologie apportent de nouvelles pistes quant à la manière dont se fait la reconnaissance. Ainsi Elder et Zucker [EZ98] ont montré que la perception des surfaces dépend fortement de la nature des contours qui les bordent. Depuis longtemps il était supposé que les propriétés régionales et les contours agissaient de façon conjointe pour donner le groupement perceptuel des formes. Or Elder et Zucker ont mis en évidence que la fermeture du contour d'un objet permet sa reconnaissance beaucoup plus rapidement. De plus, l'ajout d'une texture orientée n'influe pas sur la détection, sauf lorsque cette texture peut être confondue avec une frontière (un contour). La texture n'apporte jamais autant que la fermeture (des contours de l'objet). Le système de vision humain semble utiliser surtout les irrégularités des contours pour le groupement perceptuel, les propriétés régionales n'étant pas exploitées de la même façon. La "quantité" d'attention mobilisée est plus élevée pour les figures ouvertes. Les contours sont par conséquent un bien meilleur discriminant que les textures pour déterminer des formes.

2.3 Mécanismes attentionnels

2.3.1 Généralités

Dans le cadre de ce travail de thèse, un des objectifs annoncés (chapitre 1) est de pouvoir guider la recherche d'information dans l'image via un regard virtuel. Ainsi une approche inspirée de l'être humain est recherchée, forte du constat que les sujets

sont capables d'extraire l'information de scènes complexes grâce, entre autres, à des mécanismes attentionnels.

L'étude de ces mécanismes attentionnels doit permettre de proposer de nouvelles approches pour aider à localiser l'information pertinente dans l'image. Dans le cas particulier de cette thèse, nous souhaitons regarder la pertinence de telles approches pour pallier la complexité de la recherche visuelle dans les images de télédétection.

2.3.2 Psycho-physiologie de l'attention visuelle

L'attention en psychologie, selon [CCL97], semble se réaliser selon deux modes : le mode pré-attentionnel et le mode attentionnel. Dans le mode pré-attentionnel, l'information est traitée de manière parallèle sur l'espace de vision, il ne dépend pas du nombre d'objets dans le champ de vision : qui est très large²². Le mode attentionnel lui, est un traitement réalisé en série sur une partie beaucoup plus restreinte de l'image.

Les théories pré-attentionnelles peuvent se scinder en deux groupes. D'un côté on trouve les premiers modèles de vision comme celui de Marr et, de l'autre, les théories d'interruption, guidées par les données (ascendantes), qui traitent de la détection des différences de signal provenant d'une discontinuité de l'espace. Les discontinuités attirent naturellement l'attention sur des zones de l'espace différentes sans pour autant, à cette étape, permettre une identification de l'objet.

Les théories attentionnelles se répartissent selon les deux types d'attention suivants : l'attention sélective et l'attention spatiale. L'attention sélective est guidée par le modèle

²²Comme son nom l'indique, le mode *pré*-attentionnel ne nécessite pas de ressources attentionnelles *sensu stricto* d'après [CCL97].

qu'a le sujet, la cible étant connue ; par exemple trouver un objet d'une couleur ou d'une forme par rapport à d'autres de couleurs ou d'autres formes différentes [FG03]. L'attention spatiale, elle, est guidée par les données (reçues par l'œil²³), cela peut-être volontaire ou involontaire. Dans le cas involontaire, ce peut être par exemple, le clignotement d'une lumière qui "attire l'œil" du sujet.

2.3.2.1 Sélection attentionnelle

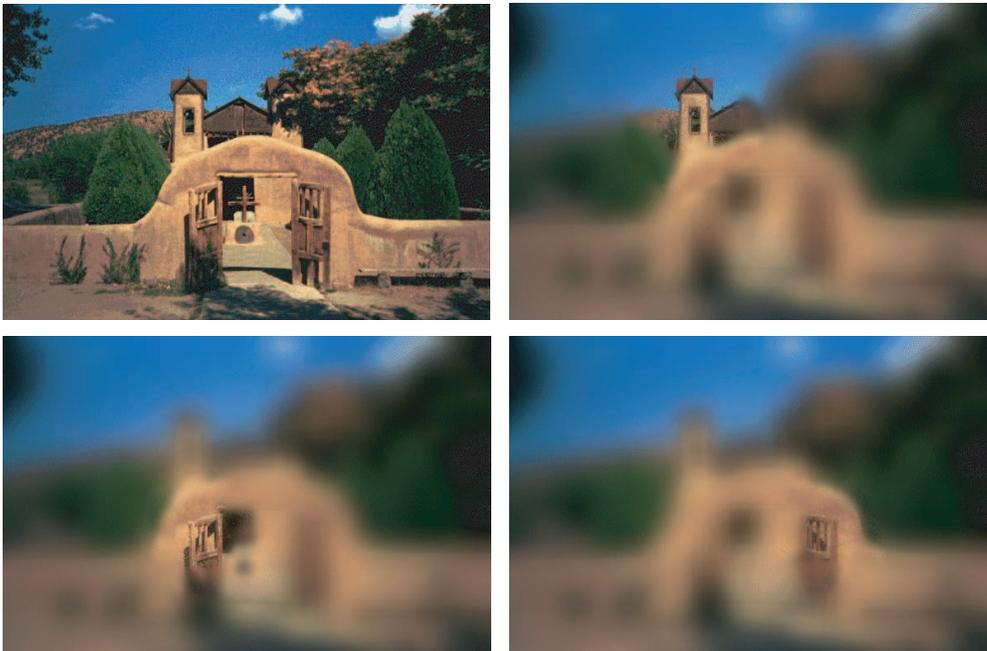


FIG. 2.15 – Principe de la vision naturelle. La première figure montre une scène visuelle ; les autres figures montrent ce que perçoit réellement l'œil d'un observateur lors d'une saccade, d'après [Mac02]

Afin de proposer une approche de la vision basée sur les mécanismes attentionnels il faut essayer d'appréhender l'attention humaine. Machrouh [Mac02] montre ce que perçoit réellement l'œil d'un observateur (figure 2.15) situé à 50 centimètres d'une image de format 13×18 cm. L'impression de voir nettement toute l'image vient de

²³Les données reçues par l'œil sont les caractéristiques physiques des stimuli.

l'exploration incessante de l'image couplée à une *mémoire à court terme* qui supplée le manque d'information perceptive directe. La vision naturelle ne traite qu'une partie infime d'une scène visuelle. Cette petite partie de la rétine est appelée région fovéale. La région périphérique ne traite que l'information basse fréquence qui permet d'orienter la *fovéa* vers les parties intéressantes de la scène visuelle.

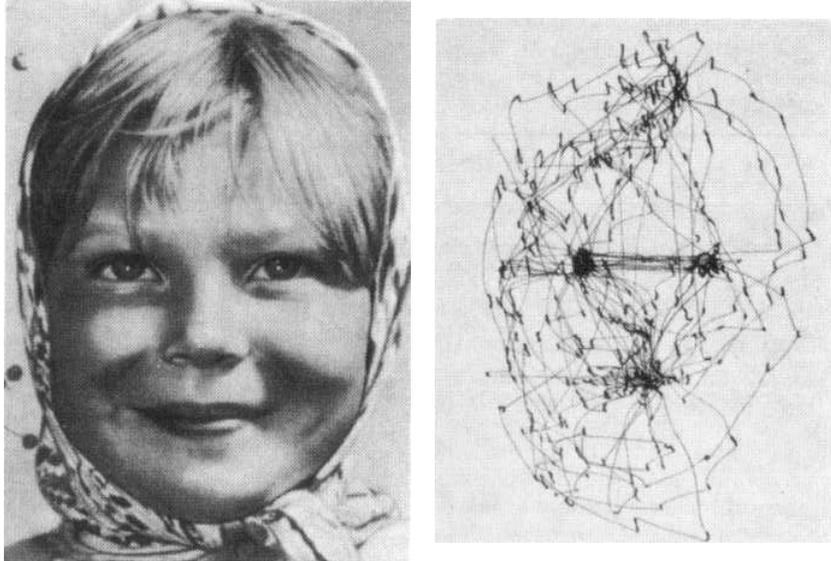


FIG. 2.16 – Mode d'exploration oculaire d'une image. Dans le cas de portraits ce sont les yeux et le nez qui sont particulièrement visés, d'après [Yar67]. Toutes les figures des travaux de Yarbus (figures 1.1, 2.16 et 2.17) présentées dans cette thèse ont été obtenues de la même manière. Yarbus utilisait une "ventouse" (membrane) qui s'ajustait au globe oculaire. Sur cette ventouse était fixé un petit miroir. En éclairant le miroir d'un rayon lumineux il était alors possible de suivre par réflexion les mouvements de l'œil en les enregistrant de manière photographique.

Le déplacement du regard ne se fait pas de manière aléatoire sur l'image mais suit les traits saillants (figure 2.16), privilégiant certains aspects de la scène ou de la figure [Gre00]. Le regard reste en général environ un quart de seconde sur un point, la scrutation à proprement parler ne durant que 5 à 15 ms.

Ainsi Yarbus [Yar67] a montré que la couleur avait peu d'influence sur l'exploration d'une image que fait un sujet. Il a montré également que le mode d'exploration de



a.



b.



c.



d.



e.



f.



g.

FIG. 2.17 – Mode d’exploration oculaire d’une image en fonction de la tâche, d’après [Yar67]. Les sujets ont eu différentes instructions en regardant la toile de Ilya Repin : “*Ils ne s’attendaient pas à le voir*”. a. Examen libre du tableau. b. Instruction d’évaluer les biens matériels de la famille. c. Instruction de donner l’âge des personnages d. Instruction de présumer de ce que faisait la famille avant l’arrivée du visiteur. e. Instruction de mémoriser la tenue vestimentaire des personnages de la toile. f. Instruction de mémoriser la position des objets et des personnages dans la pièce. g. Instruction d’estimer depuis combien de temps le visiteur inattendu était parti.

l'image était très influencé par la tâche mentale spécifique des sujets. Il a réalisé cela en comparant le parcours des yeux sur des toiles pour des sujets auxquels il avait demandé des tâches spécifiques, figure 2.17²⁴. Au regard de ces résultats (du type de la figure 2.17), Yarbus a émis l'hypothèse qu'un premier balayage était réalisé pour obtenir une impression générale, mais que "*le temps supplémentaire pour bâtir la perception n'était pas utilisé pour examiner les éléments secondaires mais pour revenir sur les éléments les plus importants*" [Yar67], page 193.

Modèle d'intégration des traits de Treisman et Gelade ²⁵

Le modèle d'intégration des traits proposé par Treisman et Gelade [TG80] est le modèle sur lequel toutes les théories sur le rapport entre les tâches parallèles²⁶ et les tâches sérielles²⁷ sont basées. De plus Treisman et Gelade ont distingué les traitements *pré-attentionnels* des traitements *attentionnels*. D'une manière générale tous les modèles actuels en psychologie cognitive de l'attention visuelle sont des raffinements du modèle de Treisman et Gelade.

Le modèle d'intégration des traits propose que l'attention soit dirigée de manière sérielle sur chaque stimulus d'une scène visuelle lorsque la conjonction de plus d'un trait (caractéristique) atomique (séparable) est nécessaire pour caractériser la cible.

Treisman et Gelade ont montré que les temps de réaction augmentaient de façon

²⁴En général ce sont des toiles de l'école réaliste Russe ; ces travaux ayant été réalisés durant la guerre froide. Bien que classiques, les travaux de Yarbus n'ont réellement été connus à l'Ouest qu'après la traduction de Lorrin Riggs [Yar67] en 1967, à cause du contexte politique de l'époque – faits rapportés par [FG03].

²⁵*Feature Integration Theory of Attention* de Treisman et Gelade [TG80] est traduit le plus souvent dans la littérature francophone [RBC⁺98] par "Modèle d'intégration des traits de Treisman" mais l'on peut trouver de nombreuses autres appellations plus ou moins ésotériques telle "recherche conjonctive de Treisman et Gelade" [Cou99], *etc.* Il est à noter que ce modèle est tellement connu qu'il est souvent simplement appelé *FIT* dans la littérature anglo-saxonne.

²⁶La recherche visuelle effectuée selon une tâche parallèle est souvent appelée "recherche de traits" (*feature search*) [FG03].

²⁷La recherche visuelle effectuée selon une tâche sérielle est souvent appelée "recherche conjonctive" (*conjunction search*) [FG03] car la recherche se fait sur la conjonction des caractéristiques de la scène visuelle.

drastique lorsque la cible recherchée est faite d'une conjonction de caractéristiques. Or cela s'oppose à un système attentionnel qui utiliserait un flux constant d'attention comme le propose Tsal [Tsa83]. En effet, un tel système ne pourrait traiter une tâche de reconnaissance aussi rapidement que le montrent les temps de réaction des sujets de Treisman et Gelade pour des cibles n'étant caractérisées que par un trait unique. Il faut par conséquent considérer un système pré-attentionnel dédié à des tâches spécifiques.

Le modèle d'intégration des traits a été étudié en utilisant le *paradigme de recherche visuelle*²⁸, proposé également par Treisman et Gelade [TG80], qui permet de définir une cible aussi bien par ses caractéristiques dissociées que par leur conjonction. Les caractéristiques simples peuvent être détectées en parallèle sans limite d'attention. La recherche de cibles simples avec peu de caractéristiques n'est que très peu affectée par le nombre de distracteurs dans la scène visuelle ; le temps de réaction sera petit. À l'opposé, une attention focalisée est nécessaire lorsqu'une cible est caractérisée par une conjonction de traits. Ce type de cible ne peut être trouvé qu'après un balayage sériel d'un nombre variable de distracteurs, ce qui augmente d'autant le temps de réaction.

Amnésie de la recherche visuelle ?

La recherche visuelle ne possède pas de mémoire selon Horowitz et Wolfe [HW98]. Ils ont en effet montré que contrairement à l'idée classiquement reçue auparavant, la recherche visuelle n'a pas de mémoire. Dans leur expérience Horowitz et Wolfe changent de manière aléatoire l'arrangement des distracteurs et de la cible toutes les 111 millisecondes ; pourtant la vitesse de recherche de la cible reste la même. Le modèle proposé par Horowitz et Wolfe est le suivant : le système de recherche visuelle génère un classement de priorité en fonction de la *saillance* (cf. 2.3.3.7 page 69) de chaque objet. Par conséquent la recherche amnésique ne garde pas

²⁸En anglais : *Visual Search Paradigm*

de trace des items visités et n'est donc pas affectée par la redistribution aléatoire des stimuli. Les résultats qui sont donnés par Horowitz et Wolfe montrent que le système visuel n'accumule pas d'information sur l'identité des objets pendant le temps que dure la recherche. Le système de recherche, au contraire, semble être dans une sorte de "présent éternel".

À cet article d'Horowitz et Wolfe, paru dans *Nature* et intitulé *Visual Search Has No Memory*²⁹ en 1998, Peterson *et al.* ont répondu en 2001 par un article intitulé *Visual Search Has Memory*³⁰ [PKFW⁺01]. Cette réponse montre combien le sujet est encore dans une phase exploratoire. Dans cet article Peterson *et al.* montrent que les sujets réexaminent bien les éléments (items) durant la tâche de recherche visuelle, mais le patron de revisitation ne semble pas en concordance avec un modèle de recherche visuelle sans mémoire. De plus, une large proportion des **fixations** est dirigée sur la cible, ceci suggérant que les revisitations seraient dues à un souvenir pour lequel les items n'ont pas été adéquatement identifiés. Les auteurs n'ont pas trouvé de preuves qui suggéraient que les fixations soient guidées par un "balayage discret"³¹ amnésique qui explorerait l'environnement à la recherche de nouveaux items au cours des fixations.

Les résultats de Peterson *et al.* ne corroborent clairement pas le modèle de Horowitz et Wolfe. La distribution des revisitations ne correspond pas d'une part, et, d'autre part, une large proportion (environ 35%) de ces revisitations sont faites sur la cible. Ceci ne serait donc pas dû à un oubli des items déjà visités mais plutôt à un retour des sujets sur des items inadéquatement traités lors de la première visite. Les résultats suggèrent même, selon Peterson *et al.*, que la recherche pourrait retenir jusqu'à 12 items. Le problème à partir de ces résultats est de savoir pourquoi ils apportent des conclusions aussi différentes de celles de Horowitz et

²⁹La recherche visuelle n'a pas de mémoire

³⁰La recherche visuelle a une mémoire

³¹*Covert scan*

Wolfe. Les auteurs avancent plusieurs explications possibles. Par exemple, selon Peterson *et al.*, les scènes visuelles qu'ils utilisent seraient écologiquement plus valides que celles utilisées par Horowitz et Wolfe. Une autre explication pourrait être que les sujets, lors de l'expérience de Horowitz et Wolfe, auraient été capables d'accumuler des preuves de la présence de la cible sur l'ensemble de la scène et ceci de manière parallèle.

L'amnésie ou la mémoire de la recherche visuelle est encore un point qui suscite de nombreuses recherches et pour laquelle les spécialistes ne s'accordent pas.

2.3.2.2 Inhibition de retour

La sélection attentionnelle guide le regard où la zone attentionnelle est la plus forte, c'est-à-dire le lieu le plus saillant de la scène. Afin de ne pas revenir toujours aux mêmes zones de la scène, l'être humain utilise un marqueur inhibant. Un tel marqueur des lieux "visités" a été très largement observé en psychologie humaine, ce phénomène s'appelle *l'inhibition de retour*.

Klein, dans son article sur l'inhibition de retour [Kle00], explique le phénomène de la façon suivante.

L'orientation est l'une des fonctions premières des êtres vivants. L'efficacité et la rapidité d'un réflexe qui contrôle l'orientation est primordiale pour la prédation et la défense ; le contrôle volontaire, par delà les réflexes d'orientation est indubitablement un développement important. Une scrutation efficace implique également d'utiliser de l'information contenue dans la mémoire qui prend en compte les comportements d'orientation antérieurs. La cause de l'inhibition de retour serait l'orientation de l'attention

vers une zone, et la réalisation du retrait subséquent de l'attention de cette zone. Des preuves convergentes suggèrent que l'inhibition de retour retarde à la fois les réponses motrices et le retour de l'attention. En revanche, des études sur le jugement temporel de l'ordre («*Laquelle des deux cibles a été présentée en premier ?*»), confirment que l'inhibition de retour n'affecte pas la vitesse avec laquelle le processus perceptuo-sensitif extrait l'information à partir d'une *zone indiquée* – déjà visitée.

L'inhibition de retour, d'après Klein et Taylor, est plus associée à la réponse du système qu'à l'attention [KT94], [DC94]. Quoiqu'il en soit, des preuves convergentes sur le fait que l'inhibition de retour entrave le retour de l'attention sur les zones déjà visitées ont été mises en évidence à partir d'études menées sur les effets de l'inhibition de retour sur la détection de cible.

D'un point de vue physiologique, l'inhibition de retour paraît durer au moins quelques secondes, et commencer après 225 millisecondes. L'inhibition de retour biaise l'orientation en la maintenant, dans l'environnement, loin des zones déjà inspectées, et ceci peut servir à faciliter le comportement d'inspection visuelle ; c'est-à-dire la recherche visuelle à proprement parler. Donc *l'inhibition de retour fonctionne pour faciliter la recherche visuelle en inhibant l'orientation vers les zones déjà visitées*.

L'inhibition de retour est générée dans un système qui est, normalement, responsable de l'orientation de la direction du regard. L'inhibition de retour inhibe l'orientation de l'attention “discrète” (*covert attention*³²), la direction du regard, et, plus généralement les réponses spatiales vers des zones ou des objets indicés (visités). En biaisant l'orientation loin des items déjà inspectés, l'inhibition de retour permet de rendre l'exploration de l'environnement plus efficace.

³²Covert Attention : expression of attention without eye movements, typically thought of as a virtual ‘spotlight’ [IK01].

L'inhibition de retour est-elle en contradiction avec les travaux d'Horowitz et Wolfe³³ [HW98] dont nous avons parlé plus haut, page 58 ? D'après Itti et Koch [IK01], la réponse est non. En effet, les résultats auxquels arrivent Horowitz et Wolfe excluent une parfaite mémorisation de tous les lieux déjà visités (sinon la recherche dans un environnement stable serait meilleure que dans un environnement qui change constamment), mais ils n'excluent pas la possibilité que les positions des derniers items visités soient mémorisées, en accord avec le fait que la période d'action de l'inhibition de retour soit courte, cf. figure 2.19 page 70.

2.3.3 Modélisation computationnelle de la vision et de l'attention visuelle

2.3.3.1 Généralités

Modéliser la vision a toujours été un défi en informatique et en intelligence artificielle. Il apparaît clairement qu'il faille étudier le cerveau afin de progresser en informatique (graphique en particulier); de la même façon qu'il faille se servir d'outils informatiques pour comprendre le fonctionnement du cerveau [Pog84]. La vision d'une manière générale est très complexe. Le système humain est une solution efficace qui permet de donner une orientation à la recherche. La vision peut être appréhendée comme un processus modulaire, c'est-à-dire qu'il existe une suite de modules qui agissent en parallèle. Ces modules produisent des représentations intermédiaires reprises ensuite par d'autres modules. Ces derniers déduisent les formes à partir des ombres portées, de la texture visuelle, du mouvement, des contours, des occlusions et de la vision stéréoscopique.

³³Ces travaux, nous le rappelons, mettent en évidence que la recherche visuelle n'a pas de mémoire.

Le cerveau utilise des impulsions d'intensités variées, alors que l'ordinateur fonctionne en terme de "tout ou rien". En outre, d'après Poggio [Pog84], le cerveau possède des connexions en trois dimensions et travaille en parallèle alors que le monde de l'informatique est en revanche en deux dimensions et travaille en série. Il faut donc raisonner en termes de traitement de l'information car la description des tâches est indépendante du matériel utilisé. Le problème peut se résumer par un découpage à trois niveaux :

- *Calculatoire* : Quels sont les calculs à effectuer ? Dans quelles conditions ?
- *Algorithmique* : Quelle est la séquence d'opérations à mettre en place ?
- *Matériel* : Comment s'exécute l'algorithme avec les *neurones* ou les circuits électroniques ?

2.3.3.2 Approches par groupement perceptuel

Des approches pour modéliser la scrutation naturelle de l'œil dans des contextes particuliers ont déjà été développées, notamment dans le travail de Fitzback à l'Université Laval [Fit99]. Ce type de modélisation propose d'utiliser les lois de la Gestalt pour guider le regard. Cette démarche originale qui prolonge les travaux de Alquier [AM96], [AM98] et [Alq98], permet une approche guidée, non-globale de type ébauche primitive complète. Ce travail a donné des résultats mitigés car les images utilisées (radar) étaient fortement bruitées ; en outre la complexité des calculs rendait le traitement long.

2.3.3.3 Approches statistiques

D'autres modèles pour le mouvement des yeux ont été proposés, en particulier une approche statistique [BG00]. Dans cette approche Brockmann et Geisel proposent que

les mouvements saccadés des yeux soient des réalisations statistiques dans un champ de saillance. Cette proposition repose sur l'hypothèse (*a priori* réaliste) que l'être humain minimiserait le plus possible le temps qu'il lui faut pour parcourir une image. Cette méthode utilise les "survol de Lévy" ou *Lévy flights*³⁴ pour générer des chemins de parcours très similaires aux chemins de parcours réels. En effet, les fixations s'accumulent dans les zones visuelles saillantes, les chemins de parcours sont différents pour chaque expérience et les *saccades* courtes sont plus fréquentes que les longues. Cette modélisation apporte deux informations. D'abord la structure géométrique des chemins de parcours, générée en utilisant une fonction en loi de puissance pour déterminer la *magnitude des saccades*, présente une ressemblance frappante avec les chemins de parcours naturels [BG00], alors que celle générée avec une échelle intrinsèque ne présente pas cette ressemblance. Ensuite, la couverture de l'*attention visuelle* se fait à travers la *vision périphérique* du système de saccades et ne possède pas de mémoire, ce qui est en accord avec le travail de Horowitz et Wolfe [HW98] page 58.

2.3.3.4 Modèles pré-attentionnels découlant de Treisman et Gelade

Le modèle d'intégration des traits de Treisman et Gelade [TG80] (cf. 2.3.2.1) a permis de proposer des modèles computationnels qui font la dichotomie entre une partie pré-attentionnelle parallèle de la vision et une partie sérielle.

Ainsi, en 1999, Cohen et Ruppin [CR99] proposent un nouveau modèle computationnel de recherche visuelle pour la partie pré-attentionnelle. Il s'agit de faire un modèle

³⁴Les survols de Lévy sont une généralisation des mouvements browniens ordinaires. Ces survols de Lévy ont fait l'objet d'études, entre autres, en physique [Fog94](dynamique des fluides, systèmes dynamiques, *etc*) et en biologie. Des études ont été menées sur le comportement des fourmis chercheuses ou le vol de la drosophile en utilisant les survols de Lévy. Plus récemment, ces mêmes survols de Lévy ont été utilisés pour décrire le comportement de recherche des albatros errants [VAB⁺96].

qui corresponde aux expériences montrant qu'il y a différents temps de réaction dans la tâche de recherche visuelle. Ces différents temps de réaction sont, nous le rappelons, dus au fait que les scènes visuelles sont dans certains cas scrutées de manière sérielle alors que pour d'autres cas elles sont balayées de manière parallèle. Le modèle proposé, qui est un réseau de neurones, opère en deux phases. D'abord, la scène visuelle est compressée de manière standard via une analyse en composantes principales (transformée de Karhunen-Loeve) pour calquer les limitations des capacités biologiques de l'être humain. Ensuite, la scène compressée est traitée afin de pouvoir identifier la cible dans la scène. Le modèle réussit à détecter rapidement des cibles sur des scènes qui ont été étiquetées³⁵ expérimentalement en parallèle mais en revanche le modèle échoue pour des scènes étiquetées de façon sérielle. Les scènes qui ont été compressées de manière parallèle contiennent de l'information de type *global* qui permet la détection instantanée de la cible.

Ce modèle a ensuite été amélioré et donne des résultats très proches des expérimentations réalisées avec des sujets humains [CLR00]. Dans ces travaux, des *cartes de caractéristiques* sont utilisées ; elles permettent de fabriquer les images compressées. Ces *cartes de caractéristiques* reflètent les limitations biologiques des sujets. Les cartes sont extraites par analyse en composantes principales ; l'analyse en composantes principales ici, ne se faisant pas classiquement de manière algébrique comme dans [CR99] mais en utilisant un réseau de *neurones* approprié³⁶. Ce type de modélisation repose donc plus sur le fait que les systèmes biologiques de traitement de l'information ont une capacité limitée ; c'est pourquoi des images compressées et des caractéristiques globales sont utilisées. La fin de la modélisation consiste en un traitement plus approfondi (en particulier, le réseau de neurones servant à la détection des cibles a été ajusté de [CR99]

³⁵Étiqueté : ici il faut comprendre que les zones d'intérêt ont été identifiées via la compression par l'analyse en composantes principales. En réalité, Cohen et Ruppin ne gardent que les sept axes principaux de l'analyse en composantes principales [CR99].

³⁶Pour plus de détails sur le réseau de neurones employé, consulter [CLR00].

à [CLR00]) pour l'identification des cibles proprement dites. La principale différence entre les modèles présentés dans [CR99] et dans [CLR00] est que dans le second cas le modèle est entièrement implémentable en utilisant des réseaux de neurones.

2.3.3.5 Modèle ascendant de Koch et Ullman

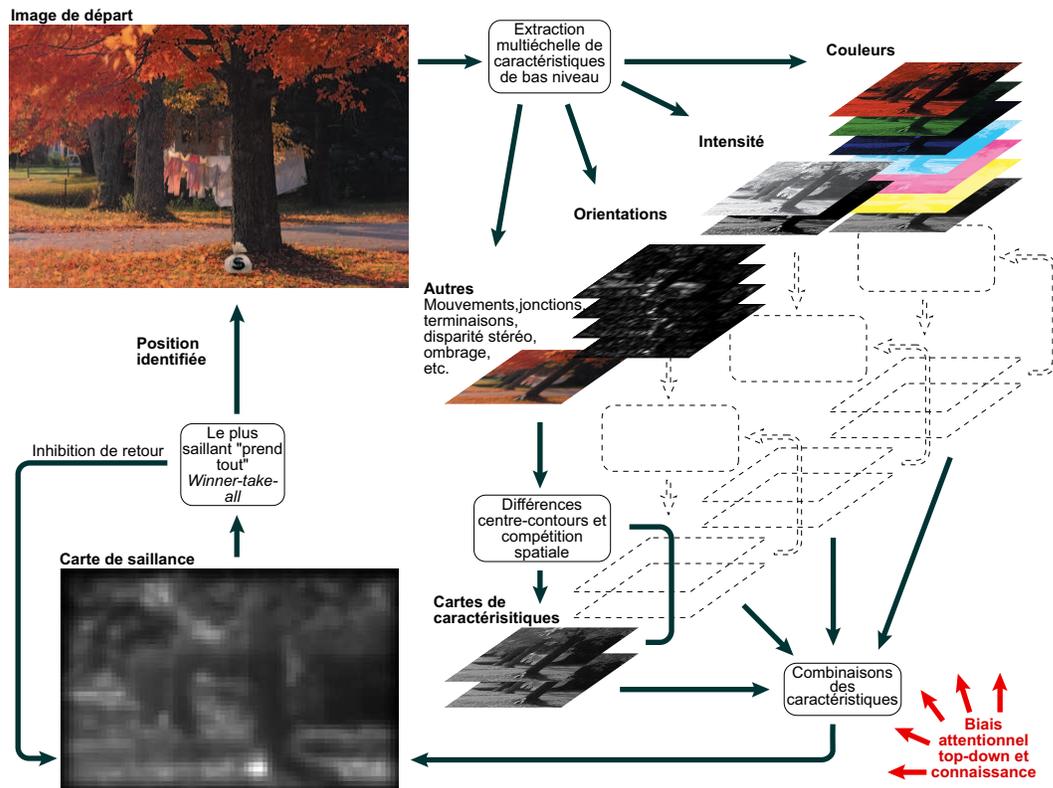


FIG. 2.18 – Diagramme d'un modèle type pour le contrôle ascendant de l'attention, d'après [KU85] modifié par [IK01]

Un des modèles à partir duquel de nombreux travaux sont issus est le modèle de contrôle de l'attention ascendante de Koch et Ullman [KU85]. Le modèle de Koch et Ullman (figure 2.18) est basé sur l'idée d'une carte de saillance (ici une carte topographique explicite en deux dimensions) qui encode l'intensité du stimulus ou la saillance pour toute position de la scène visuelle. La carte de saillance reçoit des entrées (caractéristiques de bas niveau) du processus visuel primaire et permet une stratégie de

contrôle efficace dans laquelle le *focus* de l'attention balaye simplement la carte de saillance dans un ordre décroissant de saillance.

La figure 2.18 montre comment l'image est encodée par les neurones, au travers de quelques mécanismes de détection de caractéristiques pré-attentionnelles, en cartes de contrastes pour chacune des caractéristiques. Au sein de chaque carte de caractéristiques, les neurones rivalisent spatialement pour la saillance. Les cartes de caractéristiques sont combinées ensuite pour obtenir la carte de saillance. Ce modèle est de type ascendant, la figure (en bas à droite) montre également qu'il existe une composante descendante (dépendant de la tâche) qui a été introduite ultérieurement au modèle original de Koch et Ullman. L'importance de la dépendance de l'attention visuelle vis-à-vis d'indices descendants a été étudiée plus tard, en particulier par Itti et Koch depuis la fin des années 90.

2.3.3.6 Principales clés pour une modélisation moderne

D'après Itti et Koch [IK01], cinq idées principales émergent des plus récentes publications dans le domaine de l'attention visuelle et de sa modélisation :

1. La perception de la *saillance* (cf. 2.3.3.7 page 69) dépend fortement du contexte qui l'entoure.
2. Une *carte de saillance* peut-être une bonne stratégie de contrôle pour une approche ascendante.
3. L'*inhibition de retour* est un élément critique (cf. paragraphe "Inhibition de retour" page 60) à ne pas oublier.
4. L'attention et le mouvement des yeux interagissent, ceci engendre des problèmes informatiques importants.

5. La compréhension et la reconnaissance contraignent fortement la sélection des zones de vue.

La fonction la plus importante de l'attention visuelle sélective est de diriger notre regard vers les objets d'intérêt au milieu de l'environnement visuel général. Ainsi, en terme de traitement de l'information, elle sert à restreindre le traitement en un mode "fin" du *champ visuel*, pour en réduire le coût. Depuis longtemps, comme nous l'avons vu, il est le plus couramment admis que l'attention visuelle soit guidée par des mécanismes de saillance ascendants (cf. plus haut), mais il ne faut pas négliger, selon Itti et Koch, la dépendance de l'attention visuelle vis-à-vis d'indices descendants dépendant de la tâche. La saillance est indépendante de la tâche, elle est intégrée de manière très rapide, dirigée d'abord par les mécanismes ascendants; elle est traitée de manière pré-attentionnelle, à travers l'ensemble du champ de vision.

D'une façon générale Itti et Koch distinguent deux types d'attention : l'attention volontaire et l'attention involontaire. L'attention volontaire demande un effort. Les deux attentions peuvent être menées en parallèle.

L'attention permet de simplifier le problème de la compréhension d'une scène grâce à la décomposition en une série rapide, informatiquement moins lourde, de *problèmes localisés* d'analyse visuelle. En plus de l'analyse de scène et de l'orientation, l'attention est caractérisée par une modulation, *par rétroaction*, de l'activité neuronale pour les attributs visuels, et ceci, dans la zone de la cible désirée ou sélectionnée. Cette rétroaction apparaît essentielle pour rassembler les attributs visuels d'un objet, comme la couleur et la forme par exemple.

2.3.3.7 Éléments principaux de la modélisation

Dans cette section nous abordons quatre éléments principaux qui doivent être constitutifs de toute modélisation moderne de l'attention visuelle [IK01] : la prise en compte du traitement pré-attentionnel du système de vision, l'utilisation de la saillance, l'importance de l'inhibition de retour et enfin l'intégration des aspects montant et descendant.

Traitement pré-attentionnel des caractéristiques visuelles

La première étape de traitement dans tout modèle de type ascendant pour l'attention est le traitement des *premières caractéristiques visuelles*³⁷. Les premières caractéristiques visuelles sont calculées (traitées) de manière pré-attentionnelle de façon massivement parallèle à travers tout le champ de vision. D'un autre côté, il a été récemment montré que l'attention peut clairement et vivement moduler, de façon descendante, les premiers traitements de la vision, et ce, d'une manière définie aussi bien spatialement que non spatialement. En raison de la modulation par rétroaction qui influence le traitement des premières caractéristiques ascendantes, les modèles doivent prendre en compte cet aspect non-négligeable. D'une manière générale les mécanismes pré-attentionnels qui extraient les premières caractéristiques de la vision à travers tout le champ de vision ne doivent pas être oubliés dans l'optique d'une modélisation.

Saillance et carte de saillance

Pour résoudre le problème de la représentation multiple d'une vue au sein des nombreux réseaux neuronaux, beaucoup de modèles ascendants d'attention proposent que cette représentation multiple soit regroupée en une *carte de saillance* unique. Informatiquement, une représentation explicite de la saillance dans une

³⁷En anglais : early visual features.

carte dédiée renforce l'idée qu'une partie de la sélection spatiale puisse être réalisée au cours de la détection des caractéristiques pré-attentionnelles. Le modèle de Wolfe [Wol94] élabore la carte de saillance avec, à la fois une approche ascendante, mais également avec une pondération descendante ; la carte de saillance est alors une probabilité (croyance) que la cible soit présente à un endroit donné. Ce point de vue a reçu un récent appui du côté expérimental [PLN02]. Ce qui semble important pour le traitement informatique de la saillance est le contraste des caractéristiques tout en respectant l'environnement. De plus en plus, la saillance apparaît être codée dans le cortex et séparément des autres caractéristiques visuelles [IK01]. Quelques modèles, qui reproduisent le comportement de sujets ou de singes dans des tâches de recherches visuelles, ont démontré la plausibilité des cartes de saillance ainsi que leur traitement informatique.

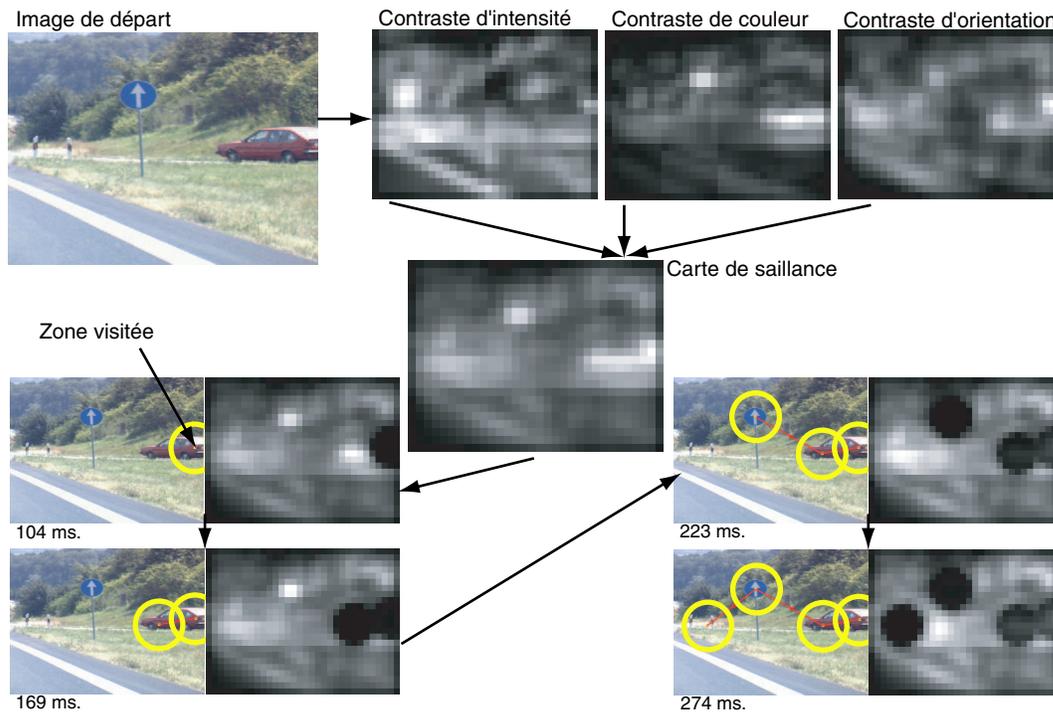


FIG. 2.19 – Modèle d'attention d'Itti et Koch [IK00] prenant en compte l'inhibition de retour. La construction de la carte de saillance ressemble à celle de la figure 2.18. Les zones d'attentions sont sélectionnées successivement dans l'ordre décroissant de leur saillance. Une fois une zone visitée, son niveau de saillance est mis au plus bas par le mécanisme d'inhibition de retour. Progressivement l'effet de l'inhibition va se dissiper.

Inhibition de retour

Bien que simple dans le principe, du point de vue informatique, l'inhibition de retour est une composante très importante de l'attention cf. 2.3.2.2 page 60. Le modèle d'Itti et Koch [IK00], figure 2.19, dans lequel l'inhibition de retour est prise en compte est de ce point de vue exemplaire.

L'inhibition de retour permet de déplacer rapidement l'attention sur des régions ayant une saillance décroissante, et non de rester lié toujours à la zone la plus saillante à un moment donné. Le rôle de l'inhibition de retour dans la vision active et dans l'attention “manifeste” (*overt attention*³⁸) est un aspect qu'il faut envisager dans un futur modèle.

Attention et reconnaissance Il apparaît évident qu'un modèle de contrôle de l'attention le plus complet possible doit intégrer une composante descendante. Le défi informatique est alors d'intégrer à la fois l'approche descendante et l'approche ascendante.

Deco et Zihl [DZ01] ont récemment proposé un modèle de ce type. Leur modèle commence par sélectionner des zones pour les objets candidats d'une manière ascendante grâce à une analyse de l'image à une échelle grossière. Un mécanisme attentionnel parcourt les zones candidates de façon sérielle et réalise la reconnaissance des objets à des échelles de plus en plus fines jusqu'à une reconnaissance suffisante par rapport à ce qui est contenu en mémoire. Autrement dit, la scène est d'abord analysée à un niveau de résolution grossière, puis le *focus* de l'attention rehausse de manière itérative la résolution sur l'emplacement de l'objet jusqu'à ce que ce dernier soit identifié.

³⁸Overt attention : expression of attention involving eye movements [IK01].

2.3.3.8 Le système complet de Machrouh

Toutes les approches citées n'ont été que très rarement intégrées les unes aux autres. Cependant, l'approche de Machrouh [MLT01b], [MLT01a], [Mac02], [MLT03] et [MT05], propose un système complet basé sur un agent logiciel doté de capacités perceptives. Cette approche est très séduisante et s'apparente beaucoup au système visuel naturel. L'agent, afin de pouvoir explorer la scène visuelle, doit pouvoir déterminer les points d'intérêt de la scène visuelle. Machrouh effectue deux types de traitements. Le premier s'opère sur un *champ large* et basse résolution pour l'exploration, c'est-à-dire la recherche des points d'intérêt ; le second consiste en une série de traitements focaux (autrement dit localisés) à haute résolution pour l'exploitation, c'est-à-dire la reconnaissance de la cible. Les points d'intérêt sont déterminés en retenant les maxima d'énergie (calculés à l'aide de *filtres* en ondelettes couvrant une gamme d'orientations et de fréquences spatiales) dans une analyse en composantes principales (transformée de Karhunen Loeve) tout comme dans l'approche proposée par Cohen *et al.* [CLR00]. Ces saillances (points d'intérêt) sont alors indexées en fonction de leurs énergies.

Selon Machrouh, un traitement partiel des scènes visuelles permet un grand gain de calcul vis-à-vis des systèmes traditionnels qui traitent toute l'image. Il faudrait donc introduire un système sélectif reposant sur des cartes de saillance tout comme le font Itti et Koch [IK00].

Afin d'extraire les caractéristiques de bas niveau, Machrouh retient entre autre des systèmes visuels des mammifères qu'il existe une distinction opérée par le système visuel entre la périphérie et le centre ; cette distinction repose sur les fréquences spatiales.

Pour travailler sur les différentes fréquences spatiales, au lieu d'utiliser des filtres de

fréquences différentes, Machrouh “*zoom*” les différents champs visuels à la taille de la fovéa (figure 2.20) et ainsi applique le même filtre. C’est ainsi le même traitement en terme de fréquences spatiales. Les différents champs visuels correspondent aux différents champs du système biologique. On rejoint ici les notions vues précédemment dans ce chapitre entre champ large traité de manière parallèle et attention focalisée traitée de manière sérielle.

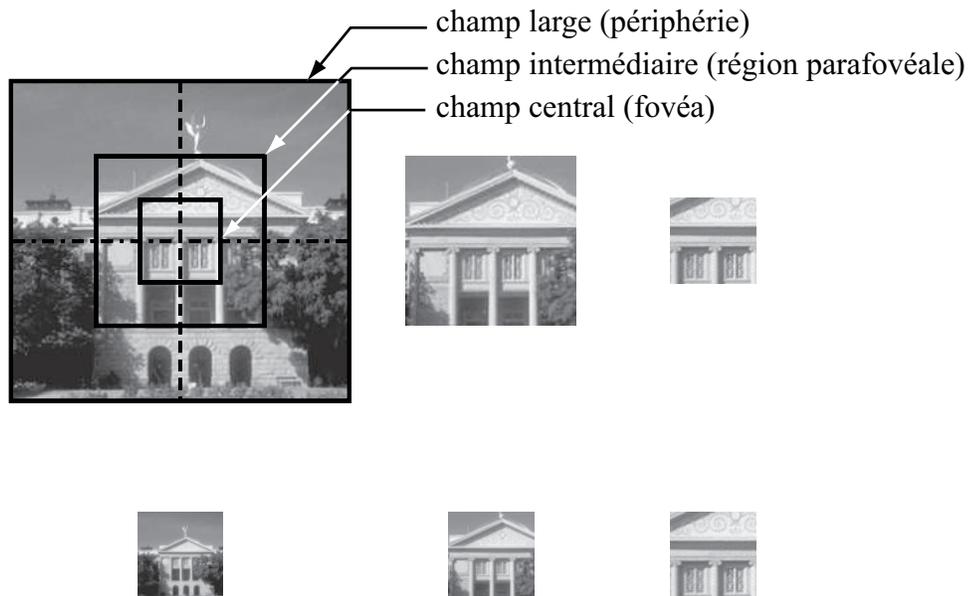


FIG. 2.20 – Les zooms appliqués par Machrouh sur une scène visuelle, d’après [Mac02]

L’analyse en composantes principales permet ensuite de déterminer un ensemble de traits aptes à guider l’exploration de scènes. Le principe d’exploration du système proprement dit repose sur le fait qu’une toute petite partie de la scène doit être traitée. La partie *fovéale* permet d’extraire les informations fines alors que la partie périphérique sert à guider le regard sur les zones saillantes de la scène. Ceci permet de réduire le temps de calcul en limitant le traitement à certaines parties sans traiter toute la scène visuelle.

Les point saillants sont déterminés par l’espace de l’analyse en composantes princi-

pales. L'ensemble de ces points est appelé *saillances naturelles* par Machrouh. Ce sont ces saillances naturelles qui permettent de guider les *saccades* effectuées par le sujet pour explorer l'image.

La carte de saillance permet de guider le système soit par une opération de recherche, soit par une opération d'exploration. La carte de saillance correspond au champ large, c'est-à-dire basse fréquence.

Le système proposé par Machrouh à l'avantage de permettre deux types d'explorations d'image [Mac02] :

1. Ascendante : guidée uniquement par la carte de saillance
2. Descendante : guidée par la carte de saillance *et* modulée par une information préalablement mémorisée.

2.4 Multirésolution, recherche et interrogation

Comme nous l'avons vu dans la section 2.3.2.1 (page 54), la vision naturelle ne traite qu'une partie infime d'une scène visuelle, la région *fovéale*. La partie périphérique ne traite que l'information basse fréquence qui permet d'orienter la fovéa vers les parties intéressantes de la scène visuelle.

La série des trois objectifs spécifiques annoncés dans le chapitre 1³⁹ s'apparente à une technique multirésolution puisqu'elle travaille à la fois au niveau local et au niveau global. Cette section explore par conséquent les techniques reposant sur des approches multirésolution pour la représentation, la recherche et l'interrogation dans les images

³⁹Synthétiquement : -1- sous échantillonnage intelligent de l'image, -2- extraction locale de l'information recherchée et -3- fusion des informations locales en objets globaux.

numériques.

D'un point de vue computationnel la multirésolution peut-être également un moyen d'augmenter l'efficacité de la recherche visuelle. La multirésolution peut-être obtenue soit par une approche locale-globale, soit en réalisant un découpage de l'image, soit en combinant ces deux approches.

2.4.1 Découpage simple

Une des méthodes que l'on peut utiliser pour découper une image est l'arborescence en "quadtree" [CW94]. Dans ce type de travaux on cherche à extraire rapidement et efficacement une courbe de l'image. Le principe est de ne garder que les "feuilles" de l'arborescence pour lesquelles une portion de courbe est présente (cf. figure 2.21 page 76). Il est possible de dégager deux étapes dans la méthode :

1. La courbe est découpée autant de fois que nécessaire pour obtenir des segments dans chaque feuille. Naturellement les feuilles peuvent avoir des tailles différentes.
2. Les segments sont regroupés selon un critère de connexion reposant sur l'orientation relative et la proximité.

Cette approche présente trois avantages. D'abord, c'est toujours la même application locale de détection qui est utilisée (quelle que soit l'échelle). Ensuite, l'efficacité est grande en terme de traitement informatique. Enfin, ce type de découpage évite une redondance de l'information pour définir la courbe.

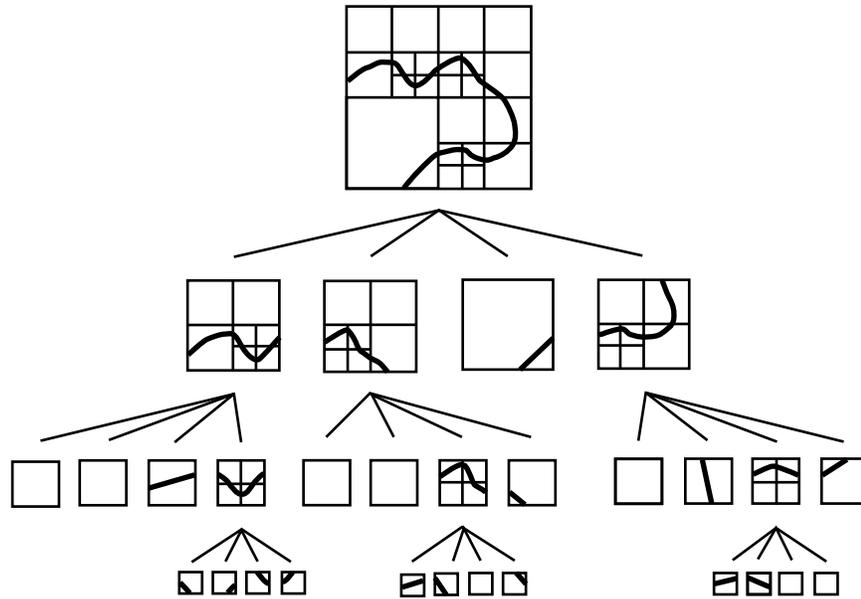


FIG. 2.21 – D’après [CW94], représentation d’une courbe par quadtree

2.4.2 Systèmes de recherche d’information visuelle

Afin de retrouver une image dans une banque d’images, de nombreuses méthodes sont développées depuis plus d’une dizaine d’années. Les systèmes de recherche d’information visuelle⁴⁰ ne sont vraiment utilisables que s’ils peuvent retrouver l’information en temps réel [DM00]. En plus des mots clés assignés par le constructeur de la base de données, les systèmes de recherche d’information visuelle peuvent utiliser le contenu des images comme index, c’est-à-dire les caractéristiques de couleur, de texture et de forme. Depuis quelques temps quelques systèmes utilisent des combinaisons d’attributs hétérogènes pour améliorer les résultats en termes de discrimination et de classification. Ces systèmes utilisent la couleur, la texture et la forme pour les interrogations d’images. Bien que l’analyse de texture ait une histoire très riche en traitement d’image et en vision, son utilisation pour des bases de données d’images est peut-être une des premières applications à grande échelle démontrant la pertinence de l’utilisation de la texture.

⁴⁰*Système de recherche d’information visuelle* en anglais : VIR (visual information retrieval) system

La forme est un autre attribut de bas niveau qui peut être utilisé pour représenter l'information locale d'une image.

Une des plus récentes approches [DM00] est la recherche par croquis. Cette approche permet de prendre en compte la possibilité d'une déformation due à la perception de l'interrogeur (la subjectivité peut induire une grande variance dans les réponses). Daoudi et Matusiak [DM00] proposent une "métrique" sur la forme des courbes : le *curvature scale space (CSS)*, "espace de courbure multi-échelle". Ils proposent également d'utiliser les relations spatiales des formes elles-mêmes. Le problème au départ est relativement compliqué : en effet les formes⁴¹ ne sont pas aisément "mathématisables" ; la difficulté réside donc dans le fait qu'il faille trouver une expression qui soit fonction de la similarité mais également trouver un moyen d'exprimer l'objet en fonction de cette similarité. Il n'est pas possible dans le contexte qui est présenté par Daoudi et Matusiak d'utiliser les mesures de distances mathématiques classiques en raison de la composante humaine. Une difficulté supplémentaire provient de la différence de nature qu'il y a entre un croquis et une image. Les auteurs préconisent alors une approche multirésolution afin de modéliser le croquis. Le travail, dans le cas présent, s'est fait sur des images déjà segmentées, ce qui allège considérablement le traitement. De par l'approche multirésolution chacune des formes (courbes) du croquis est représentée à son échelle propre, la courbe (γ) est paramétrée par la longueur de l'arc qui la constitue. La représentation par CSS possède certaines propriétés comme :

1. La représentation par CSS est invariante à l'intérieur d'une classe de similarité c'est-à-dire pour une translation, une rotation ou un changement d'échelle.
2. La "complétude" : cette propriété assure que deux contours auront la même forme si et seulement si leurs CSS sont égales.
3. Cette représentation est stable, ce qui lui donne une bonne robustesse pour des

⁴¹Ici les formes sont des contours (formes fermées).

petites distorsions dues au croquis ou à la quantification.

4. Simplicité et temps réel du traitement, propriété très importante pour les applications avec des banques d'images.

Pour résumer, l'approche préconisée par Daoudi et Matusiak a les qualités suivantes : temps réel, représentation efficace pour la perception humaine, invariance par rapport aux transformations (changements d'échelle, rotations et translations) et robustesse par rapport aux variations des formes.

2.4.3 Approches par ondelettes

S'agissant de multirésolution, la transformation la plus populaire est certainement la transformée en ondelettes. Elle est beaucoup utilisée depuis quelques années dans de nombreuses applications de la physique des signaux [Gib96]. L'analyse par ondelette *“est faite via des décompositions en séries de fonctions [...] Ces fonctions, que l'on appelle ondelettes, sont construites à partir d'une ondelette départ à laquelle on applique des dilatations et des translations.”* [Gib96]

Afin de représenter des courbes à différents niveaux de résolution⁴² il est également possible de travailler avec une approche par transformation en ondelettes. La méthode proposée par Finkelstein et Salesin [FS94] permet de ne pas avoir besoin de stockage en mémoire en supplément malgré la multirésolution. Il est ainsi possible de :

- changer la tendance sans le détail, voir figure 2.22 ;
- changer le détail sans la tendance, voir figure 2.22 ;
- éditer une courbe à n'importe quel niveau de détail ;
- permettre des niveaux continus de lissage pour enlever les caractéristiques indé-

⁴²On ne se situe pas ici précisément en vision, mais la problématique est commune.

– sirables des courbes ;

- proposer une approximation qui reste dans un taux d’erreur garanti.

L’avantage le plus important de ces algorithmes développés par Finkelstein et Salesin se situe dans le fait qu’ils sont simples et rapides.

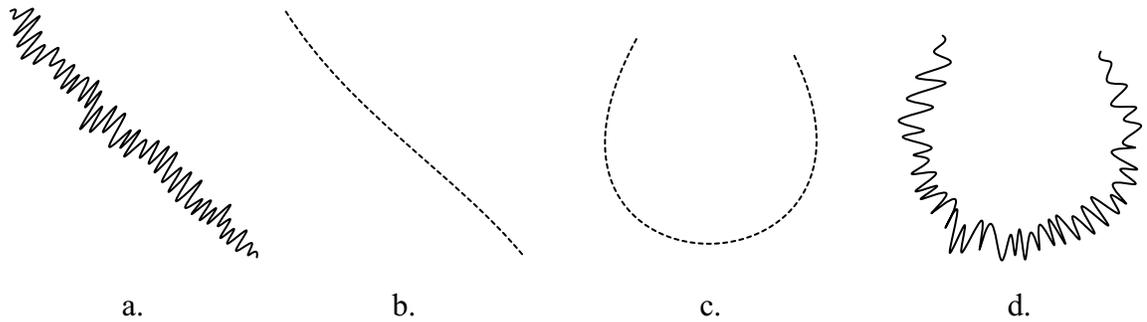


FIG. 2.22 – Tendances et détail d’une courbe. a. et b. (tout comme c. et d.) ont la même tendance mais pas le même détail. a. et d. (tout comme b. et c.) ont le même détail mais pas la même tendance, d’après [FS94].

Une des premières approches multirésolution utilisant la transformée en ondelettes pour résoudre le problème de la recherche d’un croquis (ou d’un mauvais scan) dans une base d’images est le travail de Jacobs et al. [JFS95]. Dans ce travail, les auteurs se servent d’une “métrique” qui utilise les coefficients des ondelettes. Ces métriques possèdent des paramètres qui peuvent être ajustés de manière statistique afin de s’affranchir des problèmes de distorsion qui peuvent être introduits par le dessin. Cette méthode présente de nombreux avantages :

- La décomposition par ondelettes permet d’avoir une très bonne approximation de l’image avec un nombre relativement limité de coefficients. Cette propriété est utilisée en compression d’image depuis longtemps.
- La décomposition en ondelettes peut être utilisée pour extraire et encoder l’information sur les *arêtes* . Ce sont souvent les arêtes qui sont les clés de la requête de l’utilisateur.
- Les coefficients de l’ondelette fournissent des informations qui sont indépendantes

de la résolution originale de l'image. Il est donc possible avec l'utilisation de la transformation par ondelettes de découpler la résolution de la requête et la cible proprement dite.

- La décomposition en ondelette est rapide et facile en terme d'informatique, nécessitant un temps proportionnel à la taille de l'image et peu de code.

Ce type d'approche [JFS95] s'apparente à des degrés divers à l'interrogation par contenu [BEN⁺93], [FBF⁺94], [NBE⁺93], l'interrogation par l'exemple [HK92, Kat92, KKO92], la recherche par similarité [LC93], [WTCC94], [GS93], [KZT93], [PSTT93] et enfin la recherche par croquis [KKO92]. L'algorithme est rapide et efficace. Sachant que le croquis n'est qu'une approximation grossière, le principe est de retenir une vingtaine d'images parmi les milliers de la base plutôt que le proposer *la* bonne. Un des principaux avantages puisque la méthode repose sur la décomposition par ondelettes est qu'elle est indépendante de la résolution. La métrique qui est utilisée dans [JFS95] a été conçue en tenant compte :

- de l'espace des couleurs ;
- du type d'ondelette ;
- du type de décomposition ;
- de la troncature des coefficients de l'ondelette ;
- de la quantification des coefficients de l'ondelette ;
- de la normalisation de l'ondelette.

De manière synthétique cette approche est donc rapide et efficace. Elle ne nécessite, en outre, que peu de données supplémentaires à stocker. En revanche, cette méthode fonctionne avec des patrons généraux mais pas avec de petites formes comme des icônes ou des logos.

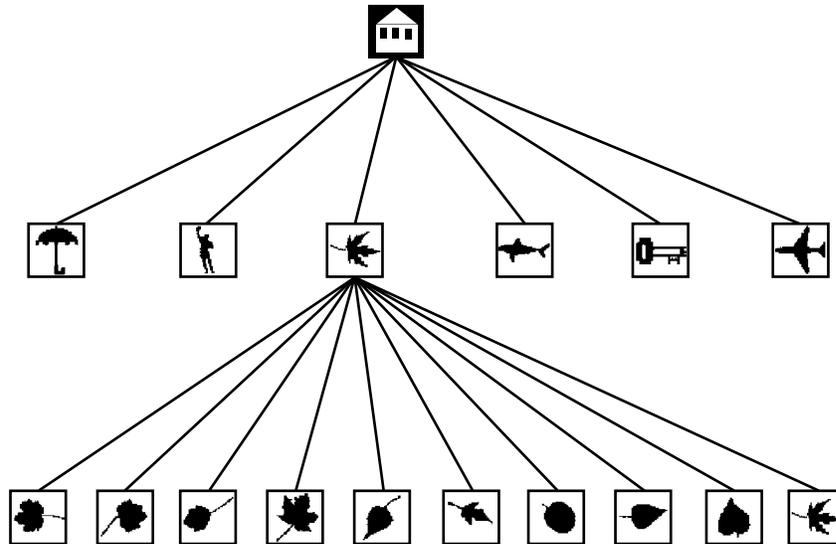


FIG. 2.23 – Exemple de graphe iconique hiérarchique [TG00].

2.4.4 Interrogation par objet

Nous avons vu qu’il existe plusieurs types de recherche et d’interrogation pour une base d’images (cf. le présent paragraphe) auxquelles s’apparente la méthode de Jacobs. Une des dernières méthodes développées par Tao et Grosky [TG00] est l’interrogation par objet. En effet par nature, une image est constituée par un assemblage d’objets. La recherche dans l’image se fait par rapport à un objet. Cette approche permet à l’utilisateur de pouvoir faire des interrogations sémantiques aussi bien de haut niveau que de bas niveau ; ces niveaux sont traduits dans un graphe “iconique” hiérarchique comme dans l’exemple de la figure 2.23. L’algorithme travaille grâce à la disposition spatiale des points de l’objet lui-même et grâce à la distribution spatiale des objets au sein de l’image.

2.5 Raisonnement spatial qualitatif

Le dernier objectif spécifique de ce travail, annoncé dans le chapitre 1 est de proposer une technique de fusion des informations locales en une structure globale. Le raisonnement spatial qualitatif (le sens commun) permet un tout à chacun de réaliser cela facilement tous les jours sans même en avoir conscience. Prenons par exemple le déplacement à l'intérieur d'une ville. Ce déplacement fournit des informations, unes à unes, très localisées ; par exemple « le panneau de circulation est juste devant la station service ». Pourtant après avoir acquis une certaine expérience de la ville il est possible de faire des connexions sur des objets qui n'ont pas de relations visuelles entre eux. Par exemple à Paris, après avoir expérimenté la ville en s'y déplaçant, on pourra dire que « l'Arc de Triomphe se trouve au nord de la Tour Eiffel⁴³ » même sans avoir vu de carte de la ville et même si aucune rue ne relie ces deux monuments⁴⁴. Même de petits insectes, dotés de très petits cerveaux, sont capables d'intégrer des informations spatiales locales pour retrouver leur chemin comme nous le verrons plus loin dans cette section. Le raisonnement spatial qualitatif, qui permet de fusionner des informations spatiales locales en une information spatiale globale, est par conséquent de premier intérêt dans le cadre du dernier objectif spécifique de la thèse.

Mettre en relation reconstruction de l'espace à partir de vues locales et appréhension globale d'une scène visuelle est une approche très originale. Nous cherchons à mettre en avant les aspects qui permettent d'établir des relations spatiales entre les objets d'un environnement dans l'idée qu'il est possible d'établir des relations de même type entre les objets (ou parties d'objets) contenus dans une image *i.e.* une scène visuelle.

⁴³Ou bien pointer la direction grossière dans laquelle se trouve le second monument quand on est près du premier.

⁴⁴Et sans être monté à leurs sommets (!)

Ainsi, l'objet de cette section 2.5 n'est pas de dresser une revue exhaustive du raisonnement spatial qualitatif mais de mettre l'accent sur les principaux points pertinents qui serviront (ont servi) à appuyer l'élaboration du modèle ELECA, chapitre 6 en particulier. Ainsi, nous insisterons plus loin sur deux types de relations entre les objets soit la distance §2.5.4 et l'orientation §2.5.3. Mentionnons qu'il existe également d'autres types de relations [BM00] [CH01] [Ren02] qui mettent en œuvre la topologie, l'alignement ou la forme.

2.5.1 Qu'est-ce que le raisonnement qualitatif ?

Depuis longtemps la description de l'espace s'est faite selon deux modes principalement. D'un côté, existent les modes très connus de la géométrie euclidienne (ou d'autres géométries découlant de cette dernière) et de l'autre les modèles qui viennent de la perception et des observations empiriques. Les modèles mathématiques qui décrivent l'espace se classifient selon deux géométries : la géométrie des mathématiciens et *la géométrie du monde sensible* [BM00].

La géométrie du monde sensible repose sur les connaissances intuitives que nous avons du monde, ce qui est appelé également le *sens commun* [HM85] [Dav90]. Ces connaissances permettent d'évaluer les relations entre les objets d'une manière qualitative, par opposition à quantitative. Le raisonnement basé sur ces relations est appelé *raisonnement qualitatif*. Lorsqu'il s'applique à l'espace, on parle de *raisonnement spatial qualitatif* [CH01]⁴⁵. Le raisonnement qualitatif présente des avantages intéressants par rapport au raisonnement quantitatif [Her94] :

Complexité : travailler avec une échelle quantitative augmente bien souvent de beau-

⁴⁵[CH01] est une version révisée et mise à jour de deux différents papiers publiés avant par Cohn : [Coh96] et [Coh97]

coup la complexité des algorithmes dus à une granularité inadaptée. En effet, on a souvent besoin de savoir qu'une table est dans le fond de la pièce, non qu'elle se trouve à 4,35 m de la porte d'entrée dans une direction de 135° par rapport au nord.

Effets de bords : dans le monde quantitatif, on force les valeurs à être discrètes ce qui peut entraîner des erreurs une fois la valeur vérifiée.

Informations partielles ou incertaines : le problème des modèles qui travaillent avec une échelle quantitative est qu'ils ne peuvent pas manipuler les informations "floues". Par exemple, la chaise est adjacente au bureau est une information spatiale sans aucun doute mais qu'il n'est pas (ou très mal pour le moins) descriptible dans une échelle quantitative.

2.5.2 Allen : du temporel au spatial

Le raisonnement spatial qualitatif a pris de plus en plus d'importance après qu'Allen ait proposé en 1983 une logique basée sur les intervalles temporels [All83]⁴⁶. Allen définit 13 relations entre ces intervalles, tableau 2.1. Ces relations sont (conjointement) exhaustives entre elles et disjointes par paires (deux à deux); autrement dit entre deux intervalles il n'existe qu'une et une seule de ces 13 relations.

La première colonne liste les termes originaux d'Allen pour le domaine temporel. Dans le cas du domaine spatial, ces intervalles deviennent des segments de lignes ou de courbes; la deuxième colonne donne une description en langage naturel dans le domaine spatial, d'après [KSJ⁺04]⁴⁷. La troisième colonne donne la notation qu'Allen

⁴⁶En réalité, d'après Renz [Ren02], on doit ce système logique d'intervalles à Nicod [Nic20], proposé dans les années 20.

⁴⁷Cette colonne est une traduction libre, les termes ne sont pas toujours en adéquation avec l'anglais.

avait proposée [All83] et qui est également utilisé dans le domaine spatial. Enfin la dernière colonne présente des exemples graphiques. Dans le domaine temporel il s’agit d’évènements ayant une durée – avec un début et une fin. Si l’on considère ces exemples dans le domaine spatial il faut se placer dans le cas 1D, l’épaisseur des rectangles n’est là que pour une représentation plus aisée.

RELATION	LANGAGE NATUREL DU DOMAINE SPATIAL	SYMBOLE	EXEMPLE GRAPHIQUE
<i>before</i>	X est à la gauche de Y	$X < Y$	
<i>meets</i>	X touche Y à gauche	$X m Y$	
<i>overlaps</i>	X chevauche la gauche de Y	$X o Y$	
<i>starts</i>	X débute Y à gauche	$X s Y$	
<i>during</i>	X est contenu dans Y	$X d Y$	
<i>finishes</i>	X termine Y à droite	$X f Y$	
<i>equals</i>	X égale Y	$X = Y$	
<i>finishing-inverse</i>	X contient Y, finit à droite	$X fi Y$	
<i>during-inverse</i>	X contient Y	$X di Y$	
<i>starts-inverse</i>	X contient Y, débute à gauche	$X si Y$	
<i>overlaps-inverse</i>	X chevauche la droite de Y	$X oi Y$	
<i>meets-inverse</i>	X touche Y à droite	$X mi Y$	
<i>after</i>	X est à la droite de Y	$X > Y$	

TAB. 2.1 – Les 13 relations d’Allen [All83], enrichies par une description en langage naturel pour le domaine spatial [KSJ⁺04] – traduction libre.

Le calcul d’Allen comprend les intervalles, les 13 relations qualitatives sur ces intervalles et une algèbre qui permet de raisonner entre ces intervalles. Le raisonnement qui repose sur les intervalles d’Allen est basé sur les compositions entre les 13 relations, ce qui fait 169 combinaisons possibles. En effet, s’il existe une relation entre A et B et une relation entre B et C alors la relation entre A et C est caractérisable – c’est

Les termes originaux sont [KSJ⁺04] *X lies to the left of Y* | *X touches Y at the left* | *X overlaps Y from the left* | *X lies left-justified in Y* | *X is completely in Y* | *X lies right-justified in Y* | *X equals Y* | *X contains Y left-justified* | *X surrounds Y* | *X contains Y left-justified* | *X overlaps Y from the right* | *X touches Y at the right* | *X lies to the right of Y*.

une composition. Par exemple si $A < B$ et $B < C$ alors $A < C$ ⁴⁸. Allen a proposé dans cette même contribution [All83] un algorithme qui calcule les compositions entre les relations par satisfaction de contraintes. Les travaux qui ont découlé du calcul d'Allen sont très nombreux, le lecteur pourra consulter [CH01] ou [Ren02] pour plus de détails. Ces travaux se poursuivent à l'heure actuelle et sont toujours plus nombreux et dans des domaines différents. Citons par exemple [LMC04] qui pose les nouvelles questions en raisonnement spatial et temporel ou bien [KSJ⁺04] qui propose de vérifier la validité du raisonnement spatial qualitatif en psychologie. Les résultats de [KSJ⁺04] montrent que le raisonnement spatial avec des relations d'intervalles est largement basé sur la construction et l'examen de représentations spatiales qualitatives ou de modèles mentaux.

2.5.3 Orientation

L'orientation est une notion très pratique à utiliser en termes qualitatifs ; c'est l'expérience que tout un chacun a du monde. Ainsi, nous dirons facilement que Brigitte est “à la droite” de Pierre, ou que Québec “est à l'ouest” de Saint-Malo, ou bien encore que la voiture roule “en direction de” la Bretagne. Par opposition, il ne viendrait pas à l'esprit de dire que l'université Laval est 2° de plus dans la direction du Sud que l'université Paris-Sud⁴⁹.

La littérature distingue trois types de cadres de référence qui permettent de donner des directions [Her94] :

⁴⁸Domaine temporel : si A précède B et B précède C alors A précède C .
 Domaine spatial (1D) : Si A est à la gauche de B et B est à la gauche de C alors A est à la gauche de C .

⁴⁹Québec est *réellement* “plus au sud” que Paris. . .

intrinsèque : l'orientation est donnée par les propriétés intrinsèques de (inhérentes à) l'objet de référence ; par exemple « Brigitte est à la droite Pierre » ;

extrinsèque : l'orientation est donnée par des facteurs extérieurs à l'objet de référence ; par exemple « la voiture roule en direction de la Bretagne » (l'orientation est donnée par le mouvement du véhicule) ;

déictique : l'orientation est donnée par un point de vue extérieur à l'objet de référence, celui de l'observateur ; par exemple « Québec est à l'ouest de Saint-Malo ».

Pour représenter l'orientation d'un point de vue qualitatif trois exemples sont donnés dans la figure 2.24.

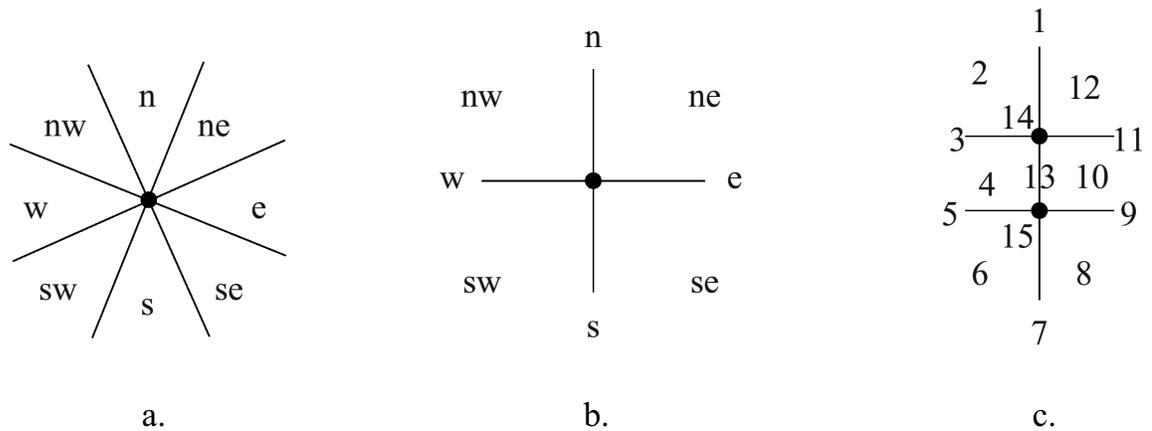


FIG. 2.24 – Représentation de l'orientation entre points, d'après [Ren02]. a. Basée sur les cônes. b. Basée sur les projections. c. Double croix.

Chaque type de représentation supporte une approche différente. Ainsi, la représentation basée sur les projections permet à Ligozat [Lig98] de proposer une “algèbre des directions cardinales” avec un formalisme qui lui permet de raisonner sur les directions cardinales et d'en étudier les propriétés informatiques – du point de vue algorithmique. Ces travaux sur les directions cardinales ont trouvé un prolongement original dans une théorie de reconstruction de l'espace à partir des perceptions locales [LE00] [EL04] sur lequel repose le modèle CLE qui est développé au chapitre 6. Ligozat et Edwards montrent qu'il est possible, grâce au raisonnement spatial qualitatif de prendre nombre

de décisions sans se référer à des directions cardinales (*sensu stricto*), à un système de coordonnées ou à des métriques⁵⁰ [LE00]. En effet, dans ces conditions l'homme n'a la plupart du temps aucune difficulté à se déplacer et à prendre des décisions. Le sujet élabore une compréhension de son environnement à partir de ses connaissances d'ordre général et des informations locales qui lui parviennent.

2.5.4 Distance

La distance est quelque peu différente de l'orientation dans la manipulation qu'en font les humains en général. En effet bien que floue, la notion de distance peut être quantitative "naturellement". Ainsi on dira en langage naturel, « Brigitte est proche de Pierre », ou bien « Saint-Malo est plus près de Paris que de Québec » qui sont des énoncés qualitatifs mais également « je me tenais à environ 20 mètres de l'ours » qui donne une information qualitative.

La représentation des distances est relativement plus simple que celle de l'orientation. On peut faire la distinction par exemple entre très loin, loin, "pas trop loin", proche et très proche. Ce peut être représenté par des cercles concentriques autour du point considéré si l'espace est isotrope en terme de distance ou alors des lignes d'iso distance si l'espace n'est pas isotrope.

Clementini *et al.* [CDFH97] ont proposé en 1997, de combiner distance et orientation

⁵⁰Une distance (ou métrique) est définie comme suit :

1. $d_{xx} = 0$
2. $d_{xy} \geq 0$
3. $d_{xy} = d_{yx}$
4. $d_{xy} + d_{yz} \geq d_{xz}$

Où $d_{\alpha\beta}$ est la distance entre α et β .

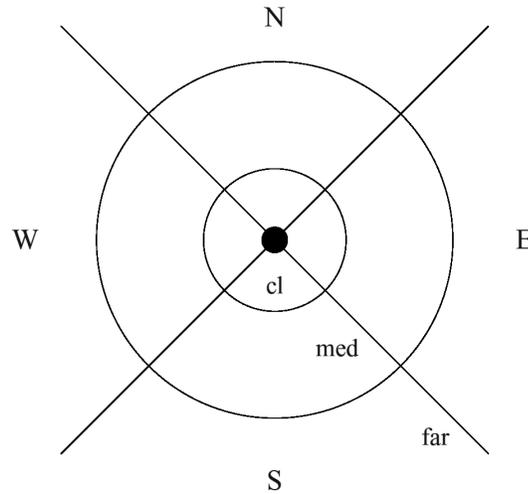


FIG. 2.25 – Représentation combinée de la distance et de l'orientation, d'après [CDFH97] et [Ren02].

dans une même représentation, figure 2.25. Une *segmentation* qualitative de l'espace de cette manière permet de fournir ce qui est appelé une *information positionnelle*. Dans l'exemple de la figure, on voit bien que l'on peut nommer les régions délimitées de cette manière. Par exemple la région proche-est, ou bien la région loin-Sud.

2.5.5 Exemple de raisonnement spatial simple : *Cataglyphis*

Enfin pour terminer nous citons l'exemple de la fourmi *Cataglyphis* (également appelée fourmi du désert) qui démontre, (si besoin était) que le raisonnement qualitatif n'est pas compliqué pour bon nombre de sujets.

Les animaux dotés de petits cerveaux arrivent à intégrer de nombreuses informations de leur environnement, ce qui leur permet d'avoir des capacités de navigation impressionnantes. Ainsi, Rüdiger Wehner [Weh99] a mis en évidence que la fourmi *Cataglyphis* (figure 2.26), qui possède un cerveau de seulement 0.1 mg, est une navigatrice hors pair.



FIG. 2.26 – La fourmi du désert *Cataglyphis* en alerte, photo de Rüdiger Wehner.

Cataglyphis se dirige principalement en intégrant des informations de polarisation de la lumière imperceptibles à l'homme mais également en utilisant des repères et finalement en pointant le but à atteindre. D'autres outils sont utilisés par *Cataglyphis* tels que l'appariement entre des vues panoramiques (type photo) et des vecteurs locaux associés à de telles vues. Des analyses comportementales et neurobiologiques montrent que le traitement de ces informations spatiales est accompli via une intégration coopérative de différents modules de bas niveau. L'information utilisée par la fourmi *Cataglyphis* est traitée et mise à jour continuellement en fonction des perceptions locales qu'elle a de son environnement [Weh03]. Wehner a pu montrer que pour retrouver son chemin (et donc "raisonner" spatialement de manière globale) *Cataglyphis* utilise une connaissance procédurale largement dépendante du contexte (acquisition des connaissances locales) plutôt que d'utiliser des représentations qui embrassent l'ensemble de l'espace.

Deuxième partie

Modèle d'extraction intelligente de données

Chapitre 3

Principes et objets de l'extraction d'objets linéaires

*Il faut commencer par éprouver ce qu'on veut exprimer.
Vincent Van Gogh.*

Ce chapitre introductif, présente tout d'abord le fil conducteur qui a permis d'élaborer le modèle ELECA : Éléments Linéaires Extraits par Calcul Attentionnel. Ensuite, une spécification des entrées et des sorties du modèle ELECA est réalisée. Enfin, la dernière partie de ce chapitre décrit un exemple type qu'il sera possible de suivre au cours des chapitres suivants (4, 5 et 6) consacrés au modèle ELECA, et ceci afin de rendre l'explication plus didactique.

Dans un monde où la gestion de l'urgence prend de plus en plus d'importance, la vitesse de réaction des systèmes s'avère être un élément crucial et central. Dans le cadre de notre projet, il s'agit de faire une interprétation rapide (à la fois précise et exacte) d'images de télédétection. Les exemples récents de catastrophes, comme le tsunami en

Asie du Sud-est ou bien les inondations que la France ou le Québec ont subies au cours de ces dix dernières années soulignent encore, s'il était besoin, l'utilité de la télédétection. De nombreux travaux sont en cours dans ce domaine. Nous renvoyons le lecteur à la thèse de doctorat de Henry [Hen04] consacrée aux systèmes d'information spatiaux pour la gestion du risque d'inondation, dans laquelle il dresse un bilan relativement complet de ce qui se fait à l'heure actuelle pour la gestion d'une catastrophe telle qu'une inondation.

Une mise à jour rapide des systèmes d'information géographique (SIG), en utilisant l'imagerie satellitaire, permet d'organiser la logistique pour la gestion de la crise ; tant pour effectuer un diagnostic que pour organiser les secours ou pour permettre la navigation d'hélicoptères de sauvetage dans la zone cible comme pour notre projet – cf. §1.1.

Le modèle ELECA¹ dont nous allons donner une description au cours des chapitres 4, 5 et 6 répond donc à un objectif d'efficacité. En effet, il n'existe aucun traitement à l'heure actuelle, totalement automatisé pour l'extraction des éléments curvilignes, en particulier sur des images bruitées telles que les images RADAR², qui soit performant [Qua04].

L'interprétation des images de télédétection qui est faite pour ce genre de finalité requiert donc toujours une intervention humaine. Ces interprètes humains, en général des spécialistes ayant une grande expertise dans les images satellitaires, utilisent des techniques qualitatives fondamentalement basées sur leur expérience. La solution que nous proposons dans ce chapitre s'appuie sur des raisonnements cognitifs et qualitatifs.

¹ELECA : *Extraction of Linear Elements by a Cognitive Approach* – Éléments Linéaires Extraits par Calcul Attentionnel.

²Le type d'image utilisé est discuté plus loin dans le chapitre.

Dans la suite de ce chapitre, nous présentons les types de données du modèle ELECA en 3.1, puis nous décrivons un exemple type auquel nous pourrions être confrontés en 3.2.

3.1 Données du modèle ELECA

Le modèle ELECA veut répondre au cas le plus général possible. Comme nous l'avons vu au premier chapitre, la figure 1.4 (page 15), présente dans ses grands traits le but du projet et les entrées et sorties du modèle. Dans cette section nous présentons les spécifications des données du modèle ELECA d'un point de vue conceptuel; les données d'un point de vue technique seront abordées au cours du chapitre 7 consacré à la conception du système logiciel qui implante le modèle; le but de cette section étant de donner une brève description de ce à quoi l'on peut s'attendre.

3.1.1 Données en entrée

Les données qui entrent dans le système que nous proposons sont, d'une part des données de type raster : il s'agit de l'image de télédétection, et, d'autre part des données vectorielles qui sont des données issues de bases de données géographiques.

3.1.1.1 Image de télédétection, données raster

Le modèle proposé doit pouvoir s'appliquer à tout type d'image, en particulier les images où le bruit ou bien les obstructions sont importantes. En revanche, certaines

restrictions s'appliquent. En effet, nous nous concentrons dans ce travail sur les éléments curvilignes, ce qui implique que nous ne nous considérons pas les éléments surfaciques. La largeur des éléments linéaires qui font l'objet de notre recherche est donc d'au plus quelques pixels. En effet, un élément linéaire "trop large" est en fait composé de deux éléments curvilignes qui sont ses bords. Il est possible de s'affranchir de ce problème, mais cela impose des traitements supplémentaires, de type *filtrage* morphologique (voir §2.1.4.1) par exemple, qui ne sont pas l'objet de cette thèse. De nombreux types d'images existent sur le marché. Dans le cas de notre étude, elles doivent représenter une réalité la plus récente possible. Le but n'est pas ici de faire une étude exhaustive des types d'images mais plutôt de montrer à quel genre de données nous pouvons être confrontés et par conséquent pourquoi nous proposons un nouveau modèle.

Dispositifs passifs L'image présentée dans la figure 3.1 est une image du satellite LANDSAT. Le capteur TM de LANDSAT fournit des images pour sept longueurs d'onde différentes. Dans cette image, sont combinées les trois longueurs d'onde du visible, *i.e.* le bleu, de 0,45 à 0,52 μm , le vert, de 0,52 à 0,60 μm , et le rouge, de 0,63 à 0,69 μm . La résolution au sol pour ces bandes est de 30 mètres \times 30 mètres.

L'échelle de cette image permet de faire une bonne cartographie des grandes structures linéaires ; par exemple les grands axes routiers comme celui numéroté 1 sur l'image 3.1. Les grandes structures de transport de courant électrique (les lignes de hautes et très hautes tensions), numérotées 2 sur la figure, sont également facilement visibles. Cependant, en raison de la végétation courte qui existe sous de telles structures, on peut remarquer que bien que larges, leur continuité sur l'image n'est pas totale. En termes algorithmiques, il faudrait ajouter à un suivi de contour classique une composante supplémentaire qui pourrait être une combinaison de la continuité et de l'orientation par exemple. Des travaux dans ce sens ont déjà été effectués mais étaient extrêmement



FIG. 3.1 – Exemple d'image LANDSAT TM (nord de la ville de Québec)

coûteux du point de vue du temps de calcul – cf. [Alq98] et [Fit99]. Lorsque les structures sont plus petites, c'est-à-dire moins larges sur l'image, la difficulté à les identifier devient plus grande. C'est le cas dans notre image pour le réseau routier secondaire (3 sur l'image) qui est visible pour un œil averti mais qui devient nettement plus compliqué à extraire par processus automatisé. Quant au réseau hydrographique, ici numéroté 4,

la difficulté à extraire l'information est vraiment plus grande, d'une part parce que la largeur des cours d'eau est faible, mais aussi à cause du relief et de la couverture végétale. La continuité n'est que rarement très grande.

Ce type de capteur, qui n'utilise pas de dispositif propre pour illuminer la scène est dit *capteur passif*. L'émission enregistrée est alors due principalement à la réflexion de la lumière du soleil sur les cibles ainsi qu'à leur propre rayonnement. Ce rayonnement propre est capté principalement dans l'infrarouge thermique puisqu'il est dû en grande partie à une émission thermique.

Les capteurs passifs possèdent maintenant de très hautes résolutions spatiales au sol. Les capteurs commerciaux (par opposition aux capteurs militaires dont on sait peu de choses) donnent des images d'une qualité et d'une précision spectaculaires ; par exemple le satellite QuickBird propose une résolution de 0,61 mètres au nadir³, on parle dans ce cas d'imagerie à très haute résolution. La démocratisation de ces images est de plus en plus importante. Les télévisions utilisent maintenant souvent ce type de données. Par exemple récemment lors du conflit en Irak, des images de Bagdad ou bien des images du Tsunami en décembre 2005 en Asie du Sud-Est ont été présentées au grand public. Malheureusement ces capteurs sont quasiment aveugles lorsqu'il y a une couverture nuageuse. Les longueurs d'ondes observées ne peuvent traverser ces amas de gouttes d'eau microscopiques en suspension que sont les nuages. Dans ces cas là il faut avoir recours à des satellites équipés de *systèmes actifs*⁴.

³Le *nadir* (selon le Centre Canadien de Télédétection) est le "point ou [l']ensemble de points de la surface du globe, situés directement sous un capteur à mesure que celui-ci se déplace le long de son orbite."

⁴Il faut néanmoins introduire une nuance ici. Pour les raisons citées ce sont plutôt des hyperfréquences qui sont cherchées. Avec des dispositifs passifs il est également possible de traverser la couverture nuageuse, en utilisant les micro-ondes passives par exemple ; mais par manque de signal la résolution est grossière. Les micro-ondes passives sont utilisées pour des observations de l'atmosphère ou bien pour déterminer la température de la surface des océans, la vitesse du vent à la surface des océans, les quantités de pluie ou les caractéristiques de la couverture neigeuse.

Dispositifs actifs Comme nous l'avons vu dans le chapitre premier, l'information de télédétection dont nous avons besoin doit être la plus récente possible afin de mettre à jour la base de donnée. Le Canada est connu pour l'âpreté de son climat hivernal. En présence de neige, de brouillard, de nuages ou d'une combinaison des trois, les seuls capteurs pouvant recueillir de l'information du sol sont alors des systèmes de type RADAR. Ceci est extrêmement intéressant dans le cas qui nous intéresse puisque la réalité du territoire peut-être complètement remodelée par la glace et/ou la neige ; dans notre contexte de pilotage d'hélicoptère en situation de sauvetage il s'avère crucial de connaître cette nouvelle réalité.

Le Canada s'est doté depuis 1995 d'un satellite équipé d'un tel type de système. Il s'agit de RADARSAT I qui utilise la technique du RADAR à Synthèse d'Ouverture (RSO) permettant obtenir des images avec une meilleure résolution géométrique qu'un système RADAR simple. Il est à noter, qu'après bien des reports, le "petit frère" de RADARSAT I, RADARSAT II devrait être lancé cette année (2005⁵) ; la meilleure résolution attendue pour cet autre satellite RSO est de trois mètres. Le but ici, encore une fois, n'est pas de faire une liste exhaustive de toutes les corrections à apporter à une image de télédétection RADAR RSO mais plutôt de souligner quels sont les problèmes d'extraction automatisée auxquels dans notre contexte, nous sommes confrontés. Autrement dit, nous n'aborderons pas ici les corrections géométriques.

L'image 3.2 présente une image RADARSAT de la vallée de la Matapédia – région de la Gaspésie, Québec. La surface noire est la partie la plus occidentale de la Baie des Chaleurs, qui correspond à l'embouchure de la rivière Restigouche. Cette image a une taille de 11327 pixels \times 7538 pixels, la sous image présentée ici fait 660 pixels \times

⁵Au moment d'écrire ces lignes, le 25 août 2005, des sources bien informées mentionnent que le lancement aurait lieu en 2006 au plus tôt, bien que le site web du Centre Canadien de Télédétection l'annonce toujours pour l'année 2005.

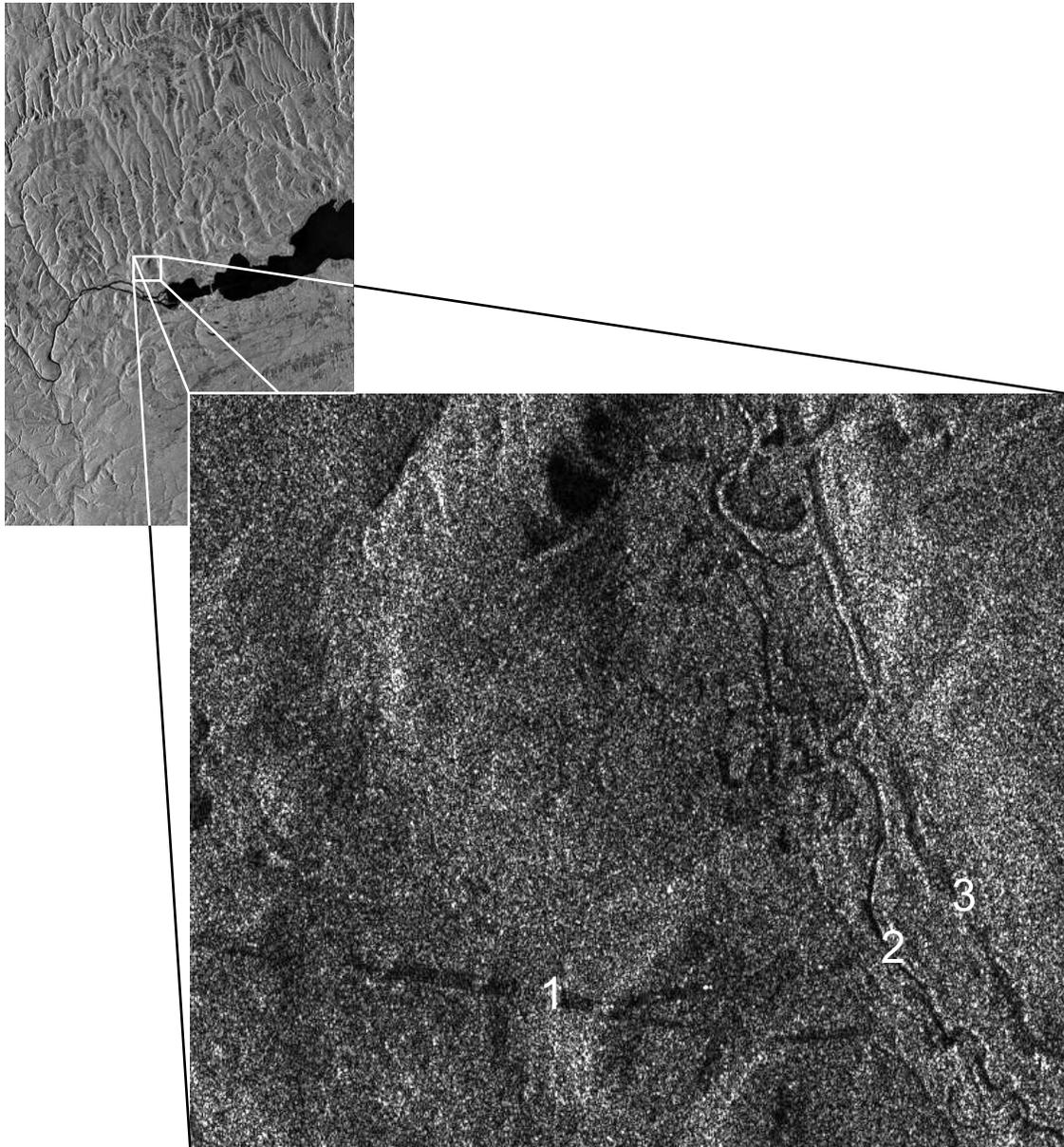


FIG. 3.2 – Exemple d'image RADARSAT I (vallée de la Matépédia, Québec)

565 pixels. La résolution géométrique d'une telle image RADARSAT en mode fin est de 10 mètres \times 10 mètres. L'échelle de la sous image est typiquement celle qui nous intéresse dans notre cas. En effet, il est possible de reconnaître des structures linéaires qui figureront dans la base de données géographiques. Numéroté 1 nous identifions une ligne à haute tension par la rectitude de la forme. Numéroté 2 nous identifions un cours d'eau : le réseau hydrographique et les plans d'eau en général sont facilement

identifiables sur des images RADAR car le signal n'est pas rétrodiffusé vers le capteur (la réflexion est spéculaire); ces objets apparaissent donc noirs. Enfin, la structure numérotée 3 semble être une route qui longe le cours d'eau, la détermination est difficile.

De cet exemple nous pouvons tirer deux enseignements importants :

- Le bruit dans l'image est très important et visible à l'œil.
- Les objets, bien que reconnaissables sur de telles images, sont, en conséquence, plus difficilement identifiables, augmentant ainsi la difficulté à traiter ce type de problème de reconnaissance avec des techniques classiques de traitement d'image.

Le bruit est principalement dû au chatoiement - *speckle*. Le chatoiement est produit par les interférences aléatoires générées pour chaque cellule de résolution [Hen04]. Ces interférences ont pour origine l'infinité de réflecteurs élémentaires qui composent chaque cellule.

La figure 3.3 est une image RSO aéroportée de la forêt Montmorency, site d'étude de l'Université Laval. La résolution de cette image est de 6 mètres \times 6 mètres. Sur une image de très bonne qualité comme celle-ci, le bruit est encore très important. En utilisant l'équation 3.1 établie par Simard [Sim98] [SGTB98], Fitzback a calculé que cette image contient environ 20% de bruit [Fit99] comparativement au signal.

$$\text{Chatoiement} = \frac{1 - \left(\frac{\Gamma(N + \frac{1}{2})}{N^{\frac{1}{2}} \Gamma(N)} \right)^2}{\left(\frac{\Gamma(N + \frac{1}{2})}{N^{\frac{1}{2}} \Gamma(N)} \right)^2} \quad (3.1)$$

Dans cette équation N est le nombre de *visées* employées et Γ est la fonction

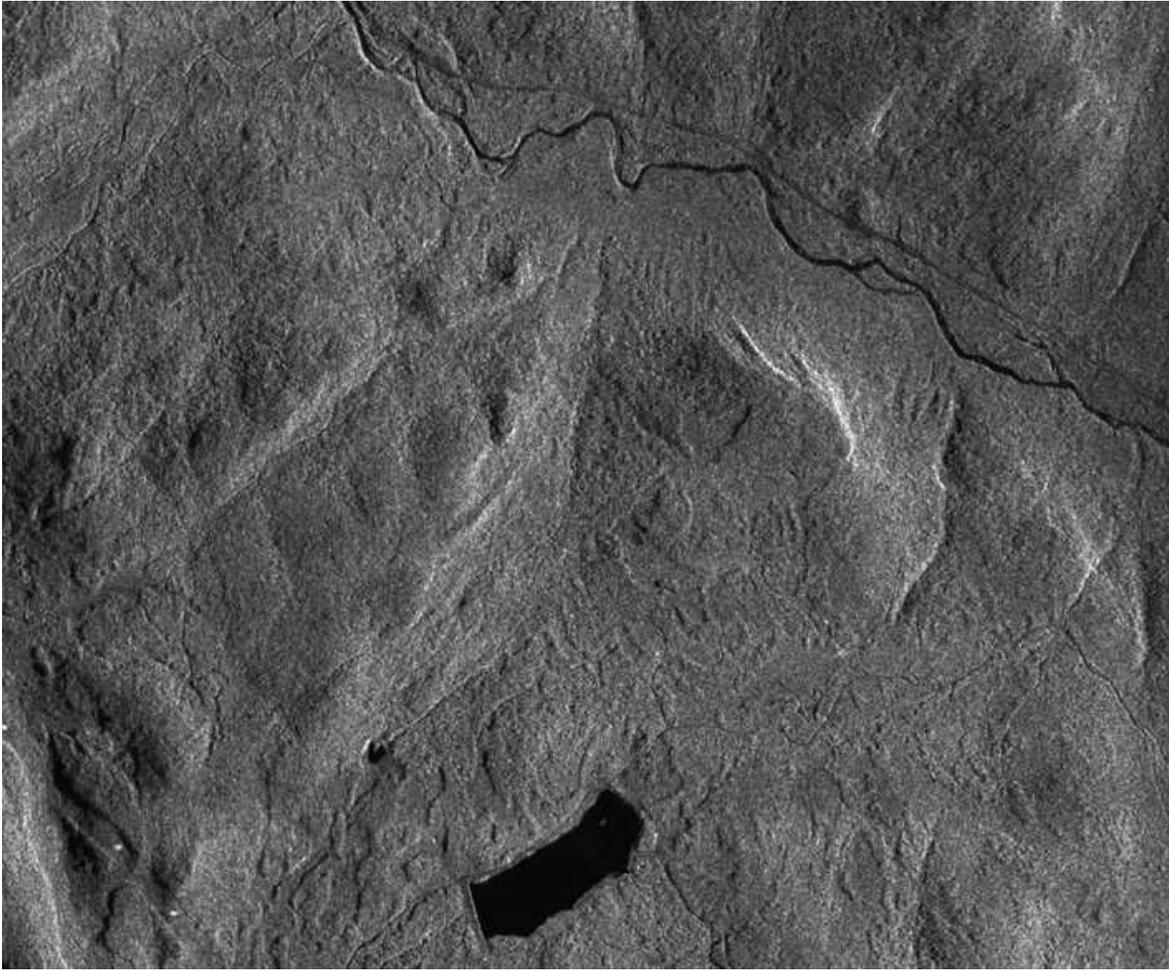


FIG. 3.3 – Exemple d'image RSO aéroportée (forêt de Montmorency, Québec)

Gamma⁶.

⁶La fonction Gamma est définie sur le demi-plan supérieur $\{s \in \mathbb{C} \text{ tel que } \text{Re}(s) > 0\}$ par :

$$\Gamma(s) = \int_0^{+\infty} t^{s-1} e^{-t} dt. \quad (3.2)$$

Dans notre cas les valeurs que peut prendre s sont seulement des réels positifs, dans ce cas on peut écrire l'égalité suivante [Bec05] :

$$\forall x \in \mathbb{R}^+, \Gamma(x+1) = x\Gamma(x) \quad (3.3)$$

Et si on se situe maintenant dans les entiers, il est possible d'écrire :

$$\forall n \in \mathbb{N}^\times, \Gamma(n) = (n-1)! \quad (3.4)$$

Ce qui permet de calculer $\Gamma(x)$ par récursivité puisque $\Gamma(1) = 1$. Pour une démonstration plus complète cf. [Wei05] et [Kre93] ; pour l'implantation informatique se référer à [PTVF02].

Il est donc important d'essayer de réduire le bruit avant toute processus d'extraction. Nous verrons comment cela est réalisé d'un point de vue théorique au chapitre 5 et d'un point de vue pratique au chapitre 8.

3.1.1.2 Informations géographiques, données vectorielles

Les données issues des systèmes d'informations géographiques (SIG) qui nous intéressent sont contenues à proprement parler dans leurs bases de données. Le moteur de représentation et de manipulation de l'information qui constitue l'interface avec l'utilisateur n'est pas la partie du SIG sur lequel porte la présente thèse, l'idée étant de prendre uniquement l'information qui nous intéresse dans la base de données.

Il existe de nombreux types de données dans une base de données géographiques [Géo97]. Les bases de données topographiques, en plus de contenir les données à proprement parler, définissent un ensemble de spécifications et normes comme le montre l'exemple de la base nationale de données topographiques du Canada (BNDT), tableau 3.1. Il faut noter que, depuis quelque temps maintenant, un effort international est fait pour standardiser l'information géographique ; cet effort est chapeauté par l'ISO⁷ et l'OGC⁸ [ISO05]. La très récente Base de Données Géospatiales (BDG) du Canada [Res03] se conforme à ces nouvelles normes.

Pour revenir au tableau 3.1, au niveau *entité ou occurrence d'entité*, dans les *représentations géométriques* on retrouve, entre autres, une description des *points* et des *lignes*. Ce sont les entités qui constituent les objets qui nous intéressent ici. Il s'agit ainsi d'éléments curvilignes (routes, cours d'eau et lignes de haute tension) identifiables

⁷ISO : International Organization for Standardization.

⁸OGC : Open GIS Consortium.

NIVEAU ORGANISATIONNEL	CONTENU
<i>Organisation de la BD</i>	<ul style="list-style-type: none"> – Description générale – Caractéristiques des données – Métadonnées du territoire
<i>Jeu de données</i>	<ul style="list-style-type: none"> – Description générale – Caractéristiques des jeux de données – Métadonnées du jeu de données
<i>Thème</i>	<ul style="list-style-type: none"> – Description générale – Métadonnées associées au thème
<i>Entité et occurrence d'entité</i>	<ul style="list-style-type: none"> – Description détaillée d'une entité : <ul style="list-style-type: none"> - Représentation géométrique - Représentation descriptive – Occurrence d'entité – Dimensions garanties
<i>Relations spatiales</i>	<ul style="list-style-type: none"> – Relation de connexion – Relation de partage – Tolérances de connexion et de partage – Intégration des données sans rel. spat. explicite – Surface d'exclusion altimétrique – Continuité spat. entre jeux de données BNDT

TAB. 3.1 – Exemple d'organisation générale d'une base de données topographiques : la Base Nationale de Données Topographiques du Canada [Géo97]

sur une image de télédétection. Les limites administratives et autres limites virtuelles invisibles sur une image de télédétection ne sont bien entendu pas prises en compte par le modèle ELECA.

Dans notre cas un point est une représentation géométrique composée d'une et d'une seule paire de coordonnées. On pourra, suivant le cas, ajouter de l'information spectrale, de l'information sur l'orientation de la ligne à cet endroit, *etc.* Cela s'approche de la représentation géographique définie par la BNDT. La BDG possède une *classe Point* qui décrit une primitive géométrique de dimension 0. Un point est alors décrit par une

paire de coordonnées x et y .

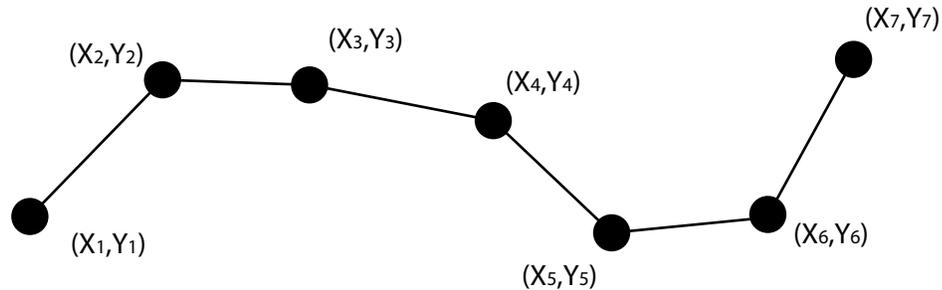


FIG. 3.4 – Ligne d’après [Géo97].

Une ligne dans le cas de la BNDT est une représentation géométrique composée d’une série d’au moins deux paires de coordonnées distinctes liées séquentiellement, d’un qualificatif de représentation géométrique et de métadonnées, figure 3.4.

Une ligne, dans le modèle ELECA est une liste ordonnée de points qui pourra, le cas échéant, avoir de l’information supplémentaire comme sa nature (route, cours d’eau, etc) par exemple. Dans le modèle ELECA de telles entités sont appelés des *éléments curvilignes*. La BDG décrit la *classe ligne* comme une spécialisation de la *classe courbe* qui suit les normes ISO-OGC.

Le document [Res03] de la BDG précise que la ligne utilise une interpolation linéaire entre chaque coordonnée. Il explique également qu’un filtrage est appliqué afin de supprimer les points superflus en utilisant l’algorithme de Douglas-Peucker [DP73]. Nous verrons dans le modèle ELECA quel profit il est possible de tirer de ce filtrage. L’annexe A.1, page 297 décrit avec plus de détail l’algorithme de Douglas-Peucker.

3.1.2 Données en sortie

Les données en sortie du modèle ELECA sont des données purement vectorielles. Ce sont des lignes comme décrites plus haut. Elles représentent les objets curvilignes présents sur l'image de télédétection (données raster en entrée) qui ont été mis à jour. Il n'y a, alors, plus besoin de "convertir" les données raster en vecteur comme c'est souvent le cas en traitement des images de télédétection. Ces données pouvaient exister au préalable dans la base de données (données vectorielles en entrée), elles ont alors été confirmées, modifiées ou supprimées. Ce peut être aussi des données qui n'existaient pas du tout dans la base de données originale.

DONNÉES EN ENTRÉE		DONNÉES EN SORTIE
<i>Raster</i>	<i>Vectorielles</i>	<i>Vectorielles</i>
Images de télédétection en général (une seule image à la fois, une seule bande, géoréférencée et corrigée)	Extraites d'une BD topographique générale	Injectables dans une BD topographique générale
Images pour lesquelles un expert est nécessaire (signal bruité et/ou résolution médiocre)	Représentations géométriques d'entités curvilinéaires tangibles (par opposition à abstraites)	Représentations géométriques d'entités curvilinéaires tangibles (par opposition à abstraites)
Objets éventuellement occultés partiellement	Objets lignes – suite de points ordonnés	Objets lignes – suite de points ordonnés
Dispositif actif principalement	Filtrés par l'algorithme de Douglas-Peucker (ou pouvant l'être)	Filtrés par l'algorithme de Douglas-Peucker
Objets curvilinéaires souvent discontinus		De même nature que les données vectorielles d'entrée

TAB. 3.2 – Spécifications des données du modèle ELECA

3.1.3 Résumé des spécifications des données du modèle

Les spécifications des données d'entrée et de sortie du modèle ELECA sont résumées dans le tableau 3.2. D'une manière générale, le modèle a pour vocation de s'attaquer à des jeux de données pour lesquelles les solutions techniques automatisées classiques échouent et nécessitent l'interprétation d'un expert.

3.2 Description d'un exemple didactique

Afin de comprendre plus facilement la manière dont fonctionne le modèle ELECA nous allons décrire un exemple type. Il expliquera le processus d'application du modèle sur les données.

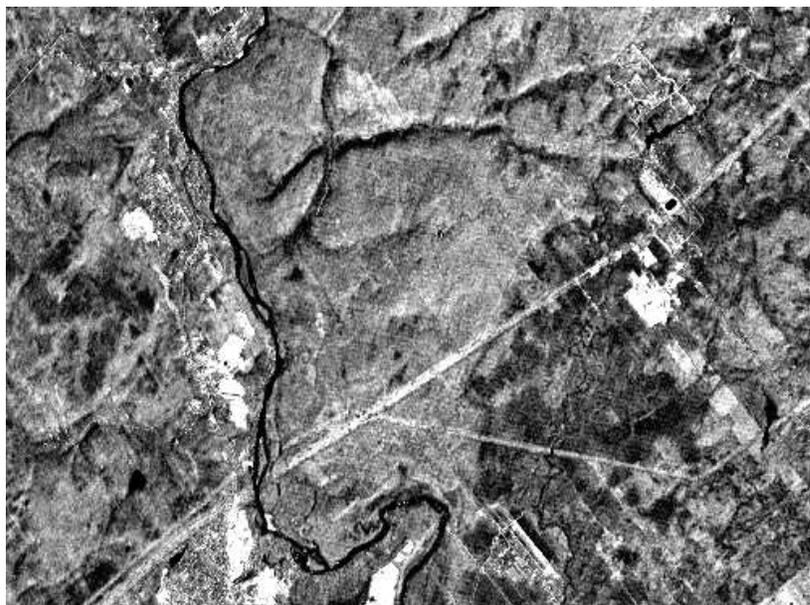


FIG. 3.5 – Exemple de données d'entrée raster.

Prenons l'exemple de la figure 3.5. L'image présentée ici est une image SPOT de la

région de Québec ; en sombre apparaissent les rivières (la plus importante étant la rivière Montmorency) et en clair (portions très rectilignes), les lignes de haute tension. Ces entités sont soulignées sur l'image suivante, figure 3.6. Cette image, qui pour les besoins de l'exemple est considérée comme la plus récente possible, contient toute l'information recherchée. Cette information est de mauvaise qualité aussi bien du point de vue la qualité de l'image que de la résolution vis-à-vis de l'échelle des éléments les plus fins.

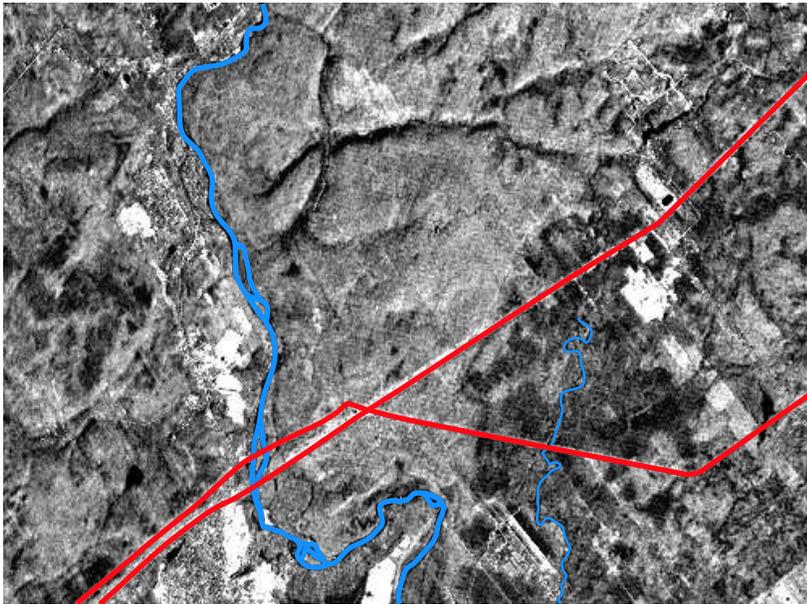


FIG. 3.6 – Exemple d'éléments curvilignes. En bleu les cours d'eau, en rouge les lignes de haute tension.

Sur une telle image l'extraction automatique des éléments linéaires est très difficile en dehors de la rivière principale. L'expertise d'un professionnel est nécessaire pour repérer les discontinuités de la ligne à haute tension au dessus de la rivière. La continuité des lignes à haute tension sur la partie droite de l'image est également difficile à suivre à cause de la texture perpendiculaire engendrée par l'activité agricole. La plus petite rivière, elle est quasi indécélable au milieu de la végétation (avec des niveaux de gris très proches et des occultations nombreuses par la couverture végétale) pour un œil non averti. Enfin toujours sur cette image il est possible d'identifier des routes⁹. Le lecteur

⁹Attention les routes ne sont pas les objets noirs dus au relief également visibles sur cette image.

peut apprécier rapidement la difficulté de l'exercice.

Ainsi avec un exemple trivial comme celui-ci, il est facile d'appréhender la difficulté qu'il existe à fournir des algorithmes robustes et fiables permettant de faire l'extraction sûre et rapide de telles entités. Jusqu'à présent il s'agit toujours d'un problème ouvert. Le modèle ELECA propose une piste de solution originale en utilisant des approches non conventionnelles en télédétection.

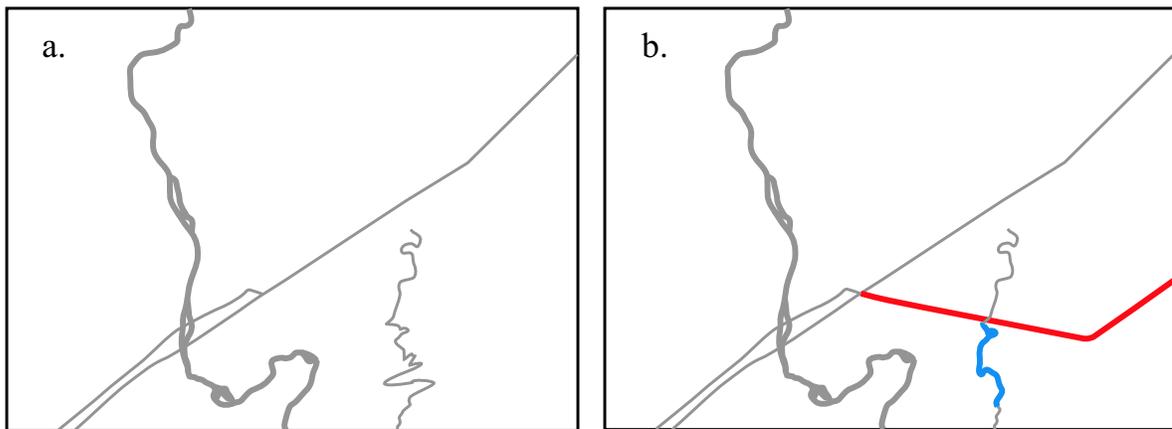


FIG. 3.7 – Exemple de données vectorielles. a. Données d'entrée b. Données de sortie.

La figure 3.7 montre ce que pourraient être les données vectorielles en entrée et en sortie dans notre exemple type. À gauche nous avons la base de données en entrée. Elle n'est pas à jour. À droite cette même base de données après sa mise à jour par le processus du modèle ELECA. En rouge, une ligne de tension qui n'était pas présente dans la base de données, et qui a été ajoutée. En bleu, une des rivières avait été mal cartographiée, elle a été corrigée. Les éléments en gris ont été confirmés. La base de données est alors cohérente avec l'image de télédétection.

Chapitre 4

Guidage du regard virtuel

*Del mirar nace el desear.
Proverbio español.*

Ce chapitre est consacré à l'élaboration d'une stratégie de recherche pertinente des objets cibles au sein de l'image. Dans un premier temps le problème est décrit par une mise en contexte générale puis également dans le cas de l'exemple type. Dans un second temps, les spécifications du problème de recherche efficace au sein de l'image pour le modèle ELECA sont explicitées en détail. Enfin, dans un troisième temps, les solutions originales de guidage d'un regard virtuel proposées par le modèle ELECA sont développées. Ces solutions propres à cette thèse sont des cycles de spécifications-solutions de plus en plus précis. Un bilan récapitulatif est dressé en conclusion de ce chapitre.

4.1 Contexte général

Dans cette partie nous nous attachons à spécifier à un niveau général le problème de la recherche de l'information dans l'image. Le tableau 3.2 page 106 résume les données qui vont entrer (et sortir) du modèle ELECA ; il s'agit de l'image de télédétection et de la base de données géographiques à mettre à jour.

La première étape pour pouvoir extraire des données d'une image, si l'on ne veut pas traiter toute l'image, consiste à mettre au point une technique de focalisation de la recherche sur les zones d'information pertinentes. Il s'agit donc de proposer une stratégie pour guider le *focus* sur la scène visuelle.

La difficulté est double : il faut d'une part extraire l'information, d'autre part il faut le faire de manière "intelligente" et rapide. L'idée est donc de s'appuyer sur les connaissances antérieures que l'on possède sur le territoire grâce à la base de données géographiques. Les bases de données, en général, n'étant pas toujours exactes (problèmes de mise à jour, de géoréférence, saisies erronées, erreurs...); il convient de prendre des précautions.

4.1.1 Objectif

L'objectif de ce chapitre est de proposer une technique efficace pour localiser l'information pertinente (ici les éléments curvilignes) dans la scène visuelle. Le modèle ELECA utilise une approche par *guidage d'un regard virtuel* afin d'éviter un traitement inutile de l'ensemble de l'image.

Cette technique exploratoire doit être en mesure, de plus, de surmonter la difficulté des images bruitées et/ou des objets partiellement occultés. Il faut donc élaborer un modèle de comportement du regard pour que ce dernier parcoure l'image de façon pertinente. Un sous-objectif important à ne jamais perdre de vue concerne l'efficacité vis-à-vis du temps de traitement de notre technique ; c'est pourquoi le modèle ELECA a pour vocation de ne pas explorer l'entièreté de l'image.

4.1.2 Idée directrice

Sachant d'une part, que l'objectif de notre technique est d'être capable d'aller chercher, où elle se trouve, l'information pertinente dans une image de télédétection même si cette information est partiellement masquée ou proche de la limite de résolution ;

et, sachant d'autre part, que les sujets humains n'ont pas de problème particulier à suivre une route partiellement masquée sur une image de télédétection, ceci est possible grâce à l'interprétation que fait le sujet de l'orientation et de l'intensité de l'objet conjointement avec le comportement "saccadique"¹ de couverture de l'image qui permet une bonne intégration des pièces du puzzle ;

l'idée directrice, à cette étape de la recherche, est donc d'étudier si un modèle inspiré du système de vision humain pourrait résoudre le problème qui consiste à guider l'extraction des informations curvilignes dans l'image de manière efficace, ceci même lorsque cette information est bruitée ou difficile d'accès.

¹Par *comportement saccadique* nous entendons un comportement fait de *saccades*. Ce n'est *pas* un comportement saccadé erratique. Ce terme est utilisé depuis longtemps, par exemple Paillard [Pai94] l'utilise en 1994 dans [RRR94]. Cette notion est explicitée plus en détail plus loin dans ce chapitre.

4.2 Exemple type

En reprenant l'exemple type présenté au chapitre 3 et en considérant l'image de la figure 3.5 page 107, l'idée est de focaliser la recherche dans les zones grisées de la figure 4.1. En effet, à cette étape du processus nous cherchons à guider la recherche dans l'image de telle façon qu'elle soit pertinente. Autrement dit, nous voulons qu'il y ait une focalisation de l'attention là où l'information en lien avec les éléments curvilignes se trouve *a priori*.



FIG. 4.1 – Exemple de zone de recherche souhaitée.

Les zones grisées de la figure 4.1 sont bien sûr les zones où la densité d'information est la plus grande, ce qui permet d'améliorer la performance de l'heuristique de recherche. Le but de cette première partie du modèle est d'établir, au sein des zones identifiées, une stratégie de recherche inspirée de la vision humaine, comme le montre la figure 4.2. Autrement dit le but est, d'une part de définir les zones grisées, et, d'autre part de définir une stratégie de recherche au sein de ces zones.

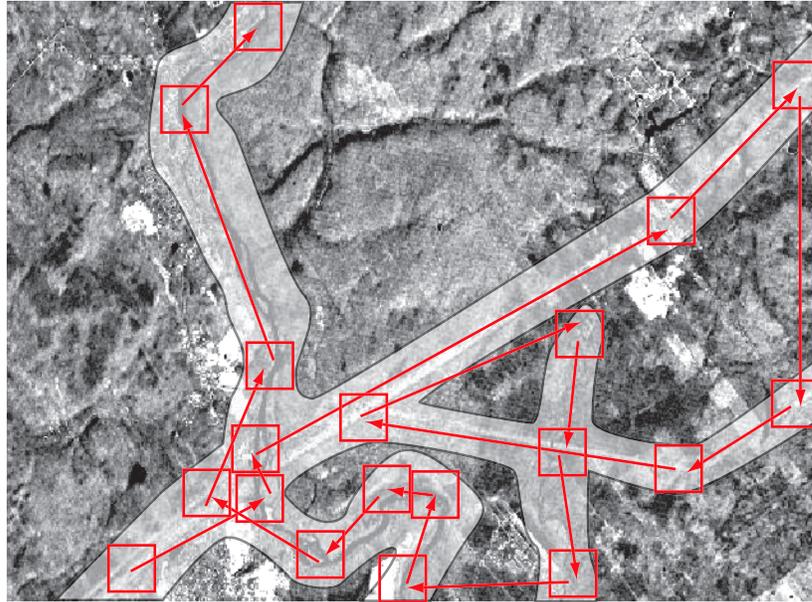


FIG. 4.2 – Exemple de parcours du regard virtuel pour la recherche d'informations pertinentes.

4.3 Modèle d'exploration de l'image

Nous avons vu, au cours du chapitre 2 (consacré à l'état de l'art) que, dans le cas des sujets humains en particulier, la saillance joue un rôle très important pour guider le regard dans une scène visuelle. Le modèle ELECA propose une solution originale qui s'appuie sur cette approche mais en utilisant un axe différent [CMEL03]. Dans le modèle ELECA, la recherche de l'information dans l'image est réalisée en combinant deux notions différentes et complémentaires : la saillance, §4.3.1, et le comportement d'un regard virtuel² dans l'image, §4.3.2.

Le but de l'utilisation de la saillance est d'identifier des zones et des points d'intérêt. L'ordonnement de ces zones et points d'intérêt fournit une carte de saillance implicite. La saillance est calculée principalement à partir de la base de données – voir plus

²Nous verrons que le comportement du regard se fait dans un champ de saillance.

bas. En revanche le comportement du regard virtuel dans l'image repose en plus sur l'information directement extraite de l'image, voir §4.3.2.

Cette façon de faire du modèle ELECA concilie deux avantages principaux : c'est une combinaison des approches *ascendante* et *descendante*. La focalisation par rapport à la base de données est foncièrement descendante puisque c'est une recherche orientée par la tâche ; mais l'orientation des données extraites de l'image, elle, est ascendante. De manière simplifiée les données d'entrées vectorielles permettent une approche descendante alors que les données raster en entrée, elles, autorisent elles une approche ascendante.

Cette approche constitue un élément de solution original de cette thèse. Dans la suite de cette section (§4.3.1), nous verrons comment il est possible de calculer une carte de saillance dans notre cas. La section suivante (§4.3.2) montre une façon originale de bâtir un modèle du regard virtuel.

4.3.1 Saillance

4.3.1.1 Corridors de recherche

Afin de déterminer une première zone de saillance supérieure au fond de l'image (autrement dit plus saillante que le fond de l'image), nous proposons d'utiliser la notion conceptuelle de *corridors de recherche*. Les corridors de recherche constituent par conséquent une solution originale pour faire une première discrimination des zones à saillance positive.

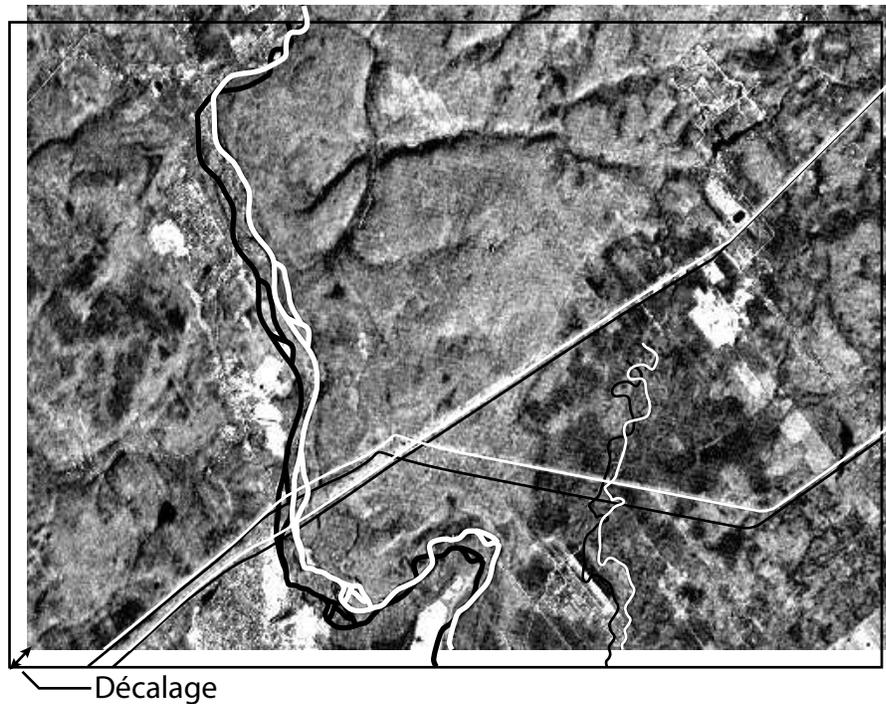


FIG. 4.3 – Illustration d'un problème de géoréférence.

Prenons par exemple le premier problème illustré par l'exemple de la figure 4.3. En blanc est figurée la réalité de terrain de l'image de télédétection, alors que les informations issues de la base de données, en noir, sont mal positionnées vis à vis de l'image. Il s'agit ici d'un cas trivial de décalage que nous montrons pour l'illustration du propos.

Les *corridors de recherche* sont des zones privilégiées de recherche. Ces corridors sont des zones tampon autour des objets de la base de données, comme l'illustre la figure 4.4. On comprend tout de suite l'avantage que l'on peut tirer de cette approche en comparant avec la figure 4.2. Nous utiliserons à partir de maintenant l'expression *base de données* pour parler des éléments pertinents du système d'information géographique comme nous l'avons vu durant la discussion sur les données vectorielles, §3.1. Les corridors de recherche sont une solution présentant de nombreux avantages pour le cas qui nous intéresse :

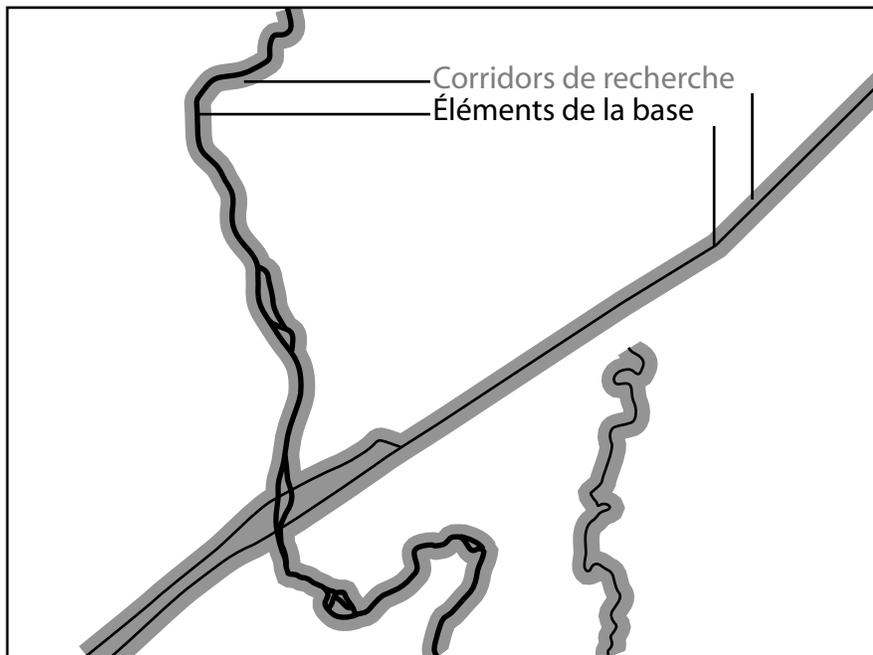


FIG. 4.4 – Corridors de recherche.

- Ces corridors permettent de focaliser la recherche sur des zones où l’information est susceptible de se trouver. C’est par conséquent une manière simple, mais très efficace de guider le regard virtuel ; ceci permet au modèle ELECA de suivre les objets tout comme le fait le regard humain, voir plus loin.
- La focalisation de l’investigation sur une zone restreinte comme ces corridors permet également d’augmenter l’efficacité de la recherche.
- Les corridors permettent également de dégager facilement la *signature spectrale* des éléments qui s’y trouvent. Il est alors relativement aisé de faire la correspondance entre le type d’objet dans la base de données et sa signature spectrale dans l’image³.
- Enfin, si l’on se réfère à l’exemple de la figure 4.3, il est intéressant de noter qu’en fonction des premières découvertes au sein des corridors, il est aisé de faire un re-

³Nous verrons plus avant dans le chapitre l’avantage que nous pouvons tirer de la caractérisation des signatures des éléments.

centrage de la zone tampon si l'on constate une transformation systématique⁴. De cette manière, la probabilité que l'information se situe dans le corridor augmente significativement.

4.3.1.2 Les modes du modèle

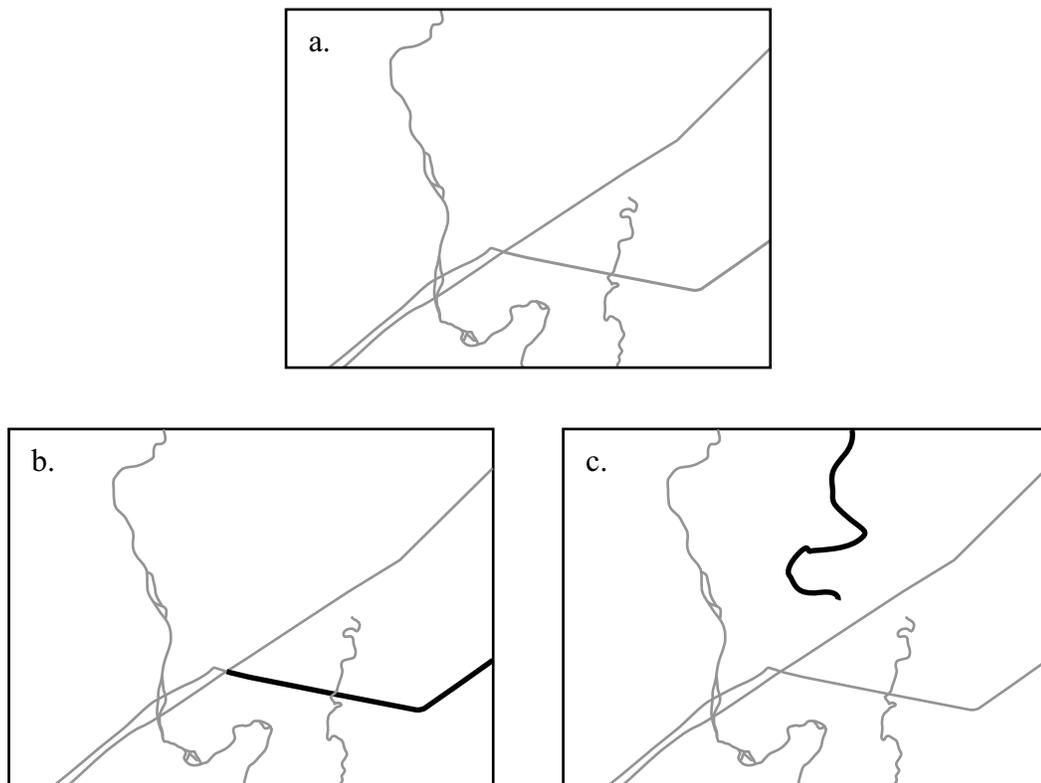


FIG. 4.5 – Les deux modes du modèle. a. Base de données antérieure. b. Mode reconnaissance : La cible est *déjà* présente dans la base de donnée. c. Mode découverte : la cible *n'existe pas* dans la base de données.

À cette étape il convient de faire la distinction entre deux situations distinctes. Le modèle ELECA utilise deux modes différents afin de pouvoir localiser les objets cibles dans l'image en fonction de leur présence ou non dans la base de données originale, voir figure 4.5. Le premier mode (figure 4.5b) est appelé *mode reconnaissance* alors que le

⁴Typiquement dans notre exemple, il s'agirait de faire une translation de l'ensemble.

second (figure 4.5c) est appelé *mode découverte*.

Mode reconnaissance L'objectif du mode de reconnaissance est de reconnaître les éléments contenus dans la base de données et de les valider (vérifier s'ils existent encore et sont situés à la bonne place – figure 4.5b) ou de les invalider – ils peuvent exister mais ne sont pas localisés correctement ou ils n'existent pas.

C'est dans le mode reconnaissance que s'effectue la recherche au sein des corridors de recherche.

Mode découverte L'objectif du mode découverte est de rechercher l'information qui serait en dehors des corridors de recherche – figure 4.5c. Par essence ce sont des informations qui ne sont pas dans la base de données de départ.

4.3.1.3 Liste de priorités

Une fois définies les zones de saillance supérieures que sont les corridors de recherche vis-à-vis de l'ensemble de l'image, il faut ensuite identifier des points ou des zones encore plus saillants au sein de ces corridors afin de diriger le regard virtuel. Dans le cas du mode découverte ces nouveaux points seront en dehors des corridors. La suite ordonnée de ces points est appelée *liste de priorités*. Cette liste de priorités est une solution propre au modèle ELECA.

Cette suite de points va du plus saillant au moins saillant des points identifiés sur l'image. La suite de cette section explique en détail son élaboration.

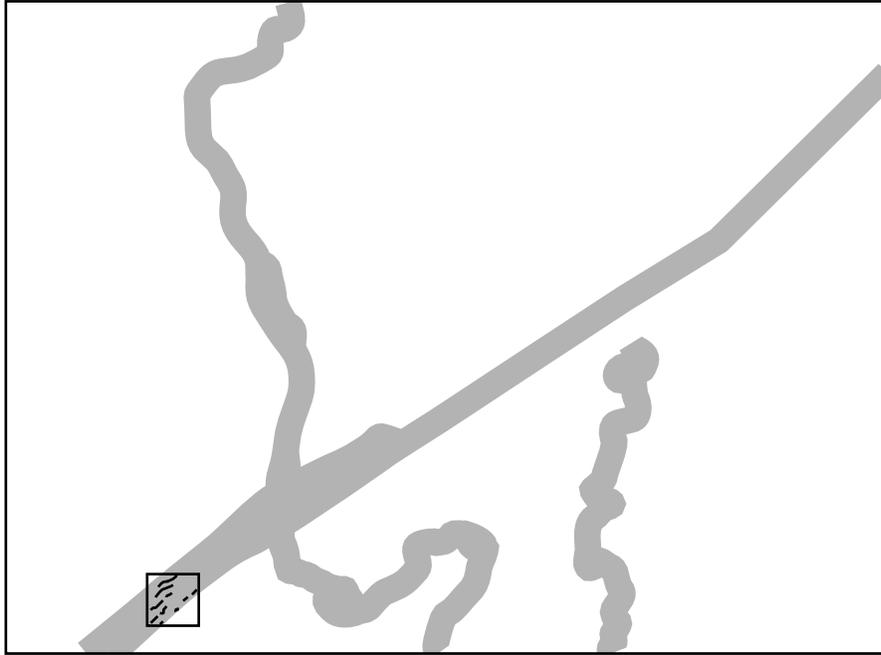


FIG. 4.6 – Première extraction.

Point de départ, sélection et priorité Afin d'extraire les éléments curvilignes de l'image le modèle utilise une petite fenêtre centrée sur une zone du corridor, qui s'apparente à une *fixation* de l'œil. Le traitement que nous effectuons sur cette petite fenêtre, et qui permet d'extraire les *arêtes* est discuté dans le chapitre suivant – *extraction de l'information locale*. La localisation de la première fenêtre est un point critique du début du processus, figure 4.6. Il conditionne une partie de l'heuristique de recherche sur l'image. Cette heuristique repose pour moitié sur la liste des priorités⁵. Pour déterminer la priorité d'un point (sa saillance), un critère de priorité est utilisé. Ce critère dépend du mode dans lequel se trouve la recherche.

Critère de priorité Une zone plus saillante autour des éléments de la base de données a été déterminée, ce sont les corridors de recherche. Dès lors il convient de mettre ces

⁵L'autre moitié étant le modèle de comportement du regard qui est abordé plus bas.

éléments dans un ordre de priorité *ad hoc*. Il faut donc construire un critère de priorité⁶ qui permette de réaliser cet ordre. Ce critère de priorité est évidemment fonction de la stratégie de recherche dans laquelle nous nous situons : soit en reconnaissance d'objets déjà présents dans la base, soit en découverte d'objets qui n'y sont pas.

L'enchaînement de spécifications-solutions pour le calcul de la saillance du modèle ELECA constitue une solution originale de la thèse. La saillance est contrainte par les corridors de recherche ; les corridors de recherche sont ordonnés par la liste des priorités ; la liste des priorités est calculée en utilisant un critère de priorité.

Détermination du critère de priorité pour le mode reconnaissance

Dans le cas de la reconnaissance, comme dans les cas de la figure 4.7, on utilise un critère qui est fonction de la facilité à reconnaître l'élément. En effet, la façon la plus logique de commencer est de reconnaître les éléments faciles à extraire de l'image et d'aller ensuite vers les éléments les plus compliqués. La valeur de saillance maximale est attribuée alors aux éléments les plus rectilignes, c'est-à-

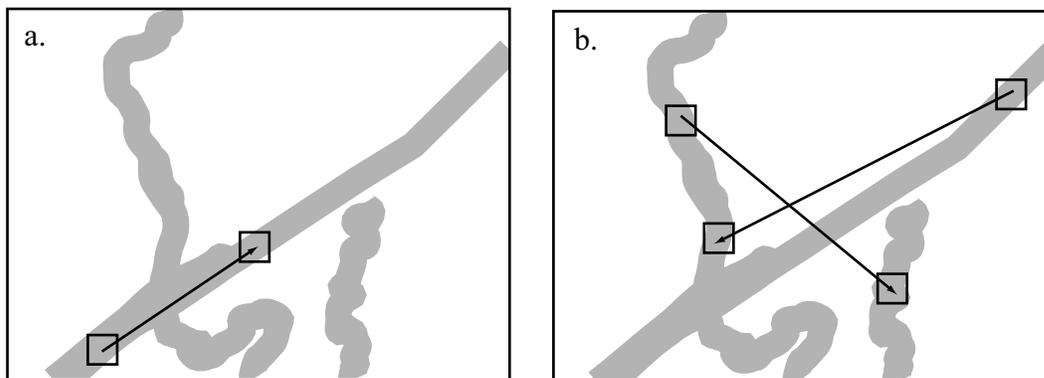


FIG. 4.7 – Différents types de déplacement au sein des corridors de recherche – mode reconnaissance.

⁶Nous utilisons ici le mot *priorité* plutôt que *priorisation* car ce dernier est un néologisme. Il faut bien comprendre que le critère de priorité est un critère qui permet de “calculer” une priorité mais n’est pas la priorité en elle-même. Les priorités en tant que telles se retrouveront dans la liste des priorités déjà évoquée plus haut et que nous verrons en détail un peu plus loin.

dire que ces éléments sont placés en tête de la liste des priorités. Ainsi de manière générale, nous avons tout d'abord les lignes hautes tensions, ensuite les routes et enfin les cours d'eau⁷. Ce critère de rectitude permet non seulement de distinguer les différents objets, mais présente aussi l'avantage d'identifier un point de fixation au sein même de l'objet ; autrement dit, quelle est la partie la plus rectiligne de l'objet le plus rectiligne ? Il faut noter que ce souci de simplification qui consiste à reconnaître les éléments les plus rectilignes en premier lieu ne suit pas exactement le comportement naturel de l'œil qui a tendance à se diriger vers les intersections d'abord (c'est-à-dire des zones un peu plus complexes) et ensuite, à aller vers les éléments rectilignes. Évidemment, dans notre cas, la sélection des intersections en premier lieu reviendrait à compliquer de beaucoup la tâche d'extraction.

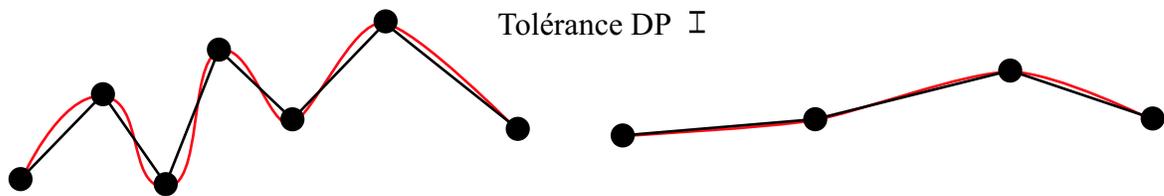


FIG. 4.8 – Différences de densités linéaires en fonction de la complexité de la courbe.

Nous avons vu précédemment que les éléments linéaires des bases de données géographiques étaient simplifiées par l'algorithme de Douglas-Peucker, cf. annexe A.1. La conséquence de ce filtrage par l'algorithme de Douglas-Peucker est la suppression de points superflus en fonction de la tolérance acceptée : ces points étaient inutiles pour décrire l'objet linéaire.

Prenons comme exemple la figure 4.8. En rouge apparaissent les objets que l'on veut représenter dans la base de données. Les points qui forment les lignes et qui sont stockés dans la base de données sont en noir. Ces points sont donc l'information à laquelle nous avons accès. Sur cet exemple les points superflus ont été filtrés en utilisant l'algorithme de Douglas-Peucker ; la tolérance utilisée

⁷Ou bien les cours d'eau d'abord et les routes ensuite s'il s'agit d'un terrain très montagneux par exemple.

est précisée. La densité linéaire de points le long de l'objet de gauche est bien plus importante que pour l'objet à droite. La densité linéaire de points le long de l'élément (après filtrage par l'algorithme de Douglas-Peucker) nous fournit donc une métrique simple et très intéressante pour quantifier la complexité d'un objet linéaire. Cela permet également au sein même d'un objet d'identifier les parties plus ou moins complexes de l'objet. Dans le cas de la reconnaissance, le critère de priorité que nous utilisons pour construire la liste des priorités est la densité linéaire de points pour les éléments de la base. Plus la densité est basse, plus l'élément se trouve haut placé dans la liste des priorités (plus sa saillance est élevée), conformément à l'approche énoncée plus haut.

Détermination du critère de priorité pour le mode découverte

Si nous nous situons maintenant dans le cas de la découverte (soit en dehors des corridors de recherche), figure 4.5c, l'heuristique de recherche repose évidemment moins sur la base de données géographiques que dans le cas de la reconnaissance. Néanmoins, un critère de priorité est construit en tenant compte des éléments curvilignes extraits de la base. En effet, les éléments curvilignes ont certaines propriétés "naturelles" desquelles nous pouvons tirer profit. Les routes, les cours d'eau et les lignes haute tension ne se créent pas de manière aléatoire. Tous ces éléments sont des parties d'un réseau [DLN94] et par conséquent y sont rattachés d'une manière ou d'une autre⁸.

La figure 4.9 illustre les deux cas possibles pour la présence de nouveaux éléments, c'est-à-dire des éléments qui ne sont pas dans la base de données originale. Dans le premier cas le point de connexion de l'élément avec le reste du réseau est *dans* l'image, alors que dans le second cas le point de connexion est *hors* de l'image. En résumé (et en caricaturant quelque peu), un élément curviligne inconnu ne peut prendre naissance qu'à partir d'un autre élément curviligne de l'image ou à partir

⁸Exceptés les cas très peu fréquents de sorties de tunnels, de routes forestières abandonnées, de pistes d'atterrissage d'un aéroport...

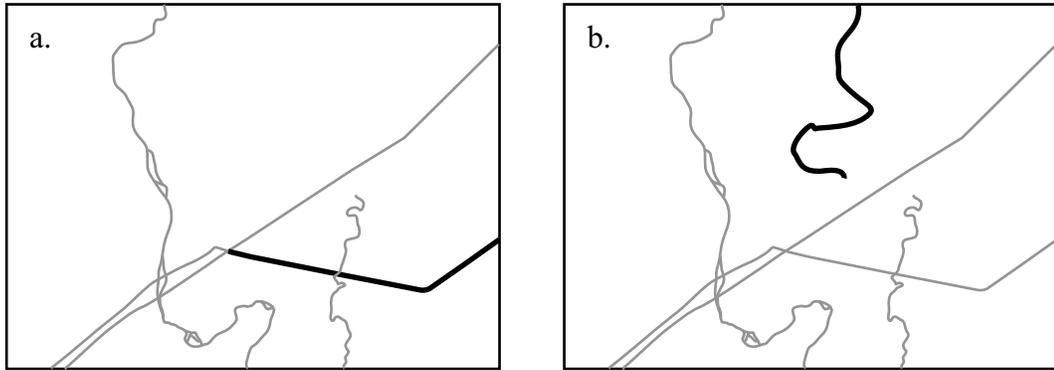


FIG. 4.9 – Cas possibles de nouveaux éléments.

d'un bord de cette même image. Il faut noter que l'élément inconnu peut-être lui-même issu d'un élément qui était inconnu à l'origine dans la base, et qui donc vient d'être reconnu – cf. figure 4.10. La construction de la liste de priorités est par conséquent en constante évolution par rétroaction et mise à jour permanente.

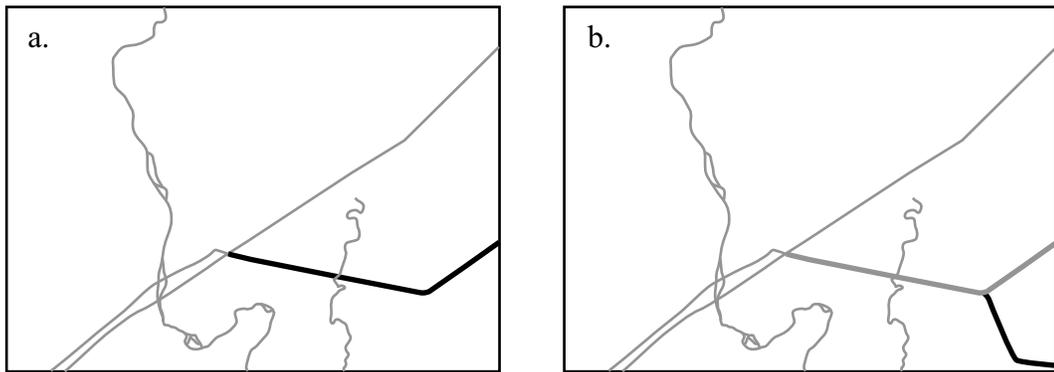


FIG. 4.10 – Nouvel élément issu d'un nouvel élément. À gauche, un nouvel élément est découvert, en noir gras; ensuite, à droite connecté à ce nouvel élément, un (second) nouvel élément est découvert, en noir gras.

Le mode découverte suit le mode reconnaissance, c'est-à-dire que l'image est déjà segmentée par les objets reconnus lors de la phase de reconnaissance. Ainsi, il est facile de partitionner l'image en différentes zones – figure 4.11. L'objectif de ce mode est de découvrir des objets qui n'étaient pas précédemment dans la base de données géographiques, c'est-à-dire les éléments curvilignes autres que ceux reconnus lors de la phase précédente. L'exploration est dès lors située en dehors

des corridors de recherche, autrement dit dans les zones hachurées de la figure 4.11.



FIG. 4.11 – Exemple de subdivision de l’image suite au mode reconnaissance.

Stratégie de couverture de l’image

La stratégie de couverture de l’image dans le mode découverte utilise également une liste de priorités, mais sa construction diffère de celle effectuée en mode reconnaissance et dépend de la stratégie de l’utilisateur. Le but est de “remplir” la liste de priorités afin que l’exploration soit la plus efficace possible. En d’autres termes, il faut un minimum de points dans la liste pour un maximum d’objets trouvés dans l’image. La stratégie que nous avons adoptée est la suivante : la stratégie d’exploration dépend de l’ordre dans lequel sont entrées les différentes catégories de points dans la liste. Ainsi, la solution originale du modèle ELECA est d’introduire les types de points suivants en fonction de leur ordre dans la liste numérotée ci-dessous :

1. Les points qui correspondent à des naissances/amorces (que ce soit de routes, de rivières ou de lignes de haute-tension) étiquetés durant le mode de recon-

naissance. Ces points sont conservés durant le mode reconnaissance dans une liste appelée *liste des découvertes* ; une description plus complète de cette liste est donnée un peu plus bas, page 128. Cela permet au processus de se focaliser directement sur des objets qui sont très certainement présents dans l'image, voir l'amorce rouge sur la figure 4.12.

2. Les points localisés en périphérie de l'image. En partant du constat que la très grande majorité des éléments curvilignes font partie d'un réseau [DLN94], même pour des objets partiellement masqués, la probabilité de trouver les éléments en périphérie est grande— zone verte sur la figure 4.12.
3. Les points choisis de manière aléatoire dans les zones hachurées. Dans des images partiellement masquées (nuages, végétation. . .) ou quand la taille des objets est trop proche de la limite de résolution, certaines amorces d'éléments peuvent être manquées par les deux ensembles de points précédents. Les éléments qui ne font pas partie d'un réseau, tels que routes forestières abandonnées, les sorties de tunnels, *etc*, doivent également être pris en compte. En conséquence de quoi, à l'intérieur des zones précédemment délimitées des points sont choisis aléatoirement pour atteindre une densité de réalisation prédéterminée par l'utilisateur, voir les croix bleutées sur la figure 4.12. Une technique proche a déjà été utilisée avec les *Lévy flights*, voir [BG00] et §2.3.3.3 page 63.

Le regard virtuel balaye la liste des points qui peut être elle-même continuellement mise à jour. Une fois la liste vide, le processus s'arrête mais une autre passe (mode découverte) peut-être refaite si par exemple de nouvelles amorces ont été trouvées.

Identification des points de fixation et construction de la liste des priorités

Grâce aux critères de priorité élaborés précédemment pour chacun des modes, une

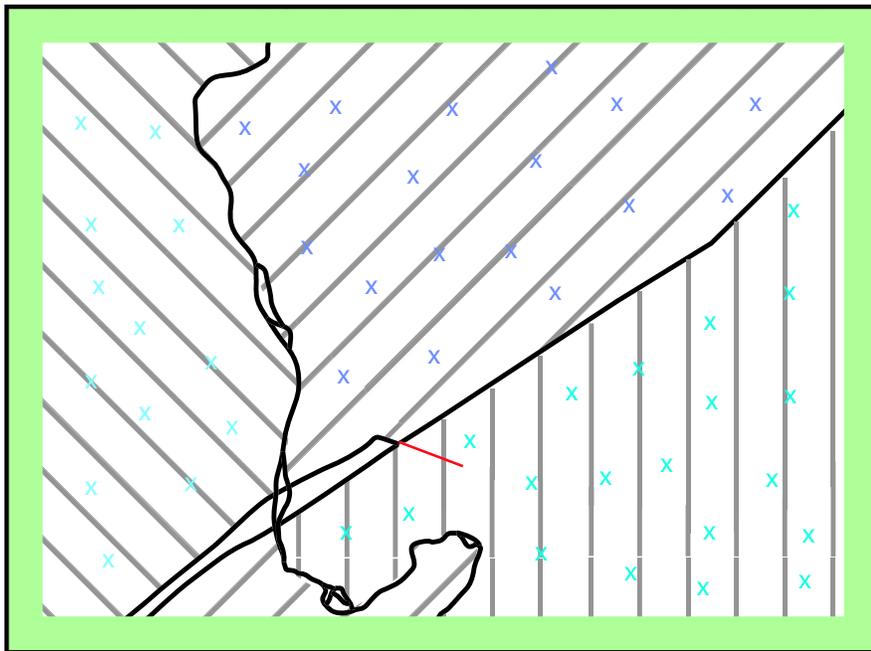


FIG. 4.12 – Stratégies en mode découverte.

identification des points de fixation au sein de l'image est menée. Cette identification permet alors de dresser une liste de priorités des lieux que le regard doit visiter, ce qui revient à affecter une saillance à chacun des points de la liste, mais dans un sens relatif plutôt qu'absolu.

Cette liste est d'abord dressée pour le mode reconnaissance. Quand celui-ci prend fin, la liste est de nouveau générée pour le mode découverte en suivant la stratégie évoquée plus haut. Cette liste s'ordonne encore une fois grâce au critère de priorité mais également en tenant compte, si nous ne sommes pas dans le cas d'une première fixation (il n'y a pas eu encore de découvertes), des découvertes qui ont été faites.

Liste des découvertes Les éléments reconnus et extraits durant les différents modes sont “stockés” dans une liste des découvertes⁹ qui constitue une solution originale du modèle. Le mécanisme de liste des découvertes existe pour quatre raisons principales. Premièrement, il est nécessaire de tenir une comptabilité de ce qui a été découvert. Deuxièmement, tout comme pour l’inhibition de retour, il faut savoir par où le regard est passé pour modifier en conséquence la liste des priorités. Troisièmement, il faut avoir la liste des amorces découvertes qui servent à la création de la liste des priorités pour le mode découverte. Quatrièmement, certaines caractéristiques des éléments aident à augmenter l’efficacité de la recherche – voir plus bas. Techniquement, nous verrons au cours du chapitre 7 qu’il s’agit d’une simple liste de pointeurs. La mise en œuvre est donc très simple.

L’information stockée comprend :

- la position de l’élément, c’est-à-dire la position de chacun des points qui le composent ;
- des caractéristiques de l’élément de type métadonnées ; en particulier leur nature (route, cours d’eau...);
- des caractéristiques de type vectoriel : l’extension et l’orientation de l’objet, l’orientation de l’élément en chaque point ;
- des caractéristiques de l’élément d’un point de vue raster, en particulier la signature spectrale de l’élément sur l’image (dans le cas du RADAR, l’intensité) ; cela permet lors de l’exploration en mode découverte d’avoir un indice supplémentaire, en fonction de l’intensité et de la nature de l’élément ciblé, pour identifier l’information contenue dans l’image ;
- des caractéristiques spécifiques et structurelles qui viennent appuyer la fusion

⁹Ici le terme *découvertes* est entendu dans le sens *vues dans l’image*. Autrement dit ce sont aussi bien des éléments qui n’étaient pas dans la base de données que des éléments qui y étaient. Ou encore, ce sont tous les éléments extraits de l’image à un moment donné.

local/global, en particulier la succession des vecteurs saccades.

Il est à noter que la fusion des éléments est différente en fonction des modes :

- Dans le cas du mode reconnaissance, le but est de confirmer/modifier/supprimer des éléments déjà présents dans la base de données en utilisant l'information de l'image. Les éléments sont donc suivis, un par un, et mis à un jour au fur et à mesure.
- Dans le cas du mode découverte tous les petits morceaux d'éléments sont mémorisés, et la fusion de ces petits éléments en une structure globale n'intervient qu'ensuite. Cette fusion fait l'objet du chapitre 6. Elle s'appuie sur des approches héritées du raisonnement spatial qualitatif.

4.3.2 Comportement du regard virtuel

Une fois établie la liste des priorités, dans quelque mode que ce soit, et contrairement aux modèles déjà existants (Koch, Ullman et Itti par exemple), l'image n'est pas balayée uniquement en suivant de manière décroissante la liste (les points de saillance maximale), mais l'exploration de l'image suit également une contrainte introduite par un modèle de *comportement de regard virtuel*.

La manière de calculer la saillance comme le fait notre modèle est en soi une contribution originale de cette thèse. La combinaison de cette saillance avec un modèle de comportement du regard pour balayer l'image constitue une deuxième originalité du modèle ELECA.

Ainsi le rôle de la saillance, qui se manifeste sous la forme des zones privilégiées et des points de la liste des priorités (*approche descendante*), est modulé par le comportement

du regard virtuel (*approche ascendante*) : le modèle de **saccades**. Cette modulation peut-être également vue comme faisant partie intégrante du processus de détermination de la saillance, ce processus est par conséquent complètement dynamique.

4.3.2.1 Saccades du regard virtuel

Dans le modèle ELECA le déplacement du regard virtuel sur l'image présente un *comportement saccadique*. Cette approche permet d'éviter le traitement entier de l'image. Elle permet également de diminuer la difficulté de suivre des éléments partiellement occultés sur l'image.

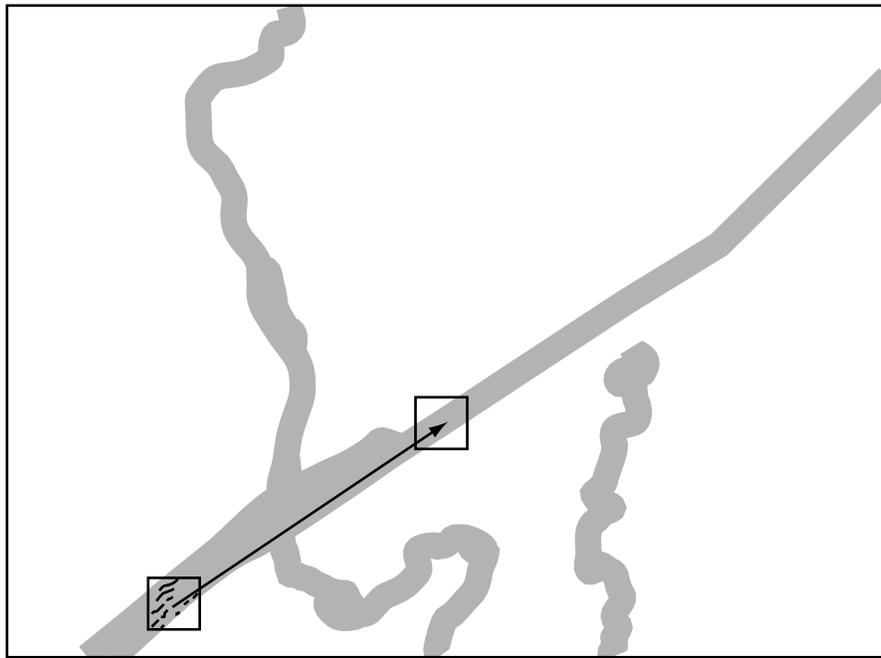


FIG. 4.13 – Exemple d'une première saccade.

Le *modèle de saccades* permet d'effectuer un saut de regard d'un point à l'autre de l'image comme illustré la figure 4.13, à l'image du regard naturel. Ainsi des séries de saccades sont réalisées afin de parcourir l'image. En chaque point de fixation, une

extraction des éléments curvilignes (s'ils existent) est réalisée, voir chapitre suivant, *extraction de l'information locale*. Chaque saccade influence la saccade suivante. Une saccade est ainsi déterminée en répondant à deux critères qui sont d'un côté la liste des priorités, et de l'autre le modèle de saccades à proprement parler.

4.3.2.2 Regards dans l'image

C'est à cette étape qu'est vraiment décidée la zone de l'image où sera effectuée la prochaine fixation. Dans le cas où il s'agit de la toute première fixation dans l'image, le modèle de saccades n'intervient pas. Ce sera alors le premier point identifié dans la liste des priorités, c'est-à-dire, dans le cas du mode reconnaissance, comme nous l'avons dit *le point* qui satisfait le mieux le critère de priorité au sein de *l'élément* qui satisfait le mieux le critère de priorité; sinon le premier point de fixation sera calculé en prenant en compte la liste des priorités *et* le modèle de saccades.

4.3.2.3 Fixation et inhibition de retour

Comme nous l'avons écrit précédemment, afin de ne pas retourner sur une zone où le calcul a déjà été effectué, la zone est étiquetée comme cela est fait pour l'inhibition de retour, cf. figure 2.19 page 70. Ce mécanisme peut naturellement influencer la liste des priorités, en particulier si la zone étiquetée en fait partie.

Nous avons vu que l'orientation du regard au sein de l'image reposait sur la liste des priorités mais également sur le modèle de saccades, lui-même dépendant de la stratégie de recherche. Les deux stratégies de recherche sont d'une part la recherche dans l'image

des éléments présents dans la base de données et d'autre part la découverte dans l'image de nouveaux éléments.

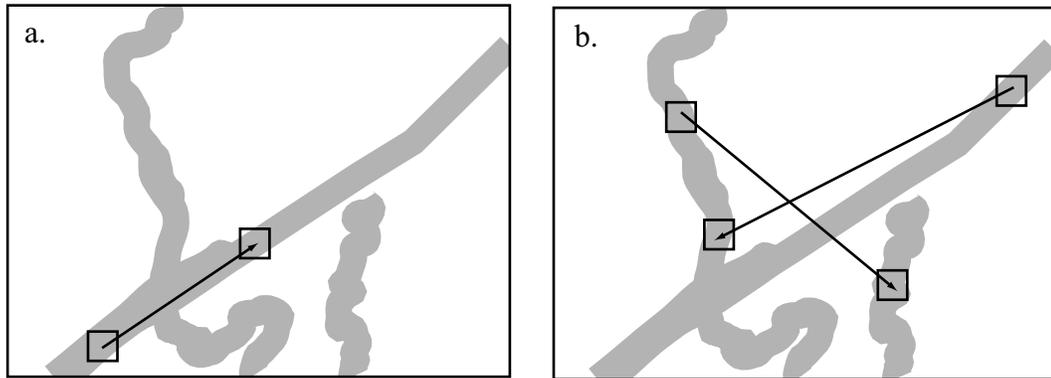


FIG. 4.14 – Deux types de saccades au sein des corridors de recherche

Mode reconnaissance Dans cette situation la recherche consiste à retrouver des éléments de la base de données au sein de l'image. Deux types de situations peuvent se présenter (figure 4.14) dans le cas où la recherche se fait à l'intérieur des corridors de recherche.

1. Dans un premier cas (figure 4.14a), la saccade se fait sur le même élément. L'orientation du regard est principalement guidée par le modèle de saccades et la connaissance de la localisation de l'objet dans la base.
2. Dans le second cas (figure 4.14b), la saccade se fait d'un élément à un autre. L'orientation du regard n'est plus du tout dirigée par le modèle de saccades mais par la liste de priorités.

Mode découverte Dans cette situation la recherche au sein de l'image porte sur des éléments qui ne seraient pas dans la base de données mais présents dans l'image. Deux

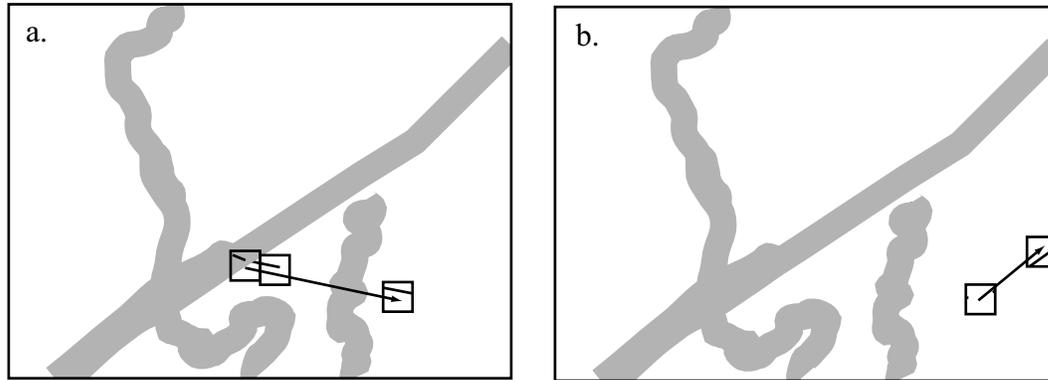


FIG. 4.15 – Deux types de saccades hors des corridors de recherche

types de situation peuvent se présenter (figure 4.15) dans le cas où la recherche se fait en dehors des corridors de recherche.

Dans un premier cas (figure 4.15a), la saccade se fait depuis le corridor vers une zone en dehors des corridors. Par exemple ici, une orientation a été trouvée à l'intérieur du corridor (présence d'une jonction) qui ne correspond pas avec l'élément de la base de données. Dans ce cas de figure l'orientation du regard sera guidée par le modèle de saccades qui détermine le comportement pour des sauts.

Dans le second cas (2 figure 4.15), la saccade part de l'extérieur des corridors vers une zone également en dehors des corridors. Dans cette configuration, tout comme pour la précédente, c'est le comportement saccadique du regard spécifié par le modèle de saccades qui détermine le saut du regard virtuel.

4.3.2.4 Modèle de saccades : comportement saccadique du regard

Comme nous l'avons vu précédemment, partant des listes de priorités pour chaque mode, la couverture de l'image se fait en suivant certaines contraintes : d'un côté la liste de priorités fait office de carte de saillance et donc est accédée dans l'ordre décroissant de saillance ; de l'autre côté, une contrainte supplémentaire est introduite de telle façon que le comportement du regard virtuel présente un comportement saccadique.

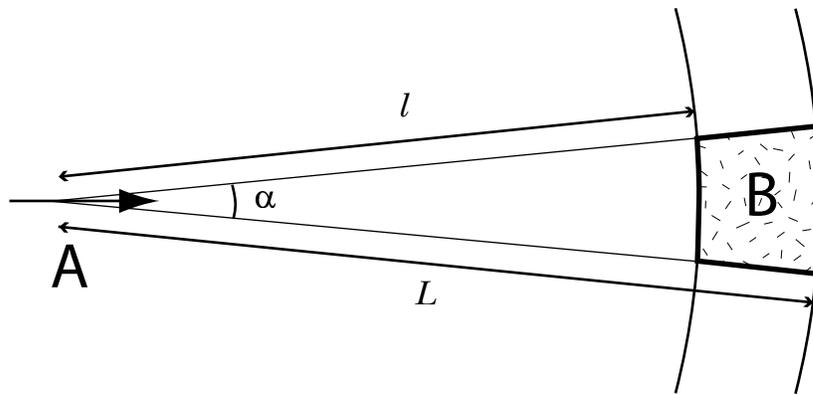


FIG. 4.16 – Zone de saccade. A Point de départ et direction de saut. B Zone cible de saccade.

Ce comportement qui constitue également une solution originale du modèle ELECA est déterminé par l'orientation (et l'intensité) de l'objet au point de départ du saut — point A figure 4.16. À partir de ce point, et, en suivant l'orientation de l'objet, une zone est déterminée (zone B figure 4.16) dans laquelle la prochaine saccade (le prochain saut) doit être localisée. Comme cela est montré dans la figure 4.16, la zone est déterminée par trois paramètres : α , l et L^{10} . Ces paramètres peuvent être calculés pour chaque saccade en fonction de la complexité de l'objet.

Nous avons vu dans la section 3.1 que les objets curvilignes sont représentés de façon

¹⁰ α détermine la taille des arcs de cercles de rayons l et L qui délimitent la zone B. l et L sont respectivement les petit et grand rayons.

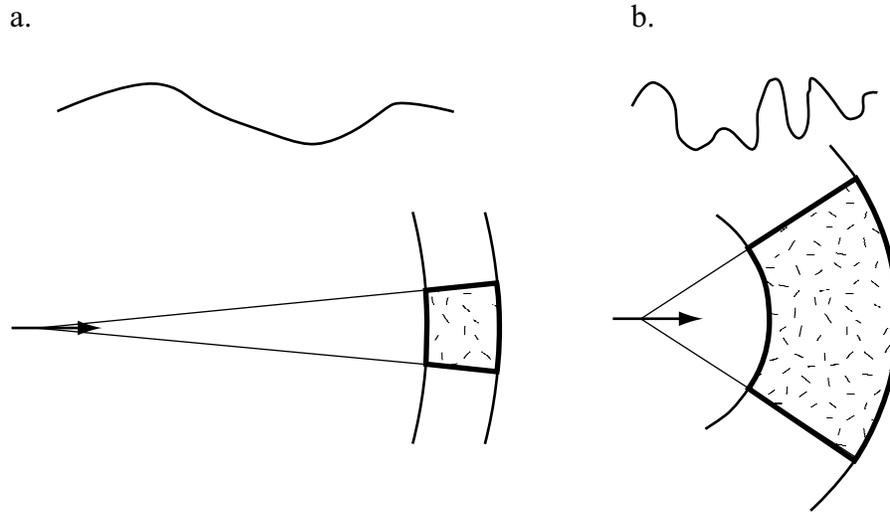


FIG. 4.17 – Différents types de zones de saccade. a. Objet curviligne simple. b. Objet curviligne complexe.

vectorielle dans les bases de données (SIG) et sont simplifiés en utilisant l'algorithme de Douglas-Peucker – cf. [DP73] et annexe A.1. On utilise une métrique de la complexité de l'objet qui est la densité linéaire des points le long de l'élément que ces derniers représentent. Plus la complexité de l'objet sera grande, et plus α sera grand et l et L seront petits. La figure 4.17 illustre ce point : dans 4.17a la zone cible est petite car l'objet curviligne est simple, c'est-à-dire que l et L sont grands, et de valeurs proches, alors que α est petit. Dans ce cas l'amplitude de la saccade est importante. Quand l'objet est plus complexe (figure 4.17b), α est grand, et, l et L sont petits mais de valeurs significativement différentes l'une de l'autre — $L \gg l$. Ceci permet d'avoir une zone cible grande et proche du point de départ. Par conséquent dans ce cas de figure, l'amplitude de la saccade est petite.

Ainsi, le point suivant doit être dans la zone précédemment déterminée. Si le premier point (courant à un moment t) de la liste de priorités se situe dans la zone, ce sera le point retenu. Si ce n'est pas le cas, en fonction de la situation (un point de – toute – la liste de priorité est situé ou non dans la zone) ce sera soit le premier point de la liste de

priorités situé dans la zone¹¹, soit le centre de la zone. La raison d’être de la zone¹² est que *si l’information n’est pas trouvée au centre (de la zone), le modèle recherche autour (du centre) et à l’intérieur de la zone.*

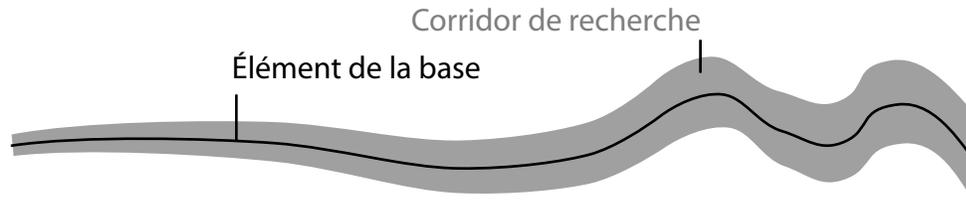


FIG. 4.18 – Différence de largeur des corridors de recherche en fonction de la complexité de l’objet.

Il est intéressant de noter que la zone déterminée (figure 4.17b) “instancie” la notion de corridor de recherche qui était jusqu’à présent conceptuelle. La largeur du corridor est donc fonction de la taille de la fenêtre d’extraction et de la zone déterminée – figure 4.17b). La figure 4.18 illustre ce propos. Lorsque l’objet est simple le corridor est plus étroit, et *vice versa*.

4.4 Conclusion

Dans ce chapitre nous avons montré comment le modèle ELECA dirige sa recherche d’information pertinente dans l’image en utilisant le guidage d’un regard virtuel. Le déplacement de ce regard virtuel est contraint à la fois par un champ de saillance et un modèle de comportement du regard.

La saillance est calculée de manière originale en prenant appui sur les spécificités du problème, c’est-à-dire en utilisant la base de données géographiques d’un côté (mode

¹¹Excepté si ce point correspond à un autre élément dans le mode de reconnaissance.

¹²C’est-à-dire pourquoi calcule-t-on une zone plutôt qu’un point.

reconnaissance) et d'un autre côté, la *segmentation* de l'image produite par le mode découverte, et les résultats intermédiaires comme les amorces.

Le modèle de comportement du regard est bâti en utilisant une approche par saccades. Cette approche par saccades permet de s'affranchir du problème de l'occultation partielle des éléments. Le calcul des saccades est fait en utilisant les données extraites de l'image au point de fixation du regard virtuel. L'extraction de cette information fait l'objet du chapitre suivant.

Le tableau 4.1 résume les cycles de spécifications-solutions qui ont conduit à l'élaboration du modèle de guidage du regard virtuel propre au modèle ELECA et qui en est la première sous partie ; les deux autres sous parties sont décrites dans les deux chapitres qui suivent.

PROBLÈME	SOLUTION
Guidage du regard de l'image vers l'information pertinente	Saillance et modèle de comportement
Saillance	Corridors de recherche
Corridors de recherche	Liste de priorités
Liste de priorités	Critère de priorité
Critère de priorité, mode reconnaissance	Algorithme de Douglas-Peucker
Critère de priorité, mode découverte	Stratégie amorce, périphérie, réalisation aléatoire
Comptabilité des découvertes	
Trace des localités explorées	
Traces des amorces	Listes des priorités
Caractéristiques des éléments	
Comportement du regard virtuel	
Parcours partiel de l'image	Modèle de saccades
Suivi d'un élément partiellement occulté	
Détermination d'une saccade	Calcul d'une zone cible à partir de l'information locale
Extraction de l'information locale	Chapitre suivant(!)

TAB. 4.1 – Cycles de spécifications-solutions pour le guidage du regard virtuel.

Chapitre 5

Extraction de l'information locale

*True genius resides in the capacity for evaluation
of uncertain, hazardous, and conflicting information.
Winston Churchill.*

Dans ce chapitre, nous élaborons une stratégie d'extraction de l'information locale au sein de l'image. Dans un premier temps le problème est décrit par une mise en contexte générale puis également dans le cas de l'exemple type. Dans un second temps, les spécifications du problème d'extraction d'informations locales pertinentes au sein de l'image pour le modèle ELECA sont explicitées. Enfin, dans un troisième temps, la solution originale par combinaison simultanée de techniques proposées dans le modèle ELECA est développée. Cette solution, obtenue par spécifications-solutions successives, permet de compléter le mouvement du regard virtuel, abordé au chapitre précédent, en y ajoutant l'extraction d'information par fixation d'un œil virtuel sur l'image. Un bilan récapitulatif est dressé en conclusion de ce chapitre.

5.1 Contexte d'extraction pour le modèle ELECA

Extraction de l'information pour guider les saccades Nous avons vu dans le chapitre précédent que le modèle ELECA proposait de s'affranchir d'un certain nombre de difficultés classiques en télédétection par une technique de sous échantillonnage de l'image. Ce sous échantillonnage de l'image est fait via un comportement de recherche de l'information inspiré de la vision humaine par *saccades*. La figure 5.1 donne un exemple du parcours saccadique sur une scène visuelle pour deux éléments curvilignes. Pour des raisons de lisibilité, seuls les emplacements où de l'information a été trouvée (carrés) sont représentés. Il nous faut maintenant spécifier le problème de l'extraction de l'information à ces emplacements particuliers.

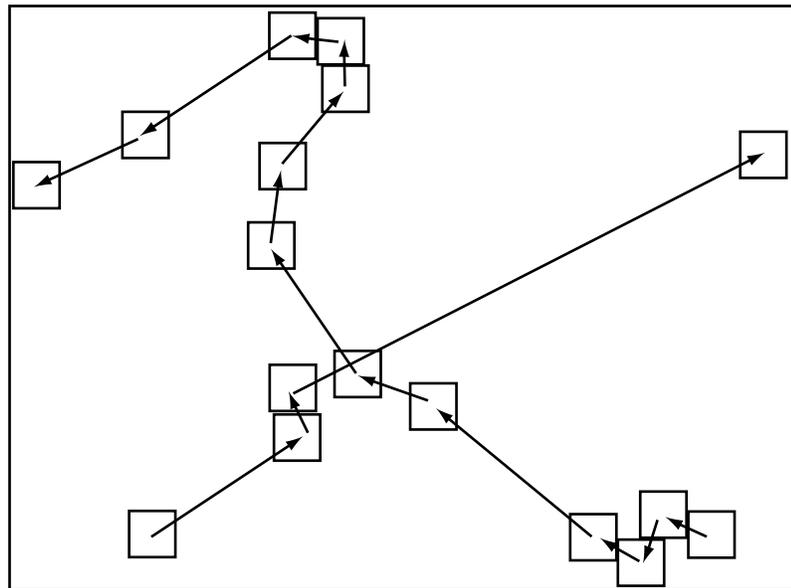


FIG. 5.1 – Exemple de parcours saccadique pour deux objets. Chaque flèche est une saccade et chaque carré (fenêtre) un point de fixation.

Dans le modèle ELECA, la localisation de la saccade suivante est déterminée par un compromis entre la liste des priorités et des contraintes dues au comportement saccadique. Dans le cas du mode de reconnaissance, une contrainte supplémentaire

impose de suivre l'élément dans la recherche le long des corridors. Afin de pouvoir réaliser une séquence de saccades comme présentée dans la figure 5.1, le modèle ELECA a besoin de l'orientation de l'élément dans l'image au point de *fixation*.

Extraction de l'information par ciblage de plus en plus fin L'information dans le modèle ELECA est “traquée” de manière de plus en plus précise. En premier lieu, au niveau le plus large, l'élément est recherché dans la zone prédéterminée — figure 4.16b page 135. Ensuite, à un niveau plus resserré, une fenêtre est centrée sur cette zone. Finalement, au niveau le plus fin, un traitement (que nous verrons ensuite) est appliqué sur cette fenêtre dans le but d'extraire les éléments linéaires.

La taille de la fenêtre dépend de plusieurs paramètres. Le problème est d'avoir suffisamment d'informations locales pour pouvoir ensuite calculer la saccade suivante. Dans ce contexte, la difficulté est de trouver un bon compromis entre le temps de traitement et le taille de la fenêtre. Une trop grande fenêtre accroîtrait le temps de traitement alors qu'une trop petite fenêtre ne permettrait pas d'extraire suffisamment d'information.

5.1.1 Objectif

L'objectif de cette étape¹ du modèle est de proposer une technique qui permette, de façon *rapide* et simple, de *fournir les “morceaux” d'éléments* (curvilignes) dont nous avons parlé précédemment, *ainsi que leur orientation*. La rapidité est un objectif très important. Rappelons ici que la finalité du travail est de retrouver les éléments curvi-

¹Autrement dit ce qui est traité dans ce chapitre.

lignes donc leurs morceaux et, que l'orientation locale de ces morceaux est cruciale pour guider le regard virtuel.

Les morceaux d'éléments doivent être extraits d'une petite imagerie qui correspond au point de fixation de l'œil. Il s'agit donc d'identifier des pixels qui font partie d'un élément linéaire. Ces pixels doivent fournir par conséquent trois informations : tout d'abord la position de l'élément, ensuite son orientation locale et enfin l'information spectrale liée à l'élément.

5.1.2 Idée directrice

L'idée directrice qui a guidé l'élaboration de cette partie du modèle ELECA est qu'il existe un sous ensemble d'informations disponibles dans l'imagerie, qui permet, bien que l'on n'exploite pas l'ensemble de l'information de l'imagerie, de dégager l'orientation de la saccade à suivre sans faire pour autant trop d'erreurs. L'erreur d'estimation de l'orientation est un cas envisageable (et envisagé) dans le modèle ELECA. Le système de contraintes, développé au chapitre précédent, couplé à la zone de "tolérance à la recherche" que représente la zone cible de la saccade, permet en cas d'échec de ne pas perdre le regard.

La démarche proposée, consiste à sélectionner des méthodes simples, ayant un comportement connu dans la mesure du possible ; l'intérêt de la recherche n'est pas dans ces aspects. La solution originale développée par le modèle ELECA réside dans l'enchaînement et l'articulation des différentes méthodes ensemble. Le modèle propose une solution élégante et simple.

5.2 Exemple type

Le comportement saccadique de l'exploration décrite dans le chapitre précédent repose sur une donnée fondamentale : l'orientation de l'objet à l'origine du saut – vecteur indiqué au point A de la figure 4.16 page 135. En reprenant l'exemple de départ, figures 4.1 et 4.2 (page 114 et suivante), il s'agit de déterminer les vecteurs orientations pour chaque saut et les morceaux de courbes présents dans l'imagette; nous verrons plus loin que vecteurs orientations et morceaux d'éléments sont intimement liés. Ainsi, sur l'image figure 4.1 il s'agirait des vecteurs rouges présentés dans la figure 5.2.

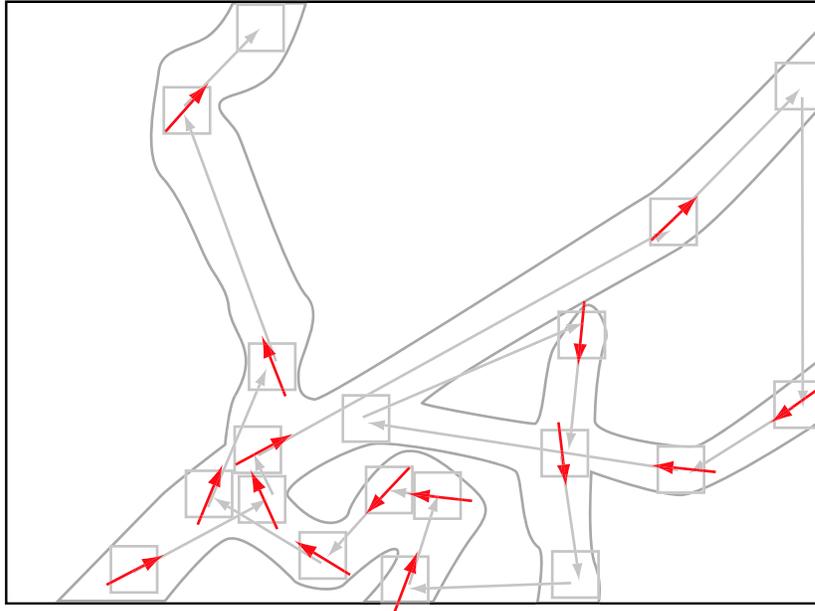


FIG. 5.2 – Exemple de vecteurs orientations à partir de la figure 4.1.

5.3 Modèle simple d'extraction de l'information locale

L'extraction au sein de l'imagerie se déroule en quatre étapes. D'abord l'imagerie est filtrée pour s'affranchir du bruit qui est très présent dans les images RADAR, cf. §3.1. Ensuite une extraction par rehaussement des *arêtes* et seuillage est menée. Finalement une solution originale du modèle ELECA pour l'identification des points connectés en même temps qu'un calcul d'orientation est présentée. La suite de cette section est entièrement consacrée à expliciter la démarche.

5.3.1 Filtrage du bruit de l'imagerie

Lorsque l'on s'attaque au *filtrage* du bruit dans une image, les références et les travaux sont extrêmement nombreux. En gardant à l'esprit l'objectif de simplicité nous avons opté pour un *filtre* robuste bien documenté dans la littérature et utilisé dans des travaux précédents au laboratoire [Fit99], le filtre de Lee [Lee81].

Outre le fait que le filtre de Lee a un fonctionnement et une mise en œuvre relativement simple, il présente trois avantages qui nous intéressent dans le cas présent [MT84] et [Fit99] :

- le filtre de Lee préserve bien les arêtes des images, ce qui nous intéresse particulièrement puisque nous en faisons l'extraction ensuite ;
- il utilise des statistiques locales pour faire l'ajustement de l'intensité des pixels, cette propriété est très intéressante puisque nous travaillons localement à cette

étape du processus ;

- son efficacité pour réduire le bruit des images RADAR est bonne.

Le filtre de Lee glisse sur l'image afin de remplacer le *pixel central* par la valeur obtenue. Le calcul est fait comme suivant l'équation 5.1 :

$$I'(i, j) = \bar{I}_f + \left(\frac{\nu_s}{\nu_s + \nu_f} \right) \cdot (I(i, j) - \bar{I}_f) \quad (5.1)$$

où :

$I'(i, j)$ est la nouvelle intensité affectée au pixel de coordonnées (i, j) ,

\bar{I}_f est l'intensité moyenne des pixels situés dans la *fenêtre de filtrage*,

ν_s est un bruit aléatoire (blanc) et multiplicatif, il se calcule en suivant l'équation 5.2 :

$$\nu_s = \left(\frac{\bar{I}_f}{\bar{I}_i} \right)^2 \nu_i \quad (5.2)$$

où :

\bar{I}_i est l'intensité moyenne des pixels de l'image à traiter,

ν_i est la variance de l'intensité des pixels de l'image à filtrer.

ν_f est la variance de l'intensité des pixels situés dans la fenêtre de filtrage,

$I(i, j)$ est l'intensité du pixel de coordonnées (i, j) .

L'équation 5.1 du filtre de Lee permet donc de lisser l'image pour atténuer le chatoie-
ment² de l'image. L'image supérieure de la figure 5.3 est une sous image de la figure 3.3
(RSO) sans le filtrage de Lee. L'image inférieure dans la figure 3.3 est la même sous
image à laquelle le filtre de Lee a été appliqué. La taille du filtre appliqué est de 5×5 .

²Chatoieement : voir page 101.

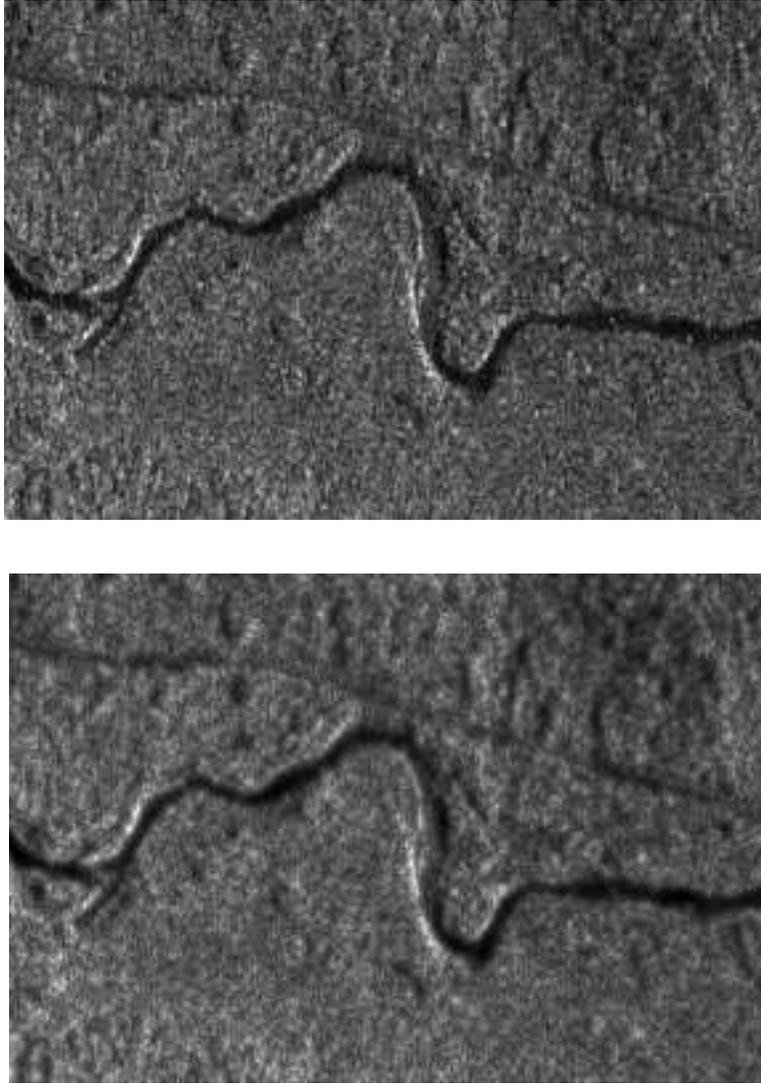


FIG. 5.3 – Application du filtre de Lee à une image RSO. L'image du haut est une image brute, sous image de l'image de la figure 3.3. L'image du dessous est l'image filtrée par le filtre de Lee 5×5 .

Il faut noter, puisqu'il s'agit d'une fenêtre glissante que la taille de l'image résultante n'est pas la même que celle de départ. On le remarque très bien sur la figure 5.3. Si l'image à filtrer a une taille de $M \times N$ et que le filtre a une taille de $m \times n$ (où m et n sont impairs) alors l'image résultante aura une taille de $[M - \frac{m-1}{2}] \times [N - \frac{n-1}{2}]$; nous verrons dans le chapitre suivant que cela a un impact dans la mise en œuvre du filtrage.

5.3.2 Détection de contours

La détection de contour est un grand classique du traitement d'image. C'est un aspect de ce domaine qui a été énormément étudié [CC01].

Le principe est de détecter les changements brusques d'intensité sur l'image. Le changement de courbure d'une fonction s'analyse en étudiant son gradient (dérivée première) et son laplacien (dérivée seconde). Pour ce type d'opération, on utilise des filtres passe-haut. Le changement brusque d'intensité implique un changement de courbure dans la fonction qui décrit l'intensité en fonction des pixels le long d'une direction donnée.

La littérature foisonne de ce genre de filtres, souvent à des fins artistiques ; lorsqu'il s'agit de rehausser les contours dans une image de télédétection la "panoplie" est nettement plus restreinte [CC01]. Suivant l'expérience de Fitzback [Fit99], nous nous sommes arrêtés sur un filtre omnidirectionnel dont la mise en œuvre est simple et dont le calcul est rapide, afin de répondre aux objectifs fixés. Il s'agit du filtre de Robinson [Rob77]. Le filtre de Robinson garde le maximum des écarts de niveaux d'intensité selon les quatre directions possibles : \uparrow , \nearrow , \rightarrow , \searrow . Les écarts de niveaux d'intensité sont obtenus en convoluant un filtre passe-haut pour chacune des directions. Ainsi le filtre de Robinson s'écrit :

$$\forall k \in \{\uparrow, \nearrow, \rightarrow, \searrow\}, \quad I'(i, j) = |\max(\mathbf{A}_{image} * \mathbf{B}_{filtre_k})| \quad (5.3)$$

où :

$I'(i, j)$ est la nouvelle intensité affecté au pixel de coordonnées (i, j) ,

\mathbf{A}_{image} est la partie de la matrice image à traiter située sur la fenêtre de filtrage,

\mathbf{B}_{filtre_k} est un filtre passe haut ; il peut prendre une des quatre valeurs suivantes :

$$\mathbf{B}_{filtre_{\uparrow}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix} \quad \mathbf{B}_{filtre_{\nearrow}} = \begin{bmatrix} 0 & 1 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\mathbf{B}_{filtre_{\rightarrow}} = \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \end{bmatrix} \quad \mathbf{B}_{filtre_{\searrow}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 0 & -1 & -1 & -1 \\ 0 & -1 & -1 & -1 & -1 \end{bmatrix}$$

L'image présentée à la figure 5.4 est l'image de la figure 5.3 filtrée par un filtre de Robinson 5×5 . On note que les arêtes ont été rehaussées. Il faut également remarquer, tout comme dans le cas du filtre de Lee, que la taille de l'image résultante est $[M - \frac{m-1}{2}] \times [N - \frac{n-1}{2}]$ par rapport à l'image résultante filtrée.

5.3.3 Seuillage et analyse par composantes connexes

Une fois que l'on a obtenu le produit du filtre de détection de contour, la difficulté est de reconnaître, à partir de pixels noirs et blancs, des éléments curvilignes. Nous avons vu au cours du chapitre 2 qu'il existe de nombreuses mesures dans le domaine perceptif pouvant aider à dégager des structures plus larges. Comme mentionné précédemment,

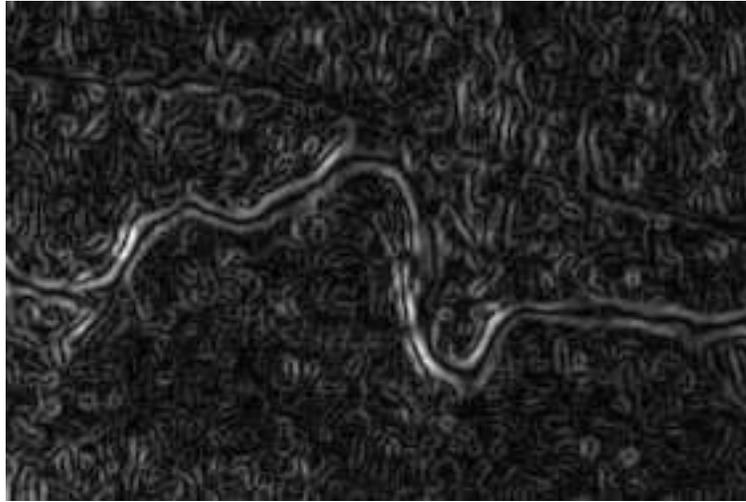


FIG. 5.4 – Application du filtre de Robinson.

l'orientation (associée à *la proximité*) est utilisée dans notre modèle. Ce sont des critères relativement simples qui répondent bien à l'idée que l'on se fait de structures linéaires. L'idée ici est de simplifier les mesures pour compenser la complexité de leur acquisition³.

Un point important est de comprendre que ces critères peuvent être différents en fonction de la nature de l'objet recherché. Autrement dit, que l'objet soit une ligne à haute tension, un cours d'eau ou une route, le critère recherché est l'adaptabilité – par exemple déterminer que leurs sinuosités sont différentes. Ceci permet de tirer avantage encore une fois de la base de données originale pour optimiser notre recherche. Les éléments qui sont extraits alimentent aussi la fusion local/global pour mettre à jour la base de données.

³En outre ces mesures sont citées par Ligozat et Edwards [LE00] dans le cas d'approches allant d'informations locales vers une intégration globale. Nous abordons ce point au chapitre 6.

5.3.3.1 Seuillage

Une fois lissée et avec ses arêtes rehaussées, l'imagette n'est pas exploitable en l'état, figure 5.4. Il faut d'une part ne prendre que les points les plus saillants, puis les regrouper en amas susceptibles d'être des "morceaux" de courbes. En pratique il n'est pas possible de regrouper les amas de pixels en une seule itération sur l'image, cela nécessite au minimum deux itérations. Nous rappelons qu'un des objectifs du modèle est la rapidité de traitement. Le modèle ELECA est contraint par conséquent par l'impératif des deux passages sur l'image.

Plutôt que de le considérer comme une contrainte, le modèle tire avantage du double passage sur l'image. Au cours du premier passage, un seuillage est réalisé en plus de la tâche de regroupement initial des pixels. Le second passage permet de calculer l'"orientation instantanée" en même temps que se fait le regroupement final.

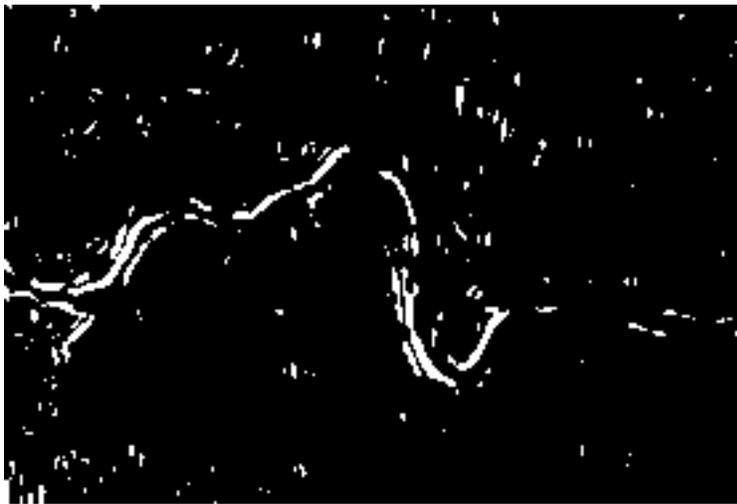


FIG. 5.5 – Image seuillée à 90.

Le seuillage est une opération très simple qui consiste à transformer l'image codée sur plusieurs niveaux⁴ d'intensité en une image binaire en ne gardant à 1 que les pixels

⁴Plus que deux, dans notre cas, la plupart du temps classiquement ce seront 256 niveaux de gris.

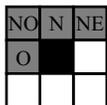
de l'image supérieurs à certain seuil, les autres étant mis à 0. L'image de la figure 5.5 est l'image de la figure 5.4 seuillée à 90. Cela veut dire que sur les 256 niveaux de gris qui composent l'image de départ, on ne garde que ceux supérieurs à 90, et on leur affecte à tous la valeur 1. Les pixels dont la valeur est supérieure au seuil correspondent à de probables points de contours. Il s'agit ensuite de regrouper ces pixels blancs ensemble.

5.3.3.2 Analyse par composantes connexes

Le regroupement se fait via une *analyse en composantes connexes* [GW02]. L'algorithme est illustré par l'algorithme 5.1. La figure "a." de l'algorithme 5.1 est l'image contenant les pixels à regrouper. Ces pixels sont identifiés en noir dans notre exemple ; ils forment un U.

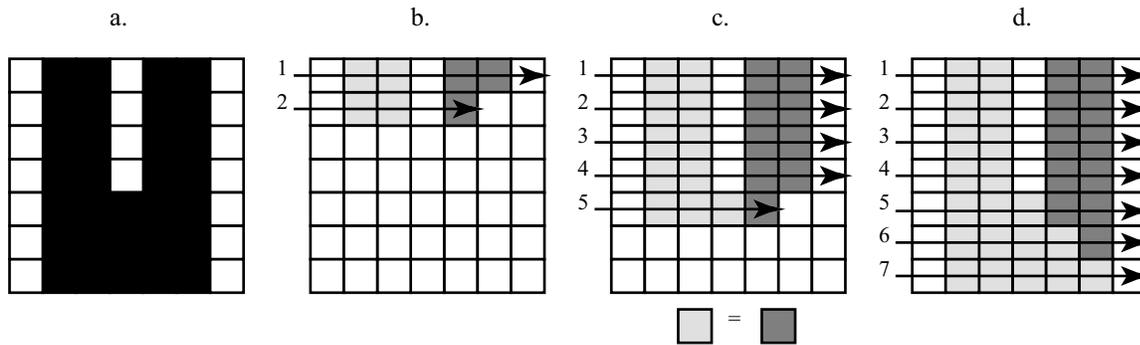
Dans un premier temps on commence à parcourir l'image (ici, dans l'algorithme 5.1, de gauche à droite et de haut en bas) jusqu'à ce que l'on rencontre un pixel candidat. Dans notre cas, il s'agit d'un pixel dont la valeur d'intensité est supérieure à la valeur du seuil. C'est ainsi que l'on peut seuiller l'image en même temps que l'on fait le premier passage pour l'analyse en composantes connexes – ACC. À chaque fois que l'on rencontre un pixel candidat on vérifie si les pixels qui lui sont connectés ont déjà été parcourus⁵ et sont déjà étiquetés. Si tel n'est pas le cas, on attribue une étiquette au pixel ; c'est le cas du premier pixel clair et du premier pixel foncé de la première ligne en "b." de l'algorithme. Les pixels qui ont une connexion avec un pixel étiqueté auparavant se voient attribuer l'étiquette du pixel auquel ils sont connectés. Lorsqu'un pixel candidat vient à être connecté à deux pixels d'étiquettes différentes (cas "c." de

⁵C'est-à-dire les quatre pixels nord-ouest, nord, nord-est et ouest.



l'algorithme), l'équivalence des étiquettes est mémorisée.

Il ne reste plus ensuite qu'à faire une seconde itération pour affecter la même étiquette aux pixels ayant des étiquettes équivalentes et ainsi isoler chaque regroupement de pixels sur l'image. Nous verrons que bien que très simple à décrire l'implantation de cet algorithme n'est pas aisée, en particulier dans notre contexte.



ALG. 5.1 – Algorithme général d'analyse en composantes connexes – premier passage.

5.3.4 Extraction “instantanée” de l'orientation

Le modèle de saccades que nous utilisons pour parcourir l'image, nécessite l'orientation des structures locales au point de fixation de l'œil virtuel – sur l'imagette. Nous avons mentionné que l'imagette est parcourue une seconde fois au cours de l'ACC pour l'affectation des étiquettes équivalentes. On peut souligner également que l'imagette que nous utilisons pour faire nos mesures est de taille très réduite, ce qui restreint la présence de structures trop complexes sur une aussi petite surface, sachant que ce sont les arêtes qui ont été extraites de l'image.

L'orientation “instantanée” est prise directement durant la seconde itération de l'ACC. Elle est calculée entre le premier point étiqueté et le dernier. Cette approche

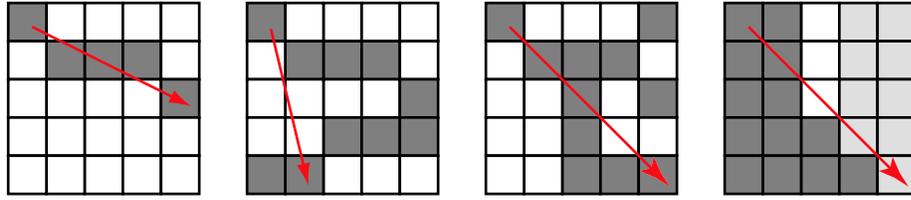


FIG. 5.6 – Orientations instantanées.

est extrêmement rapide puisqu'elle évite tout parcours supplémentaire, que ce soit de l'image ou bien de la structure vectorielle créée. En revanche, dans des cas particuliers l'information peut être erronée. La figure 5.6 montre quelques exemples. Les cas où l'orientation est erronée apparaissent très minoritaires dans le cas général. De plus, ces orientations erronées tendent à se produire lorsque le nombre de pixels étiquetés dans l'imagette est important, ce qui est peu fréquent puisque l'on recherche des objets linéaires – par opposition aux objets surfaciques. Enfin, comme expliqué précédemment, le modèle ELECA est suffisamment robuste dans le guidage du regard virtuel pour qu'une orientation erronée n'ait pas pour conséquence de perdre le regard virtuel. La liste des priorités recentrera automatiquement un œil virtuel qui n'extraîrait aucune information dans les zones cibles.

5.4 Conclusion

Dans ce chapitre nous avons montré comment le modèle ELECA extrait l'information locale pertinente dans l'image en utilisant une combinaison originale de techniques robustes connues. L'information extraite sert à alimenter le modèle de guidage du regard virtuel vu au chapitre 4. *Le présent chapitre a décrit ce que réalise l'œil virtuel lors d'une fixation du regard virtuel.*

L'extraction locale se déroule en trois phases :

1. Filtrage du bruit de l'imagette à l'aide du filtre de Lee.
2. Rehaussement des arêtes à l'aide du filtre de Robinson.
3. Seuillage, regroupement des points de contours⁶ et calcul "instantané" de leur orientation par une ACC.

La troisième phase de l'extraction tire avantage de manière originale des contraintes liées à la technique d'identification des objets rehaussés en calculant leur orientation de manière rapide et simple. De par la robustesse de la méthode de recherche de l'information par saccades, la relative grossièreté du calcul de l'orientation locale s'en trouve atténuée.

Une fois extraites ces informations locales nécessitent, dans le mode découverte, d'être fusionnées dans une perspective de carte globale. Cet aspect est discuté dans le chapitre suivant.

Le tableau 5.1 résume les cycles de spécifications-solutions du modèle d'extraction des informations locales pour le modèle ELECA.

⁶Les très petits groupes de pixels sont rejetés.

PROBLÈME	SOLUTION
Extraction de l'information locale pertinente pour guider le regard virtuel	Fixation d'un œil virtuel sur une sous image (imagerie)
Information bruitée dans la sous imagerie	Application d'un filtre approprié (Lee) uniquement sur l'imagerie
Recherche de l'information locale (dans l'imagerie)	Rehaussement des arêtes uniquement sur la sous imagerie par application d'un filtre omnidirectionnel approprié (Robinson)
Identification des pixels candidats	ACC modifiée judicieusement. Seuillage et première passe ACC simultanés.
Étiquetage initial des pixels candidats dans l'imagerie	Puis deuxième passe ACC et calcul de l'orientation locale simultanés.
Regroupement des pixels étiquetés	
Orientation des groupes de pixels étiquetés	
Fusion de l'informations locales	Chapitre suivant(!)

TAB. 5.1 – Cycles de spécifications-solutions pour l'extraction locale.

Chapitre 6

Fusion des informations locales

*It is quality rather than quantity that matters.
Lucius Annaeus Seneca, Epistles.*

Ce chapitre est consacré à l'élaboration d'une stratégie de fusion des informations locales dans le but de les intégrer dans une perspective de carte globale. Ce chapitre s'applique par conséquent au mode découverte. Dans un premier temps, le problème est spécifié par une mise en contexte qui justifie l'utilisation d'approches développées en raisonnement spatial qualitatif ; puis le problème est présenté dans le cas de l'exemple type. Dans un second temps, une lecture spécifique de travaux en raisonnement spatial qualitatif nous permet de proposer une articulation spécifique (nommée modèle LE) de reconstruction d'un environnement à partir d'informations perçues localement par un agent. Le modèle LE élabore un modèle de relations spatiales entre les objets perçus. Enfin, dans un troisième temps, s'appuyant sur des formalismes appropriés, le modèle CLE, un sous modèle du modèle ELECA, est proposé pour guider la fusion des petits éléments locaux extraits précédemment en un schéma général. Le modèle CLE établit des équivalences formelles originales avec le modèle LE pour construire un modèle de relations spatiales des objets extraits en vue de la reconstruction globale des structures géométriques linéaires de l'image. Un bilan récapitulatif est dressé à la fin du chapitre.

6.1 Raisonnement spatial qualitatif et appréhension globale

6.1.1 Objectif

L'objectif de cette partie du travail est de proposer un "portrait clair" de la manière de formaliser le problème qui nous est posé, à savoir *fusionner les petites informations locales en des entités globales qui puissent être mises (injectées) dans la base de données*. Il s'agit, pour les buts de la thèse, de montrer qu'il y existe des pistes de solutions possibles pour formaliser le problème.

Dans un cadre général l'approche développée plus bas doit pouvoir s'attaquer à :

1. Déterminer si deux morceaux courbes font partie d'un même élément curviligne, en particulier si ces deux morceaux ne sont pas géométriquement proches, par exemple dans les cas d'obstructions évoqués précédemment.
2. Traiter les changements brusques de direction des éléments comme les lignes de haute tension de l'exemple type.
3. Traiter les phénomènes d'aiguillage convergent – (\succ) deux lignes se regroupent en une – ou divergent – (\prec) une courbe en devient deux.

Il s'agit donc de fournir une description qualitative de la scène en termes de lignes. Autrement dit, il s'agit d'être capable de répondre aux questions suivantes concernant ces lignes :

- Quelles sont elles ?
- Comment se comportent-elles (géométriquement parlant) entre elles ?
- À laquelle appartient un morceau donné ?

- Comment s’arrangent-elles entre elles de manière semi-globale ? Par exemple la ligne L_1 part vers le nord-est puis oblique brusquement vers l’ouest, coupant le chemin L_2 , qui lui est nord-sud, à la hauteur de...

En résumé, l’objectif est de montrer que l’on peut proposer dans un cas concret comme celui-ci un *modèle des relations spatiales* qui existent dans la scène visuelle pour permettre ultérieurement la fusion cohérente des informations.

6.1.2 Idée directrice

6.1.2.1 Cadre structurant

Nous avons vu lors du chapitre 2 que les travaux en raisonnement spatial qualitatif proposent des formalismes qui permettent de manipuler l’information spatiale de type qualitatif. Un autre aspect intimement lié au raisonnement est la représentation qualitative spatiale [Ren02] qui elle aussi autorise une meilleure emprise sur l’information spatiale qualitative. Forts du constat, rappelé tout au long de cette thèse et qui sous-tend le modèle ELECA, qui est que l’humain “se sort” de situations complexes facilement [HM85] [Dav90] en utilisant des raisonnements foncièrement qualitatifs (par opposition à quantitatifs), le raisonnement et la représentation qualitatifs apparaissent comme des outils privilégiés permettant de proposer des solutions à la fusion d’informations locales vers une perspective globale, cas qui nous intéresse dans le cadre du mode découverte.

Ceci étant dit, en raisonnement spatial qualitatif, deux cadres généraux peuvent être utilisés pour notre travail. D’une part, les travaux sur le calcul des dipôles (des segments de droites orientés) proposé par Schlieder [Sch95] et le formalisme élaboré

par Moratz *et al.* [MRW00] sur les segments de lignes mais aussi sur les points orientés [Mor04] offrent des formalismes sur lesquels le modèle ELECA peut s'appuyer. D'autre part, les travaux initiés par Ligozat en 1998 [Lig98] sur les directions cardinales et poursuivis par Ligozat et Edwards d'abord sur la proximité et l'alignement [LE00] et ensuite sur la structuration des perceptions locales de l'espace environnant [EL04] apportent un éclairage nouveau sur *le passage de connaissances spatiales locales vers une appréhension spatiale globale*. Ces deux cadres sont abordés dans la section 6.3.1.

6.1.2.2 Cas du modèle ELECA

Une fois que le problème a été situé vis-à-vis du raisonnement spatial qualitatif d'une part, et que les objectifs ont été spécifiés en détail dans le contexte de la thèse d'autre part, l'idée directrice de la recherche, ici, est de regarder comment le modèle ELECA peut tirer avantage des travaux précités. Ainsi, nous pensons que le raisonnement qualitatif spatial offre une alternative aux approches purement algorithmiques permettant de fusionner les informations locales (comme celles du modèle ELECA) dans une perspective de carte globale. L'approche qualitative offre un cadre structurant pour le problème de fusion.

Plus précisément, il est possible de s'appuyer, d'une part sur des formalismes existants en raisonnement spatial et, d'autre part sur les travaux de reconstruction en deux dimensions des environnements locaux afin d'élaborer un cadre de travail permettant d'obtenir une description qualitative des relations spatiales, première étape vers un *modèle des relations spatiales*.

6.2 Exemple

Dans le cadre de l'exemple type de la page 107 et suivantes, la situation à ce point du modèle ELECA peut être décrite comme suit. Les données issues de la base de données (mode reconnaissance) ont été identifiées et mises à jour. Des informations locales ont été extraites de l'image en mode découverte, il s'agit maintenant de fusionner ces informations locales pour obtenir une perspective globale.

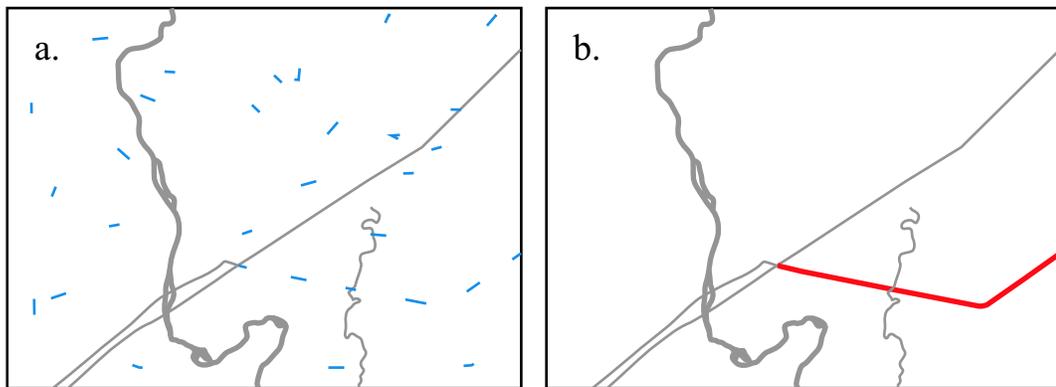


FIG. 6.1 – Fusion des informations locales dans une perspective de carte globale.

La figure 6.1 spécifie le problème. La figure 6.1a présente l'état des connaissances à l'issue de la recherche en mode découverte. C'est-à-dire qu'il y a d'une part, en grisé, la base de données originale corrigée de ses erreurs; et d'autre part en bleu l'ensemble des petites informations locales (des morceaux de courbes) collectées par l'œil virtuel. Toute la difficulté consiste donc à pouvoir discriminer les informations pertinentes de celles qui ont été collectées à tort et à fusionner ces informations pertinentes pour obtenir une mise à jour définitive de la base de données comme le montre la figure 6.1b.

6.3 Vers un modèle de relations spatiales

6.3.1 Relations spatiales et segments de lignes

L'approche développée dans ce chapitre repose sur des modèles proposés dans des contextes différents. Ces travaux portent, d'une part (§6.3.1.1), sur la reconstruction d'environnements à partir de relations spatiales locales et, d'autre part (§6.3.1.2), sur le raisonnement spatial qualitatif qu'il est possible de faire sur des petits segments de lignes.

6.3.1.1 Environnement 2D : relations spatiales et reconstruction

Edwards et Ligozat ont proposé un formalisme afin de décrire qualitativement les *relations spatiales dans des espaces à deux dimensions*. Ces travaux ont fait l'objet de trois publications successives qui décrivent un aspect différent à chaque fois permettant la description (et la reconstruction) qualitative d'environnement à deux dimensions.

Tout d'abord Ligozat, à partir des travaux de Frank [Fra91], Freska [Fre92] et Hernández [Her94], propose un *calcul des directions cardinales*¹ qui permet, grâce à neuf relations basiques, de représenter la position d'un point vis-à-vis d'un autre point fixe comme un repère [Lig98]. Ces neuf relations sont $n, s, e, w, ne, se, sw, nw$ et eq ; elles correspondent respectivement à des relations nord, sud, est, ouest, nord-est, sud-est, sud-ouest, nord-ouest et équivalente². Ligozat a montré qu'à partir de ces relations il était possible de mener des opérations et donc d'en déduire une algèbre, nommée

¹*Cardinal Direction Calculus*

²Un exemple de représentation de ces relations est donné plus loin, figure 6.3.

*algèbre des relations cardinales*³. Ces relations cardinales sont représentables facilement, en particulier via un treillis.

Ensuite Ligozat et Edwards [LE00] ont montré que même sans référence spatiale explicite il était possible de reconstruire une partie de l'environnement global. En utilisant la proximité et l'orientation et sans se servir des directions cardinales, ils ont montré qu'il était possible, pour et à partir d'un point donné, de :

- décrire l'espace ;
- décrire la localisation du point ;
- trouver la sortie d'un espace ;
- trouver une localisation donnée, en fonction des repères locaux⁴ entourant le point ;
- raisonner sur les orientations relatives entre les repères locaux ;
- raisonner sur la proximité entre les repères locaux.

Outre l'exploration de l'environnement qui permet de construire le modèle de relations spatiales qui existe entre les repères locaux, Ligozat et Edwards introduisent la notion de *panorama*. Ce panorama est la liste cyclique des directions autour d'un point donné vers les repères locaux. Ce panorama permet une description cyclique de l'environnement d'un point donné.

Enfin, récemment, Edwards et Ligozat [EL04] ont décrit un modèle formel pour structurer les perceptions (locales) de l'environnement local⁵. Ces travaux, comme les précédents, se limitent à deux dimensions. Edwards et Ligozat ont montré, en utilisant le langage du calcul des directions cardinales de Ligozat, que l'on pouvait identifier dans

³ *Cardinal Relation Algebra*

⁴ Les repères locaux sont appelés *agents* par Ligozat et Edwards

⁵ Toujours récemment il est intéressant de noter que des efforts ont été entrepris pour donner un cadre général au calcul qualitatif [LR04]. En particulier ces travaux s'attachent à proposer un cadre algébrique commun.

l'espace différents types de zones par exploration de l'environnement. Prenons l'exemple de la figure 6.2. A, B, C, D, et E sont des objets, appelons les repères locaux comme plus haut. Les barres noires sont des "murs" qui sont également des obstacles à la vue. Si l'on se déplace selon la trajectoire x_1, x_2, x_3, x_4, x_5 et x_6 on peut décrire à chaque point $x_1 \dots x_6$ par le biais du langage du calcul des directions cardinales, figure 6.3.

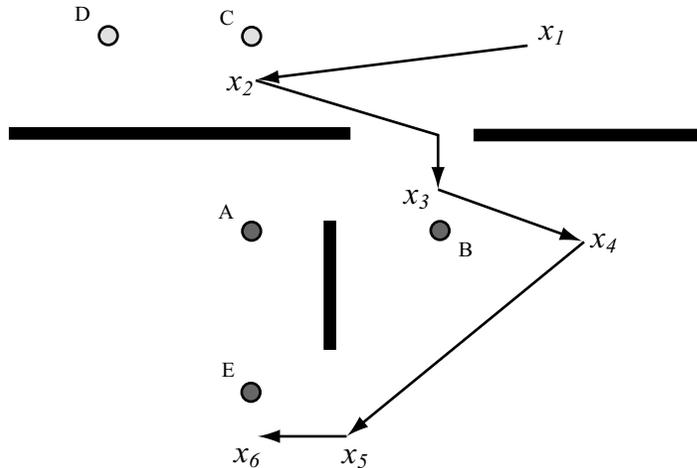


FIG. 6.2 – Exploration d'un environnement avec des occlusions, d'après [EL04].

Edwards et Ligozat ont mis en évidence par ce biais quatre types de zones :

1. Les *zones de stabilité*, dans lesquelles les vues ne changent que peu. C'est le cas par exemple lors du déplacement de x_1 à x_2 , figures 6.2 et 6.3. Les trois autres types de zones se retrouvent entre les zones de stabilité.
2. Les *barrières* qui servent en partie à délimiter les zones de stabilité. C'est le cas de la barrière verticale sur la figure 6.2. x_4 et x_6 sont dans deux zones de stabilités différentes délimitées par cette barrière, figures 6.2 et 6.3.
3. Les *zones de transition* où le passage d'une zone de stabilité à une autre est progressif. Le passage de x_4 à x_6 se fait par le passage par une zone de transition (une barrière étant un obstacle à la navigation) ; ainsi le point x_5 se situe dans une zone de transition, voir le passage progressif des vues $x_4 \rightarrow x_5 \rightarrow x_6$, figure 6.3.

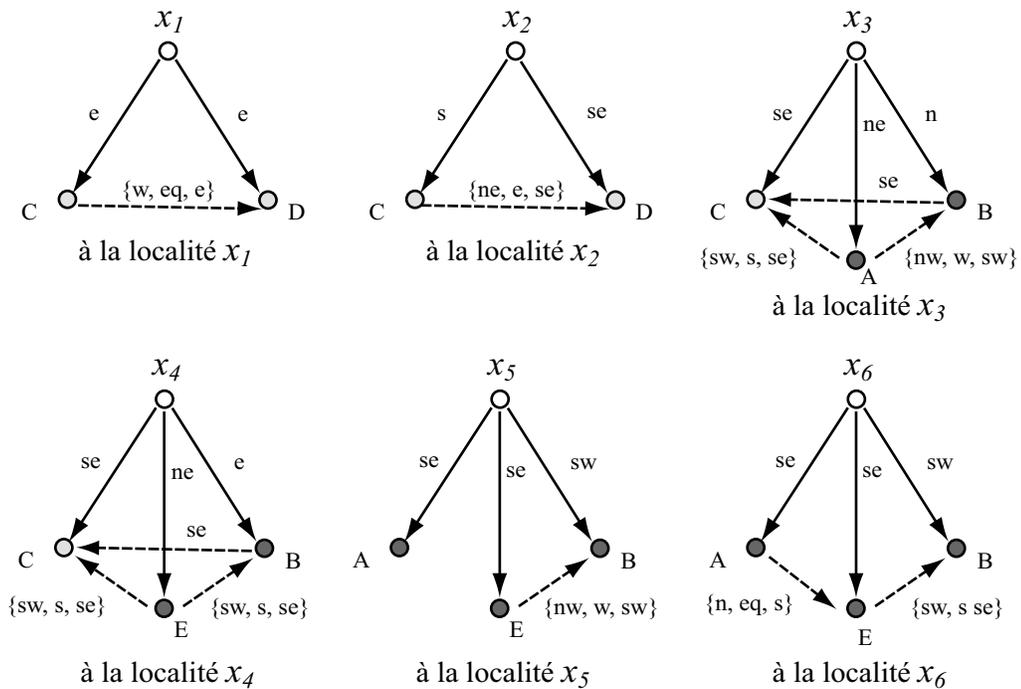


FIG. 6.3 – Les six vues issues de la figure 6.2, d’après [EL04] et selon le formalisme de [Lig98].

4. Les *portes*⁶ où le passage d’une zone à l’autre est brutal. Dans l’exemple des figures 6.2 et 6.3, il s’agit typiquement de l’ouverture entre deux murs. Cette porte sépare deux zones de stabilité.

Pour la suite nous appellerons *modèle LE* la description des travaux de Ligozat et Edwards telle que nous l’avons faite. C’est-à-dire l’articulation que nous venons de donner entre les trois contributions.

⁶[EL04] parle de *gateway*.

6.3.1.2 Raisonement spatial qualitatif et segments de lignes

Moratz *et al.* proposent un *calcul des dipôles* [MRW00] qui offre un formalisme très intéressant dans notre cas⁷. Les dipôles de Moratz sont des segments de lignes orientés (figure 6.4), notés A, B, C, \dots , et formés de deux points. Pour un dipôle A , ces deux points sont notés \mathbf{s}_A pour le point de début et \mathbf{e}_A pour le point de fin.

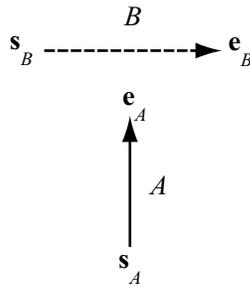


FIG. 6.4 – Relations d’orientations entre deux dipôles, d’après [MRW00].

Afin de pouvoir manipuler les différentes configurations spatiales entre deux dipôles, [MRW00] introduit une représentation de base des relations entre les dipôles. On peut décrire l’arrangement spatial entre deux dipôles A et B selon quatre relations spatiales point–dipôle de la manière suivante :

$$A R \mathbf{s}_B \wedge A R \mathbf{e}_B \wedge B R \mathbf{s}_A \wedge B R \mathbf{e}_A \quad (6.1)$$

où R est une relation spatiale qui peut prendre une des trois valeurs $\{r, o, l\}$ et \wedge le ET logique. Ces valeurs correspondent respectivement aux relations spatiales dans \mathbb{R}^2 «à la droite de», «dans le prolongement de» et «à la gauche de». Ainsi, la relation spatiale qui existe entre les deux dipôles de la figure 6.4 peut s’écrire :

$$A l \mathbf{s}_B \wedge A r \mathbf{e}_B \wedge B r \mathbf{s}_A \wedge B r \mathbf{e}_A \quad (6.2)$$

⁷Moratz propose également une approche par *points orientés* [Mor04]. Cette approche est très proche de la précédente et repose sur le calcul *OPRA*.

Pour alléger l'écriture on pose :

$$A \text{ lrrr } B := A \text{ l s }_B \wedge A \text{ r e }_B \wedge B \text{ r s }_A \wedge B \text{ r e }_A \quad (6.3)$$

Ainsi, Moratz *et al.* peuvent réécrire les quatorze relations de Schlieder [Sch95] possibles lorsque les points sont distincts, figure 6.5.

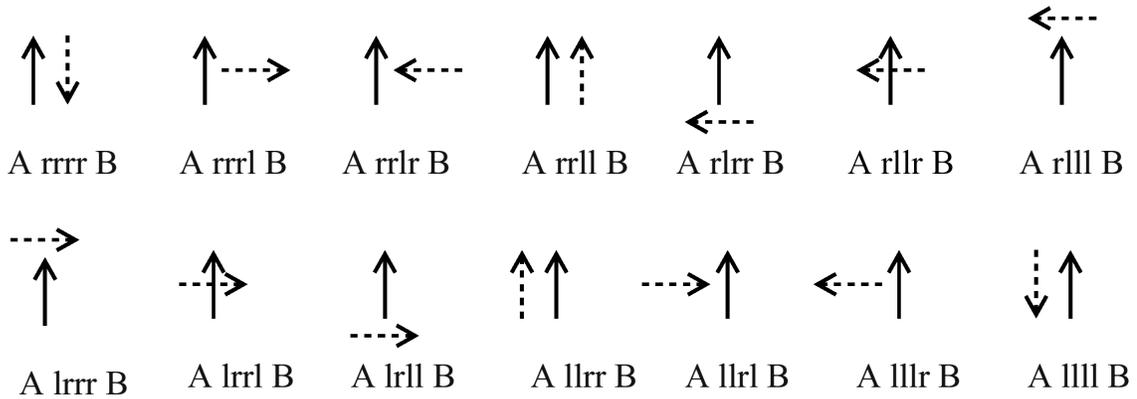


FIG. 6.5 – Les 14 relations de Schlieder entre deux dipôles selon le formalisme de Moratz *et al.*, d'après [MRW00].

En plus des quatorze relations précitées, on identifie dix relations supplémentaires lorsque les dipôles ont un ou deux points communs. Ces relations sont présentées sur la figure 6.6. La relation *s* correspond au point de départ commun, la relation *e* au point de fin commun. On peut remarquer que la relation *sese* est la relation d'identité. On note \mathcal{D}_{24} ces 24 relations atomiques et \mathcal{DRA}_{24} l'ensemble des parties qui contient les 2^{24} unions possibles entre ces relations atomiques.

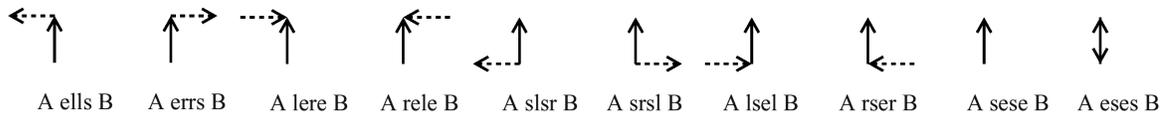


FIG. 6.6 – Dix relations supplémentaires, d'après [MRW00].

Moratz *et al.* démontrent que ces relations forment une algèbre relationnelle au sens

de [LM94]⁸. L'avantage est alors qu'il est possible de raisonner sur les relations entre les dipôles en utilisant des techniques de raisonnement basées sur les contraintes, ce qui n'était pas possible avec le calcul de Schlieder car il ne contient pas la relation d'identité. Moratz *et al.* montrent qu'il est possible d'utiliser le calcul des dipôles dans des cas de navigation.

6.3.2 Scène visuelle et relations spatiales

Nous proposons d'étudier les parallèles possibles entre le modèle LE et la problématique à laquelle nous sommes confrontés puisque, de prime abord il existe plusieurs ressemblances formelles. Il s'agit par conséquent de fournir une description de la *notion d'espace* en utilisant le *raisonnement qualitatif* pour proposer un *modèle des relations spatiales*. Dans la suite de ce chapitre, nous appellerons cette sous partie du modèle ELECA, basée sur le modèle LE, *modèle CLE*⁹. Fondamentalement le modèle des relations spatiales doit pouvoir répondre à la question : ces deux petits morceaux de courbes (amas de pixels) font-ils partie ou non d'un même élément curviligne sur l'image ?

Ainsi, pour atteindre l'objectif fixé à la page 160, il convient de spécifier alors un objectif particulier incontournable, à savoir *de chercher à établir les équivalences formelles entre le modèle LE et le modèle CLE*. Ainsi, si le modèle CLE peut être jugé comme étant formellement équivalent au modèle LE, il est possible de se servir de l'approche par contrainte du modèle LE pour faire le travail du modèle CLE, c'est-à-dire la reconstruction d'un environnement.

⁸Pour former une algèbre relationnelle les relations doivent être stables selon l'intersection (\cap), l'union (\cup), la composition (\circ) et la conversion (\curvearrowright) en plus de comporter les relations vide, universelle et identité.

⁹CLE : Cotteret–Ligozat–Edwards.

6.3.2.1 Évolution de la connaissance de l’environnement

Avant tout parallèle particulier, il faut constater que la progression de l’appréhension de la connaissance de l’environnement se fait de manière très différente dans notre cas vis-à-vis de celui du modèle LE. En effet, l’appréhension de la connaissance de l’environnement se fait à partir d’un point de manière circulaire dans le cas du modèle LE, figure 6.7, alors que dans notre cas il se fait en suivant un “fil”, figure 6.8, (contraint par le comportement saccadique de l’œil virtuel) ; autrement dit, la collecte de l’information locale se fait d’un côté depuis une “localité” 0D alors que dans le cas qui nous intéresse elle se fait depuis une “localité” 1D.

6.3.2.2 Du modèle LE vers la construction du modèle CLE

En reprenant le modèle LE il est possible d’identifier des aspects pertinents à notre domaine. Le tableau 6.1 résume les points les plus importants du modèle LE susceptibles d’être transposés dans le champ de notre problématique – voir figure 6.1a, page 163.

DOMAINE GÉNÉRAL	ASPECTS PARTICULIERS
<i>Espace</i>	– Voisinage, proximité spatiale
	– Trajectoire, localisation → cumul de relations
	– Panorama - cycle (visibilité)
	– Modèle pour chaque localisation
	– Barrières de visibilité, barrières de navigation
	– Description <i>cohérente</i>
<i>Qualitatif</i>	– Notion d’ordre unique, pas de valeur numérique
<i>Modèle de relations</i>	– Caractérisation des zones de stabilité

TAB. 6.1 – Points porteurs de LE vis-à-vis de la problématique

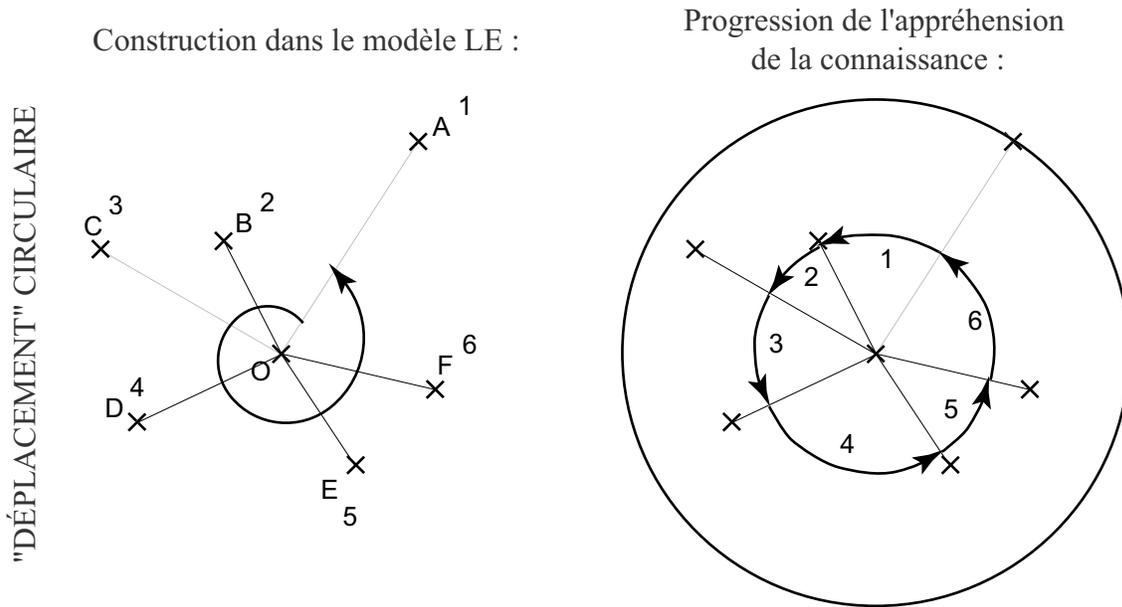


FIG. 6.7 – Progression de l'appréhension de la connaissance dans le modèle LE.

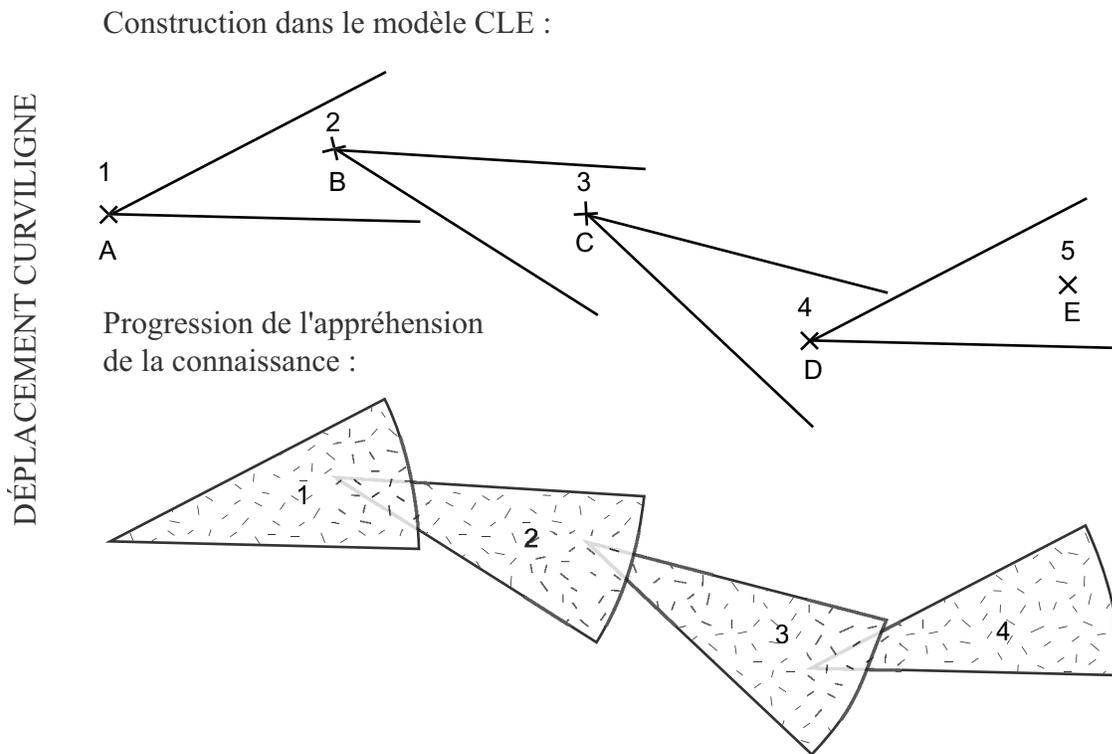


FIG. 6.8 – Progression de l'appréhension de la connaissance dans le modèle CLE.

Trois informations de départ se dégagent alors :

1. Le modèle LE a été construit en partant des notions de proximité et de voisinage, sans oublier l'orientation.
2. L'unité atomique de construction de l'environnement du modèle LE est un point (0D), alors que dans le cas de CLE¹⁰ il s'agit de morceau de courbe 1D.
3. L'évolution de la connaissance atomique (*i.e.* locale) se fait pour LE¹⁰ autour d'un point sans dimension alors que dans le cas de CLE elle se fait en suivant un parcours en une dimension sur l'image.

6.3.2.3 Décor formé d'informations locales

Unités atomiques et formalisme L'orientation instantanée (cf. page 154) dont nous avons parlé précédemment se calcule à partir de deux points. Nous pouvons faire très facilement l'analogie avec les dipôles de Moratz *et al.*, en gardant toujours à l'esprit que le dipôle ne représente que les extrémités de l'amas de points. Dans la suite de cette section nous utiliserons ce formalisme, illustré figure 6.9. Il s'agit de l'unité atomique du modèle CLE.

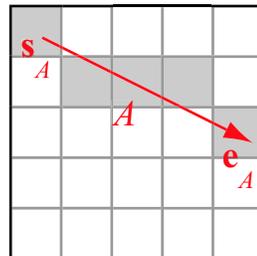


FIG. 6.9 – Transposition du formalisme de Moratz *et al.* à CLE.

En suivant le formalisme du calcul des dipôles il est possible d'ores et déjà de re-

¹⁰Afin d'alléger le texte on parlera indifféremment de *modèle LE* ou de *LE* pour désigner la même chose ; de la même façon nous utiliserons également l'acronyme *CLE* pour désigner le *modèle CLE*.

marquer deux propriétés spécifiques de CLE. D'une part, deux relations de \mathcal{D}_{24} seront difficilement identifiables dans le modèle CLE ; ce sont les relations *sese* et *eses*, où les deux points sont partagés par les deux dipôles. En effet, sur l'image, il n'est pas possible de distinguer deux éléments superposés¹¹. D'autre part, les huit relations de \mathcal{D}_{24} autres que *sese* et *eses* (figure 6.6) contenant la relation *e* ou *s* sont des cas particuliers de \mathcal{D}_{24} . En effet, ces relations $\{\text{ells}, \text{errs}, \text{lere}, \text{rele}, \text{slsr}, \text{srls}, \text{lsel}, \text{rser}\}$ présentent la particularité, par définition, que les deux dipôles partagent un point en commun. Lorsque, dans l'imagette, deux points appartiennent à deux amas de points, ils sont regroupés au sein d'un même élément par analyse en composantes connexes. Ces cas ne peuvent pas intervenir non plus lorsque deux imagettes sont connexes, figure 6.10a. La seule configuration alors les autorisant correspond au cas où deux imagettes se chevauchent et partagent un pixel candidat, figure 6.10b. L'identification de ces configurations est évidemment très importante car il convient de faire la fusion entre les 2 morceaux partageant un pixel commun.

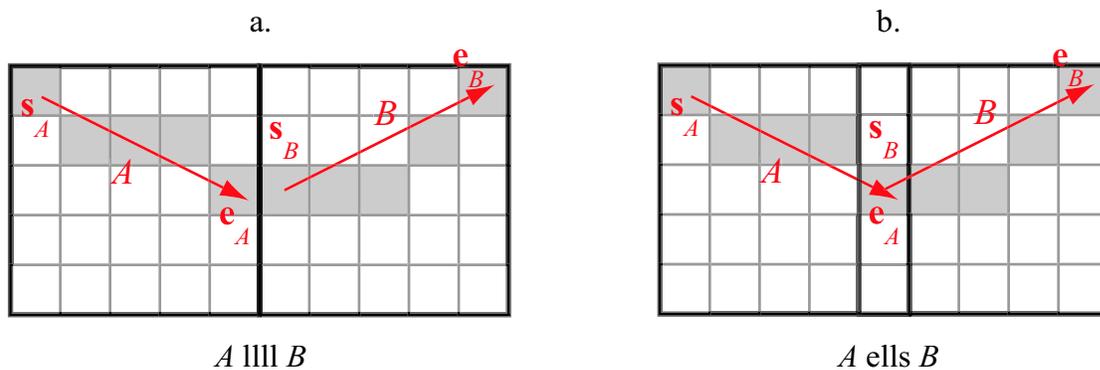


FIG. 6.10 – Cas de points communs à deux dipôles dans le modèle CLE.

Voisinage et proximité Le modèle LE a été bâti en partant des notions de voisinage et de proximité spatiales. En outre, ces notions construisent un modèle de relations spatiales entre un point et ses voisins ; par exemple selon un découpage Voronoï [Vor08].

¹¹Dans la très grande majorité des cas, il est toutefois possible de trouver des exceptions.

D'autre part, sachant que les unités atomiques sont les points dans le modèle LE et les dipôles dans le modèle CLE, on peut écrire :

$$\text{si } x_i \mapsto A_j \text{ alors } V_{LE} \mapsto \Omega_{c\hat{o}nes} \quad (6.4)$$

Où :

x_i est une localité du modèle LE, comme x_1, x_2, \dots, x_6 dans la figure 6.2

A_j est un dipôle extrait de l'image et $j \in [1..n]$ avec n le nombre de dipôles extraits de l'image en mode découverte.

\mapsto traduit la transposition du modèle LE vers le modèle CLE

V_{LE} est le voisinage défini dans le modèle LE

$\Omega_{c\hat{o}nes}$ est l'ensemble des cônes autour de la courbe qui délimitent les variations possibles de la courbe. Un cône est défini comme une oscillation possible autour de la tangente à la courbe.

Il faut noter une différence importante lors de la transposition des point x_i en dipôles A_j . Dans les deux modèles il s'agit de localités mais dans le cas de LE la notion de localité est plus restrictive. Les x_i sont des positions (voir figure 6.2) mais non des repères¹² comme les objets A, B, C, D, E et les barrières de la figure 6.2. En revanche les dipôles A_j sont des éléments faisant partie ou non du "paysage" dépendamment qu'ils sont des parties ou non d'un élément curviligne de l'image.

De ces définitions découlent deux choses. Premièrement, de par la définition des dipôles A_j qui portent intrinsèquement une orientation, $\Omega_{c\hat{o}nes}$ est différent suivant que l'on oriente le dipôle dans un sens ou dans l'autre, voir figure 6.11. Deuxièmement, nous parlons de "cônes" et non de "secteurs de disque" (plats) car le domaine d'étude n'est pas bidimensionnel. En effet, les dipôles que ne sont pas définis dans \mathbb{R}^2 l'espace plan

¹²*landmarks*

décrit par les coordonnées x et y mais bien dans \mathbb{R}^n , où $n > 2$ est un entier naturel¹³. Les dimensions supplémentaires sont par exemple la valeur spectrale, l'orientation, le contraste, la nature (supposée ou non) de l'objet, *etc.* On passe donc d'un modèle LE à deux dimensions à un modèle CLE à n dimensions. Pour une courbe donnée, en fonction de ces dimensions de définition propres, nous noterons son espace de définition \mathbb{D} ; \mathbb{D} est un sous espace¹⁴ de l'espace \mathbb{R}^n .

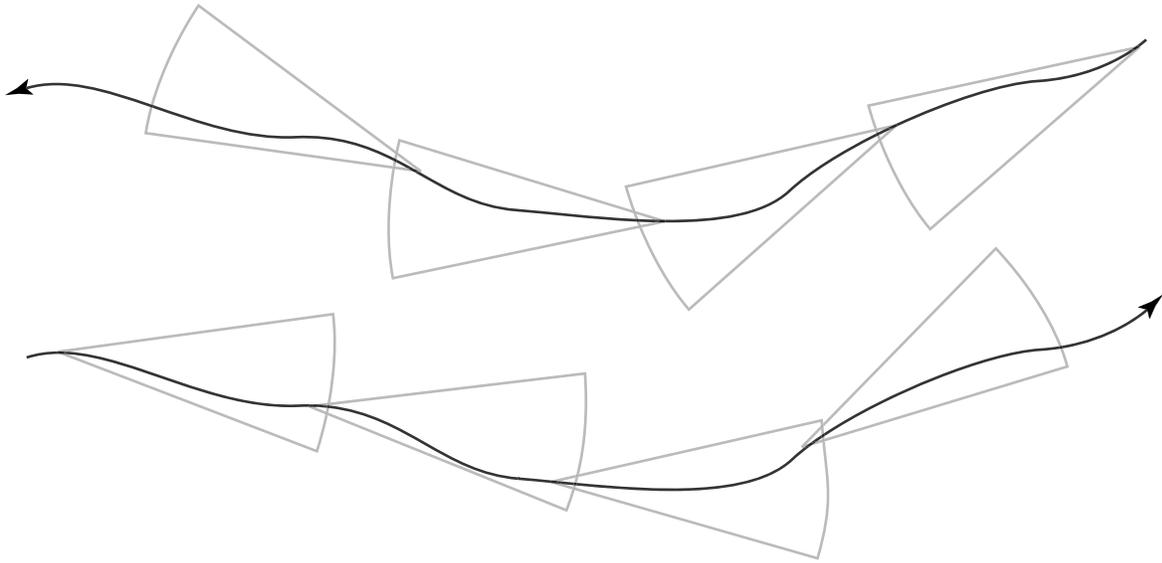


FIG. 6.11 – Deux voisinages pour une même courbe ayant deux orientations possibles.

Il est donc possible de donner une autre définition du voisinage, plus générale que celle illustrée dans la figure 6.11. Si l'on se place désormais dans l'espace à n dimensions évoqué plus haut alors le voisinage correspond à ce qui est illustré pour la figure 6.12, c'est-à-dire à une variation de la pente du dipôle le long de ce dernier – le long de son abscisse curviligne dans son espace de définition. Cette nouvelle définition diffère de la définition précédente car la contrainte est moins forte. En effet la courbe que décrit le dipôle pendant son déplacement (voir paragraphe suivant) peut complètement diverger et pourtant faire partie du même voisinage, tout cela par des déplacements

¹³Remarque : lorsque $n > 3$ ce ne sont plus des cônes mais nous garderons ce nom.

¹⁴Ou l'espace dans son ensemble.

infinitésimaux de la courbe.

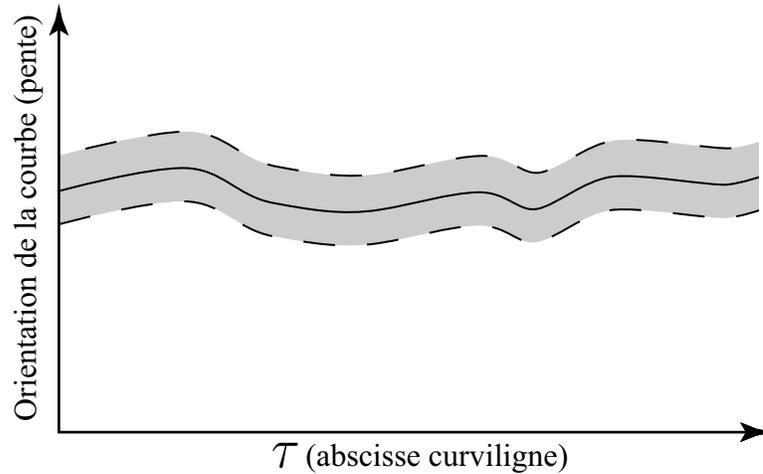


FIG. 6.12 – Voisinage (en gris) d'une courbe, cas général.

Il faut noter ici qu'il existe une divergence conceptuelle importante entre le voisinage du modèle LE et celui du modèle CLE. Le voisinage est déterminé, dans le modèle LE, par la relation spatiale entre les repères, ce qui logiquement devrait correspondre à des relations spatiales entre des dipôles et non autour de courbes solutions. Cela tient au fait que les dipôles A_j peuvent être à la fois des repères (et faire partie du paysage) et des localités. Ainsi, deux dipôles peuvent faire partie du même voisinage s'ils sont portés par le même élément curviligne. Nous distinguons par conséquent deux types de dipôles (encore une fois dipôle et amas de pixels sous-tendus) : d'un côté ceux qui font partie d'un élément curviligne de l'image et de l'autre, ceux qui ont été extraits "à tort", c'est-à-dire sans faire partie d'un élément curviligne. Nous noterons ces dipôles, respectivement, A^s et A^b , puisque certains sont solutions de notre modèle global et les autres sont assimilables à du bruit.

Déplacement, exploration de l'environnement Dans le modèle LE, un déplacement entraîne (en certaines circonstances) un changement de panorama, figures 6.2 et 6.3. Analytiquement parlant, un déplacement dans le sens de LE est un changement

continu des coordonnées x et y . Le déplacement, dans le modèle CLE est alors de manière logique un *changement continu des paramètres du dipôle*. Si l'on considère l'amas de pixels qui sous-tendent le dipôle comme un segment de courbe et si $c_{\mathbb{D}}$ est la fonction qui décrit la courbe sur son espace de définition \mathbb{D} (les dimensions vues plus haut), alors $\delta c_{\mathbb{D}}$ est un déplacement (infinitésimal) de la courbe dans \mathbb{D} . À partir de là, définir le panorama et les barrières devient beaucoup plus complexe que dans le cas en deux dimensions.

Le déplacement dans le modèle LE permet de passer d'une zone de stabilité à une autre. Les passages d'une zone de stabilité à une autre (que ce soit via une zone de transition ou via une porte) sont marqués par un changement à la fois de la *visibilité/panorama* et de *voisinage/proximité*.

Panorama Si l'on revient au problème initial, qui est de fusionner les petits morceaux de courbes extraits de l'image pour en arriver à une carte globale, c'est-à-dire à un ensemble de courbes (les éléments curvilignes) d'échelle supérieure (de grande taille) sur l'ensemble de l'image, il est possible de proposer un modèle de panorama. Les amas de pixels (qui sont des petits morceaux de courbes) sont stockés dans une liste de découvertes (voir figure 6.13), autrement dit, ils sont ordonnés. L'ordre de cette liste reflète le parcours que l'œil virtuel a réalisé sur l'image¹⁵. Ainsi, le parcours n'ayant pas été aléatoire (comportement en saccades, suivi d'un élément, *etc*), l'ordre de la liste contient intrinsèquement en soi de l'information. Plus les éléments sont proches dans la liste, plus la probabilité qu'ils appartiennent à la même courbe est grande, comme illustré dans la figure 6.13. On note O_n l'ordre du n ème point, alors pour le i ème point le panorama est défini comme $\|O_i - O_j\|$, Appelons *ordre relatif du dipôle A_j au dipôle A_i* cette différence. On la note $O_{i,j}$

¹⁵uniquement là où de l'information a été identifiée

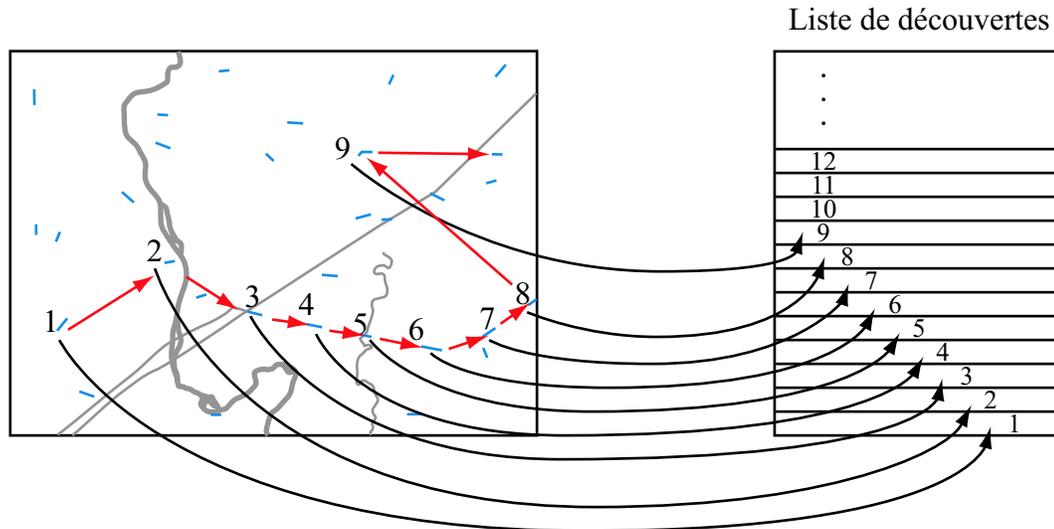


FIG. 6.13 – Reflet du parcours de recherche dans la liste des découvertes. 1, 2, 3... est l'ordre dans lequel s'est faite l'exploration de l'image.

Par conséquent, avec cette définition, toute valeur de l'ordre relatif à un amas de pixels proche définit le panorama. Par exemple, si i et k sont proches alors $O_{i,k}$, $O_{i,k-1}$ et $O_{i,k+1}$ ont des valeurs petites et proches, donc A_i , A_{k-1} et A_{k+1} font partie du même panorama, si tant est que ces derniers appartiennent à la même courbe solution.

Le voisinage ($\Omega_{cônes}$), le déplacement, et le panorama pour CLE ayant été définis plus haut, un changement de voisinage et donc de zone de stabilité (c'est-à-dire l'équivalent du passage d'une cellule Voronoï à l'autre) est un changement *brusque* d'un des paramètres de la courbe. Autrement dit, analytiquement parlant il s'agit d'une discontinuité de la fonction qui décrit le paramètre. Un "saut" de suivi de l'œil virtuel (c'est-à-dire d'un élément curviligne à un autre) est un bon indice d'une discontinuité; par exemple un changement drastique de la *signature spectrale*, de l'orientation, de la complexité de la courbe...

La figure 6.14 illustre ce dernier point. Dans l'espace $\mathbb{R}^2(x, y)$ 1, 2, 3 et 4 sont dans le même voisinage, mais si l'on ajoute une dimension à l'espace de définition de la fonction

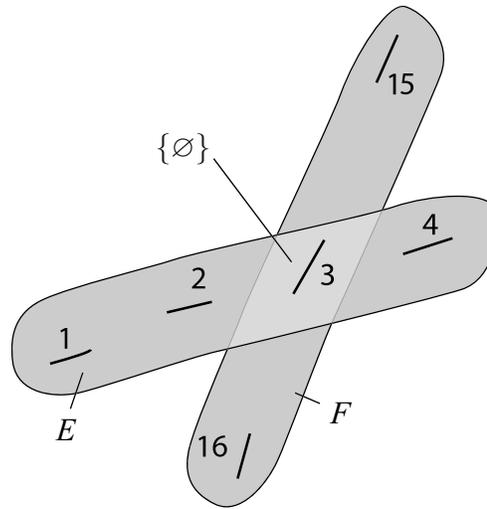


FIG. 6.14 – Deux voisinages possibles pour le même “morceau de courbe” 3. Finalement le morceau 3 appartient uniquement à l’ensemble F sur $\mathbb{R}^3(x, y, o)$

qui décrit la courbe, en l’occurrence l’orientation, 3 n’est plus alors dans le voisinage de 1, 2 et 4 mais dans celui de 15 et 16. L’intersection des ensembles E et F dans l’espace $\mathbb{R}^3(x, y, o)$ où o est l’orientation est donc l’ensemble vide. $E \cap F = \{\emptyset\}$.

Si maintenant $\{1, 2, 3, 4, \dots, 15, 16\}$ représente l’ordre O_n de la liste des découvertes, alors 1, 2, 3 et 4 sont dans le même panorama tout en n’étant pas dans le même voisinage. Autrement dit, ils ont peu de chance d’être portés par la même courbe solution – élément curviligne de l’image. En revanche, les amas 16, 3 et 15 sont à l’inverse : éléments d’un même voisinage mais pas d’un même panorama. Ainsi l’on voit que les définitions que nous avons données du voisinage et du panorama répondent bien au sens commun. En effet il est possible d’avoir, dans la vie courante, un objet qui fait partie de notre panorama au loin mais qui n’est pas dans notre voisinage. Inversement, un objet peut être proche, dans notre voisinage mais caché, absent de notre panorama.

6.4 Synthèse et équivalences

6.4.1 Ébauche du modèle CLE

MODÈLE LE	Notion	MODÈLE CLE
\mathbb{R}^2	<i>Espace</i>	\mathbb{R}^n
x_i	<i>Localité</i>	A_j (A^s et A^b)
Barrières et objets	<i>Repère</i>	A^s
V_{LE}	<i>Voisinage</i>	$\Omega_{cônes}$
$\{x_1, x_2, \dots, x_n\}$	<i>Déplacement</i>	$\delta_{C_{\mathbb{R}^n}}$: chgt <i>continu</i> des paramètres du dipôle
Cycle	<i>Panorama</i>	$O_{i,x}$: ordre relatif au dipôle A_i
De visibilité	<i>Barrière</i>	Changement de panorama
De navigation	<i>Barrière</i>	Changement de voisinage \rightarrow discontinuité

TAB. 6.2 – Transposition des notions de LE vers CLE

À la lumière de ce qui a été dit plus haut, le modèle CLE vis-à-vis du modèle LE (tableau 6.2) s’articule donc comme suit :

1. Le groupe de points, dans le modèle CLE, est l’unité atomique : on l’appelle dipôle A_j ; cela correspond aux localités notées x_i dans le modèle LE.
2. À une localité (liée à une position) on associe un panorama unique.
3. Le schéma global que l’on recherche est une collection de dipôles A^s avec des relations spatiales *déterminées* entre eux.
4. La notion de zone de stabilité est un “voisinage” de dipôles A^s ayant de manière approximative le même panorama.
5. La démarche de construction d’un modèle de relations spatiales (dans le but d’aboutir à une fusion des informations locales) consiste à :
 - (a) se servir de déplacements de dipôles A^s de l’un à l’autre pour constituer les relations spatiales, ceci en adaptant le calcul des dipôles de Moratz *et al.*

- (b) considérer qu'une courbe solution est le sous ensemble des relations spatiales au sein d'un même voisinage.

Une manière de voir le schéma global (point 3) selon un angle plus analytique est la suivante. Les courbes que nous cherchons à identifier (les éléments curvilignes) sont l'ensemble de solutions dans l'espace \mathbb{R}^n ; un exemple visuel est donné dans la figure 6.15. Construire le schéma global revient donc à identifier l'appartenance ou non d'un dipôle

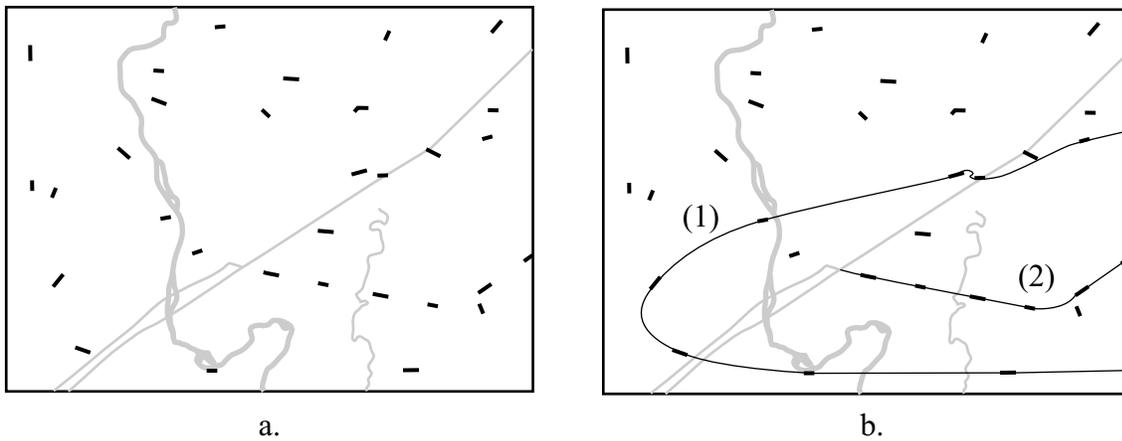


FIG. 6.15 – Exemple de courbe solution. a. Exemple habituel (figure 3.5) sur lequel les amas (dipôles) issus du mode découverte sont identifiables. b. Deux courbes pouvant lier les dipôles; conformément à l'exemple (figure 3.6), (1) *n'est pas* une courbe solution alors que (2) *est une courbe solution*.

A_j à une grande courbe. Les relations spatiales déterminées dont on parle dans le point 3 sont donc :

1. pour un dipôle A_j donné son appartenance ou non à l'ensemble des courbes solutions;
2. pour deux dipôles A_j leur appartenance ou non à la même courbe solution.

6.4.2 Perspectives pour le modèle CLE

Le modèle ELECA permet de “générer” un ensemble d’informations locales au sein de l’image. Ces informations sont obtenues à partir d’un processus particulier de recherche de l’information contrôlée par un survol de l’image inspiré du regard humain. En outre ces informations sont extraites d’une image segmentée au préalable, voir les deux chapitres précédents.

Nous nous sommes attachés dans le présent chapitre à bâtir un sous modèle au modèle ELECA¹⁶, le modèle CLE afin de fournir une stratégie de fusion judicieuse des informations locales. La finalité de cette fusion est de générer les éléments curvilignes de l’image en rassemblant de manière appropriée les informations locales pertinentes d’une part et en écartant les artéfacts issus des deux premières étapes de modèle ELECA d’autre part.

Pour réaliser cette fusion nous avons montré qu’il est possible d’établir des équivalences formelles entre un modèle de raisonnement qualitatif spatial, le modèle LE, et le modèle CLE. La description originale du problème et l’utilisation d’un modèle issu du raisonnement spatial qualitatif ouvre des perspectives intéressantes pour la fusion d’informations lorsqu’il est possible d’établir un modèle de relations spatiales entre les objets. Le cas de cette thèse est de ce point vue intéressant car il essaye de le faire dans un cas concret. L’établissement complet des équivalences formelles entre les deux modèles est néanmoins nécessaire pour pouvoir utiliser l’approche par contraintes du modèle LE qui permet de reconstruire l’environnement. L’objectif de cette transposition est d’aboutir à une reconstruction des éléments curvilignes du modèle CLE.

¹⁶Le modèle ELECA comporte par conséquent trois parties : 1 - guidage de la recherche d’information dans l’image, 2 - extraction locale d’informations dans l’image et 3 - fusion des informations locales pour génération de la carte globale.

Troisième partie

Développement logiciel

Chapitre 7

Conception du système

*Imagination is more important than knowledge.
Albert Einstein, On Science.*

Ce chapitre est consacré à la conception d'un système logiciel qui supporte le modèle ELECA et les orientations prônées. Dans un premier temps, nous montrons rapidement comment il est possible de passer du modèle ELECA à une solution informatique. Dans un deuxième temps, la portée du logiciel est présentée brièvement. Puis, dans un troisième temps nous présentons le cheminement de conception du logiciel à proprement parler. Cette solution logicielle, baptisée VISION, a été développée suivant le formalisme UML. Conformément à l'approche UML, le travail a été réalisé par cycles et en commençant par les spécifications. En d'autres termes, et selon la présentation qui en est faite dans ce chapitre, le diagramme d'activité détaillé a été spécifié, puis à partir de ce dernier l'architecture de VISION a été conçue en s'appuyant sur un diagramme de classes. Des boucles de rétroaction ont été menées constamment sur ces deux diagrammes.

7.1 Du modèle ELECA au modèle informatique

7.1.1 Synthèse du modèle ELECA

Le modèle ELECA¹, qui a été développé au cours des chapitres précédents, peut être synthétisé de manière graphique par la figure 7.1. Cette vue d'ensemble résume correctement la philosophie qui a sous-tendu l'élaboration du modèle.

D'une manière générale sur la figure 7.1, il est facile d'appréhender les boucles de rétroaction du modèle aussi bien au niveau des données (mises à jour) que de l'exploration dans l'image. La stratégie employée permet par conséquent, d'une part une grande souplesse quant au "parcours" dans le modèle², et, d'autre part, une grande robustesse du modèle – puisque des retours sont autorisés limitant ainsi la propagation des erreurs.

Le cœur du modèle ELECA apparaît comme un compromis constant entre les influences ascendante et descendante. Ainsi, pour explorer l'image (données raster) on note, d'un côté, l'influence des données vectorielles (et la chaîne de traitements présentée en haut à gauche sur la figure) qui amènent de l'information *a priori* et, de l'autre, l'influence du modèle de *saccades* qui est contraint par l'information extraite directement de l'image – information postérieure à la création de la base de données.

Enfin le dernier point à souligner ici est l'importance qui est accordée à l'information locale. Visuellement sur la figure 7.1 on voit que le traitement de l'information locale est au centre du modèle. Dans un premier temps, l'information globale cherche

¹ELECA : Éléments Linéaires Extraits par Calcul Attentionnel – *Extraction of Linear Elements by Cognitive Approach*.

²Par *parcours dans le modèle* nous entendons la succession de traitements qui sont faits. Graphiquement sur la figure 7.1, par quelles boîtes et flèches va-t-on passer ?

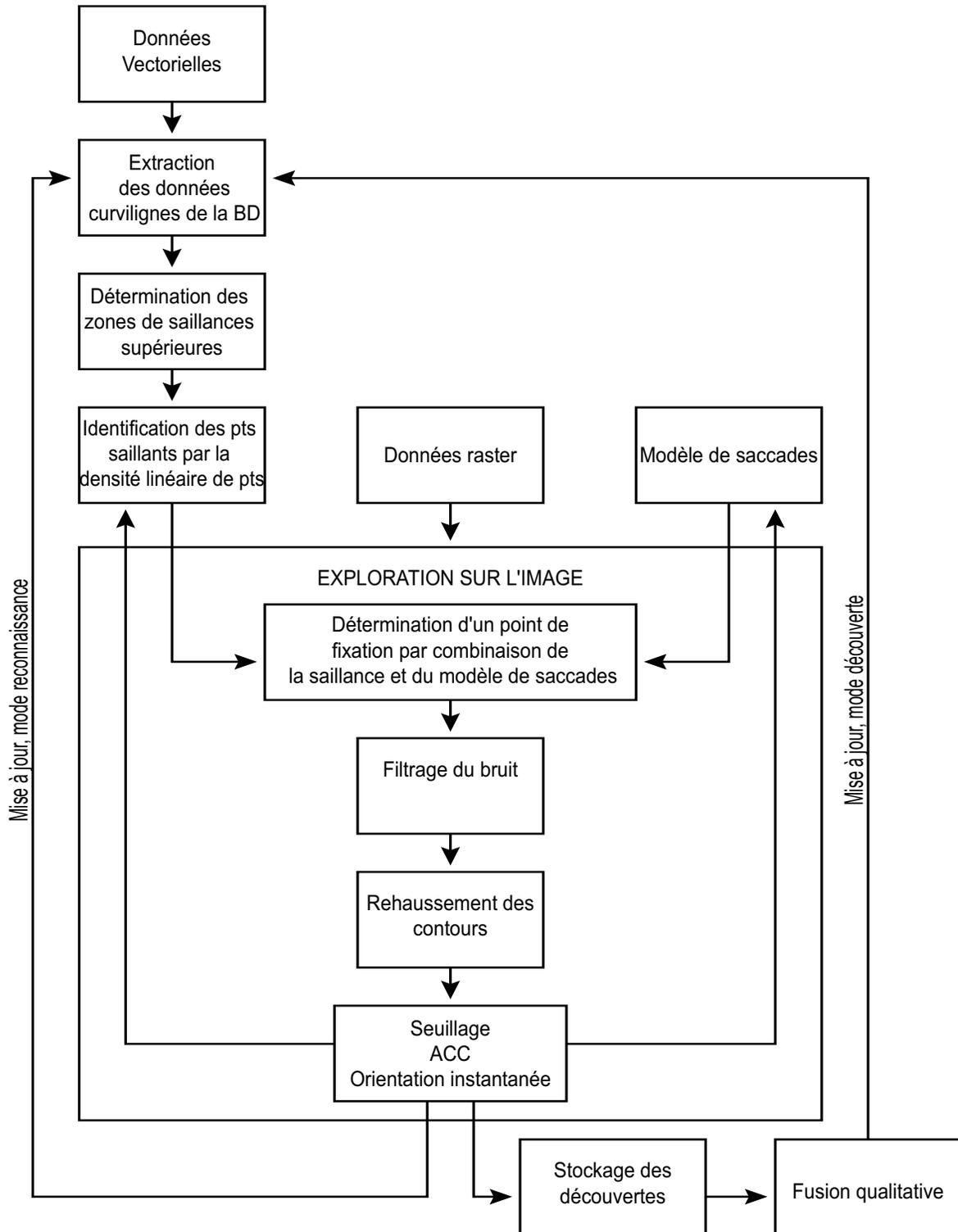


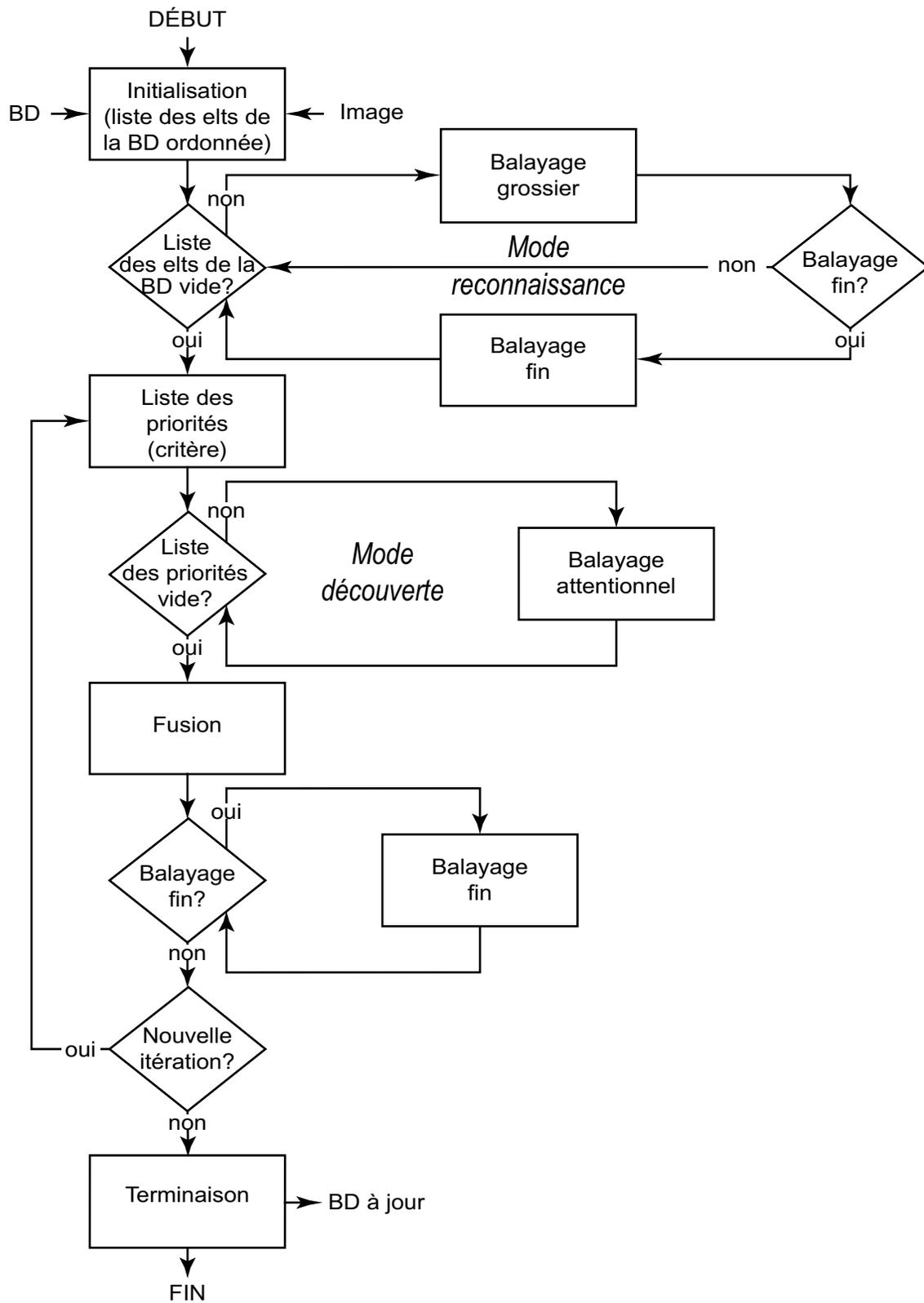
FIG. 7.1 – Vue d'ensemble du modèle ELECA.

à aider au guidage le regard vers l'information locale, puis, dans un second temps c'est l'information locale qui sert à mettre à jour l'information globale. L'algorithme général est une transposition de cela, voir section suivante.

7.1.2 Algorithme général tiré du modèle

À partir de la lecture que nous venons de faire de la figure 7.1, l'algorithme général de la solution logicielle a été élaboré, algorithme 7.1. Cela a été fait en "basculant" les traitements globaux des informations de manière linéaire ; sur l'algorithme 7.1 il s'agit de la colonne qui va de "DÉBUT" à "FIN". Cette stratégie a été adoptée pour répondre à la finalité du logiciel, à savoir mettre à jour une base de données géographiques. Les entrées sont la base de données à mettre à jour d'une part et l'image d'autre part. Les sorties sont les données de la base mises à jour – voir aussi le but général du projet figure 1.4.

Les boucles sur la partie droite de l'algorithme 7.1 correspondent aux traitements qui sont effectués localement. La séquence (l'enchaînement) entre les différentes parties du modèle ELECA qui n'apparaît pas sur la figure 7.1 est mise en évidence dans cet algorithme. L'algorithme 7.1 est une transposition de l'architecture générale du modèle ELECA à des fins informatiques.



ALG. 7.1 – Algorithme général.

7.2 Portée du logiciel

La solution logicielle qui a été développée est une étape vers la validation du modèle ELECA. Trois sous étapes ont été réalisées dans notre travail : premièrement transposition du modèle en algorithme général (section précédente), deuxièmement élaboration d'une architecture dédiée (section suivante) et enfin troisièmement (chapitre suivant) développement d'algorithmes spécifiques directement implantables. La solution logicielle qui est présentée ici, couvre une grande partie du modèle et a pour vocation de pouvoir supporter les développements ultérieurs du modèle.

La conception du logiciel (nous appellerons le logiciel également *VISION* dans la suite du document) a été menée en utilisant le formalisme et les techniques de modélisation UML³ [Mou02] [FL99] [EP98]. *VISION* a été développé en langage C++ pour permettre, d'une part, d'appliquer la conception objet de UML, et, d'autre part, d'avoir une vitesse d'exécution la meilleure possible. Nous rappelons que le contexte général est la mise à jour d'une base de données embarquée pour la recherche et le sauvetage. Cette base de données devant être mise à jour extrêmement rapidement, il importe que non seulement le modèle soit simple comme nous l'avons vu au cours des chapitres précédents mais également que la solution logicielle soit développée avec des solutions efficaces dans un langage efficace.

En outre, le logiciel a été conçu et développé avec un souci constant de réutilisation et de portabilité⁴ ; ainsi, aucune bibliothèque spécifique n'a été utilisée. Le code source est le plus standard possible pour répondre aux exigences précitées. Ce travail de doctorat

³U.M.L. : *Unified Modeling Language*.

⁴Ici *portabilité* s'entend dans le sens que le code n'est pas dépendant d'un système d'exploitation particulier. Cette thèse faisant partie du réseau d'un projet beaucoup plus large, il importe que la solution ne soit pas liée à une plate-forme particulière.

s'inscrit dans un projet beaucoup plus grand⁵ qui vise (entre autre) à aider le pilotage des hélicoptères de sauvetage, et ceci depuis la mise à jour de la base de données embarquée (notre projet) jusqu'à la vision augmentée du pilote. Dans ce cadre, une documentation technique de spécifications architecturales et fonctionnelles complète a été rédigée en anglais. L'annexe B donne des exemples de ce document.

7.3 Conception logicielle

La conception de l'architecture⁶ logicielle s'est faite en partant des entrées et des sorties identifiées du programme. Les entrées et sorties ont été identifiées pour notre modèle dans le chapitre 3 ; elles sont également résumées dans la figure 1.4, page 15.

Un *use case*⁷ très simple peut être tiré de ces paramètres, figure 7.2. Ce *use case* se décline en quatre étapes de base :

1. L'utilisateur démarre le programme.
2. L'utilisateur passe les données géographiques vectorielles au programme.
3. L'utilisateur passe l'image de télédétection (raster) au programme.
4. L'utilisateur obtient des données mises à jour du même format que les données géographiques introduites à l'étape 2.

Il est ainsi facile de donner un diagramme d'activité⁸ concis, figure 7.3. Le programme charge les données, les traite et retourne un jeu de données mis à jour.

⁵Projets ENV#19 et MNG#ELD du réseau de centres d'excellence du Canada GEOIDE.

⁶“*Une architecture est la structure organisée d'un système*” [Mou02], traduction libre.

⁷*Use case* : cas d'utilisation ; c'est le terme anglais qui est le plus généralement utilisé.

⁸Un diagramme d'activité permet de saisir des actions et leurs résultats [Mou02].

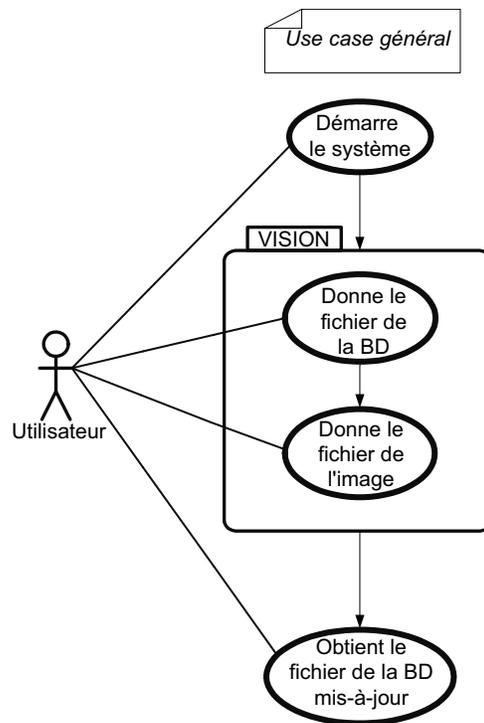


FIG. 7.2 – Use case général

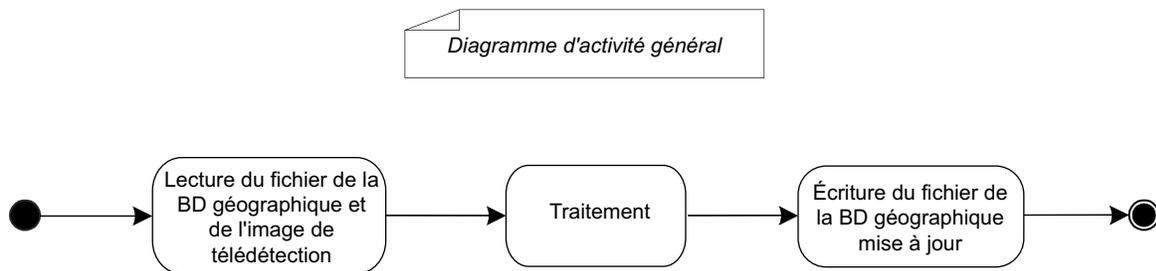


FIG. 7.3 – Diagramme d'activité général

7.3.1 Étapes de traitement — Diagramme d'activité

À partir des deux schémas précédents et de l'algorithme général du Modèle ELECA, figure 7.1, nous avons établi le diagramme d'activité figure 7.4 qui permet de définir des modules spécifiques. En dehors de la phase de chargement, huit modules numérotés de M1 à M8 sont identifiés ; les décisions sont symbolisées par les losanges et numérotées

de T1 à T4.

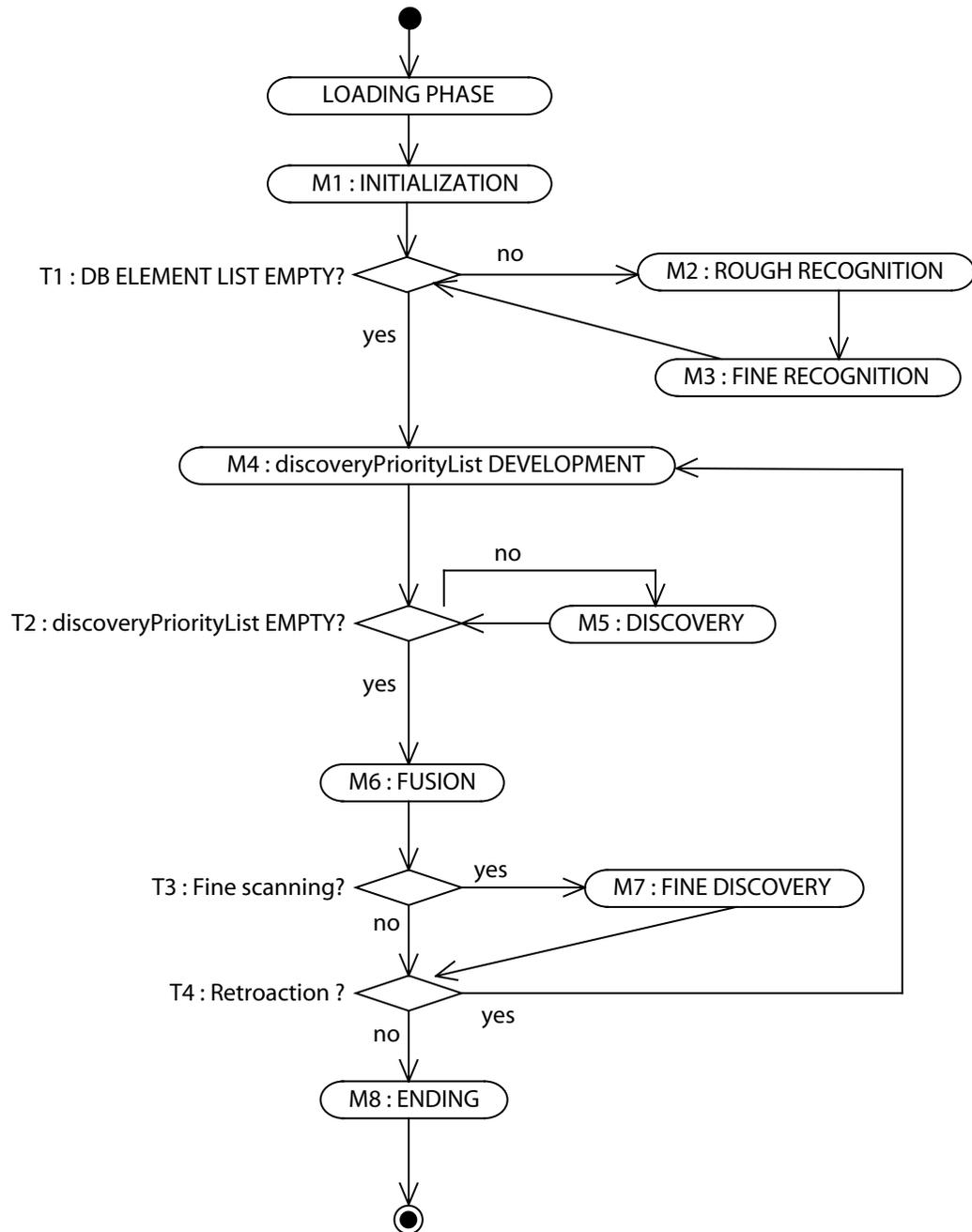


FIG. 7.4 – Diagramme d'activité présentant les différents modules de VISION

M1 : Initialisation des données (intermédiaires ou non) pour le traitement.

T1 : La liste des éléments pertinents de la base de données originale a-t-elle été entièrement parcourue ?

- M2** : Exploration de l'image en mode reconnaissance grossier.
- M3** : Exploration de l'image en mode reconnaissance fin.
- M4** : Développement de la liste des priorités pour le mode découverte.
- T2** : La liste des priorités pour le mode découverte a-t-elle été entièrement parcourue ?
- M5** : Exploration de l'image en mode découverte – balayage attentionnel figure 7.1.
- M6** : Fusion des informations locales collectées durant le mode découverte.
- T3** : Parcourir de manière fine les éléments fusionnés ?
- M7** : Exploration de l'image en mode fin pour les éléments fusionnés.
- T4** : Mener une nouvelle exploration en mode découverte à partir des informations obtenues précédemment ?
- M8** : Terminer le programme, comprenant entre autre l'écriture de la base de données mise à jour.

La suite de cette partie 7.3.1 revient plus en détail sur chacun des modules principaux. Les figures sont des extraits d'un très grand diagramme d'activité (format A0) non représentable dans cette thèse. Certaines figures se trouvent dans l'annexe C pour plus de lisibilité. Nous décrivons les étapes de déroulement de VISION, les spécificités d'implémentation (en particulier algorithmiques) et d'implantation sont abordées dans la partie 8.1. Les étapes de certains modules se retrouvent parfois dans plusieurs modules. Nous ne les décrivons que la première fois qu'ils sont rencontrés dans notre lecture du diagramme.

7.3.1.1 Module de chargement

Le module de chargement permet de démarrer le programme, figure 7.5.

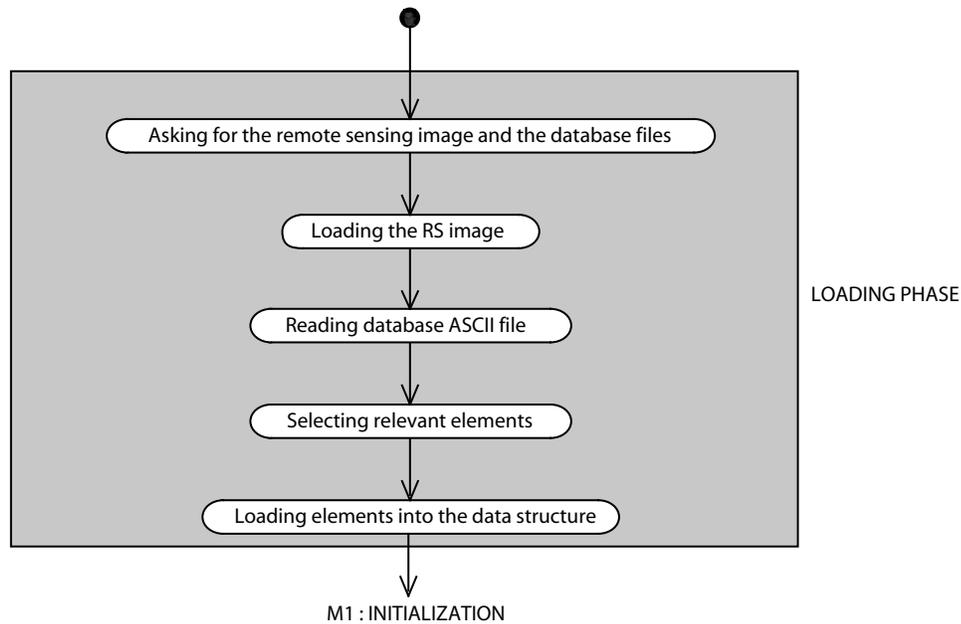


FIG. 7.5 – Module de chargement

Lecture des paramètres d'entrée Cette première phase permet tout d'abord de recueillir l'information pour trouver les fichiers décrivant la base de données et l'image de télédétection. Nous verrons plus loin dans ce chapitre que cela se fait via un fichier XML. Dans l'environnement *windows*, une petite interface a été codée en *visual basic*, figure C.1 page 312. De nombreux autres paramètres sont également chargés; ils comprennent la stratégie algorithmique générale (quels sont les modules utilisés pour une exécution donnée?), des variables (taille de l'œil virtuel, comportement général de celui-ci, *etc*), des traitements ou non sur les données (application de l'algorithme de Douglas-Peucker aux données vectorielles par exemple) ...

Chargement de l'image de télédétection (RS Image) L'image de télédétection est chargée en mémoire. Elle restera résidente en mémoire (plutôt que rester sur le disque) durant toute l'exécution du programme pour gagner en rapidité en limitant le nombre d'accès disque.

Lecture de la base de données géographiques La base de données est lue. On utilise pour cela un format d'exportation ASCII classique d'un SIG.

Sélection des éléments pertinents *A priori* la base de données géographiques contient de nombreux objets ; notre application se concentre sur les éléments curvilignes. Une sélection des routes, des cours d'eau et des lignes à haute tension est effectuée⁹.

Chargement des éléments dans la structure de données Tout comme pour l'image de télédétection, les éléments sélectionnés à l'étape précédente sont chargés en mémoire et restent résidents durant toute l'exécution du programme – s'ils n'ont pas été détruits au cours du traitement car non identifiés sur l'image. La structure de données est expliquée en profondeur au paragraphe 7.3.2.1.

7.3.1.2 Module M1 : Initialisation

Le module d'initialisation (figure 7.6) permet de calculer les paramètres intermédiaires dont VISION a besoin pour commencer l'exploration de l'image en mode reconnaissance¹⁰.

⁹D'autres éléments curvilignes peuvent être envisagés comme les limites de très grands éléments surfaciques (lacs, parcelles, *etc*) mais cela n'est pas traité à l'heure actuelle, le problème étant de déterminer l'échelle des objets pour savoir s'il est pertinent de faire le traitement à l'échelle de l'image donnée.

¹⁰Nous rappelons que, fondamentalement, le mode reconnaissance sert à identifier les éléments de la base de données sur l'image.

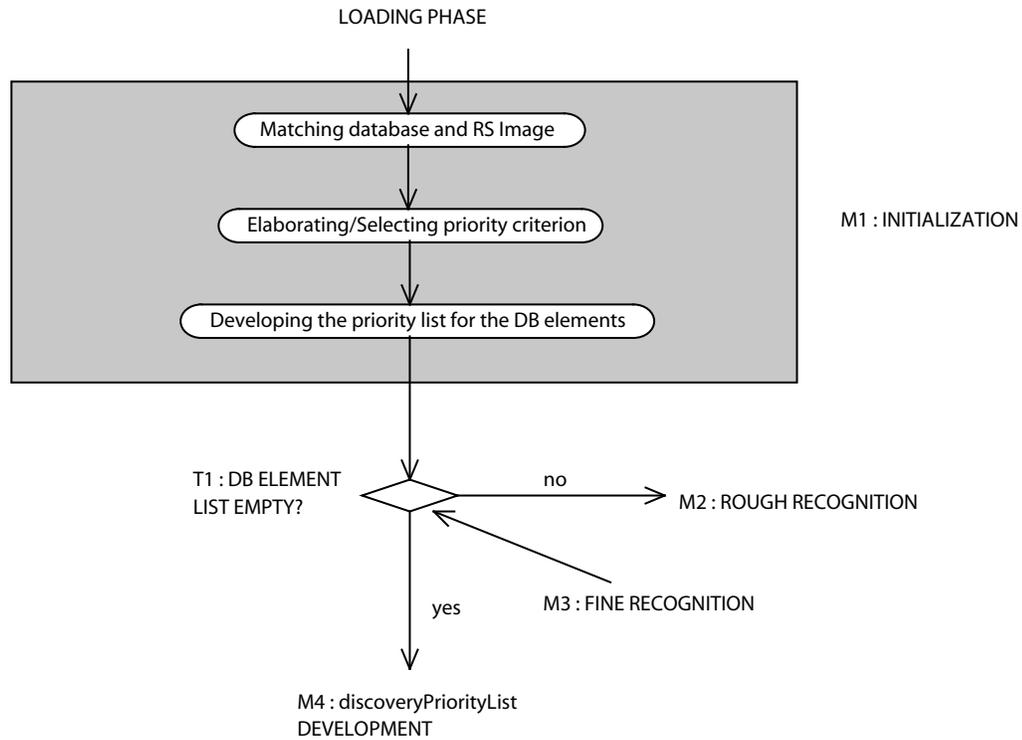


FIG. 7.6 – Module M1 : initialisation

Mise en concordance de la base de données et de l'image de télédétection

Comme nous l'avons évoqué précédemment, les systèmes de coordonnées et datums¹¹ peuvent être différents pour l'image et la base de données. Il s'avère donc indispensable de faire une transformation des coordonnées si c'est le cas pour l'une ou l'autre des données.

Élaboration/sélection du critère de priorité

Cette partie conditionne la stratégie de recherche de l'information dans l'image en mode reconnaissance. Il s'agit de déterminer (et/ou de calculer), en fonction des paramètres passés par l'utilisateur, quel va être le critère de priorité que l'on va appliquer aux données vectorielles chargées en mémoire. Autrement dit comment va-t-on calculer la saillance de ces éléments sur

¹¹un datum est système de référence géodésique. C'est-à-dire un "système de référence constitué de l'ensemble des conventions qui permettent d'exprimer, de façon univoque, la position de tout point sur la surface terrestre" [Off05a].

l'image. La première approche que nous avons adoptée dans ce travail est d'utiliser les propriétés des courbes simplifiées par l'algorithme de Douglas–Peucker – cf §4.3.2.4, [DP73] et annexe A.1. En pratique l'utilisateur demande de simplifier la base de données par l'algorithme de Douglas-Peucker si la base ne l'est déjà pas, ou qu'il l'ignore ou encore que la tolérance appliquée ne lui convient pas. Les données vectorielles en mémoire sont, par conséquent, modifiées si besoin en appliquant l'algorithme de Douglas-Peucker.

Développement de la liste des priorités pour les éléments en mémoire La liste des priorités pour les éléments chargés en mémoire est générée à cette étape. Il s'agit de faire une liste ordonnée des éléments, le premier étant le plus saillant. Dans notre cas les éléments sont classés par ordre croissant de densité linéaire, conformément au modèle – cf §4.3.2.4.

7.3.1.3 Module M2 : Reconnaissance grossière

Le module M2 de reconnaissance grossière est le “cœur” du mode reconnaissance, figure 7.7. Il s'agit d'une boucle qui permet de parcourir les points d'un élément dans la liste des priorités. Cette boucle est elle-même imbriquée dans une boucle plus grande $T1 \rightarrow M2 \rightarrow M3 \rightarrow T1$ qui permet de visiter tous les éléments chargés dans la liste des priorités en fonction de leur saillance décroissante.

Développement d'une liste de priorité pour les points de l'élément courant

À cette étape de l'exécution du programme, la liste des priorités contient des éléments curvilignes ordonnés selon leur complexité. Il faut, pour chaque élément, ordonner les

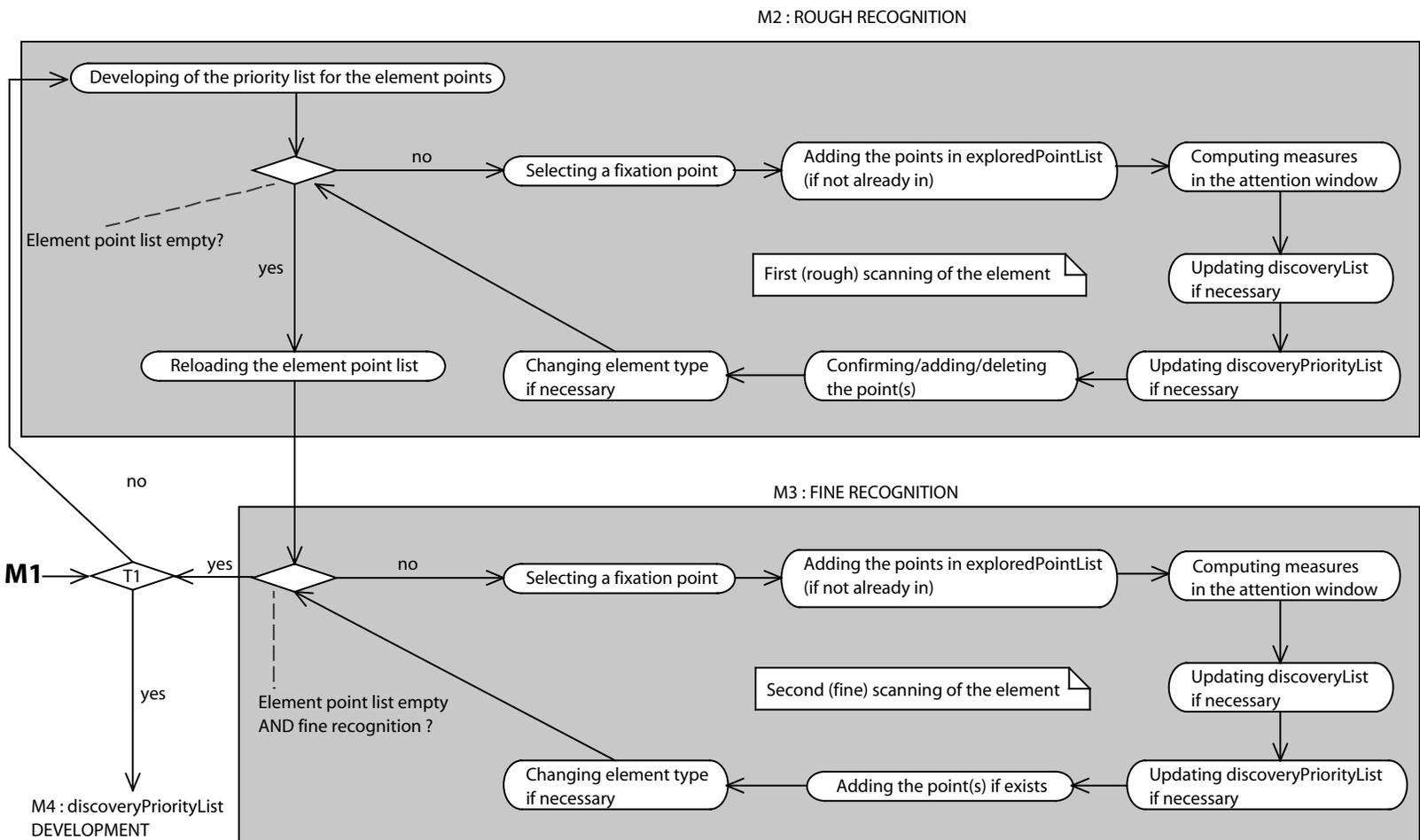


FIG. 7.7 – Modules M2 –Reconnaissance grossière– et M3 –Reconnaissance fine.

points à parcourir. En fonction de ce que désire l'utilisateur¹², il peut commencer par le premier point par défaut ou bien par l'extrémité la plus ou la moins complexe – le calcul se fait toujours en fonction de la densité linéaire de points, mais cette fois ci au sein d'un même élément.

Nœud décisionnel La boucle se fait tant que les points de l'élément n'ont pas tous été parcourus

Détermination du point de fixation Le programme, à ce point, possède les deux éléments, vus dans le chapitre 4, et nécessaires à la détermination du point de *fixation* du regard. En effet, conformément au modèle ELECA, le guidage du regard virtuel se fait en fonction du modèle de comportement saccadique et de la liste des priorités¹³. Selon les contraintes évoquées page 135 et suivantes, le point de fixation du regard sur l'image est déterminé.

Ajout du point dans la liste des points explorés Les points de l'image explorés par le regard virtuel peuvent être mémorisés. Cela permet de soutenir un mécanisme d'inhibition de retour sur les points visités.

Calculs sur l'image Une fois déterminé le point de fixation, il convient de faire le traitement de l'image à ce point ; autrement dit sur l'imagette. Le traitement suit le modèle élaboré dans le chapitre 5. Parallèlement au traitement d'image et de manière dynamique se fait la création éléments des extraits, dans la structure de données du

¹²Dans tous les cas se sont des paramètres donnés avant l'exécution du programme ; bien évidemment plus l'utilisateur demande de traitements plus le temps de réponse sera élevé.

¹³Le programme peut également vérifier que le point n'a pas déjà été exploré.

programme. Cette étape est l'une des plus importantes de VISION : elle se retrouve plusieurs fois au cours de l'exécution. Nous revenons plus en détail sur ce point dans la section consacrée à l'implantation.

Mise à jour de la liste des découvertes Si de nouveaux éléments, qui ne font pas partie de l'élément recherché, ont été identifiés¹⁴, ils sont stockés dans la liste des découvertes qui servira ultérieurement dans le mode découverte.

Mise à jour de la liste des priorités pour le mode découverte Dans le cas où des découvertes ont été faites, la liste des priorités pour ce mode est mise à jour.

Mise à jour de la structure de données En fonction de ce qui a été trouvé sur l'image, la structure de données est mise à jour. Autrement dit, pour le point de l'élément considéré à un moment donné, il est confirmé, ajouté ou supprimé de la structure. Les cas possibles sont les suivants :

Confirmation : L'élément a été trouvé sur l'image. Il s'agit toujours d'un amas de points par définition. Le point est donc confirmé et les points qui "l'accompagnent" sont ajoutés.

Ajout : Le point n'existait pas dans la base de données. Typiquement il s'agit d'une fixation du regard déterminée par le comportement du regard et non par la liste des priorités. Il est ajouté. Il peut également s'agir d'un point faisant partie d'un amas englobant un point de la base de données, ou bien d'un point d'un amas

¹⁴Cette identification se fait par rapport à une gamme de valeurs. Autrement dit, par exemple si l'orientation de l'amas est trop différente de celle au point de fixation précédent, l'amas sera considéré comme ne faisant pas partie de l'élément cible (suivi) ; le même critère (d'appartenance à une gamme de valeurs proches) est utilisé pour les autres dimensions de l'élément – intensité, nature supposée. . .

très proche (*i.e.* dans la même imagerie) du point attendu de la base de données, ce dernier n'apparaissant pas dans l'image il n'aura pas été confirmé¹⁵.

Suppression : L'élément qui était dans base de données ne retrouve pas dans l'image, il est supprimé.

Une séquence suppression-ajout sur des points très proches est la traduction que l'élément recherché n'était pas à l'emplacement attendu (*i.e.* l'emplacement fourni par la base de données) mais proche dans le corridor de recherche. Ce cas de figure traduit une imprécision de la base de données ou une erreur de mise en concordance (mauvais géoréférencement par exemple) des données vectorielles avec la donnée raster.

Mise à jour des attributs En fonction des calculs menés sur l'imagerie, il est possible que la nature supposée de l'élément ait changé ou ait été déterminée en fonction de son comportement spectro-spatial sur l'image. Un exemple typique est la traversée de reliefs de manière rectiligne qui correspond à des lignes de haute tension.

7.3.1.4 Module M3 : Reconnaissance fine

Le module M3 de reconnaissance est "facultatif", figure 7.7. Il s'agit de suivre l'élément précédemment reconnu sur l'image de manière plus fine. La plupart des étapes de ce module ressemblent à celle du module précédent. La différence principale vient du comportement d'exploration adopté. Deux stratégies sont possibles pour ce faire. Soit on applique un suivi de contour classique en traitement d'image, avec des points d'ancrage déterminés par les résultats du module précédent. Soit on sous échantillonne de place en place entre deux points d'ancrage.

¹⁵On se situe ici dans un cas d'erreur de géoréférencement entre l'image et la base de données.

Il est évident que ce module affecte énormément la performance du logiciel. Il pourra être utilisé dans les cas où le traitement peut prendre plus de temps. Le résultat évidemment s'en trouvera amélioré.

La dernière étape différente du module M3 vis-à-vis du module M2 est qu'il ne peut, par définition, y avoir qu'ajout de points. La suppression ou la confirmation sont par essence impossibles dans ce cas de figure.

7.3.1.5 Module M4 : Développement de la liste des priorités pour le mode découverte

Ce module séquentiel très simple permet de générer la liste des priorités pour le mode découverte, figure C.2 page 313. Ce module met en œuvre la partie du modèle que nous avons vue précédemment sur l'exploration en mode découverte et illustré par la figure 4.12, page 128. Cela se fait sur la liste des priorités pour le mode découverte déjà existante (vide ou non) issue des modules M2 et M3¹⁶ et ceci en deux étapes. D'abord une liste de points le long de la périphérie est générée. Cette liste de points est insérée à la suite des points qui existaient éventuellement dans la liste. Autrement dit, le pourtour de l'image est moins saillant que les amorces d'éléments supposés. Enfin une réalisation aléatoire de points dans les zones segmentées (cf. chapitre 4) est insérée également à la suite des points précédents.

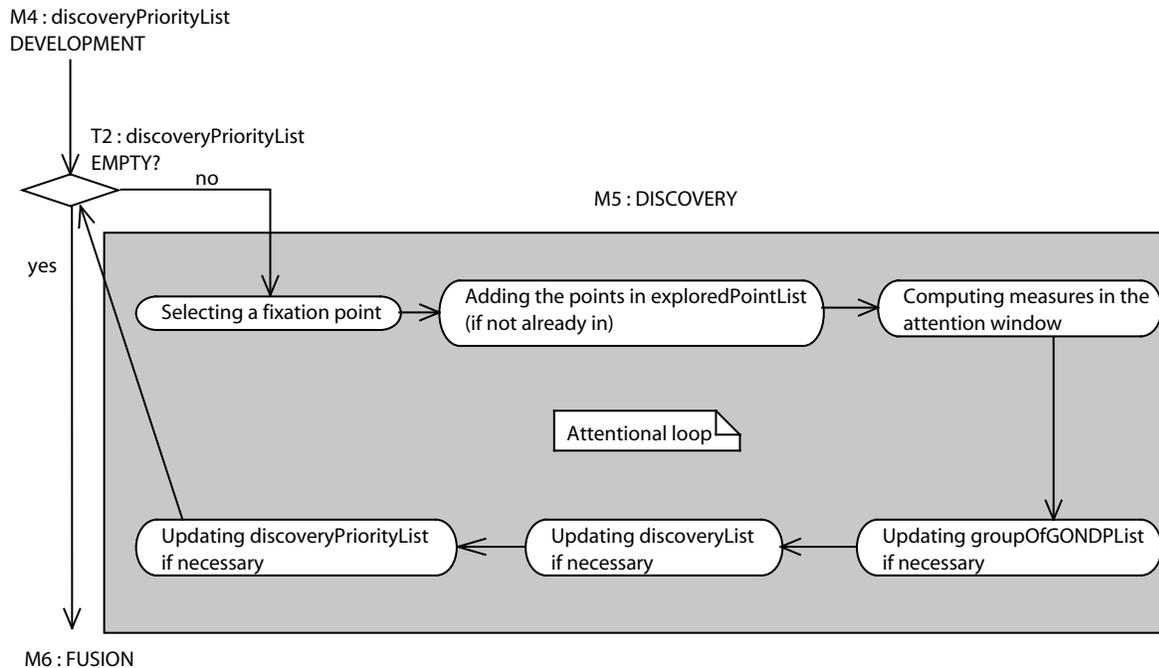


FIG. 7.8 – Module M5 : Découverte — balayage attentionnel.

7.3.1.6 Module M5 : Découverte — balayage attentionnel

La figure 7.8 décrit le balayage attentionnel réalisé pour le mode découverte. Le module M5 est en réalité une boucle sur l'ensemble des points de la liste des priorités pour le mode découverte; cette boucle est également appelée *boucle attentionnelle* de par la stratégie de déplacement du regard virtuel dans l'image. Ce module comporte des points communs avec le module M2 dans sa construction. En revanche l'approche sur l'image est très différente comme nous l'avons vu au chapitre 4.

La stratégie est de “suivre” chaque amorce d'élément quand elle est trouvée. Ce suivi se fait bien évidemment en fonction du modèle de comportement du regard.

¹⁶En réalité la liste peut être issue du mode découverte par la boucle de rétroaction, cf. figure 7.4. Cette situation ne peut intervenir au premier passage dans le module M4.

Sélection du point de fixation La sélection du point de fixation ne se fait à partir de la liste des priorités que quand le suivi attentionnel de l'élément précédent est terminé. Autrement dit, c'est le modèle de saccades qui la plupart du temps détermine ce point. Si un élément est extrait, une nouvelle saccade est calculée en fonction des informations extraites, sinon une recherche autour du point de fixation est effectuée dans le but de trouver l'élément recherché. Si dans ce dernier cas aucune information n'est trouvée, une deuxième saccade est calculée et la recherche s'effectue comme pour la précédente. Si une fois encore l'information n'est pas trouvée, alors on passe au point suivant de la liste des priorités. Un autre cas d'arrêt est évidemment l'atteinte du bord de l'image.

Mise à jour de la liste des groupes de groupes de points Au cours du chapitre 6 il a été démontré que l'on pouvait proposer un modèle¹⁷ de fusion des informations locales pour les intégrer globalement sur l'ensemble de l'image. De plus ces informations locales sont extraites aux points de fixation du modèle de saccades ; la probabilité de bifurquer d'un élément à un autre est par conséquent non négligeable. La stratégie adoptée est donc, contrairement au mode reconnaissance (module M2), de finir l'exploration de tous les points de la liste de priorités avant de faire la fusion – module M6. Autrement dit, l'exploration se fait sur tous les éléments possibles supposés. Puis ensuite, seulement ces derniers sont mis à jour dans la structure, alors que le mode reconnaissance fait la reconnaissance et la mise à jour élément par élément. Évidemment l'ordre dans lequel est parcourue l'image est important puisqu'il sert à construire le panorama – cf. §Panorama page 178. VISION mémorise le parcours de l'œil virtuel sur l'image en regroupant dans une liste tous les éléments extraits d'une même fenêtre. C'est pourquoi il existe des *groupes de groupes de points*¹⁸. La structure informatique est discutée plus loin.

¹⁷Le modèle CLE.

¹⁸Nous attirons l'attention du lecteur sur le fait que ces *groupes de points* sont les *amas de points* qui sous tendent les *dipôles* du modèle CLE.

7.3.1.7 Module M6 : Fusion

Le modèle que nous avons établi dans le chapitre 6 pour la fusion des informations locales basées sur le raisonnement spatial qualitatif jette les bases pour la conception d'une nouvelle théorie même si cette théorie doit être complétée. Il est toutefois possible à ce stade d'implémenter le point de départ de cette théorie. C'est ce qui est fait dans le module M6, figure 7.9.

Chargement des groupes (amas) de points Les points qui sont stockés dans les listes créées par le module M5 sont chargés séquentiellement dans une liste unique suivant le parcours que l'œil virtuel a effectué sur l'image. La distance entre deux amas dans la liste est *l'ordre relatif*¹⁹ d'un amas par rapport à l'autre.

Sélection de la paire d'amas dans la liste Une paire d'amas dans la liste est sélectionnée. Lors de la première entrée dans la boucle il s'agit des deux premiers amas de la liste. Si cette paire a été rejetée dans l'étape suivante, la paire est "déplacée" c'est-à-dire qu'on incrémente de un l'indice d'un des amas ou des deux.

Validation de la paire d'amas Comme le modèle CLE le précise, le "*schéma global [recherché] est une collection [d'amas] avec des relations spatiales déterminées entre eux*". La relation spatiale que l'on cherche concrètement est l'appartenance ou non à un même élément curviligne sur l'image. Une comparaison entre les attributs des amas est menée (orientation, *signature spectrale*, proximité spatiale). Si la distance entre les attributs est trop grande, la paire est rejetée. Dans le cas contraire les deux amas sont

¹⁹ *Ordre relatif* au sens du modèle CLE, c'est-à-dire une traduction du *panorama*.

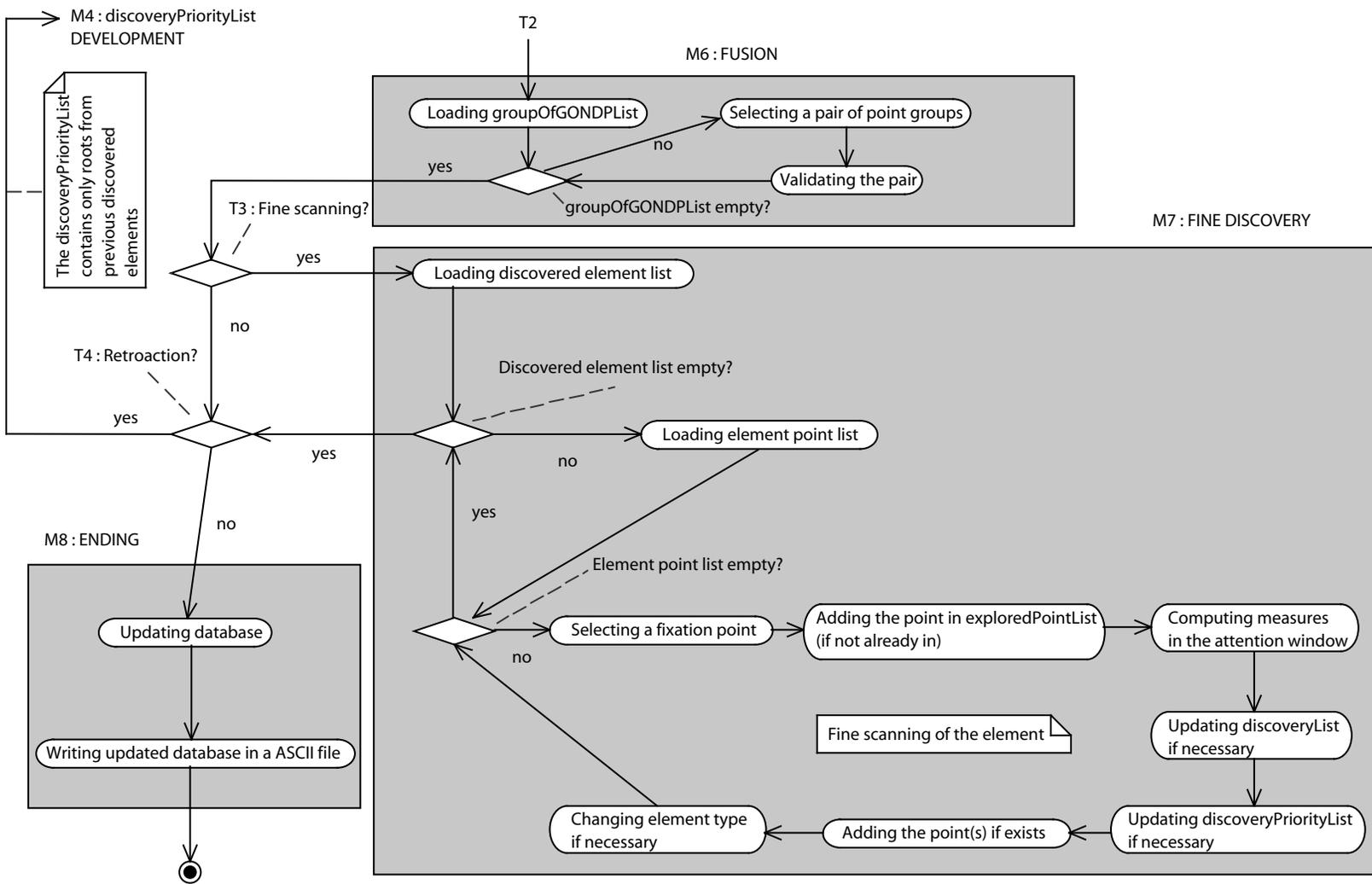


FIG. 7.9 – Modules M6 à M8.

considérés comme faisant partie du même élément curviligne dans l'image. Ces deux amas sont alors placés dans un même élément curviligne dans la structure de données.

7.3.1.8 Module M7 : Découverte fine

Le module M7 de découverte fine (figure 7.9) s'apparente au module M3. La différence principale réside dans le fait que le module M7 effectue la recherche fine pour l'ensemble des éléments découverts par les modules M5 et M6 ; le module M3 lui ne le fait que pour un élément. Une boucle supplémentaire sur l'ensemble des éléments est donc introduite. Enfin, tout comme le module M3, ce module est facultatif et influe fortement sur la performance de VISION. Les éléments suivis ont donc déjà été fusionnés auparavant, il s'agit d'avoir une meilleure connaissance des ces éléments²⁰.

7.3.1.9 Module M8 : Terminaison

Ce dernier module effectue deux traitements principaux, figure 7.9, afin de terminer le programme en produisant les sorties attendues.

Mise à jour de la base de données Les données mises à jour et découvertes au cours de l'exécution sont, à ce point du programme, résidentes en mémoire. Elles sont structurées (cf. la structure de données plus bas) mais non filtrées. Autrement dit, *tous* les points (ici les pixels reconnus) identifiés comme faisant partie d'un élément sont mémorisés. Avant l'écriture de cette base de données il convient donc de la filtrer pour

²⁰éventuellement des inconsistances peuvent être soulevées ici, introduisant alors une nouvelle fusion.

retirer les points superflus. C'est ce qui est fait à cette étape avec notre désormais très familier algorithme de Douglas-Peucker, annexe A.1.

Écriture de la base de données mise à jour La base de données est écrite dans un fichier de la même forme que le fichier d'entrée ASCII. Ce fichier est lisible par un SIG classique. Dans notre cas il s'agit du logiciel MapInfo.

7.3.2 Architecture logicielle — Diagramme de classes

Toutes les étapes de traitement développées dans le paragraphe précédent impliquent que le logiciel soit relativement important en terme de développement informatique. D'autre part, étant dans une phase exploratoire du concept, il importe que le logiciel soit modulable. Une *architecture* logicielle *orientée objet* s'impose donc pour un tel outil. Cette architecture a été développée sous le formalisme UML. La figure 7.10²¹ est une représentation de l'architecture de VISION par son diagramme de classes²².

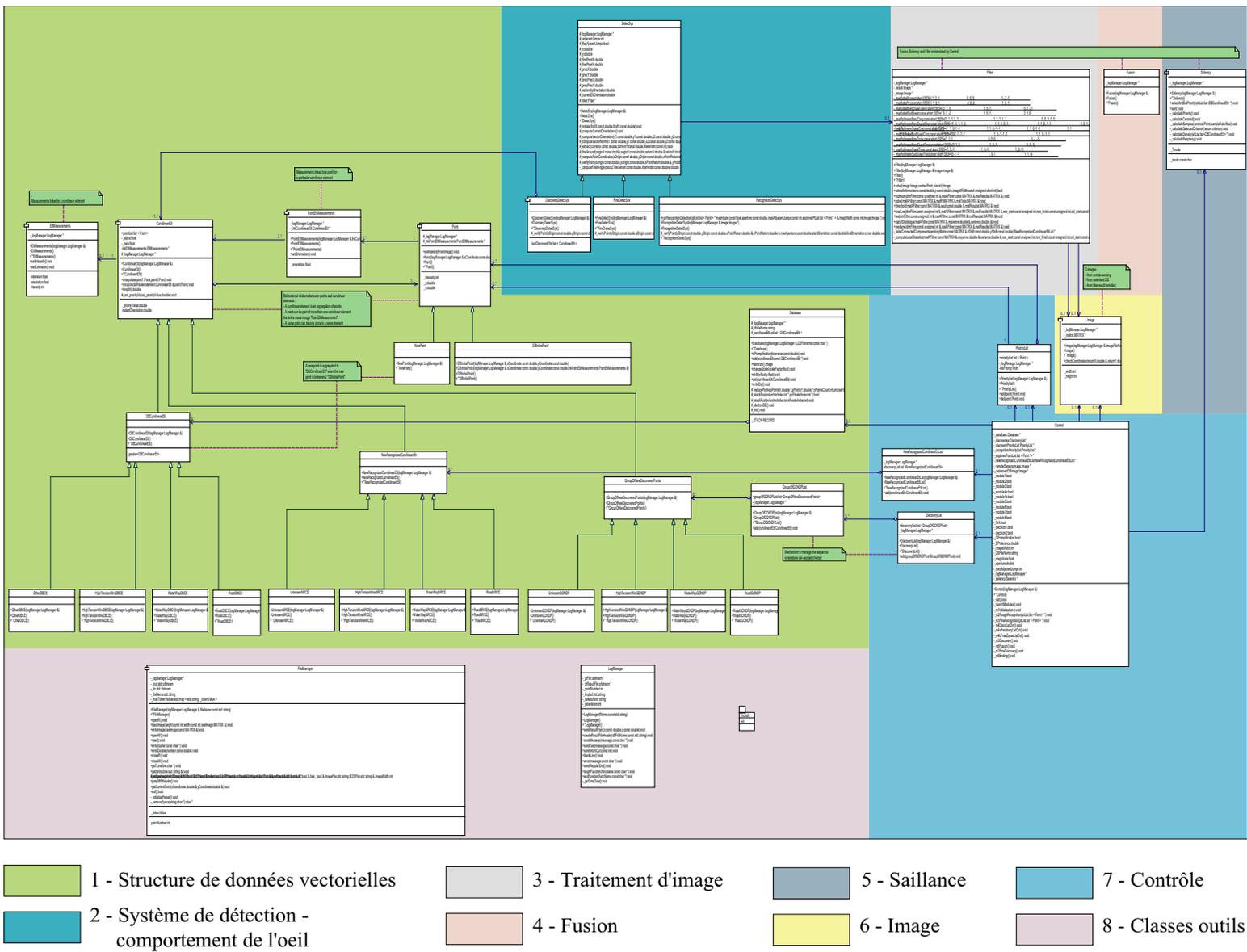
L'objectif de cette partie n'est pas de faire une description exhaustive mais de donner les points clés de l'architecture logicielle de VISION. Comme le montre la figure 7.10, le logiciel comporte une petite quarantaine de classes que l'on peut regrouper en huit pôles :

1. Structure de données vectorielles.
2. Système de détection qui supporte le comportement de l'œil virtuel.

²¹NB : la version électronique de la thèse permet au lecteur curieux de “zoomer” sur la figure.

²²La charte des couleurs utilisées dans ce diagramme est reprise d'un document du *Geological Survey of Canada* présentant un grand diagramme de classes pour ses données [BJTB99]. Les points les plus importants sont explicités plus bas dans le document selon l'approche adoptée par l'ISO et l'OGC dans leurs documents techniques [ISO05]. Le détail du diagramme est une information supplémentaire non essentielle à la compréhension de l'architecture logicielle dans son ensemble

FIG. 7.10 – Diagramme de classes, structure logique. N.B. : La version électronique est “zoomable”.



3. Traitement d'image qui permet les traitements menés sur le point de fixation.
4. Fusion qui contient principalement l'implémentation du module M6.
5. Saillance qui permet de calculer les priorités sur les points, autrement dit leur saillance.
6. Image.
7. Contrôle qui gère l'ensemble du programme, en particulier le déroulement de l'algorithme général.
8. Classes outils qui contiennent des classes et un *package* permettant de manipuler les fichiers, produire des *logs* du programme, mais aussi faire des manipulations d'algèbre linéaire – calcul matriciel.

Dans la suite de cette section nous revenons plus en détail sur les points 1, 2 et 7 car il s'agit des aspects les plus importants du point de vue de l'architecture. La description de ces points n'est pas exhaustive.

7.3.2.1 Structure de données vectorielles

La structure de données vectorielles, comme son nom l'indique, supporte tout le stockage des données vectorielles. Il s'agit de tous les points et éléments de la base de données, mais également de tous les points reconnus et découverts sur l'image au cours de l'exécution. Évidemment, l'instanciation et la manipulation de la structure sont entièrement gérées dynamiquement. Le nombre de points pouvant être très important, les classes sont conçues de la manière la plus légère possible afin de supporter un grand nombre d'instances de (mêmes) classes simultanées²³. Tout comme nous l'avons dit précédemment pour le modèle, l'idée est toujours de faire de la manière la plus simple

²³Par *simultanées*, nous entendons des instances de classes, résidentes en mémoire à un même moment t de l'exécution du programme.

possible.

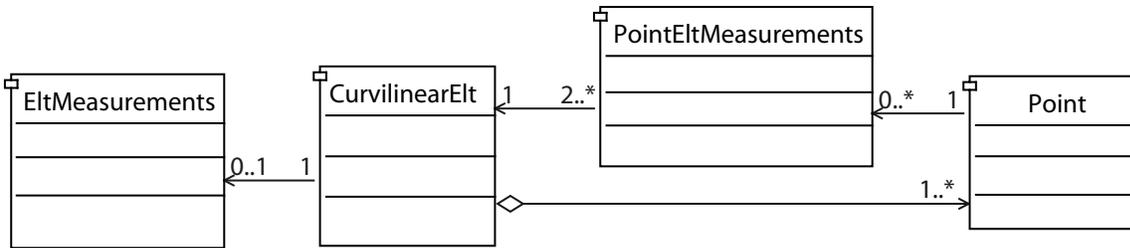


FIG. 7.11 – Relation entre points et éléments dans la structure logique de données.

Relation élément curviligne–point La figure 7.11 décrit la relation entre points et éléments curvilignes au niveau général. La double liaison entre les deux classes est la particularité la plus remarquable²⁴. Cette double liaison s’avère indispensable pour notre structure. D’une part un élément curviligne est une agrégation de points, mais d’autre part lorsqu’un point est rencontré au cours du traitement on doit être capable d’identifier à quel(s) élément(s) il appartient. Cette double liaison ce fait via la classe `PointEltMeasurements` qui permet de gérer les propriétés (mesures) d’un point particulier pour un élément particulier. Un exemple de propriété est l’orientation locale de l’élément à l’emplacement du point, *i.e.* la tangente à la courbe au point considéré. Nous avons vu que cette mesure est très importante pour le modèle.

Un point peut faire partie de plusieurs éléments curvilignes à la fois dans le cas d’intersections par exemple. En revanche un même point ne peut être présent qu’une seule fois au sein d’un même élément curviligne.

²⁴En mémoire, une telle structure en boucle est bien sûr compliquée à gérer en termes informatiques, en particulier lors de la création et de la destruction des objets. Avantageusement, le C++ permet de créer ce type de structure en “pré déclarant” dynamiquement les instances de classes. La difficulté est de gérer les *cycles d’objets* (terme consacré), surtout lors de leur destruction ; il s’agit de prendre garde à détruire tout le cycle et non les pointeurs puisque le C++ ne dispose pas de ramasse-miettes (*garbage collector*) comme le C# par exemple. Une fuite mémoire en cours d’exécution pourrait entraîner une sortie catastrophique de l’application.

Point La classe `Point` permet la gestion des points. Cette classe a pour attributs²⁵ principaux ses coordonnées x et y et l'intensité de l'image à ce point. Comme le montre la figure 7.12, la classe `Point` accepte comme filles les classes `DBInitialPoint` et `NewPoint`. Elles permettent de gérer, respectivement, les points issus de la base de données originale et les points nouvellement créés.

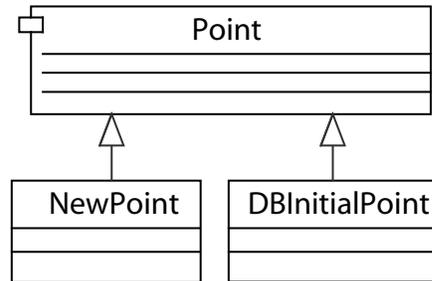


FIG. 7.12 – Héritages de la classe `Point`.

Élément curviligne Trois classes héritent de la classe `CurvilinearElt`, figure 7.13. Elles répondent aux exigences du modèle.

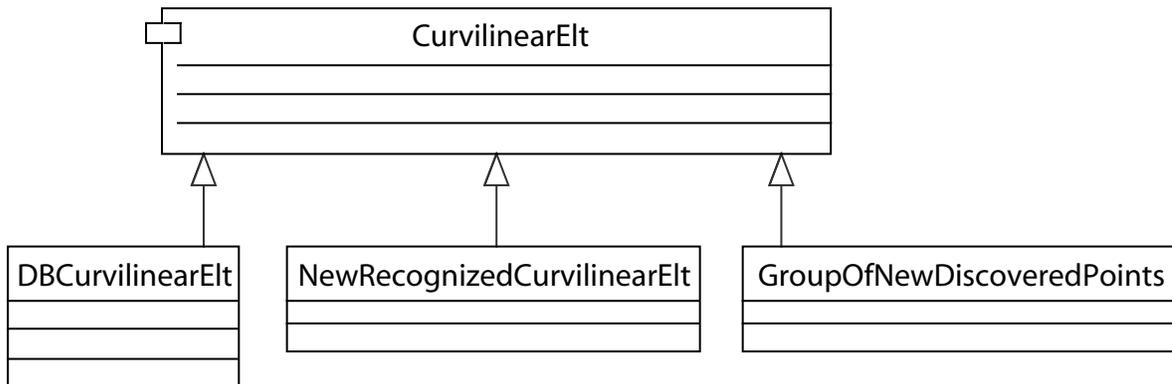


FIG. 7.13 – Héritages de la classe `CurvilinearElt`.

1. Classe `DBCurvilinearElt` : il s'agit des points qui étaient à l'origine dans la base

²⁵Ces attributs sont des *propriétés* au sens de certains outils case. C'est-à-dire que ce sont des attributs auxquels sont associés des méthodes de type *accesseur* et *modificateur*. Attention la définition de *propriété* en UML peut différer dans la littérature. Ce peut être également, pour certains auteurs, une description des propriétés intrinsèques comme par exemple les noms des éléments [EP98].

de données et qui sont regroupés au sein de cette classe. Ce type d'élément est constitué uniquement de points de la classe `DBInitialPoint`.

2. Classe `NewRecognizedCurvilinearElt` : il s'agit des éléments reconnus au cours du mode reconnaissance.
3. Classe `GroupOfNewDiscoveredPoints` : il s'agit des groupes de points qui sont extraits de l'imagette durant la phase de découverte – module M5. C'est cette classe qui supporte en réalité les dipôles au sens du modèle CLE. Comme expliqué précédemment (§7.3.1.6), ces groupes de points sont eux-mêmes regroupés, afin de pouvoir avoir accès au parcours qu'a fait l'œil sur l'image durant la fusion. Une classe spécialisée, `GroupOfGONDPList`, autorise ce mécanisme par agrégation, figure 7.14. Ces groupes de groupes sont enfin stockés dans une classe liste (`DiscoveryList`) gérée dans le pôle contrôle. Passer d'un groupe à l'autre par incrémentation de un dans cette liste est la traduction d'une saccade²⁶ qui s'est produite durant le mode découverte.

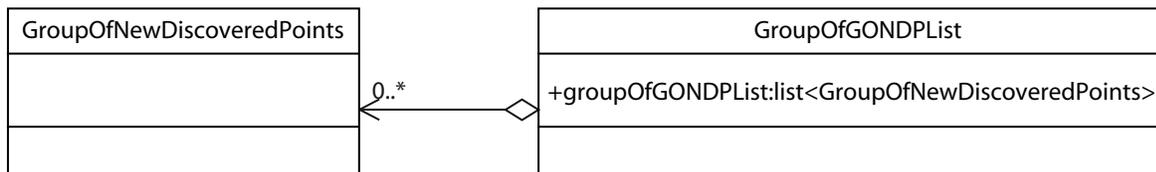


FIG. 7.14 – Mécanisme d'agrégation de la classe `GroupOfNewDiscoveredPoints` en `GroupOfGONDPList`. Ce mécanisme permet d'avoir la mémoire du parcours de l'œil sur l'image.

Base de données L'ensemble des éléments curvilignes de la base de données originale sont agrégés dans une classe `Database`, figure 7.15. Cette classe fait l'objet de l'exemple de spécifications fonctionnelles, annexe B.

²⁶Il s'agit de saccade avec succès, c'est-à-dire que les saccades à l'issue desquelles aucune information n'a été trouvée ne se retrouvent pas traduites dans cette liste ; et ceci par définition même de la liste.

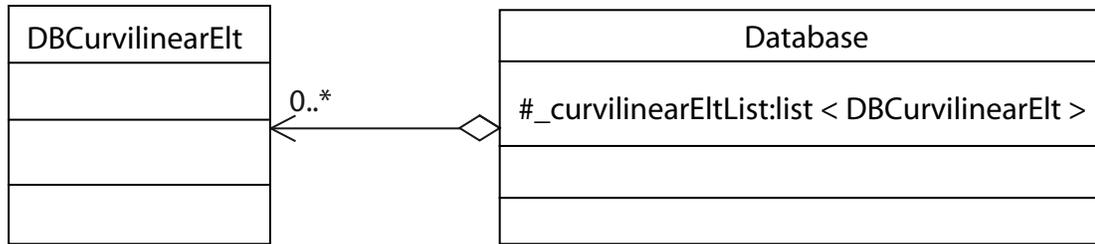


FIG. 7.15 – Classe `Database` : agrégation de `DBCurvilinearElt`.

En plus de contenir la liste des éléments curvilignes de la base de données vectorielles originales, la classe `Database` comporte des méthodes pour manipuler les données. En particulier, une méthode de cette classe est l'implémentation de l'algorithme de Douglas-Peucker. Les méthodes qui permettant l'appariement de la base avec l'image se trouvent également ici. De manière générale, les méthodes de la classe `Database` sont des méthodes appelées au début de l'exécution de VISION c'est-à-dire dans les modules d'initialisation et M1, afin de préparer les données pour le traitement ultérieur.

7.3.2.2 Contrôle — Moteur logiciel

Le système de contrôle du logiciel est dirigé principalement par une classe, la classe `Control`, figure 7.16. Cette classe implémente le déroulement du diagramme d'activité de la figure 7.4. Ceci est réalisé en positionnant correctement, les nœuds de décision T1, T2, T3 et T4. La séquence des modules est ainsi contrôlée.

L'information est passée par l'utilisateur au logiciel via le fichier XML évoqué plus haut – cf §C.1. Cette information est représentée par des attributs de la classe `Control`, comme `_module4a`, `_decision1`, *etc.* Ces attributs sont utilisés par `_launchModules()`, méthode qui implémente ni plus ni moins le diagramme d'activité de la figure 7.4. `_launchModules()` lance, en fonction du choix de l'utilisateur, les méthodes associées

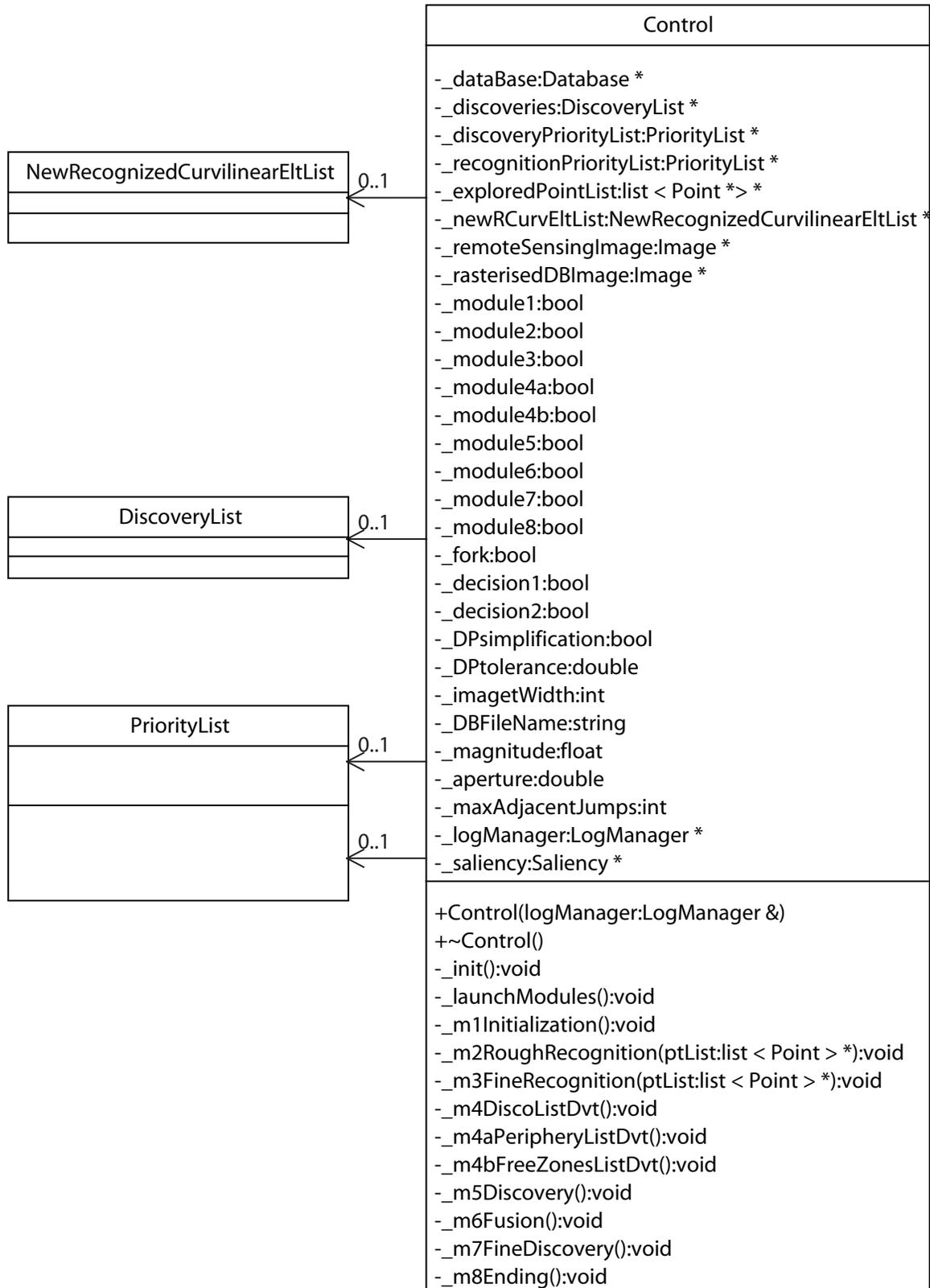


FIG. 7.16 – Classes du système de contrôle de VISION.

au modules : `_m1Initialization()`, `_m2RoughRecognition()` et ainsi de suite.

Afin d'exercer le contrôle du logiciel, la classe `Control` possède également comme attributs des pointeurs sur des instances de classes – qui font partie également du pôle de contrôle ou non. Ce sont les listes de priorités²⁷, la liste des découvertes²⁸, des éléments reconnus, la base de données originale, l'image raster²⁹ mais aussi une classe permettant de faire les calculs liés à la saillance³⁰ comme la simplification par l'algorithme de Douglas-Peucker ou le calcul des densités linéaires *etc.*

7.3.2.3 Système de détection — Comportement de l'œil

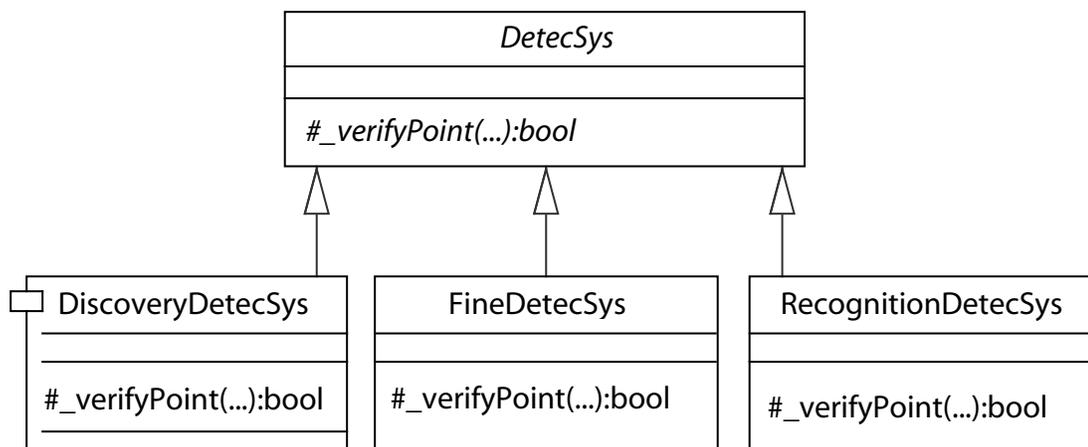


FIG. 7.17 – Classes des systèmes de détection.

Enfin le dernier groupe de classes sur lequel nous revenons est celui qui supporte le système de détection, autrement dit le comportement de l'œil virtuel. Il se compose de quatre classes (figure 7.17) : une classe abstraite `DetectSys` de laquelle héritent trois classes, soient `DiscoveryDetectSys`, `FineDetectSys` et `RecognitionDetectSys`.

²⁷Classe `PriorityList`.

²⁸Classe `Discoverylist`.

²⁹Classe `Image`.

³⁰Classe `Saliency`.

La classe *DetectSys* est une classe abstraite qui n'est jamais instanciée mais propose un nombre de méthodes communes aux trois classes filles pour les calculs sur les déplacements. Par exemple des calculs de normes, d'orientations, de coordonnées, *etc* ; mais aussi une méthode qui lance l'extraction des pixels sur l'imagette. Les trois autres classes (*DiscoveryDetectSys*, *FineDetectSys*, *RecognitionDetectSys*) implémentent principalement l'algorithme de comportement de l'œil proprement dit pour trois modes, respectivement le mode découverte, module M5, le mode fin, modules M3 et M7 et le mode reconnaissance, module M2. La méthode `_verifyPoint(...)`, donnée en exemple dans la figure 7.17, illustre le fait que l'on peut se servir de certaines méthodes de la classe *DetectSys* (depuis une des classes filles) devant vérifier si une saccade est autorisée en un point. La validité de la localisation de cette saccade diffère en fonction du mode. C'est ainsi que la bonne méthode sera toujours appelée, même si c'est depuis une méthode de *DetectSys*.

7.4 Conclusion

L'architecture du logiciel VISION, développée dans ce chapitre, montre qu'une solution logicielle peut supporter le modèle ELECA. De par l'originalité du modèle ELECA l'enchaînement des traitements informatiques dégagés ici est inédit. En outre, l'architecture de VISION, conçue pour répondre aux exigences du modèle ELECA, est également une contribution originale de cette thèse.

Chapitre 8

Réalisation logicielle et enseignements tirés

*La réalisation réside dans la pratique.
Bouddha.*

Ce chapitre est consacré à l'implantation du modèle ELECA et aux leçons l'on peut tirer de cette implantation. Dans un premier temps, nous nous attardons sur les aspects algorithmiques de VISION. Des algorithmes originaux sont développés pour supporter, d'une part, le guidage du regard virtuel dans l'image, et, d'autre part, l'extraction d'informations comme présentée au chapitre 5. Nous montrons qu'il est possible d'implanter une solution qui mêle dès le départ les données vectorielles et matricielles. Cette implantation se fait dans l'architecture élaborée au chapitre précédent. Dans un second temps, un bilan est dressé pour les aspects logiciels de ce travail, en en dégageant les points forts et les points faibles puis en comparant VISION avec les logiciels commerciaux actuels.

8.1 Implantation

La première partie de ce chapitre se focalise sur certaines spécificités d'implantation et d'implémentation propres au modèle ELECA, en particulier les choix qui ont été faits en termes algorithmiques. Les parties du modèle ELECA directement implantables sont celles développées dans les chapitres 4 et 5. Une première section est consacrée aux aspects algorithmiques du guidage du regard sur l'image, puis la seconde section est consacrée, à l'extraction des informations locales.

8.1.1 Couverture de l'image par le comportement saccadique

Algorithmiquement parlant, les modes reconnaissance et découverte sont très similaires en ce qui concerne le comportement du regard sur l'image. En pratique, le mode découverte est un sous ensemble du mode reconnaissance.

8.1.1.1 Mode de reconnaissance

Déplacement le long d'un élément L'algorithme 8.1 présente une solution implantée pour le modèle de comportement saccadique développé tout au long du chapitre 4 pour le mode découverte. Du point de vue de la conception UML, cela couvre une partie du module M2. Pour cet algorithme et ceux qui suivent, les termes utilisés sont explicités dans le tableau 8.1 page 225.

Le but du mode reconnaissance est de valider, par rapport à l'image, les éléments

début

```

pour tous les points de l'élément faire
  calcul(vecteur); /* distance entre point origine et point suivant */
  calcul(magnitude); /* magnitude de saccade maximale */
  si vecteur ≤ magnitude
    /* si présent dans le grand cercle */
    alors
      extraction ← extract(pointi);
      si ¬extraction alors extraction ← rechercheAutour; fsi
    sinon
      /* vecteur > magnitude */
      tant que vecteur > magnitude
        /* tant qu'on est à l'extérieur du grand cercle on se rapproche */
        faire
          calcul(orientation); /* entre point précédent et point suivant */
          calcul(centrezone); /* coordonnées & vecteur */
          extraction ← extract(centrezone);
          si ¬extraction alors extraction ← rechercheAutour; fsi
          si extraction
            alors
              | incrémenter les indices; /* point courant devient point précédent */
            sinon
              | incrémenter avec centrezone
            fsi
          calcul(vecteur); /* avec nouveau point et point suivant */
        ftq
        /* vecteur ≤ magnitude : le point est atteint */
        extraction ← extract(pointi);
        si ¬extraction alors extraction ← rechercheAutour; fsi
      fsi
    si ¬extraction
      alors
        | méthodeDécouverte; /* Algorithme de type découverte, fig. 8.3 */
      fsi
      māj(liste des découvertes, des priorités...);
      confirmation/ajout/retrait du point de la base;
      incrémenter les indices; /* point courant devient point précédent */
      extraction ← faux;
    fpour
  fin

```

ALG. 8.1 – Algorithme simplifié de comportement de l'œil pour le mode reconnaissance.

présents dans la base de données. Pour cela le regard virtuel suit chaque élément selon un comportement saccadique. Si l'élément n'est pas trouvé à la première *fixation*, la recherche est effectuée autour, à l'intérieur des corridors de recherche.

La première étape consiste à déterminer si le saut du point courant au point suivant n'est pas trop grand. Autrement dit est-ce qu'il ne dépasse pas la *magnitude* maximale autorisée par le comportement ? La figure 8.1 illustre ceci. Les points 1, 2 et 3 sont les points de la base de données. Dans un premier temps, l'œil est en 1 – fig. 8.1a. La distance entre le point courant et le suivant est notée *vecteur*, la magnitude maximale *magnitude*.

<code>calcul()</code>	Procédure effectuant le calcul du terme entre parenthèses
<i>vecteur</i>	Distance entre point origine et point suivant, figure 8.1
<i>magnitude</i>	Magnitude de saccade maximale figure 8.1
<i>extraction</i>	Booléen vrai si des pixels ont été extraits au point de fixation
←	Traduit l'affectation de l'expression de droite à celle de gauche
<code>extract()</code>	Procédure d'extraction locale sur le point entre parenthèses et retournant vrai dans le cas où des pixels ont été extraits
<i>point_i</i>	Point courant considéré
¬	Non logique : ¬vrai = faux ; ¬faux = vrai
<code>rechercheAutour</code>	Procédure permettant de chercher autour d'un point si rien n'a été trouvé au point même. Il s'agit de l'algorithme 8.2
<i>orientation</i>	Orientation entre le point précédent et le point courant
<i>centre_{zone}</i>	Centre de la zone cible pour la saccade
<code>méthodeDécouverte</code>	Algorithme suivant une stratégie de type découverte. Il s'agit de l'algorithme 8.3
<code>màj()</code>	Mise à jour des termes entre parenthèses
<i>centre_{zoneDipôle}</i>	Centre de la zone cible pour la saccade, calculée en utilisant l'orientation du dipôle
<i>centre_{zoneExtrémitéÉlt}</i>	Centre de la zone cible pour la saccade, calculée en utilisant

TAB. 8.1 – Termes utilisés dans les algorithmes (fin du tableau page 225)

	l'orientation de l'extrémité de l'élément
$centre_{zoneÉlément}$	Centre de la zone cible pour la saccade, calculée en utilisant l'orientation du dipôle
γ	Pas angulaire d'incrément de l'angle, figure 8.2
$décalageAngle$	Décalage d'angle par rapport à l'orientation pour la prochaine saccade
$intérieur_+$	Booléen vrai si les sauts précédents du côté positif étaient autorisés, figure 8.2
$intérieur_-$	Booléen vrai si les sauts précédents du côté négatif étaient autorisés, figure 8.2
$verifyPoint()$	Procédure permettant de vérifier que le point entre parenthèses est autorisé pour une saccade ; c'est à dire qu'il n'ai pas en dehors de l'angle maximal autorisé, de l'image...
$point_{décalageAngle+}$	Centre de la zone cible pour la saccade, calculée avec un saut suivant la direction : $orientation + décalageAngle$
$point_{décalageAngle-}$	Centre de la zone cible pour la saccade, calculée avec un saut suivant la direction : $orientation - décalageAngle$
MC	Matrice de correspondance des étiquettes de l'imagerie
MC^*	Matrice de correspondance MC sur laquelle l'algorithme de Roy-Warshall (algorithme A.2) a été appliqué
p_e	Pixel étiqueté courant
i	Compteur d'étiquette pour parcours de la matrice MC
$sortie$	Drapeau de sortie de boucle
$surnombre$	Compteur d'étiquettes surnuméraires
$étiquette()$	Procédure retournant l'étiquette de l'argument
$nbreÉlément$	Nombre d'éléments différents identifiés dans l'imagerie
$ajout(\#1, \#2)$	Procédure qui permet l'ajout du point $\#2$ à l'élément $\#1$
$éléments_i$	Élément correspondant à l'étiquette i

TAB. 8.1 – Termes utilisés dans les algorithmes du chapitre 8

La figure 8.1a illustre le cas où $vecteur \leqslant magnitude$, c'est-à-dire que le point 2 (point suivant au point courant dans la base de données) est situé à l'intérieur du grand cercle¹ décrit par le rayon $magnitude$. Dans ce cas le saut est fait directement sur le

¹On parle ici de *grand* cercle par opposition au *petit* cercle qui traduit la distance minimale pour

point suivant, ici 2. À ce point l'extraction est menée (`extract()`) sur 2; si cela échoue on recherche autour, algorithme `rechercheAutour`, figure 8.2 plus bas.

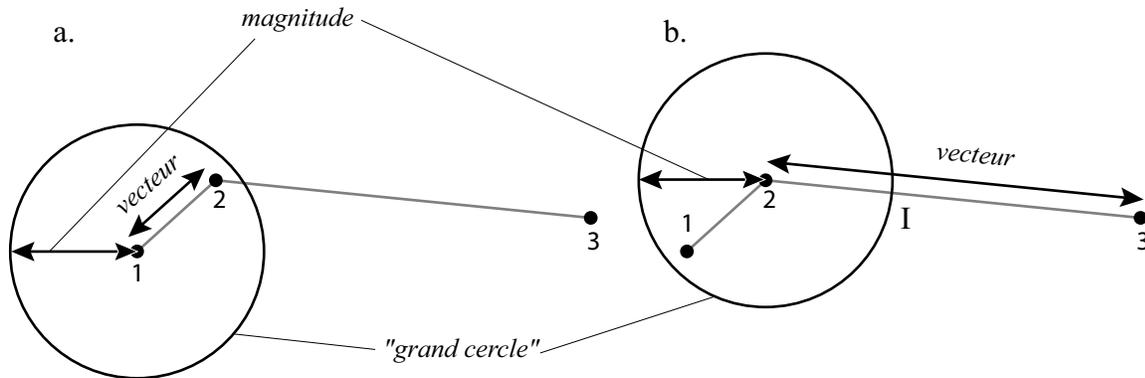


FIG. 8.1 – Magnitude, vecteur et “grand cercle” pour le mode reconnaissance.

Dans le cas contraire (figure 8.1b), *vecteur* est strictement supérieur à *magnitude*. Donc le point suivant (ici 3) est en dehors du grand cercle. Dans ce cas, le regard va se rapprocher du point suivant par des *saccades* de taille *magnitude*, ceci est décrit par la boucle “tant que”. Concrètement le saut suivant sera en I, intersection du segment [2,3] et du grand cercle de centre 2 et de rayon *magnitude*. Encore une fois si rien n’est trouvé à ce point, une recherche est menée autour du point².

Si aucun amas n’a été identifié (extraît) dans aucun des cas vus précédemment, le regard va adopter un comportement de type découverte décrit plus bas par l’algorithme 8.3, *méthodeDécouverte*.

Enfin, dépendamment du résultat, les différentes listes sont mises à jour, ainsi que la base de données. Pour terminer, le point extrait (ou bien l’emplacement du point suivant si rien n’a été trouvé) est considéré comme le nouveau point courant. Le point

une saccade, voir figure 4.16, page 135.

²Par la suite à chaque fois qu’un point de fixation est identifié et qu’à ce point aucun amas de pixels n’est extrait, systématiquement on recherche autour du point. Pour alléger le texte ceci ne sera pas reprécisé chaque fois. L’algorithme de recherche autour d’un point, *rechercheAutour*, fait l’objet de la figure 8.2.

courant devient le point précédent. Le point précédent devient l'antépénultième, *etc.* Cet aspect est important car ces points interviennent dans les calculs de l'algorithme méthodeDécouverte.

Recherche d'informations autour d'un point La recherche d'informations autour d'un point lorsque celui-ci est dépourvu d'informations locales (aucun amas de pixels extrait) est décrite dans l'algorithme 8.2. Le but est de trouver de l'information qui n'est pas exactement à la place attendue. Il s'agit de mener la recherche à l'intérieur de la zone cible définie dans le chapitre 4. L'algorithme va permettre de rechercher de part et d'autre de la direction initiale (*orientation* figure 8.2) en faisant osciller l'angle du saut autour de cette direction. Le pas angulaire de saut est appelé γ . Ainsi, tant que la recherche s'effectue à l'intérieur de la zone autorisée (zone B, figure 4.16), le regard sonde dans les directions $orientation + \gamma$, $orientation - \gamma$, $orientation + 2\gamma$, $orientation - 2\gamma$, *etc.* Tant qu'aucune information n'est trouvée et que la recherche est autorisée au moins dans l'un des cotés *intérieur₊* ou *intérieur₋*, le sondage se poursuit comme la séquence $\{1,2,3,4,5\}$ montrée en exemple à la figure 8.2.

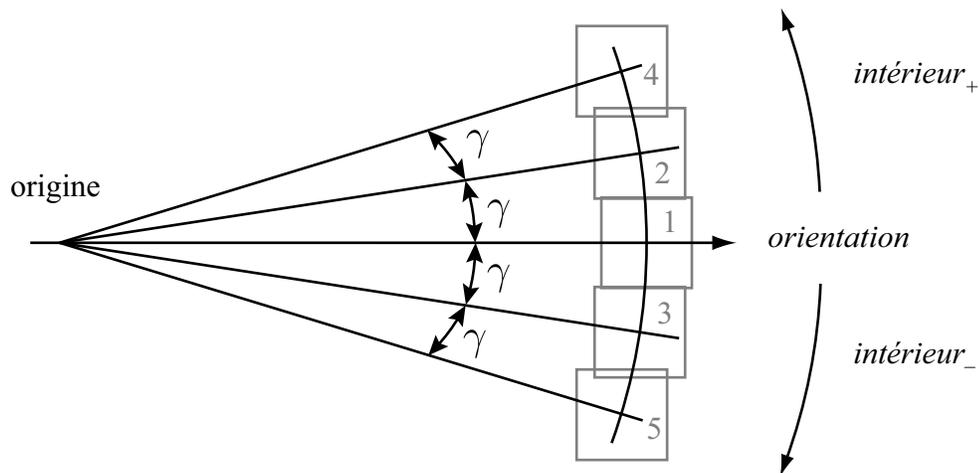


FIG. 8.2 – Technique de recherche autour d'un point.

```

début
calcul( $\gamma$ );
décalageAngle  $\leftarrow \gamma$ ; /* décalage d'angle */
extraction  $\leftarrow$  faux; /* Initialisation */
intérieur+  $\leftarrow$  vrai; /* Initialisation des valeurs de vérification */
intérieur-  $\leftarrow$  vrai;
tant que  $\left( (i\grave{e}n\grave{e}r\grave{i}e\grave{u}r_{+} \text{ ou } i\grave{e}n\grave{e}r\grave{i}e\grave{u}r_{-}) \text{ et } (\neg extraction) \right)$ 
  /* tant qu'encore à l'intérieur au moins d'un côté et que rien trouvé */
faire
  si intérieur+ /* premier côté */
  alors
    si  $\left( i\grave{e}n\grave{e}r\grave{i}e\grave{u}r_{+} \leftarrow \text{verifyPoint}(point_{d\acute{e}calageAngle+}) \right)$ 
    alors
      extraction  $\leftarrow$  extract(pointdécalageAngle+);
    fsi
  fsi
  si  $(i\grave{e}n\grave{e}r\grave{i}e\grave{u}r_{-} \text{ et } \neg extraction)$  /* second côté */
  alors
    si  $\left( i\grave{e}n\grave{e}r\grave{i}e\grave{u}r_{-} \leftarrow \text{verifyPoint}(point_{d\acute{e}calageAngle-}) \right)$ 
    alors
      extraction  $\leftarrow$  extract(pointdécalageAngle-);
    fsi
  fsi
  décalageAngle  $\leftarrow$  décalageAngle +  $\gamma$ ;
ftq
valeurRetour  $\leftarrow$  extraction;
fin

```

ALG. 8.2 – Algorithme rechercheAutour simplifié.

8.1.1.2 Mode de découverte

```

début
  /* calcul avec l'orientation du dipôle s'il existe */
  calcul(centrezoneDipôle);
  extraction ← extract(centrezoneDipôle);
  si ¬extraction alors extraction ← rechercheAutour; fsi
  si ¬extraction
  alors
    /* calcul avec l'orientation de l'extrémité de l'élément s'il existe */
    calcul(centrezoneExtrémitéÉlt);
    extraction ← extract(centrezoneExtrémitéÉlt);
    si ¬extraction alors extraction ← rechercheAutour; fsi
    si ¬extraction
    alors
      /* calcul avec l'orientation générale de l'élément s'il existe */
      /* et si différente de l'extrémité */
      calcul(centrezoneÉlément);
      extraction ← extract(centrezoneÉlément);
      si ¬extraction alors extraction ← rechercheAutour; fsi
    fsi
  fsi
fin

```

ALG. 8.3 – Algorithme `méthodeDécouverte` simplifié. Dans le cas de l'inclusion de cet algorithme dans l'algorithme 8.1 (mode reconnaissance), aucun élément n'a été identifié sur la zone déterminée à partir de la base de données. Dans le cas du mode découverte "pur" le point origine est soit le premier point de la liste des priorités (il est alors dépilé), soit un point issu de ce même algorithme s'il en a été identifié un.

Comme nous l'avons vu dans les chapitres précédents, le mode découverte, par rapport au mode reconnaissance, ne peut se reposer sur des éléments de la base de données. L'algorithme 8.3 décrit le comportement général de l'œil sur l'image pour ce mode³.

Dans un premier temps le centre de la zone cible est calculé avec les données du

³L'algorithme 8.3 est également utilisé dans le mode reconnaissance lorsqu'un objet n'a pas été retrouvé à un point. La différence réside dans le fait qu'au point suivant, dans le cas du mode reconnaissance, l'algorithme va se rattacher à la base de données.

dipôle sur lequel l'œil se trouve. Comme toujours, si rien n'est trouvé au centre de la zone, la recherche est menée de part et d'autre, algorithme 8.1⁴. Dans un second temps, si la recherche précédente a été infructueuse, la zone cible est calculée en fonction de l'orientation de l'extrémité de l'élément poursuivi. Un exemple est donné un peu plus bas. Enfin si jusque là toute recherche a été négative, l'algorithme va guider l'œil pour qu'il cherche dans l'axe global de l'élément.

8.1.1.3 Exemple de déroulement de l'algorithme le long d'un élément fictif de la base de données

La figure 8.3⁵ donne un exemple du déplacement du regard le long d'un élément fictif de la base de données. Afin que cela soit didactique, il n'y a pas d'image, autrement dit, l'extraction est toujours négative, ainsi cela permet de passer par la plupart des branches de l'algorithme.

Les points de la base de données sont les points A, B, C et F. Autrement dit l'élément est représenté en rouge. Le parcours de l'œil est figuré en bleu, chaque point de fixation (où est menée une extraction) est un petit carré bleu.

La liste de priorités contient donc en haut de la pile {A, B, C, F, ...}. Le point de départ est en A. Le regard saute sur le point B. Comme rien n'est trouvé en B, la recherche est menée autour – algorithme 8.2. C'est ainsi que le regard se fixe successi-

⁴Tous les calculs pour tous les algorithmes ne sont effectués que s'ils ont raison d'être. Par exemple ici, la branche de l'algorithme qui permet le calcul de la zone cible avec les données du dipôle, n'est pas prise si le dipôle n'existe pas au point origine. Dans le même ordre d'idée, tout comme l'inhibition de retour du regard naturel (§2.3.2.2), une zone déjà explorée n'est pas ré explorée.

⁵Il s'agit d'une sortie intermédiaire de VISION, au format MapInfo. Le logiciel permet d'obtenir, si l'utilisateur le souhaite, des résultats intermédiaires comme celui présenté ici. Ces résultats sont utilisés à des fins de recherche et de débogage.

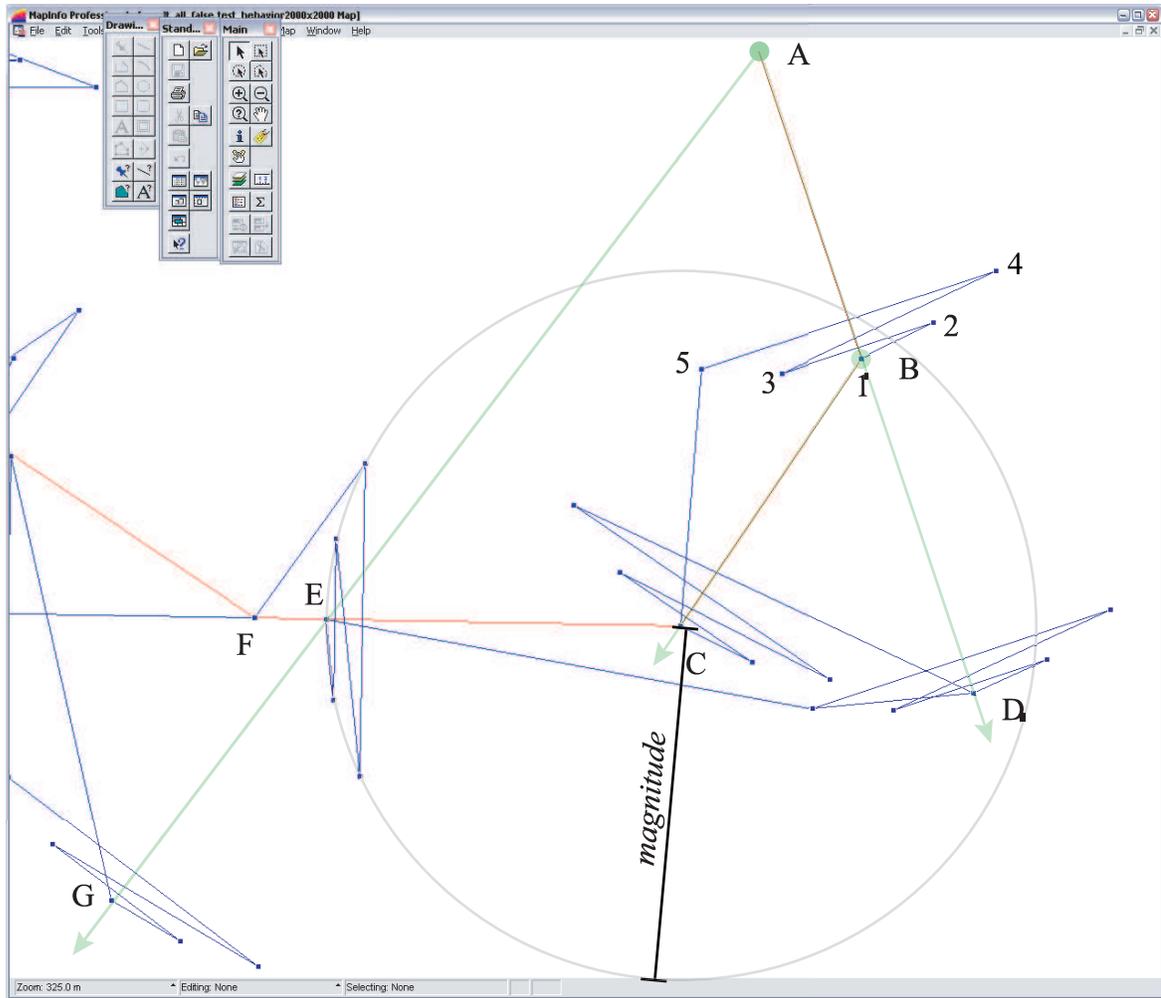


FIG. 8.3 – Comportement de l'œil le long d'un élément de la base de données.

vement aux points 1, 2, 3, 4 et 5. La recherche étant toujours négative, le regard passe au point suivant de la base de données, le point C, en prenant par défaut comme point d'origine le point B.

La recherche en C et autour s'avère négative, on cherche alors dans l'axe de l'extrémité de l'élément, soit en D. C'est pourquoi les points A, B et D sont alignés. La recherche dans l'axe de l'extrémité de l'élément n'avait pas lieu d'être au point précédent puisque l'élément ne comportait qu'un seul point – le point A. De nouveau rien n'est trouvé, on passe au point suivant de la base de données le point F en prenant par défaut

le point C comme point d'origine.

Il s'avère qu'il est impossible de passer directement du point C au point F puisque la distance CF est plus grande que la magnitude maximale autorisée pour une saccade. C'est-à-dire que F est à l'extérieur du cercle de rayon *magnitude* et de centre C. Le regard saute donc au point situé à l'intersection du segment [C, F] et du cercle soit le point E. La recherche étant infructueuse, le regard passe au point suivant de la base de données.

La distance [E, F] est maintenant inférieure à la magnitude, le saut est autorisé en F. Il est intéressant de noter ici que la recherche autour de F n'a pas été menée. En effet la distance EF étant très courte, l'angle pour la recherche aurait été très grand et en dehors de l'angle maximal autorisé. Comme rien n'est extrait en F, la recherche est alors menée, tout comme précédemment dans l'axe de l'extrémité – non montré sur la figure.

Enfin comme rien n'est trouvé, on recherche dans l'axe général de l'élément⁶, soit en F. De nouveau c'est la raison pour laquelle les points A, E et G sont alignés.

8.1.2 Extraction de l'information locale

Conformément au modèle établi dans le chapitre 5, l'algorithme pour l'extraction des informations locales (algorithme 8.4) se déroule en trois étapes. Ces trois traitements sont menés *sur* l'imagette. Autrement dit, le **filtrage** du bruit par le filtre de Lee et le rehaussement des **arêtes** par le filtre de Robinson sont appliqués *localement*, évitant

⁶Pour simplifier les calculs, l'axe général de l'élément est la direction entre le premier et le dernier point de l'élément.

ainsi un (double) filtrage de l'image dans son entièreté⁷. Dans cette section nous nous attardons sur la troisième étape, qui est une analyse en composantes connexes modifiée pour les besoins de notre problème.

```

début
| Filtrage du bruit ;
| Rehaussement des arêtes ;
| Analyse en composantes connexes modifiée ;
fin

```

ALG. 8.4 – Algorithme général pour un point de fixation.

8.1.2.1 Analyse en composantes connexes classique

Algorithme général Nous avons vu de manière théorique l'algorithme général d'analyse en composantes connexes, algorithme 5.1 page 154. Nous nous penchons un peu plus en détail sur cet algorithme pour en expliquer les modifications qui y ont été apportées.

```

début
| Étiquetage des pixels candidats d'une image binaire ;
| calcul( $MC^*$ ) ;
| Affectation d'une même étiquette aux étiquettes équivalentes ;
fin

```

ALG. 8.5 – Algorithme classique d'analyse par composantes connexes.

On peut écrire l'algorithme 5.1 illustré selon l'algorithme 8.5⁸. Cet algorithme prend en entrée une image binaire *seuillée au préalable* et étiquette les pixels blancs en

⁷Le fait d'appliquer deux filtres consécutifs sur la même imagerie réduit deux fois sa taille sur les bords comme il a été dit dans le chapitre 5. Il faut bien sûr alors en tenir compte dans l'algorithme 8.2 pour qu'il n'y ait pas de défaut de recouvrement ("blancs") entre les imageries.

⁸Les termes des algorithmes qui suivent sont toujours explicités dans le tableau 8.1.

construisant une matrice de correspondance MC (voir plus loin) des étiquettes équivalentes. Ensuite il applique l'algorithme de Roy-Warshall⁹ de fermeture transitive de la matrice d'un graphe sur MC pour obtenir la matrice MC^* . Enfin il affecte la même étiquette aux pixels équivalents.

Matrice de correspondance et sa fermeture transitive Prenons la figure 8.4 comme exemple. Il s'agit d'une imagerie 10×10 pixels résultant de la première itération de l'ACC soit l'étiquetage des pixels candidats. Les lettres a, b, c, d , correspondent aux étiquettes des différents pixels. Parallèlement à la lecture des pixels et à leur étiquetage on construit la matrice de correspondance MC , équation 8.1.

a					b				c
a				b	b		d		c
	a	a	b				d		c
				b	b	d			c
									c
								c	
							c		
							c		
							c		
						c	c		

FIG. 8.4 – Étiquetage des pixels après le premier passage.

⁹L'algorithme de Roy-Warshall fait l'objet de l'annexe A.2.

$$MC = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} a \quad b \quad c \quad d \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array} \right] \end{array} \quad (8.1)$$

La matrice MC permet de mettre en correspondance les étiquettes équivalentes de l'image. Les lignes et les colonnes de la matrice sont par conséquent les étiquettes de l'image. Par définition une étiquette correspond à elle-même c'est pourquoi la matrice contient systématiquement des 1 sur sa diagonale. La construction de la matrice se fait au fur et à mesure que se fait l'étiquetage. Ainsi lorsqu'une connexion entre deux groupes de pixels étiquetés est identifiée¹⁰, un 1 est ajouté entre les 2 étiquettes. Ainsi, dans notre exemple la matrice MC rend compte du fait que a touche b , que b touche d et que c ne touche pas d'autres étiquettes.

$$MC^* = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} a \quad b \quad c \quad d \\ \left[\begin{array}{cccc} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array} \right] \end{array} \quad (8.2)$$

Maintenant il s'agit de modifier la matrice pour qu'elle traduise le fait que a , b et d sont en réalité un et un seul élément. On souhaite donc obtenir la matrice MC^* ,

¹⁰Voir page 154.

équation 8.2. Le calcul est fait en appliquant un algorithme de la théorie des graphes, l'algorithme de Roy-Warshall¹¹. L'annexe A.2, explicite ce calcul en détail. Il est important de remarquer pour la suite que cette matrice est symétrique. C'est à partir de cette matrice MC^* que l'affectation des étiquettes équivalentes est faite dans l'analyse en composantes connexes classique.

8.1.2.2 Analyse en composantes connexes modifiée

L'algorithme 8.6, présente notre version modifiée de l'algorithme 8.5. Cet algorithme a été modifié pour deux raisons principales : d'abord pour que le traitement soit le plus rapide possible, afin de répondre à l'exigence de rapidité de ce travail de thèse ; ensuite pour pouvoir instancier les éléments issus des pixels. L'analyse en composantes connexes classique travaille de raster à raster ; notre *analyse en composantes connexes modifiée part d'une image raster pour fournir des données vectorielles*.

Dans un premier temps, contrairement à l'algorithme classique, et pour économiser un passage supplémentaire sur l'imagette, c'est une imagette en niveaux de gris qui est considérée en entrée de l'algorithme. De cette manière le seuillage, l'étiquetage et la création de la matrice de correspondance MC se font successivement et en une fois pixel par pixel – et non en 2 fois.

Dans un second temps, l'algorithme de Roy-Warshall (algorithme A.2) est appliqué classiquement sur MC pour obtenir sa fermeture transitive, MC^* .

Enfin, lors du deuxième et dernier passage sur l'imagette (toujours par souci de

¹¹L'algorithme de Roy-Warshall réalise la fermeture transitive du graphe décrit par la matrice MC .

```

début
  pour tous les pixels de l'image filtrée et rehaussée faire
  | Seuillage du pixel ;
  | Étiquetage du pixel si candidat ;
  fpour
  calcul( $MC^*$ );
   $nbreÉléments \leftarrow 0$ ; /* initialisation */
  pour tous les pixels étiquetés  $p_e$  faire
  |  $i \leftarrow 0$ ; /* initialisation */
  |  $sortie \leftarrow \text{faux}$ ; /* initialisation */
  |  $surnombre \leftarrow 0$ ; /* initialisation */
  | faire
  | |  $i \leftarrow i + 1$ ;
  | | si  $MC^*[étiquette(p_e), i] = 1$ 
  | | alors
  | | |  $sortie \leftarrow \text{vrai}$ ;
  | | fsi
  | jqa  $sortie$ 
  | si  $étiquette(p_e) \neq i$ 
  | alors
  | |  $surnombre \leftarrow surnombre + 1$ ;
  | | /* pour éviter la création intempestive d'éléments */
  | fsi
  | si  $i > (nbreÉléments + surnombre)$ 
  | alors
  | | Création d'un nouvel élément avec le point courant ;
  | |  $nbreÉléments \leftarrow nbreÉléments + 1$ ;
  | sinon
  | | ajout( $élément_i, p_e$ );
  | fsi
  fpour
fin

```

ALG. 8.6 – Algorithme d'analyse par composantes connexes modifié.

rapidité) les éléments (les amas de pixels) sont instanciés à la volée. L'algorithme utilise avantageusement le fait que la matrice MC^* soit diagonale. En effet la condition d'arrêt de la boucle "faire" est atteinte automatiquement, dans le pire des cas, à la diagonale de la matrice. Dans le cas le moins favorable, c'est-à-dire quand toutes les étiquettes correspondent à un amas de pixels différent, la moitié de la matrice est parcourue. Ainsi un même amas sera étiqueté avec l'étiquette la plus petite de celles qui lui avaient été affectées. En reprenant l'exemple de la figure 8.4, cela veut dire que l'amas étiqueté a , b et c recevra comme étiquette finale a .

Pour chaque nouvelle étiquette finale (dans notre exemple a et c) rencontrée un nouvel élément est créé à la volée avec le point. L'instanciation se fait dynamiquement. Si en revanche l'étiquette n'est pas nouvelle, alors le point est ajouté à l'élément correspondant également à la volée.

8.2 Enseignements et bilan

8.2.1 Considérations générales

La conception et la réalisation d'un système tel que VISION ne sont pas, comme on peut se l'imaginer, choses aisées. La difficulté, outre la mise au point du modèle ELECA, vient du fait que le logiciel en lui-même est très gros et rassemble de nombreuses composantes et données, même en simplifiant plusieurs éléments en s'appuyant sur la littérature scientifique. Ainsi, la manière inédite de formuler le problème a engendré une architecture de plus grande envergure que prévu.

En effet, *VISION* gère de manière imbriquée des données vectorielles et des données raster. La navigation de l'œil dans l'image (et en fonction de l'image) est sous-tendue complètement par les données vectorielles. Ainsi, le grand nombre de données impose une très grande rigueur dans la conception de la structure de données ; d'autant plus que tout le processus est dynamique. Pour ne pas dupliquer les données, ce sont des liens (pointeurs en C++) qui sont utilisés pour maintenir la connaissance (sur les données) à un moment donné t de l'exécution. Il en résulte, au cours de l'exécution de *VISION*, une véritable "toile d'araignée" virtuelle reliant les données. Cette manière de procéder permet, d'une part, une grande souplesse pour l'évolution du système et, d'autre part, une grande rapidité d'exécution. La contre partie, puisque le langage C++ est un langage objet de bas niveau, est que le temps de débogage est important dès qu'il s'agit de travailler avec les adresses mémoire que sont les pointeurs.

8.2.2 Solution adaptée au problème

Notre approche a été de construire totalement le système, sans le programmer dans une extension d'un logiciel déjà existant pour trois raisons. La raison première a été la grande imbrication prévue des différentes parties du logiciel, aussi bien du point de vue des données, comme nous l'avons dit, que de la relation avec le moteur logiciel. La seconde raison a été que *VISION* devait s'inscrire dans un projet de grande envergure (GEOIDE ENV#19 et ENV#ELD). Une solution "clé en main" était alors souhaitable car il s'agit de l'amont de la chaîne de recherche¹². Enfin, le dernier argument pour la conception de *VISION* dans son entièreté a été un souci d'efficacité du code. Ainsi, une structure de données ne contenant que les attributs et les méthodes qui servent au

¹²Le projet consiste à utiliser la réalité augmentée pour aider le pilotage d'hélicoptères de sauvetage. La création de la base de données qui est chargée à bord de l'hélicoptère est par conséquent réalisée pendant la première phase de déclenchement de la mission. C'est pourquoi cette recherche se situe en amont des autres recherches traitées dans ce projet.

logiciel évite tout chargement ou traitement intempestif ou superflu.

Effectivement ces trois points se sont révélés fondamentaux lors de la conception et du codage de VISION. Ainsi, comme nous l'avons vu dans le chapitre 7, le nombre de classes impliquées est grand, et par conséquent le nombre de lignes de code en découlant est important. À l'heure actuelle, le développement de VISION est en phase de débogage de la partie *“extraction de l'information locale”* et contient plus de 10 000 lignes de code. Le document de spécifications logicielles, dont un extrait est présenté en annexe B, fait plus de 120 pages.

Il est difficile de concevoir un logiciel “plus léger” pour des tests de validation du modèle ELECA car la taille du logiciel est contrainte en partie par la création de la structure de données. Liant très intimement les aspects vectoriels et les aspects raster, *le traitement n'autorise pas une approche “plus vectorielle” ou “plus raster”* tirant ainsi partie le plus possible de chaque type de données.

Il s'est également révélé exact que la grande “intimité” des modes vectoriel et raster impliquait un traitement en commun des deux types de données. Par exemple, le modèle ELECA impose, lors de la fixation de l'œil virtuel, une gestion dynamique des données vectorielles extraites. Cela implique que le traitement de l'image ne soit pas découplé des données vectorielles : c'est pourquoi l'entreprise est si grande.

8.2.3 Comparaison avec l'existant

Pour terminer, mentionnons qu'il existe sur le marché des logiciels commerciaux qui traitent les deux types de structures de données géométriques que sont les for-

mats matriciel et vectoriel. Il est possible d'identifier trois logiciels commerciaux très importants sur le marché et qui présentent trois philosophies différentes : AutoCAD, Geomatica et ArcInfo. Aucun de ces logiciels ne propose une mise à jour automatique de la base de données (géométriques et descriptives) par l'image. Comme nous le soulignons précédemment, il s'agit d'un problème toujours ouvert.

8.2.3.1 AutoCAD

AutoCAD (AutoDesk) est un logiciel qui au départ a été conçu sur une base vectorielle et sur une structure spaghetti (sans topologie) pour l'ingénierie, le dessin industriel. Plus tard ont été rajoutés des modules pour le traitement de données raster, le géoréférencement et le traitement de bases de données. Ce sont des couches qui ont été rajoutées ensuite et ne sont par conséquent pas optimisées. Enfin, dans ce logiciel de dessin qu'on essaie d'utiliser pour les systèmes d'information géographique (SIG), la structure de donnée est très lourde et la mise à jour est fastidieuse.

8.2.3.2 Geomatica

Geomatica (PCI Geomatics) est un logiciel qui découle du logiciel PCI. Ce sont des logiciels dédiés à l'origine au traitement d'images de télédétection. La couche de données vectorielles est venue ensuite. Les traitements vectoriels sont limités. Ils se limitaient, au début, à la superposition de ces données géoréférencées au dessus des images. Un effort important a été mis sur l'aspect vectoriel dans la dernière version – Geomatica 9. Mais cela reste toujours une couche (même relativement importante) sur les données raster qui sont la base du logiciel.

8.2.3.3 ArcInfo

La solution qui semble la plus avancée dans l'intégration des types de données raster et vecteur est le SIG ArcInfo. ArcInfo (ESRI) s'appuie sur une philosophie inverse de Geomatica : il s'appuie des données vectorielles. ArcInfo est un puissant logiciel d'analyse spatiale, certainement l'un des plus populaires. Il est possible d'ajouter de nombreux modules permettant de manipuler des données aussi bien vectorielles que raster. Il est également possible de programmer directement dans le logiciel sa propre extension. Il est possible également de créer et d'éditer les données vectorielles de manière très poussée. Malheureusement, en ce qui nous concerne, la complétude du logiciel joue contre lui puisque cela implique une lourdeur des données à manipuler en raison des nombreuses fonctionnalités inutiles dans le cas qui nous intéresse. On reste tributaire de la structure imposée par le logiciel.

8.2.3.4 Bilan

En conclusion peu de logiciels sont bien aussi à l'aise en traitement vectoriel qu'en matriciel. Pour ceux qui, comme ArcInfo le sont le plus, on reste tributaire de la structure lourde du logiciel avec une mise à jour périlleuse des données. Ainsi il apparaît qu'il existe un chaînon manquant entre les différents logiciels ; c'est-à-dire un logiciel qui soit conçu à l'origine sur les deux types de données géométriques (vecteur et raster) avec une structure suffisamment légère pour faire facilement la mise à jour de données. VISION s'inscrit donc dans cette philosophie.

Ainsi, une solution logicielle fournissant l'intégration dès le départ des deux types de données comme nous le proposons, n'est pas, à notre connaissance, encore sur le

marché. Trois raisons se dégagent principalement pour expliquer cela :

1. Historiquement les producteurs de logiciels sont spécialisés dans un des deux segments mentionnés précédemment qui sont conditionnés par le type de données.
2. La complexité vient de la simplicité ! : il est complexe (nous l'avons vu) de concevoir un système simple et léger qui permette d'intégrer les deux aspects raster et vecteur¹³.
3. Les algorithmes qui traitent en commun les deux types de données sont très peu nombreux et encore au stade de la recherche. Cette thèse en est un exemple.

En conséquence, le développement logiciel proposé ici s'affranchissant de ces trois points (qui sont également des difficultés) est un pas en avant.

8.3 Conclusion

Le logiciel VISION présente les défauts de ses qualités. Il s'affranchit de la barrière traditionnelle entre les données dès sa conception. En contrepartie, l'architecture est complexe, mais reste malgré tout légère ; c'est-à-dire que VISION comporte de nombreuses classes (et instances de ces classes durant l'exécution) mais ces dernières ne sont pas très importantes en termes d'attributs et de méthodes.

Enfin, bien que la mise au point et l'implémentation soient longues, les algorithmes proposés dans ce logiciel et soutenus par le modèle ELECA sont très novateurs car ils utilisent des techniques inusitées dans le domaine de la géomatique. Ces algorithmes permettent de s'affranchir des difficultés traditionnellement rencontrées en traitement

¹³De plus il faut considérer que les logiciels commerciaux sont faits pour être vendus à un public, si possible, le plus large possible. Les solutions proposées dans ces logiciels sont donc les plus génériques possibles.

automatique des images de télédétection.

Quatrième partie

Bilan

Chapitre 9

Contributions

*All good things which exist are the fruits
of originality.
John Stuart Mill, On Liberty, 3.*

Ce chapitre est consacré à un survol général des contributions de ce travail doctoral. Après une brève introduction sur le contexte de la recherche passée et présente, nous montrons, dans un premier temps, que le modèle ELECA répond de manière originale aux objectifs spécifiés au début de cette thèse. Dans un second temps nous montrons qu'il est possible de concevoir et de réaliser une solution logicielle originale pour supporter le modèle. Enfin dans un troisième temps, des considérations plus larges sont abordées. Nous montrons comment ce travail exploratoire est novateur dans l'appréhension générale qui est faite du problème mais également vis-à-vis du traitement des images de télédétection.

9.1 Introduction

9.1.1 Le passé

L'appariement, la fusion de données et/ou la mise à jour de données de type SIG, avec des images de télédétection, est depuis longtemps un problème et reste de nos jours un sujet de recherche important [GGPD04]. Il est intéressant de noter que les ambitions de la recherche concernant ce sujet sont à l'heure actuelle plus modestes qu'il y a une quinzaine d'années. En effet, au début des années 90 la recherche s'est penchée sur l'intégration des données pensant que là était la clé pour l'intégration des systèmes [Wha91]. Bien que les outils logiciels actuels permettent de manipuler relativement facilement les différents types de formats, le problème d'intégration des données raster et vectorielles (et par conséquent la mise à jour des unes par rapport aux autres) reste ouvert.

À la fin des années 90, il est apparu que le problème d'extraction d'objets d'images de télédétection était très complexe et qu'il fallait s'y attaquer par morceaux – bâtiments, cours d'eau, routes, *etc* [SWH00]. À cette époque les chercheurs ont constaté que les experts réalisaient facilement les mises à jour des systèmes d'informations géographiques (SIG) à partir d'images de télédétection. Aussi se sont-ils tournés vers la réalisation de systèmes experts [PMS⁺99], [PST⁺99], qui modélisent les règles des experts. Cependant, les résultats n'ont pas été à la hauteur des attentes car l'extraction des connaissances est complexe [Fig01] et il faut que les règles qui en découlent couvrent tous les cas possibles [Qua04]. Or, dans ce type de travail les exceptions représentent la majorité des cas rencontrés.

9.1.2 Le présent

Les ambitions visées présentement par les chercheurs sont nettement moins grandes qu’au début des années 90. Derrière des titres comme “*Road Detection Statistics for Automated Quality Control of GIS Data*”^{1, 2} se cache souvent une très grande spécialisation de la méthode utilisée. Dans ce cas précis, la méthode se veut générique mais fonctionne uniquement pour des images de télédétection à très haute résolution spatiale. De plus, il s’avère que la méthode ne fonctionne pas dans les zones rurales à cause du bruit dans les images. Ainsi, les publications récentes s’attaquent principalement à la mise à jour de SIG en utilisant des images à très hautes résolutions spatiales et en zones urbaines. Les approches sont différentes : ondelettes [ZC04], graphes [GB03], mais toujours spécialisées. Les approches globales de type Alquier [AM96] sont maintenant minoritaires dans la littérature [SWH00]. De plus, nous avons vu qu’il n’existait pas sur le marché d’outil logiciel intégré qui permette l’implantation d’un modèle comme le modèle ELECA.

9.1.3 *Quid* du futur ?

L’approche que nous préconisons pour cette thèse est quelque peu à contre-courant des recherches actuelles en télédétection, mais plus proche en certains points des développements récents en vision. Les interactions entre la vision numérique et la vision biologique sont de plus en plus grandes et s’alimentent mutuellement [Itt05] [HB05]. En outre, le raisonnement spatial commence à être envisagé comme partie intégrante d’un système de vision [CMG⁺03]. La télédétection certainement, comme par le passé,

¹[GGPD04]

²Statistiques de détection de routes pour le contrôle automatique de la qualité des données SIG – traduction libre.

adaptera ces approches avec un délai de quelques années.

9.2 Le modèle ELECA

Le modèle ELECA³ propose une approche intégrée de l'image et de la base de données géographiques pour la mise à jour de cette dernière en utilisant des techniques inspirées de la vision humaine.

L'angle d'attaque du problème est, à notre connaissance, original. Le modèle ELECA répond à trois questions fondamentales (et successives) :

1. Comment retrouve-t-on l'information dans une image de télédétection qui n'est pas traitable par des méthodes classiques à cause du bruit et/ou des occultations ?
2. Quelle information locale extrait-on et comment l'extrait-on une fois celle-ci localisée ?
3. Quelle avenue de solution permet de proposer une intégration des informations locales dans une perspective de carte globale ?

En outre, le modèle s'attaque au problème de l'efficacité de traitement des données en terme de temps de calcul en proposant des solutions simples d'un point de vue de la manipulation et du traitement des données spatiales.

³ELECA : Éléments Linéaires Extraits par Calcul Attentionnel — *ELECA : Extraction of Linear Element by Cognitive Approach.*

9.2.1 Recherche de l'information dans l'image

Le modèle ELECA propose une voie nouvelle d'exploration de l'image qui s'appuie au maximum sur les connaissances *a priori* du territoire. Pour rechercher l'information dans l'image (et pouvoir mettre à jour les données SIG) le modèle utilise une technique inspirée de la vision humaine. Pour guider le regard virtuel, le modèle se sert de la saillance et d'un modèle de comportement saccadique.

La saillance est un "grand classique" en vision depuis Treisman et Gelade en 1980 [TG80] (puis Koch et Ullman en 1985 [KU85]) mais surtout depuis les travaux de Itti et Koch en 2001 [IK01] qui intègrent une approche descendante au précédent modèle ascendant de Koch et Ullman. Notre approche du calcul de la saillance, qui s'appuie sur les connaissances *a priori* du territoire (données SIG) est original.

Le couplage avec un modèle de comportement saccadique de l'œil virtuel propre au modèle ELECA est également novateur et permet de proposer une technique, d'une part, de recherche de l'information dans l'image qui s'affranchit des problèmes de résolution de l'image et de continuité des éléments et, qui, d'autre part, évite un traitement global de l'image. En cela le modèle répond au premier objectif spécifique que nous nous étions fixés lors du chapitre premier.

9.2.2 Extraction de l'information de l'image

L'extraction des arêtes, nous l'avons vu (chapitre 2), est un problème auquel de nombreux travaux se sont attaqués et qui est bien connu maintenant⁴. Le modèle ELECA

⁴Ce qui ne veut pas dire résolu.

propose d'utiliser une technique d'extraction de l'information locale qui s'appuie sur des filtres éprouvés, robustes et couramment utilisés en télédétection. L'originalité de la méthode réside dans la succession (l'articulation) des traitements qui permet de dégager rapidement les segments de lignes en même temps que leurs orientations et qui permet également de manipuler simultanément les données raster et les données vectorielles.

Tout comme l'œil humain le fait, l'extraction est menée localement. Les points cibles ont été calculés par le modèle de guidage du regard virtuel. Ainsi, si l'information extraite est erronée (par exemple une mauvaise orientation), le regard virtuel cherchera dans la "mauvaise direction". Mais si aucune information n'est dégagée par la suite (ce qui est le plus probable dans le cas où l'orientation est erronée), le regard repartira sur les valeurs de plus haute saillance – voir figure 8.3 page 231. Cela permet par conséquent un retour automatique aux zones de probabilité maximale de présence de l'information, évitant ainsi que le regard ne se perde dans l'image. Le modèle ELECA, en couplant ainsi deux méthodes (rapides) de déplacement de l'œil dans l'image et d'extraction de l'information locale apparaît robuste par construction. Le modèle répond donc au deuxième objectif spécifique fixé au premier chapitre.

9.2.3 Fusion des informations locales en une carte globale, modèle CLE

Le modèle ELECA propose une avenue de solution pour réaliser la fusion du nuage d'informations locales extraites par l'œil virtuel⁵ lorsque la base de données géographiques ne contient pas d'information, c'est-à-dire dans le mode découverte.

⁵Mais qui pourrait être également, dans un cas différent de traitement d'image, le résultat d'une extraction globale des composantes curvilignes de l'image par le filtre de Sobel par exemple.

De manière originale, le modèle s'appuie sur des travaux en raisonnement spatial qualitatif. La méthode proposée (baptisée modèle CLE) est inspirée de travaux liés à la navigation. L'ensemble de ces travaux [Lig98] [LE00] [EL04], que nous avons regroupés, selon la lecture que nous en avons faite, sous le nom de modèle LE, montre qu'il est possible de reconstruire l'environnement global d'un agent local lorsque celui-ci se déplace et explore son environnement. Le modèle ELECA montre qu'il existe des équivalences entre le modèle LE et le modèle CLE, et que, par conséquent, il est possible de fusionner les informations locales en courbes globales (modèle CLE) comme un agent⁶ reconstruit son environnement global à partir de ses vues locales. Le modèle CLE (sous partie du modèle ELECA) fournit ainsi une méthode pour répondre au troisième objectif spécifique annoncé en introduction de cette thèse.

9.2.4 Modèle ELECA et objectif général de la thèse

Le modèle ELECA, dans son ensemble, fournit une méthode pour répondre à l'objectif général de la thèse, c'est-à-dire une méthode qui permette la mise à jour de bases de données géographiques en utilisant des images de télédétection. Ce modèle s'affranchit des problèmes classiques en télédétection (à savoir le bruit, les éléments à la limite de résolution et/ou discontinus) en utilisant une technique de sous échantillonnage de l'image par saccades. Le sous échantillonnage de l'image introduit une difficulté, à savoir le morcellement des informations dans la scène visuelle. Pour pallier ce morcellement, le modèle ELECA propose une technique de fusion des données locales qui se base sur des approches du raisonnement spatial qualitatif.

⁶ *Agent* dans le sens de [EL04].

9.3 Aspects logiciels

En plus de proposer une méthode originale qui réponde à l'objectif général de la thèse, via le modèle ELECA, ce travail comporte un volet logiciel important avec l'élaboration du logiciel VISION qui supporte la solution ELECA.

9.3.1 Conception d'un système original

Le modèle ELECA mêle très intimement les aspects vectoriels et les aspects matriciels, tant du point de vue des données que du point de vue de leurs traitements. VISION, contrairement aux autres logiciels disponibles sur le marché, propose dès le départ de traiter les deux aspects conjointement. Ainsi, la structure de données vectorielles a été conçue en fonction des traitements qui seront effectués sur l'image; de la même manière, les traitements effectués sur l'image impliquent automatiquement, à la volée, la structure de données vectorielles.

De notre point de vue, un logiciel tel que VISION, avec une conception reposant à l'origine sur les deux aspects matriciel et vectoriel, sera incontournable pour réaliser des manipulations efficaces⁷ des données géospaciales dans les années à venir. En effet, bien que les capacités de traitement des machines augmentent rapidement, la masse de données à traiter est toujours plus considérable.

⁷dépendamment de son efficacité réelle après implantation complète.

9.3.2 Algorithmes et implantation du système

La dernière étape de ce travail doctoral a été d'implanter la solution originale conçue précédemment. Le programme VISION est construit selon l'architecture dessinée pour supporter le modèle ELECA. En outre, des solutions algorithmiques permettant d'implanter le modèle ont été proposées pour s'insérer dans le programme VISION. Ces solutions originales montrent qu'il est possible de réaliser informatiquement une implantation du modèle qui a été développé. Certains de ces algorithmes ont même été testés "à vide" (voir figure 8.3 page 231) puisque l'implantation du modèle de comportement du regard a été réalisée. À la date de rédaction de cette thèse, l'implantation de l'extraction des informations locales est en cours, en particulier l'aspect dynamique de la gestion des données vectorielles en lien avec l'analyse en composantes connexes.

9.4 Considérations générales

9.4.1 Aspect transversal de la recherche

D'une manière générale, il existe deux types de doctorats : les doctorats qui se spécialisent dans un domaine très pointu et les doctorats que nous qualifierons de transversaux. Tout comme la tendance qu'a la recherche actuelle à établir des ponts entre les différentes disciplines, ce travail est "à cheval" sur plusieurs domaines faisant ainsi partie des doctorats plutôt transversaux. Les sciences géomatiques sont par essence pluridisciplinaires, nous nous sommes appuyés également sur des disciplines qui ne font pas partie de la définition traditionnelle de la géomatique – voir chapitre 1.

La difficulté la plus importante de ce travail a été par conséquent d'appréhender les différents domaines⁸ que couvre cette recherche de manière à les intégrer de façon cohérente et rigoureuse. *Ce travail exploratoire jette les bases pour l'intégration de différents domaines traditionnellement hors de la télédétection* comme la vision naturelle (mécanismes attentionnels) et le raisonnement qualitatif (spatial) dans le but d'apporter des solutions à des problèmes auxquels les méthodes traditionnelles n'ont pas apporté de réponses satisfaisantes⁹. De ce point de vue, l'approche qui est développée dans ce travail suit la tendance que la vision numérique a amorcé depuis quelques années à savoir l'intégration de principes de la vision naturelle.

9.4.2 Aspect traitement d'image de télédétection

Si l'on se place du point de vue du traitement d'images de télédétection ce travail propose plusieurs aspects originaux. Nous en citerons trois ici ; ils concernent le sous échantillonnage de l'image, sa segmentation et l'intégration du raisonnement spatial qualitatif.

9.4.2.1 Sous échantillonnage et multi résolution

Le constat a été dressé, précédemment, que les images de télédétection présentaient le désavantage de n'avoir souvent qu'une information morcelée. Ceci étant dû, en par-

⁸Citons par exemple dans le désordre : la télédétection, les systèmes d'information géographique, le traitement d'images, la fusion de données, le raisonnement spatial qualitatif, le génie logiciel, la vision numérique et biologique, l'informatique, les mécanismes attentionnels. . .

⁹Il y a déjà des travaux qui intègrent la vision naturelle en télédétection mais ce sont des aspects liés au groupement perceptuel (en s'appuyant sur le théorie de la Gestalt entre autres) et non les mécanismes attentionnels, citons [AM96] et [Fit99] ; quant à l'intelligence artificielle elle a également été déjà utilisée en télédétection mais pas dans son volet raisonnement spatial qualitatif.

ticulier lorsque l'on parle de capteurs actifs et/ou de basse résolution, au bruit et aux obstructions¹⁰ entre le capteur et la cible. Il résulte que le traitement complet de l'image, outre le fait qu'il soit coûteux en terme de temps de calcul, est inutile puisqu'en de nombreuses localités l'information n'est pas disponible.

Le modèle ELECA ouvre donc une voie originale pour l'exploration des images de télédétection en s'appuyant sur une recherche par saccades. Cette manière de faire permet d'une part de ne pas traiter l'image en entier, et donc évite des traitements inutiles, et d'autre part évite de perdre le contact avec un objet partiellement occulté ou bruité en utilisant une "résolution supérieure" issue de la connaissance de la base données.

9.4.2.2 Segmentation de l'image

Nous avons dit précédemment que les techniques actuelles échouaient souvent parce que chaque image est un cas particulier, voir section 2.1.5. Une intégration des différentes approches peut apporter des pistes de solution mais cela s'avère compliqué.

L'approche que nous avons présentée dans le modèle ELECA, permet, de manière originale, de concilier plusieurs philosophies qui sous-tendent certaines approches classiques. Ainsi, nous avons vu dans la section précédente que la multi résolution peut être appréhendée de manière complètement différente de la manière traditionnelle – §2.1.4.3. En outre, l'appui du guidage du regard sur la base de données présente deux avantages importants. D'une part, cela permet une intégration des connaissances *a*

¹⁰Il faut noter que dans le cas de capteurs passifs, l'image étant acquise de manière oblique, les obstructions sont souvent plus importantes. Il suffit de penser à une route en forêt. Si le capteur est à la verticale de la route, cette dernière aura plus de chance d'être visible que lorsque le capteur n'est pas à la verticale (RADAR) car dans ce cas elle peut complètement disparaître sous la couverture végétale.

priori – une autre approche de la télédétection §2.1.4.4. D’autre part, mais cela permet également de segmenter l’image de manière originale. Cette segmentation, qui souvent se fait en prétraitement en utilisant la classification¹¹, §2.1.4.5, aide à dégager des zones contextuelles – champ, eau, zone urbaine, *etc.* Cette manière originale de réaliser la segmentation de l’image permet, dans notre cas, de s’affranchir d’un pré-traitement tout en gardant les avantages. Il convient ici de noter que la segmentation qui découle de notre approche est une *conséquence* du modèle ELECA, ce n’est pas une finalité en soi. Ainsi, une segmentation réalisée de cette manière aura la limitation qu’elle est réalisée uniquement vis-à-vis des éléments curvilignes, d’autres aspects de type surfacique et/ou multi spectral n’étant pas directement pris en compte par définition.

Le modèle ELECA concilie avantageusement par conséquent trois philosophies, souvent distinctes, de la télédétection, *id est* la multi résolution, l’intégration des connaissances et la classification¹² de l’image.

9.4.2.3 Raisonnement spatial qualitatif

Fusion de données Le modèle ELECA s’appuie de manière originale sur des travaux en raisonnement spatial qualitatif pour “reconstruire” la structure spatiale de la scène visuelle. Il est intéressant de noter à ce point, que le raisonnement spatial qualitatif reçoit depuis quelque temps une justification cognitive grâce aux travaux de Knauff [KSJ⁺04]. Ainsi, l’exploration de l’image est guidée par des mécanismes inspirés de la vision humaine mais également la reconstruction de la structure globale qui s’appuie sur le raisonnement spatial qualitatif, peut être vue comme s’approchant du raisonnement

¹¹Nous nous plaçons dans le cas des images de télédétection.

¹²*Classification* par abus de langage puisque dans le cas qui nous intéresse, l’extraction d’éléments curvilignes, elle sert à réaliser une segmentation de l’image – §2.1.4.5.

humain.

L'approche du modèle CLE propose de transposer un modèle de reconstruction de l'environnement par une série de vues locales vers la reconstruction de structures curvilignes (environnement) à partir de petits morceaux de ces éléments (vues locales). Cette manière de mener la fusion de données grâce à des méthodes *spatiales* qualitatives telles que celles présentées ici (c'est-à-dire en utilisant des outils de reconstruction d'espaces globaux) n'avait jamais été explorée en télédétection avant ce travail, à notre connaissance. En revanche, il convient de remarquer que d'autres méthodes qualitatives de fusion de données, très différentes de celle présentée dans cette thèse, comme la logique floue ont déjà été utilisées en télédétection introduisant par là même des méthodes spatiales qualitatives.

Vision Dans le domaine de la vision, le raisonnement *spatial* qualitatif commence à faire son entrée [CMG⁺03]. L'analyse qualitative permet de produire des classes de généralisation du comportement plus faciles à manipuler¹³. Elle évite la propagation des erreurs à cause du bruit qui peut exister dans les données, quand elles sont traitées quantitativement. Cohn *et al.* [CMG⁺03] vont plus loin qu'une simple intégration d'aspects qualitatifs dans ce qui se fait déjà mais proposent une ébauche d'architecture dès la début basée sur la prise en compte de l'aspect qualitatif.

De ce point de vue, le modèle ELECA se situe dans la lignée des travaux les plus récents, mais en appliquant de manière originale le domaine à la télédétection, à la fusion d'informations et aux les problèmes de mise à jour cartographiques.

Enfin, si l'on considère le modèle de saccades et en particulier sa zone cible (voir

¹³Ici ce sont plutôt des scènes dynamiques qui sont étudiées.

figures 4.16, 4.17 et 8.2), son comportement intègre une composante qualitative. En effet, le calcul du point de chute de la saccade est mené de manière qualitative, puisqu'il s'agit d'une zone et non d'un point. Nous avons décrit, de ce point de vue, un algorithme simple de recherche dans le "camembert". Ces zones s'apparentent à celles décrites par Clementini *et al.* [CDFH97] dans leurs travaux sur la représentation qualitative de l'information positionnelle, voir §2.5. Le modèle ELECA utilise de manière originale un type de découpage qui s'apparente à ces travaux mais en l'appliquant à la saillance. Ainsi, le calcul de la saillance, lors d'une saccade, est judicieusement modulé de manière qualitative pour éviter de "passer à côté" de l'information.

9.4.3 Stratégies générales

En plus de la stratégie d'intégration transversale de différents domaines de recherche, ce travail utilise des stratégies originales ou peu utilisées dans le domaine d'application concerné. Ces stratégies concernent d'abord l'utilisation mixte qui est faite de l'information (raster et vecteur), ensuite la méthode mise en place pour mener à bien ce travail et enfin le choix de la simplicité et de la modularité

9.4.3.1 Approche mixte ascendante et descendante

La stratégie définie pour extraire l'information de l'image a été résolument d'utiliser toutes les informations connues *a priori*. Ce qui nous a amené à proposer un modèle (ELECA) à la fois ascendant et descendant. Ce type d'approche, qui est utilisée en vision depuis 2001 et les travaux de Itti et Koch [IK01], est original en télédétection. Nous rappelons ici que la composante descendante se situe dans la construction de la

liste des priorités, en cherchant *a priori* les éléments grâce à l'information de la base de données¹⁴ et que la composante ascendante apparaît dans le calcul de la saccade puisque celui-ci est réalisé grâce à l'information locale extraite directement de l'image.

9.4.3.2 Méthode de recherche

La méthode de recherche qui a été appliquée pour la construction du modèle ELECA a été déterminée par le problème général. Sachant que le problème de la mise à jour des données géographiques par les images de télédétection restait un problème ouvert, et que l'extraction en général était toujours un sujet de recherche sans solution générique, nous avons pris résolument le parti d'utiliser les approches non conventionnelles en télédétection dont il a été fait mention tout au long de cette thèse.

Il a fallu, par conséquent, adopter l'approche par spécification-solution qui a été évoquée au cours du document. En effet, en proposant une solution complètement nouvelle comme le modèle ELECA, nous avons dû spécifier à chaque fois plus précisément les problèmes à résoudre. Cette stratégie exploratoire (et peu utilisée à notre connaissance) a permis d'élaborer le modèle ELECA qui atteint les trois objectifs spécifiques fixés au début du travail, §1.4, page 14. En outre les spécifications larges au début puis de plus en plus fines au fur et à mesure des itérations de recherche ont conduit à l'élaboration de l'architecture du logiciel VISION de manière la plus générique possible.

¹⁴Le fait que la base de données ne fournisse pas d'information dans certaines zones est également une information à part entière.

9.4.3.3 Simplicité et modularité

Une autre stratégie utilisée dans cette thèse s'applique à la conception du modèle, de l'architecture logicielle et des algorithmes, ainsi qu'à la modularité du programme. Ce travail étant conditionné au départ par un souci d'efficacité en terme de temps de traitement, les solutions ont été pensées de manière à être les plus simples possibles, ce qui est un défi lorsque l'on s'attaque à ce genre de problème et qui est relativement peu courant. En outre, l'architecture modulaire du programme telle qu'elle a été décrite doit permettre de moduler facilement les stratégies de recherche dans l'image ; par exemple le nombre de rétroactions après la fusion, le suivi de contours ou non, *etc.*

9.5 Conclusion

Ce travail doctoral apporte une contribution générale originale à deux problèmes récurrents majeurs de la géomatique, à savoir la mise à jour des bases de données géographiques par l'information tirée de la télédétection et plus généralement à l'extraction d'objets curvilignes sur des images de télédétection. L'approche préconisée s'appuie sur différents domaines traditionnellement peu utilisés en géomatique et en télédétection en particulier. Cette méthode nous a permis d'élaborer un modèle ELECA complètement nouveau. L'élaboration de ce modèle en passant par des phases de spécifications de plus en plus précises a généré une série de contributions spécifiques originales. Enfin, à partir du modèle ELECA, nous avons proposé une solution logicielle légère, souple et adaptative qui intègre complètement les deux types de données raster (*i.e.* image de télédétection) et vectoriel (*i.e.* données du système d'informations géographiques).

Chapitre 10

Conclusions et perspectives

*Ah ! quel talent je vais avoir demain.
Hector Berlioz, Mémoires.*

10.1 Conclusions

Suite à ce travail doctoral, différentes conclusions peuvent être tirées. Nous mettrons en exergue les principales :

- Le problème de la mise à jour des bases de données à l'aide des images de télédétection reste un problème ouvert. Ce travail de doctorat, et en particulier le modèle ELECA, fournit une réponse originale qui ouvre de nouvelles perspectives de recherche. L'intégration transversale de différents domaines qui ne sont pas propres à la géomatique ont permis de montrer que des voies de solutions peuvent contourner le problème des images de basse résolution et/ou bruitées en

utilisant un modèle comme le modèle ELECA. Ce modèle montre qu'il est possible de concevoir une solution intégrée, à vocation automatique, pour des images génériques de télédétection, pour des éléments curvilignes.

- Dans ce travail nous proposons une solution au problème de recherche de l'information dans l'image, par une technique inspirée de la vision humaine. Cette technique par saccades est une manière originale en télédétection de sous échantillonner l'image, évitant ainsi un traitement inutile de l'image d'une part, et, réduisant la charge de calcul d'autre part, autorisant ainsi une plus grande efficacité en terme de temps de traitement. Nous insistons sur le fait que l'idée ici n'était pas de copier le regard humain, mais de regarder les aspects qui pouvaient être intéressants pour notre problème.
- Le modèle ELECA propose de lier intimement les deux types de données utilisés en général en géomatique à savoir les données vectorielles et les données raster. Dans le cas qui nous intéresse ici, les calculs d'un type alimentent ceux de l'autre type et vice et versa. Nous soulignons ici le fait que le modèle ELECA a vocation à utiliser toutes les informations à sa disposition : informations vectorielles, raster, connaissances *a priori*, absence d'information sur une région...
- L'approche qui est préconisée pour réaliser la fusion des informations locales s'appuie sur des modèles de reconstruction de l'environnement à partir de vues locales. Ces modèles sont des modèles qualitatifs. Le raisonnement spatial qualitatif est utilisé pour reconstruire, à partir d'informations parcellaires, la structure spatiale des éléments (curvilignes) qui constituent la scène visuelle.
- Dans ce travail, nous avons élaboré un modèle qui propose des solutions au problème d'extraction posé. De plus, nous avons montré que ce modèle était réalisable d'un point de vue informatique. L'architecture logicielle proposée pour supporter le modèle ELECA est originale car elle autorise la manipulation simultanée des données raster et vecteur, ce qui n'est pas proposé dans les logiciels présents

sur le marché. Les étapes de traitement (diagramme d'activité) et l'architecture (diagramme de classes) ont été entièrement spécifiées.

- Les solutions proposées dans le modèle ELECA, pour ce qui est des parties du guidage et de l'extraction locale, ont été traduites en termes d'algorithmes. Ces algorithmes originaux présentent des solutions originales directement implantables. Une implantation a été réalisée pour le guidage, l'implantation de l'extraction d'information locale est en cours. Ces implantations prennent place dans le logiciel VISION qui est construit selon la conception architecturale précédemment évoquée.

10.2 Perspectives

De part la nature exploratoire de cette thèse, les perspectives de recherche que l'on peut dégager sont très larges. En voici les principales :

- La première perspective de recherche de ce travail est de terminer l'implantation du modèle ELECA et de procéder à une analyse des performances ensuite. La comparaison avec un expert est un aspect très important pour mesurer d'une part le taux d'automatisation et d'autre part la qualité de l'information extraite. Comme nous l'avons évoqué, la tâche est en cours et elle est importante. La rigueur qui a caractérisé le développement de VISION (programme en anglais, code au plus près de normes ANSI et documentation complète de 120 pages) permet de continuer le travail que ce soit par le concepteur ou par une tierce personne.
- L'intégration des différents domaines de recherche au sein de ce travail ouvre des perspectives séduisantes de collaboration avec d'autres chercheurs. Ainsi les trois

aspects les plus importants du modèle ELECA peuvent être raffinés :

- Concernant le premier aspect, le modèle de guidage de l’œil virtuel, d’autres stratégies de comportement de l’œil peuvent être intégrées en utilisant d’autres modèles issus de la littérature ou de collaboration avec des psychologues en cognition. Cette approche devrait permettre de raffiner les algorithmes de recherche de l’information dans l’image.
- Le second aspect du modèle ELECA à considérer est l’extraction locale d’information. Une réflexion sur la manière d’obtenir l’orientation le plus rapidement possible tout en quantifiant l’erreur serait judicieuse. Un indicateur de “fiabilité” de l’information extraite pourrait aider à mieux conditionner la recherche au cours de la saccade suivante ; par exemple insister plus ou moins sur la recherche dans une zone en fonction de la fiabilité de l’information qui a permis d’arriver à la zone en question. Le premier indicateur de qualité qui vient à l’esprit est le ratio nombre de pixels candidats sur nombre de pixels total. Plus ce ratio est important, moins la probabilité que l’objet soit curviligne est importante, entraînant par là même une plus grande possibilité d’erreur.
- Le troisième aspect important du modèle ELECA concerne la fusion de données en utilisant des techniques issues du raisonnement spatial qualitatif. Ce travail doctoral a ouvert des perspectives en essayant de jeter les bases d’une équivalence formelle entre deux modèles. Cette équivalence formelle doit être réalisée maintenant au complet pour pouvoir appliquer l’approche par contraintes du modèle de départ.
- Mentionnons pour terminer une dernière perspective de recherche parmi d’autres. Nous avons vu comment segmenter l’image de manière originale avec le modèle ELECA. Cette manière de faire ouvre toute une série de perspectives pour segmenter les images sans pour autant appliquer de pré-traitement. On peut imaginer des stratégies de recherche différentes en fonction des zones segmentées de l’image afin

d'augmenter l'efficacité de la méthode. Si l'on regarde maintenant à l'extérieur du cas de notre travail, il est possible de penser que ce type de segmentation pourra aider à faire la classification d'images de télédétection en apportant une nouvelle méthode pour ajouter de l'information *a priori* portant sur les frontières dans l'image.

L'approche originale que nous avons adoptée dans cette thèse, ouvre, par son caractère exploratoire de nouvelles perspectives de recherche en télédétection et dans la mise à jour automatique des bases de données géographiques. Il convient de noter que tout pas dans la direction d'une automatisation des méthodes d'extraction, apporte des nouvelles perspectives quant à la semi automatisation. Nous souhaitons que ce travail soit une contribution qui suscite l'intérêt et que cette avancée originale soit une étape novatrice dans le domaine.

Appendice

Bibliographie

- [All83] J. F. ALLEN : Maintaining knowledge about temporal intervals. *In Communications of the ACM*, volume 26, pages 832–843, novembre 1983. [84, 85, 86]
- [Alq98] L. ALQUIER : *Analyse et Représentation de Scènes Complexes par Groupement Perceptuel*. Thèse de doctorat, Université de Montpellier II, Montpellier, France, septembre 1998. [63, 97]
- [AM96] L. ALQUIER et Ph. MONTESINOS : Perceptual grouping and active contour functions for the extraction of roads in satellite pictures. *In Image and Signal Processing*, décembre 1996. [4, 48, 63, 249, 256]
- [AM98] L. ALQUIER et Ph. MONTESINOS : Representation of linear structures using perceptual grouping. *In IEEE Computer Society Workshop on Perceptual Organization in Computer Vision, June 26*, pages 1–19, 1998. [63]
- [ASL90] M. M. ANSOULT, P. J. SOUILLE et J. A. LOODTS : Mathematical morphology : A tool for automated GIS data acquisition from scanned thematic maps. *Photogrammetric Engineering & Remote Sensing*, 56(9):1263–1271, 1990. [32, 33]
- [BCZC97] M. BARZO HAR, M. COHEN, I. ZISKIND et D. COOPER : *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, chapitre

- Fast Robust Tracking of Curvy Partially Occluded Roads in Clutter in Aerial Images, pages 277–286. Birkhauser Verlag, Switzerland, 1997. [4]
- [Bec05] A. BECHATA : La fonction gamma. <http://abdellah.bechata.free.fr/telechargement/gamma/pdf/gamma.pdf>, mai 2005. [102]
- [BEN⁺93] R. BARBER, W. EQUITZ, W. NIBLACK, D. PETKOVIC et P. YANKER : Efficient query by image content for very large image databases. *In Digest of Papers. COMPCON Spring '93*, pages 17–19, San Francisco, CA, USA, 1993. [80]
- [BG90] V. BRUCE et P. R. GREEN : *Visual perception, Physiology, Psychology and Ecology*. Lawrence Erlbaum Associates Ltd. Publishers, 1990. [39, 41, 45, 50, 51]
- [BG00] D. BROCKMANN et T. GEISEL : The ecology of gaze shifts. *Neurocomputing*, 32-33:643–650, 2000. [5, 63, 64, 127]
- [BHN97] G. BREWKA, Ch. HABEL et B. NEBEL, éditeurs. *KI-97 : Advances in Artificial Intelligence, 21st Annual German Conference on Artificial Intelligence, Freiburg, Germany, September 9-12, 1997, Proceedings*, volume 1303 de *Lecture Notes in Computer Science*. Springer, 1997. [274]
- [BJTB99] B. BRODARIC, M. JOURNEYAY, S. TALWAR et E. BOISVERT : NADM variant fact sheet – CORDLink. Conceptual model v5.2, Geological Survey of Canada, http://www.nadm-geo.org/dmdt/CORDLink_Variant_Description.html, 1999. [211]
- [BM00] Ph. BALBIANI et Ph. MULLER : Le raisonnement spatial. *In Henri PRADE, Robert JEANSOULIN et Catherine GARBAY, éditeurs : Le temps, l'espace et l'évolutif en sciences du traitement de l'information, Marseille, France*, pages 33–53. CEPAD, 1-30 septembre 2000. [83]

-
- [BSM⁺99] A. BAUMGARTNER, C. STEGER, H. MAYER, W. ECKSTEIN et H. EBNER : Automatic road extraction based on multi-scale grouping and context. *Photogrammetric Engineering & Remote Sensing*, 65(7):777–785, 1999. [29, 37]
- [Car00] C. CARON : Gestion de données spatiales – Concepts de base. Graduate course, Université Laval, département des Sciences Géomatiques, Québec, Canada, 2000. [16]
- [CC01] R. CALOZ et C. COLLET : *Précis de télédétection*, volume 3. Traitements numériques d’images de télédétection. Agence Universitaire de la Francophonie, Québec, Canada, 2001. [4, 31, 32, 149, 293, 294]
- [CCL97] V. CANTONI, G. CAPUTO et L. LOMBARDI : *Attentional Mechanisms in Vision Systems*, chapitre Active Vision, pages 3–42. In CANTONI et al. [CLR97], 1997. [53]
- [CCP96] J. CALMET, J. A. CAMPBELL et J. PFALZGRAF, éditeurs. *Artificial Intelligence and Symbolic Mathematical Computation, International Conference AISMC-3, Steyr, Austria, September 23-25, 1996, Proceedings*, volume 1138 de *Lecture Notes in Computer Science*. Springer, 1996. [274]
- [CDFH97] E. CLEMENTINI, P. DI FELICE et D. HERNÁNDEZ : Qualitative representation of positional information. *Artificial Intelligence*, 95(2):317–356, 1997. [88, 89, 260]
- [Cen05] CENTRE CANADIEN DE TÉLÉDÉTECTION : Glossaire de télédétection du CCT. http://www.ccrs.nrcan.gc.ca/ccrs/learn/terms/glossary/glossary_f.html, août 2005. [293, 295]
- [CH01] A. G. COHN et S. M. HAZARIKA : Qualitative spatial representation and reasoning : An overview. *Fundamenta Informaticæ*, 46(1-2):1–29, 2001. [83, 86]

- [CLR97] V. CANTONI, S. LEVIALDI et V. ROBERTO, éditeurs. *Artificial Vision : Image Description, Recognition and Communication*. Signal Processing and its Applications. Academic Press, 1997. [273]
- [CLR00] E. COHEN, N. LEVY et E. RUPPIN : Global vs. local processing of compressed representations : A computational model of visual search. *Neurocomputing*, 32-33:667–671, 2000. [65, 66, 72]
- [CMEL03] G. COTTERET, B. MOULIN, G. EDWARDS et G. LIGOZAT : Using a guided virtual gaze for the analysis of low resolution remote sensing images. In LEVACHKINE *et al.* [LSE03]. CD-ROM. [13, 115]
- [CMG⁺03] A. G. COHN, D. R. MAGEE, A. GALATA, D. HOGG et S. M. HAZARIKA : Towards an architecture for cognitive vision using qualitative spatio-temporal representations and abduction. In FREKSA *et al.* [FBHW03], pages 232–248. [249, 259]
- [Coh96] A. G. COHN : Calculi for qualitative spatial reasoning. In CALMET *et al.* [CCP96], pages 124–143. [83]
- [Coh97] A. G. COHN : Qualitative spatial representation and reasoning techniques. In BREWKA *et al.* [BHN97], pages 1–30. [83]
- [Cot02] G. COTTERET : Extraction d'éléments curvilignes guidée par des mécanismes attentionnels pour des images de télédétection : approche par fusion de données. Projet de doctorat, Département des Sciences Géomatiques, Québec, Canada, février 2002. [11, 16]
- [Cou99] D. COUSINEAU : *Analyses et modèles des temps de réponse, de l'apprentissage, et de l'automatisation de la recherche mnésique et visuelle*. Thèse de doctorat, Université de Montréal, 1999. [57]
- [CR99] E. COHEN et E. RUPPIN : From parallel to serial processing : A computational study of visual search. *Perception & Psychophysics*, 61(7):1449–

-
- 1461, 1999. [64, 65, 66]
- [CW94] A. D. CALWAY et R. WILSON : Curve extraction in images using a multi-resolution framework. *Computer Vision, Graphics, and Image Processing. Image Understanding*, 59(3):359–366, mai 1994. [75, 76]
- [Dav90] E. DAVIS : *Representations of Commonsense Knowledge*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990. [83, 161]
- [DC94] D. DAGENBACH et T. H. CARR, éditeurs. *Inhibitory Processes in Attention, Memory and Language*. Academic Press, 1994. [61, 281]
- [DLN94] M. DOBIE, P. LEWIS et M. NIXON : Curvilinear network extraction from remotely sensed image. *In IAPR Workshop on Machine Vision Applications (IAPR MVA '94), December 13–15*, pages 123–126, Kawasaki, Japan, 1994. [4, 124, 127]
- [DM00] M. DAOUDI et S. MATUSIAK : Visual image retrieval by multiscale description of user sketches. *Journal of Visual Languages and Computing*, 11:287–301, 2000. [76, 77]
- [DNR02] C. DILLABAUGH, K. NIEMANN et D. RIDCHARSON : Semi-automated extraction of rivers from digital imagery. *GeoInformatica*, 6(3):263–284, 2002. [4]
- [DP73] D. H. DOUGLAS et T. K. PEUCKER : Algorithms for the reduction of the number of points required to represent a line or its caricature. *The Canadian Geographer*, 10(2):112–122, 1973. [105, 136, 200, 297]
- [DZ01] G. DECO et J. ZIHL : A neurodynamical model of visual attention : Feedback enhancement of spatial resolution in a hierarchical system. *Journal of Computational Neuroscience*, 10:231–253, 2001. [71]

- [EL04] G. EDWARDS et G. LIGOZAT : A formal model for structuring local perceptions of environmental space. *Cognitive Processing*, 5(1):3–9, mars 2004. [87, 162, 165, 166, 167, 253]
- [EP98] H.-E. ERIKSON et M. PENKER : *UML Toolkit*. John Wiley & Son, Inc., New York, NY, USA, 1998. [192, 215]
- [EZ98] J. H. ELDER et S. W. ZUCKER : Evidence for boundary-specific grouping. *Vision Research*, 38(1):143–152, 1998. [52]
- [FBF⁺94] C. FALOUTSOS, R. BARBER, M. FLICKNER, J. HAFNER, W. NIBLACK, D. PETKOVIC et W. EQUITZ : Efficient and effective querying by image content. *Journal of Intelligent Information Systems : Integrating Artificial Intelligence and Database Technologies*, 3(3-4):231–262, 1994. [80]
- [FBHW03] Ch. FREKSA, W. BRAUER, Ch. HABEL et K. F. WENDER, éditeurs. *Spatial Cognition III, Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Learning*, volume 2685 de *Lecture Notes in Computer Science*. Springer, 2003. [274]
- [FC98] N. C. FITTON et S. J. D. COX : Optimising the application of the Hough transform for automatic feature extraction from geoscientific images. *Computers & Geosciences*, 24(10):933–951, décembre 1998. [34]
- [FCF92] A. U. FRANK, I. CAMPARI et U. FORMENTINI, éditeurs. *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, International Conference GIS - From Space to Territory : Theories and Methods of Spatio-Temporal Reasoning, Pisa, Italy, September 21-23, 1992, Proceedings*, volume 639 de *Lecture Notes in Computer Science*. Springer, 1992. [277]
- [FG03] J. M. FINDLAY et I. A. GILCHRIST : *Active Vision : The Psychology of Looking and Seeing*. Oxford University Press, New York, NY, USA, 2003. [54, 57, 293]

-
- [Fit99] J. FITZBACK : Détection des cours d'eau par image radar selon les stratégies de la vision humaine. Mémoire de D.E.A., Université Laval, mars 1999. [4, 32, 63, 97, 101, 146, 149, 256]
- [FK95] A. U. FRANK et W. KUHN, éditeurs. *Spatial Information Theory : A Theoretical Basis for GIS, International Conference COSIT '95, Semmering, Austria, September 21-23, 1995, Proceedings*, volume 988 de *Lecture Notes in Computer Science*. Springer, 1995. [287]
- [FL99] R. FANNADER et H. LEROUX : *UML Principes et modélisation*. Dunod, Paris, France, 1999. [192]
- [Fog94] H. C. FOGEDBY : Lévy flights in random environments. *Physical Review Letters*, 73(19):2517–2520, novembre 1994. [64]
- [Fra91] A. U. FRANK : Qualitative spatial reasoning about cardinal directions. *In Proceedings of the 7th Austrian Conference on Artificial Intelligence*, pages 157–167, Vienna, Austria, 1991. [164]
- [Fre92] Ch. FREKSA : Using orientation information for qualitative spatial reasoning. *In FRANK et al. [FCF92]*, pages 162–178. [164]
- [FS94] A. FINKELSTEIN et D. H. SALESIN : Multiresolution curves. *In SIGGRAPH 94*, 1994. [78, 79]
- [FTW81] M. FISCHLER, J. TENENBAUM et H. WOLF : Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. *Computer Graphics and Image Processing*, 15:201–223, 1981. [4]
- [GB03] S. GAUTAMA et A. BORGHGRAEF : Using graph matching to compare VHR satellite images with GIS data. *In Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, volume 1, pages 173–175, juillet 2003. [249]

- [GGPD04] W. GOEMAN, S. GAUTAMA, W. PHILIPS et J. D’HAEYER : Road detection statistics for automated quality control of GIS data. *In Geoscience and Remote Sensing Symposium, 2004. IGARSS ’04. Proceedings. 2004 IEEE International*, volume 5, pages 2826–2829, 2004. [30, 248, 249]
- [Gib96] D. GIBERT : Éléments de traitement du signal. Graduate course, Équipe de géophysique interne (UPR 4661 CNRS), Rennes, France, janvier 1996. [78]
- [Gib03] L. GIBSON : Finding road networks in IKONOS satellite imagery. *In Proceedings of the ASPRS Annual Conference, 05-09 May, Anchorage, AK, USA, 2003. CD-ROM*. [26]
- [GJ96] D. GEMAN et B. JEDYNAK : An active testing model for tracking roads in satellite images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):1–14, 1996. [4]
- [GL97] A. GRUEN et H. LI : Semi-automatic linear feature extraction by dynamic programming and LSB-snakes. *Photogrammetric Engineering & Remote Sensing*, 63(8):985–995, 1997. [29, 37]
- [Géo97] GÉOMATIQUE CANADA : Normes et spécifications de la base nationale de données topographiques. Document technique, Ministère des Approvisionnements et Services Canada, Sherbrooke, Canada, mai 1997. [103, 104, 105]
- [Gre00] R. L. GREGORY : *L’œil et le cerveau : la psychologie de la vision*. Neurosciences & cognition. De Boeck Université, Bruxelles, 2000. [6, 55]
- [GS93] T. GEVERS et W. M. SMUELDERS : An approach to image retrieval for image databases. *In V. MARIK, J. LAZANSKY et R. R. WAGNER, éditeurs : Database and Expert Systems Applications (DEXA ’93)*, pages 615–626, Prague, Czechoslovakia, 1993. [80]

-
- [Gui99] B. GUINDON : The development of perceptual grouping operators to aid in the extraction of planimetric features from high resolution satellite imagery. *In Proceedings of the ASPRS Annual Conference, 17-21 May, Portland, OR, USA, 1999.* [29]
- [GW02] R. C. GONZLÁEZ et R. E. WOODS : *Digital Image Processing.* Prentice Hall, Upper Saddle River, N.J., 2nd édition, 2002. [153, 293, 294]
- [HB05] M. HAYHOE et D. BALLARD : Eye movements in natural behavior. *TRENDS in Cognitive Sciences*, 9(4):188–194, avril 2005. [6, 249]
- [Hen04] J. B. HENRY : *Systèmes d'information spatiaux pour la gestion du risque d'inondation de plaine.* Thèse de doctorat, Université de Strasbourg I, avril 2004. [94, 101]
- [Her94] D. HERNÁNDEZ : *Qualitative Representation of Spatial Knowledge*, volume 804 de *Lecture Notes in Artificial Intelligence.* Springer-Verlag, Berlin, Germany, 1994. [83, 86, 164]
- [HK92] K. HIRATA et T. KATO : Query by visual example — content based image retrieval. *In Advances in Database Technologies (EDTB '92)*, pages 56–71, Vienna, Austria, 1992. [80]
- [HM85] J. R. HOBBS et R. C. MOORE, éditeurs. *Formal Theories of the Commonsense World.* Greenwood Publishing Group Inc., Westport, CT, USA, 1985. [83, 161]
- [Hor00] W. HORN, éditeur. *ECAI 2000, Proceedings of the 14th European Conference on Artificial Intelligence, Berlin, Germany, August 20-25, 2000.* IOS Press, 2000. [284]
- [HW98] T. S. HOROWITZ et J. M. WOLFE : Visual search has no memory. *Nature*, 394:575–577, août 1998. [58, 62, 64]

- [IK00] L. ITTI et C. KOCH : A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, mai 2000. [70, 71, 72]
- [IK01] L. ITTI et C. KOCH : Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, mars 2001. [6, 7, 61, 62, 66, 67, 69, 70, 71, 251, 260]
- [ISA95] J. IISAKA et T. SAKURAI-AMANO : Automated terrain feature detection from remotely sensed images integrating spectral, spatial and geometrical attributes of objects. *In Proceedings of the GIS/LIS Conference, 14-16 May, Nashville, TE, USA, 1995*. [37]
- [ISO05] ISO/TC211 : Geographic information/geomatics. <http://www.tc211.org/>, mai 2005. [103, 211, 291]
- [Itt03] L. ITTI : Visual attention. *In M. A. ARBIB, éditeur : The Handbook of Brain Theory and Neural Networks, 2nd Ed.*, pages 1196–1201. MIT Press, 2003. [5, 7]
- [Itt05] L. ITTI : Quantitative modeling of perceptual salience at human eye position. *Visual Cognition*, 2005. Accepté. [249]
- [JEC01] R. JEANSOULIN, G. EDWARDS et G. COTTERET : Kalman filter, and logical revision : comparing them on geographic information. *In GI-FaR'01 (Geographic Information Fusion and Revision)*, Québec, Canada, avril 2001. [16]
- [JFS95] C. E. JACOBS, A. FINKELSTEIN et D. H. SALESIN : Fast multiresolution image querying. *In SIGGRAPH 95*, 1995. [79, 80]
- [Kat92] T. KATO : Database architecture for content-based image retrieval. *In Proceedings of the SPIE — The International Society for Optical Engineering*, volume 1662, pages 112–113, San Jose, CA, USA, 1992. [80]

-
- [KKOH92] T. KATO, T. KURITA, N. OTSU et K. HIRATA : A sketch retrieval method for full color image database — query by visual example. *In Proceedings of the 11th IAPR International Conference on Pattern Recognition*, pages 530–533, Los Alamitos, CA, USA, 1992. [80]
- [Kle00] R. M. KLEIN : Inhibition of return. *Trends in Cognitive Sciences*, 4(4):138–147, avril 2000. [60]
- [Kre93] E. KREYSZIG : *Advanced Engineering Mathematics*. John Wiley & Sons, Inc., 7 édition, 1993. [102]
- [KSJ+04] M. KNAUFF, G. STRUBE, C. JOLA, R. REINHOLD et Ch. SCHLIEDER : The psychological validity of qualitative spatial reasoning in one dimension. *Spatial Cognition & Computation*, 4(2):167–188, 2004. [84, 85, 86, 258]
- [KT94] R. M. KLEIN et T. L. TAYLOR : *Inhibitory Processes in Attention, Memory and Language*, chapitre Categories of Cognitive Inhibition, with Reference to Attention, pages 113–150. *In* DAGENBACH et CARR [DC94], 1994. [61]
- [KU85] C. KOCH et S. ULLMAN : Shifts in selective visual attention : Towards the underlying neuronal circuitry. *Human Neurobiology*, 4:219–227, 1985. [6, 66, 251]
- [KZT93] A. KITAMOTO, C. ZHOU et M. TAKAGI : Similarity retrieval on NOAA satellite imagery by graph matching. *In Storage and Retrieval for Image and Video Databases*, pages 60–73, San Jose, CA, USA, 1993. [80]
- [LC93] J. LIANG et C. C. CHANG : Similarity retrieval on pictorial databases based upon module operation. *In* S. MOON et H. IKEDA, éditeurs : *Database Systems for Advanced Applications*, pages 19–26, Taejon, South Korea, 1993. [80]
- [LE00] G. LIGOZAT et G. EDWARDS : Implicit spatial reference systems using proximity and alignment knowledge. *Spatial Cognition and Computation*,

- 2:373–392, 2000. [87, 88, 151, 162, 165, 253]
- [Lee81] J. S. LEE : Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing*, pages 24–31, 1981. [146]
- [Lig98] G. LIGOZAT : Reasoning about cardinal directions. *Journal of Visual Languages & Computing*, 9:23–44, mars 1998. [87, 162, 164, 167, 253]
- [LM94] P. B. LADKIN et R. D. MADDUX : On binary constraint problems. *Journal of the Association for Computing Machinery*, 41(3):435–469, 1994. [170]
- [LMC04] G. LIGOZAT, G. MITRA et J.-F. CONDOTTA : Spatial and temporal reasoning : beyond Allen’s calculus. *AI Communications*, 17(4):223–233, 2004. [86]
- [LN72] P. H. LINDSAY et D. A. NORMAN : *Human Information Processing*. Academic Press, New York, NY, USA, 1972. [50]
- [LR04] G. LIGOZAT et J. RENZ : What is a qualitative calculus? A general framework. In ZHANG *et al.* [ZGY04], pages 53–64. [165]
- [LRSO⁺00] R. LEPAGE, R. G. ROUHANA, B. ST.-ONGE, R. NOUMEIR et R. DESJARDINS : Cellular neural network for automated detection of geological lineaments on RADARSAT images. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1224–1233, mai 2000. [37]
- [LSE03] S. LEVACHKINE, J. SERRA et M. EGENHOFER, éditeurs. *Research on Computing Science*. National Polytechnic Institute, 2003. [274]
- [Mac02] J. MACHROUH : *Perception attentive et vision en intelligence artificielle*. Thèse de doctorat, University Paris XI, Orsay, France, décembre 2002. [5, 54, 72, 73, 74]

-
- [Mar76] D. MARR : Early processing of visual information. *In Philosophical Transactions of the Royal Society of London*, volume 275 de B, pages 483–524, 1976. [39]
- [Mar82] D. MARR : *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman and Co, San Francisco, 1982. [5, 38, 43]
- [MD88] D. MCKEOWN et J. DENLINGER : Cooperative methods for road tracking in aerial imagery. *In IEEE Computer Vision and Pattern Recognition, June 5–9*, pages 662–672, Ann Arbor, Michigan, USA, 1988. [4]
- [MH80] D. MARR et E. HILDRETH : Theory of edge detection. *In Proceedings of the Royal Society of London*, volume 197 de B, pages 441–475, 1980. [41, 42]
- [MJMC⁺96] D. M. MCKEOWN JR, C. MCGLONE, S. D. COCHRAN, Y. C. HSIEH, M. ROUX et J. A. SHULFELT : *Manual of Photogrammetry Addendum*, chapitre Feature Extraction and Object Recognition, pages 195–211. American Society for Photogrammetry and Remote Sensing, Bethesda, MA, USA, 1996. [29]
- [MLT01a] Y. MACHROUH, J. S. LIÉNARD et Ph. TARROUX : Exploration de scènes en vision artificielle. *In IVème Colloque des Jeunes Chercheurs en Sciences Cognitives, Lyon, France, 2-4 mai 2001*, 2001. [72]
- [MLT01b] Y. MACHROUH, J. S. LIÉNARD et Ph. TARROUX : Multiscale feature extraction from the visual environment in an active vision system. *In Springer Verlag Lecture Notes in COMPUTER SCIENCE*, éditeur : *4th International Workshop on Visual Form (IWVF4) Capri, Italy May, 28-30, 2001*, pages 388–397, Berlin, Germany, mai 2001. [34, 72]
- [MLT03] J. MACHROUH, J. S. LIÉNARD et Ph. TARROUX : Exploration de scènes en vision artificielle. *In Cognito, in revision*, pages 1–12, 2003. [72]

- [MN78] D. MARR et H. K. NISHIHARA : Representation and recognition of the spatial organization of three dimensional shapes. *In Proceedings of the Royal Society of London*, volume 207 de B, pages 187–216, 1978. [51]
- [Mor04] R. MORATZ : Qualitative spatial reasoning about oriented points. SFB/TR 8 Report 003-10/2004, Universität Bremen / Universität Freiburg, 2004. Report Series of the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition. [162, 168]
- [Mou02] B. MOULIN : Object-Oriented Project (Design and Management). Graduate course, Université Laval, département d’informatique, Québec, Canada, janvier 2002. [18, 192, 193]
- [MRW00] R. MORATZ, J. RENZ et D. WOLTER : Qualitative spatial reasoning about line segments. *In HORN [Hor00]*, pages 234–238. [162, 168, 169]
- [MT84] N. MACFARLANE et M. H. B. THOMAS : Speckle reduction algorithms and their applications to SAR images. *In 10th International Conference of Remote Sensing Society*, pages 391–398, 1984. [146]
- [MT05] J. MACHROUH et Ph. TARROUX : Attentional mechanisms for interactive image exploration. *EURASIP Journal on Applied Signal Processing*, 2005(14):1–6, 2005. [72]
- [MZ96] N. MERLET et J. ZERUBIA : New prospects in line detection by dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):426–431, 1996. [4]
- [NBE+93] W. NIBLACK, R. BARBER, W. EQUITZ, M. FLICKNER, E. GLASMAN, D. PETKOVIC, D. PETKOVIC, P. YANKER, C. FALOUTSOS et G. TAUBIN : The QBIC project : Querying images by content using color, texture and shape. *In Storage and Retrieval for Image Video Databases*, pages 173–187, SPIE, 1993. [80]

-
- [NI02] V. NAVALPAKKAM et L. ITTI : A goal oriented attention guidance model. *Lecture Notes in Computer Science*, 2525:453–461, novembre 2002. [5]
- [Nic20] J. NICOD : *La géométrie dans le monde sensible*. Thèse de doctorat, Université de la Sorbonne, Paris, France, 1920. [84]
- [Off05a] OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE : Cap sur la géomatique. http://www.olf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/terminologie_geomatique/lex_geomatique.html, juin 2005. [199]
- [Off05b] OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE : Grand dictionnaire terminologique. <http://www.granddictionnaire.com/>, août 2005. [294]
- [Pai94] J. PAILLARD : *L'intégration sensori-motrice et idéomotrice*, chapitre III.6, pages 925–961. In RICHELLE *et al.* [RRR94], 1994. [113]
- [Pig01] L. PIGEON : *Fusion d'informations appliquée à l'extraction et à l'identification de structures cartographiques stratégiques à partir d'images satellitaires*. Thèse de doctorat, Université Laval, avril 2001. [35, 248]
- [PKFW⁺01] M. PETERSON, A. KRAMER, R. FRANCES WANG, D. IRWIN et J. MCCARLEY : Visual search has memory. *Psychological Science*, 12(4):287–292, juillet 2001. [59]
- [PLN02] D. PARKHURST, K. LAW et E. NIEBUR : Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1):107–123, 2002. [70]
- [PMS⁺99] L. PIGEON, B. MOULIN, B. SOLAIMAN, T. TOUTIN et K. P. B. THOMSON : Human-experts features extraction strategies for topographical map production and updating. In *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. 1999 IEEE International*, volume 1, pages 296–298, juillet 1999. [25, 35, 248]

- [Pog84] T. POGGIO : *La perception visuelle*, chapitre Vision humaine et vision par ordinateur, pages 144–159. Belin, Paris, pour la science édition, 1984. [62, 63]
- [PR85] D. E. PEARSON et J. A. ROBINSON : Visual communications at very low data rates. *In Proceedings of the IEEE*, volume 73, pages 785–815, 1985. [40]
- [PST⁺99] L. PIGEON, B. SOLAIMAN, K. P. B. THOMSON, B. MOULIN et T. TOUTIN : Human-experts rules modeling for linear planimetric features extraction in a remotely sensed images data fusion context. *In Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. 1999 IEEE International*, volume 5, pages 2369–2371, juillet 1999. [29, 248]
- [PSTT93] G. PETRAGLIA, M. SEBILLO, M. TUCCI et G. TORTORA : Rotation invariant iconic indexing for image database retrieval. *In Proceedings of the 7th International Conference on Image Analysis and Processings*, pages 271–278, Monopoli, Italy, 1993. [80]
- [PTVF02] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING et B. P. FLANERY : *Numerical Recipes in C++ : The Art of Scientific Computing*. Cambridge University Press, Cambridge, MA., 2.1 édition, 2002. [102]
- [Qua04] L. J. QUACKENBUSH : A review of techniques for extracting linear feature from imagery. *Photogrammetric Engineering & Remote Sensing*, 70(12): 1383–1392, décembre 2004. [4, 24, 25, 26, 28, 29, 30, 32, 33, 35, 36, 37, 94, 248]
- [RBC⁺98] J. ROULIN, C. BONNET, J. CAMUS, P. CRADDOCK, D. GAONAC'H, A. GUERRIEN, S. NICOLAS, J. PASSERAULT et T. RIPOLL : *Psychologie Cognitive*, chapitre L'Attention, pages 138–170. Bréal, 1998. [57]

-
- [Ren02] J. RENZ : *Qualitative Spatial Reasoning with Topological Information*, volume 2293 de *Lecture Notes in Artificial Intelligence*. Springer-Verlag, Berlin, Germany, 2002. [83, 84, 86, 87, 89, 161]
- [Res03] RESSOURCES NATURELLES CANADA : Norme et catalogue de la base de données géospatiale (BDG). Document technique, Ministère des Approvisionnements et Services Canada, avril 2003. [103, 105]
- [Rev05] A. REVEL : Graphes et algorithmes. Graduate course, École Nationale Supérieure de l'Électronique et de ses Applications, Cergy-Pontoise, France, juillet 2005. <http://www.etis.ensea.fr/~revel/html/graphes/graphes.html>. [300]
- [RLB98] M. R. ROSENZWEIG, A. L. LEIMAN et S. M. BREEDLOVE : *Psychobiologie. Neurosciences & cognition*. De Boeck Université, Bruxelles, 1998. [293, 294]
- [Rob77] G. S. ROBINSON : Edge detection by compass gradient masks. *Computer Graphics, Image Processing*, 6(5):492–501, octobre 1977. [149]
- [Roy59] B. ROY : Transitivité et connexité. *Comptes rendus de l'Académie des Sciences*, 249:216–218, 1959. [303]
- [RRR94] M. RICHELLE, J. REQUIN et M. ROBERT, éditeurs. *Traité de psychologie expérimentale*. Presses universitaires de France, 1994. [113, 285]
- [Sch95] Ch. SCHLIEDER : Reasoning about ordering. In FRANK et KUHN [FK95], pages 341–349. [161, 169]
- [Sel59] O. G. SELFRIDGE : *The Mechanisation of Thought Processes*, chapitre Pandemonium : A Paradigm for Learning. H.M.S.O., London, UK., 1959. [50]
- [SGTB98] M. SIMARD, G. GEGRANDI, K. P. B. THOMSON et G. B. BÉNIÉ : Analysis of speckle noise contribution on wavelet decomposition of SAR images.

- IEEE Transactions of Geoscience and Remote Sensing*, 36(6):1953–1962, novembre 1998. [101]
- [SI04] C. SIAGIAN et L. ITTI : Biologically-inspired face detection : Non-brute-force-search approach. *In First IEEE-CVPR International Workshop on Face Processing in Video*, pages 62–69, juin 2004. [5]
- [Sim98] M. SIMARD : *Analyse de texture multiéchelle par transformée par ondelettes des images RSO : application aux images de la forêt tropicale*. Thèse de doctorat, Université Laval, février 1998. [101]
- [SWH00] B.-M. STRAUB, C. WIEDEMANN et C. HEIPKE : Towards the automatic interpretation of images for GIS update. *In International Archives of Remote Sensing and Photogrammetry*, volume XXXIII, pages 521–532, juillet 2000. [248, 249]
- [TG80] A. TREISMAN et G. GELADE : A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980. [57, 58, 64, 251]
- [TG00] Y. TAO et W. I. GROSKY : Image indexing and retrieval using object-based point feature maps. *Journal of Visual Languages and Computing*, 11:323–343, 2000. [81]
- [Tou90] J.-J. TOUMAZET : *Traitement d'image par l'exemple*. Sybex, Paris, France, 1990. [4]
- [Tsa83] Y. TSAL : Movement of attention across the visual field. *Journal of Experimental Psychology : Human Perception & Performance*, 9:523–530, 1983. [58]
- [VAB⁺96] G. D. VISWANATHAN, V. AFANASYEV, S. V. BULDYREV, E. J. MURPHY, P. A. PRINCE et H. E. DTANLEY : Lévy flight search patterns of wandering albatrosses. *Nature*, 381:413–415, mai 1996. [64]

-
- [VC95] C. VOIRON-CANICIO : *Analyse spatiale et analyse d'images par morphologie mathématique*. GIP-RECLUS, Montpellier, France, 1995. [4]
- [Vor08] G VORONOI : Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire : Recherches sur les paralléloèdres primitifs. *Journal für die reine und angewandte Mathematik*, 134:198–287, 1908. [174]
- [WA00] P. R. WOLF et Dewitt B. A. : *Elements of Photogrammetry : With Applications in GIS*. McGraw Hill, Boston, MA, USA, 3 édition, 2000. [25]
- [War62] S. WARSHALL : A theorem on boolean matrices. *Journal of the Association for Computing Machinery*, 9(1):11–12, 1962. [303]
- [Wat87] R. J. WATT : Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A : Optics, Image Science, and Vision*, 4(10):2006–2021, octobre 1987. [41]
- [Wat88] R. J. WATT : *Visual Processing : Computational, Psychological and Cognitive Research*. Lawrence Erlbaum Associates Ltd., London, 1988. [41]
- [Weh99] R. WEHNER : Large-scale navigation : The insect case. In *Spatial Information Theory – Cognitive and Computational Foundations of Geographic Information Science – International Conference COSIT'99*, pages 1–20, Stade, Germany, août 1999. [89]
- [Weh03] R. WEHNER : Desert ant navigation : How miniature brains solve complex tasks. *Journal of Comparative Physiology A : Sensory, Neural, and Behavioral Physiology*, 189(8):579–588, août 2003. [90]
- [Wei05] E. W. WEISSTEIN : Gamma function. <http://mathworld.wolfram.com/GammaFunction.html>, mai 2005. [102]

- [WH89] J. WANG et P. J. HOWARTH : Edge following as graph searching and hough transform algorithms for lineament detection. *In IGARSS'89, May 10-14*, volume 1, pages 93-96, Vancouver, B.C., Canada, 1989. [4]
- [Wha91] F. WHANG : Integrating GIS's and remote sensing image analysis systems by unifying knowledge representation schemes. *IEEE Transactions on Geoscience and Remote Sensing*, 29(4):656-664, juillet 1991. [248]
- [Wol94] J. M. WOLFE : Visual search in continuous, naturalistic stimuli. *Vision Research*, 34:1187-1195, 1994. [5, 70]
- [WTCC94] Ch. WU TZONG et Ch. CHEN CHANG : Application of geometric hashing to iconic database retrieval. *Pattern Recognition Letter*, 15(9):871-876, 1994. [80]
- [WZ00] J. WANG et Q. ZHANG : Application of gradient profile algorithm for road network extraction : Sensor resolution and background considerations. *Canadian Journal of Remote Sensing*, 26(5):428-439, 2000. [25]
- [Yar67] A. YARBUS : *Eye movements and vision*. Plenum, New York, 1967. [5, 6, 7, 55, 56, 57]
- [Yee87] B. YEE : An expert system for planimetric feature extraction. *In Proceedings of IGARSS'87 Symposium, May 18-21*, pages 321-325, Ann Arbor, Michigan, USA, 1987. [4, 26, 29, 37]
- [ZC93] A. ZLOTNICK et P. CARNINE : Note : Finding road seeds in aerial images. *Computer Vision, Graphics, and Image Processing : Image Understanding*, 57(2):243-260, 1993. [4]
- [ZC04] Q. ZHANG et I. COULOIGNER : Automatic road change detection and GIS updating from high spatial remotely-sensed imagery. *Geo-Spatial Information Science*, 7(2):89-95, 2004. [30, 249]

-
- [ZGY04] C. ZHANG, H.W. GUESGEN et W.-K. YEAP, éditeurs. *PRICAI 2004 : Trends in Artificial Intelligence, 8th Pacific Rim International Conference on Artificial Intelligence, Auckland, New Zealand, August 9-13, 2004, Proceedings*, volume 3157 de *Lecture Notes in Computer Science*. Springer, 2004. [282]
- [Zha00] Y. ZHANG : A method for continuous extraction of multispectrally classified urban rivers. *Photogrammetric Engineering and Remote Sensing*, 66(8):991–999, 2000. [4]
- [ZTB04a] N. ZLATOFF, B. TELLEZ et A. BASKURT : Image understanding and scene models : a generic framework integrating domain knowledge and Gestalt theory. *In Proceedings of International Conference on Image Processing (ICIP), October 24-27, 2004, Singapore*, 2004. [48]
- [ZTB04b] N. ZLATOFF, B. TELLEZ et A. BASKURT : Image understanding using domain knowledge. *In Proceedings of Recherche d'information Assistée par Ordinateur (RIAO), April 26-28 2004, Avignon, France*, 2004. [48]
- [ZTB04c] N. ZLATOFF, B. TELLEZ et A. BASKURT : Vision Gestalt et connaissance : une approche générique à l'interprétation d'images. *In Proceedings of COmpression et REprésentation des Signaux Audiovisuels (CORESA), 25-26 Mai 2004, Lille, France*, 2004. [48]

NB : Les dates associées aux entrées électroniques, comme [ISO05], correspondent à la date de consultation de la référence.

Glossaire

Arête (*edge*) Ensemble de pixels connectés entre eux qui suivent la frontière entre deux régions [GW02].

Attention visuelle Processus permettant de sélectionner une partie de l'information visuelle disponible pour une analyse plus détaillée et/ou ultérieure [FG03].

Champ central Voir *vision fovéale*.

Champ large Voir *vision périphérique*.

Champ visuel Aire globale qui est vue sans bouger ni la tête ni les yeux [RLB98].

Élément structurant (en morphologie mathématique) Masque binaire qui sert au filtrage morphologique. À proprement parler l'élément structurant est constitué des valeurs 1 du filtre [CC01] – le reste du filtre étant mis à 0.

Filtrage (en traitement d'image) Élimination de certaines fréquences spectrales ou spatiales afin d'accentuer des caractéristiques particulières sur l'image par passage d'un filtre [Cen05].

Filtre (en traitement d'image) Fenêtre mobile dont le pixel central est modifié selon le paramètre à filtrer. Le filtre est habituellement constitué d'un nombre impair de pixels (fenêtres de 3×3 , 5×5 ou 7×7 , *etc*). Plus le noyau est de grande dimension, plus l'effet de filtrage est important. Appelé aussi *noyau* [Cen05].

Fixation Période stationnaire entre les saccades [FG03].

Fovea ou Fovéa Petite dépression du centre de la rétine où l'acuité visuelle est maximale [RLB98].

Lignes de force d'une image Ossature des objets présents dans l'image [CC01].

Magnitude d'une saccade Amplitude de la saccade.

Mémoire à court terme Forme de mémoire qui ne persiste que quelques secondes, généralement, ou aussi longtemps que dure la répétition [RLB98].

Mouvement saccadique Voir *saccades*.

Neurone (en biologie) Unité fondamentale du système nerveux. Chaque neurone est composé d'un corps cellulaire, de prolongements récepteurs (dendrites) et d'un prolongement transmetteur (axone). Appelé aussi *cellule nerveuse* [RLB98].

Neurone (en informatique) Unité élémentaire de traitement des *réseaux de neurones* utilisée en intelligence artificielle, imitant les neurones du système nerveux humain et conçue pour simuler les méthodes de traitement d'information, d'apprentissage et de mémorisation propres au cerveau [Off05b].

Ontologie (en informatique) Ensemble d'informations dans lequel sont définis les concepts utilisés dans un langage donné, pour un domaine donné, et qui décrit les relations logiques qu'ils entretiennent entre eux [Off05b].

Pixel De *picture element*. Élément atomique constitutif d'une image numérique.

Saccades Série de mouvements rapides de l'œil qui se produisent régulièrement pendant la vision normale. Appelées aussi *mouvement saccadique* [RLB98].

Segmentation d'une image Partition d'une image en ses parties et/ou objets constituants [GW02].

Signature spectrale d'un objet Distribution des luminances de l'objet dans chacune des bandes concernées [CC01].

Visée (en télédétection) La terminologie RADAR définit les visées individuelles comme des paquets d'échantillons de signal dans un processeur RSO qui divisent l'ou-

verture synthétique totale en plusieurs sous-ouvertures, chacune représentant une visée individuelle de la même scène. L'image résultante est formée de la somme incohérente de ces visées et se caractérise par un chatoiement réduit et une résolution spatiale dégradée [Cen05].

Vision fovéale Vision centrale correspondant à une réception par la fovéa. Appelée également *région fovéale* ou *champ central*.

Vision périphérique Vision hors de la région fovéale. Appelée également *région périphérique* ou *champ large*.

Annexe A

Compléments algorithmiques

A.1 Algorithme de Douglas-Peucker

L'algorithme de Douglas-Peucker [DP73] permet de simplifier une ligne¹ en ôtant les points superflus pour une granularité donnée. Ainsi les points gardés ne sont que les points “essentiels”, figure 4.8 page 123.

L'algorithme de Douglas-Peucker fonctionne avec une tolérance donnée. Nous appellerons δ cette tolérance. L'hypothèse de départ est simple : toute ligne peut être approximée de la manière la plus grossière en éliminant tout point autre que les extrémités, figure A.1a. Ainsi la ligne $\{P_1, P_2, P_3, P_4, P_5, P_6, P_7\}$ est réduite en $\{P_1, P_7\}$.

Une fois que la ligne a été réduite ainsi à deux points, il est possible de considérer

¹Le terme *ligne* est employé ici selon la définition que nous en avons donnée dans le chapitre 3. Il est à noter que le terme *polyligne* est également utilisé dans la littérature ; dans ce cas une polyligne est composée de lignes, lignes qui ne sont alors composées que de deux points.

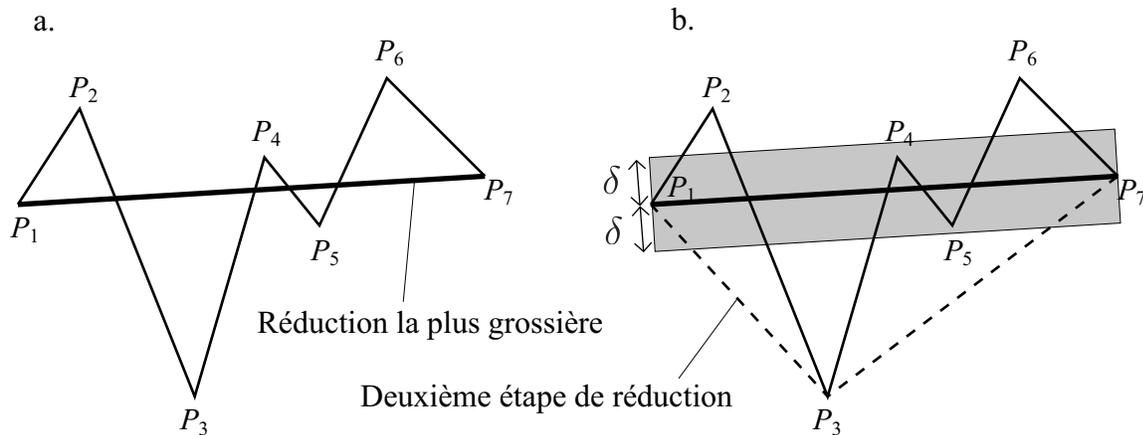


FIG. A.1 – Principe de base de l’algorithme de Douglas-Peucker.

la tolérance δ . δ est la distance minimum à la ligne pour laquelle les points en deçà de cette distance seront éliminés. Ainsi, sur la figure A.1b si la ligne originelle n’avait contenu que les points P_1 , P_4 , P_5 et P_7 , la ligne aurait été réduite seulement à $\{P_1, P_7\}$ car la distance du segment $[P_1, P_7]$ aux points P_4 et P_5 est inférieure à δ ; respectivement on note $d_{[P_1, P_7], P_4} < \delta$ et $d_{[P_1, P_7], P_5} < \delta$.

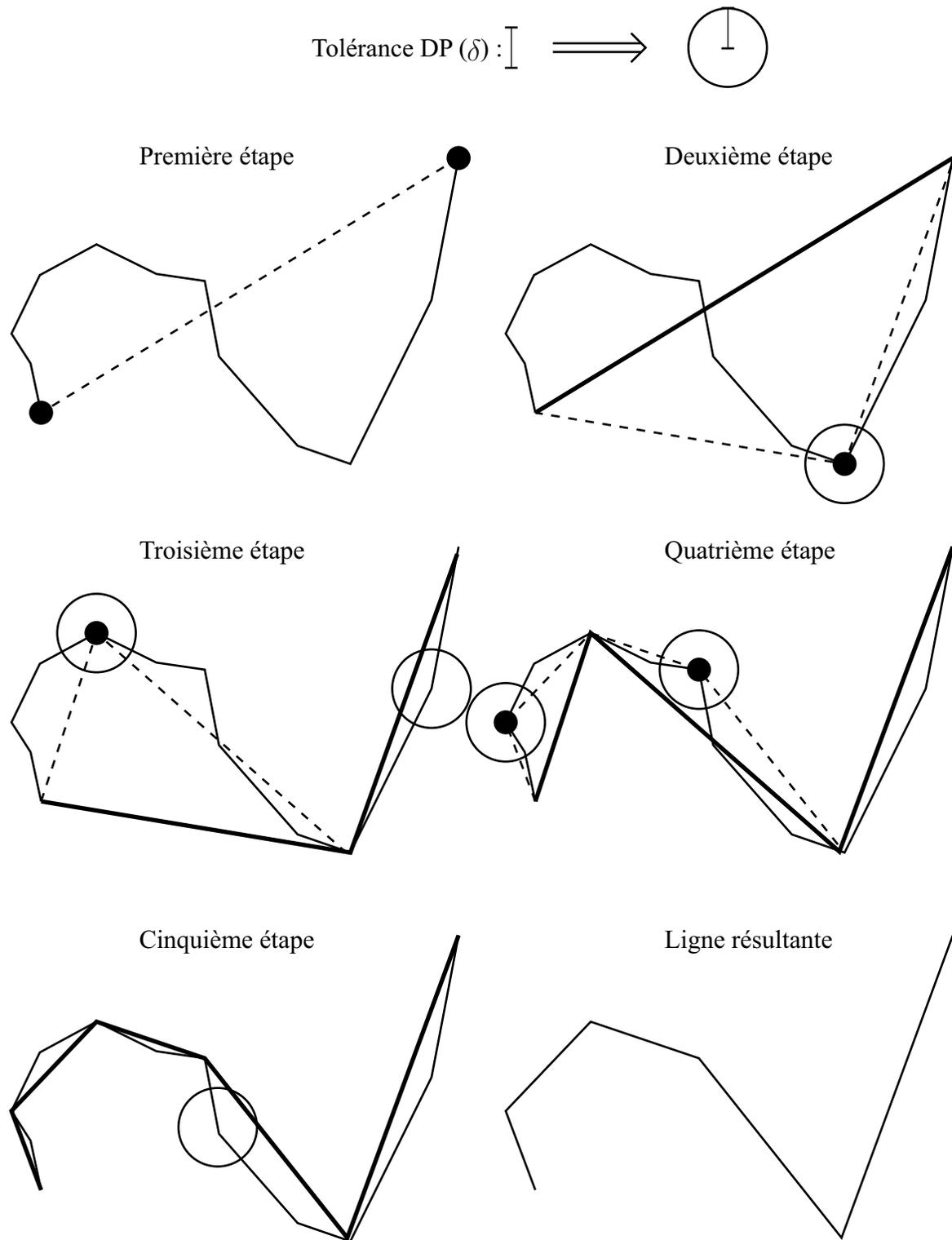
En revanche dans le cas de la ligne $\{P_1, P_2, P_3, P_4, P_5, P_6, P_7\}$ on a :

$$\left\{ \begin{array}{l} d_{[P_1, P_7], P_2} > \delta \\ d_{[P_1, P_7], P_3} > \delta \\ d_{[P_1, P_7], P_6} > \delta \end{array} \right. \quad (\text{A.1})$$

et

$$d_{[P_1, P_7], P_3} > d_{[P_1, P_7], P_6} > d_{[P_1, P_7], P_2} \quad (\text{A.2})$$

D’après le système A.1 on sait que les points P_2 , P_3 et P_6 sont à une distance du segment $[P_1, P_7]$ supérieure à la tolérance autorisée. Or P_3 est le plus éloigné des trois points de $[P_1, P_7]$ (inéquation A.2), ce sera ce point qui sera retenu. La ligne est donc réduite à cette étape à $\{P_1, P_3, P_7\}$, figure A.1b.



ALG. A.1 – Algorithme de Douglas-Peucker pas à pas.

L'étape suivante consiste à considérer les lignes $\{P_1, P_2, P_3\}$ et $\{P_3, P_4, P_5, P_6, P_7\}$ comme distinctes et appliquer le même principe que pour $\{P_1, P_2, P_3, P_4, P_5, P_6, P_7\}$ et ceci jusqu'à l'étape où il ne reste plus que des point en deçà de δ , cinquième étape de l'algorithme A.1 déroulé pas à pas.

A.2 Algorithme de fermeture transitive de la matrice d'un graphe

A.2.1 Fermeture transitive d'un graphe

A.2.1.1 Fermeture transitive (d'après [Rev05])

Étant donné un graphe orienté $G = (S, A)$,² on désire savoir s'il existe un chemin dans G de i vers j , $\forall(i, j)$. La *fermeture transitive* de G est définie par le graphe $G^* = (S, A^*)$ tel que :

$$A^* = \{(i, j) : \exists \text{ un chemin du sommet } i \text{ au sommet } j \text{ dans } G\} \quad (\text{A.3})$$

A.2.1.2 Illustration par l'exemple

Reprenons l'exemple de la figure 8.4 page 234. Le but final est d'obtenir deux éléments distincts ; ici un élément regroupant tous les pixels étiquetés a , b et d et un

² $G = (S, A)$, où S est un ensemble de sommets et A un ensemble d'arcs. A est une relation binaire sur S . Pour un arc (u, v) , on dit que l'arc part de u et arrive en v

autre avec les pixels étiquetés c .

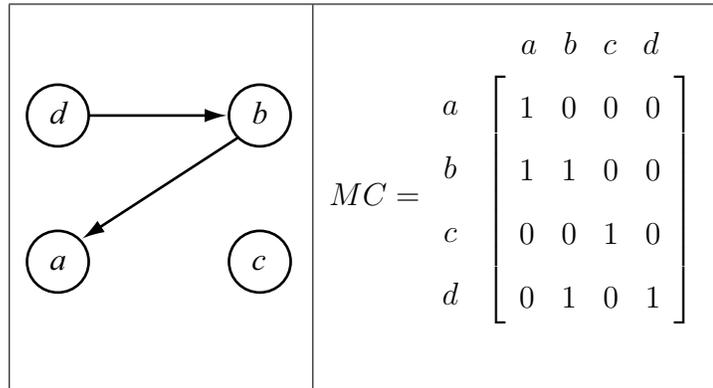


FIG. A.2 – Graphe et sa matrice associée.

Le graphe de la figure A.2 reflète la connectivité entre les pixels étiquetés lors du premier passage – voir chapitre 8. Les arcs (a, b) et (b, d) traduisent le fait que des connexions ont été identifiées respectivement entre les pixels étiquetés a et b et les pixels étiquetés b et d . La même figure présente la matrice de correspondance associée à ce graphe. La diagonale ne contient que des 1 puisqu’un pixel est “connecté à lui-même”. Les boucles³ associées à cette diagonale ne sont pas montrées sur le graphe pour ne pas alourdir la figure.

Étant donné que pour tout pixel p d’une image connecté à un pixel q de cette même image, le pixel q est connecté au pixel p , alors dans notre cas :

$$\forall (u, v) \in A \implies \exists (v, u) \quad (\text{A.4})$$

Ce qui se traduit dans la figure A.3 par les flèches pointillées. Du point de vue de l’écriture matricielle, cela implique que la matrice soit symétrique.

³Un arc qui part d’un sommet et arrive sur le même sommet est appelé une *boucle*.

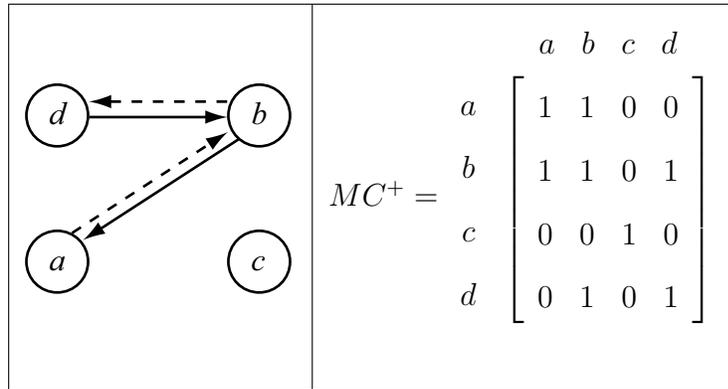


FIG. A.3 – Connexion réciproque des pixels.

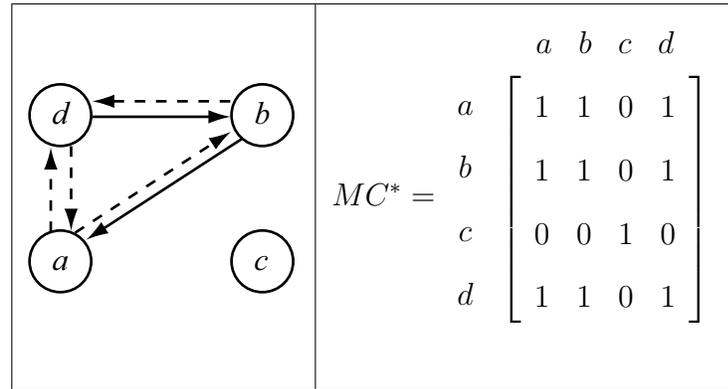


FIG. A.4 – Fermeture transitive du graphe.

Enfin, si l'on applique l'équation A.3 de fermeture transitive d'un graphe, on doit ajouter les arcs (a, d) et (d, a) car il existe des chemins du sommet a au sommet d dans MC^+ ainsi que du sommet d au sommet a . Respectivement ces chemins sont $\langle a, b, d \rangle$ et $\langle d, b, a \rangle$.

La figure A.4 illustre ceci du point de vue du graphe et de son écriture matricielle. Ainsi dans notre exemple, MC^* se lit la façon suivante :

Colonne a : Les étiquettes b et d sont connectées à a . L'étiquette c ne l'est pas.

Colonne b : Les étiquettes a et d sont connectées à b . L'étiquette c ne l'est pas.

Colonne c : Aucune étiquette n'est connectée à c .

Colonne d : Les étiquettes a et b sont connectées à d . L'étiquette c ne l'est pas.

On vérifie la cohérence de l'information fournie. En langage naturel, MC^* se formule comme suit : *“Les pixels étiquetés a , b et d forment un élément unique, distinct, sur l'imagette; les pixels étiquetés c forment un second élément distinct de l'imagette.”*

A.2.2 Algorithme de Roy-Warshall

Roy [Roy59] et Warshall [War62] ont traduit algorithmiquement l'équation A.3 de fermeture transitive d'un graphe. On construit la fermeture transitive G^* en effectuant sur l'ensemble des sommets s le traitement $\Phi_s(G)$ qui consiste à ajouter les arcs (y, z) tels que (y, x) et (x, z) existent. L'algorithme de Roy-Warshall est donc le suivant : algorithme A.2.

```

début
  |
  | pour tout  $s$  faire
  | | pour tout  $s'$  faire
  | | | si  $\exists(s, s')$ 
  | | | alors
  | | | | pour tout  $s''$  faire
  | | | | | si  $\exists(s, s'')$ 
  | | | | | alors
  | | | | | | ajouter l'arc  $(s', s'')$ ;
  | | | | | fsi
  | | | | fpour
  | | | fsi
  | | fpour
  | fpour
fin

```

ALG. A.2 – Algorithme de Roy-Warshall.

Annexe B

Exemple de spécifications du logiciel

Cette annexe présente un exemple du document technique de spécifications de VISION.

Cas de la classe Database

Nous prenons ici les spécifications établies pour la classe `Database`. Cette classe est abordée dans le chapitre 7 ; en particulier page 216.

Page 20 of 123	
Class Database	
<p>class Database</p> <p>This class implements the database used in VISION.</p> <p>It also computes methods in order to manipulate the database.</p> <p>The database is made by reading a text file and saved by writing into a text file.</p> <p>The database consists of a list of curvilinear element from the class DBCurvilinearElt</p> <p>Author: Gilles Cotteret</p> <p>Version: 1.0</p>	
Inner Class Summary	
protected final static class	Database._STACK_RECORD Stack record structure for DP algorithm
Field Summary	
protected list < DBCurvilinearElt >	_curvilinearEltList List of the curvilinear elements contained in the database
protected string	_dbfileName Original database file
protected LogManager	_logManager Log manager
public int	nAnchorIndex
public int	nFloaterIndex
public Database._STACK_RECORD	precPrev
Constructor Summary	
public	Database (LogManager logManager, const char * DBfilename) Constructor with log manager.
Destructor Summary	
public synchronized	~Database ()
Method Summary	
public void	add (DBCurvilinearElt curvilinearElt) This function adds a new curvilinear element into the database.
public void	changeScale (float scaleFactor) This function realizes the transformation of the database coordinates by a scale factor to fit the image and the database.
public void	del (CurvilinearElt curvilinearElt) This method is used to remove an element from the database.

FIG. B.1 – Spécifications pour la classe Database – 1/5.

Method Summary	
public void	dpsimplification (const double tolerance) This function launches the Douglas Peucker simplification on the database.
public Image	rasterize () This function makes the rasterization of the database into an image.
public void	shift (float x, float y) This function performs the transformation of the database coordinates by a shift to fit the image and the database.
public void	writeOut () This function writes to an ASCII file the content of the database.
protected void	_destroyDB () This function deallocates the memory used in the data structure of the database (attribute <code>_curvilinearEltList</code>).
protected void	_init () This function puts the database file into the data structure.
protected void	_reducePoints (double * pPointsX, double * pPointsY, int nPointsCount, int * pnUseFlag, double dTolerance) This function returns the points to delete in order to simplify a polyline by the Douglas Peucker algorithm.
protected bool	_stackPop (int * pnAnchorIndex, int * pnFloaterIndex) This function allows to pop anchor from the stack.
protected void	_stackPush (int nAnchorIndex, int nFloaterIndex) This function allows to push anchor in the stack.

Field Detail**_curvilinearEltList**

protected list <DBCurvilinearElt > _curvilinearEltList
List of the curvilinear elements contained in the database

_dbfileName

protected string _dbfileName
Original database file

_logManager

protected LogManager _logManager
Log manager

Constructor Detail**Database**

public Database(LogManager logManager, const char * DBFilename)
Constructor with log manager.

Parameters:
logManager Log manager
DBFileName Database file name

FIG. B.2 – Spécifications pour la classe Database – 2/5.

Page 22 of 123
<i>Destructor Detail</i>
~Database public synchronized ~Database()
<i>Method Detail</i>
add public void add(DBCurvilinearElt curvilinearElt) This function adds a new curvilinear element into the database. Presently work for a MIF file (Map Info). Parameters: curvilinearElt The curvilinear element to put in the database Postconditions: "curvilinearEltList" has one more element The database contains "curvilinearElt"
changeScale public void changeScale(float scaleFactor) This function performs the transformation of the database coordinates by a scale factor to fit the image and the database. Parameters: scaleFactor Scale factor to apply on the database Postconditions: Scale changed, "curvilinearEltList" updated
del public void del(CurvilinearElt curvilinearElt) This method is used to remove an element from the database. This function is used when an element has not been recognised during the recognizing mode. Parameters: curvilinearElt The curvilinear element to delete Preconditions: "curvilinearElt" is in "curvilinearEltList" Postconditions: "curvilinearEltList" has one element less "curvilinearElt" removed from "curvilinearEltList"

FIG. B.3 – Spécifications pour la classe Database – 3/5.

dPsimplification

```
public void dPsimplification(const double tolerance)
```

This function launches the Douglas Peucker simplification on the database.

Parameters:

tolerance Douglas Peucker tolerance in pixels

Preconditions:

tolerance ≥ 0

rasterize

```
public Image rasterize()
```

This function makes the rasterization of the database into an image.

The image produced by this function is a binary matrix representing the curvilinear elements of the original database.

Returns:

An image corresponding to the rasterization of the database.

shift

```
public void shift(float x, float y)
```

This function performs the transformation of the database coordinates by a shift to fit the image and the database.

Parameters:

x Abscissa coordinate of the shift vector

y ordinate coordinate of the shift vector

Postconditions:

The database is shifted

writeOut

```
public void writeOut()
```

This function writes to an ASCII file the content of the database. The function copies the data structure into the file named dbFileName+"Out".

This out file dbFileName+"Out" is readable by classical database software.

Returns:

A file named "out"+dbFileName" is written and contained the database

_destroyDB

```
protected void _destroyDB()
```

This function deallocates the memory used in the data structure of the database (attribute `_curvilinearEltList`).

Returns:

Memory is free

FIG. B.4 – Spécifications pour la classe Database – 4/5.

Page 24 of 123
<p><code>_init</code></p> <p><code>protected void _init()</code></p> <p>This function puts the database file into the data structure.</p> <p>Parameters: <code>fileName</code> Name of the file to read</p> <p>Returns: The data structure is loaded</p>
<p><code>_reducePoints</code></p> <p><code>protected void _reducePoints(double * pPointsX, double * pPointsY, int nPointsCount, int * pnUseFlag, double dTolerance)</code></p> <p>This function returns the points to delete in order to simplify a polyline by the Douglas Peucker algorithm.</p> <p>For more explanations see Cotteret, Ph.D. thesis 2005.</p> <p>Parameters: <code>pPointsX</code> Vector of the polyline X points <code>pPointsY</code> Vector of the polyline Y points <code>nPointsCount</code> Number of polyline points <code>pnUseFlag</code> Vector result of the function (0 = point to delete) <code>dTolerance</code> Tolerance of the algorithm (in pixels)</p> <p>Preconditions: <code>sizeof (nPointsX,nPointsY,pnUseFlag) = nPointsCount</code></p> <p><code>pnUseFlag</code> must be zeroed out before invoking the function</p> <p><code>dTolerance >= 0</code></p>
<p><code>_stackPop</code></p> <p><code>protected bool _stackPop(int * pnAnchorIndex, int * pnFloaterIndex)</code></p> <p>This function allows to pop the anchor from the stack.</p> <p>For more explanations see Cotteret, Ph.D. thesis 2005.</p> <p>Parameters: <code>pnAnchorIndex</code> Anchor index <code>pnFloaterIndex</code> Floater index</p> <p>Returns: True if exists</p>
<p><code>_stackPush</code></p> <p><code>protected void _stackPush(int nAnchorIndex, int nFloaterIndex)</code></p> <p>This function allows to push the anchor in the stack.</p> <p>For more explanations ssee Cotteret, Ph.D. thesis 2005.</p> <p>Parameters: <code>nAnchorIndex</code> Anchor index <code>nFloaterIndex</code> Floater index</p>

FIG. B.5 – Spécifications pour la classe Database – 5/5.

Annexe C

Compléments d'architecture

C.1 Extrait du fichier de paramètres XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!--Configuration file for vision software-->
<VisionConfig>
  <imageFile>
    E:\These\Implementation\Vision\Boite1314x1086.raw
  </imageFile>
  <imageWidth>
    1314
  </imageWidth>
  <imageHeight>
    1086
  </imageHeight>
  <databaseFile>
    Boite1314x1086.MIF
  </databaseFile>
  <module1>
    1
  </module1>
  </GazeJumpMagnitude>
  <GazeJumpAperture>
    1.5707963267948966192313216916398
  </GazeJumpAperture>
```

C.2 Interface de saisie des paramètres sous windows

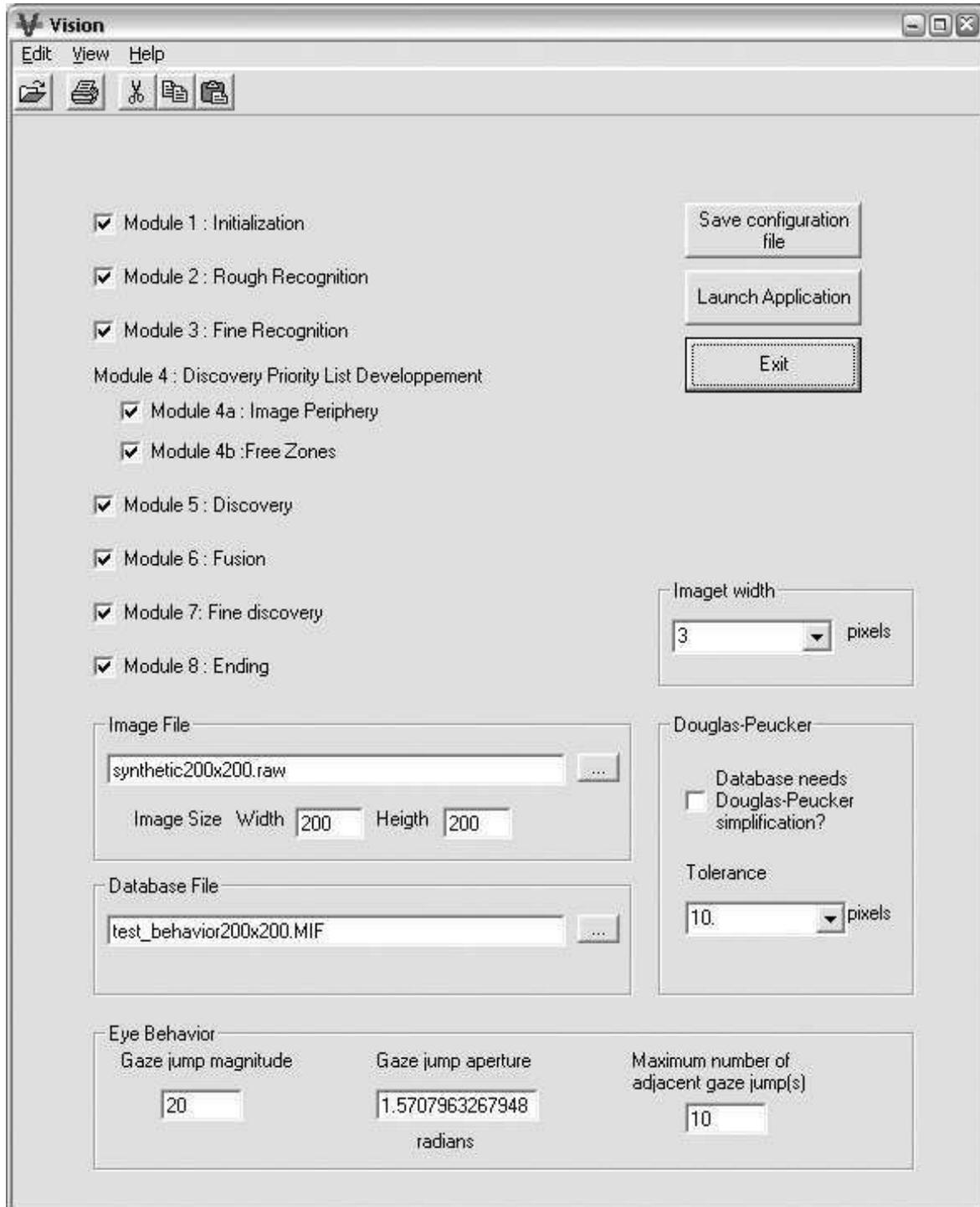


FIG. C.1 – Interface

C.3 Complément sur le diagramme d'activité

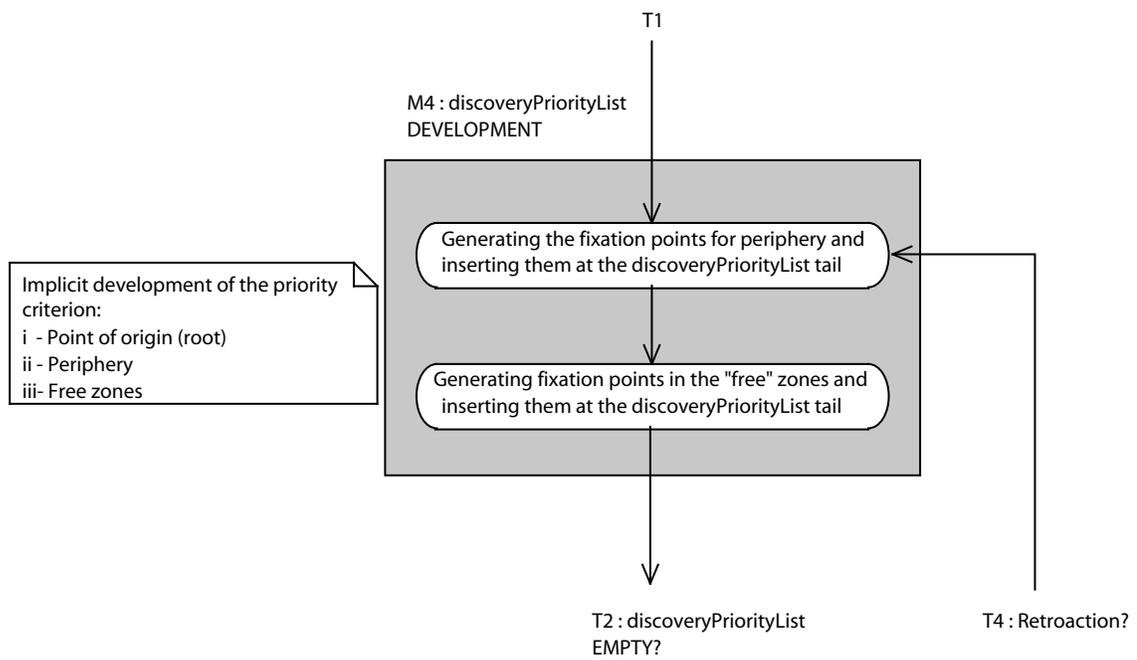


FIG. C.2 – Module M4 : Développement de la liste des priorités en mode découverte

Index

- A**
amnésie de la recherche visuelle 58–62
analyse
 en composantes connexes 153–156,
 174, 232–238
 en composantes principales . 65, 72, 73
 par traits 49
appariement à un gabarit 49
ArcInfo 241, 242
ascendant 6, 39, 53–71, 116, 131, 188, 251,
 260, 261
attention visuelle 5, 7, 11, 13, 18, 52–74,
 114
 sélective 53–62
 spatiale 53
AutoCAD 241
- B**
balayage attentionnel 206
bottom-up voir ascendant
- C**
carte de saillance voir saillance
corridors de recherche . . 116–126, 134, 204,
 224
- D**
découverte fine 210
descendant 6, 7, 39, 47, 67–74, 116, 130,
 188, 251, 260
dilatation 31
dipôle 168–183, 207, 216, 230
- E**
échantillonnage . . voir sous échantillonnage
érosion 31
- F**
fermeture 32
fermeture transitive 234, 236, 300–303
filtre 293
- de Lee 146, 147, 150, 156, 232
de Robinson 149, 150, 156, 232
- G**
Geomatica 241, 242
Gestalt 43–48, 63, 256
 bonne continuation 43
 destin commun 43
 environnement 43
 fermeture 43
 orientation 43
 proximité 43
 similitude 43
 symétrie 43
 taille relative 43
groupement perceptuel . . 4, 48, 52, 63, 256
- H**
hélicoptère 8, 9, 11, 14, 94, 99, 193, 239
- I**
Ikonos 27, 28, 34, 36
inhibition de retour 60–62, 67, 69, 71, 129,
 132, 230
- L**
LANDSAT 24, 25, 28, 36, 96
Lévy flights 64
liste de priorités 120–130, 132, 133,
 135–137, 155, 196, 200, 202, 203,
 205–207, 229, 230, 261, 313
- M**
Marr 5, 38–52
 Ébauche 2,5D 39
 Ébauche primitive 39–47
 modèle 3D 40
 rassemblement curvilinéaire 47
 rassemblement thêta 47
 regroupement 47
matrice de correspondance . . 234, 236, 301

- mécanismes attentionnels ... voir attention visuelle
- mode
 découverte 120, 124–128, 130, 138, 139, 156, 159, 161, 163, 175, 182, 196, 203, 205, 206, 216, 220, 222, 229, 252, 313
 reconnaissance 119, 120, 122, 125–128, 130, 138, 139, 163, 196, 198–200, 207, 216, 220, 222, 223, 226, 229
- Modèle
 ascendant de Koch et Ullman 66
 CLE 170–183, 207, 208, 216, 252
 complet voir système complet de Machrouh
 d'intégration des traits .. 57–58, 64–66
 LE 167–183, 253
- multirésolution 13, 30, 34–35, 74–80
- O**
- ondelettes 35, 78–80
- ouverture 32
- P**
- panorama 165, 177–181, 207, 208
- perception visuelle .. 4, 38, 52, 57, 67, 162, 165
- Q**
- QuickBird 12, 28, 34, 98
- R**
- RADAR 26, 94, 98–103, 129, 146, 147, 257, 294
- RADARSAT 99, 100
- raisonnement spatial qualitatif 21, 83–90, 159–183, 208, 253, 256, 258–260, 264, 266
- reconnaissance
 fine 201, 204
 grossière 200
- relation spatiale ... 28, 37, 77, 82, 159–183, 208
- S**
- sélection attentionnelle voir attention sélective
- saillance 6, 58, 64, 66–74, 116–131, 135, 137, 199, 200, 213, 219, 251, 252, 260
- seuillage 146, 152–153, 236
- signature spectrale .. 34, 35, 104, 118, 129, 144, 176, 179, 208, 294
- sous échantillonnage . 13, 74, 142, 253, 256
- SPOT 24, 25, 28, 36, 107
- squelettisation 32
- système complet de Machrouh 72–74
- T**
- top-down* voir descendant
- U**
- UML 18, 19, 187, 192, 211, 215, 222
- V**
- vision
 augmentée 9
 naturelle 38, 55, 74, 256
 numérique .. 4, 34, 38, 39, 48, 249, 256
 synthétique 9
 système de 52, 69, 113, 249
- voisinage 173–182
- Y**
- Yarbus 7, 55, 57