



Université PARIS DESCARTES  
Centre universitaire des Saints-Pères  
**UFR DE MATHÉMATIQUES ET INFORMATIQUE**

Thèse présentée pour l'obtention du grade de Docteur  
de l'université PARIS DESCARTES  
Spécialité : **Informatique**

Sujet de thèse :

# **Reconnaissance d'objets cartographiques dans les images satellitaires à haute résolution**

**Güray ERUS**

Soutenue le 12 juin 2008, devant le jury composé de :

---

Pr. Nicole Vincent, Univ. Paris Descartes	Directeur de thèse
Dr. Florence Tupin, ENST, Paris	Rapporteur
Pr. Alain Faure, Univ. du Havre	Rapporteur
Pr. Volkan Atalay, METU, Ankara	Examineur
Dr. Nicolas Paparoditis, IGN, Paris	Examineur
Dr. Nicolas Loménie, Univ. Paris Descartes	Examineur
Pr. Georges Stamon, Univ. Paris Descartes	Examineur

---



*à Selma ...*



# Remerciements

Je tiens spécialement à remercier très chaleureusement Prof. George Stamon qui m'a accueilli au sein de son équipe et qui a tout mis en oeuvre pour que ma thèse puisse continuer dans de bonnes conditions.

Je remercie Prof. Nicole Vincent qui m'a accordé sa confiance et a accepté d'être ma directrice de thèse.

Toutes mes expressions de gratitude vont à Nicolas Loménie qui a encadré ma thèse et m'a assuré de son soutien et de son assistance sans faille pendant ces quatre années. Je le remercie également pour sa grande disponibilité, ses encouragements et ses conseils.

Mes remerciements vont également à Florence Tupin et à Alain Faure pour l'intérêt qu'ils ont porté à mon travail et pour avoir accepté d'en être les rapporteurs. Je voudrais remercier Nicolas Paparoditis pour avoir accepté de faire partie de mon jury.

Merci à Prof. Nese Yalabik et Prof. Volkan Atalay qui m'ont permis d'effectuer une partie de mes travaux de recherche en coopération scientifique avec l'université METU à Ankara.

Je témoigne mes remerciements au gouvernement français et spécialement à M. Zouheir Hamrouni au service culturel à l'ambassade de France à Ankara pour le soutien financier accordé à mes travaux au début de ma thèse.

Je remercie Jordi Inglada qui a proposé la problématique principale de cette thèse dans le cadre d'une coopération entre le CNES et le laboratoire SIP-CRIP5, et qui a fourni les bases d'images satellitaires.

Je tiens également à remercier ici l'ensemble de l'équipe SIP parmi laquelle j'ai eu la chance de travailler dans un cadre agréable. Merci spécialement à Benoit et à Franck avec qui j'ai partagé les moments difficiles ainsi que les bons moments.

Enfin, je remercie Selma, son soutien et sa présence étaient indispensables pendant ces quatre années.



# Résumé

L'imagerie satellitaire présente de nombreux avantages par rapport aux autres sources d'observation du sol. Les satellites fournissent un très grand nombre d'images de la Terre utilisées dans des domaines très variés avec des forts enjeux scientifiques, économiques et stratégiques. Le Centre National d'Etudes Spatiales (CNES) se prépare à la mise en orbite de satellites équipés de capteurs optiques de nouvelle génération avec une résolution spatiale optimale de l'ordre de 0,7 mètres par pixel. Cette avancée technologique nécessite un important travail de préparation pour pouvoir exploiter les futures images satellitaires à très haute résolution et en même temps **ouvre la voie à l'utilisation des méthodes qui sont très peu abordées jusqu'à présent en imagerie satellitaire.**

L'extraction manuelle d'objets cartographiques est une tâche à la fois difficile et longue, vu la quantité et les dimensions importantes des images fournies par les satellites. Le CNES a lancé plusieurs groupes méthodologiques dont un projet de recherche sur la reconnaissance automatique d'objets cartographiques dans les images satellitaires. Une base d'images SPOT5 contenant différentes catégories d'objets a été préparée pour tester et évaluer les méthodes développées. Notre thèse s'inscrit dans le cadre de ce projet et porte spécifiquement sur **la reconnaissance d'objets cartographiques à forte structure prononcée : les ronds-points et les ponts.**

Nous proposons d'explorer l'utilisation des méthodes récentes de reconnaissance d'objets qui suivent une approche de haut niveau prenant en compte la forme de l'objet et les relations spatiales de ses composants, au lieu de rester à l'échelle pixel et signal pour comparer les images selon la couleur ou la texture. Sur les images satellitaires à basse résolution les objets disparaissent et font partie de la texture. Par contre, en haute résolution les caractéristiques discriminantes géométriques et structurelles des objets apparaissent et un traitement à un haut niveau d'abstraction devient possible et même nécessaire pour la reconnaissance.

Dans les dernières années on observe un intérêt croissant pour la reconnaissance d'objets dans les images, et particulièrement **la reconnaissance générique d'objets**, où l'objectif est de reconnaître les objets appartenant à des catégories d'objet au lieu de rechercher les instances d'un objet spécifique. Notre problème fait donc partie de la reconnaissance générique d'objets. Les objets cartographiques sont groupés dans des catégories qui représentent des concepts sémantiques généraux. La principale difficulté de notre problème réside dans la variance intra-classe très importante des objets-types qui sont souvent constitués de composants très similaires mais articulés suivant des organisations spatiales différentes.

Etant des constructions humaines, les objets cartographiques ont des caractéristiques géométriques régulières qui les distinguent des autres objets. Nous proposons d'exploiter principalement ces caractéristiques pour obtenir une représentation de leur structure inhérente.

Dans un premier temps, nous avons développé une méthode de génération de **modèle explicite d'objet structurel** représenté par des graphes relationnels attribués (*ARGs*) à partir des images segmentées par un expert. Les sommets des graphes correspondent aux primitives géométriques simples obtenues à partir de l'image et les arêtes aux relations spatiales entre les

primitives. On apprend le modèle en utilisant des algorithmes d'appariement de graphes. A la fin de cette étape préliminaire, nous avons réussi à générer des modèles d'objets pour les catégories d'objets ronds-points et ponts. Pourtant ces modèles sont trop rigides pour être utilisés dans le cadre de la reconnaissance d'objets dans une application réelle à cause de la difficulté de sélection des primitives appartenant à l'objet dans les images non-segmentées. En revanche cette étape aura montré la faisabilité d'une approche structurale dans l'absolu à ce niveau de résolution.

En vue de développer un système opérationnel de reconnaissance, nous avons proposé, dans un deuxième temps, l'apprentissage d'un modèle plus souple fondé principalement sur **l'apparence des composants locaux de l'objet** par une méthode statistique d'apprentissage. On apprend **un modèle implicite d'objet** par la fusion de classifieurs faibles obtenus à partir des primitives géométriques en utilisant l'algorithme *Adaboost*. Les primitives sont extraites par une approche par contours et une approche par régions en parallèle.

Nous avons également proposé une méthode de reconnaissance par un modèle implicite construit par une approche **composants et structure**. On représente les composants par des primitives géométriques. La structure est représentée par les propriétés géométriques et l'organisation spatiale des primitives. Le modèle est appris par l'algorithme de regroupement (*clustering*) *Mean-Shift*.

Enfin, les méthodes de reconnaissance sont validées sur un ensemble d'images réelles fournies par le CNES dans le cadre d'une compétition nationale et d'une application cartographique.

**Mots clés :** Imagerie satellitaire à haute résolution, objets cartographiques, modèle d'objet structurel, modèle d'objet implicite, apprentissage de modèle, classification

# Abstract

Satellite imagery has many advantages compared to other sources of land observation. Satellites provide an important number of images of the Earth used in various domains with important scientific, economical and strategic concerns. The French National Space Agency (CNES) is preparing to launch new satellites equipped with new generation optical components with an optimal spatial resolution of 0.7 meters per pixel. This technological advance requires an important work of preparation in order to exploit future very high resolution satellite images and at the same time **opens the way to the use of methods seldom addressed to date in satellite imagery.**

Manual extraction of cartographic objects is a process both hard and time consuming when one considers the quantity and the dimensions of images provided by satellites. CNES has founded several methodological work groups and has started a research project on automatic recognition of cartographic objects on satellite images. An image-base of SPOT5 images containing different object categories is prepared in order to test and evaluate developed methods. Our thesis' subject is determined in the frame of this project and considers specifically **the recognition of cartographic objects with a highly composite structure : the roundabouts and the bridges.**

We propose to explore the use of recent methods of object recognition that follow a high-level approach by taking into consideration the object's shape and the spatial relations of its parts, instead of working on pixel and signal level and comparing images based on their color and texture. On low-resolution satellite images, objects disappear and become part of the texture. On the other hand, on high-resolution images, geometrical and structural discriminant features of objects become visible and processing them using a high level of abstraction is possible and even necessary for the recognition.

In recent years we witnessed a growing interest on object recognition in images and particularly on **generic object recognition**, where the aim is to recognize objects belonging to object categories instead of searching for instances of a specific object. Our problem belongs to the generic object recognition. Cartographic objects are grouped in categories that correspond to general semantic concepts. The main difficulty of our problem is the important intraclass variance of the objects that are in most of the time composed of similar parts articulated according to different spatial organizations.

Being man-made constructions, cartographic objects have regular geometrical features that distinguish them from other objects. We propose to exploit principally these features in order to obtain a representation of their inherent structure.

A method to generate **an explicit structural object model** represented by Attributed Relational Graphs (ARGs) from images segmented by an expert is first developed. The vertices of the graphs correspond to geometrical primitives extracted from the image and the edges to the spatial relations between the primitives. The model is learned using graph matching algorithms. At the end of this preliminary stage we succeeded to generate object models for the roundabout

and bridge categories. However these models are too rigid to be used for the recognition on a practical application because of the difficulty of selecting the primitives belonging to the object on non-segmented images. On the other hand this step showed the feasibility of a structural approach in this level of resolution.

With the aim of developing an operational recognition system, we then proposed to learn a more flexible model based on **the appearances of the local parts of the objects** using a statistical learning method. **An implicit object model** is learned by the fusion of weak classifiers obtained from geometrical primitives using the Adaboost algorithm. The primitives are obtained by an edge based and a region based approach in parallel.

An object recognition method using an implicit model constructed by a **parts and structures** approach is then proposed. The parts are represented by geometrical primitives. The structure is represented by the shape properties of the primitives and their spatial organizations. The model is learned using the Mean-Shift clustering algorithm.

Finally, the methods are validated on satellite images provided by the CNES in the frame of a national competition and a cartographic application.

**Keywords :** High-resolution satellite imaging, cartographic objects, structural object model, implicit object model, model learning, classification

# Table des matières

<b>I</b>	<b>Introduction et état de l’art</b>	<b>1</b>
<b>1</b>	<b>Introduction et motivations</b>	<b>3</b>
1.1	Le programme Pléiades . . . . .	5
1.2	Le programme d’accompagnement ORFEO . . . . .	6
1.3	Les objets cartographiques à forte structure prononcée . . . . .	7
1.3.1	La base d’images d’objets cartographiques du CNES . . . . .	8
1.3.2	Le projet <i>ROBIN</i> . . . . .	10
1.4	Le contexte général de notre recherche . . . . .	11
1.5	L’organisation de la thèse . . . . .	12
<b>2</b>	<b>La reconnaissance d’objets dans les images</b>	<b>15</b>
2.1	Introduction . . . . .	15
2.2	Les principaux domaines d’application . . . . .	15
2.3	La reconnaissance spécifique versus générique . . . . .	17
2.4	Les défis majeurs de la reconnaissance d’objets . . . . .	17
2.5	La vision humaine . . . . .	18
2.5.1	<i>Particulate feature theory</i> . . . . .	19
2.5.2	La théorie de la reconnaissance par composants . . . . .	19
2.6	Le schéma général des méthodes de reconnaissance d’objets . . . . .	21
2.7	Les traitements de bas niveau : extraction des descripteurs . . . . .	22
2.7.1	Les points d’intérêt . . . . .	23
2.7.1.1	Détection des coins . . . . .	23
2.7.1.2	Les descripteurs <i>SIFT</i> . . . . .	23
2.7.1.3	<i>Les blobs</i> . . . . .	25
2.7.1.4	<i>Maximally Stable Extremum Regions (MSER)</i> . . . . .	25
2.7.2	Les descripteurs contours . . . . .	25
2.7.2.1	Détection des contours . . . . .	26
2.7.2.2	Extraction des descripteurs contours . . . . .	27
2.7.2.2.a	Les fragments de contours . . . . .	28
2.7.2.2.b	Les primitives géométriques . . . . .	29
2.7.3	Les descripteurs régions . . . . .	30
2.7.3.1	La segmentation de l’image . . . . .	30
2.7.3.2	La segmentation multi-échelle . . . . .	31
2.7.3.3	Les descripteurs de formes . . . . .	32
2.7.3.3.a	Les descripteurs obtenus par une approche région . . . . .	32
2.7.3.3.b	Les descripteurs obtenus par une approche contours . . . . .	34
2.7.4	Les descripteurs de texture et de couleur . . . . .	35

2.8	Les traitements de haut niveau : reconnaissance d'objets . . . . .	36
2.8.1	L'apprentissage structurel . . . . .	36
2.8.2	L'approche fondée sur l'apparence . . . . .	37
2.8.3	Les algorithmes d'apprentissage . . . . .	38
2.8.4	La représentation de l'objet . . . . .	40
2.8.4.1	Les modèles par l'apparence seule . . . . .	40
2.8.4.2	Les modèles par l'apparence et la structure . . . . .	42
2.8.4.2.a	<i>Pictorial Structures</i> . . . . .	42
2.8.4.2.b	Le modèle de constellation . . . . .	43
2.9	La reconnaissance d'objets cartographiques . . . . .	46
2.10	Notre approche . . . . .	49
 <b>II Génération de modèle d'objet explicite</b>		<b>51</b>
 <b>3 Modèles explicites d'objets cartographiques</b>		<b>53</b>
3.1	Introduction . . . . .	53
3.1.1	La position du problème et la solution proposée . . . . .	53
3.1.2	La base d'images d'apprentissage . . . . .	54
3.2	Notre approche . . . . .	54
3.3	Les traitements de bas niveau : extraction des primitives . . . . .	56
3.3.1	Extraction des primitives "cercles" . . . . .	56
3.3.2	Extraction des primitives "rectangles" . . . . .	57
3.3.2.1	Approximation polygonale . . . . .	57
3.3.2.2	Sélection des segments parallèles et des rectangles . . . . .	58
3.3.2.3	Détection des connections entre les composants de l'objet . . . . .	58
3.3.3	Extraction des primitives "segments de droites" . . . . .	59
3.3.3.1	L'algorithme d'amincissement ( <i>thinning</i> ) . . . . .	59
3.3.3.2	Détection des segments . . . . .	60
3.4	Les traitements de haut niveau : construction du modèle . . . . .	60
3.4.1	Définitions principales . . . . .	61
3.4.2	Construction des <i>ARGs</i> . . . . .	62
3.4.3	Détection des prototypes . . . . .	63
3.4.4	Construction du modèle . . . . .	64
3.4.5	Le sous-graphe commun maximal des prototypes . . . . .	64
3.4.6	Le sur-graphe commun minimal des prototypes . . . . .	65
3.4.7	Distance d'édition des exemples au modèle . . . . .	65
3.5	Les résultats expérimentaux . . . . .	66
3.6	Conclusion . . . . .	67
 <b>III Modèles implicites d'objets fondés sur l'apparence</b>		<b>69</b>
 <b>4 Extraction des primitives</b>		<b>71</b>
4.1	Introduction . . . . .	71
4.2	Extraction par régions . . . . .	72
4.2.1	Segmentation par lignes de partage des eaux ( <i>Watershed</i> ) . . . . .	72
4.2.2	<i>Watershed</i> avec marqueurs . . . . .	73
4.2.3	Segmentation par <i>Mean-Shift</i> . . . . .	75

4.2.4	Fusion des régions . . . . .	75
4.2.5	Les résultats de l'extraction . . . . .	76
4.3	Extraction par contours . . . . .	77
4.3.1	Les étapes de l'extraction . . . . .	78
4.3.2	Détection des contours sub-pixelliques . . . . .	78
4.3.3	Chaînage des contours . . . . .	80
4.3.4	Détection des <i>blobs</i> . . . . .	80
4.3.5	Détection des segments de droites . . . . .	82
4.3.6	Détection des arcs de cercles . . . . .	82
4.4	Conclusion . . . . .	85
<b>5</b>	<b>La modélisation par la fusion des classifieurs faibles</b>	<b>87</b>
5.1	Introduction . . . . .	87
5.2	Classifieur <i>Adaboost</i> . . . . .	88
5.3	Construction du vecteur de caractéristiques . . . . .	90
5.4	Résultats expérimentaux . . . . .	93
5.5	Conclusions . . . . .	95
<b>6</b>	<b>Le modèle “composants et structure”</b>	<b>99</b>
6.1	Introduction . . . . .	99
6.2	Le schéma général . . . . .	100
6.3	Représentation des primitives . . . . .	101
6.4	<i>Clustering</i> . . . . .	102
6.5	La sélection des <i>clusters</i> . . . . .	103
6.6	La classification . . . . .	106
6.7	Résultats expérimentaux . . . . .	108
6.8	Conclusion . . . . .	114
<b>IV</b>	<b>Applications, évaluations, conclusions et perspectives</b>	<b>117</b>
<b>7</b>	<b>Applications et Evaluations</b>	<b>119</b>
7.1	Introduction . . . . .	119
7.2	La compétition <i>ROBIN</i> . . . . .	119
7.3	La détection des ronds-points . . . . .	123
7.4	Conclusion . . . . .	126
<b>8</b>	<b>Conclusion et perspectives</b>	<b>129</b>



# Table des figures

1.1	Exemples d'images satellitaires SPOT5. . . . .	4
1.2	Image satellitaire SPOT5 et la simulation d'image satellitaire Pléiades. . . . .	5
1.3	Imagettes SPOT5 d'objets cartographiques à forte structure prononcée, un rond-point et un pont. . . . .	9
1.4	Imagettes panchromatiques SPOT5 d'objets cartographiques. . . . .	9
2.1	Les géons. (a). Exemples de géons, (b). Quelques objets composés de géons. . . . .	20
2.2	Les groupes formés par les propriétés non-accidentelles. . . . .	20
2.3	Les coins détectés sur une image satellitaire par le détecteur de Harris et Stephens [1988]. . . . .	23
2.4	Les points d'intérêt et les descripteurs locaux, dans Schmid et Mohr [1997]. . . . .	24
2.5	SIFT, calcul de la signature d'un point d'intérêt. . . . .	24
2.6	Les fragments de contours de quelques objets courants, dans Shotton <i>et al.</i> [2007].	26
2.7	Les 35 <i>kAS</i> les plus fréquents du dictionnaire de codes, dans Ferrari <i>et al.</i> [2008].	28
2.8	Les segments linéaires par la détection des événements significatifs, dans Desolneux <i>et al.</i> [2000]. . . . .	30
2.9	Illustration de l'algorithme <i>Watershed</i> sur un signal 1D. . . . .	32
2.10	Les images d'apprentissage artificielles créées par une grammaire, dans Cordella <i>et al.</i> [2002]. . . . .	37
2.11	La conception d'objet dans les approches basées sur l'apparence. . . . .	38
2.12	Illustration du modèle Sac de Mots. . . . .	41
2.13	Les 56 premiers mots du sac de mots obtenu par un quadrillage régulier, dans Li et Perona [2005]. . . . .	41
2.14	Les structures de graphes. . . . .	43
2.15	La représentation d'un visage par le modèle <i>Pictorial Structures</i> . . . . .	43
2.16	Illustration de la méthode de Leibe <i>et al.</i> [2004] pour la détection des voitures. . . . .	44
2.17	Illustration du modèle hiérarchique de Bouchard et Triggs [2005]. . . . .	45
2.18	Illustration des approches par contours. . . . .	46
2.19	Les descripteurs radiaux et angulaires, dans Bordes [1997]. . . . .	47
3.1	Schéma général de génération de modèle structurel. . . . .	55
3.2	Détection de cercles sur une image de rond-point par la transformée de Hough. . . . .	57
3.3	Illustration de l'algorithme de la corde. . . . .	57
3.4	Les segments détectés par l'algorithme de la corde. . . . .	58
3.5	Décomposition d'images de ronds-points par des rectangles et des cercles. . . . .	59
3.6	Décomposition d'images de ponts par des rectangles. . . . .	59
3.7	Extraction des primitives segments de droites à partir du squelette d'objet sur une image de pont. . . . .	60

3.8	Décomposition d'images de ronds-points et de ponts par des cercles et des segments de droites. . . . .	61
3.9	Représentation d'un pont par un <i>ARG</i> . . . . .	63
3.10	Les prototypes obtenus. . . . .	66
3.11	Les modèles obtenus. . . . .	67
4.1	Exemples d'imagettes et leur segmentation réalisée manuellement par un expert. . . . .	72
4.2	Segmentation par <i>Watershed</i> sur une image de rond-point et la fusion des régions. . . . .	73
4.3	Image d'un rond-point. (a). image panchromatique, (b). image multispectrale. . . . .	74
4.4	Les marqueurs obtenus sur une image de rond-point. . . . .	74
4.5	Les objets détectés par <i>Watershed</i> avec marqueurs. . . . .	75
4.6	Segmentation par <i>Mean-Shift</i> d'une image de rond-point. . . . .	76
4.7	Les régions sélectionnées après la segmentation initiale. . . . .	77
4.8	L'ensemble des primitives obtenues par l'approche par régions. . . . .	77
4.9	Le pyramide gaussien d'une image de rond-point. . . . .	78
4.10	Illustration de la détection du point de contour subpixellique <i>M</i> . . . . .	79
4.11	Les contours obtenus par un détecteur de Canny et les contours subpixelliques. . . . .	80
4.12	Les chaînes de contours après la fusion des points de contours. . . . .	81
4.13	Les <i>blobs</i> détectés par l'approche par contours. . . . .	82
4.14	Les segments détectés par l'approche par contours. . . . .	83
4.15	Les arcs de cercles détectés par l'approche par contours. . . . .	85
4.16	L'ensemble des primitives obtenues par l'approche par contours. . . . .	86
5.1	Les 3 motifs rectangulaires utilisés dans [Viola et Jones, 2004]. . . . .	89
5.2	Les deux premiers classifieurs faibles sélectionnés par l'algorithme <i>Adaboost</i> pour détecter les visages, dans Viola et Jones [2004]. . . . .	90
5.3	Les primitives composés. . . . .	91
5.4	Les courbes <i>ROC</i> de la classification par le $\mathcal{C}_{\mathcal{RP}}$ et le $\mathcal{C}_{\mathcal{PT}}$ . . . . .	94
5.5	Les 4 images de ronds-points et les 8 images de ponts qui ont obtenu les scores les plus bas pendant la classification. . . . .	96
6.1	Le vecteur de caractéristiques. . . . .	102
6.2	Illustration du <i>clustering Mean-Shift</i> sur les données bi-dimensionnelles . . . . .	103
6.3	Le <i>clustering</i> des primitives pour les ronds-points. . . . .	104
6.4	Le <i>clustering</i> des primitives pour les ponts. . . . .	105
6.5	La fonction sigmoïde et la fonction d'activation. . . . .	107
6.6	Les courbes <i>ROC</i> de la classification par le $\mathcal{C}_{\mathcal{RP}}$ et le $\mathcal{C}_{\mathcal{PT}}$ . . . . .	109
6.7	Les 5 images de ronds-points avec les scores les plus élevés d'après le score $s_{RP}$ . . . . .	110
6.8	Les 5 images de ronds-points avec les scores les plus bas d'après le score $s_{RP}$ . . . . .	111
6.9	Les 5 images de ponts avec les scores les plus élevés d'après le score $s_{PT}$ . . . . .	112
6.10	Les 5 images de ponts avec les scores les plus bas d'après le score $s_{PT}$ . . . . .	113
6.11	La correspondance entre la segmentation obtenu par l'approche fondée sur l'apparence et le modèle généré par l'approche structurelle. . . . .	115
7.1	Les résultats de la classification. . . . .	121
7.2	Exemples d'erreurs de classification. . . . .	122
7.3	Les ronds-points détectés par un expert sur l'image satellitaire utilisée pour le test. . . . .	124
7.4	Les arcs de cercles détectés sur l'image satellitaire utilisée pour le test. . . . .	125
7.5	Les ronds-points détectés sur l'image satellitaire utilisée pour le test. . . . .	127

# Liste des tableaux

5.1	Les résultats de la classification par la fusion des classifieurs faibles. . . . .	93
5.2	Le nombre d'objets de chaque catégorie dans les 36 premières images obtenues par le $\mathcal{C}_{\mathcal{RP}}$ . . . . .	95
5.3	Le nombre d'objets de chaque catégorie dans les 49 premières images obtenues par le $\mathcal{C}_{\mathcal{PT}}$ . . . . .	95
6.1	Les moyennes (les centres et les écart-types des caractéristiques des <i>clusters</i> sélectionnés). . . . .	106
6.2	Les résultats de la classification par l'approche composants et structure. . . . .	108
6.3	Le nombre d'exemples positifs dans les $n$ premières images d'après les scores triés obtenues par le $\mathcal{C}_{\mathcal{RP}}$ . . . . .	108
6.4	Le nombre d'exemples positifs dans les $n$ premières images d'après les scores triés obtenues par le $\mathcal{C}_{\mathcal{PT}}$ . . . . .	109
6.5	Le nombre d'objets de chaque catégorie dans les 36 premières images obtenues par le $\mathcal{C}_{\mathcal{RP}}$ . . . . .	114
6.6	Le nombre d'objets de chaque catégorie dans les 49 premières images obtenues par le $\mathcal{C}_{\mathcal{PT}}$ . . . . .	114
7.1	Les taux de classification à l'incertitude minimal. . . . .	121
7.2	La matrice de confusion. . . . .	122
7.3	Les résultats de la classification pour la détection des ronds-points. . . . .	126



# Première partie

## Introduction et état de l'art

Cette première partie vise avant tout à décrire le sujet de notre thèse, son contexte et le cadre dans lequel elle s'inscrit.

Nous introduisons dans **le chapitre 1** le problème de reconnaissance que nous avons traité tout au long de nos travaux de recherche. Ce problème correspond à un réel besoin des principaux acteurs du domaine de l'imagerie satellitaire. Nous précisons clairement les avancées technologiques qui ont rendu de nouvelles méthodes de traitement en imagerie satellitaire nécessaires, et expliquons nos motivations et nos objectifs.

**Le chapitre 2** est consacré à une étude de l'état de l'art en reconnaissance d'objets, d'abord dans les images en général et puis en imagerie satellitaire. Cette étude nous permet de positionner notre problème par rapport aux travaux de recherche existants et de constater la rapide évolution des méthodes de reconnaissance d'objets dans les dernières années. Nous nous concentrons particulièrement sur les approches les plus récentes qui donnent des résultats très encourageants mais qui sont très peu explorées en imagerie satellitaire. Nous présentons finalement, à la lumière de ces nouvelles approches, les composants principaux de notre thèse.



# Chapitre 1

## Introduction et motivations

L'objet principal de cette thèse concerne la reconnaissance des objets cartographiques à forte structure prononcée dans des images satellitaires à très haute résolution. Aujourd'hui environ 2500 satellites tournent autour de la terre et une très grande quantité d'images de la terre est transmise aux stations terrestres chaque jour. On se sert des images satellitaires dans des domaines très variés avec des enjeux scientifiques, économiques et stratégiques comme la défense, l'agriculture, la géologie, la foresterie, l'organisation régionale et urbaine, le suivi de l'environnement et la cartographie. Par exemple, la Communauté Européenne se sert d'un système de surveillance par satellite dans le cadre de sa politique agricole commune, les systèmes d'information géographiques utilisent les images satellitaires pour extraire les données géographiques, et un programme pour la surveillance des glaces marines et fluviales par satellite est lancé au Canada. Le traitement rapide et efficace des images satellitaires est un préalable primordial pour leur utilisation opérationnelle.

L'imagerie satellitaire présente en plus de nombreux avantages par rapport aux autres sources d'observation du sol comme l'imagerie aérienne, les relevés de terrain et les cartes géographiques : Elle est rapide : le temps nécessaire pour obtenir les images satellitaires d'une région spécifique dépend exclusivement de la planification de l'orbite du satellite. Elle est économique : les images satellitaires ont une couverture très large. Elle est universelle : elle ne connaît pas de frontière géographique ou politique. Elle est facilement mise à jour : on peut renouveler l'information systématiquement avec des délais courts. Elle est détaillée : grâce à la sensibilité des capteurs à un très large spectre de la lumière on obtient des informations variées qui caractérisent l'occupation du sol.

La résolution spatiale d'un capteur est la distance minimum entre deux sources ponctuelles permettant de les distinguer. Plus on a une résolution élevée, plus on peut détecter les objets de petites tailles. Sur une image satellitaire, la résolution spatiale correspond à la superficie au sol représentée par chaque pixel. Par exemple, une image de résolution 30 mètres a des pixels qui représentent une superficie de 30 mètres par 30 mètres au sol. Pendant longtemps les images satellitaires avaient une résolution beaucoup plus basse que celle des images aériennes. Cela ne permettait qu'une utilisation limitée, pour classifier l'occupation du sol grossièrement

ou pour détecter les structures de grande étendue. Avec les progrès technologiques continus du développement de nouveaux capteurs, on a des images satellitaires avec une résolution de plus en plus haute, et les images satellitaires actuelles à très haute résolution s'approchent de la résolution des images aériennes. En conséquence, il est possible d'exploiter les images satellitaires pour des applications qui faisaient partie du domaine de l'imagerie aérienne jusqu'à ce jour.

Prenons un exemple concret pour mieux comprendre le rôle de la résolution dans le choix des traitements qu'on effectue sur l'image : un objet en forme d'un carré de 10 mètres de côté serait représenté par un seul pixel sur une image à résolution spatiale de 10 mètres, quoique sur une image à résolution spatiale de 0,7 mètres, les côtés de ce même objet correspondraient à des chaînes de contours de 15 pixels. Dans le deuxième cas, un traitement qui exploite les propriétés géométriques de l'objet serait bien adapté pour le reconnaître.

On utilise deux types de capteurs dans l'imagerie satellite : Les capteurs radars sont des systèmes d'acquisition actifs qui transmettent un signal radar et mesurent l'intensité du signal que réfléchit la surface de la terre. Par contre, les capteurs optiques sont des appareils passifs qui mesurent le rayonnement électromagnétique provenant du soleil réfléchi par la surface de la terre. Le rayonnement électromagnétique est représenté par un spectre électromagnétique divisé en bandes en fonction de la longueur d'onde. Il existe deux types d'images optiques selon la partie du spectre électromagnétique mesurée : une image panchromatique (Figure 1.1.(a)) est acquise par un capteur qui mesure une seule bande électromagnétique. En général cette bande correspond à la partie visible du spectre. On la représente par une image en niveaux de gris. Une image multi-spectrale (Figure 1.1.(b)) se compose de plusieurs bandes spectrales. On crée une image couleur à partir des valeurs des différentes bandes. Une image multi-spectrale sert principalement à obtenir des informations qualitatives sur l'occupation du sol. La résolution des images multi-spectrales issues d'un même système de capteurs est en général plus basse que celle des images panchromatiques.

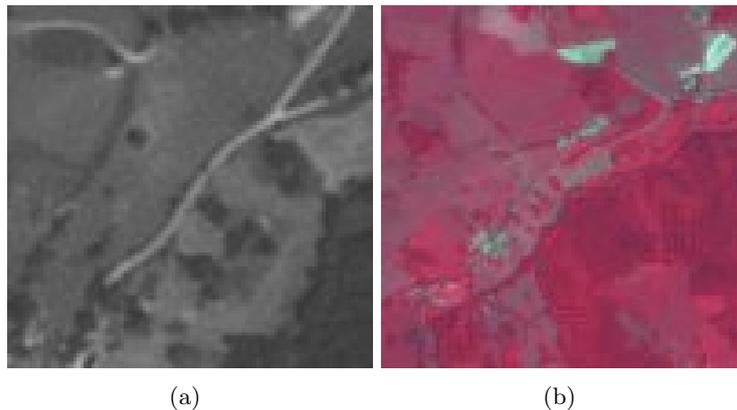


FIG. 1.1 – Exemples d'images satellitaires SPOT5. (a). Une image panchromatique à résolution 2,5 mètres et de taille 100x100 pixels, (b). L'image multi-spectrale de la même région à une résolution de 10 mètres et une taille de 100x100 pixels.

## 1.1 Le programme Pléiades

Le Centre National d'Etudes Spatiales (CNES) se prépare à la mise en orbite de satellites équipés de capteurs optiques de nouvelle génération à partir de l'année 2010. Ces satellites appelés Pléiades auront une capacité de couverture d'images de 2 500 000 km<sup>2</sup> par an qui correspond à 450 images par jour et par satellite. Elles fourniront des images d'une superficie plus petite que celle des images des satellites actuels SPOT5 du CNES, mais à résolution spatiale optimale de 0,7 mètres par pixel qui dépasse la barre symbolique d'un mètre par pixel. Pour montrer l'avancée très importante au niveau de la résolution, il faut rappeler que les satellites SPOT5 ont une résolution de 2,5 mètres par pixel pour les images panchromatiques et 10 mètres par pixel pour les images multi-spectrales (Figure 1.2)



(a)



(b)

FIG. 1.2 – Image satellitaire SPOT5 et la simulation d'image satellitaire Pléiades. (a). Vue de la région du viaduc de Millau prise par Spot 5. Crédits : CNES-dist. Spot Image, 2004, (b). Simulation Pléiades réalisée à partir d'images aéroportées. Exemple de la qualité de détails (résolution de 70 cm) que l'on pourra obtenir avec le premier satellite Pléiades. Crédits : CNES-dist. Spot Image, 2004.

## 1.2 Le programme d'accompagnement ORFEO

Le développement des nouveaux satellites Pléiades rend nécessaire un important travail de préparation pour pouvoir exploiter les futures images de ces satellites. En vue de cet objectif, le CNES a lancé un programme d'accompagnement qui a démarré mi-2003 et qui va durer jusqu'en 2009, juste avant le lancement des premiers satellites.

Ce programme envisage principalement de déterminer les besoins des utilisateurs dans divers domaines, de développer des méthodes adaptées surtout à la haute résolution des nouveaux capteurs (et en même temps aux autres atouts des capteurs comme la capacité d'accès et la qualité des données) et de suivre les nouveaux développements dans le domaine d'imagerie en général et d'imagerie satellite spécifiquement.

On a organisé les travaux de préparation en deux volets qui sont étroitement liés. Cette organisation correspond à l'objectif d'assurer une haute qualité scientifique et une activité de recherche de pointe pour le traitement des images futures, et leur exploitation d'après les besoins des utilisateurs dans un plus grand nombre de domaines. Le volet méthodologique vise à développer des méthodes et des outils nécessaires au traitement d'images. Ce volet est fondé principalement sur le développement des projets de recherche en coopération avec des laboratoires nationaux et européens experts en imagerie numérique. L'objectif du volet thématique est de déterminer les produits et les services dont les utilisateurs non-spécialistes en traitement d'images dans différents domaines ont besoin. On a composé 8 groupes thématiques qui réunissent des experts essentiellement institutionnels autour des thèmes mer et littoral, cartographie et aménagement du territoire, risques et aide humanitaire, géologie et géophysique, hydrologie, forêt, agriculture et défense. Le CNES organise des réunions périodiques qui réunissent les participants des deux volets afin de spécifier et valider les applications qui seront développées dans le cadre du programme ORFEO.

Les objectifs méthodologiques majeurs définis d'après les demandes des experts et les évaluations des scientifiques qui participent au volet thématique sont les suivantes :

- Les pré-traitements ;
- La détection des changements ;
- La segmentation ;
- La détection et la reconnaissance d'objets.

Les trois premiers de ces objectifs correspondent à des sujets de recherche bien connus et bien étudiés en traitement d'images satellitaires, sur lesquels un très grand nombre de travaux scientifiques existe. Par contre, la reconnaissance d'objets se différencie des autres par la relative nouveauté du sujet. Même si c'est un thème d'actualité et d'une très grande importance en traitement d'images, le problème de la reconnaissance d'objets sur les images satellitaires est presque négligé jusqu'à ce jour, surtout à cause de la basse résolution des images satellitaires qui rendaient la reconnaissance d'objets pratiquement impossible. Les travaux existants se concentrent

surtout sur la détection des réseaux routiers, des structures de grande taille et des régions urbaines au lieu de détecter des objets spécifiques comme les bâtiments, les ronds-points, les ponts et les carrefours. Les experts dans différents domaines qui font partie du volet thématique ont indiqué leur intérêt croissant pour le développement de méthodes de reconnaissance d'objets.

Cet objectif a donné lieu à trois projets de recherche interconnectés qui ont les sujets précisés ci-dessous :

- La détection d'objets cartographiques à forte structure prononcée ;
- La reconnaissance géométrique supervisée ;
- Le raisonnement spatial pour la détection d'objet.

Notre équipe, Systèmes Intelligents de Perception (SIP-CRIP5) de l'université Paris Descartes a participé au volet méthodologique du programme d'accompagnement ORFEO depuis le début du programme. L'équipe s'est concentré spécifiquement sur la détection d'objets à forte structure prononcée, ce qui était relativement nouveau à l'époque dans le contexte des images satellitaires et restant encore assez marginal aujourd'hui. Le sujet de notre thèse a été défini dans ce contexte.

### 1.3 Les objets cartographiques à forte structure prononcée

Le projet de recherche sur lequel cette thèse est basée est destiné à détecter automatiquement les objets cartographiques à forte structure prononcée dans les images satellitaires. Dans les images satellitaires nouvelle génération les constructions humaines sont identifiables à l'œil. Pourtant, les images ayant une taille très importante, dans la plupart des applications pratiques, l'extraction manuelle d'information à partir de ces images est extrêmement fastidieuse. D'où provient la nécessité des outils de traitement d'image pour l'extraction automatique d'information.

Ce sujet de recherche correspond à un important changement de paradigme dans le domaine du traitement des images satellitaires. Les méthodes existantes de traitement utilisent en général des algorithmes de bas niveau, qui fonctionnent au niveau pixellique. Par conséquent, ces méthodes ne sont pas capables d'exploiter les propriétés d'objets à un plus haut niveau sémantique, comme la forme, la géométrie et les relations spatiales des composants d'un objet. Nous proposons une approche au niveau "objet" au lieu de rester au niveau "pixellique". C'est un vaste terrain à explorer et nous visons principalement à développer des réflexions sur les éventuels chemins à suivre, les évaluer et les valider sur un problème de reconnaissance concret. Nous croyons qu'avec le progrès technologique au niveau des capacités des systèmes optiques satellitaires, ces nouvelles approches gagneront une importance croissante dans un avenir très proche.

Afin de formaliser le problème de la reconnaissance d'objets dans les images satellitaires, les experts du CNES ont défini 10 catégories d'objets cartographiques. Ces dix catégories sont les autoroutes, les routes moyennes, les routes secondaires, les chemins et pistes, les voies ferrées, les ronds-points, les bâtis-isolés, les lotissements, les ponts et les carrefours. Parmi ces dix catégories,

les cinq premiers correspondent aux différents types de routes. Nous pouvons affirmer que la détection des routes sur les images satellitaires est un sujet de recherche arrivé à maturité. Les méthodes existantes ont des résultats assez satisfaisants pour les applications pratiques et ces méthodes sont en général directement applicables aux images qui ont une résolution spatiale plus haute (même avec des taux de détection plus élevés). Pour cette raison, nous avons exclu la reconnaissance de ces catégories d'objet dans notre recherche. Nous nous sommes intéressés exclusivement à la reconnaissance d'**objets cartographiques à forte structure prononcée, les ronds-points et les ponts**. Ces deux catégories d'objets ont la propriété d'avoir une structure composée de l'assemblage d'éléments géométriques réguliers selon une organisation spatiale bien définie, mais très variable. Leur reconnaissance nécessite une approche de haut niveau qui exploite les propriétés structurelles des objets. Dans un deuxième temps, nous avons appliqué nos méthodes de reconnaissance à la classification des objets cartographiques compacts dans le cadre de la compétition *ROBIN* (Section 1.3.2) et à la détection des ronds-points sur une image satellitaire de grande taille.

Nous présentons dans les deux sections suivantes les bases d'images que nous avons utilisées pour tester nos méthodes de reconnaissance.

### 1.3.1 La base d'images d'objets cartographiques du CNES

Pour valider les méthodes de reconnaissance d'objets sur les images satellitaires, on a besoin d'une plate-forme de validation commune. Le CNES a préparé une base d'images satellitaires qui est constituée d'images d'objets appartenant aux catégories d'objets recherchés. Comme les images des satellites Pléiades ne sont pas encore disponibles et la construction d'une large base d'images par la simulation est une opération coûteuse, on a créé la base à partir des images satellitaires SPOT5 de la région Toulousaine. Les images ont une taille de 100x100 pixels. Sur chaque image l'objet cible est presque exactement centré. Dans chaque catégorie il y a près de 100 objets. Pour chaque objet dans chaque catégorie, cette base contient les images suivantes :

- L'image panchromatique de l'objet ;
- L'image multi-spectrale de l'objet ;
- L'image segmentée, sous la forme d'un masque obtenu par l'extraction manuelle de l'objet par un expert pour les deux types d'images.

Les images panchromatiques ont une résolution de 2,5 mètres par pixel et les images multi-spectrales 10 mètres par pixel. Elles couvrent une superficie au sol de 250 mètres x 250 mètres et 1 km x 1 km respectivement. Les images panchromatiques sont sensibles à une large bande du spectre de 0,48  $\mu\text{m}$  à 0,71  $\mu\text{m}$  et sont représentées par 256 niveaux de gris codés sur 8 bits. Les images multi-spectrales sont sensibles aux bandes verte (0,50-0,59  $\mu\text{m}$ ), rouge (0,61-0,58  $\mu\text{m}$ ) et proche infrarouge (0,78-0,89  $\mu\text{m}$ ) du spectre et sont représentées par trois canaux de couleurs avec 256 valeurs chacun. La figure 1.3 présente les différents types d'images contenant

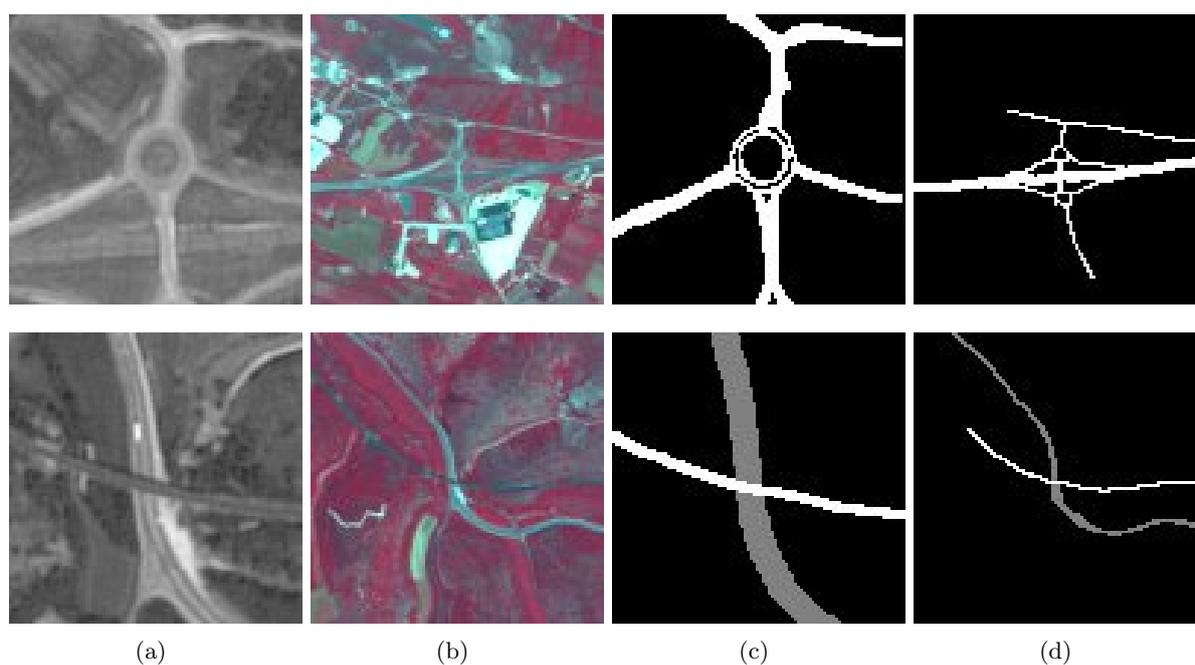


FIG. 1.3 – Imagettes SPOT5 d'objets cartographiques à forte structure prononcée, un rond-point et un pont. (a). Les imagettes panchromatiques, (b). les imagettes multi-spectrales, (c)-(d). les imagettes segmentées correspondantes

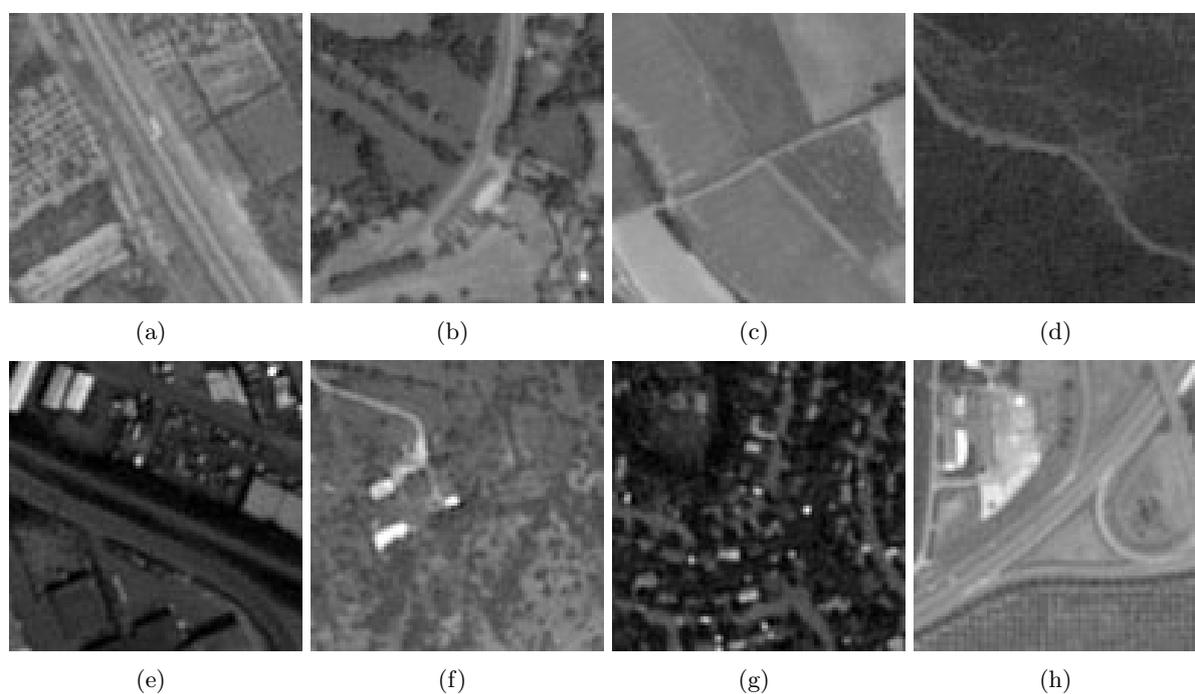


FIG. 1.4 – Imagettes panchromatiques SPOT5 d'objets cartographiques. (a). Les autoroutes, (b). les routes moyennes, (c). les routes secondaires, (d). les chemins et pistes, (e). les voies ferrées, (f). les bâtis-isolés, (g). les lotissements, (h). les carrefours.

un rond-point et un pont. La figure 1.4 présente les imagettes panchromatiques contenant les autres catégories d'objets.

### 1.3.2 Le projet *ROBIN*

Dans le cadre du programme Techno-Vision, un programme national lancé par le Ministère de la Recherche et le Ministère de la Défense pour la création d'une dynamique de l'évaluation de technologies de vision par ordinateur, le projet *ROBIN*<sup>1</sup> vise à mettre en place une structure d'évaluation d'algorithmes de reconnaissance d'objets sur des données d'intérêt opérationnel. Cela concerne la production des bases d'images, des vérités terrains, des compétitions, et des métriques et protocoles d'évaluation. Le projet réunit les acteurs industriels et institutionnels pour la préparation des bases d'images, et les chercheurs pour le développement des algorithmes de reconnaissance d'objets.

On a construit des bases d'images dans trois domaines d'imagerie pour la reconnaissance d'une dizaine de catégories d'objets généraux. Les catégories d'objets et les acteurs industriels qui les ont proposées sont les suivants :

- Imagerie aérienne : EADS, les avions et les véhicules. MBDA, les avions et les hélicoptères. THALES, les véhicules, les voitures et les bateaux
- Imagerie satellitaire : CNES, les objets cartographiques
- Imagerie terrestre : Bertin Technologies, les véhicules et les piétons. SAGEM, les voitures (images infra-rouges)

Une trentaine de laboratoires nationaux en vision par ordinateur ont participé au projet et une dizaine ont participé à la compétition finale. Le défi principal de cette plate-forme est de refléter les besoins des utilisateurs réels. Les autres bases d'images existantes ont été créées le plus souvent par des universités qui ont développé leurs propres méthodes de reconnaissance, et par conséquent leurs contenus ont été influencés par les particularités et les contraintes des méthodes développées. Au contraire, pour le projet *ROBIN*, on a fait un appel aux groupes industriels ayant des activités en vision par ordinateur pour qu'ils proposent des bases d'images correspondant aux problèmes de reconnaissance qu'ils rencontrent dans des applications réelles, indépendamment des méthodes et des algorithmes. Cette approche qui donne la priorité aux besoins opérationnels vise à assurer l'évaluation des algorithmes dans un cadre réaliste. La participation d'importants groupes industriels ainsi que d'un grand nombre de laboratoires de recherche au projet montre l'intérêt applicatif et scientifique du sujet.

On a défini un protocole d'évaluation unifié pour toutes les compétitions. La recherche d'objet se décline en deux tâches :

- La détection : déterminer les occurrences possibles d'un objet appartenant à une catégorie sur l'image de test.

---

<sup>1</sup>Recherche d'Objets dans des Bases d'Images Numériques

- La catégorisation : classifier l'image de test en une des catégories d'objet.

Dans une première étape d'apprentissage (qui peut se réaliser hors ligne), les algorithmes utilisent comme entrée une série d'images étiquetées qui représentent les objets exemplaires.

A l'étape de test, l'entrée est une image ou une série d'images dans lesquelles on doit rechercher l'objet. Pour la détection, la sortie est la liste des positions des objets dans l'image appartenant à une catégorie d'objet unique, présentée initialement au système. Pour la catégorisation, la sortie est la catégorie de l'objet qui se trouve dans l'image parmi un ensemble de catégories possibles. On demande également à l'algorithme de retourner si possible un score de confiance du résultat pour chaque image de test afin de fournir une courbe *ROC*.

Le CNES a participé au projet avec une base d'images semblable à celle présentée à la section 1.3.1, mais avec un plus grand nombre d'images (150 images pour l'apprentissage et 150 pour le test) et sans les images multispectrales et les images segmentées. Cette base contient les images panchromatiques de 10 catégories d'objets cartographiques de taille 125x125 pixels. On a groupé les catégories d'objet dans deux classes : les objets linéaires et les objets compacts. 3 compétitions de détection et 3 compétitions de catégorisation ont été lancées, pour les objets linéaires, pour les objets compacts et pour les objets cartographiques en général. Les compétitions de détection devraient être testées sur des images satellitaires de grande taille qui couvrent des régions contenant plusieurs objets cartographiques. Pourtant, ces images ne sont pas fournies pour la compétition, et par conséquent on a redéfini la détection comme la classification d'une image comme objet linéaire, objet compact ou non-objet.

Nous avons participé à la compétition sur la catégorisation des objets compacts. Le problème proposé est de classifier une image dans une des quatre catégories d'objets. On présente comme entrée une image qui représente un rond-point, un pont, un carrefour ou un bâti isolé. L'algorithme retourne la catégorie de l'image avec un score de confiance.

Nous présentons un bilan de cette évaluation dans le chapitre consacré aux applications (Section 7.2).

## 1.4 Le contexte général de notre recherche

La reconnaissance d'objets dans les images est un sujet actuel de recherche qui est en plein développement. Dans de plus en plus d'applications on utilise les images comme la première source d'information et le traitement manuel de ces images n'est pas envisageable en pratique dans la plupart des cas. Les méthodes de traitement de bas niveau sont en général insuffisantes pour les applications qui nécessitent l'extraction d'information à partir des images à un niveau plus haut. C'est pourquoi la recherche actuelle en vision par ordinateur s'est concentrée sur des approches de haut niveau, qui visent à représenter les images par les objets qui les constituent et leurs relations afin d'avoir une interprétation sémantique de l'image.

Dans le domaine de l'imagerie satellitaire le problème de la reconnaissance d'objets est en

général négligé, ou est limité à la détection de certains objets classiques comme les routes et les régions urbaines. Cela est surtout dû à la basse résolution d’images qui ne permettait pas la détection d’objets dans d’autres catégories plus spécifiques. Aujourd’hui, on est à un point de croisement entre l’amplification importante des travaux de recherche en reconnaissance d’objets et l’avancement technologique au niveau des capteurs en imagerie satellite qui implique le développement de nouvelles méthodes dans l’imagerie satellitaire en utilisant des traitements de haut niveau. Plusieurs travaux de recherche lancés par le CNES ou par des laboratoires nationaux se concentrent exclusivement sur le problème d’interprétation d’images satellitaires, notamment par l’utilisation des représentations symboliques, du raisonnement spatial qualitatif, des outils de l’intelligence artificielle et des techniques d’apprentissage.

Notre thèse s’est inscrit dans ce contexte actuel. Dès le début de notre travail, deux décisions principales nous ont guidés :

- Suivre une approche de haut niveau en visant à construire un modèle d’objet qui est ensuite utilisé pour la reconnaissance d’objets cibles.
- Développer des idées, lancer des discussions et explorer les pistes éventuelles même si celles-ci n’aboutissent pas à des résultats directement utilisables au développement d’un système fonctionnel pour résoudre le problème de reconnaissance.

Nos réflexions se sont concentrées principalement autour de la question “Quel est le modèle d’objet le plus approprié pour représenter les objets cartographiques dans les images satellitaires?”. La distinction entre **les modèles explicites** et **les modèles implicites** a joué un rôle important dans l’avancement de nos travaux.

Par **modèle explicite**, nous appelons les modèles qui essaient de représenter précisément toutes les formes qu’un objet peut prendre. **A partir de ces modèles il est possible de générer des instances de l’objet et de le visualiser.**

Par contre, **les modèles implicites** essaient de définir les formes de l’objet “autorisées” en terme d’apparences locales qui sont consistantes entre eux.

Nos travaux de recherche ont débuté par la génération d’un modèle explicite d’objet représenté par des graphes relationnels attribués (*ARGs*) en utilisant une approche purement structurelle. Ils ont évolué vers un modèle implicite d’objet construit à partir des composants locaux et leurs organisations spatiales, et appris par des méthodes d’apprentissage statistiques. Les chapitres qui suivent racontent le récit de notre odysée à la recherche du modèle d’objet.

## 1.5 L’organisation de la thèse

Cette thèse s’organise autour de quatre parties principales :

**La première partie, Introduction et état de l’art**, vise avant tout à décrire le sujet de notre thèse, son contexte et le cadre dans lequel elle s’inscrit. Dans le chapitre 2, nous

présentons une vue d'ensemble des domaines d'applications de reconnaissance d'objets dans les images et les différentes approches et méthodes proposées. Nous nous focalisons ensuite sur les travaux de recherche en traitement d'images satellitaires. Cette étude bibliographique nous permet de nous positionner par rapport aux travaux existants et de proposer une démarche pour la reconnaissance d'objets cartographiques dans les images satellitaires.

**La deuxième partie, Génération d'un modèle d'objet explicite**, présente une première approche d'apprentissage structurel pour construire un modèle d'objet de ronds-points et de ponts à partir d'un ensemble d'images segmentées. Ce modèle explicite est fondé sur une représentation par des graphes relationnels attribués (*ARGs*) générés par des méthodes d'appariement de graphes.

**La troisième partie, Modèles implicites d'objets fondés sur l'apparence**, est constitué de trois chapitres : dans le chapitre 4 on développe une chaîne de traitement visant à extraire des primitives géométriques à partir des images satellitaires en utilisant une approche par contours et une approche par régions en parallèle. Dans le chapitre 5, nous présentons un modèle d'objet implicite. La méthode est basée sur la fusion des classifieurs faibles obtenus à partir des primitives en utilisant l'algorithme *Adaboost* pour construire un classifieur fort. Le chapitre 6 est consacré à l'apprentissage d'un modèle d'objet implicite par une approche composants et structure. On représente les composants d'un objet par des primitives géométriques et sa structure par les propriétés géométriques et l'organisation spatiale des primitives. On apprend les composants et la structure simultanément par l'algorithme de *clustering Mean-Shift*.

**La quatrième partie, Applications, évaluations, conclusion et perspectives**, présente deux applications concrètes que nous avons développées dans le cadre de notre thèse, donne les conclusions principales que nous avons retirées de nos travaux de recherche et indique les principales directions de recherche pour les travaux futurs.



## Chapitre 2

# La reconnaissance d'objets dans les images

### 2.1 Introduction

La reconnaissance automatique d'objets dans les images est un des problèmes les plus difficiles en vision par ordinateur. Elle est en même temps une étape primordiale pour la mise en oeuvre de plusieurs applications actuelles qui nécessitent une interprétation de haut niveau d'images. Par conséquent, il existe un intérêt croissant sur ce domaine de recherche dans les dernières années et une vaste littérature. Notre objectif, dans ce chapitre, est de présenter d'abord les approches et les méthodes principales de reconnaissance d'objets en général et d'évaluer leur applicabilité à notre problème. Dans un deuxième temps, nous allons nous concentrer spécifiquement sur les images satellites, et présenter les travaux de recherches en reconnaissance d'objets dans ces images. Finalement nous allons expliquer nos choix de méthodes pour le développement de notre système de reconnaissance d'objets cartographiques à la lumière des travaux existants.

Le problème de la reconnaissance d'objets est de décider si un objet spécifique ou un objet appartenant à une catégorie d'objets prédéfinie existe dans une image. On peut poser le problème de la reconnaissance comme un appariement entre le modèle de l'objet cible et un ensemble de descripteurs qu'on extrait à partir de l'image de test. Cette généralisation, aussi simple qu'elle semble être, est capable d'expliquer la multitude d'approches qui se différencient par le choix des descripteurs qui représentent l'objet, le type et la complexité du modèle, et les méthodes utilisées pour l'apprentissage du modèle d'objet et l'appariement.

### 2.2 Les principaux domaines d'application

Au delà d'être un problème général de la vision par ordinateur, la reconnaissance d'objets dans des catégories très variables est une étape indispensable pour de nombreuses applications spécifiques. On peut citer, parmi d'autres, quelques applications exemplaires pour illustrer les

enjeux des travaux de recherche sur ce problème.

- La vidéo-surveillance : la reconnaissance et la détection automatique des personnes, des visages et des objets d'intérêt sur la très grande quantité d'images vidéo utilisées pour la surveillance.
- La recherche dans des bases d'images numériques : les systèmes actuels effectuent en général des requêtes à partir des mots clés associés aux images. On voudrait exploiter le contenu visuel des bases d'images sans avoir besoin de les indexer textuellement.
- La navigation des robots mobiles et des véhicules : on a besoin de reconnaître les obstacles, les objets divers dans l'environnement ou les autres véhicules pour la navigation automatique.

Pour pouvoir évaluer les différentes méthodes et pour pouvoir les comparer, il est nécessaire de créer des plate-formes communes de test. Ont été définies, par des équipes de recherche qui travaillent sur ce problème, des bases d'images consacrées à la reconnaissance d'un ensemble limité de catégories d'objets. Ces bases contiennent en général un grand nombre d'images d'objets étiquetées et éventuellement des informations associées (la position, le rectangle englobant, la segmentation de l'objet). En considérant qu'elles indiquent les principales directions de recherche, nous voulons mentionner les bases d'images publiques les plus connues et les catégories d'objets répertoriés les plus communes :

- Bases d'images FERET, CMU/MIT, MIT-CBCL, Caltech, Yale : les visages ;
- Bases d'images ETHZ, Caltech, UIUC : les voitures, les motocyclettes ;
- Bases d'images Pascal, CMU : les bicyclettes ;
- Bases d'images Graz 02, Pascal : les personnes, les vaches ;
- Base d'images Weizmann : les chevaux ;

Le choix des catégories d'objets dans ces bases reflète principalement les deux priorités de la communauté scientifique : la nécessité de détecter l'objet cible pour des applications réelles (les visages, les voitures, les personnes) et une forme bien précise qui distingue les objets des autres catégories similaires (les vélos, les vaches, les chevaux). Nous avons testé nos méthodes de reconnaissance sur une base d'images étiquetées préparée par le CNES dans le cadre du programme ORFEO et la compétition ROBIN du programme national Technovision (Sections 1.3.1 et 1.3.2). La principale difficulté liée à cette base par rapport aux autres est la variance très importante des objets d'une même catégorie et la similarité radiométrique des différentes catégories. Inglada [2007] a démontré cette difficulté par un test de classification humaine des objets qui a donné des taux d'erreur élevés. A l'inverse, un test similaire sur les autres catégories mentionnées donnerait un taux de reconnaissance de près de 100%.

## 2.3 La reconnaissance spécifique versus générique

On peut faire la distinction entre deux problèmes de reconnaissance différents : **la reconnaissance spécifique** on cherche à détecter un objet précis qui a éventuellement subi des déformations perspectives ou des transformation d'illumination. Les exemples sont la reconnaissance d'un jouet d'enfant [Lowe, 2004], la reconnaissance d'un produit spécifique d'une marque [Auclair *et al.*, 2007], la reconnaissance des panneaux de signalisation ou l'identification d'une personne. Dans **la reconnaissance générique**, par contre, l'objectif est de détecter les objets appartenant à une certaine catégorie. Ce problème est considérablement plus difficile que le précédent. La catégorie d'un objet n'est pas un attribut intrinsèque de l'objet, elle est souvent apprise par l'observation des similarités et des différences d'objets distincts. L'homme est capable d'attribuer une catégorie à un objet vu pour la première fois et aussi de créer des nouvelles catégories pour de nouveaux objets. Par conséquent, les objets dans une même catégorie peuvent avoir des apparences très variables et la difficulté majeure du problème de la reconnaissance générale est de trouver une description de la catégorie qui permet de représenter ces variances. **Les travaux récents se concentrent de plus en plus souvent sur le problème de la reconnaissance générique** [Agarwal et Roth, 2002; Borenstein et Ullman, 2002; Csurka *et al.*, 2004; Dorkó et Schmid, 2003; Li et Perona, 2005; Felzenszwalb et Huttenlocher, 2005; Fergus *et al.*, 2003, 2005; Heisele *et al.*, 2001; Leibe *et al.*, 2004; Murphy *et al.*, 2006; Opelt *et al.*, 2004, 2006; Schmid et Mohr, 1997; Shotton *et al.*, 2007; Sivic *et al.*, 2005; Torralba *et al.*, 2006; Weber *et al.*, 2000]. Les thèses récentes de Fergus [2005] et Opelt [2006] se concentrent exclusivement sur ce problème d'une grande importance actuelle.

Notre problème correspond à la reconnaissance générique d'objets. Les objets cartographiques appartiennent à des catégories qui représentent des concepts sémantiques généraux. La catégorisation est principalement basée sur la structure : les objets ont presque tous les mêmes composants avec des structures géométriques mais ces composants présentent des organisations spatiales différentes.

## 2.4 Les défis majeurs de la reconnaissance d'objets

La difficulté majeur du problème de la reconnaissance d'objets repose sur les variations d'un objet ou des objets d'une même catégorie sur des images différentes. On peut grouper les sources de variations dans sept catégories principales :

- Les variations de point de vue : les objets réels sont des entités physiques tri-dimensionnelles. Leur apparence change de façon significative en fonction du point de vue.
- Les variations d'illumination : en fonction de changements de l'environnement et des conditions d'acquisition d'une image, un objet peut subir des variations d'illumination importantes. Les objets dans les images des scènes naturelles sont très sensibles aux variations d'illumination qui modifient les couleurs et les valeurs radiométriques des objets considé-

rablement.

- Les occlusions : dans une image les objets peuvent être partiellement masqués par d'autres objets ou par leurs propres composants.
- Les variations d'échelle : les objets d'une même catégorie peuvent avoir des tailles physiques différentes ou les images d'un même objet peuvent avoir des tailles différentes en fonction de la résolution de l'image.
- Les déformations : le plupart des objets réels n'ont pas une forme rigide. Particulièrement les objets articulés subissent généralement des déformations linéaires et non-linéaires de leurs composants.
- Complexité du fond : en général, dans une image les objets ne sont pas isolés de leur environnement. Sur un fond complexe il est difficile de distinguer les régions appartenant à l'objet et au fond.
- Variations intra-classe : c'est le plus important des problèmes envisagés pour la reconnaissance générique d'objets. Les objets d'une même classe peuvent avoir des apparences très variables, même quand on se limite aux catégories à un niveau élémentaire. Dans le cas extrême, les catégories **fonctionnelles** comme la classe "chaise" sont définies par un haut niveau d'abstraction et les indices visuels seuls ne sont pas suffisants pour les reconnaître.

L'objectif des méthodes de reconnaissance d'objets est de trouver des descripteurs et des représentations qui soient invariants aux variations de l'objet présentées ci-dessus. Cet objectif étant très ambitieux, pendant la phase initiale de la définition du problème de reconnaissance et de la construction de la base d'images on applique souvent des contraintes qui limitent les éventuelles variations des objets .

Les images satellites offrent un point de vue constant. Elles n'ont pas des variations d'illumination importantes. Les objets cartographiques ne sont pas en général sujets à des occlusions. Dans notre base toutes les images ont la même résolution spatiale. Pourtant les objets ont des tailles physiques assez différentes. Ce sont des objets articulés qui ont subi des déformations importantes et qui ont une grande variation intra-classe. Finalement il est très difficile de séparer les objets du fond à cause de la ressemblance des composants d'objets cibles à des autres objets environnants. Ces considérations ont eu un impact crucial pour la sélection des méthodes que nous avons proposées.

## 2.5 La vision humaine

L'œil humain est capable de reconnaître rapidement une multitude d'objets différents avec des variations de point de vue, d'illumination et de forme, même dans les contextes les plus inhabituels. On reconnaît les catégories d'objets aussi bien que les instances d'objets uniques, et cela avec une capacité de généralisation sur plusieurs échelles et sous plusieurs sortes de transformations. On peut ajouter à cela l'apprentissage rapide d'objets qu'on voit pour la première fois.

On est loin de connaître le modèle complet de la perception humaine qui réunit des mécanismes cognitifs complexes et de l'imiter pour développer un système automatique capable d'approcher la performance humaine. Pourtant, les expériences psycho-cognitives sur les étapes primales (la reconnaissance rapide sans utiliser les processus cognitifs de haut niveau) de la vision humaine et sur la vision des pigeons ont révélé des indices qui peuvent être exploités par les systèmes automatiques. Il nous semble pertinent de citer deux travaux expérimentaux importants.

### 2.5.1 *Particulate feature theory*

Cerella [1986] propose que la vision chez les pigeons fonctionne comme des perceptrons. Chaque perceptron correspond à une caractéristique simple locale de l'objet et la reconnaissance est liée aux réponses des perceptrons individuels. Pour justifier sa théorie, il présente une expérience où les pigeons reconnaissent de la même façon un objet quand ses composants sont mélangés. Cela montre que ce qui est prioritaire pour la reconnaissance d'objets est la détection des caractéristiques locales mais non pas l'organisation spatiale des primitives.

Cette théorie, beaucoup discutée parce qu'elle ne peut pas expliquer la reconnaissance en cas de changement de point de vue, est quand même très importante pour éclairer les étapes primitives de la reconnaissance. Notre système de reconnaissance fait l'extraction d'un ensemble de primitives géométriques simples à partir de l'image. Cette étape se conforme bien aux conclusions de Cerella sur les mécanismes de la vision aviaire. Une question intéressante qui n'a pas été abordée par Cerella serait "Est-ce que les pigeons procèdent à la séparation de l'objet du fond ou est-ce qu'ils considèrent l'image comme un ensemble de toutes les primitives?".

### 2.5.2 La théorie de la reconnaissance par composants

Biederman [1987] a développé la théorie de la **reconnaissance par composants** pour expliquer la reconnaissance d'objets par les humains. Selon sa théorie, les objets sont constitués de composants primitifs et on les reconnaît en détectant ces composants et leurs relations. Les composants élémentaires sont appelés "les géons" (*geometrical ions*). Ce sont des formes géométriques tri-dimensionnelles, obtenues par la déformation d'un cylindre (Figure 2.1), avec les propriétés suivantes :

- Invariantes au point de vue : ils sont identifiables de plusieurs points de vue différents ;
- Discriminantes : il est difficile de confondre un géon avec un autre ;
- Résistantes au bruit : on peut les détecter même en cas d'imperfection de la scène.

Le modèle de Biederman ressemble au système de la construction des mots à partir d'un nombre limité de phonèmes en utilisant des règles de construction. Un objet est représenté dans le cerveau par les géons qui le constituent et les relations spatiales entre les géons. La reconnaissance est faite par la détection des géons et par un appariement avec le modèle d'ob-

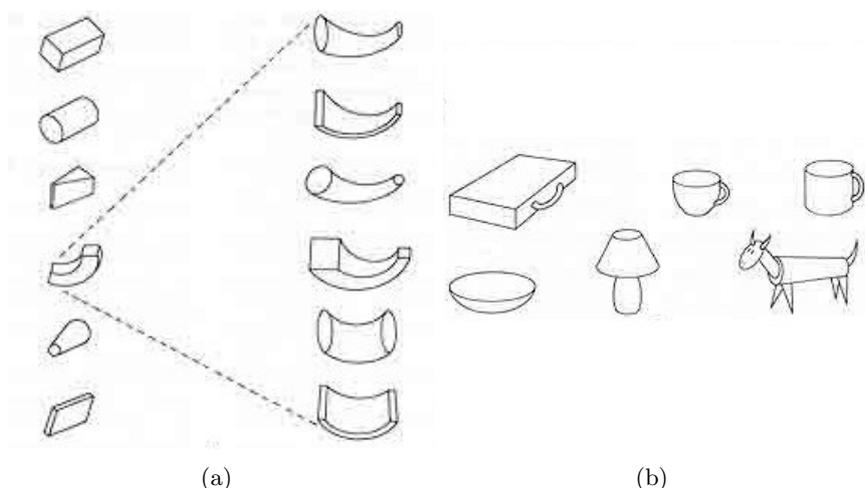


FIG. 2.1 – Les géons. (a). Exemples de géons, (b). Quelques objets composés de géons.

jet correspondant. Biederman propose que sur une image bi-dimensionnelle, les géons peuvent être détectés grâce aux “propriétés non-accidentelles” (*NAPs* : *Non-Accidental Properties*) des contours, les propriétés de l’image qui ne changent pas en général par le changement de perspective (contrairement aux propriétés métriques, comme la longueur des contours). Ce sont la courbure, la colinéarité, la symétrie, le parallélisme et la cotermination. La figure 2.2 [Price, 1996] montre comment l’ajout intentionnel de quelques segments de droites sur un ensemble aléatoire de lignes crée des groupements non-accidentels. Dans un article récent, Biederman [2007] présente des évidences neurales qui justifient la théorie de la description structurale par les géons et les *NAPs* chez les macaques.

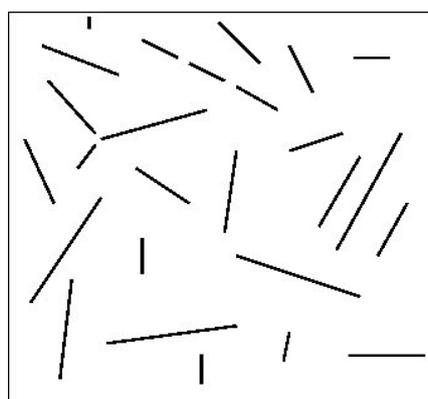


FIG. 2.2 – Les groupes formés par les propriétés non-accidentelles. L’œil humain perçoit rapidement les groupes des parallèles, des coterminations et des segments de droites colinéaires.

La théorie de Biederman explique bien la reconnaissance d’objets sous différents points de vue, avec des changements d’illuminations et même quand l’image de l’objet n’est pas complète. En même temps, elle est bien appropriée pour être adaptée à un modèle informatique. Ce qui nous a intéressé le plus pour notre recherche dans cette théorie, est la décomposition de l’objet

par des primitives géométriques, l'extraction de ces primitives à partir des contours de l'image (en utilisant les propriétés non-accidentelles) et la construction du modèle d'objet à partir des relations spatiales des primitives.

## 2.6 Le schéma général des méthodes de reconnaissance d'objets

Le fossé sémantique est un terme utilisé pour décrire la difficulté des systèmes de vision par ordinateur pour obtenir une interprétation de haut niveau à partir des descriptions visuelles de bas niveau d'une image. Le problème de la reconnaissance d'objets automatique est un exemple parfait pour illustrer cette difficulté. On cherche dans l'image, qui n'est qu'un ensemble de pixels ayant des valeurs d'intensité numériques et des coordonnées dans un espace bi-dimensionnel, des objets qui ont des descriptions correspondantes aux concepts sémantiques. Les premiers travaux de recherche en vision par ordinateur se sont surtout concentrés sur les traitements de bas niveau, les traitements qui opèrent uniquement sur l'intensité des pixels sans considérer les entités qu'ils représentent, comme la segmentation et la détection des contours. Après être arrivé à une certaine maturité aux traitements de bas niveau, on est confronté aujourd'hui à un intérêt croissant de la communauté scientifique pour des traitements de haut niveau qui visent à interpréter une image par des entités symboliques et leurs relations spatiales. Les méthodes de reconnaissance suivent en général ces deux niveaux de traitements, les traitements de bas niveau et les traitements de haut niveau qui correspondent aux étapes suivantes :

1. *Extraction* : l'extraction d'un ensemble de descripteurs à partir de l'image, qui reflètent les informations pertinentes pour la reconnaissance de l'objet ;
2. *Représentation* : la construction d'un modèle d'objet pour représenter une catégorie d'objet ;
3. *Apprentissage* : l'apprentissage des paramètres du modèle à partir des images d'apprentissage étiquetées ;
4. *Reconnaissance* : l'appariement avec le modèle ou la classification en utilisant le modèle pour reconnaître l'objet cible dans une image de test.

Les méthodes de reconnaissance d'objets se différencient principalement au niveau de :

- La définition du problème de reconnaissance : les catégories d'objets recherchées, le nombre de catégories ;
- Le nombre d'images utilisées pour l'apprentissage : une extrémité étant l'apprentissage à partir d'une seule image, on peut aussi utiliser un très grand nombre d'images pour l'apprentissage ;
- Le niveau de supervision : en cas d'un apprentissage non-supervisé on ne dispose pas d'images étiquetées, et on essaie de grouper les images dans des catégories d'après leurs similarités (*clustering*). En cas d'apprentissage supervisé, on a un ensemble d'images d'ap-

prentissage étiquetées. Il s'agit d'un niveau de supervision plus élevé quand on connaît aussi la segmentation manuelle des parties de l'objet sur l'image ;

- La modélisation : ici, nous utilisons le terme modèle dans le sens général de “la représentation d'un objet”. Un très grand nombre de modèles d'objets différents ont été proposés pour représenter les objets dans les images ;
- Les méthodes et les algorithmes choisis pour l'extraction des descripteurs, la construction du modèle et la reconnaissance.

La reconnaissance d'objets nécessitant une chaîne de traitement qui se constitue de plusieurs étapes, il est difficile de présenter la multitude d'approches utilisées sans analyser chacun des principaux composants individuellement. Même si les traitements de haut niveau ont une grande importance pour la reconnaissance, elles se sont construites sur les résultats d'une étape de traitement de bas niveau qui est également importante et toujours d'actualité en vision par ordinateur. Nous introduisons ci-dessous les approches existantes pour ces étapes majeures, avant de présenter les méthodes de reconnaissance d'objets les plus significatives.

## 2.7 Les traitements de bas niveau : extraction des descripteurs

L'utilisation directe de l'image pour la reconnaissance n'est pas envisageable parce qu'elle contient une grande quantité d'information impertinente ou redondante. C'est pourquoi, la première étape de presque toutes les méthodes de reconnaissance est d'extraire un ensemble de descripteurs qui représentent les propriétés discriminantes de l'objet cible. Les descripteurs sont un ensemble de nombres scalaires générés pour décrire un objet (ou une forme). Un objet ne peut pas être reconstruit en général à partir des descripteurs, mais les descripteurs pour des objets distincts doivent avoir des valeurs suffisamment différentes pour les discriminer, et les descripteurs pour les objets similaires doivent avoir des valeurs proches. On peut extraire un nombre infini de descripteurs à partir d'une image, pourtant l'objectif est d'en sélectionner un nombre moins important que le nombre de pixels dans l'image. Pour une reconnaissance robuste il est souhaitable que les descripteurs soient invariants aux changements d'échelle et d'illumination et aux transformations affines.

On peut distinguer deux types de descripteurs : **les descripteurs globaux** sont obtenus à partir de l'image entière. Par contre, **les descripteurs locaux** caractérisent un ensemble de régions locales significatives de l'image qui sont riches en information. Les travaux récents en reconnaissance d'objets se concentrent le plus souvent sur l'extraction des descripteurs locaux qui sont plus efficaces pour la reconnaissance générique. On utilise principalement les contours, les régions ou finalement les propriétés de texture et de couleur de l'image pour extraire les descripteurs. Ci-dessous, nous présentons les principaux types de descripteurs et les méthodes utilisées pour les extraire.

### 2.7.1 Les points d'intérêt

Les points d'intérêt sont des régions de l'image riches en terme de contenu de l'information locale et stables sous des transformations affines et des variations d'illumination. Ils sont des indicateurs des régions susceptibles de contenir un objet, et en même temps des parties importantes de l'objet. Les points d'intérêts sont plus stables que les régions ou les contours de l'image et leur extraction est plus simple.

#### 2.7.1.1 Détection des coins

Le détecteur des coins de Harris et Stephens [1988] est l'un des plus populaires des détecteurs de points d'intérêt. Un coin, l'intersection de deux contours, a un contenu informatif important. Sur l'image les coins correspondent aux pixels qui ont des changements d'intensité élevés dans toutes les directions sur un voisinage local (par contre sur les régions plates le changement d'intensité est nul dans chaque direction, et sur les contours il est élevé dans une seule direction). L'idée principale est donc de trouver les points à partir desquels un petit déplacement donne un changement d'intensité élevée. Pour cela, on calcule  $M$ , la matrice de covariance du gradient dans une fenêtre de  $2n + 1$  autour d'un pixel  $p$ . Les deux valeurs propres de  $M$  correspondent au changement minimal et au changement maximal d'intensité au voisinage du  $p$ . Deux valeurs propres qui sont supérieures à un seuil (précisé empiriquement) indiquent l'existence d'un coin. La figure 2.3 montre les coins détectés sur une image satellitaire.

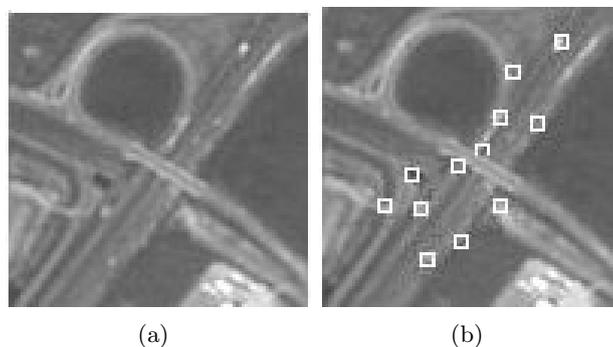


FIG. 2.3 – Les coins détectés sur une image satellitaire par le détecteur de Harris et Stephens [1988].

Les coins de Harris sont très sensibles au changement d'échelles et ils ne contiennent pas d'information sur la région locale autour du point d'intérêt. Par conséquent il n'est pas possible d'apparier correctement les points d'intérêt correspondants sur deux images sauf si les objets sur les images ont des formes précises et identiques.

#### 2.7.1.2 Les descripteurs *SIFT*

Schmid et Mohr [1997] ont proposé de calculer un descripteur local pour chaque point d'in-

térêt. Le descripteur local est calculé à l'aide des invariants différentiels de l'intensité des pixels [Koenderink et van Doorn, 1987] sur un voisinage local autour du point d'intérêt. Pour obtenir l'invariance à l'échelle on utilise un ensemble de dérivées Gaussiennes avec des valeurs de  $\sigma$  différentes. La figure 2.4 illustre l'approche utilisée. L'association des descripteurs locaux aux points d'intérêt rend possible un nombre d'appariement correct élevé entre deux images qui contiennent le même objet. Cela assure la robustesse de la reconnaissance d'objets.

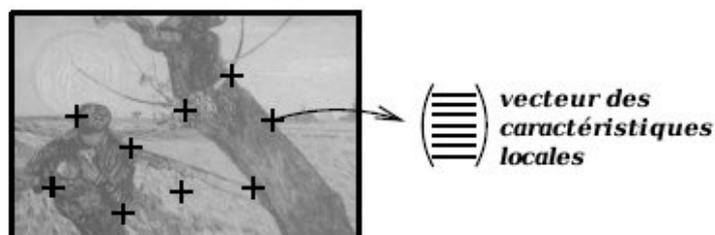


FIG. 2.4 – Les points d'intérêt et les descripteurs locaux, dans Schmid et Mohr [1997].

La méthode *SIFT* (*Scale Invariant Feature Transform*) de Lowe [2004] est basée sur la même idée. C'est une méthode actuellement très populaire qui détecte des points d'intérêt et leurs descripteurs locaux invariants à l'échelle et aux rotations, et invariants partiellement aux changements d'illumination. Les descripteurs obtenus sont très distinctifs. On peut appairer un descripteur correctement avec une probabilité très élevée parmi une grande base de descripteurs obtenus d'autres images.

L'algorithme procède en quatre étapes principales : la détection des extrema (qui correspondent aux points d'intérêt) à un espace d'échelle créé par des filtres *DoG* sur une pyramide multi-échelle ; la localisation et le filtrage des points d'intérêt ; le calcul des histogrammes des directions des gradients locaux autour du point d'intérêt dans un voisinage qui varie en fonction de l'échelle du point, et le calcul d'une signature à partir des histogrammes en divisant la zone autour du point en quatre sous-zones. La signature est un code de taille 32 qui indique la répartition des 8 orientations principales pour chaque zone (Figure 2.5, Lowe [2004]).

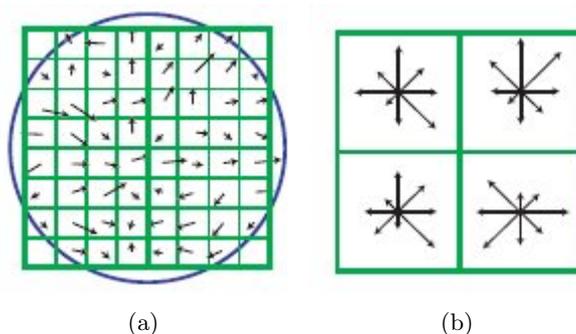


FIG. 2.5 – SIFT, calcul de la signature d'un point d'intérêt. (a). Les gradients de l'image autour du point d'intérêt, (b). Les histogrammes correspondants.

L'objet recherché est représenté finalement par un ensemble de points d'intérêt et les signatures associées, et la reconnaissance est effectuée par un appariement des points d'intérêt entre les images. Grâce aux propriétés d'invariance et à la haute capacité descriptive des points d'intérêt, les résultats sont excellents pour la reconnaissance d'un objet spécifique dans une autre image. L'appariement de seulement trois points suffit pour une détection fiable.

Pourtant, les méthodes de reconnaissance par la correspondance des points d'intérêts ne sont pas adaptées pour la reconnaissance générique surtout quand la variation intra-classe est élevée. Les descripteurs locaux sont trop spécifiques pour apparier deux objets dans la même catégorie mais avec des variations. Pour notre problème, les objets appartenant à une classe n'ont pas de caractéristiques locales similaires, mais ont des caractéristiques communes structurelles à un plus haut niveau d'abstraction. Pour cette raison, nous n'avons pas décidé d'utiliser ces méthodes.

### 2.7.1.3 Les blobs

*Les blobs* sont des régions de l'image qui sont plus claires ou plus sombres que leurs environs. On peut les considérer comme des points d'intérêt qui signalent les objets ou les composants des objets. Le détecteur le plus fréquemment utilisé est le filtre *LoG* qui donne des valeurs positives élevées pour *les blobs* sombres et des valeurs négatives élevées pour *les blobs* claires. Pourtant la réponse de l'opérateur dépend de l'échelle du *LoG*. Lindeberg [1998] a proposé d'utiliser un opérateur Laplacien normalisé sur un espace d'échelle et de trouver les maxima locaux dans l'espace et dans l'échelle simultanément. De cette façon on détecte *les blobs* à l'échelle correspondante.

#### 2.7.1.4 Maximally Stable Extremum Regions (MSER)

Les *MSER* proposés par Matas *et al.* [2004] est une technique pour détecter *les blobs* dans une image. On obtient une série d'images noires et blanches en appliquant un seuillage pour chaque niveau de gris dans l'image. *Les MSER* sont *les blobs* qui restent stables sur plusieurs images consécutives.

## 2.7.2 Les descripteurs contours

Les contours d'un objet donnent suffisamment d'information sur sa forme. D'après la définition du dictionnaire Le Petit Robert [1973] le mot "forme" signifie :

*Forme : Ensemble des contours d'un objet, d'un être, résultant de la structure de ses parties.*

L'oeil humain est capable de reconnaître un objet à partir des fragments de ses contours (Figure 2.6).

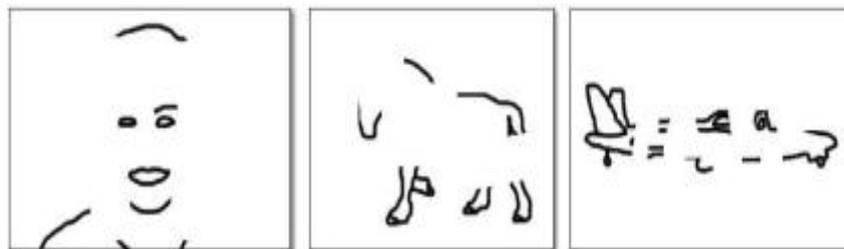


FIG. 2.6 – Les fragments de contours de quelques objets courants, dans Shotton *et al.* [2007].

Les descripteurs obtenus à partir des contours ont l'avantage d'être invariants aux changements d'intensité, de texture et de couleur de l'image. Les descripteurs contours sont des candidats idéaux pour représenter les propriétés structurales des objets.

### 2.7.2.1 Détection des contours

Les contours correspondent aux changements brusques de l'intensité de la lumière sur l'image. Sur une image numérique ces changements peuvent être détectés à l'aide des dérivées du premier et du second ordre de la fonction d'intensité. Un point de contour est défini comme le maximum du module du gradient dans la direction du gradient. Avant la détection des contours on procède à des pré-traitements pour supprimer le bruit provenant des imperfections des capteurs. Le filtre le plus fréquemment utilisé pour supprimer le bruit est le filtre gaussien  $G_\sigma(x, y)$  :

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)}$$

qui pondère la valeur de chaque pixel selon une distribution normale autour de ce pixel. Cela résulte en un lissage de l'intensité de l'image. La dimension du filtre dépend du  $\sigma$  et plus on a un  $\sigma$  élevé, plus le lissage est important. En convoluant l'image par une série de filtres gaussiens avec des valeurs de  $\sigma$  croissantes, on peut obtenir une représentation multi-échelle de l'image qu'on appelle la pyramide gaussienne.

Une méthode de filtrage adaptatif est la diffusion anisotropique [Perona et Malik, 1990] où on contrôle le niveau de la diffusion, donc du lissage, par un paramètre d'échelle qui la réduit quand l'intensité du gradient augmente.

La détection est souvent suivie d'une étape de post-traitement qui sert à réduire le nombre de points de contours et à les relier. Très souvent, on élimine les points de contour qui ont un gradient faible par un seuillage. Des algorithmes de suivi de contours peuvent ensuite être appliqués pour connecter les points de contours adjacents.

**Les dérivées du deuxième degré :** On peut se servir des dérivées du deuxième degré de la fonction d'intensité de l'image  $f(x, y)$  pour trouver les maxima du gradient. Les passages par zéro du Laplacien

$$\Delta = \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

correspondent aux maxima du gradient si le contour a une courbure faible. Les passages par zéro du Laplacien ont l'avantage de donner des contours fermés. En pratique, on filtre initialement l'image par un filtre gaussien. On peut composer ces deux filtres pour obtenir le filtre Laplacien de Gaussienne [Marr, 1980], *LoG* :

$$\Delta [G_\sigma(x, y) * f(x, y)] = [\Delta G_\sigma(x, y)] * f(x, y) = LoG * f(x, y)$$

Un filtre similaire, différence de gaussiennes (*DoG*) est obtenu par la différence de deux gaussiennes avec des valeurs de  $\sigma$  différentes mais proches.

**Le détecteur de Canny :** Le détecteur de Canny [1986] est considéré par plusieurs comme le détecteur optimal. Canny a précisé trois critères pour un bon détecteur et a proposé une méthode qui satisfait ces trois critères. Ce sont : la bonne détection (faible taux d'erreur), la bonne localisation (minimisation des distances entre les contours réels et contours détectés) et la clarté de la réponse (une seule réponse par contour).

**Les modèles déformables :** Une autre approche de détection, très souvent utilisée pour les applications médicales, est la méthode des modèles déformables. L'idée principale est de faire évoluer itérativement une courbe initiale afin qu'elle s'approche à chaque itération du contour réel jusqu'à ce qu'elle reste constante. L'évolution de la courbe est faite par la minimisation d'une fonction d'énergie qui est la somme d'énergies internes et externes. L'énergie interne assure la régularisation de la forme du contour et l'énergie externe l'attache au contour réel (aux valeurs élevées du gradient). Les deux grandes familles des modèles déformables sont les contours actifs (*snakes*) [Kass *et al.*, 1988] où on utilise un modèle paramétrique pour représenter la courbe, et les ensembles de niveau (*level sets*) [Andrew, 2000] où on utilise un modèle géométrique. Ces méthodes sont très sensibles à l'initialisation de la courbe et sont mieux adaptées quand on dispose de connaissances *a priori* sur la forme de l'objet recherché.

### 2.7.2.2 Extraction des descripteurs contours

Les contours d'un objet ne sont pas utilisables directement pour représenter l'objet mais on procède en général à une étape d'extraction des descripteurs.

### 2.7.2.2.a Les fragments de contours

Shotton *et al.* [2007] propose d'utiliser des fragments de contours comme descripteurs. A partir d'un petit nombre d'images d'objets segmentées on construit un dictionnaire des fragments. Le dictionnaire est constitué des fragments des contours extérieurs de l'objet et des positions des fragments par rapport au centre de l'objet. Pour construire le dictionnaire, on génère aléatoirement un grand nombre de régions qui contiennent une partie de l'objet et du fond. On perturbe légèrement ces régions par des transformations aléatoires pour obtenir un dictionnaire plus représentatif, et on détecte les contours sur chaque région.

Ferrari *et al.* [2006] représente un objet par un réseau de segments de droites. Après l'extraction des contours, on procède à une approximation polygonale et un chaînage des segments par une recherche locale à leurs extrémités. On définit un ensemble de règles pour connecter les contours. On représente finalement les contours de l'objet par un graphe de segments de contour.

Suivant la même idée principale, Ferrari *et al.* [2008] apporte une amélioration majeure en proposant un codage (par un descripteur numérique) des propriétés géométriques de chaque groupe de  $k$  segments adjacents ( $kAS$ ). Le codage est invariant à l'échelle, aux translations et aux rotations. Le nombre de  $kAS$  obtenus étant très élevé pour l'appariement, on construit un dictionnaire de codes visuels ( $\cdot$ ) en groupant les  $kAS$  similaires par l'algorithme de partition des cliques. Au delà de  $k = 3$  les  $kAS$  représentent des structures très spécifiques et leur nombre devient très important. Pratiquement on construit le dictionnaire de codes pour  $k = 2$  et  $k = 3$  (Figure 2.7).

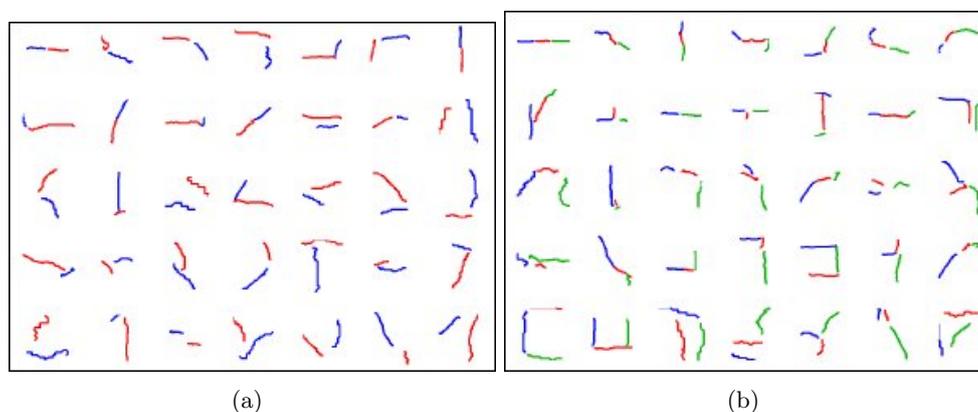


FIG. 2.7 – Les 35  $kAS$  les plus fréquents du dictionnaire de codes, dans Ferrari *et al.* [2008]. (a).  $k = 2$ , (b).  $k = 3$ .

Les  $kAS$  ont les bonnes qualités de couvrir des chaînes de contours très courtes, d'avoir une complexité intermédiaire, d'être basés sur la connexité des segments (les segments connexes ont plus de chance de se trouver sur les frontières d'un objet) et d'avoir des descripteurs invariants localement. Il est intéressant de voir que les dictionnaires de codes construits à partir d'images différentes sont assez similaires. Cela sert à montrer la généralité des  $kAS$ .

### 2.7.2.2.b Les primitives géométriques

Une alternative pour obtenir des descripteurs à partir des contours est de rechercher des formes qui ont des propriétés géométriques bien définies. Les objets construits par l'homme sont en général constitués de composants qui ont des formes géométriques simples, contrairement aux objets naturels qui ont des formes plus irrégulières. Les formes géométriques les plus générales sont les segments de droites, les rectangles et les cercles.

La transformée de Hough [Duda et Hart, 1972] est une méthode pour détecter les formes simples à partir d'un ensemble de points. L'application la plus classique est la détection des segments de droites. Le principe est de représenter les segments entre chaque paire de points dans un espace de coordonnées défini par l'angle et la distance du segment à l'origine et d'accumuler les nouvelles coordonnées dans une matrice. Les maxima locaux de cette matrice indiquent les segments. On peut appliquer la transformée de Hough également pour détecter les cercles. Cette fois-ci on fait la transformation vers un espace à trois coordonnées, position  $x$ , position  $y$  et le rayon. La recherche des cercles dans cet espace est difficile surtout quand on ne connaît pas les rayons des cercles recherchés.

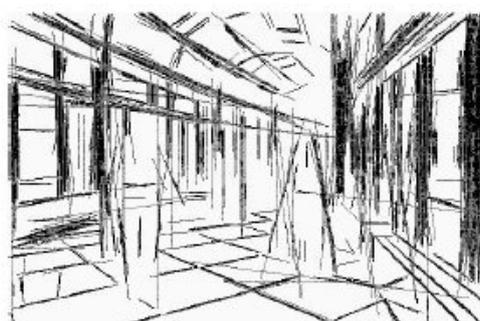
La transformée de Hough traite les contours comme un ensemble de points indépendants. A la fin de l'étape de l'extraction des contours on dispose souvent des informations sur les connections entre les points de contours et la direction du gradient sur ces points. On peut exploiter ces informations pour une meilleure extraction des primitives. Venkateswar et Chellappa [1992] ont détecté les segments de droites dans les images aériennes en parcourant l'image pour affecter une étiquette de segment à chaque point de contour d'après la direction du gradient à ce point. On regroupe ensuite les points qui ont les mêmes étiquettes et on détermine les extrémités. Dans une méthode similaire, Desolneux *et al.* [2000] utilisent le principe d'évènement significatif : dans une image, les directions des gradients de deux points voisins sont des variables aléatoires. Les déviations de ce caractère aléatoire sont considérées comme des événements significatifs qui correspondent aux structures. Ils ont calculé la direction du gradient très localement, dans un voisinage de  $2 \times 2$  et recherché les événements significatifs. Ils ont utilisé leur méthode pour détecter les structures linéaires dans les images (Figure 2.8).

Une autre approche est l'utilisation des algorithmes d'approximation polygonales qui transforment une chaîne de contours en une suite de segments de droites. Garnesson et Giraudon [1992] présentent une vue d'ensemble du problème et les méthodes utilisées. Deux algorithmes polygonaux classiques par découpage sont l'algorithme de la corde [Douglas et Peucker, 1973] et l'algorithme de Dunham [Dunham, 1986]. On commence par un segment grossier qui représente la chaîne de contour, on trouve le point de découpage (qui satisfait les critères spécifiques à l'algorithme) et on divise le segment en deux par ce point. On répète cette étape récursivement jusqu'à ce qu'on obtienne des segments avec une erreur d'approximation acceptable.

Iqbal et Aggarwal [2002] ont proposé d'extraire d'autres primitives par les principes de regroupement perceptuel des segments de droites, la proximité, la similarité, la continuation, la



(a)



(b)

FIG. 2.8 – Les segments linéaires par la détection des événements significatifs, dans Desolneux *et al.* [2000].

fermeture et le symétrie. Les primitives obtenus sont les coterminations, les jonctions L et T (Les jonctions L sont les paires de segments qui s'articulent à leurs extrémités et les jonctions T sont les paires de segments où l'articulation est entre l'extrémité d'un segment et le milieu de l'autre), les segments parallèles, les groupes parallèles et les polygones.

### 2.7.3 Les descripteurs régions

L'objectif de ces approches est de détecter dans l'image les régions qui appartiennent à l'objet et qui ont une forme particulière. On peut ensuite représenter les régions détectées par un ensemble de descripteurs de formes.

#### 2.7.3.1 La segmentation de l'image

La segmentation est la partition d'une image en zones homogènes. Elle a été depuis toujours un des principaux problèmes de la recherche en vision par ordinateur et de nombreux algorithmes de segmentation ont été développés. La segmentation étant un processus hautement subjectif, et dépendant des images, de l'application et des connaissances *a priori*, une méthode de segmentation universelle qu'on peut appliquer sur toutes les images n'existe pas. Il est nécessaire de trouver la méthode qui correspond le mieux au problème qu'on traite. Nous voulons donner une

taxonomie générale des méthodes de segmentation. Pour une présentation détaillée des méthodes de segmentation on peut consulter [Haralick et Shapiro, 1985; Serra, 2003].

On effectue en général la segmentation en utilisant quatre approches principales :

- Les approches globales utilisent l’histogramme des intensités des pixels et appliquent un seuillage pour classifier les pixels de l’image. Ces méthodes sont simples à implémenter, pourtant elles ne sont pas robustes au bruit parce qu’on n’utilise pas l’information spatiale.
- Les méthodes par modèle ou motif (*template*) recherchent dans l’image des régions qui correspondent à des motifs prédéfinis. Les approches les plus simples utilisent des filtres qui déterminent les motifs recherchés dans l’image [Brunelli et Poggio, 1997; McQueen, 1981]. Il est aussi possible d’utiliser des modèles déformables [Jain *et al.*, 1996].
- Les approches frontières se servent des contours de l’objet pour détecter une région. L’extraction des contours est suivie des algorithmes de fermeture des contours par la relaxation des contours, recherche dans un graphe des contours ou la programmation dynamique. Il est pourtant difficile d’obtenir une segmentation complète de l’image en utilisant ces approches.
- Les méthodes de *clustering*, divisent l’image en régions en regroupant les pixels voisins qui ont des niveaux d’intensités proches ou qui satisfont d’autres critères numériques.

### 2.7.3.2 La segmentation multi-échelle

Nous nous intéressons particulièrement aux méthodes de *clustering* qui donnent la possibilité de segmenter une image en plusieurs niveaux d’échelles : en haut de l’échelle, on obtient une segmentation grossière avec peu de régions et en bas de l’échelle une segmentation détaillée avec un grand nombre de régions. On ne peut pas parler d’une échelle unique de segmentation qui représente une image. Les régions obtenues aux différentes échelles correspondent aux objets ayant des tailles plus petites ou plus grandes.

Les approches ascendantes utilisent une sur-segmentation de l’image et fusionnent itérativement les régions d’après une métrique de similarité. On utilise le plus souvent la similarité radiométrique des pixels comme métrique. On peut également utiliser la texture, la couleur, ou les propriétés géométriques des régions [Chehata, 2005].

Pour initialiser la fusion, on peut considérer chaque pixel de l’image comme une région. Pourtant avec cette approche la fusion dure considérablement longtemps, surtout si la mesure de similarité utilisée n’est pas simple. Galland *et al.* [2003] propose d’utiliser des régions appartenant à un quadrillage régulier comme germes et de les fusionner par le principe de longueur de description minimale (*Minimum Description Length*).

La segmentation par les lignes de partage des eaux (*Watershed*) développée par Meyer et Beucher [1990] est une méthode de segmentation populaire pour obtenir une sur-segmentation initiale. Le gradient de l’image est considéré comme une surface topographique et une analogie

d'immersion d'eaux est appliquée. Les lignes de rencontre des eaux des minima divisent l'image en régions homogènes. La figure 2.9 illustre l'algorithme *Watershed*.

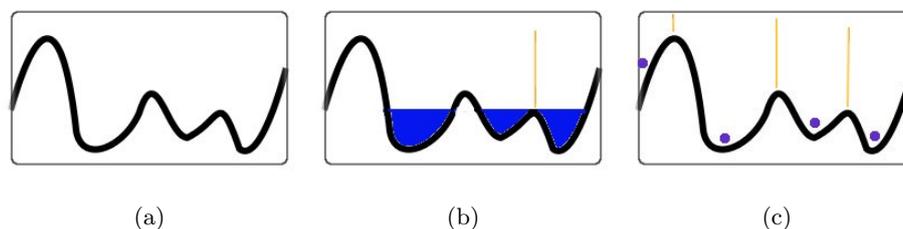


FIG. 2.9 – Illustration de l'algorithme *Watershed* sur un signal 1D. (a). L'image, (b). L'immersion et la première ligne de partage, (c). Segmentation finale, on obtient 4 régions (les bassins marqués par les points) séparées par 3 lignes de partage.

Pour réduire le nombre de régions initiales, une approche proposée par Beucher *et al.* [1990] est d'utiliser des marqueurs internes et externes qui indiquent respectivement l'objet et le fond dans l'image, et de modifier l'image avec un pré-traitement tel qu'elle n'exhibe des minima locaux que sur les marqueurs. De cette façon, quand on applique l'algorithme *Watershed*, le nombre de régions est réduit au nombre de marqueurs. La difficulté principale de cette approche est de trouver des marqueurs pertinents. On utilise souvent une intervention humaine pour détecter les marqueurs.

Une alternative efficace est la segmentation *Mean-shift* [Comaniciu et Meer, 2002], une adaptation de l'algorithme de *clustering Mean-shift* pour la segmentation. *Mean-shift* (Cheng [1995]) est un algorithme de *clustering* des données par l'estimation de leur densité. Dans une fenêtre de recherche de taille constante on trouve le centroïde (position moyenne des données) et la fenêtre est centrée sur le centroïde. Cette procédure est répétée jusqu'à la convergence et le point de convergence est considéré comme le centre de *cluster* pour les données visitées. Pour la segmentation l'image est considérée comme une lattice bi-dimensionnelle des valeurs des pixels, et l'algorithme *mean-shift* est appliqué pour le domaine joint spatial-intensité. L'avantage de cette méthode est la possibilité d'ajuster le nombre de régions en choisissant la taille appropriée pour la fenêtre de recherche.

### 2.7.3.3 Les descripteurs de formes

Une fois l'image segmentée, on peut représenter les régions obtenues par des descripteurs de formes. Une analyse très détaillée des descripteurs de formes est donnée dans Zhang et Lu [2004]. On représente une forme comme une région ou par ses contours pour extraire les descripteurs.

#### 2.7.3.3.a Les descripteurs obtenus par une approche région

Pour calculer ces descripteurs la forme est considérée comme l'ensemble des pixels qui constituent sa région interne.

**Les descripteurs géométriques simples :**

- L’aire, le périmètre ;
- La compacité ou la circularité : le rapport du carré du périmètre à l’aire de la forme  $(\frac{Périmètre^2}{4*\pi*Aire})$ . La forme la plus compacte est un cercle avec une compacité égale à 1 ;
- La rectangularité : c’est le rapport de l’aire de la forme à l’aire du plus petit rectangle englobant ;
- L’excentricité : la mesure de la circularité d’une ellipse. Pour une ellipse c’est le rapport de la distance entre les foyers à la longueur de l’axe majeur. On calcule l’excentricité de l’ellipse qui a les mêmes moments d’ordre deux que la forme ;
- L’élongation : le rapport de la longueur à la largeur du plus petit rectangle englobant ;
- Les axes principaux : l’axe majeur et l’axe mineur de l’ellipse qui a les mêmes moments d’ordre deux que la forme ;
- L’orientation : la direction de l’axe majeur.

**Les descripteurs topologiques :** Deux formes sont topologiquement identiques si on peut obtenir une en déformant l’autre par des transformations continues. Les descripteurs topologiques de base sont le nombre de composants connexes et les trous. Le plus connu de ces descripteurs est le nombre d’Euler qui est égal au nombre de composants connexes moins le nombre de trous.

**L’enveloppe convexe :** Une région  $R$  est convexe si pour chaque paire de points  $p_1, p_2$  de  $R$  le segment  $[p_1p_2]$  est dans  $R$ . L’enveloppe convexe est la plus petite région convexe qui contient la forme.

**Les profils :** On peut obtenir une signature de l’objet en le projetant sur l’axe horizontal ou vertical et en comptant le nombre de pixels sur chaque colonne ou ligne. On utilise ces descripteurs le plus souvent pour la reconnaissance des caractères [Kavallieratou *et al.*, 2002].

**Les moments :** La fonction d’intensité de la forme peut être représentée comme une densité de probabilité d’un variable aléatoire bi-dimensionnelle. Les moments géométriques décrivent les caractéristiques statistiques de cette variable aléatoire :

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

**Les moments centraux** sont des moments géométriques invariants à la translation. On les obtient en utilisant comme origine le centre de gravité de la forme  $(x_c, y_c)$  où  $x_c = \frac{m_{10}}{m_{00}}$  et  $y_c = \frac{m_{01}}{m_{00}}$ . En normalisant les moments centraux, on calcule **les moments centraux normalisés** qui sont invariants à l’échelle et à la translation. Hu [1962] a défini 7 moments invariants à la rotation à partir des combinaisons des moments centraux normalisés jusqu’à l’ordre trois. **Les moments de Zernike** [Khotanzad et Hong, 1990] sont construits en utilisant une famille de polynômes

complexes qui forment une base orthogonale définie sur le disque unité. Ils sont invariants à la translation, à l'échelle et à la rotation.

**Le squelette :** Le squelette de l'objet est un ensemble de segments connexes d'épaisseur un pixel obtenu par l'amincissement d'une forme. On l'appelle également l'axe médian, l'ensemble des points qui ont la même distance minimale aux frontières de la forme pour au moins deux points de frontière distincts. On peut obtenir le squelette d'une forme par les opérateurs de la morphologie mathématique ou par une transformée de distance qui associe à chaque pixel la valeur de la distance minimale à la frontière.

### 2.7.3.3.b Les descripteurs obtenus par une approche contours

Ce sont des descripteurs obtenus à partir des contours externes de l'objet.

#### Les descripteurs géométriques simples :

- Le périmètre ;
- La courbure : le contour est représenté par l'angle de la tangente à chaque point de contour. On peut ensuite calculer la dérivée des courbures sur chaque point de contour ou on peut détecter les courbures importantes du contour comme descripteur ;
- L'énergie de flexion (*bending*) : la somme des carrées des courbures à chaque point de contour ;
- La distribution des cordes : une corde est un segment de droite qui joint deux points de contours. On peut mesurer la distribution des cordes par les histogrammes des longueurs et des angles de toutes les cordes du contour.

**Le codage de chaîne :** A partir d'un point de contour initial, on traverse le contour en codant les points consécutifs par la direction du déplacement. Les chiffres de 0 à 7 sont utilisés pour coder les 8 directions principales. Le codage est plus économique que représenter le contour par les coordonnées des points de contours, et il est invariant à la translation. Pour assurer l'invariance à la rotation on utilise le codage de chaîne différentiel en calculant les différences des directions des déplacements consécutifs en mode 8. La sélection du point de départ qui donne le code avec la plus petite valeur numérique rend le code invariant à la sélection du point de départ.

**Approximation polygonale :** Les contours de la forme peuvent être représentés par des segments de droites obtenus par une approximation polygonale (section 2.7.2.2.b).

**Descripteurs Fourier :** Si on considère les  $n$  points du contour consécutifs comme une fonction discrète  $\mathbf{x}(n) = (x_1(n), x_2(n))$ , on peut définir une fonction discrète complexe  $u(n) =$

$x_1(n) + jx_2(n)$ ). Cette fonction peut être transformée dans le domaine fréquentiel par une transformée de Fourier discrète. Les coefficients de cette transformation, appelés les descripteurs de Fourier sont utilisés pour représenter la forme.

**Shape contexts :** C'est une façon de représenter le contour par rapport à un point de contour choisi. On prend  $n$  points sur le contour. Pour chaque point  $p_i$  on construit les vecteurs qui joignent  $p_i$  à ces points. Le shape context de  $p_i$  est l'histogramme des coordonnées relatives des  $n - 1$  points.

#### 2.7.4 Les descripteurs de texture et de couleur

La couleur est probablement le plus expressif mais le moins objectif parmi tous les descripteurs de l'image. La couleur sur les images numériques est représentée par un espace colorimétrique qui associe des nombres aux couleurs. Par exemple l'espace de couleurs RBV(Rouge-Bleu-Vert) est défini par les intensités des trois couleurs principales rouge, vert et bleu. L'espace TSV est définie en fonction des trois composantes teinte, saturation et valeur (brillance) de la couleur. Les descripteurs les plus fréquemment utilisés sont basés sur des histogrammes des couleurs qui caractérisent la distribution des couleurs dans une image. Après la sélection d'un espace colorimétrique, d'une quantification de l'espace de couleurs et d'une quantification de l'espace des valeurs de l'histogramme, on construit l'histogramme en comptant le nombre de couleurs de chaque pixel. Les descripteurs standards définis par le groupe MPEG7 pour limiter le nombre de descripteurs et assurer l'interopérabilité des différents systèmes sont présentés dans [Cieplinski, 2001].

Les descripteurs de textures sont utilisés dans plusieurs domaines du traitement d'image, comme l'imagerie médicale, le traitement des documents et la télédétection. Contrairement à la couleur, la texture n'est pas la propriété d'un pixel mais d'une région. Définir exactement ce qu'est la texture n'est pas facile (on peut voir [Tuceryan et Jain, 1998] pour une discussion sur les différentes définitions de la texture). Par conséquent, il existe un nombre très important de travaux pour définir, segmenter et classifier les textures, groupés en 4 approches principales. Dans l'approche structurelle, on distingue deux éléments principaux, les primitives à utiliser et les relations spatiales qui lient les primitives entre elles. Dans l'approche statistique, la texture est vue comme la réalisation d'un processus stochastique. Le but est alors d'en extraire des attributs statistiques. Les données sources de ces attributs peuvent être les pixels eux-mêmes (cas le plus classique, premier ordre) ou des couples de pixels, comme c'est le cas pour la matrice de co-occurrence [Haralick, 1979] ou l'auto-corrélation (mesures du second ordre). L'approche modèle repose sur des modèles stochastiques. Les paramètres du modèle sont estimés et utilisés pour l'analyse de la texture. Dans la pratique ces méthodes sont relativement coûteuses en temps de calcul. Dans l'approche s'appuyant sur les transformées, on cherche à représenter la texture dans une autre base que le domaine spatial de l'image. Le but est alors le plus souvent de trouver

des vecteurs de la base qui soient les plus porteurs d'information pour reconnaître la texture. On retrouve notamment dans cette catégorie les méthodes s'appuyant sur la transformée de Fourier, les filtres de Gabor et les ondelettes.

## 2.8 Les traitements de haut niveau : reconnaissance d'objets

Au début des travaux de recherche en reconnaissance d'objets dans les années 1970 les représentations géométriques ont joué un rôle important. On a souvent ciblé la reconnaissance d'objets rigides, et la géométrie, extraite à partir des contours de l'image, a fourni une base théorique solide pour gérer les variations d'illumination et de perspective de ces objets. L'approche **centrée sur l'objet** de Marr [1980] au début des années 1980 a eu une influence majeure sur les travaux de recherche pendant deux décennies. Selon Marr la reconnaissance passe par la reconstruction de la structure tri-dimensionnelle de l'objet physique à partir des différentes vues bi-dimensionnelles de l'objet. Cette approche a donné lieu à des méthodes fondées sur l'identification de la structure de l'objet par la correspondance entre les régions et les contours extraits de l'image et un modèle hypothétique d'objet tri-dimensionnel.

### 2.8.1 L'apprentissage structurel

**L'interprétation d'images** (*image understanding*) [Crevier et Lepage, 1997] est un domaine de recherche devenu très populaire dans les années 90, qui unit deux disciplines proches, le traitement d'images et l'intelligence artificielle. Elle vise principalement à associer les régions d'une image numérique à des objets sémantiques pour obtenir une interprétation de haut niveau du contenu d'une image en essayant d'intégrer des connaissances externes (représentées par exemple par des ontologies) dans le traitement. **La reconnaissance de formes structurelle** est souvent vue comme un moyen de combler le fossé sémantique entre les descripteurs de bas niveau et les concepts de haut niveau. Elle est caractérisée par la représentation des formes en terme de structures de données symboliques pour révéler une description de la structure d'un objet. Cette description est en général basée sur les attributs des composants de l'objet et leurs relations dans l'espace représentés par un modèle structurel d'objet, et la méthode principale de la reconnaissance de formes structurelle est l'appariement des structures.

La représentation par des graphes est principalement utilisée pour la recherche d'images dans des bases d'images de grande taille. Petrakis et Faloutsos [1995] proposent une méthode de modélisation et indexation en utilisant des graphes relationnels attribués (*ARGs*) pour trouver les images similaires dans une base de donnée images par une métrique de distance entre deux *ARGs*. [Shao et Kittler, 1999] présentent une représentation projective invariante aux transformations pour les *ARGs* utilisée pour l'appariement des *ARGs* des objets soumis à des transformations géométriques. Bardin et al. [2000] utilisent le squelette d'une image pour le représenter par un *ARG*. On applique un algorithme d'appariement de graphes pour détecter la correspondance

entre les graphes. Dans [Di Ruberto, 2004] on utilise les *ARGs* issus de squelettes pour détecter les objets d'une même classe en utilisant une technique d'optimisation non-linéaire pour appairer les graphes. Le système est testé sur un problème de classement d'images binaires d'objets simples. Siddiqi *et al.* [1999] et Sebastian *et al.* [2004] utilisent les graphes-chocs, une représentation basée sur l'axe médial de l'objet, pour détecter la similarité de deux objets. Mauro *et al.* [2003] représentent les objets par des graphes acycliques orientés et ordonnés. On utilise ensuite un réseau de neurones récuratif pour obtenir un vecteur qui caractérise l'image.

La perspective de certains travaux est d'apprendre un modèle structurel représentatif des différents objets d'une même classe : dans [Pope et Lowe, 1993] on apprend à modéliser l'apparence d'un objet en représentant les variations de l'apparence par une distribution de probabilité. Dans [Hong et Huang, 2004] le but est l'apprentissage automatique d'un modèle paramétrique pour modéliser un ensemble d'échantillon représenté par des *ARGs*. [Sangineto, 2003] décrit une classe d'objet par un modèle qui représente les invariants de formes utilisé pour classifier des objets par un algorithme de satisfaction de contraintes. Dans [Cordella *et al.*, 2002] on présente une méthode d'apprentissage d'un modèle générique pour des figures représentées par des *ARGs*.

On effectue la modélisation souvent pour des problèmes jouets avec des objets très simples ou synthétiques (un exemple est donné dans la figure 2.10). Les modèles obtenus sont simples et souvent trop rigides pour permettre la reconnaissance sur les images réels.

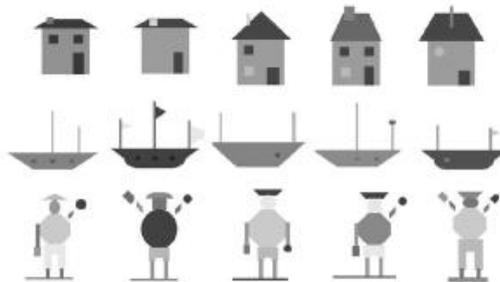


FIG. 2.10 – Les images d'apprentissage artificielles créées par une grammaire, dans Cordella *et al.* [2002].

### 2.8.2 L'approche fondée sur l'apparence

Depuis plus d'une dizaine d'années on est confronté à un changement majeur de paradigme vers les approches **fondées sur l'apparence**, dans lesquelles un objet est représenté comme une collection d'apparences sur les images prises selon différents points de vue et on vise à apprendre cette représentation à partir d'exemples par une approche statistique en négligeant la forme de l'objet physique tri-dimensionnelle.

On peut utiliser des descripteurs globaux ou locaux pour révéler l'apparence d'un objet sur une image. La détection des visages en utilisant les *eigenfaces* est **l'approche globale** la plus remarquable [Turk et Pentland, 1991]. L'idée principale est de projeter les images de visages sur

l'espace des caractéristiques qui encode le mieux les variations entre les images. Cet espace est défini par les *eigenfaces*, les vecteurs propres de l'ensemble des images de visages. On calcule les *eigenfaces* par l'analyse en composantes principales (ACP), une méthode de réduction de dimension bien connue en analyse de données.

Les approches globales, bien qu'elle soient rapides, ne sont pas adaptées pour la reconnaissance des objets non-rigides qui ont subi des déformations, et des objets sur des fonds complexes. Les méthodes récentes sont fondées sur **une approche locale**, le plus souvent sur les régions locales significatives de l'image dont l'apparence reste stable sur plusieurs images du même objet. Un objet est souvent conçu comme un ensemble de ses composants qui peuvent être décrits par des caractéristiques (Figure 2.11). La reconnaissance est alors formulée comme un problème d'apprentissage supervisé : on extrait les caractéristiques de l'objet à partir d'un ensemble d'images d'apprentissage positives et négatives pour construire un classifieur par une approche statistique.



FIG. 2.11 – La conception d'objet dans les approches basées sur l'apparence.

### 2.8.3 Les algorithmes d'apprentissage

Le problème de la reconnaissance d'objets a une relation étroite avec le domaine de l'apprentissage automatique. On vise à construire, à partir d'un ensemble de descripteurs, un classifieur qui va affecter une image dans une des deux classes "objet" et "non-objet". Dans le cas de l'apprentissage supervisé, le problème de l'apprentissage est d'estimer une fonction de décision (ou la probabilité *a posteriori*  $P(y|\mathbf{x})$ ) en utilisant un ensemble d'apprentissage étiqueté  $\{\mathbf{x}_i, y_i\}$  où  $\mathbf{x}$  est un vecteur de caractéristiques (les données) et  $y$  est la classe associée (les étiquettes). On peut faire la distinction entre deux approches principales d'apprentissage :

- **Dans l'approche discriminatoire**, l'objectif est d'apprendre directement une fonction de décision qui sépare les deux classes. Dans le domaine probabiliste cela correspond à calculer directement les probabilités *a posteriori*  $P(y|\mathbf{x})$ . On introduit un modèle paramétrique qui exprime les probabilités *a posteriori* et on obtient les valeurs des paramètres à partir des données étiquetées. Les méthodes les plus classiques sont la régression logistique, les k-plus proches voisins, l'apprentissage par des réseaux de neurones, les machines à vecteurs de support et le *boosting*.
- **Dans l'approche générative**, par contre, on vise à construire un modèle pour chaque classe et on classe un objet d'après sa ressemblance au modèle. On modélise la distribution jointe  $p(\mathbf{x}, y)$  des données et des étiquettes, en apprenant les probabilités de classe *a priori*  $p(y)$  et les densités conditionnelles de classe  $p(\mathbf{x}|y)$ . On classe un échantillon  $x$  en calculant la probabilité *a posteriori* par l'application de la règle de Bayes.

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}|y)p(y)}{p(x)}$$

Les méthodes d'apprentissage génératives les plus fréquemment utilisées sont le classifieur naïf de Bayes, le modèle de Markov caché et l'algorithme espérance-maximisation.

Les nombreuses méthodes de reconnaissance d'objets reposent souvent sur quelques familles de méthodes principales d'apprentissage automatique :

**Les machines à vecteurs de support (*Support Vector Machines, SVM*) :** Les machines à vecteurs de support ou séparateurs à vaste marge est une technique assez récente et très populaire de classification supervisée binaire [Vapnik, 1998; Burges, 1998]. L'idée principale est de trouver l'hyperplan optimal qui sépare les données en deux classes. Cela correspond à trouver les frontières qui séparent les échantillons étiquetés en deux classes avec la distance maximale. Les frontières qui satisfont ce critère sont appelées les vecteurs supports. Le problème de trouver les vecteurs supports est résolu en le formulant comme un problème d'optimisation quadratique. Le plus simple des classifieurs statistiques est un classifieur linéaire qui trouve un hyperplan qui sépare les deux classes obtenu par une combinaison linéaire des caractéristiques. Pourtant, sauf dans le cas des problèmes jouets, les classes ne sont pas linéairement séparables et on a besoin d'utiliser des approches plus sophistiquées. Le grand avantage du *SVM* est sa capacité de classifier les données qui ne sont pas linéairement séparables à l'aide des noyaux, en utilisant une méthode appelée *kernel trick*. On projette les données en utilisant un noyau dans un espace de dimension plus élevée (éventuellement infinie) où la séparation linéaire est possible. L'utilisation des noyaux réduit les calculs dans cet espace à des produits vectoriels simples à calculer. La sélection du noyau approprié est en général réalisée en testant les différents noyaux.

**Boosting :** *Boosting* est une technique d'apprentissage par la combinaison des classifieurs binaires. L'algorithme le plus souvent utilisé en reconnaissance d'objets est *Adaboost* (*adaptive boosting*) [Freund et Schapire, 1995]. On l'utilise pour sélectionner un sous-ensemble des descripteurs à partir d'un ensemble d'images positives et un ensemble d'images négatives. L'idée principale de l'algorithme *Adaboost* est de combiner une collection de classifieurs faibles pour construire un classifieur fort. On appelle "faible" un classifieur simple qui ne classifie pas bien l'ensemble d'apprentissage même avec la meilleure fonction de classification. Chaque caractéristique unique peut être considérée comme un classifieur faible. Dans ce cas l'apprentissage du classifieur faible correspond à la détection d'un seuil et d'une parité qui indique la direction du symbole de l'inégalité qui séparent le mieux les exemples positifs et négatifs. A chaque itération d'*Adaboost*, on trouve le classifieur faible optimal et on calcule l'erreur de classification. En fonction de l'erreur, on calcule le poids du classifieur faible, et on met à jour les poids des données (initialement égaux) de telle sorte que les poids des exemples mal classifiés soient augmentés et les poids des exemples bien classifiés soient réduits. Cela garantit de trouver à l'itération suivante

un autre classifieur faible qui classe mieux les exemples mal-classifiés. Le classifieur fort final est constitué de l'ensemble des classifieurs faibles sélectionnés, pondérés par leur poids.

**Le classifieur naïf de Bayes :** C'est un classifieur probabiliste basée sur l'application de la règle de Bayes. L'inconvénient de l'application directe de la règle de Bayes est sa complexité à cause du nombre élevé de paramètres indépendants qu'il faut calculer. Le classifieur naïf de Bayes fait la supposition que les attributs  $x_1, \dots, x_n$  sont tous mutuellement indépendants. On obtient alors

$$p(x_1, \dots, x_n | y) = \prod_{i=1}^n p(x_i | y)$$

et on réduit le nombre de paramètres considérablement.

**Espérance-Maximisation (*Expectation-Maximization, EM*) :** L'*EM* est une méthode itérative utilisée pour trouver le maximum de vraisemblance des paramètres d'un modèle en présence de données incomplètes [Dempster *et al.*, 1977]. Elle est bien appropriée pour la reconnaissance d'objets où on a souvent des données manquantes à cause de l'insuffisance de la segmentation ou de l'extraction des primitives. L'algorithme *EM* alterne une étape d'estimation de paramètres et une étape de maximisation itérativement. A l'étape *E* on calcule l'espérance de la vraisemblance en utilisant les données inconnues comme si elles étaient connues, et à l'étape *M* on estime le maximum de vraisemblance des paramètres en maximisant l'espérance calculée à l'étape *E*. On répète les deux étapes itérativement jusqu'à la convergence des paramètres.

## 2.8.4 La représentation de l'objet

Le modèle d'objet utilisé pour la reconnaissance peut avoir des structures très variables, principalement en fonction de la représentation des relations entre les composants de l'objet. Aux deux extrêmes se situent **les approches fondées sur l'apparence seule** où on néglige toute sorte de relations entre les composants et **les approches fondées sur l'apparence et la structure** où on représente également les positions et les arrangements spatiaux des composants, généralement par différents types de graphes.

### 2.8.4.1 Les modèles par l'apparence seule

L'approche par sac de mots est un modèle efficace pour la reconnaissance générique d'objets. C'est un modèle initialement utilisé pour la classification des textes où on représente un texte uniquement par les mesures statistiques du vocabulaire utilisé. On construit un vecteur de caractéristiques en comptant les mots dans le texte sans prendre en compte ni le grammaire ni l'ordre des mots. Pour deux textes de même catégorie la probabilité de trouver les mêmes mots

est élevée et par conséquent les vecteurs de caractéristiques sont similaires. Cette approche est utilisée pour la première fois en vision par Leung et Malik [2001] pour la reconnaissance des images de textures. Csurka *et al.* [2004] ont utilisé le modèle sac de mots pour la reconnaissance d'objets dans les images en établissant une analogie directe entre le vocabulaire d'un texte et les régions locales significatives d'une image. Le modèle est fondé uniquement sur l'apparence des composants locaux de l'objet. La figure 2.12 illustre l'idée principale de cette approche.

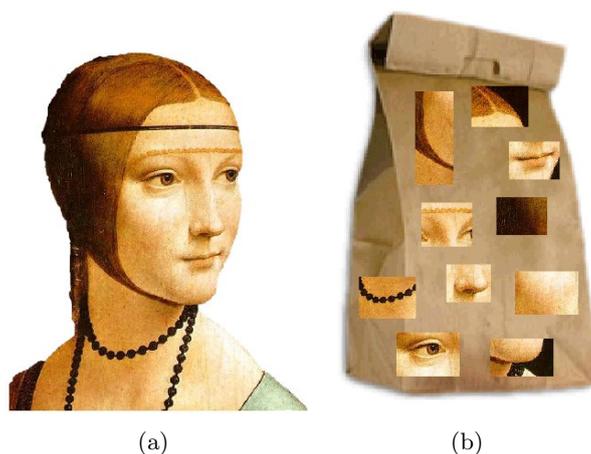


FIG. 2.12 – Illustration du modèle Sac de Mots. (a). L'objet, (b). Le sac de mots

Pour construire le dictionnaire de codes on peut extraire des régions locales par les détecteurs de points d'intérêt (on utilise le plus souvent le détecteur SIFT) [Csurka *et al.*, 2004; Sivic *et al.*, 2005], on peut utiliser les régions obtenues par un quadrillage régulier [Li et Perona, 2005; Vogel et Schiele, 2002] ou une segmentation [Barnard *et al.*, 2003]. Le nombre des régions qu'on peut extraire à partir d'une image étant trop important, l'extraction est suivie d'une étape de *clustering* pour grouper les descripteurs similaires en un même mot (Figure 2.13).



FIG. 2.13 – Les 56 premiers mots du sac de mots obtenu par un quadrillage régulier à partir de 650 images dans 13 catégories, dans Li et Perona [2005]. Les mots sont triés dans l'ordre décroissant d'après le nombre d'apparition dans les images.

Pour la classification on utilise le plus souvent les approches génératives probabilistes. [Csurka *et al.*, 2004] ont utilisé le classifieur naïf de Bayes qui a donné de très bons résultats de classification malgré la supposition forte d'indépendance.

Li et Perona [2005], et Sivic *et al.* [2005] ont utilisé les approches bayésiennes hiérarchiques, *probabilistic latent semantic analysis* (pLSA) et *latent Dirichlet allocation* (LDA), qui ont été

développées initialement pour l'analyse de texte. Dans ces approches on propose un modèle probabiliste hiérarchique où l'image est constituée de thèmes et les thèmes sont constitués de mots, et on apprend les paramètres de ce modèle hiérarchisé pour classifier une image.

Il est aussi possible de classifier les images par une approche discriminatoire comme le SVM [Csurka *et al.*, 2004; Zhang *et al.*, 2007].

Opelt *et al.* [2004] ont proposé d'extraire un ensemble de descripteurs locaux à partir de l'image en utilisant les points d'intérêt de Harris, les points d'intérêt *SIFT*, et deux méthodes de segmentation, la première par *Mean-Shift* et la deuxième par une mesure de similarité. Un classifieur *Adaboost* a été utilisé pour construire les hypothèses finales à partir d'ensemble de descripteurs.

Dorkó et Schmid [2003] ont utilisé des descripteurs Harris-Laplace et *DoG* invariants à l'échelle. Ils ont groupé les descripteurs similaires et appliqué deux types de classifieurs, les *SVM* et le modèle de mixture gaussienne pour apprendre les descripteurs les plus significatifs.

Les modèles par l'apparence seule sont indifférents aux positions des composants et à leurs relations spatiales. Cela apporte certes une classification rapide et robuste aux occlusions. Mais, d'un autre côté, le modèle n'est pas capable de distinguer les images d'objets qui possèdent des composants similaires mais avec des configurations spatiales différentes.

#### 2.8.4.2 Les modèles par l'apparence et la structure

Ces modèles visent à apporter une solution au problème mentionné ci-dessus. Elles prennent en compte en même temps l'apparence des composants et la forme de l'objet. L'apparence est représentée par les descripteurs des composants locaux et la forme par les positions relatives des composants et leurs relations spatiales qui sont exprimées en général par des graphes. Cette approche est une combinaison des approches fondées sur l'apparence pure et les approches géométriques. Contrairement à un sac de mots qui contient plusieurs régions locales, on représente l'objet par un nombre réduit de régions significatives.

On peut utiliser différentes structures de graphes qui se distinguent par le type de connections entre les composants. Ce sont principalement les graphes complets, les structures en étoile ou les arbres (Figure 2.14). La représentativité et la complexité du modèle augmentent en fonction du nombre de connections.

##### 2.8.4.2.a *Pictorial Structures*

L'idée de représenter un objet par l'ensemble de ses composants organisés dans une configuration déformable a été initialement proposée par Fischler et Elschlager [1973]. La configuration des composants est représentée par des connections élastiques (Figure 2.15). La reconnaissance est effectuée en minimisant une fonction de coût calculé par la somme du coût d'appariement des composants et le coût de la déformation par rapport au modèle.

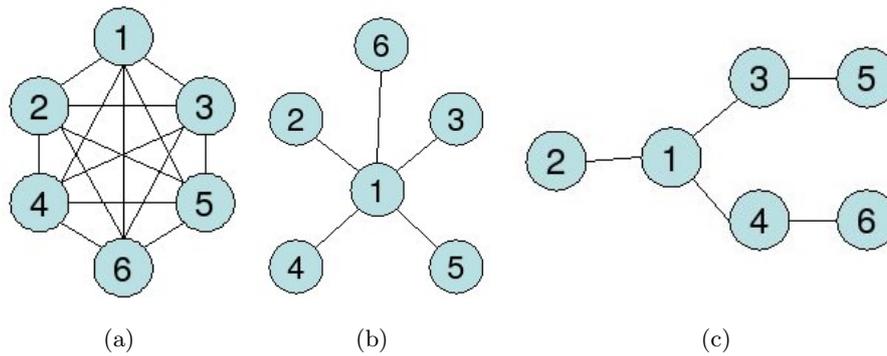


FIG. 2.14 – Les structures de graphes. (a). Graphe complet, (b). Etoile, (c). Arbre.

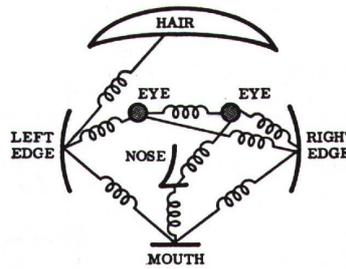


FIG. 2.15 – La représentation d'un visage par le modèle *Pictorial Structures*, dans [Fischler et Elschlager, 1973] (*hair* : cheveux, *eye* : oeil, *nose* : nez, *left-right edge* : côté gauche-droite, *mouth* : bouche).

Récemment Felzenszwalb et Huttenlocher [2005] ont repris cette idée pour reconnaître les personnes en mouvement, en représentant le corps humain par des composants articulés. Leur approche se concentre exclusivement sur l'apprentissage de la structure par l'utilisation des images d'apprentissage sur lesquelles les composants du corps sont segmentés et étiquetés.

#### 2.8.4.2.b Le modèle de constellation

C'est un modèle similaire au modèle *Pictorial Structures*. La différence principale est la représentation de la structure par un graphe en forme d'étoile : la structure de l'objet est représentée par les positions des composants relatives à un point constant, souvent **le centre de l'objet**. On utilise généralement un cadre probabiliste pour exprimer cette représentation.

Une approche évidente est d'apprendre d'abord les composants de l'objet et puis la structure.

Dans Weber *et al.* [2000], dans une première étape d'apprentissage des composants, on détecte les régions autour des points d'intérêt et on les regroupe par l'algorithme de *clustering* k-moyennes pour réduire le nombre des composants. Le modèle d'objet est composé des types des composants, de leurs positions et de leur appartenance à l'objet ou au fond. On apprend les paramètres du modèle en utilisant l'algorithme *EM* appliqué itérativement sur des différentes configurations des composants.

L'étape d'apprentissage des composants est très similaire dans Leibe *et al.* [2004]. On utilise un détecteur de Harris [Harris et Stephens, 1988]. On regroupe les régions itérativement d'après une métrique de similarité obtenue par la corrélation normalisée au niveau de gris (*Normalized Greyscale Cross Correlation*), et on les trie par ordre décroissant du nombre de régions dans chaque *cluster*. La figure 2.16.(a) montre le dictionnaire de codes obtenu à partir des images de voitures vues de coté. On voit clairement que les premiers *clusters* correspondent aux parties des voitures les plus communes comme les roues et les coins des fenêtres. La structure est représentée par les positions des composants relatives au centre de l'objet, ce qui est le centre de l'image dans leurs expériences. Pour chaque élément du dictionnaire on garde les positions de toutes les régions qui appartiennent au même *cluster*. Pour la détection on extrait les régions significatives de la même façon et on les apparie aux éléments du dictionnaire de codes. Chaque région appariée vote pour les positions éventuelles du centre de l'objet (Figure 2.16.(b)). On détecte les meilleures hypothèses pour le centre de l'objet par une transformée de Hough. La méthode rend également possible la segmentation de l'objet en réunissant les régions qui contribuent à l'hypothèse sélectionnée.

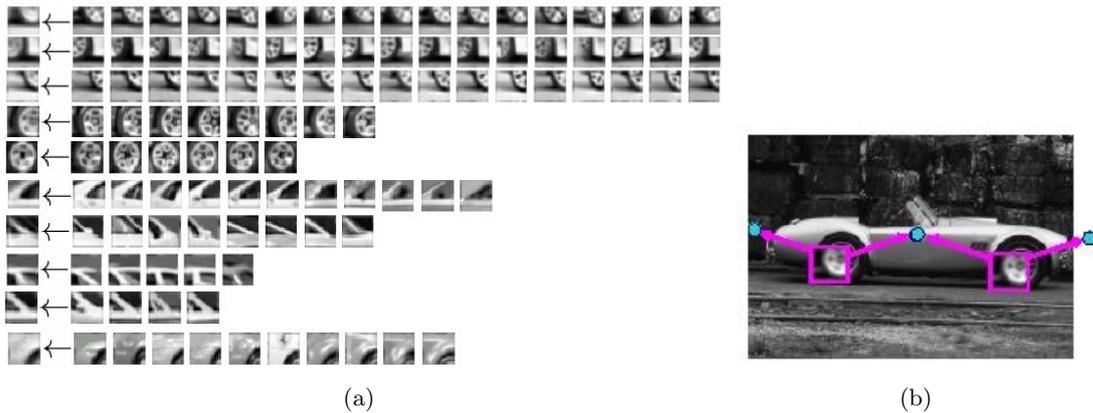


FIG. 2.16 – Illustration de la méthode de Leibe *et al.* [2004] pour la détection des voitures. (a). Le dictionnaire de codes obtenu à partir de 2519 régions significatives détectées sur les images de voitures, (b). Les votes des régions détectées indiquant le centre d'objet.

Dans Agarwal et Roth [2002] on construit le dictionnaire de codes d'une façon très similaire, mais à partir des exemples positifs et négatifs. Pour représenter la structure on a défini 20 relations binaires entre les paires de composants en discrétisant leur distance en 5 intervalles et leur angle en 4 intervalles de 45 degrés. On construit un vecteur des caractéristiques en comptant dans l'image le nombre d'occurrence  $P_n$  de chaque *cluster*  $n$  du dictionnaire de codes (dans leurs expériences  $n = \{1, \dots, 270\}$ ) et le nombre d'occurrence de chaque relation  $R_m(P_{n1}, P_{n2})$  entre chaque paire de *clusters* différents ( $m = \{1, \dots, 20\}$ ,  $n_1 = \{1, \dots, 270\}$ ,  $n_2 = \{1, \dots, 270\}$ ). On obtient un vecteur creux (qui contient beaucoup de zéros) d'une très grande taille. Pour la classification, on utilise le classifieur *SNoW* (*Sparse Network of Windows*) [Carlson *et al.*, 1999] qui est bien approprié pour la classification des objets représentés par des vecteurs de caractéristiques creux.

Différemment de ces approches, Fergus *et al.* [2003] a proposé d'apprendre l'apparence des composants et la structure de l'objet simultanément en utilisant l'algorithme *EM*. On ne connaît ni les paramètres du modèle ni l'affectation des régions aux parties de l'objet pour l'apprentissage. On considère les affectations comme des variables cachées et on apprend les affectations et les paramètres du modèle en même temps. A l'étape d'estimation on estime les affectations et à l'étape de maximisation on met à jour les paramètres du modèle de la structure.

Bouchard et Triggs [2005] ont proposé un modèle de constellation hiérarchique qui représente les relations spatiales probabilistes entre les composants, les sous-composants et les descripteurs locaux. Les paramètres du modèle ont été appris par l'algorithme *EM*. Figure 2.17 illustre le modèle hiérarchique sur une image de cheval.

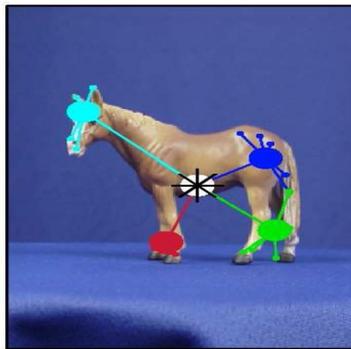


FIG. 2.17 – Illustration du modèle hiérarchique de Bouchard et Triggs [2005].

On peut également utiliser un apprentissage discriminatoire pour apprendre le modèle. Mohan *et al.* [2001] et Heisele *et al.* [2001] ont utilisé le SVM pour l'apprentissage du modèle sur des exemples d'apprentissage sur lesquelles les composants de l'objet sont étiquetés manuellement. Amores *et al.* [2005] ont proposé d'utiliser l'algorithme Adaboost pour l'apprentissage.

Finalement nous voudrions citer deux méthodes qui utilisent les contours comme descripteurs :

Dans [Ferrari *et al.*, 2008] on détecte et on regroupe les  $k$ -adjacent segments (*kAS*) pour obtenir un dictionnaire de codes. La construction du modèle est inspirée de la construction des descripteurs *SIFT* ; on calcule les histogrammes des *kAS* dans des régions obtenues en divisant un rectangle glissant en sous-zones. De cette façon on représente grossièrement les positions relatives des composants (Figure 2.18.(a)). On effectue la classification par un classifieur *SVM*.

Shotton *et al.* [2007] construisent un modèle de constellation à partir d'un dictionnaire de codes obtenu des fragments de contours et des positions des fragments par rapport au centre de l'objet (Figure 2.18.(b)). On détecte le centre de l'objet par un *clustering Mean-Shift* et on apprend le modèle par l'algorithme *Adaboost*.

L'ensemble de ces méthodes qui tentent d'apprendre une catégorie d'objets à partir de l'apparence des composants locaux et de leur organisation constitue un cadre méthodologique tout à fait adapté aux objets cartographiques. En revanche, elles ont été toujours appliquées à des bases

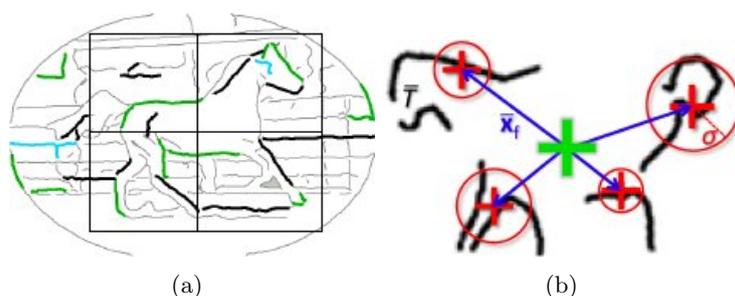


FIG. 2.18 – Illustration des approches par contours. (a). Ferrari *et al.* [2008], (b). Shotton *et al.* [2007]

d'images pour lesquelles les catégories présentent peu de variations structurelles (par exemple les vélos, les automobiles, les visages).

## 2.9 La reconnaissance d'objets cartographiques sur les images satellites et aériennes

En ce qui concerne les images satellitaires, la plupart des travaux aboutis concernent l'extraction de structures étendues (linéaires comme les réseaux routiers ou surfaciques comme les zones urbaines) (Lorette *et al.* [2000]; Tupin *et al.* [1998]). Cette étape d'analyse des images satellitaires correspondait à un niveau de résolution peu élevé où les structures devenaient des textures.

Un livre dédié au sujet [A. Gruen et Henricsson, 1998] recueille 36 articles. L'analyse des sujets des articles révèle des indices pour mieux comprendre où se focalisent les travaux de recherche dans les années 90 : 17 articles traitent de la détection des bâtiments, 6 articles traitent de la détection des routes, 4 articles sont basés sur les méthodes qui utilisent une carte ou un *GIS* pour l'extraction d'information à partir de l'image, et 3 articles traitent de la visualisation 3D.

Presque toutes les méthodes de détection de routes suivent la même approche : la détection des fragments des routes évidents et le suivi de ces fragments pour construire le réseau routier. Ruskoné [1996] a travaillé sur l'extraction automatique du réseau routier sur les images aériennes. Les étapes du système sont la détection de points d'amorces (les fragments de routes évidents) à partir des contours de l'image, le suivi des points d'amorces et la reconstruction des fragments de routes qui manquent d'après un certain nombre de règles de connexions. Le point original pour nous dans son travail est qu'il a souligné l'importance de l'identification d'objets cartographiques pour la validation des hypothèses de routes. Dans ce cadre il a fait une analyse des différents types de carrefours (et les ronds-points qui sont considérés comme une sorte de carrefour). Pourtant, il a limité son travail à la détection des carrefours simples à branches droites. On utilise une approche heuristique pour la détection des carrefours : on calcule des profils radiométriques concentriques autour d'un point candidat qui sont discriminants pour les carrefours.

Bordes [1997] a travaillé sur le problème spécifique de la détection des ronds-points sur les

images satellites. Cette approche ressemble à celle que [Ruskoné, 1996] a utilisé pour détecter les carrefours. On a défini deux descripteurs locaux, un circulaire qui analyse les pixels sur des cercles autour du point et un autre radial sur des rayons (Figure 2.19)

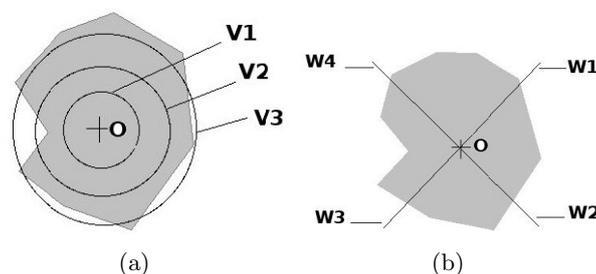


FIG. 2.19 – Les descripteurs radiaux et angulaires, dans Bordes [1997].

Forssen [1997] a traité le sujet de reconnaissance d'objets construits par l'homme. Les objets cibles sont les routes et les zones urbaines. L'auteur a précisé la nécessité d'impliquer des relations spatiales sans les utiliser.

Caron *et al.* [2002] ont développé un système pour détecter les objets artificiels dans des environnements naturels. L'idée principale est que les objets artificiels ont des surfaces plus uniformes que les objets naturels. Pourtant cette méthode suit une approche de bas niveau fondée sur l'uniformité des textures de l'image.

Bhattacharya [2008] a proposé une méthode de classification des images satellites haute résolution. Sur des imagerie obtenues par le découpage d'une image satellitaire, on détecte le réseau routier et les zones urbaines pour calculer des caractéristiques géométriques et topologiques. On classe les images en 7 classes générales en utilisant ces caractéristiques. On obtient une erreur de classification faible pour les classes cibles (villages, champs, montagnes, etc.) qui correspondent à des classes objets d'une grande taille.

Au sein du laboratoire SIP on a développé différents systèmes de reconnaissance d'objets cartographiques à la fois en utilisant des approches radiométriques et des approches géométriques. Ces travaux se sont limités surtout à la détection des ponts, les considérant comme un cas spécifique du problème général.

Dans Sanz et Loménie [2003] on a développé un système dans lequel chaque pixel est d'abord classifié selon ses propriétés radiométriques en utilisant un réseau de neurones. De cette façon, on obtient une segmentation préliminaire des images. On précise ensuite un ensemble de règles fixes pour identifier un pont. Les résultats de la segmentation sont utilisés pour déterminer les régions qui satisfont ces règles, c'est à dire les ponts. L'inconvénient de cette approche est la nécessité d'avoir une large base d'images segmentées manuellement. En même temps l'utilisation des règles manuelles très spécifiques rend difficile l'application de la méthode sur des catégories d'objet génériques.

Loménie *et al.* [2003] a proposé une approche géométrique. Un pont est modélisé comme

deux régions rectangulaires placées dans une certaine relation spatiale. Dans chacun des images, on détecte les segments parallèles et on détecte les formes rectangulaires à partir des parallèles. On compare finalement chaque paire de rectangle avec le modèle du pont défini par des règles manuelles pour détecter les ponts. L'originalité de cette approche est l'utilisation des primitives géométriques pour la reconnaissance, en revanche l'application est limitée à la reconnaissance des ponts.

Dans le cadre du programme ORFEO, Inglada [2007] a proposé une méthode de classification d'objets cartographiques en utilisant des descripteurs par une approche région et des descripteurs géométriques. A l'étape initiale on segmente les images contenant les objets cartographiques et on détecte les structures linéaires. On calcule des descripteurs de régions en utilisant les moments invariants et des descripteurs correspondant aux propriétés géométriques des régions et des structures linéaires. On construit un classifieur SVM à partir des images d'apprentissages pour classifier les images.

Le CNES a développé récemment une boîte à outils appelé **ORFEO Toolbox**<sup>1</sup> dans le cadre du programme ORFEO. Le logiciel est construit à partir de la bibliothèque de traitement d'image *ITK* en y ajoutant les fonctionnalités adaptées au traitement d'images satellites, particulièrement les images à très haute résolution. Le logiciel est distribué sous la forme d'une librairie source ouverte. Cela garantit l'utilisation par un plus grand nombre de personnes et la possibilité pour les utilisateurs de contribuer au développement du logiciel en ajoutant des fonctionnalités qui leur sont nécessaires. La première version du système contient des fonctionnalités de base comme l'accès aux images, la visualisation, le filtrage, l'extraction d'information, la segmentation et la détection des changements. La deuxième version qui a été lancée récemment contient les fonctionnalités de recalage, et de corrections géométriques et radiométriques. Les perspectives pour les deux prochaines versions est l'intégration d'**un module de reconnaissance et la détection d'objets** pour un premier temps, et de **raisonnement spatial** par la suite. A l'heure actuelle on est dans une phase d'analyse et de réflexion sur ce problème qui est assez nouveau pour l'imagerie satellitaire et qui nécessite l'utilisation des méthodes assez différentes des traitements classiques.

Deux thèses qui vont débiter dans les jours suivants montrent l'intérêt des acteurs principaux du domaine de l'imagerie satellitaire pour la recherche d'objets à partir de nouvelles approches qui permet la très haute résolution des images à venir : le CNES se prépare à entamer une thèse sur le sujet '**Relations spatiales pour la détection d'objet**. L'ONERA<sup>2</sup> propose actuellement une thèse sur le sujet **Détection d'objets de petite taille sur séquences aériennes ou satellitaires**.

---

<sup>1</sup>ORFEO Toolbox : [http://smc.cnes.fr/PLEIADES/Fr/lien3\\_vm.htm](http://smc.cnes.fr/PLEIADES/Fr/lien3_vm.htm)

<sup>2</sup>Office National d'Études et de Recherches Aérospatiales

## 2.10 Notre approche

Au début de nos travaux de recherche nous nous sommes intéressés exclusivement aux méthodes d'apprentissage structurel. Ces méthodes qui visent à construire un modèle explicite de l'objet fondé sur une représentation symbolique nous semblent être convenables pour représenter les images satellitaires à un haut niveau d'abstraction. Nous avons utilisé les *ARGs* construits à partir des primitives géométriques pour modéliser les objets cartographiques, et développé une méthode basée sur l'appariement de graphes pour apprendre un modèle commun aux objets appartenant à une même classe. Nous avons réussi à obtenir des modèles pour représenter les ronds-points et les ponts. Les résultats de cette étape nous ont montré qu'une approche qui exploite les propriétés géométriques de composants d'objets est appropriée pour définir et distinguer les objets cartographiques.

La partie 2 est une étude de faisabilité d'une approche purement structurelle au niveau de résolution des images satellitaires actuellement atteint. Nous verrons en revanche que son application sur les imagerie non-segmentées présente encore des difficultés dans la partie de reconnaissance.

Ainsi, nous nous sommes orientés, tout en préservant la représentation des objets par des primitives géométriques, vers les approches fondées sur l'apparence de l'objet, un domaine de recherche relativement nouveau mais en plein développement, qui est appliqué sur plusieurs problèmes de reconnaissance d'objets actuels avec succès.

Nous avons donc adopté dans un deuxième temps un modèle plus souple de constellation qui représente les composants de l'objet et sa structure. Les primitives géométriques sont extraites par une approche par régions et une approche par contours en parallèle. On a proposé des descripteurs qui révèlent les propriétés géométriques des primitives et leurs organisations spatiales.

Deux modèles implicites construits à partir d'un ensemble d'images d'apprentissage par deux méthodes différentes sont utilisés pour reconnaître les imagerie de ronds-points et de ponts.

Dans la première méthode on construit le modèle d'objet par la fusion des classifieurs faibles obtenus des primitives géométriques en utilisant un classifieur *Adaboost*.

Dans la deuxième méthode on construit un dictionnaire de codes à partir des primitives géométriques par le *clustering Mean-Shift*.



## Deuxième partie

# Génération de modèle d'objet explicite

L'objectif principal de cette deuxième partie est d'explorer l'utilisation des méthodes de reconnaissance de formes structurelle pour la représentation des objets cartographiques. L'approche structurelle vise à construire un modèle d'objet constitué d'un ensemble d'entités et des relations entre ces entités, qui décrit explicitement la forme des objets appartenant à une même catégorie. Cette représentation de haut niveau de l'objet est généralement construite en utilisant des graphes, dont les sommets correspondent aux entités de l'objet et les arêtes aux relations spatiales des entités.

Nous présentons une méthode d'apprentissage structurel pour générer des modèles d'objet pour les catégories ronds-points et ponts. Le modèle d'objet est représenté par un graphe relationnel attribué (*ARG*), et construit à partir des formes géométriques détectées dans les images segmentées par un expert. On apprend ensuite un modèle plus général qui couvre les *ARGs* les plus fréquents appartenant aux objets dans une même catégorie. Les modèles d'objet sont explicites dans la mesure où ils permettent :

- une expressivité quasi symbolique de l'objet ;
- une visualisation de l'objet qui ouvre la voie à une éventuelle interactivité avec l'utilisateur final.

Nous verrons que les modèles obtenus ne sont pas encore mature d'un point de vue théorique mais mériteront d'être mieux étudiés dans leur capacité à intégrer les relations spatiales entre les objets, un sujet émergent actuel.

Nos travaux de recherche sur la génération de modèle d'objet explicite ont donné lieu à deux publications dans des congrès internationaux ([Erus et Loménie, 2005a] et [Erus et Loménie, 2005b])



## Chapitre 3

# Modèles explicites d'objets cartographiques

### 3.1 Introduction

Le problème d'apprentissage structurel consiste à construire automatiquement des prototypes qui soient capables de représenter les différents objets appartenant à une même classe. C'est un problème très difficile et non résolu jusqu'à ce jour, parce qu'on veut construire des modèles qui décrivent explicitement la forme d'un objet en utilisant une approche symbolique au lieu d'une approche statistique.

Comme point de départ pour la reconnaissance d'objets cartographiques, nous proposons une méthode d'apprentissage structurel pour générer un modèle d'objet à partir d'un ensemble d'images segmentées et étiquetées. Notre méthode vise à générer des modèles explicites de ronds-points et de ponts. L'originalité de cette partie de notre travail repose principalement sur l'application de techniques d'apprentissage structurel sur des objets cartographiques complexes et très variés.

Cette partie doit être considérée comme une étude de faisabilité de l'approche structurelle avec la résolution actuelle des images. Nous verrons les limites actuelles de ces approches en fin de chapitre pour la reconnaissance dans des images non-présegmentées.

#### 3.1.1 La position du problème et la solution proposée

Nous nous sommes posés quelques questions essentielles tout au début de notre recherche :

- Quelle est la représentation appropriée pour construire le modèle d'une classe d'objets cartographiques ?
- Peut-on apprendre automatiquement les modèles génériques pour différentes classes d'objets ?
- Comment peut-on apprendre ces modèles à partir d'exemples ?

La recherche des réponses à ces questions nous a conduit à développer un système d'apprentissage d'un modèle d'objet structurel par une approche fondée sur les graphes. Nous avons utilisé des graphes relationnels attribués (*Attributed Relational Graphs, ARGs*), en tant que structure de donnée principale pour représenter les caractéristiques géométriques et spatiales des objets cartographiques. Puis nous avons appliqué des algorithmes d'appariement de graphes pour générer le modèle d'objet. La modélisation des relations spatiales était l'enjeu principal de cette étude.

### 3.1.2 La base d'images d'apprentissage

Pope et Lowe [1993, page 2] précisent clairement la difficulté d'apprendre un modèle d'objet structurel à partir d'images :

Peu de systèmes de reconnaissance d'objets sont conçus pour acquérir leur modèle directement à partir des images d'intensité. Au lieu de cela, on donne à la plupart des systèmes tout simplement des modèles en forme de descriptions de formes construits manuellement ... Comme il est difficile de modéliser la formation de l'image, la plupart des systèmes de reconnaissance d'objets se restreignent aux modèles d'objet relativement simples et grossiers<sup>1</sup>.

La construction du modèle structurel nécessite avant tout de détecter un ensemble de primitives communes aux différents objets d'une même classe sur les images. Cette information est obtenue généralement en effectuant l'apprentissage sur des images synthétiques d'objets simples ou en utilisant un niveau de supervision élevé en marquant les parties de l'objet sur les images. Sur les imagerie d'objets cartographiques il est très difficile de séparer l'objet du fond et de sélectionner les primitives qui appartiennent à l'objet. Pour cette raison, nous avons utilisé comme images d'apprentissage les imagerie segmentées par un expert (Section 1.3.1). Sur ces imagerie les objets cibles sont marqués par un masque.

## 3.2 Notre approche

Les propriétés radiométriques des images ne permettent pas de distinguer les objets cartographiques dans des catégories distinctes. En même temps les images appartenant à une même catégorie peuvent avoir des distributions radiométriques différentes. Par exemple, un pont qui passe sur une rivière a des propriétés radiométriques assez différentes de celles d'un pont d'autoroute. Il est donc nécessaire de trouver une représentation plus abstraite pour définir les catégories d'objets. L'information discriminante est sans doute structurelle, dissimulée d'abord

---

<sup>1</sup>Few object recognition systems have been designed to acquire their models directly from intensity images. Instead, most systems are simply given models in the form of manually-produced shape descriptions ... Because modeling image formation has proven difficult, most object recognition systems that follow this approach have been restricted to object models that are relatively simple and coarse.

dans les composants primitifs géométriques de l'objet et puis dans l'organisation spatiale de ces composants.

Dans la plupart des travaux d'apprentissage structurel existants on utilise les *ARGs*, les graphes enrichis d'un ensemble d'attributs associés aux sommets et aux arêtes, pour représenter les objets. C'est la structure de donnée la plus appropriée pour révéler l'information structurelle parce que les *ARGs* permettent de représenter les propriétés des composants de l'objet et aussi les propriétés des relations entre les composants.

Cordella *et al.* [2002] ont travaillé sur une question qui évoque notre problème : comment peut on représenter les objets qui ont des structures similaires ? Pour répondre à cette question, ils ont introduit une nouvelle structure de donnée, l'*ARG* Généralisé (*GARG*) construit à partir des sommets, des arêtes et des attributs généralisés qui peuvent avoir plusieurs valeurs en même temps. L'objectif d'apprentissage est de générer un ensemble de *GARGs* qui couvrent tous les exemples d'une même catégorie. Pour construire un prototype on initialise le prototype à l'ensemble vide et on le spécialise jusqu'à ce qu'il couvre un certain nombre d'exemples. On a évalué la méthode sur des objets artificiels créés par une grammaire.

Nous avons suivi un schéma général similaire. Notre système se constitue de deux modules principaux (Figure 3.1) :

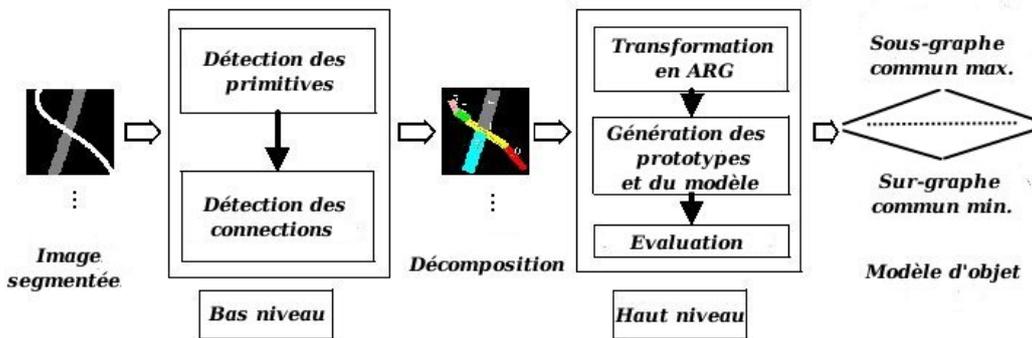


FIG. 3.1 – Schéma général de génération de modèle structurel.

Le module d'extraction utilise comme entrées les imagettes segmentés et étiquetés qui contiennent l'objet cible. On décompose les imagettes en formes géométriques primitives. On sélectionne les primitives pertinentes et on détermine leurs connexions.

Le deuxième module utilise la sortie du premier module pour construire un *ARG* pour chaque image. Les caractéristiques des composants et les relations spatiales binaires entre les composants sont stockées respectivement dans les attributs des sommets et les attributs des arêtes du graphe. On regroupe les graphes appartenant à une même classe d'objet en utilisant un algorithme d'appariement de graphe exact. De cette façon, on détermine les prototypes les plus fréquents de l'objet. On génère ensuite un treillis de graphe qui représente un modèle général de l'objet. Pour évaluer la qualité du résultat et pour comparer le degré d'appartenance des autres images, on calcule une distance d'édition de chacun des *ARGs* au modèle obtenu.

### 3.3 Les traitements de bas niveau : extraction des primitives

L'objectif de ce module est de décomposer l'image d'objet en primitives géométriques simples connectées. Deux décompositions alternatives ont été développées. Une approche par régions décompose l'image en cercles et en rectangles. Une approche par squelette la décompose en cercles et en segments de droites. La décomposition des imagerie en primitives géométriques se constitue de plusieurs étapes successives présentées ci-dessous.

#### 3.3.1 Extraction des primitives "cercles"

La première étape du traitement est la détection des cercles dans l'image. La transformée de Hough est une méthode bien connue basée sur la projection de l'image dans un espace de paramètres représentatif de la forme recherchée. L'accumulation des évidences dans cet espace indique l'existence ou non de la forme dans l'image. Dans la transformée de Hough pour les cercles, l'espace de paramètres  $h(I)$  se compose de trois paramètres  $(c_x, c_y, r)$  où  $(c_x, c_y)$  est le centre et  $r$  est le rayon du cercle recherché. On applique la transformée sur les contours de l'objet. La projection d'un point de contour  $(x, y)$  de l'image  $I$  est l'ensemble de tous les points sur  $h(I)$ , de centre  $x, y$  et de rayon  $r$ , pour toutes les valeurs possibles de  $r$ . Pour chaque point de contour sur l'espace  $I$ , on obtient donc un cône dans l'espace  $h(I)$ . La projection des points appartenant à un cercle sur  $I$  devrait donner une accumulation importante en un même point sur  $h(I)$  qui caractérise les paramètres de ce cercle. Les étapes de notre implémentation sont données dans l'algorithme 1.

---

**Algorithm 1** Algorithme de détection de cercles

---

**Entrées:**

$\mathbf{p} = \{\mathbf{p}_x, \mathbf{p}_y\}$  : la liste des points de contour,  $s$  : seuil d'acceptation

**Sorties:**

$\mathbf{h}(\mathbf{I}) \in N^3$  : l'accumulateur final,  $\mathbf{c} = \{\mathbf{c}_x, \mathbf{c}_y, \mathbf{c}_r\}$  : la liste des cercles détectés

**Pour** chaque point de contour  $p_i \in \mathbf{p}$  **Faire**

Calculer la projection de  $p_i$  sur  $h(I)$  et accumuler les valeurs correspondants

**Fin Pour**

Normaliser  $\mathbf{h}(\mathbf{I})$  par la valeur qu'on obtiendrait pour un cercle complet de rayon  $r$

Trouver  $\mathbf{m}$  les maxima locaux de  $\mathbf{h}(\mathbf{I})$

**Retourner**  $\mathbf{c} = \mathbf{m}$  dont  $\mathbf{h}(\mathbf{m}) > s$

---

On a défini le seuil d'acceptation empiriquement. Les figures 3.2.(b-c) présentent  $I$  et la section transversale de  $h(I)$  où on a obtenu l'accumulation maximale sur une image de rond-point. Après la détection d'un cercle, la région intérieure du cercle est effacée de l'image (Figure 3.2.(c))

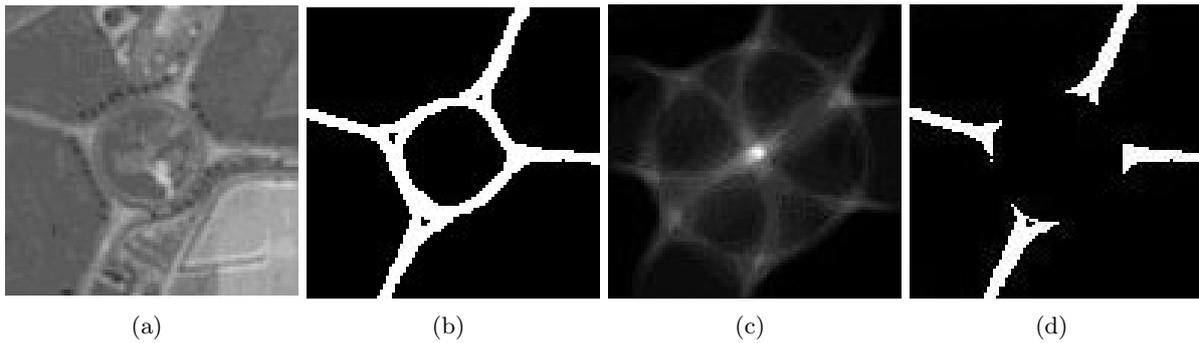


FIG. 3.2 – Détection de cercles sur une image de rond-point par la transformée de Hough. (a). Imagerie originale, (b). Imagerie segmentée par un expert, (c). La section transversale de l'espace de Hough pour  $r = 19$ . On a une accumulation de 232 (normalisé entre 0 et 255) au point (50,50), ce qui indique l'existence d'un cercle de rayon 19 aux coordonnées (50, 50), (d). L'image segmentée après l'extraction du cercle.

### 3.3.2 Extraction des primitives “rectangles”

On détecte les rectangles à partir des segments de droites parallèles obtenus à partir des contours de l'objet. Sur les images segmentées les routes qui se croisent ont été marquées par des niveaux de gris différents. Nous avons créé une image binaire pour chaque niveau de gris pour les traiter séparément. Les irrégularités des routes ont été éliminées par une ouverture morphologique. Comme on a des images binaires, on peut détecter les contours par le gradient morphologique qui consiste à trouver la différence entre l'image originale et l'image dilatée.

#### 3.3.2.1 Approximation polygonale

On peut représenter les chaînes de contours par des segments de droites par une approximation polygonale. Nous avons utilisé l'algorithme de la corde [Douglas et Peucker, 1973]. L'algorithme est proposé originalement pour la simplification des cartes cartographiques. L'idée principale est d'approximer initialement la courbe par un segment qui relie les deux extrémités et de diviser récursivement ce segment au point de la courbe qui est le plus loin au segment (Figure 3.3).

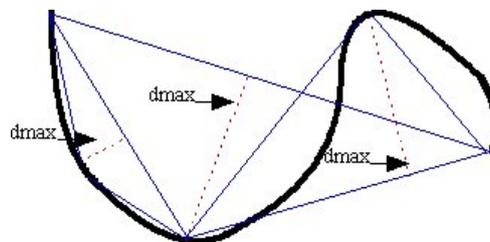


FIG. 3.3 – Illustration de l'algorithme de la corde.

La figure 3.4 montre les segments de droites détectés sur une image de rond-point et sur une

image de pont.

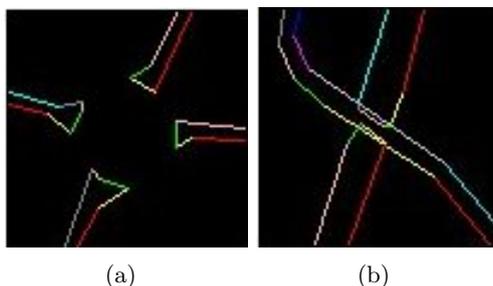


FIG. 3.4 – Les segments détectés par l'algorithme de la corde. (a). Rond-point, (b). Pont

### 3.3.2.2 Sélection des segments parallèles et des rectangles

Les parallèles sont représentés par les rectangles qui correspondent à leur recouvrement minimal comme dans Ruskoné [1996]. Parmi tous les segments détectés, nous avons choisi ceux qui sont parallèles entre eux et éliminé les autres. Pour détecter le parallélisme, on a précisé un seuil de tolérance  $a$ . Après la détection des parallèles, on trouve le rectangle déterminé par chaque paire de parallèles. Il est possible que deux segments parallèles ne déterminent pas un rectangle parce que leur recouvrement minimal est vide. En même temps on élimine les rectangles qui ne satisfont pas les critères de taille, c'est à dire que la distance entre les parallèles est plus important que la longueur de leur recouvrement.

### 3.3.2.3 Détection des connections entre les composants de l'objet

A la fin de l'étape précédente on a obtenu des formes géométriques qui ne sont pas en général connectées. C'est la conséquence naturelle des approximations lors de la détection des primitives. Pourtant, les connections entre les figures sont cruciales pour la construction des graphes. Pour cette raison, nous avons appliqué un dernier traitement heuristique qui vise à déterminer les primitives qui pouvaient être connectées. On prolonge les deux extrémités de chaque rectangle jusqu'à ce qu'on atteigne un seuil d'élargissement. On connecte les primitives qui s'intersectent.

Un objet peut être décomposé de plusieurs façons différents est l'évaluation de la décomposition est très subjective. A cause de cette raison nous avons fait une analyse qualitative par une inspection visuelle. Nous avons observé qu'on a correctement détecté les ronds des ronds-points sauf pour quelques ronds-points qui ont une petite taille et les routes principales sauf quelques routes étroites avec une courbure importante. Ce manque de détection sera compensé à l'étape de sélection des prototypes. Figures 3.5 et 3.6 représentent des images de ronds-points et de ponts décomposées.

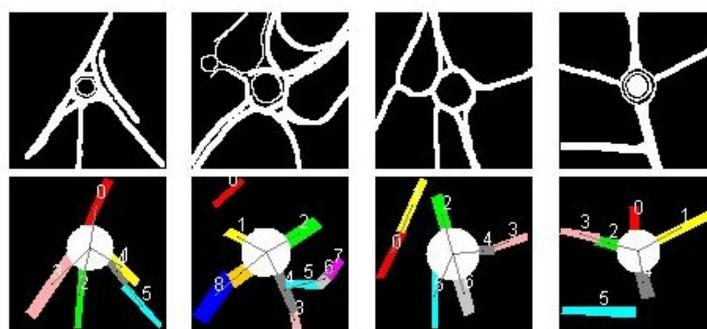


FIG. 3.5 – Décomposition d’images de ronds-points par des rectangles et des cercles.

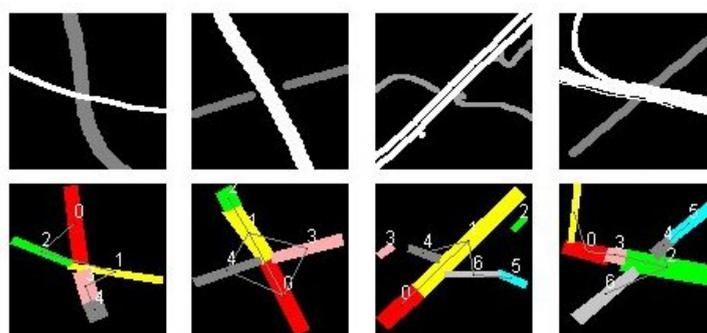


FIG. 3.6 – Décomposition d’images de ponts par des rectangles.

### 3.3.3 Extraction des primitives “segments de droites”

Une approche alternative pour décomposer un objet en primitives est l’extraction du squelette de l’objet. Le squelette est surtout approprié pour représenter les régions allongées comme les routes. Nous avons appliqué une méthode d’amincissement pour extraire le squelette.

#### 3.3.3.1 L’algorithme d’amincissement (*thinning*)

L’amincissement est un ensemble d’érosions successives appliquées sur l’image jusqu’à l’obtention d’un squelette connexe d’épaisseur minimale. Nous avons choisi d’implémenter l’algorithme de Zhang et Suen [1984] pour extraire le squelette. C’est un algorithme à deux itérations. Les pixels qui composent l’objet sont marqués par noir et le fond par blanc. Les voisins d’un pixel sont les pixels 8-connexes à ce pixel. A la première itération, on marque un pixel  $I(i, j)$  noir pour le supprimer plus tard si ce point satisfait les conditions suivantes :

1. Le pixel est connecté à un seul groupe de pixels noirs. C’est-à-dire, il n’est pas le connecteur entre deux groupes de pixels noirs ;
2. Il a au moins 2 et au plus 6 voisins noirs ;
3. Au moins un des pixels  $I(i, j - 1)$ ,  $I(i - 1, j)$  et  $I(i, j + 1)$  est noir ;

4. Au moins un des pixels  $I(i - 1, j)$ ,  $I(i + 1, j)$  et  $I(i, j - 1)$  est noir.

A la fin de l'itération on supprime les pixels marqués. A la deuxième itération on marque un pixel  $I(i, j)$  noir pour le supprimer plus tard si ce point satisfait les conditions suivantes :

1. Le pixel est connecté à un seul groupe de pixels noirs. C'est à dire, il n'est pas le connecteur entre deux groupes de pixels noirs ;
2. Il a au moins 2 et au plus 6 voisins noirs ;
3. Au moins un des pixels  $I(i - 1, j)$ ,  $I(i + 1, j)$  et  $I(i, j + 1)$  est noir ;
4. Au moins un des pixels  $I(i, j - 1)$ ,  $I(i + 1, j)$  et  $I(i, j + 1)$  est noir.

A la fin de la deuxième itération on supprime les pixels marqués et on obtient le squelette.

La figure 3.7.(a) représente le résultat de l'algorithme d'amincissement sur une image de pont.

### 3.3.3.2 Détection des segments

A partir des squelettes qu'on a obtenus, on a appliqué les traitements suivants pour obtenir les segments de droites :

- Algorithme de la corde pour l'approximation des courbes par des segments de droites (3.7.(b)) ;
- Filtrage des segments qui sont plus court qu'un seuil ;
- Algorithme de prolongement des segments (similaire à celui utilisé pour les rectangles) pour déterminer les connections entre les segments de droites (3.7.(c)).

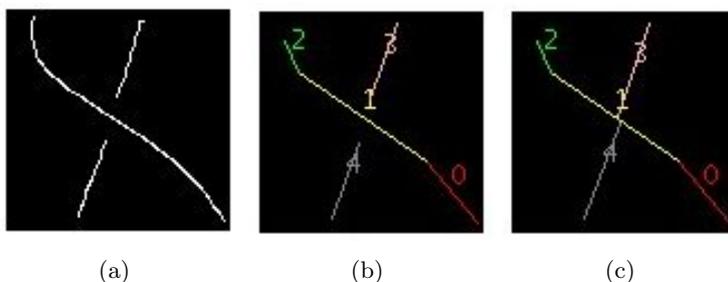


FIG. 3.7 – Extraction des primitives segments de droites à partir du squelette d'objet sur une image de pont. (a). Le squelette (b). Approximation par des segments (c). Les segments connectés.

La figure 3.8 montre la décomposition sur quelques imagerie de ronds-points et de ponts.

## 3.4 Les traitements de haut niveau : construction du modèle

La première étape des traitement de haut niveau est la transformation des images décomposées en *ARGs*. On détecte ensuite les prototypes parmi les graphes d'une même classe d'objet.

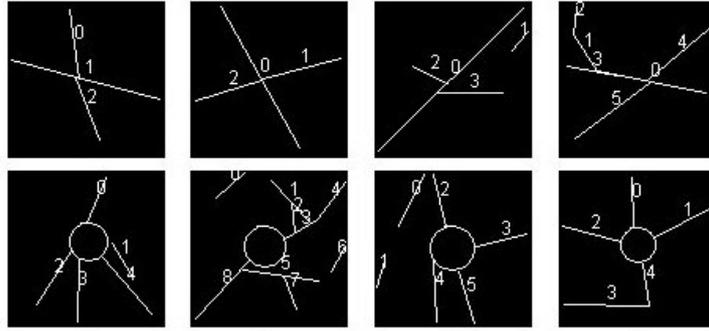


FIG. 3.8 – Décomposition d'images de ronds-points et de ponts par des cercles et des segments de droites.

On trouve le **sous-graphe commun maximal** (*SousGMax*) et le **sur-graphe commun minimal** (*SurGMin*) des prototypes pour déterminer le modèle. Finalement, on calcule la distance d'édition de chaque graphe au modèle trouvé pour évaluer la valeur représentative du modèle.

### 3.4.1 Définitions principales

Pour rendre la lecture de la partie suivante plus claire nous présentons les définitions des notions principales sur les graphes et les graphes relationnels attribués :

**Définition 1 :** Un graphe est un 4-tuple  $G = (S, A, \alpha S, \alpha A)$  où :

$S$  est un ensemble fini de sommets,

$A \in S \times S$  est un ensemble fini d'arêtes,

$\alpha S : S \rightarrow EtiquetteS$  est une fonction qui assigne l'étiquette d'un sommet,

$\alpha A : A \rightarrow EtiquetteA$  est une fonction qui assigne l'étiquette d'une arête.

Un *ARG* peut être défini d'une façon similaire en remplaçant les étiquettes des sommets et des arêtes par un ensemble d'attributs

**Définition 2 :** Un graphe relationnel attribué est un 6-tuple  $G = (S, A, A_s, A_a, \alpha S, \alpha A)$  où :

$S$  est un ensemble fini de sommets,

$A \in S \times S$  est un ensemble fini d'arêtes,

$A_s$  est un ensemble fini des attributs de sommets,

$A_a$  est un ensemble fini des attributs d'arêtes,

$\alpha S : S \rightarrow A_s$  est une fonction qui assigne les attributs d'un sommet,

$\alpha A : A \rightarrow A_a$  est une fonction qui assigne les attributs d'une arête.

Le nombre d'attributs d'un sommet ou d'une arête est variable en fonction du type du sommet ou de l'arête. Ce qui rend l'*ARG* spécifique à une application est la détermination de ces attributs. La structure de nos *ARGs* est expliquée à la section suivante.

**Définition 3 :**  $G = (S, A, A_s, A_a, \alpha S, \alpha A)$  étant un *ARG*, le sous-graphe de  $G$  est  $G' = (S', A', A'_s, A'_a, \alpha S', \alpha A')$  tels que :

$$\begin{aligned} S' &\subseteq S, \\ A' &= A \cap (S' \times S') \text{ et,} \\ \forall s \in S', \alpha S'(s) &= \alpha S(s), \\ \forall a \in A', \alpha A'(a) &= \alpha A(a). \end{aligned}$$

**Définition 4 :** Une fonction bijective  $f : S \rightarrow S'$  est un isomorphisme de graphe d'un *ARG*  $G = (S, A, A_s, A_a, \alpha S, \alpha A)$  à un *ARG*  $G' = (S', A', A'_s, A'_a, \alpha S', \alpha A')$  si :

$$\begin{aligned} \forall s \in S', \alpha S'(s) &= \alpha S'(f(s)), \\ \forall a = (s_1, s_2) \in A \text{ il existe une arête } a' &= (f(s_1), f(s_2)) \text{ tels que } \alpha A(a) = \alpha A'(a') \text{ et,} \\ \forall a' = (s'_1, s'_2) \in A' \text{ il existe une arête } a &= (f(s'_1), f(s'_2)) \text{ tels que } \alpha A'(a') = \alpha A(a). \end{aligned}$$

**Définition 5 :** Une fonction injective  $f : A \rightarrow A'$  est un isomorphisme de sous-graphe de  $G$  à  $G'$  s'il existe un sous-graphe  $S$  de  $G'$  tels que  $f$  est un isomorphisme de graphe de  $G$  à  $S$ .

**Définition 6 :** Soient  $G, G_1$  et  $G_2$  des *ARGs*.  $G$  est le sous-graphe commun de  $G_1$  et  $G_2$  s'il existe des isomorphismes de sous-graphe de  $G$  à  $G_1$  et de  $G$  à  $G_2$ .

**Définition 7 :** Le sous-graphe commun  $G$  de  $G_1$  et  $G_2$  est maximal si pour tout sous-graphe commun  $G'$  de  $G_1$  et  $G_2$ , le nombre de sommets de  $G$  est supérieur ou égal au nombre de sommets de  $G'$ .

**Définition 8 :** Soient  $G, G_1$  et  $G_2$  des *ARGs*.  $G$  est le sur-graphe commun de  $G_1$  et  $G_2$  s'il existe des isomorphismes de sous-graphe de  $G_1$  à  $G$  et de  $G_2$  à  $G$ .

**Définition 9 :** Le sur-graphe commun  $G$  de  $G_1$  et  $G_2$  est minimal si pour tout sur-graphe commun  $G'$  de  $G_1$  et  $G_2$ , le nombre de sommets de  $G$  est inférieur ou égal au nombre de sommets de  $G'$ .

### 3.4.2 Construction des *ARGs*

Nous avons utilisé des *ARGs* pour représenter les images décomposées. Les primitives constituent les sommets d'un graphe, et chaque connexion entre deux primitives est une arête entre les sommets correspondants. Les attributs des sommets sont

- Le type ( $t$ ) d'un sommet, choisi parmi un alphabet fini  $T = \{\text{cercle}, \text{rectangle}, \text{segment}\}$  de types possibles.
- Les rectangles ont les attributs *centre*, *longueur*, *largeur* et *angle avec l'horizontale*. Les segments ont les mêmes attributs que les rectangles sauf la *largeur*. Les cercles ont les attributs *centre* et *rayon*.

Les valeurs de ces attributs sont représentées par des entiers naturels. Les attributs des arêtes sont :

- Le *type de l'arête*, un tuple de types des sommets connectés.
- La *direction de connexion* qui est discrétisée par l'ensemble  $\{\textit{perpendiculaire}, \textit{parallèle}\}$  en utilisant un seuil de 45 degrés.

Figure 3.9 présente la transformation d'un objet en *ARG* correspondant.

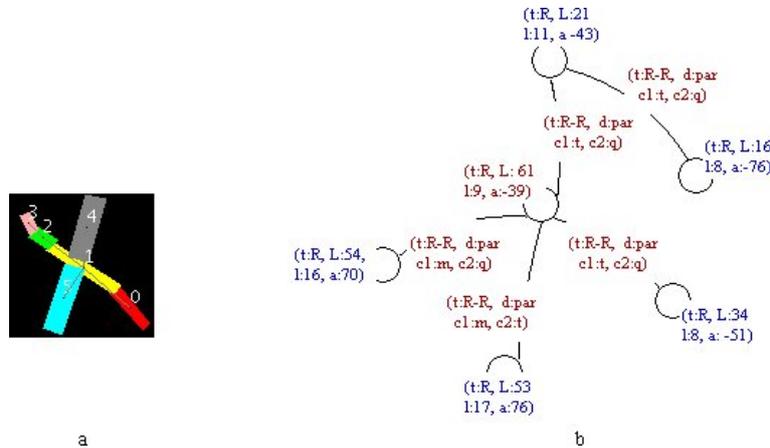


FIG. 3.9 – Représentation d'un pont par un *ARG*. (a). l'image décomposée (b). l'*ARG*.

### 3.4.3 Détection des prototypes

Au lieu de traiter tous les graphes, nous proposons une méthode heuristique qui vise à réduire le nombre de graphes en détectant les prototypes entre eux. Nous avons supposé que parmi les graphes appartenant à une classe d'objet, plus le nombre d'occurrence d'un graphe augmente, plus ce graphe devient représentatif de l'objet. En groupant les graphes qui sont isomorphiques et en triant les groupes par le nombre d'individus, nous avons construit une liste dans l'ordre de fréquence. Nous avons considéré les prototypes comme les graphes qui occupent les premiers rangs dans cette liste. Le nombre de prototypes est déterminé en comparant le nombre de graphes couverts par chaque prototype. On considère que les prototypes obtenus représentent les caractéristiques de l'objet commun et les autres graphes sont des variations du modèle général de l'objet.

Nous avons relaxé les conditions nécessaires pour l'isomorphisme en utilisant un sous-ensemble des attributs pour la comparaison, au lieu de comparer tous les attributs. En considérant que les dimensions des formes étaient variables et n'étaient pas représentatives de la définition d'un objet, nous avons négligé ces attributs. Pour préserver les relations spatiales, nous avons utilisé la direction de connexion entre les formes. En conséquence, les attributs utilisés pour comparer deux *ARGs* sont le type de sommet, le type de l'arête et la direction de connexion.

Notre algorithme d'appariement exact se compose de deux étapes :

Soient  $G$  et  $H$  deux graphes,

1. On sépare les sommets de  $G$  et de  $H$  en classes d'équivalences  $c_G$  et  $c_H$  par leurs degrés : la création des classes d'équivalences empêche les comparaisons inutiles entre les sommets qui ont des degrés différents et de cette façon réduit la complexité de l'algorithme.
2. On fait une recherche exhaustive en appariant les sommets de deux graphes qui sont dans la même classe d'équivalence. On trouve toutes les permutations possibles d'appariements entre les sommets  $s_G$  de  $G$  et  $s_H$  de  $H$ , tels que  $c_G(s_G)$  est égal à  $c_H(s_H)$ . Pour chaque permutation, si
  - les types de sommets qui s'apparient,
  - les arêtes entre ces sommets, et
  - la direction de connexion pour chaque arête
 sont les mêmes, on considère cette permutation comme un isomorphisme entre les deux graphes.

L'algorithme est appliqué aux paires de graphes. On ne compare pas toutes les paires, chaque fois qu'on détecte un isomorphisme, on construit un groupe et on ajoute les deux graphes dans le groupe. Un nouveau graphe est comparé avec seulement un des éléments de ce groupe. Les prototypes détectés manquent de valeur pour des attributs qui ne sont pas utilisés pour la comparaison des graphes. On trouve le graphe qui est le plus proche de la moyenne cumulative de ces attributs et les valeurs de ce graphe sont utilisées pour déterminer les valeurs du prototype pour ces attributs. De cette façon, on utilise certains des attributs pour déterminer l'équivalence entre les graphes, et les autres pour trouver l'attribut correspondant du prototype.

### 3.4.4 Construction du modèle

Nous considérons comme modèle le treillis déterminé par le *SousGMax* et le *SurGMin* des prototypes, c'est à dire tous les graphes  $G$  tels que  $SousGMax \subseteq G \subseteq SurGMin$ .

### 3.4.5 Le sous-graphe commun maximal des prototypes

Nous avons implémenté un algorithme récursif pour trouver le *SousGMax* de deux *ARGs*  $G_1$  et  $G_2$ . On vérifie d'abord si le deuxième graphe est un sous-graphe du premier. Si c'est le cas, on obtient le résultat. Sinon on appelle la même fonction récursivement avec tous les sous-graphes du deuxième graphe obtenus en supprimant un des sommets à la fois, et on obtient une liste des sous-graphes communs des deux graphes. Parmi ces graphes, on choisit celui avec le plus grand nombre de sommets. La condition d'arrêt de la récursion est l'obtention d'un graphe vide ou bien la détection d'un sous-graphe. Algorithme 2 présente une définition formelle de l'algorithme. L'opérateur  $\ominus$  entre deux graphes  $G_1$  et  $G_2$  exprime la suppression des sommets de  $G_1$  appariés aux sommets de  $G_2$ , et les arêtes qui sont liés aux sommets supprimés.

Pour obtenir le *SousGMax* de tous les prototypes, on range les prototypes dans une liste, on prend les deux premiers graphes et on les remplace par leur *SousGMax*. On continue cette

---

**Algorithm 2** Algorithmme pour calculer le *SousGMax* de deux *ARGs*  $G_1$  et  $G_2$ 


---

**Entrées:** $G_1, G_2$  : les graphes**Sorties:** $M(G_1, G_2)$  : le SousGMax de  $G_1$  et  $G_2$ **Si**  $G_2 \subseteq G_1$  **alors****Retourner**  $G_2$ **Sinon****Pour**  $\forall s_i \in \text{Sommets}(G_2)$  **Faire** $G_i = G_2 \ominus s_i$  $M_i = M(G_1, G_i)$ **Fin Pour****Retourner**  $\max_{\|M_i\|}$ **Fin Si**


---

procédure jusqu'à ce qu'il ne reste qu'un seul graphe dans la liste.

### 3.4.6 Le sur-graphe commun minimal des prototypes

Le *SurGMin* de deux graphes  $G_1$  et  $G_2$  est trouvé à partir de leur *SousGMax*. Une définition formelle de l'algorithme est présentée dans l'Algorithme 2. L'opérateur  $\oplus$  entre deux graphes  $G_1$  et  $G_2$  exprime l'ajout au  $G_1$  des sommets de  $G_2$  non appariés au  $G_1$ , et les arêtes qui sont liés à ces sommets.

---

**Algorithm 3** Algorithmme pour calculer le *SurGMin* de deux *ARGs*  $G_1$  et  $G_2$ 


---

**Entrées:** $G_1, G_2$  : les graphes**Sorties:** $m(G_1, G_2)$  : le SurGMin de  $G_1$  et  $G_2$  $G^M = \text{SousGMax}(G_1, G_2)$  $G^m = \text{SurGMin}(G_1, G_2)$ **Retourner**  $(G_1 \ominus M) \oplus (G_2 \ominus M) + M$ 


---

Les attributs numériques des modèles obtenus sont déterminés en utilisant les attributs des prototypes comme on l'a fait à l'extraction des prototypes.

### 3.4.7 Distance d'édition des exemples au modèle

Pour évaluer la puissance représentative du modèle généré, nous avons calculé la distance d'édition de chacun des graphes au modèle en utilisant la métrique suivante proposé par Bunke et Shearer [1998] :

$$d(G_1, G_2) = 1 - \frac{\|(SousGMax(G_1, G_2))\|}{\max(\|G_1\|, \|G_2\|)}$$

Nous avons adapté cette métrique aux *ARGs*, en utilisant l'algorithme pour calculer le *SousGMax* des *ARGs*. La métrique est définie pour trouver la distance entre deux graphes. Pour calculer la distance d'un graphe  $G$  au modèle  $M$  nous avons généré  $\bar{G}$  l'ensemble de tous les graphes couverts par le modèle et trouvé le minimum des distances de  $G_1$  à chaque élément de  $\bar{G}$  (Algorithme 4).

---

**Algorithm 4** Algorithme pour calculer la distance d'édition d'un *ARG*  $G$  au modèle  $M$

---

**Entrées:**

$G$  : le graphe,  $M$  : le modèle

**Sorties:**

$d_M$  : distance d'édition

**Pour**  $\forall G_i \in G$  **Faire**

$d_i = d(G, G_i)$

**Fin Pour**

**Retourner**  $\min_{d_i}$

---

### 3.5 Les résultats expérimentaux

Nous avons testé notre système sur 4 groupes de test : images de pont décomposées par rectangles ( $P_R$ ) ou par segments ( $P_S$ ), et images de ronds-points décomposées par cercles et rectangles ( $RP_R$ ) ou par cercles et segments ( $RP_S$ ). Nous avons utilisé 62 images de ponts et 54 images de ronds-points.

Les prototypes obtenus sont présentés à la figure 3.10. La figure 3.11 présente les modèles obtenus à partir des prototypes. Les modèles sont simples, expriment la structure des catégories d'objets cibles et représentent les variations des objets dans une certaine mesure. Les modèles de pont obtenus par les deux différents algorithmes de décomposition sont similaires.

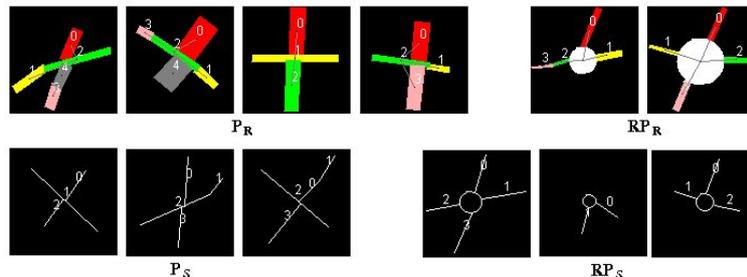


FIG. 3.10 – Les prototypes obtenus.

Pour évaluer la qualité des modèles quantitativement, nous avons calculé la distance d'édition

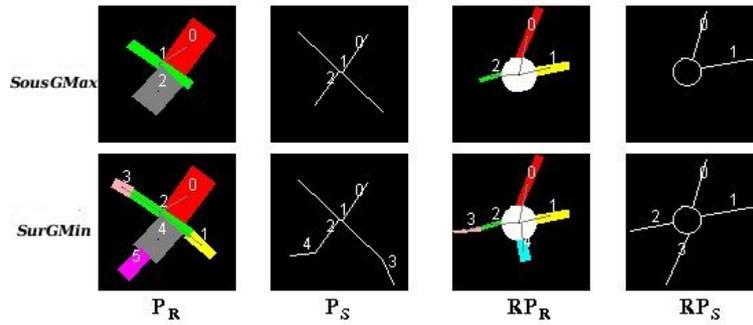


FIG. 3.11 – Les modèles obtenus.

de l'ARG de chaque objet au modèle de sa catégorie. La distance moyenne est 0.098 pour  $P_S$ , 0.1572 pour  $RP_S$ , 0.165 pour  $P_R$  et 0.1874 pour  $RP_R$ .

### 3.6 Conclusion

Dans ce chapitre nous avons présenté une méthode de génération de modèle explicite structural pour les ponts et les ronds-points. Notre principal objectif était de trouver un modèle général qui englobe les ARGs représentant les objets d'une même catégorie. Nous avons atteint notre objectif en sélectionnant les ARGs les plus représentatifs par une approche heuristique et en introduisant une architecture de treillis qui correspond à un intervalle d'ARGs.

Tout au début de notre travail, lors de nos premières réflexions sur le problème envisagé, nous nous étions posé une question : “Si on demande à quelqu'un de dessiner un rond-point (ou un pont) quelle serait la forme que cette personne tracerait ?” La réponse, ce que nous appelons un modèle d'objet manuel, était en général une esquisse constituée des formes géométriques simples. Les modèles que nous avons générés sont visuellement semblables aux modèles d'objets manuels.

La sélection des attributs des ARGs est sans doute un critère très important de la représentation. La gestion des relations spatiales en particulier sans référentiel connu est un problème majeur. Par exemple une représentation formelle de la notion de “autour” est difficile à construire, d'autant plus quand le nombre d'éléments composants varie. De cette problématique très complexe, nous avons dû choisir une représentation simplifiée en utilisant un nombre réduit d'attributs discrétisés. Par exemple les relations spatiales sont représentées par l'attribut “orientation” qui prend seulement deux valeurs différentes. Par conséquent les modèles obtenus sont très minimalistes. Cela peut présenter un obstacle important pour l'utilisation des modèles pour la reconnaissance des objets cibles dans les images réelles. En effet, l'étape d'extraction est alors susceptible de produire plusieurs fausses détections, surtout des primitives qui n'appartiennent pas à l'objet. Dans cette étape de la thèse, nous avons éprouvé les difficultés d'une représentation spatiale explicite efficace pour le problème de la reconnaissance dans les images réelles. Pourtant,

nous pensons que dans l'avenir ces modèles explicites seront nécessaires pour l'interaction avec les utilisateurs novices (les thématiciens). Le modèle d'objet en tant que tel ne nous permet pas de combler le fossé sémantique.

On pourrait enrichir le modèle principalement de deux manières :

- Par l'utilisation de la logique floue pour exprimer les valeurs numériques. De cette façon on peut convenablement intégrer les propriétés des primitives dans une représentation symbolique.
- Par l'utilisation du raisonnement spatial qualitatif pour représenter les relations entre les primitives. Deux très bons exemples sont présentés dans [Colliot, 2003] où on se sert du raisonnement spatial pour la segmentation de structures cérébrales, et dans [Bloch *et al.*, 2006] où on modélise la relation “entre” en utilisant les outils du raisonnement spatial.

En tout état de cause, les algorithmes proposés sont génériques pour de telles problématiques d'une nature structurelle.

En conclusion, une alternative est de rechercher un autre type de modèle d'objet, plus souple, qui ne vise pas à représenter l'objet explicitement. Les méthodes de reconnaissance d'objets récentes utilisent de plus en plus souvent une approche statistique qui considère le problème de la reconnaissance comme un problème d'apprentissage où on apprend un modèle implicite de l'apparence des composants de l'objet à partir d'un ensemble d'images. Nous avons choisi de suivre cette alternative qui est très peu explorée en imagerie satellitaire.

## Troisième partie

# Modèles implicites d'objets fondés sur l'apparence

Dans les dernières années les méthodes de reconnaissance d'objets fondées sur l'apparence de l'objet ont donné des résultats très encourageants pour résoudre de nombreux problèmes essentiels de reconnaissance. Cette approche relativement nouvelle correspond à un changement radical de paradigme dans ce domaine : contrairement à l'approche centrée-objet qui cherche le modèle de l'objet réel tri-dimensionnel à partir des différentes images de l'objet, l'approche fondée sur l'apparence représente un objet par l'apparence des régions locales significatives qui subsistent dans plusieurs images. Cela permet de poser le problème de reconnaissance comme un problème d'apprentissage supervisé et d'utiliser des méthodes statistiques pour l'apprentissage du modèle d'objet.

Cette troisième partie montre comment on peut appliquer cette approche sur notre problème de reconnaissance.

Dans le **chapitre 4**, on présente une étape initiale d'extraction des primitives géométriques. Ces primitives représentent les composants éventuels de l'objet et seront ensuite utilisées par les deux méthodes que nous avons développées pour calculer un ensemble de descripteurs. Les primitives sont des formes géométriques simples et sont extraites par deux approches complémentaires en parallèle, une approche par régions et une approche par contours.

Dans le **chapitre 5** on construit un modèle implicite d'objet en utilisant l'algorithme *Ada-boost* qui fusionne des classifieurs faibles pour en obtenir un classifieur fort. Les classifieurs faibles sont obtenus en groupant les primitives dans des mots visuels et en comptant la fréquence des mots différents.

Le **chapitre 6** est consacré à l'apprentissage d'un modèle d'objet fondée sur une approche "composants et structure". On sélectionne un ensemble de primitives significatives qui correspondent aux composants de l'objet et on représente la structure par les propriétés géométriques et l'organisation spatiale des primitives. L'apprentissage est effectué par un algorithme de *clus-*

*tering.*

Les deux méthodes proposées sont testées sur un problème de classification où on vise à détecter les ronds-points et les ponts parmi les images appartenant aux dix catégories d'objets.

Nous avons présenté les résultats de nos travaux de recherche concernant cette partie dans trois congrès internationaux ([Erus et Loménie, 2006], [Erus et Loménie, 2007] et [Erus et Loménie, 2008])

# Chapitre 4

## Extraction des primitives

### 4.1 Introduction

Dans ce chapitre, nous allons présenter la chaîne de traitement développée pour extraire des primitives géométriques à partir des images d'objets. Ces primitives seront utilisées dans les étapes suivantes pour l'apprentissage du modèle et la reconnaissance d'objets.

Nous proposons d'extraire les primitives structurelles d'objets cartographiques pour deux raisons principales :

1. Il y a une variance importante de taille, de rotation, d'illumination et de forme entre les objets d'une même classe. Les définitions des classes appartiennent à des concepts sémantiques et il n'y a pas de correspondance radiométrique directe entre les composants d'objets. Pourtant les objets cartographiques ont des structures bien définies et en général géométriques.
2. A basse résolution, les objets disparaissent et deviennent une partie de la texture. Les méthodes de bas niveau, au niveau pixel, sont appropriées pour les images à basse résolution. A haute résolution les propriétés structurelles discriminantes apparaissent et les approches de haut niveau sont envisageables, voire nécessaires.

L'extraction des primitives est effectuée en général par deux approches classiques :

1. **L'approche par régions**, où on cherche à détecter des régions qui appartiennent à l'objet et qui ont une forme géométrique particulière ;
2. **L'approche par contours**, où on cherche les primitives géométriques dans les contours de l'objet.

Nous avons développé deux méthodes d'extraction en parallèle en utilisant les deux approches. Dans la première approche les primitives détectées sont **les segments de droites** et **les cercles**. Dans la deuxième approche on extrait **les segments de droites**, **les arcs de cercles** et **les blobs**. De cette façon, on obtient des informations complémentaires sur la structure des objets.

## 4.2 Extraction par régions

Nous avons rencontré des difficultés importantes pour une segmentation directe de nos images :

- Sur les images panchromatiques les objets n’ont pas une radiométrie homogène ;
- Il est souvent difficile même pour l’œil humain de séparer l’objet du fond.

La 4.1 montre deux images qui illustrent bien ces difficultés. On observe que pour s’approcher de la performance de l’opérateur humain il est nécessaire d’inclure dans le traitement une importante quantité d’informations contextuelles de haut niveau.

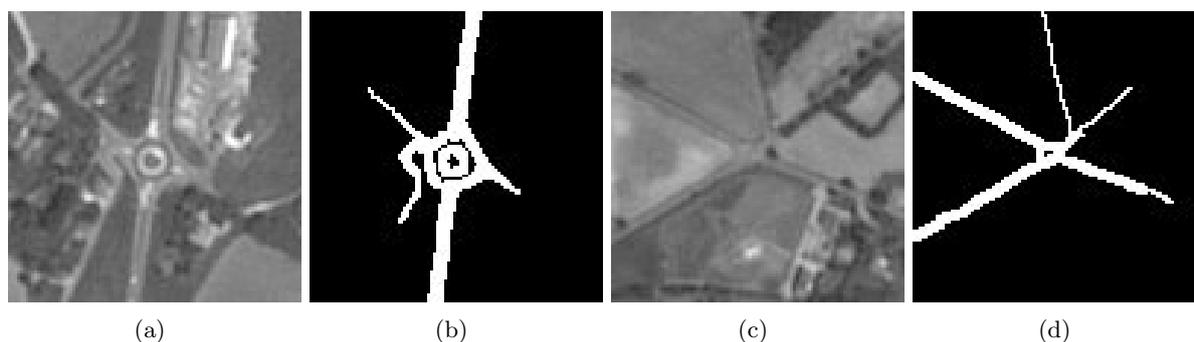


FIG. 4.1 – Exemples d’imagettes et leur segmentation par un expert. (a). Image de rond-point, (c). image de carrefour routier, (b)-(d). segmentation par un expert.

Dans un premier temps, nous avons essayé d’obtenir une segmentation qui est proche de la segmentation de l’expert. En observant que cet objectif était trop ambitieux pour une partie importante des images, dans un deuxième temps, nous avons cherché à détecter les régions qui, éventuellement, font partie de l’objet, d’après leurs propriétés géométriques, au lieu de séparer l’objet du fond. Pour cela, nous proposons d’utiliser des algorithmes de fusion où on fait une sur-segmentation de l’image et on fusionne les régions en utilisant un critère de similarité. Ces méthodes nous semblent être bien adaptées à notre problème, car elles fournissent une segmentation multi-échelle de l’image qui couvre des régions de différentes tailles. Nous avons cherché les régions qui ont une forme géométrique régulière pour détecter les composants d’objets cartographiques. Ce sont **les cercles** et **les rectangles allongés** qui correspondent aux régions circulaires et linéaires.

### 4.2.1 Segmentation par lignes de partage des eaux (*Watershed*)

L’algorithme de lignes de partage des eaux [Meyer et Beucher, 1990] est très souvent utilisé pour une segmentation initiale. L’inconvénient majeur de cette approche est le nombre élevé de régions initiales dû à la sur-segmentation de l’image. On obtient une région pour chaque minimum local et par conséquent, l’algorithme est très sensible au bruit. Un lissage gaussien peut améliorer le résultat, mais au coût d’une réduction de la résolution de l’image, donc cela

n'est pas préférable. En général, la segmentation est suivie d'une fusion de régions itérative. Pour l'image de rond-point exemplaire, nous avons obtenu 876 régions initiales. La figure 4.2 montre la segmentation initiale et les régions obtenues par la fusion aux itérations 776 et 836. Cette opération, étant très coûteuse, ne constitue pas une solution envisageable.

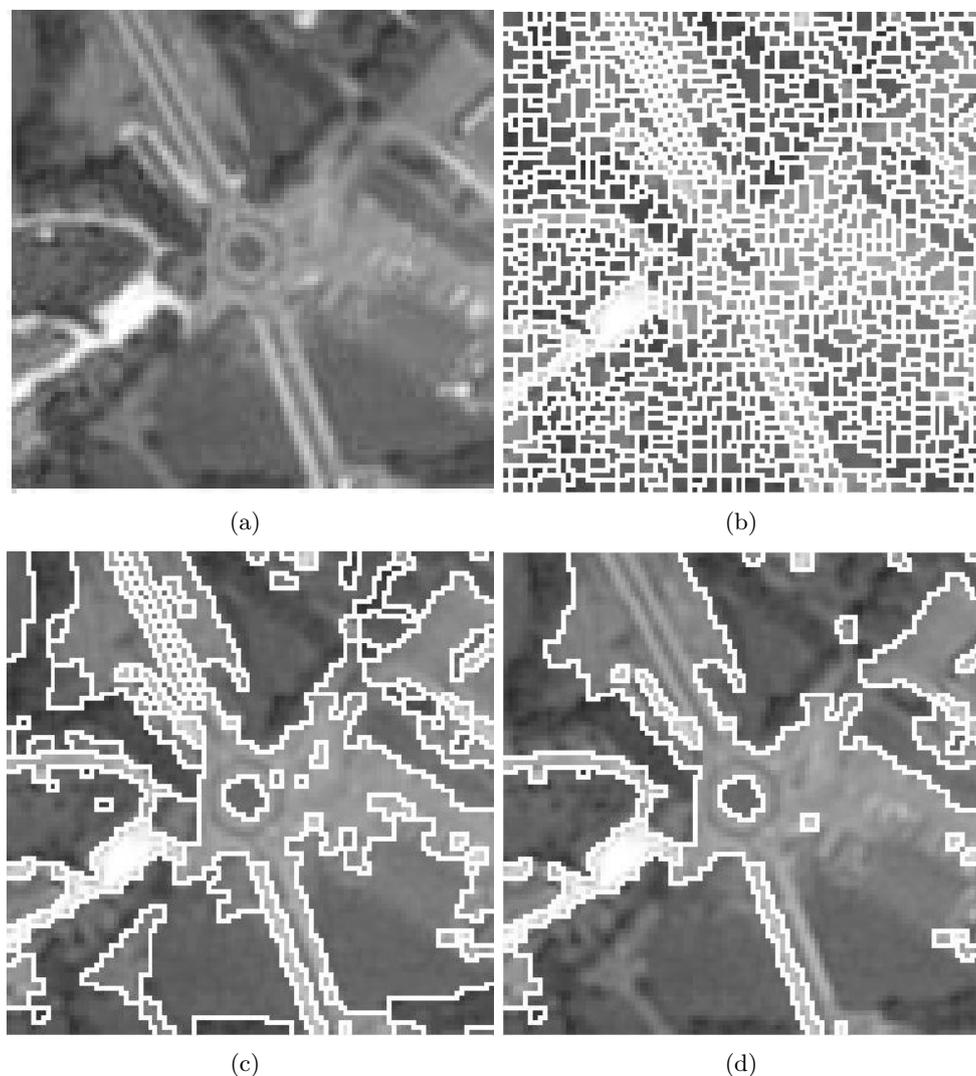


FIG. 4.2 – Segmentation par *Watershed* sur une image de rond-point et la fusion des régions. (a). Image originale, (b). segmentation initiale, (c). fusion avec 100 régions, (d). fusion avec 40 régions.

#### 4.2.2 *Watershed* avec marqueurs

Dans [Erus et Loménie, 2006] nous avons proposé une méthode de segmentation d'images satellitaires en utilisant les images multispectrales ( $I_m$ ) pour trouver les marqueurs (Figure 4.3). Sur l' $I_m$ , contrairement aux images panchromatiques ( $I_p$ ), il est possible de détecter l'objet en utilisant les caractéristiques radiométriques des trois canaux. Pourtant la détection est grossière à

cause de la résolution de 10 mètre par pixel. L'idée principale est d'utiliser  $I_m$  pour la détection et de raffiner le résultat en utilisant l'algorithme de *Watershed* avec marqueurs sur  $I_p$ . On trouve un masque correspondant à l'objet par un seuillage par hystérésis sur les trois canaux de  $I_m$ . On localise le masque sur  $I_p$ . Le marqueur externe est le contour du masque dilaté, et le marqueur interne est le squelette du masque. On applique le *Watershed* avec marqueurs pour détecter l'objet. Les marqueurs obtenus sont illustrés sur la figure 4.4.

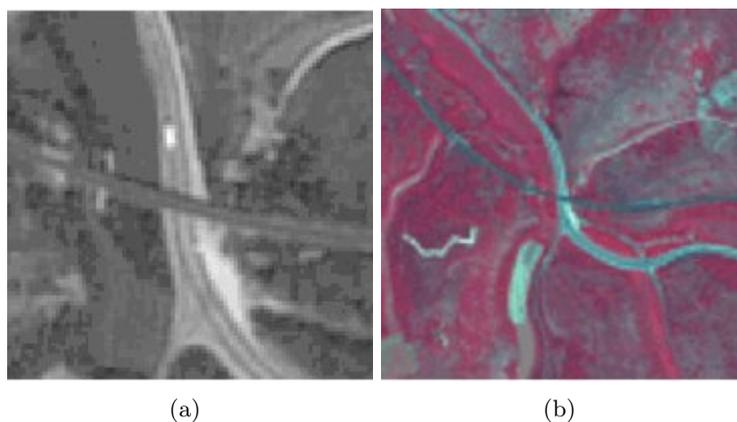


FIG. 4.3 – Image d'un rond-point. (a). image panchromatique, (b). image multispectrale.

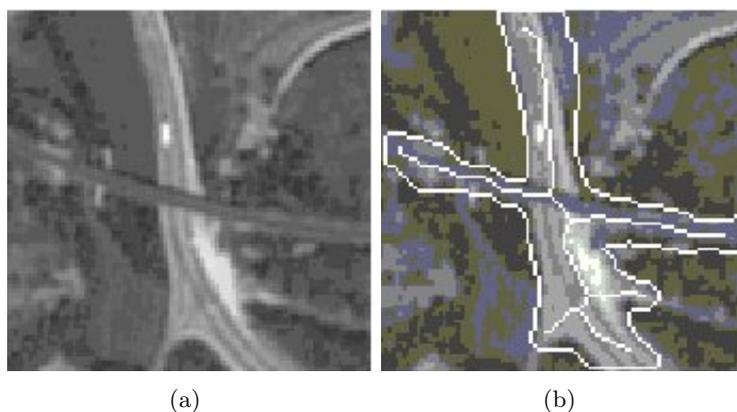


FIG. 4.4 – Les marqueurs obtenus sur une image de rond-point. (a). Image panchromatique, (b). les marqueurs

En utilisant cette méthode on arrive à extraire l'objet cible sur l'image (Figure 4.5). Pourtant l'extraction finale présente des imprécisions, particulièrement quand l'objet se trouve dans un environnement urbain, à cause des similitudes radiométriques entre l'objet et le fond, et quand l'objet est de très petite taille pour être détecté sur  $I_m$ . Un autre inconvénient est l'utilisation des images multispectrales qui ne sont pas toujours disponibles. En conséquence, nous avons conclu qu'on a besoin d'une approche plus souple qui ne cherche pas à isoler l'objet du fond pour en extraire les primitives géométriques, mais à détecter les composants éventuelles de l'objet.

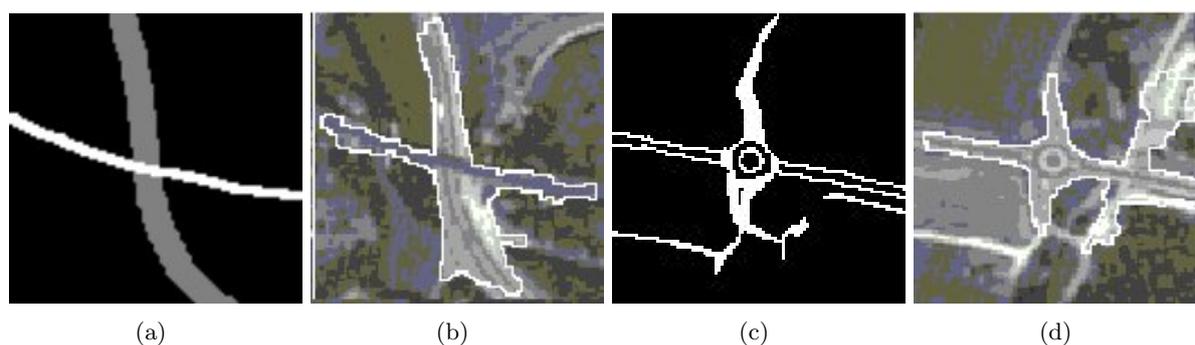


FIG. 4.5 – Les objets détectés par *Watershed* avec marqueurs. (a)-(c) Segmentation par un expert, (b)-(d) extraction par *Watershed* avec marqueurs.

### 4.2.3 Segmentation par *Mean-Shift*

Nous avons appliqué une segmentation *Mean-Shift* en utilisant le système EDISON [2003]. Avec une fenêtre de recherche de taille 6 dans le domaine d'intensité et de 15 dans le domaine spatial, la segmentation a donné 162 régions pour l'image de rond-point. La figure 4.6.b montre les régions obtenues après la segmentation initiale.

Nous avons utilisé les deux métriques suivantes pour détecter les primitives :

- *Circularité* pour la détection des primitives circulaires ;
- *Excentricité* pour la détection des primitives rectangulaires.

Les régions sont triées d'après leurs scores de circularité et d'excentricité, et les régions régulières ont été sélectionnées par un simple seuillage. La figure 4.7 montre les régions sélectionnées à la fin de la segmentation initiale d'une image de rond-point.

### 4.2.4 Fusion des régions

Les composants des objets sont généralement segmentés en plusieurs régions. Une fusion itérative des régions est proposée pour obtenir une segmentation plus grossière. Pour chaque paire de régions adjacentes, une métrique de similarité est calculée. A chaque itération, les deux régions les plus similaires sont fusionnées jusqu'à ce que le nombre total des régions soit inférieur à un seuil (La fusion peut itérer jusqu'à l'obtention d'une seule région mais nous avons choisi d'arrêter à 20 régions par souci d'efficacité, car à ce niveau la taille des régions obtenues dépasse la taille de l'objet). En tant que métrique, nous avons utilisé la similarité radiométrique. Une métrique plus sophistiquée, par exemple en utilisant les propriétés géométriques des régions est concevable, pourtant cela augmenterait le temps de fusion. La matrice d'adjacence et la matrice de similarité sont calculées une seule fois après la segmentation initiale, et sont mises à jour à chaque itération. La figure 4.6.c,d montre la segmentation aux itérations 60 et 120.

En ajoutant aux primitives initiales les primitives obtenues par la fusion des régions (celles qui correspondent à des primitives à un espace d'échelle supérieur) nous avons construit un

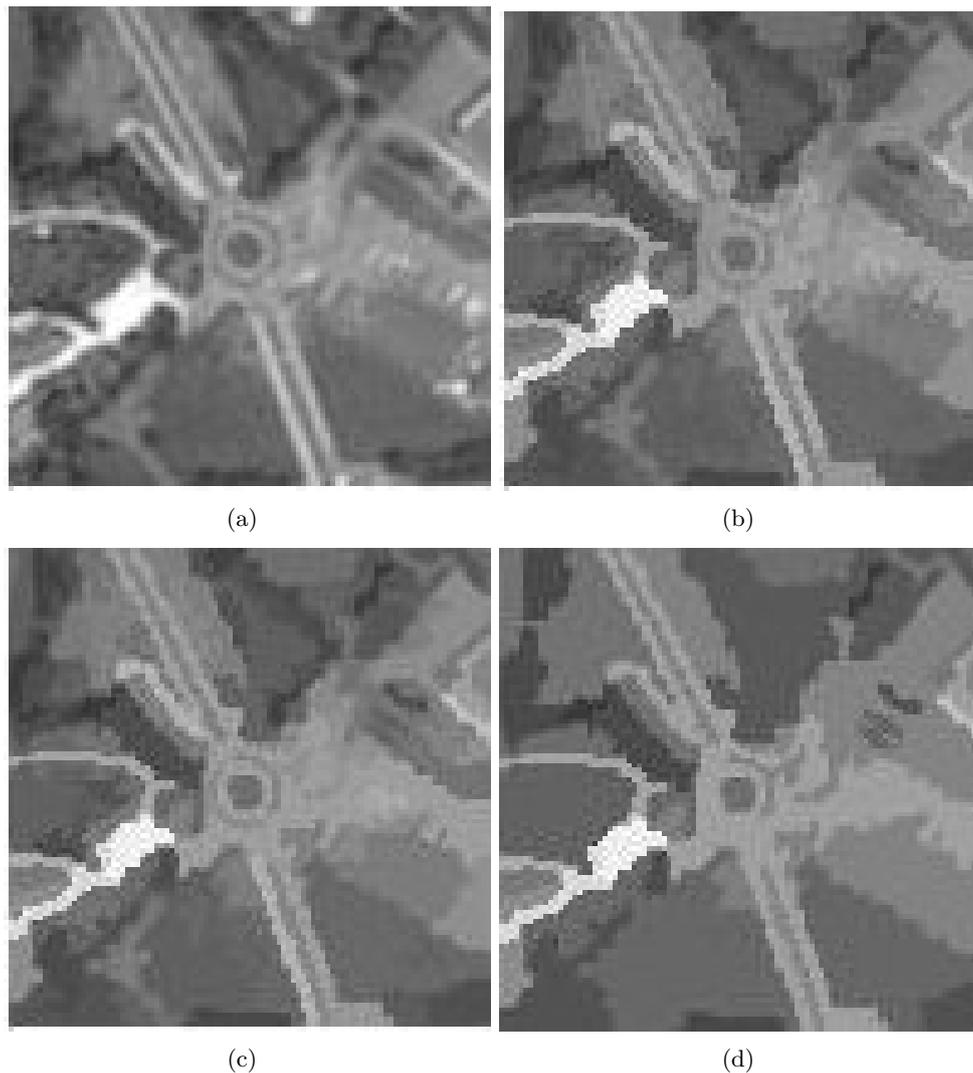


FIG. 4.6 – Segmentation par *Mean-Shift* d'une image de rond-point. (a). Image originale, (b). image segmentée, (c). fusion à l'itération 60 avec 102 régions, (d). fusion à l'itération 120 avec 42 régions.

ensemble de primitives circulaires et linéaires.

#### 4.2.5 Les résultats de l'extraction

Les résultats de l'extraction sont très bonnes : on a détecté les ronds centraux pour toutes les images de ronds-points sauf quelques unes avec une taille très petite. On a détecté toutes les routes principales. On a eu des détections multiples à cause de l'approche multi-échelle et on a détecté des primitives qui n'appartiennent pas aux objets. Mais cela était prévisible et les fausses détections seront éliminées par les algorithmes utilisés aux étapes suivantes. Les résultats de l'extraction sur deux images exemplaires sont présentés à la figure 4.8.

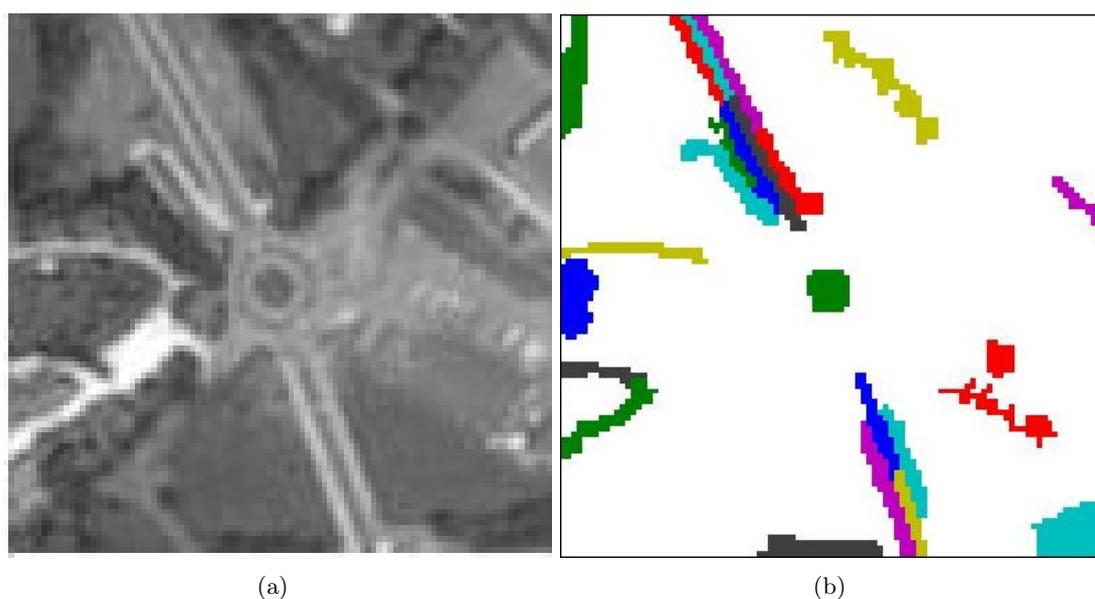


FIG. 4.7 – Les régions sélectionnées après la segmentation initiale. (a). Image originale, (b). les régions circulaires et linéaires.

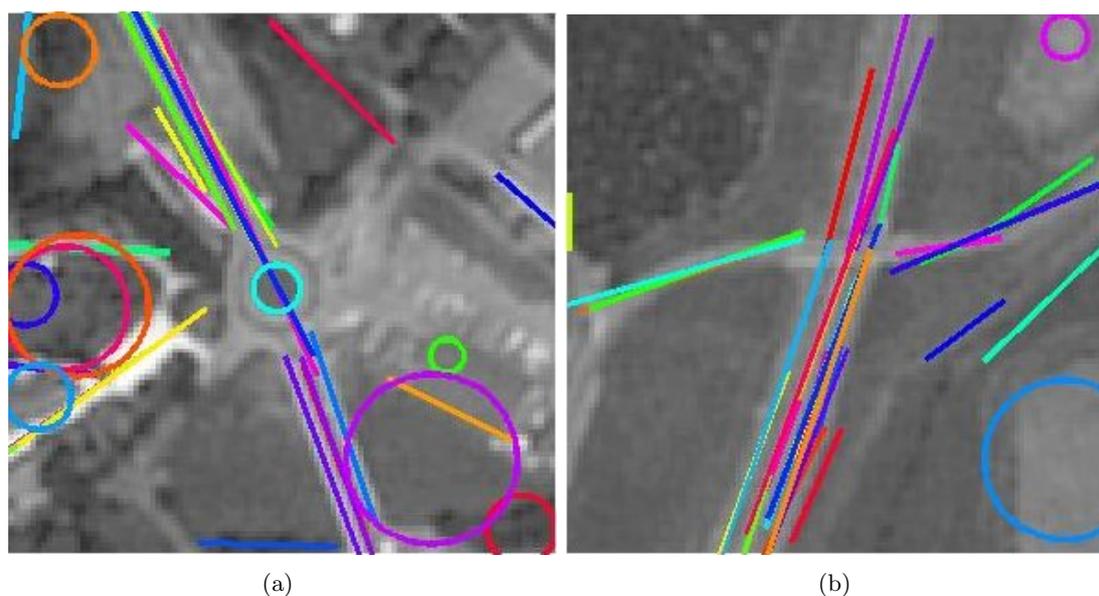


FIG. 4.8 – L'ensemble des primitives obtenues par l'approche par régions. (a). Image d'un rond-point, (b). image d'un pont.

### 4.3 Extraction par contours

Les catégories d'objets cibles ont des composants à une forte structure prononcée correspondant à des formes géométriques simples. En utilisant les contours il est possible de trouver des descripteurs qui représentent les propriétés structurelles des objets. Nous appelons nos descripteurs "les primitives géométriques simples". Ce sont des formes géométriques, *les segments*, *les*

*cercles* et *les blobs*. Il existe une relation directe entre les primitives et les objets recherchés : les formes linéaires correspondent aux routes, *les blobs* aux bâtiments isolés et les cercles aux centres des ronds-points.

Pour assurer l'invariance à l'échelle, nous avons fait l'extraction sur une pyramide d'images gaussiennes. La pyramide est obtenue par la convolution de l'image par une série de filtres gaussiens  $G_\sigma(x, y)$  avec des valeurs de  $\sigma$  qui augmentent. Nous avons utilisé trois niveaux d'échelle avec  $\sigma = \{1, 2, 3\}$ . La figure 4.9 montre le pyramide construit à partir d'une image de rond-point pour les trois niveaux d'échelles.

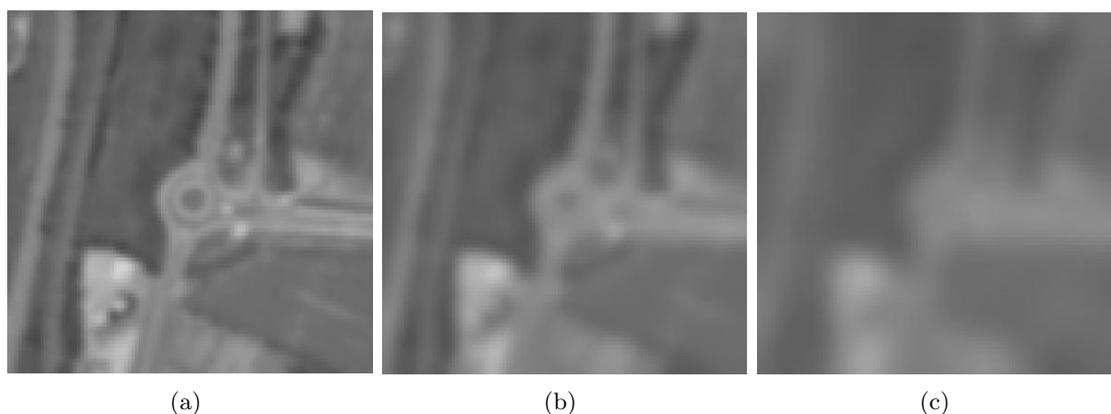


FIG. 4.9 – Le pyramide gaussien d'une image de rond-point. (a).  $\sigma = 1$ , (b).  $\sigma = 2$ , (c).  $\sigma = 3$ .

### 4.3.1 Les étapes de l'extraction

L'extraction est effectuée par la chaîne de traitement suivante :

1. La détection des contours au niveau sub-pixellique.
2. L'enchaînement des points de contours
3. La détection des *blobs* : les chaînes de contours fermées.
4. La polygonisation : l'approximation des chaînes de contours par des segments.
5. La détection des arcs de cercles.

Nous explicitons ces étapes ci-après.

### 4.3.2 Détection des contours sub-pixelliques

Nous avons utilisé le détecteur de Canny [1986] pour détecter les contours. Le détecteur de Canny a trois étapes principales :

- **Calcul du gradient de l'image** : on trouve la norme et la direction du gradient sur chaque pixel. La norme du gradient est faible sur les régions uniformes, et élevée sur les zones de transition.

- **La suppression des non-maxima locaux** : on élimine les pixels qui n'ont pas une norme maximale dans la direction du gradient. Cela garantit l'obtention des contours d'épaisseur un pixel. Pour savoir si un point  $A$  est un maximum local, on trouve les deux voisins  $B$  et  $C$  à la direction du gradient dans un voisinage de  $3 \times 3$ , on calcule la norme à  $B$  et  $C$  par une interpolation linéaire ou quadratique de leurs voisins, et on compare la norme de  $A$  avec les normes des trois points  $A$ ,  $B$  et  $C$ .
- **Le seuillage par hystérésis** : on utilise deux seuils  $t_{haut}$  et  $t_{bas}$  pour choisir les points de contours finals. Les pixels qui ont une intensité supérieure à  $t_{haut}$  sont directement acceptés, les pixels qui ont une intensité inférieure à  $t_{bas}$  sont directement rejetés. Les autres pixels sont acceptés s'ils sont connectés à des pixels déjà acceptés. Cela évite d'avoir des pixels isolés tout en étant plus souple qu'un seuillage direct.

La qualité des primitives obtenues dépend de la précision de l'approximation des contours par les segments. Comme la taille des imageries est petite (de 100 à 125 pixels), les résultats d'un détecteur de Canny classique manquent de précision. Une solution est d'augmenter la taille de l'image par un facteur de  $n$  en calculant les valeurs des pixels qui manquent sur la nouvelle image par une interpolation. Le grand inconvénient de cette solution est le temps de calcul qui est multiplié par un facteur de  $n^2$  pour toutes les opérations suivantes.

Une solution efficace et rapide est de détecter les contours subpixeliques. Nous avons implémenté un algorithme de détection de contours qui détecte la position des contours avec une précision inférieure au pixel. La méthode, proposée par Devernay [1995] consiste à ajouter à l'algorithme de Canny une étape d'estimation de la position sub-pixelique des points de contour par une interpolation quadratique, pendant la suppression des non-maxima locaux.

- Soit  $A$  un maxima local et  $B$ ,  $C$  les deux voisins dans la direction du gradient, la position du point de contour  $M$  est estimée par le maximum de l'interpolation quadratique des normes de  $A$ , de  $B$  et de  $C$  (la figure 4.10)

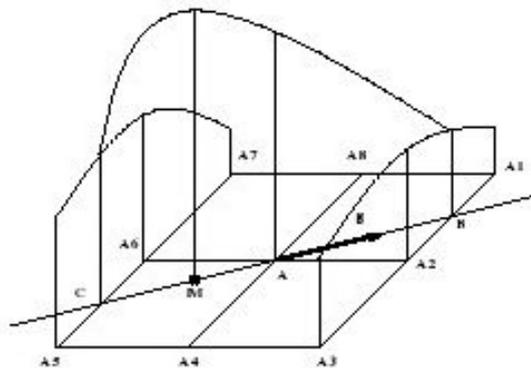


FIG. 4.10 – Illustration de la détection du point de contour subpixelique  $M$ .

A côté des contours subpixeliques on garde aussi une image binaire représentant les contours discrets de l'image. L'utilisation de cette carte de contours facilite la mise en oeuvre de l'algo-

rithme de chaînage suivant, sans avoir besoin de faire des calculs sur les contours subpixelliques. La figure 4.11 montre les contours obtenus avec l’algorithme de Canny classique (a) et au niveau subpixellique (b).

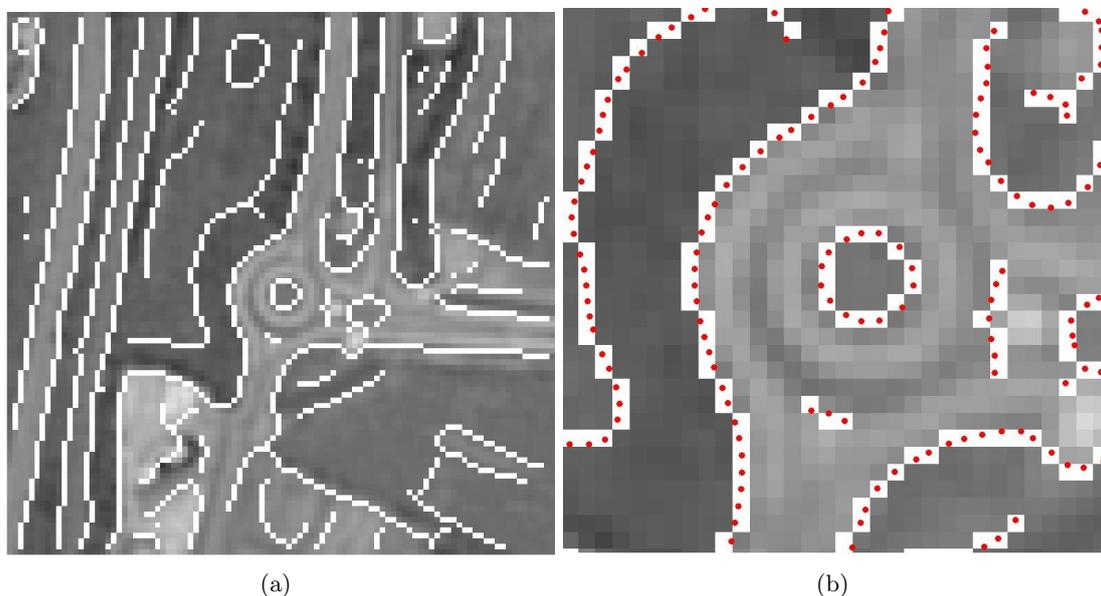


FIG. 4.11 – Les contours obtenus par un détecteur de Canny et les contours subpixelliques. (a) Les contours obtenus par un détecteur de Canny (les pixels blancs) sur une image de rond-point, (b). les contours subpixelliques superposés (les points rouges) à la région centrale de l’image.

### 4.3.3 Chaînage des contours

Le détecteur de Canny donne des points de contours qui ne sont pas connectés. Pour extraire les segments à partir des contours de l’image, nous avons besoin de grouper les points de contours en chaînes de contours. Nous avons utilisé l’algorithme de chaînage de contours proposé par Kovese [2000](Algorithme 5).

La figure 4.12 présente les chaînes de contours obtenus à la fin de l’étape de chaînage pour les contours subpixelliques extraits aux trois niveaux d’échelles.

### 4.3.4 Détection des *blobs*

Les *blobs* sont surtout utiles pour détecter les objets de petite taille, particulièrement les bâtiments. On considère les chaînes de contours fermés ou quasi-fermés comme des *blobs*. On les détecte en comparant la distance entre les deux extrémités de chaque chaîne de contour par un seuil de distance. Après la détection des *blobs*, on fait un deuxième test pour sélectionner ceux qui ont une forme circulaire. Les *blobs* qui ont un score de circularité élevé sont considérés comme des cercles. Pour les *blobs* de petites tailles, le score de circularité est en général élevé et a une variance très faible. Nous avons choisi de ne pas classifier ces *blobs* en tant que cercle

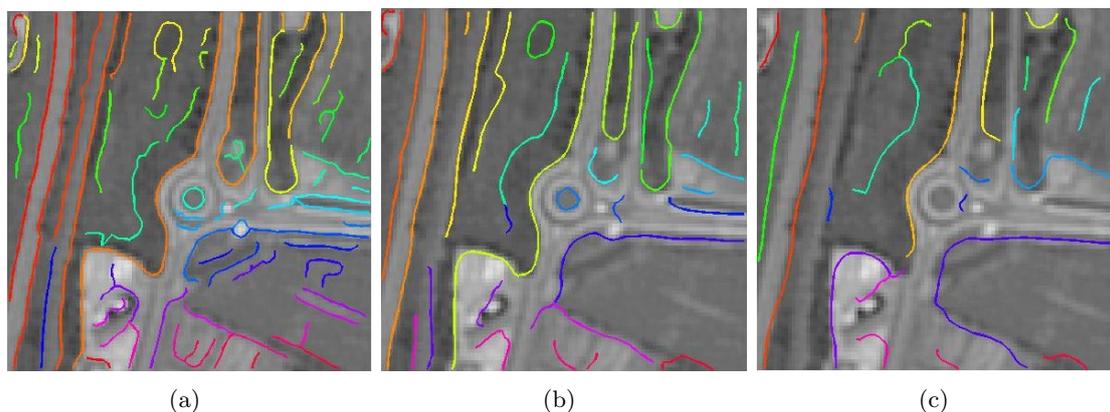
**Algorithm 5** Algorithme de chaînage de contours**Entrées:** $\mathbf{p} = \{\mathbf{p}_x, \mathbf{p}_y\}$  : la liste de points de contour**Sorties:** $\mathbf{c}$  : la liste de chaînes de contours**Pour**  $\forall p_i \in \mathbf{p}$  **Faire**Calculer  $t_i$  : le nombre de passages de 0 à 1 ou de 1 à 0 autour de  $p_i$  ((Une fenêtre glissante de 3x3 pixels est utilisée pour calculer  $t_i$ )Trouver  $e_i$  l'étiquette de  $p_i$ , tels que  $e_i \in \{ \text{extrémité } (E), \text{ corps } (C), \text{ jointure } (J) \}$ .**Si**  $t_i \geq 6$  **alors** $e_i = J$ ;**Sinon****Si**  $t_i = 6$  **alors** $e_i = E$ ;**Sinon** $e_i = C$ ;**Fin Si****Fin Si****Fin Pour****Pour**  $\forall p_i \in \mathbf{p}$  **Faire****Si**  $e_i \geq J$  ou  $e_i \geq E$  **alors** $c_i$  = la chaîne de contour à partir de  $p_i$  jusqu'à une autre extrémité ou jointure.**Sinon** $c_i$  = la fusion de deux chaînes de contours de  $p_i$  jusqu'à une autre extrémité ou jointure.**Fin Si**Supprimer les points appartenant à  $c_i$  dans  $p$ **Fin Pour****Retourner**  $\mathbf{c}$ 

FIG. 4.12 – Les chaînes de contours après la fusion des points de contours pour trois niveaux d'échelles. (a).  $\sigma = 1$ , (b).  $\sigma = 2$ , (c).  $\sigma = 3$ .

pour éviter d’avoir un nombre important de faux cercles, en fixant un seuil de rayon minimal de 3 pixels. A une résolution de 2,5 mètre par pixel cela correspond à un seuil de rayon de 7,5 mètres. La figure 4.13 montre les *blobs* détectés sur une image de rond-point et une image de bâti-isolé.

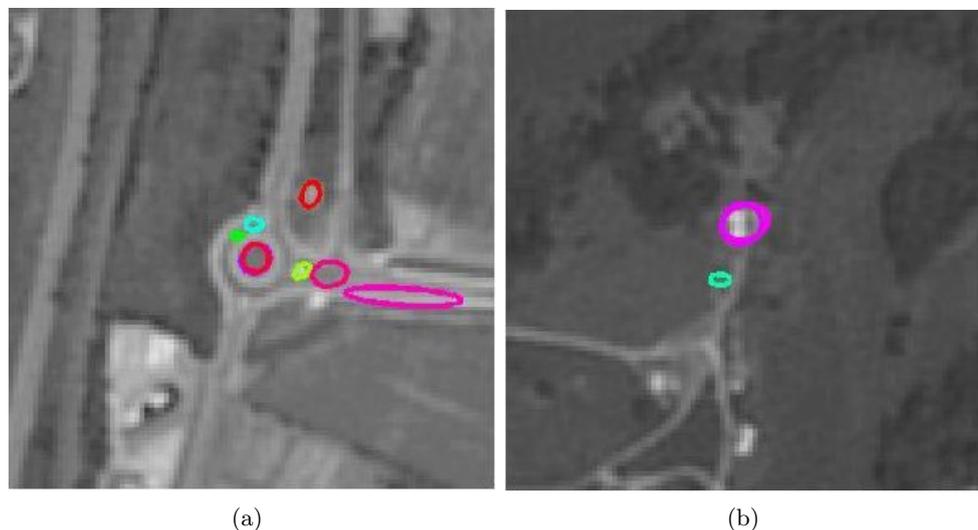


FIG. 4.13 – Les *blobs* détectés par l’approche par contours. (a). Image de rond-point, (b). image de bâti-isolé.

#### 4.3.5 Détection des segments de droites

Les chaînes de contours obtenues à la fin de l’étape précédente sont des courbes. Nous avons utilisé l’algorithme de la corde pour une approximation polygonale pour obtenir des segments de droites. Le paramètre de la tolérance est déterminé empiriquement mais en fonction de l’échelle de l’image. Les segments détectés sont présentés à la figure 4.14).

#### 4.3.6 Détection des arcs de cercles

Une bonne détection des arcs de cercles est nécessaire pour distinguer les ronds-points et les bâtis isolés. Nous proposons un algorithme de régression de points par un cercle (*circle fit*). L’idée principale est de calculer un score d’approximation par un cercle pour chaque paire de segments adjacents et de le comparer avec l’erreur d’approximation par un segment.

Il est possible d’approximer une chaîne de contour par un cercle par la résolution d’une équation de moindres carrés qui minimise la somme des carrés de la différence entre la distance de chaque point au centre hypothétique du cercle et le rayon du cercle.

L’équation d’un cercle avec le centre  $(a, b)$  et le rayon  $r$  est

$$(x - a)^2 + (y - b)^2 = r^2$$

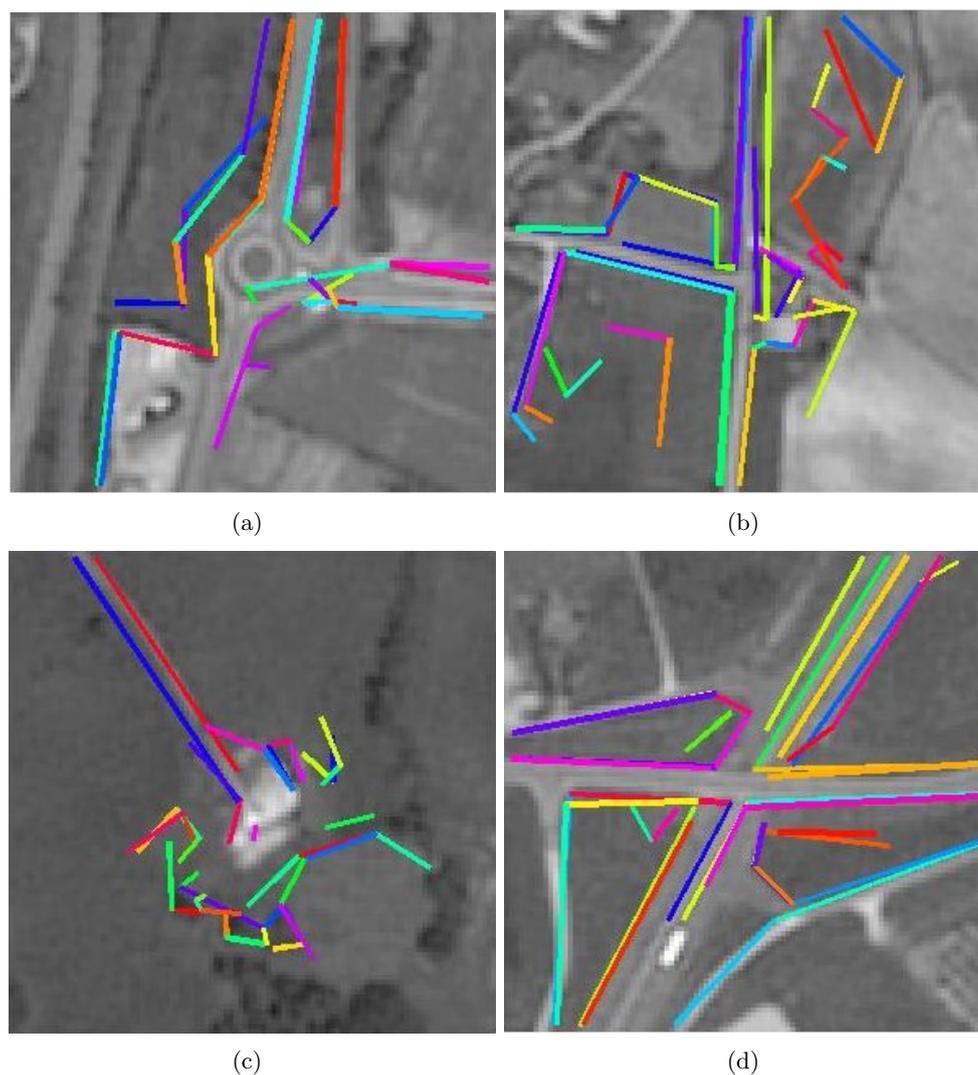


FIG. 4.14 – Les segments détectés par l'approche par contours. (a). Image d'un rond-point, (b). image d'un carrefour routier, (c). image d'un bâti-isolé, (d). image d'un pont.

En développant cette équation, on a :

$$\begin{aligned}
 (x - a)^2 + (y - b)^2 - r^2 &= 0 \\
 x^2 - 2ax + a^2 + y^2 - 2by + b^2 &= r^2 \\
 x^2 + y^2 - 2ax - 2by + a^2 + b^2 - r^2 &= 0 \\
 -2ax - 2by + a^2 + b^2 - r^2 &= x^2 + y^2
 \end{aligned}$$

Par le changement de variables

$$p = -2a, q = 2b, t = a^2 + b^2 - r^2$$

on obtient

$$px + qy + t = x^2 + y^2$$

L'expression matricielle de cette équation pour  $n$  points avec  $\overline{X} = \{x_1, \dots, x_n\}$  et  $\overline{Y} = \{y_1, \dots, y_n\}$  est

$$\begin{bmatrix} \overline{X} & \overline{Y} & \mathbf{1} \end{bmatrix} \cdot \begin{bmatrix} p & q & t \end{bmatrix}^T = \begin{bmatrix} \overline{X}^2 + \overline{Y}^2 \end{bmatrix}$$

On calcule les variables inconnues par

$$\begin{bmatrix} p & q & t \end{bmatrix}^T = \begin{bmatrix} \overline{X} & \overline{Y} & \mathbf{1} \end{bmatrix} \setminus \begin{bmatrix} \overline{X}^2 + \overline{Y}^2 \end{bmatrix}$$

où l'opérateur  $x = A \setminus B$  signifie la solution par les moindres carrés de l'équation  $Ax = B$ . On calcule d'abord  $\begin{bmatrix} p & q & t \end{bmatrix}^T$ , et puis  $(x, y, r)$ . De cette façon, on trouve une approximation d'un ensemble de points par un arc de cercle et on calcule l'erreur d'approximation.

Cette approximation est ensuite utilisée dans l'algorithme de détection des arcs (Algorithme 6).

---

**Algorithm 6** Algorithme de détection des arcs de cercles par l'approche par contours.

---

**Entrées:**

**s** : la liste des segments

**Sorties:**

**c** : la liste de chaînes de contours

**Pour**  $\forall (s_i, s_j)$  consécutives **Faire**

$p_i = pixels(s_i)$ ,  $p_j = pixels(s_j)$

$e_i = erreur\_d'approximation(s_i)$ ,  $e_j = erreur\_d'approximation(s_j)$

$s' = joindre(s_i, s_j)$

$c' = approximation\_cercle(p_2)$ ,  $e' = erreur\_d'approximation(c)$

**Si**  $e' \leq max(e_i, e_j)$  **alors**

remplacer  $s_i$  et  $s_j$  par  $c'$

Ajouter  $c'$  dans **c**

**Sinon**

remplacer  $s_i$  et  $s_j$  par  $s'$

**Fin Si**

**Fin Pour**

**Retourner c**

---

On a détecté avec succès les arcs de cercles à partir des chaînes de contours en utilisant cet algorithme. L'algorithme a une tendance à représenter les contours des petits objets par des arcs de cercles au lieu des segments. Cela est préférable parce qu'elle garantit de détecter des petits cercles, même si on a un nombre élevé de faux positifs. La figure 4.15 montre les arcs détectés

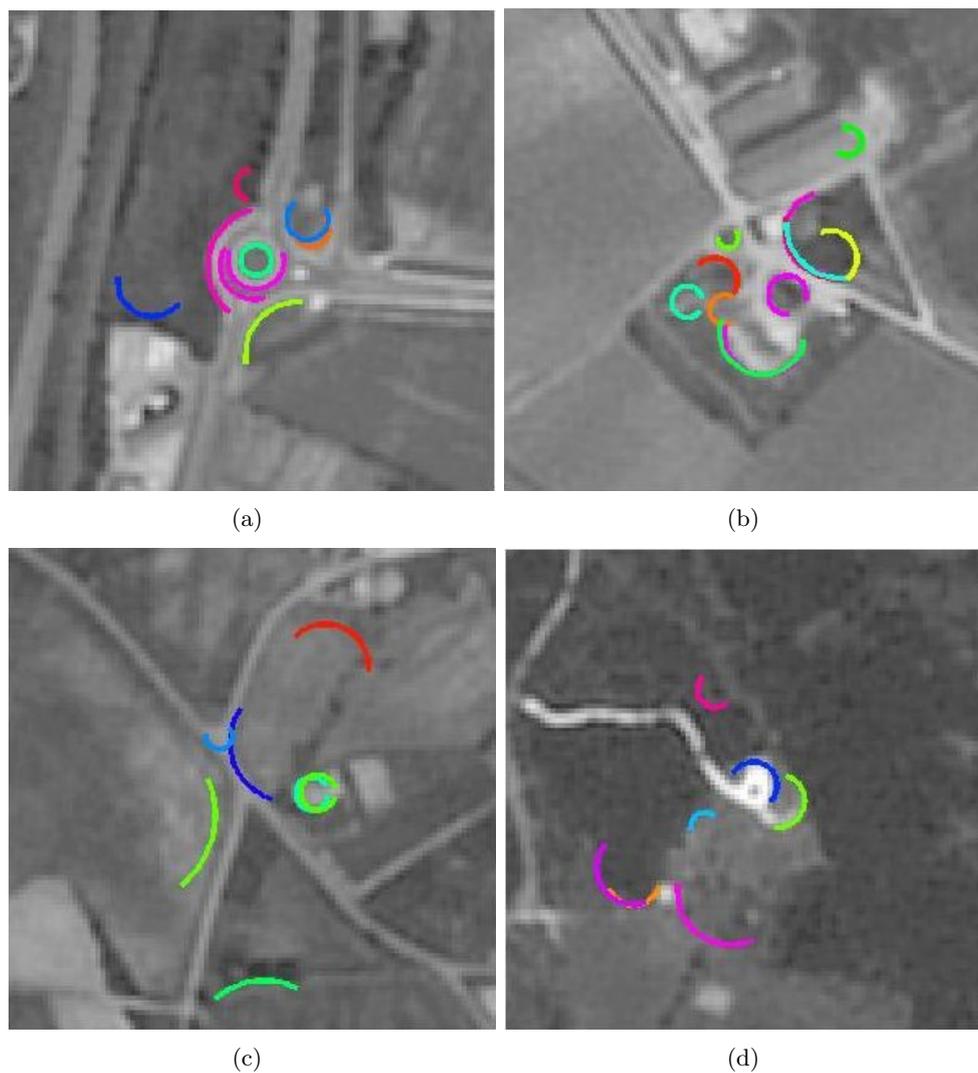


FIG. 4.15 – Les arcs de cercles détectés par l’approche par contours. (a)-(b). Image d’un rond-point, (c). image d’un carrefour routier, (d). image d’un bâti isolé.

sur l’image d’un rond-point.

La figure 4.16 montre les trois types de primitives détectées sur une image de rond-point et sur une image de pont.

## 4.4 Conclusion

Nous avons présenté dans ce chapitre deux approches complémentaires, une approche par régions et une approche par contours, pour extraire un ensemble de primitives géométriques à partir de l’image d’un objet.

L’isolement de l’objet du fond étant très difficile, nous avons visé à extraire le plus grand nombre de primitives appartenant à l’objet aux dépens d’un nombre de fausses détections élevé.

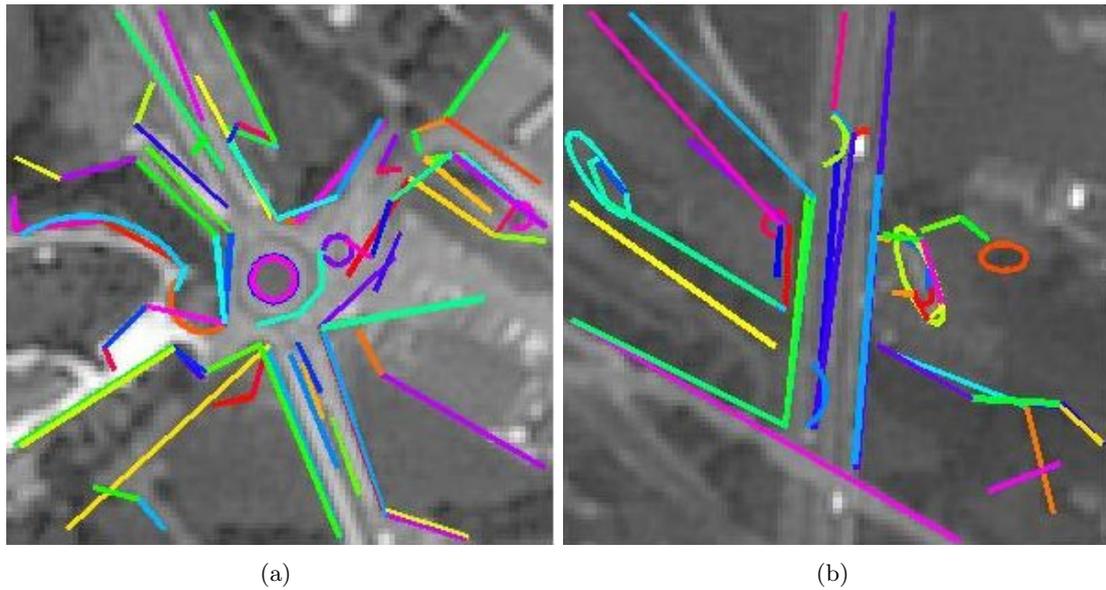


FIG. 4.16 – L'ensemble des primitives obtenues par l'approche par contours. (a). Image de rond-point, (b). Image de pont.

La sélection des primitives pertinentes sera faite aux étapes suivantes, pendant la classification. En effet on suppose que l'accumulation aléatoire des primitives fausses ne devrait pas avoir un mauvais effet systématique pour la classification de l'objet-type.

Nous avons détecté avec succès des primitives qui correspondent aux composants géométriques principaux des objets cibles.

## Chapitre 5

# La modélisation par la fusion des classifieurs faibles

### 5.1 Introduction

Nous proposons dans ce chapitre une méthode de classification supervisée d'objets cartographiques ronds-points et ponts. La méthode est fondée sur un modèle implicite construit par un classifieur *Adaboost*.

Les méthodes de classification cherchent généralement à obtenir des descripteurs compliqués qui représentent à eux seuls le plus grand nombre de propriétés de l'objet. Par contre, on peut très souvent utiliser des indices très simples pour classifier rapidement certains objets d'une catégorie même si ces indices ne sont pas suffisants pour obtenir un taux de classification élevé sur l'ensemble d'objets. Par exemple on peut facilement reconnaître les maisons (mais pas toutes) sur une image par leur toit rouge.

Le classifieur *Adaboost* nous permet d'utiliser un grand nombre de descripteurs simples, qui ne sont pas capables individuellement de bien classifier les exemples, mais qui peuvent ensemble construire un classifieur puissant.

Les étapes principales de la méthode sont :

1. l'extraction des caractéristiques,
2. l'apprentissage du classifieur, et
3. La classification.

On apprend le modèle à partir d'exemples positifs et d'exemples négatifs. On calcule un vecteur de caractéristiques à partir des primitives géométriques détectées sur la base d'images d'apprentissage. Le classifieur final est construit principalement par la fusion des classifieurs faibles obtenus des caractéristiques. Le classifieur est testé sur un ensemble d'images de test. Une définition formelle de la méthode est donnée dans l'Algorithme 7.

---

**Algorithm 7** Algorithme global de la méthode fondée sur la fusion des classifieurs faibles
 

---

**Entrées:** $\mathbf{I}_A$  : Ensemble d'images d'apprentissage positives et négatives $\mathbf{I}_T$  : Ensemble d'images de test**Sorties:** $\mathbf{c}$  : les classes des images de test

APPRENTISSAGE

**Pour**  $\forall I_i \in \mathbf{I}_A$  **Faire** $\mathbf{p}_i$  : les primitives géométriques détectées dans  $I_i$  $\mathbf{d}_i$  : les descripteurs calculés à partir de  $\mathbf{p}_i$ **Fin Pour** $\mathbf{m}$  : le dictionnaire de codes obtenu à partir des histogrammes des  $\mathbf{p}_i$  d'après les valeurs des  $\mathbf{d}_i$  $\mathbf{V}$  : le vecteur de caractéristiques obtenu à partir de  $\mathbf{m}$  $\mathbf{C}$  : le classifieur fort obtenu à partir de  $\mathbf{V}$  en utilisant l'algorithme *Adaboost*

CLASSIFICATION

**Pour**  $\forall I_i \in \mathbf{I}_T$  **Faire** $\mathbf{p}_i$  : les primitives géométriques détectées dans  $I_i$  $\mathbf{d}_i$  : les descripteurs calculés à partir de  $\mathbf{p}_i$  $c_i$  : la classe de l'image  $i$  déterminée à partir de  $\mathbf{d}_i$  en utilisant le  $\mathbf{C}$ **Fin Pour****Retourner**  $\mathbf{c}$ 

## 5.2 Classifieur *Adaboost*

Dans leur article séminal, Viola et Jones [2004] ont proposé un système de détection de visages à la fois rapide et efficace, avec un taux de détection élevé. Le système est basé sur une classification supervisée binaire, où on apprend à classifier une image de taille 24x24 pixels en tant que "visage" ou "non-visage". La détection est effectuée par la classification de toutes les images incluses dans l'image à l'aide d'une fenêtre glissante.

Ce qui rend le système rapide est la fusion de plusieurs classifieurs simples. A l'apprentissage, pour chaque image, on construit un vecteur de caractéristiques très large. Trois sortes de caractéristiques simples sont utilisées. Ce sont des motifs qui ressemblent à des fonctions de la base de Haar, en formes rectangulaires divisés en 2, 3 ou 4 régions noires et blanches (Figure 5.1). La valeur d'une caractéristique est la différence entre la somme des pixels blancs et la somme des pixels noirs. Le vecteur de caractéristiques (qui contient 45963 éléments dans leur expérience) se compose de valeurs des trois motifs pour toutes les configurations possibles dans l'image.

Un classifieur de type *Adaboost* (Algorithme 8) est utilisé pour sélectionner un nombre réduit de caractéristiques à partir d'un ensemble d'images positives et d'images négatives.

Les auteurs ont aussi développé une cascade de classifieurs faibles qui est équivalente au classifieur final et qui est plus efficace en temps de calcul. La figure 5.2 montre les caractéristiques

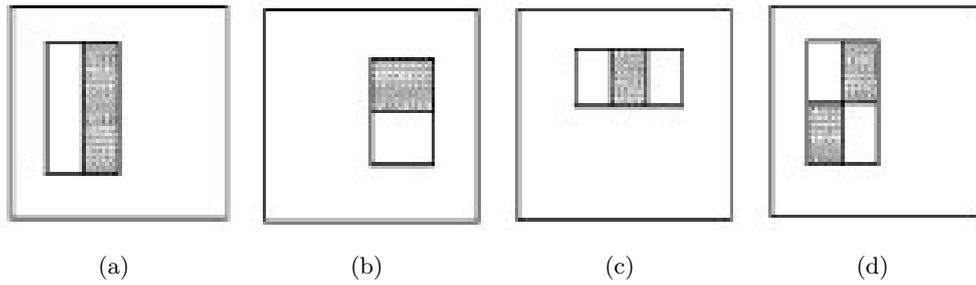


FIG. 5.1 – Les 3 motifs (a-b, c, d) rectangulaires utilisés dans Viola et Jones [2004].

**Algorithm 8** Algorithme de classification *Adaboost***Entrées:** $(x_1, \dots, x_n)$  : les images $(y_1, \dots, y_n)$  : les étiquettes tel que  $y_i = 0, 1$  pour les images négatives et positives**Sorties:** $h(x)$  : le classifieur finalInitialiser les poids,  $w_{1,i} = \frac{1}{n}$ **Pour**  $t = 1$  to  $T$  **Faire**Normaliser les poids :  $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,i}}$ **Pour** chaque caractéristique  $j$  **Faire**Trouver le classifieur faible optimal  $h_{t,j}$ ,Calculer l'erreur du classifieur pondérée par le  $w_{t,i}$ ,  $\epsilon_{t,j} = \sum_i w_{t,i} |h_{t,j}(x_i) - y_i|$ **Fin Pour**Sélectionner le classifieur qui minimise l'erreur :  $h_t = \underset{h_{t,j}}{\operatorname{argmax}} \epsilon_{t,j}$ Mettre à jour les poids :  $w_{t+1,i} = w_{t,i} \beta_t^{1-\epsilon_i}$ , où  $\epsilon_i = 0$  si l'exemple  $x_i$  est classifié correctement,  $\epsilon_i = 1$  sinon, et  $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$ **Fin Pour****Retourner** :

$$h(x) = \begin{cases} 1, & \text{if } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0, & \text{sinon} \end{cases}$$

où  $\alpha_t = \log \frac{1}{\beta_t}$ 

qui correspondent aux deux premiers classifieurs faibles sélectionnés par l'algorithme pour la détection de visages. En utilisant seulement ces deux caractéristiques on obtient un taux de détection de 100 % de visages avec un rejet de 60 % de non-visages.

On a proposé plusieurs méthodes fondées sur la même approche générale, surtout pour améliorer la sélection des classifieurs faibles. Par exemple Abramson *et al.* [2007] utilisent un ensemble (de taille constante) de pixels individuels au lieu de régions rectangulaires pour construire le vecteur de caractéristiques. L'explosion combinatoire est maîtrisée en cherchant le meilleur classifieur faible à l'aide d'un algorithme génétique.

Nous avons choisi d'utiliser l'algorithme *Adaboost* pour la classification des objets cartogra-

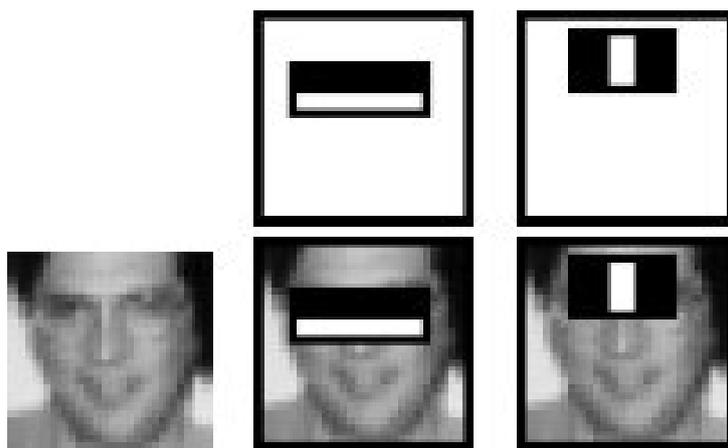


FIG. 5.2 – Les deux premiers classifieurs faibles sélectionnés par l'algorithme *Adaboost* pour détecter les visages, dans Viola et Jones [2004].

phiques avec deux motivations principales :

- Il est possible de distinguer les objets cartographiques par des règles structurelles générales. Par exemple un rond-point peut être décrit par les règles
  - Les objets qui ont une région centrale circulaire ;
  - Les objets qui ont des régions périphériques linéaires orientées vers le centre.
- Les objets d'une même classe ayant une variance intra-classe importante, il serait approprié d'exprimer ces règles en forme d'une combinaison de règles simples qui couvre le plus grand nombre d'exemples et qu'on peut apprendre à partir des primitives de bas niveau.

L'algorithme *Adaboost* nous permet de construire un modèle d'objet implicite qui se constitue d'un ensemble de composants simples pondérés par des poids d'après leurs importances pour la définition de l'objet.

### 5.3 Construction du vecteur de caractéristiques

Notre contribution à la méthode de reconnaissance fondée sur *Adaboost* est principalement au niveau de la construction de classifieurs faibles. Nous proposons d'utiliser les primitives géométriques et leurs organisations spatiales pour construire un vecteur de caractéristiques.

Aux primitives géométriques simples détectées dans le chapitre 4) nous ajoutons un ensemble de primitives composées, obtenues par les principes de regroupement perceptuel ([Iqbal et Aggarwal, 2002]). Ce sont **les jonctions L**, **les jonctions T** et **les parallèles**. Ces primitives révèlent les relations spatiales entre les paires de primitives.

**Les jonctions L** : Pour chaque paire de segments de droites on calcule les distances entre les extrémités des segments. Si le minimum de ces distances est en dessous d'une tolérance

de distance  $d_L$ , on ajoute le point de jonction à la liste des jonctions L avec les attributs des segments qui le composent.

**Les jonctions T** : Pour chaque paire de segments qui n'appartient pas à la liste des jonctions L, on calcule les distances des extrémités du premier segment au deuxième segment et vice versa. Si le minimum de ces distances est en dessous d'une tolérance de distance  $d_T$ , on ajoute le point de jonction à la liste des jonctions T avec les attributs des segments qui le composent.

**Les parallèles** : On détecte les parallèles en utilisant la méthode utilisée dans la section 3.3.2.2.

Les primitives composées détectées dans une image de rond-point et dans une image de pont sont présentées dans la figure 5.3

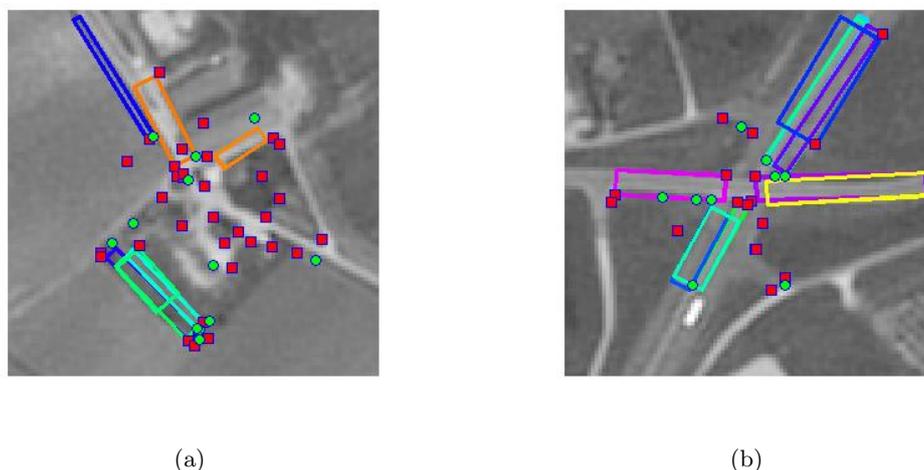


FIG. 5.3 – Les primitives composés (les jonctions L en rouge, les jonctions T en vert). (a). Image d'un rond-point (b). Image d'un pont.

Dans Iqbal et Aggarwal [2002] on a détecté les mêmes types de primitives pour construire un vecteur de caractéristiques composé de trois éléments scalaires calculés par un simple comptage des primitives. Ce vecteur est utilisé pour classifier les images cartographiques dans les trois classes  $\{\mathbf{structure}, \mathbf{non\text{-}structure}, \mathbf{intermédiaire}\}$  à l'aide d'un classifieur de type **les k plus proches voisins**. Le vecteur calculé n'est suffisant que pour une classification dans des catégories très vagues.

Nous proposons d'extraire un vecteur d'une taille beaucoup plus importante qui représente les propriétés géométriques, spatiales et relationnelles des primitives. L'approche adoptée pour construire le vecteur de caractéristiques est inspiré de la construction d'un **dictionnaire de codes visuel** dans Ferrari *et al.* [2008]. Dans ce travail, on détecte dans les images d'apprentissage, les groupes de  $k$  segments adjacents. Nous avons regroupé les primitives géométriques en **mots visuels** qui représentent leurs similarités **géométriques**, **spatiales** et **relationnelles**.

Pour ce faire, on calcule un ensemble d'attributs de primitives, on divise les valeurs des attributs en intervalles réguliers et on compte le nombre de primitives dans chaque intervalle. Cela est équivalent à construire un histogramme pour chaque attribut. Les attributs géométriques correspondent au type et à la forme des primitives, les attributs spatiaux correspondent à leur position et à leur orientation par rapport au centre de l'image et les attributs relationnels correspondent aux propriétés des primitives composées.

Les attributs pour chaque type de primitives sont les suivants :

- $A_{arc} = \{\text{distance au centre, rayon, angle d'ouverture de l'arc}\}$
- $A_{segment} = \{\text{distance au centre, longueur, angle par rapport au centre}\}$
- $A_{blob} = \{\text{distance au centre, longueur de l'axe majeur, compacité}\}$
- $A_{jonctionL} = \{\text{distance au centre, angle entre les segments, longueur minimal, longueur maximal}\}$
- $A_{jonctionT} = \{\text{distance au centre, angle entre les segments, longueur minimal, longueur maximal}\}$
- $A_{parallèle} = \{\text{distance au centre, longueur, angle par rapport au centre}\}$

Les valeurs des attributs sont normalisées et divisées en intervalles réguliers. Le nombre d'intervalles est un variable libre qu'on peut préciser en fonction du compromis entre la taille du vecteur final et la granularité du regroupement. Soient  $m$  le nombre de type de primitives,  $n_i$  le nombre d'attributs pour un type de primitive  $i$  et  $p_{i,j}$  le nombre d'intervalles pour un attribut  $j$  du primitive  $i$ , la taille du vecteur de caractéristiques est égale à :

$$\sum_{i=1}^m \sum_{j=1}^{n_i} p_{i,j}$$

Les mots visuels représentent des concepts sémantiques tels que :

- Le nombre d'arcs de cercles centraux, à une distance de 0 à 5 pixels au centre ;
- Le nombre d'arcs de cercles qui ont un angle d'ouverture de 270 à 360 degrés ;
- Le nombre de segments périphériques, à une distance de 20 à 30 pixels au centre ;
- Le nombre de jonctions qui ont un angle droit, de 70 à 90 degrés ;
- Le nombre de *blobs* compacts, avec une compacité de 0,8 à 1.

Dans nos expériences nous avons créé un vecteur de caractéristiques de taille 96. Un classifieur faible est construit à partir de chaque élément du vecteur, à chaque itération de l'algorithme *Adaboost*, en y associant une valeur de seuil et une parité qui indique la direction de l'inéquation. Par exemple un classifieur faible pourrait être "classifier l'image en tant que positive si le nombre d'arcs de cercles centraux est supérieur à 0,7" (les caractéristiques étant normalisées entre 0 et 1)

## 5.4 Résultats expérimentaux

La méthode est testée sur la base d'images du CNES pour détecter les images de ronds-points et de ponts. Cette base contient **962** images appartenant aux dix catégories d'objets. Parmi ces images, **72** appartiennent à la classe ronds-points et **99** à la classe ponts. On a divisé la base en deux ensembles d'images, un ensemble d'apprentissage et un ensemble de test, qui contiennent chacun approximativement 50% des images dans chaque catégorie. **L'ensemble de test contient 481 images, dont 36 ronds-points et 49 ponts.**

Nous avons construit un classifieur pour les ronds-points ( $\mathcal{C}_{\mathcal{RP}}$ ) et un classifieur pour les ponts ( $\mathcal{C}_{\mathcal{PT}}$ ) en utilisant les primitives obtenues par l'approche contours et les primitives obtenues par l'approche régions conjointement.

Pour l'apprentissage du classifieur, on a utilisé toutes les images d'apprentissage, en considérant les images qui contiennent l'objet cible comme des exemples positifs et les autres images comme des exemples négatifs. On a ensuite calculé un score pour chaque image de la base de test et on a classifié les images en tant que positive ou négative.

Le tableau 5.1 présente les résultats de la classification.

TAB. 5.1 – Les résultats de la classification par la fusion des classifieurs faibles (VP=vrais positifs, FP=faux positifs, FN=faux négatifs, VN=vrais négatifs, P=précision, R=rappel, S=spécificité).

Classe	Seuil	VP	FP	FN	VN	P	R	S	f-mesure
Ronds-points	$s_{Adaboost}$	29	2	7	443	0,9355	0,8056	0,9995	0,8657
Ronds-points	$s_{meilleur}$	33	6	3	439	0,8462	0,9167	0,9865	0,8800
Ponts	$s_{Adaboost}$	18	13	31	419	0,5806	0,3600	0,9698	0,4444
Ponts	$s_{meilleur}$	27	30	22	402	0,4737	0,5400	0,9304	0,5047

Nous avons mesuré les valeurs suivantes pour l'évaluation des résultats :

- La précision (**P**) est le pourcentage des réponses correctes dans les images classifiées en tant que positive. Elle est calculée par la formule :  $P = VP/(VP + FP)$  ;
- Le rappel (**R**) ou la sensibilité mesure la capacité du classifieur de classifier une image appartenant à la classe d'objet recherchée en tant que positive. Elle est calculée par la formule :  $R = VP/(VP + FN)$  ;
- La spécificité (**S**) mesure la capacité du classifieur de classifier une image qui n'appartient pas à la classe d'objet recherchée en tant que négative. Elle est calculée par la formule  $S = VN/(VN + FP)$  ;
- La **f-mesure** est la moyenne harmonique pondérée de la précision et du rappel :

$$f\text{-mesure} = \frac{(\beta^2 + 1) \times P \times R}{P + R}$$

Nous supposons que  $P$  et  $R$  ont le même poids. On obtient alors  $\beta = 1$  et  $F = 2 \times P \times R /$

$(P + R)$ . La *f-mesure* est surtout appropriée pour mesurer la performance d'un classifieur en cas d'exemples positifs et négatifs déséquilibrés, comme c'est le cas pour notre base de test.

La première ligne du tableau pour chaque classifieur présente les valeurs obtenues d'après le seuil de décision retourné par le classifieur ( $s_{Adaboost}$ ). Nous avons également calculé (dans la deuxième ligne) les valeurs correspondant au meilleur seuil de décision ( $s_{meilleur}$ ) qui maximise la *f-mesure*.

La courbe *ROC* (Receiver Operating Characteristic) est un très bon outil pour visualiser la performance d'un classifieur binaire et elle est largement utilisée dans le domaine d'apprentissage par machines. C'est le graphique du taux des vrais positifs par rapport au taux de faux positifs d'un classifieur binaire pour les différentes valeurs du seuil de décision.

L'aire sous la courbe *ROC* indique la performance d'un classifieur. Elle peut être interprétée comme la probabilité d'assigner un score plus élevé à un exemple positif quand on choisit aléatoirement un exemple positif et un exemple négatif.

La figure 5.4 présente les courbes *ROC* obtenues pour les deux classifieurs. Les valeurs de l'aire sous la courbe *ROC* sont 0,9815 et 0,7853 respectivement.

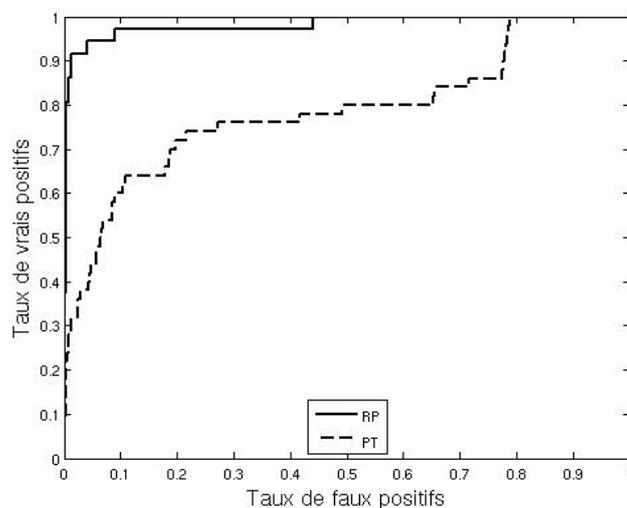


FIG. 5.4 – Les courbes *ROC* de la classification par le  $\mathcal{C}_{RP}$  et le  $\mathcal{C}_{PT}$ .

Les tableaux 5.2 et 5.3 montrent le nombre d'objets de chaque catégorie dans les  $n$  premières images où  $n$  correspond au nombre d'exemples positifs dans l'ensemble de test. Ces tableaux révèlent une des raisons principales pour les mauvaises classifications des ponts. La similarité structurelle des ponts avec les autoroutes et les carrefours routiers étant très élevée la méthode qui est fondée sur les caractéristiques géométriques des objets a un taux de confusion important pour ces trois catégories.

On a obtenu de très bons résultats de classification pour les ronds-points. Pourtant les résul-

TAB. 5.2 – Le nombre d’objets de chaque catégorie dans les 36 premières images obtenues par le  $\mathcal{C}_{\mathcal{RP}}$ .

Catégorie d’objet	Rond-point	Bâti isolé	Lotissement	Carrefour routier
Nombre d’objet	31	2	2	1

TAB. 5.3 – Le nombre d’objets de chaque catégorie dans les 49 premières images obtenues par le  $\mathcal{C}_{\mathcal{PT}}$ .

Catégorie d’objet	Pont	Autoroute	Carrefour	V. ferrée	R. moyenne	B. isolé
Nombre d’objet	24	10	7	4	3	1

tats de classification pour les ponts sont moins bons. Pour comprendre les causes de mauvaises classifications nous avons effectué deux autres analyses.

Nous avons d’abord testé le classifieur pour les ponts sur **les images d’apprentissage**. Ce test a donné un taux de vrais positifs de 100% et un taux de vrais négatifs de 95,79%. Ce résultat signale un problème de sur-entraînement (*overtraining*). Le classifieur, au lieu de généraliser, s’attache aux exemples et retourne un hyperplan trop complexe qui sépare presque tous les exemples positifs des exemples négatifs.

Dans un deuxième temps nous avons analysé visuellement les exemples mal-classifiés. La figure 5.5 montre les 4 images de ronds-points et les 8 images de ponts qui ont obtenu les scores les plus bas. On voit que pour le  $\mathcal{C}_{\mathcal{RP}}$  les erreurs sont principalement dues à la taille très petite des ronds-points. Pour le  $\mathcal{C}_{\mathcal{PT}}$  les 9 ponts qui ont eu les scores les plus bas sont tous des ponts sur des voies fluviales. Ces résultats peuvent être expliqués par le manque d’indices géométriques et structurels pour reconnaître ces objets.

Les mêmes types d’images de ponts sur les voies fluviales existent également dans la base d’apprentissage. Pendant la classification d’images d’apprentissage, nous avons observé que la plupart de ces images sont bien classifiées. Cela justifie notre hypothèse de manque de généralisation. Ces objets étant trop variés (surtout en termes de caractéristiques structurelles), un modèle qui classe bien une partie d’eux n’est pas suffisamment général pour reconnaître les autres instances.

## 5.5 Conclusions

Dans ce chapitre nous avons présenté une méthode de classification pour reconnaître les images de ronds-points et de ponts. Cette méthode est fondée sur la construction d’un modèle implicite d’objet composé de classifieurs faibles fusionnés par l’algorithme *Adaboost*.

Nous avons construit les classifieurs faibles à partir de primitives géométriques. L’apprentissage du classifieur correspond effectivement à la sélection des primitives qui ont les propriétés

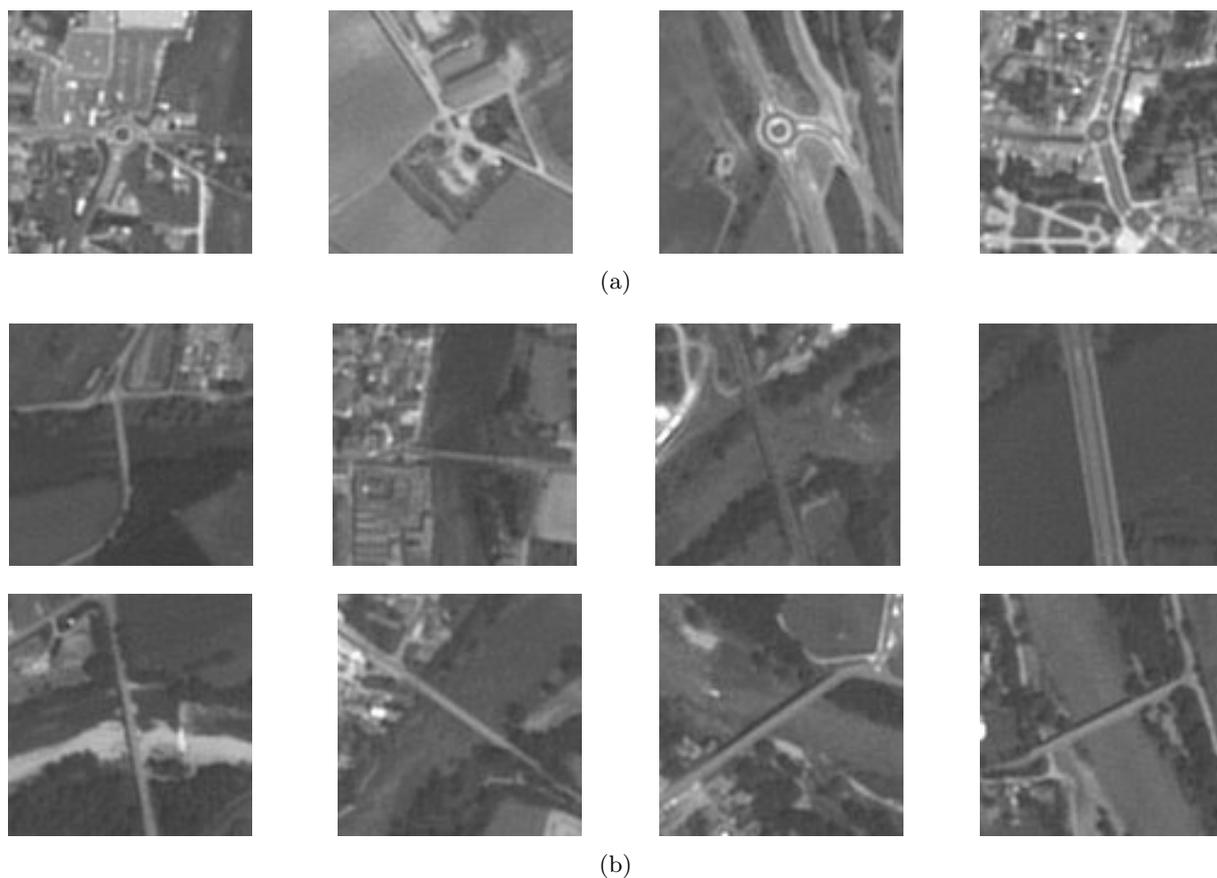


FIG. 5.5 – Les 4 images de ronds-points et les 8 images de ponts qui ont obtenu les scores les plus bas pendant la classification. (a). Les ronds-points, (b). Les ponts.

géométriques et les organisations spatiales les plus significatives pour représenter l'objet. Nous pensons que la méthode proposée est une adaptation remarquable d'une méthode de reconnaissance d'objets actuelle sur notre problème.

La méthode a eu des difficultés à la classification des ponts. Une analyse détaillée des résultats nous a permis de clarifier les causes des mauvaises classifications, surtout pour les ponts.

Nous avons observé que les ponts sont confondus avec un taux important de confusion avec les objets qui ont des structures similaires, comme les autoroutes et les carrefours. Cette confusion est surtout due au manque de représentation des relations spatiales binaires entre les composants d'objets par le modèle implicite.

Un autre problème est observé à la classification des ponts qui passent sur des voies fluviales. Ce résultat provient principalement de la forme irrégulière des régions d'eaux. L'existence de ces objets très différents des autres objets dans la même catégorie cause un sur-entraînement pendant l'apprentissage.

Nous proposons dans le chapitre suivant une autre méthode de reconnaissance qui utilise les mêmes types de descripteurs structurels mais qui construit un dictionnaire visuel plus général

avec moins d'éléments. Cette méthode apprend le dictionnaire visuel sans qu'on ait besoin de le construire par des histogrammes d'attributs de primitives.



## Chapitre 6

# Le modèle “composants et structure”

### 6.1 Introduction

Dans les dernières années on a obtenu des avancées considérables en détection d’objets génériques dans les images par l’utilisation des méthodes fondées sur les composants et la structure. L’idée principale de ces méthodes est de considérer l’objet comme une collection de ses composants (souvent représentés par des régions locales significatives de l’image détectées à partir des propriétés radiométriques) et de représenter la structure de l’objet en modélisant les positions des composants relatives à un point de repère. La reconnaissance est vue comme un problème d’apprentissage où on apprend un modèle d’objet qui est constitué de deux éléments principaux, **les composants** et **la structure**, à partir d’un ensemble d’images d’apprentissage. On reconnaît un objet par la détection des composants de l’objet dans l’image, l’appariement de ces composants avec les composants du modèle et la recherche de la configuration des composants appariés qui correspond au modèle d’objet.

Les approches fondées sur les composants sont bien appropriées pour la détection d’objets cartographiques pour deux raisons principales :

- Les objets cartographiques sont des objets articulés. Par exemple un rond-point est composé d’un rond central et des routes articulées autour du rond.
- La variation intra-classe est très importante et cette variation est souvent due aux transformations affines des composants de l’objet.

Nous présentons dans ce chapitre une méthode de reconnaissance d’objets cartographiques **ronds-points** et **ponts** par un modèle fondé sur les composants et la structure. L’originalité de notre méthode par rapport aux autres est l’utilisation des primitives géométriques pour la détection des composants afin de saisir la similitude structurelle des objets, et l’apprentissage de la structure en même temps que les composants de l’objet.

## 6.2 Le schéma général

Nous nous sommes particulièrement intéressés à deux méthodes récentes de reconnaissance d’objets qui utilisent l’approche fondée sur les composants et la structure.

Dans ces deux méthodes fondamentalement similaires [Leibe *et al.*, 2004; Agarwal et Roth, 2002], on vise à reconnaître les voitures vues de coté. On détecte des régions significatives autour des points d’intérêts et on regroupe les régions les plus similaires dans des *clusters* pour construire un dictionnaire de codes qui représente les composants de l’objet. On encode ensuite les positions ou les relations spatiales des composants pour créer un modèle d’objet utilisé pour la reconnaissance. Dans [Leibe *et al.*, 2004] on apprend le modèle à partir d’images positives seulement. Agarwal et Roth [2002] utilisent les images positives ainsi que les images négatives pour l’apprentissage.

Dans les deux approches, les objets cibles ont des formes assez régulières avec une variation intra-classe faible. L’orientation des objets est toujours la même, les variations d’échelle et de taille sont très limitées et les positions relatives des composants sont similaires pour des objets différents dans une même classe. Les objets cartographiques ne satisfont pas en général ces critères désirables.

Nous avons adopté une approche radicalement différente pour détecter les composants de l’objet. On extrait à partir de l’image des primitives géométriques qui correspondent aux composants de l’objet. Ces primitives ont l’avantage d’avoir une forme bien définie qui révèle la structure géométrique inhérente des objets cartographiques. Pourtant comme elles sont très générales, leur contenu informatif est faible, et par conséquent l’appariement de deux primitives entre deux images n’indique pas nécessairement que les deux images contiennent le même composant. A cause de cela il n’est pas possible d’appliquer directement une approche qui apprend d’abord les composants et puis la structure.

Nous proposons d’apprendre les composants et la structure simultanément en associant aux primitives des attributs qui représentent leurs propriétés géométriques et spatiales, et en sélectionnant, à l’aide des attributs associées, les primitives les plus pertinentes pour représenter un objet.

A partir d’un ensemble d’images d’apprentissage contenant l’objet cible, nous construisons un dictionnaire de codes en groupant les primitives d’après la similarité de leur position, leur orientation et leurs propriétés géométriques. On sélectionne les meilleurs *clusters* automatiquement.

Une image est classifiée en appariant les primitives extraites de l’image avec les *clusters* sélectionnés et en calculant un score d’appartenance à la classe. Nous avons donné une définition formelle de notre méthode dans l’Algorithme 9. Dans les sections suivantes nous présentons les étapes de la méthode en détail.

---

**Algorithm 9** Algorithme global de la méthode fondée sur les composants et la structure
 

---

**Entrées:** $\mathbf{I}_A$  : Ensemble d'images d'apprentissage positives $\mathbf{I}_T$  : Ensemble d'images de test**Sorties:** $\mathbf{s}$  : les scores des images de test

APPRENTISSAGE

**Pour**  $\forall I_i \in \mathbf{I}_A$  **Faire** $\mathbf{p}_i$  : les primitives géométriques détectées dans  $I_i$  $\mathbf{d}_i$  : les descripteurs calculés à partir de  $\mathbf{p}_i$ **Fin Pour** $\mathbf{c}$  : les *clusters* obtenus par le regroupement des  $\mathbf{d}$  $\mathbf{c}'$  : les *clusters* sélectionnés

CLASSIFICATION

**Pour**  $\forall I_i \in \mathbf{I}_T$  **Faire** $\mathbf{p}_i$  : les primitives géométriques détectées dans  $I_i$  $\mathbf{d}_i$  : les descripteurs calculés à partir de  $\mathbf{p}_i$ **Pour**  $\forall d_{ij} \in \mathbf{d}_i$  **Faire** $s_{D_{ij}}$  : le score du descripteur calculé à partir de  $\mathbf{c}'$ **Fin Pour** $s_i : \sum s_{D_{ij}}$ **Fin Pour****Retourner**  $\mathbf{s}$ 


---

### 6.3 Représentation des primitives

Nous utilisons deux types de primitives géométriques, les arcs de cercles et les segments. Après l'extraction des primitives, on calcule pour chaque primitive un ensemble de caractéristiques qui représentent ses propriétés géométriques et spatiales.

Comme les objets cartographiques sont tous centrés sur les imageries, les caractéristiques de position et d'orientation sont calculés relativement au centre de l'image. Cette représentation autour du centre de l'image assure l'invariance des primitives aux rotations de l'image.

On représente les primitives par les caractéristiques suivantes :

- $f_{Cercle} = \{d_C, r_C\}$ ,  $d_C$  = la distance entre le centre du cercle et le centre de l'image,  $r_C$  = le rayon du cercle
- $f_{Segment} = \{d_S, \theta_S, l_S\}$ ,  $d_S$  = la distance entre le centre du segment et le centre de l'image,  $\theta_S$  = l'angle entre le segment et la droite qui joint le centre à l'extrémité du segment le plus loin du centre,  $l_S$  = la longueur du segment.

Figure 6.1 illustre les caractéristiques d'un arc de cercle et les caractéristiques d'un segment de droite.

Nous avons utilisé les primitives obtenues par l'approche contours et par l'approche régions

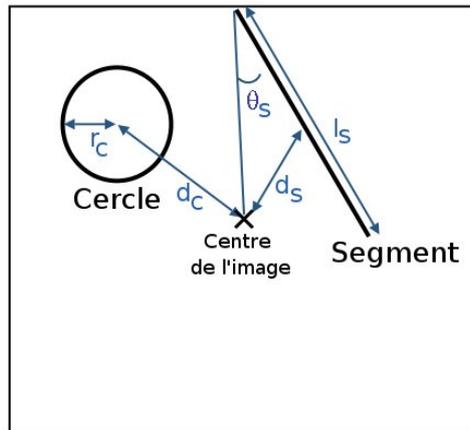


FIG. 6.1 – Le vecteur de caractéristiques.

conjointement. On construit un vecteur de caractéristiques ( $V_C$ ) de taille  $n_C \times 2$  pour les arcs de cercles et un vecteur de caractéristiques ( $V_S$ ) de taille  $n_S \times 3$  pour les segments de droite, en regroupant les primitives de même type sur les images d'apprentissages appartenant à la classe d'objet cible.  $n_C$  et  $n_S$  sont le nombre total d'arcs de cercles et de segments respectivement. On normalise les vecteurs en divisant la valeur de chaque caractéristique par la valeur maximale dans la même colonne.

L'étape suivante est la sélection des primitives les plus significatives pour la détection d'objet, en utilisant les vecteurs de caractéristiques obtenus.

## 6.4 Clustering

On peut représenter les vecteurs de caractéristiques comme un nuage de points bi-dimensionnel pour les cercles et un nuage de points tri-dimensionnel pour les segments. Nous construisons un dictionnaire de codes en regroupant les points les plus proches. La méthode *Mean-Shift* est bien appropriée pour le *clustering* parce que c'est une méthode non-paramétrique qui ne nécessite pas la connaissance *a priori* du nombre de *clusters* ni la forme de la distribution, et qui est rapide pour les données de petites dimensions.

**Mean Shift Clustering** *Mean-shift* (Cheng [1995]) est un algorithme de *clustering* des données par l'estimation de leur densité. On considère les points dans un espace de dimension  $d$  comme une fonction de densité de probabilité et on recherche les régions denses qui correspondent aux maxima locaux de la distribution.

L'algorithme *Mean-Shift* a deux étapes : l'estimation du gradient de la fonction de densité et l'utilisation des résultats pour former les *clusters*. Sur chaque point de donnée on applique une procédure d'ascension de gradient sur la densité locale estimée, jusqu'à la convergence. Les points stationnaires représentent les modes de la distribution. L'ensemble des points qui

convergent vers le même mode sont regroupés dans un même *cluster*. La figure 6.2 illustre les itérations de l'algorithme à partir de deux points différents jusqu'à la convergence.

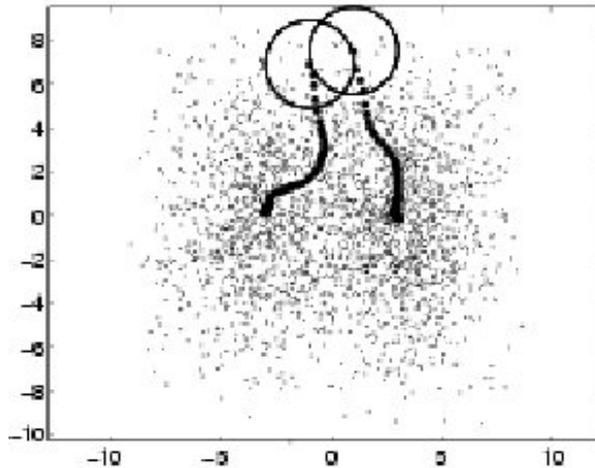


FIG. 6.2 – Illustration du *clustering Mean-Shift* sur les données bi-dimensionnelles. La convergence de deux points initiaux vers les maxima de la densité.

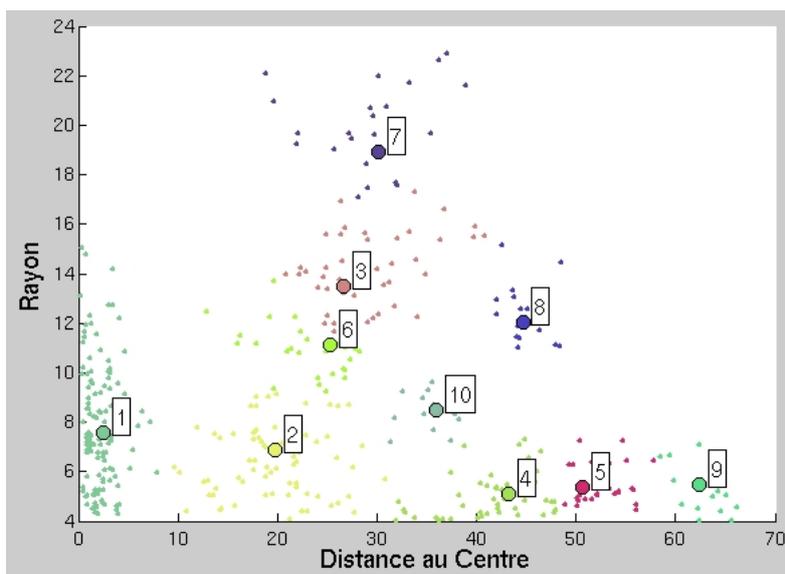
Nous avons appliqué l'algorithme *Mean-Shift* sur les vecteurs  $V_C$  et  $V_S$  obtenus à partir des imagettes de ronds-points et des imagettes de ponts. L'algorithme a un seul paramètre libre  $t$ , la taille de la fenêtre dans laquelle on estime la densité. Nous avons utilisé  $t = 0,1$  (sur des valeurs normalisées entre 0 et 1) dans nos expériences. Les figures 6.3 et 6.4 montrent les 10 *clusters* qui contiennent le plus grand nombre de points, et le nombre de points dans ces *clusters*, obtenus à partir des images d'apprentissages appartenant à la classe ronds-points et à la classe ponts.

## 6.5 La sélection des *clusters*

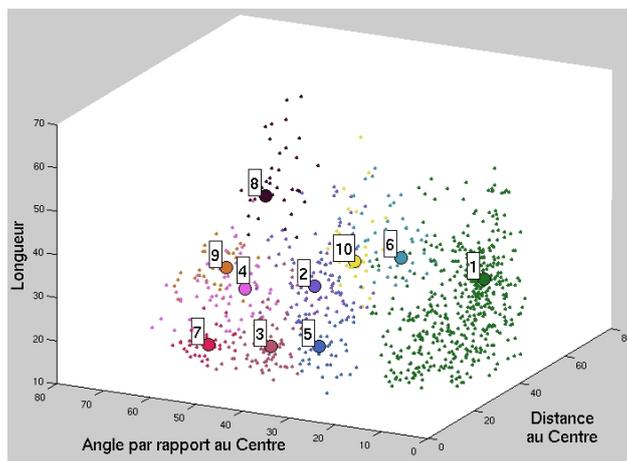
Pour sélectionner les *clusters* les plus significatifs nous avons utilisé une approche heuristique basée sur le nombre de points dans un *cluster* normalisé par le nombre d'images d'apprentissage.

Soit  $n$  le nombre d'images d'apprentissage et  $p_i$  le nombre de points dans un *cluster*  $i$ . On calcule  $\bar{n}_i = p_i/n$ , la moyenne du nombre de points par image dans le *cluster*  $i$ . On accepte les *clusters*  $i$  dont  $\bar{n}_i > d$ , où  $d$  est le seuil de densité. Nous avons choisi  $d = 2$ , ce qui veut dire qu'on observe la primitive au moins une fois en moyenne dans chaque image (la valeur de  $d$  est doublée parce qu'on utilise deux ensembles de primitives obtenus par une approche contours et une approche régions). Nous avons testé l'influence des différentes valeurs de  $d$  sur les résultats et nous avons observé que les résultats sont stables dans l'intervalle  $[1, 5; 2, 5]$ .

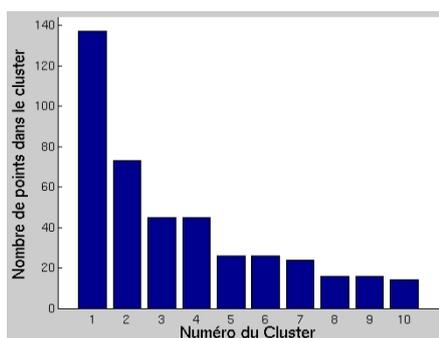
On a sélectionné le premier *cluster* de cercles et le premier *cluster* de segments pour les imagettes de ronds-points et les quatre premiers *clusters* de segments pour les imagettes de ponts. Le tableau 6.1 montre les *clusters* sélectionnés, leur centre de gravité et leur écart-type. Le centre de gravité d'un *cluster* correspond aux moyennes des valeurs de chaque caractéristique



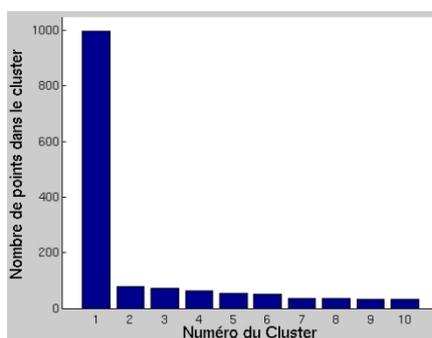
(a)



(b)

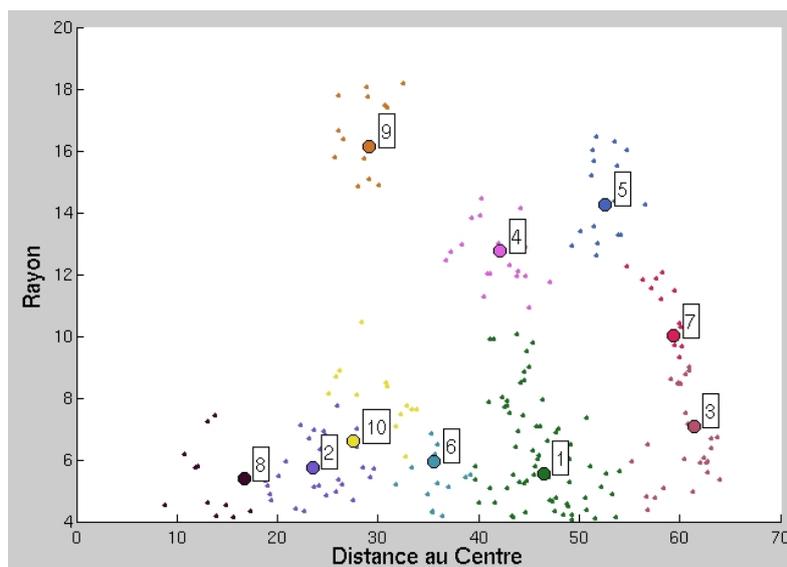


(c)

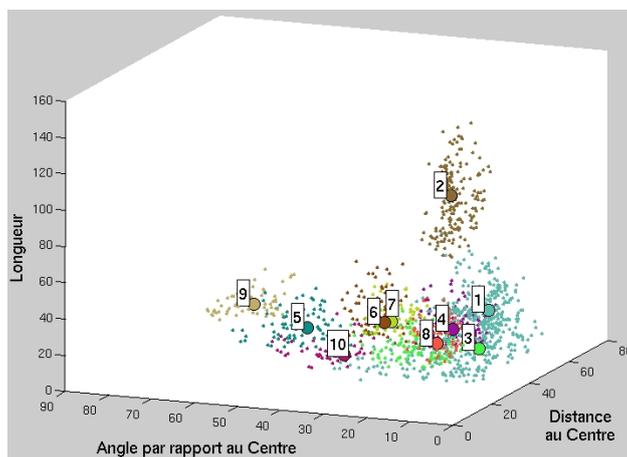


(d)

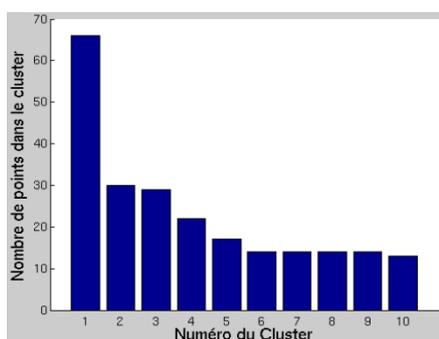
FIG. 6.3 – Le *clustering* des primitives pour les ronds-points. (a). La distribution des données et les *clusters* obtenus pour les primitives arcs de cercles, (b). pour les primitives segments de droites. (c). Le nombre de points dans les *clusters* obtenus pour les primitives arcs de cercles, (d). pour les primitives segments de droites.



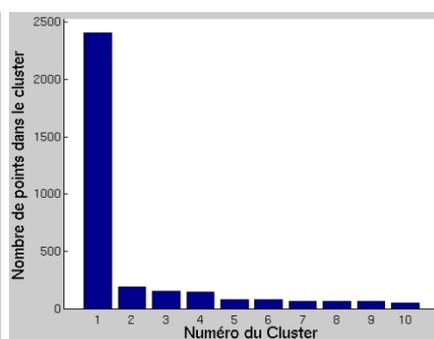
(a)



(b)



(c)



(d)

FIG. 6.4 – Le *clustering* des primitives pour les ponts. (a). La distribution des données et les *clusters* obtenus pour les primitives arcs de cercles, (b). pour les primitives segments de droites. (c). Le nombre de points dans les *clusters* obtenus pour les primitives arcs de cercles, (d). pour les primitives segments de droites.

pour tous les points dans le *cluster*. Pour les cercles, les deux caractéristiques  $c_1$  et  $c_2$  sont la distance au centre et le rayon. Pour les segments, les trois caractéristiques  $c_1$  et  $c_2$  et  $c_3$  sont la distance au centre, l’angle par rapport au centre et la longueur.

TAB. 6.1 – Les moyennes (les centres et les écart-types des caractéristiques des *clusters* sélectionnés).

Classe	Primitif	Cluster	Moyennes			Ecart-types		
			$\mu_1$	$\mu_2$	$\mu_3$	$\sigma_1$	$\sigma_2$	$\sigma_3$
RP	Cercle	1	2.3620	7.7988		1.5023	2.3373	
	Segment	1	33.7203	8.6485	30.4808	8.3736	6.8744	10.5808
PT	Segment	1	29.7787	7.4545	39.0685	11.3769	6.1487	17.5508
	Segment	2	8.2442	3.2656	119.5884	5.1273	2.6033	16.5983
	Segment	3	25.6388	20.0696	26.0300	3.3961	5.1335	8.1287
	Segment	4	46.3790	21.0096	24.1159	6.1384	4.1542	6.9641

On peut interpréter le tableau ci-dessus en considérant la distribution des primitives comme une distribution normale. Statistiquement 68% des valeurs tirées au sort d’une distribution normale sont à une distance de  $1 \sigma$  de la moyenne  $\mu$ . En supposant qu’ils font partie de cet intervalle de confiance, on peut exprimer les ronds-points de la façon suivante :

Un rond-point est construit par :

- des cercles qui ont une distance au centre de 0.8597 à 3.8643 pixels (de 2.1493 mètres à 9.6607 mètres) et un rayon de 5.4615 à 10.1361 pixels (de 13.6536 mètres à 25.3402 mètres), et
- des segments qui ont une distance au centre de 25.3467 à 42.0939 pixels (de 63.3669 mètres à 105.2348 mètres), un angle par rapport au centre de 1.7741 à 15.5229 degrés et une longueur de 19.9000 à 41.0617 pixels (de 49.7500 mètres à 102.6541 mètres).

## 6.6 La classification

Nous calculons  $p(k|p)$ , la probabilité d’appartenance d’un point  $p$  à un *cluster*  $k$ , par la distance de Mahalanobis du point  $p$  au *cluster*  $k$ . Soient  $\mathbf{p} = (p_1, p_2, \dots, p_n)$  le vecteur de dimension  $n$  qui représente le point  $p$ ,  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$  les moyennes de  $k$ , et  $\boldsymbol{\Sigma}$  la matrice de covariance de  $k$  définie par :

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1n}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_n^2 \end{pmatrix}$$

la distance de Mahalanobis du point  $p$  au *cluster*  $k$  est définie par la formule

$$d_M(\mathbf{p}) = \sqrt{(\mathbf{p} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{p} - \boldsymbol{\mu})}$$

Contrairement à la distance Euclidienne, la distance de Mahalanobis tient compte des covariances des variables. On normalise la distance dans chaque dimension en fonction de la variance des données sur cette dimension. Pour mieux comprendre la différence on peut comparer la représentation visuelle des deux distances : quand on utilise la distance Euclidienne, l'ensemble des points équidistants d'un point donné forme un sphère. Par contre la distance de Mahalanobis étire ce sphère en fonction des covariances entre les variables et on obtient une ellipsoïde dont les axes majeurs correspondent aux vecteurs propres de  $\boldsymbol{\Sigma}$ .

Pour chaque primitive  $p_i$  d'une image  $I$  on calcule un score d'appartenance  $s_i$  au modèle d'objet. Le score  $s_i$  est le maximum des distances de Mahalanobis de  $p_i$  à chaque *cluster* sélectionné. Pour calculer le score total de l'image  $I$  nous avons utilisé une fonction analogue à la fonction d'activation utilisée dans les réseaux de neurones artificiels pour calculer la valeur de sortie d'un neurone.

La fonction sigmoïde (Figure 6.5.(a)) est définie par la formule

$$f(x) = \frac{1}{1 + e^{-\lambda x}}$$

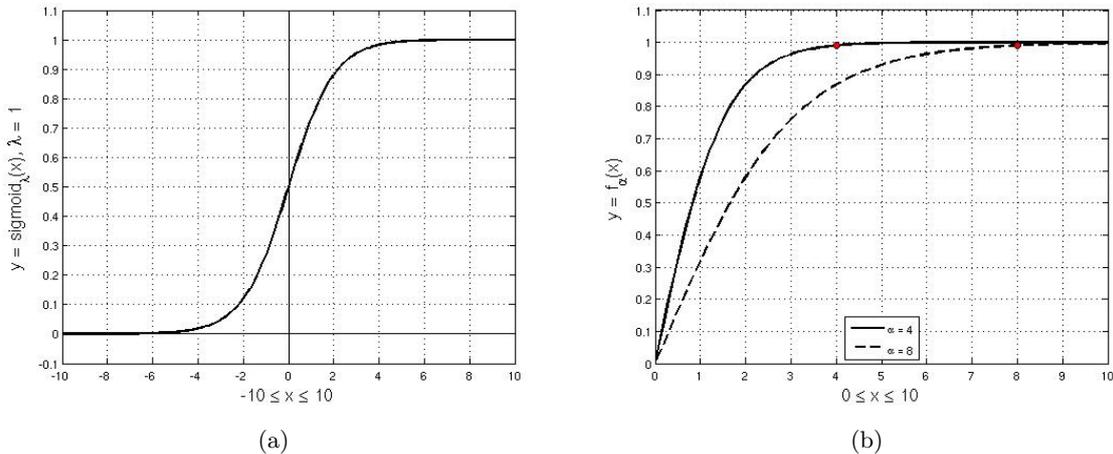


FIG. 6.5 – La fonction sigmoïde et la fonction d'activation.

A partir de la fonction sigmoïde nous avons défini une fonction d'activation  $f_\alpha(x)$  (Figure 6.5.(b)) qui prend la valeur 0 pour  $x = 0$  et la valeur 0.9 pour  $x = \alpha$ . On calcule  $f_\alpha(s_i)$  pour chaque primitive. Cette fonction effectue un seuillage souple, où les primitives qui ont un score  $\alpha$  obtiennent un score d'activation élevé. La valeur de  $\alpha$  est calculée automatiquement pour chaque *cluster* sélectionné en fonction de la moyenne du nombre de primitives dans un *cluster*. Le score total d'une image est la somme des scores d'activation de toutes les primitives détectées dans

cette image.

## 6.7 Résultats expérimentaux

Nous avons testé notre méthode sur la même base d’images (Section 5.4) partitionnée en un ensemble d’apprentissage et un ensemble de test de la même manière.

On a construit deux classifieurs  $\mathcal{C}_{\mathcal{RP}}$  et  $\mathcal{C}_{\mathcal{PT}}$ , un pour les ronds-points et un pour les ponts, en utilisant la base d’apprentissage. Pour chaque image dans la base de test, on a calculé le score d’appartenance à la classe ronds-points ( $s_{RP}$ ) et à la classe ponts ( $s_{PT}$ ).

Nous avons obtenu de très bons résultats de classification, comparables aux résultats obtenus par la méthode précédente pour les ronds-points et beaucoup plus élevés pour les ponts. Nous présentons d’abord ces résultats en détail et ensuite une analyse des principales causes des mauvaises classifications pour les exemples positifs qui ont eu des scores faibles.

Le tableau 6.2 présente les résultats de la classification pour le seuil de décision optimal qui correspond à la valeur de *f-mesure* la plus élevée.

TAB. 6.2 – Les résultats de la classification par l’approche composants et structure (VP=vrais positifs, FP=faux positifs, FN=faux négatifs, VN=vrais négatifs, P=précision, R=rappel, S=sensibilité).

Classe	Seuil	VP	FP	FN	VN	P	R	S	<i>f-mesure</i>
Ronds-points	$s_{Adaboost}$	30	3	6	442	0,9091	0,8333	0,9933	0,8696
Ponts	$s_{Adaboost}$	31	18	18	414	0,6327	0,6327	0,9583	0,6327

La *f-mesure* pour les ronds-points est assez proche à celui obtenu par la méthode précédente. En revanche la *f-mesure* pour les ponts est 25% plus élevé. On a obtenu une nette amélioration à la classification des ronds points.

Nous avons trié les scores par ordre décroissant pour obtenir un classement d’appartenances d’objets à la catégorie recherchée. Les tableaux 6.3 et 6.4 montrent les nombres d’exemples positifs dans les  $n$  premières images pour différentes valeurs de  $n$ .

TAB. 6.3 – Le nombre d’exemples positifs dans les  $n$  premières images d’après les scores triés obtenues par le  $\mathcal{C}_{\mathcal{RP}}$ .

Nombre d’images	25	29	33	37
Nombre de ronds-points	25	28	30	31

Dans les 49 images qui ont obtenu les meilleurs scores par le  $\mathcal{C}_{\mathcal{PT}}$ , on a 31 ponts au lieu de 24 ponts obtenus à la méthode précédente.

TAB. 6.4 – Le nombre d'exemples positifs dans les  $n$  premières images d'après les scores triés obtenues par le  $\mathcal{C}_{PT}$ .

Nombre d'images	12	29	38	49
Nombre de ponts	12	21	25	31

La figure 6.6 montre les courbes *ROC* correspondant à  $\mathcal{C}_{RP}$  et à  $\mathcal{C}_{PT}$ . L'aire sous la courbe *ROC* a les valeurs **0,9699** et **0,8432** respectivement.

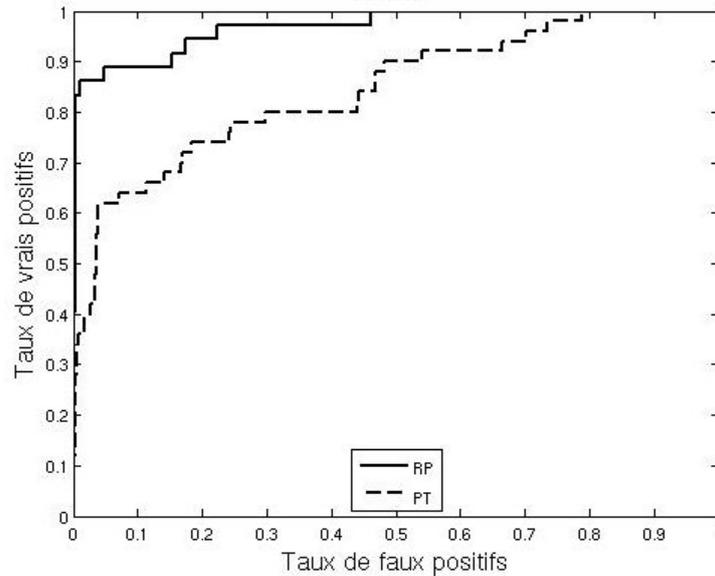


FIG. 6.6 – Les courbes *ROC* de la classification par le  $\mathcal{C}_{RP}$  et le  $\mathcal{C}_{PT}$ .

Le modèle appris nous permet de visualiser les résultats de la détection. Les figures 6.7 et 6.8 présentent toutes les primitives détectées et les primitives sélectionnées, sur les 5 premières et les 5 dernières images de ronds-points, d'après les scores  $s_{RP}$ . Les figures 6.9 et 6.10 présentent les images correspondantes pour les ponts, d'après les scores  $s_{PT}$ .

Sur les images qui ont eu les meilleurs scores on a détecté tous les composants avec très peu de fausses détections. Les ronds-points qui ont eu les scores les plus bas sont ceux qui ont la plus petite taille (sauf un qui est de très grande taille). Sur ces images il est difficile même pour l'œil humain de détecter le rond central. Par conséquent on n'a pas pu détecter les cercles centraux et la détection seule des routes qui s'articulent au rond central n'a pas suffi à obtenir un bon score. Avec l'augmentation de la résolution des images on aurait certainement des meilleurs résultats.

On observe un résultat qui confirme les résultats obtenus au chapitre précédent pour les images de ponts qui ont eu les scores les plus bas : ce sont tous des images de ponts sur des voies fluviales. Le modèle nous permet de voir les principales raisons de mauvaise classification clairement (Figure 6.10) : sur ces images le nombre de primitives géométriques sélectionnées est très insuffisant pour les classifier correctement.

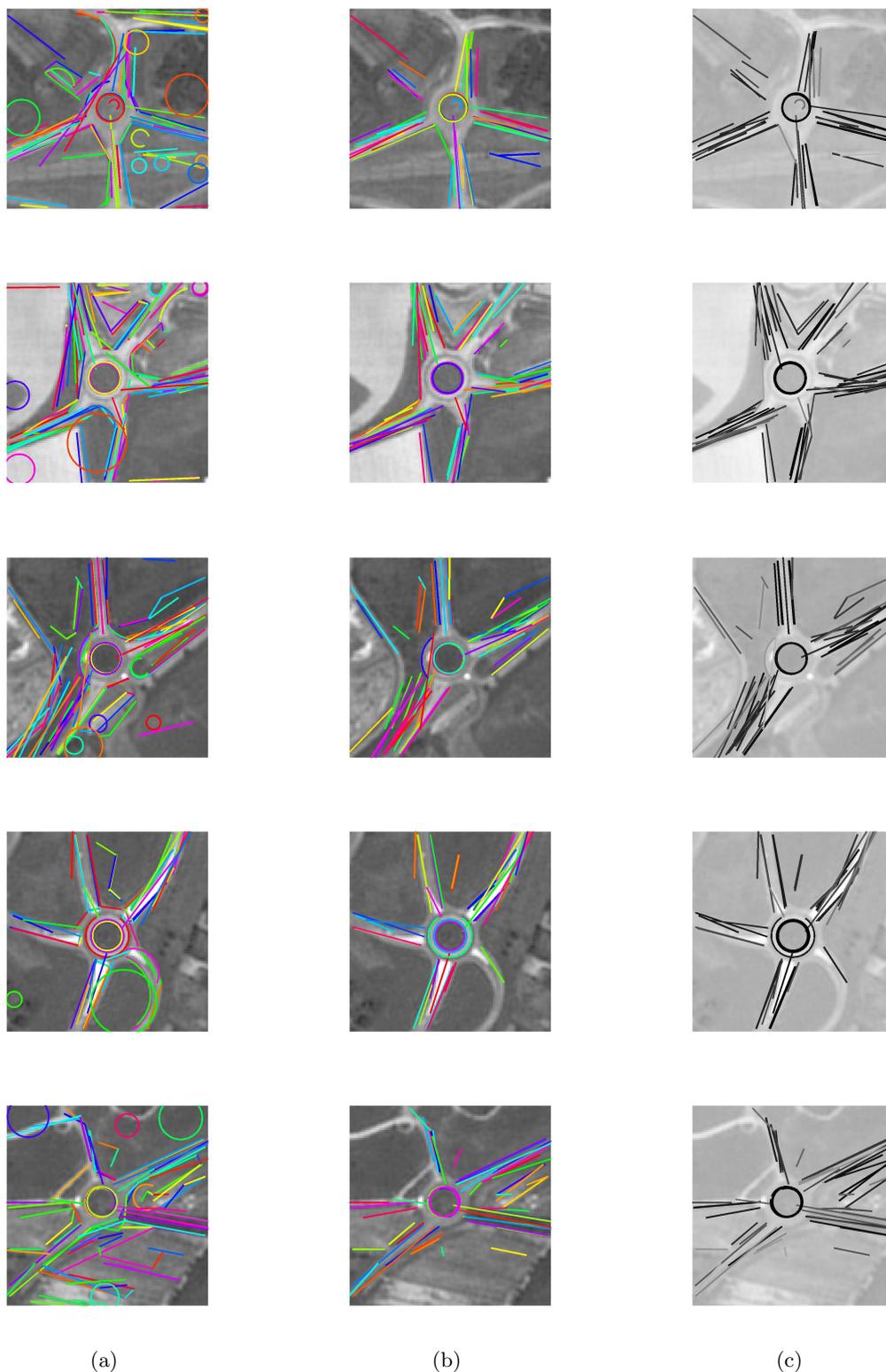


FIG. 6.7 – Les 5 ronds-points avec les scores les plus élevés d’après le score  $s_{RP}$ . (a). Toutes les primitives détectées, (b). Les primitives sélectionnées, (c). Les primitives sélectionnées, représentées par un niveau de gris correspondant à leurs scores d’appartenance.

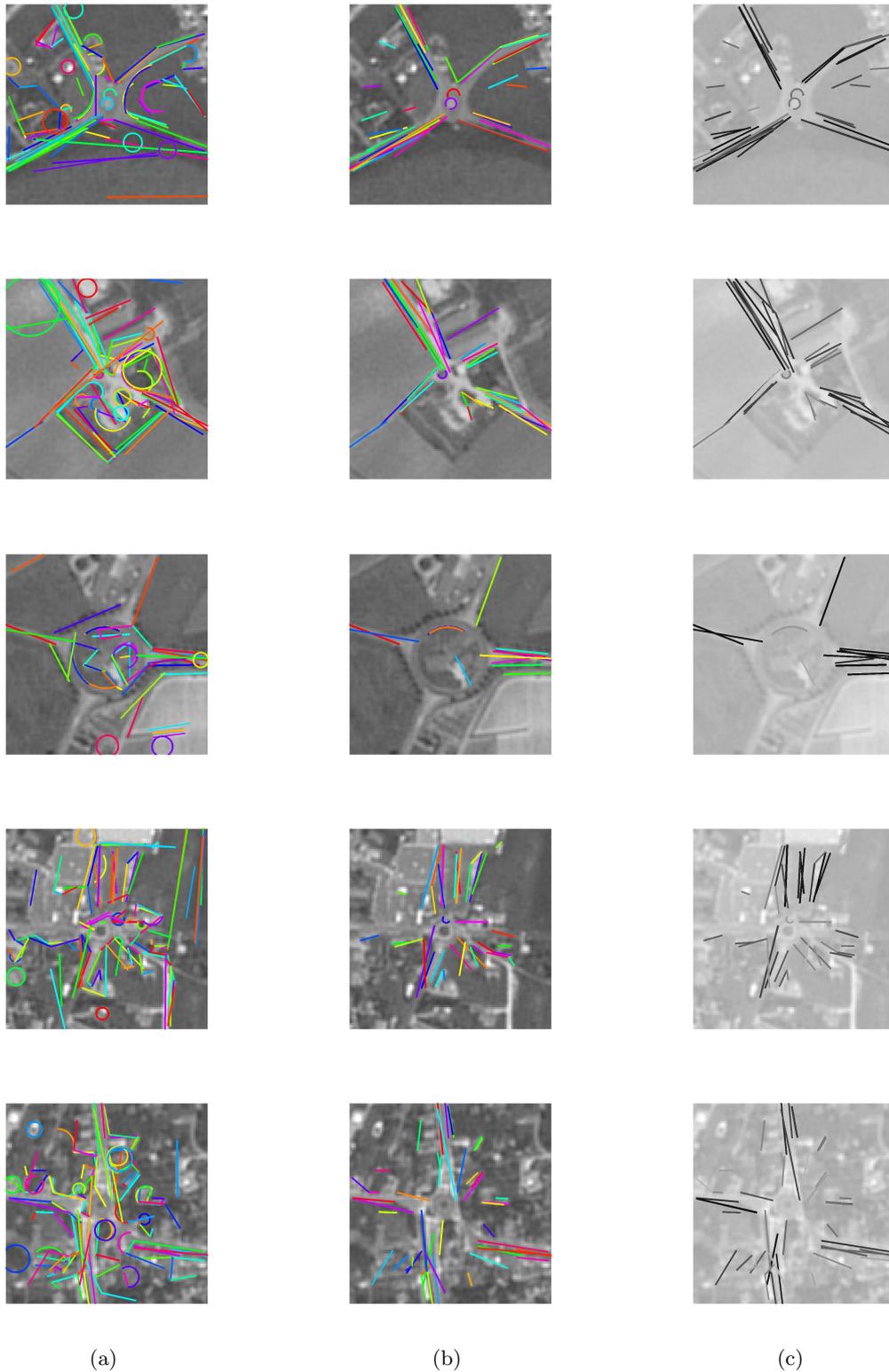


FIG. 6.8 – Les 5 ronds-points avec les scores les plus bas d'après le score  $s_{RP}$ . (a). Toutes les primitives détectées, (b). Les primitives sélectionnées, (c). Les primitives sélectionnées, représentées par un niveau de gris correspondant à leurs scores d'appartenance.

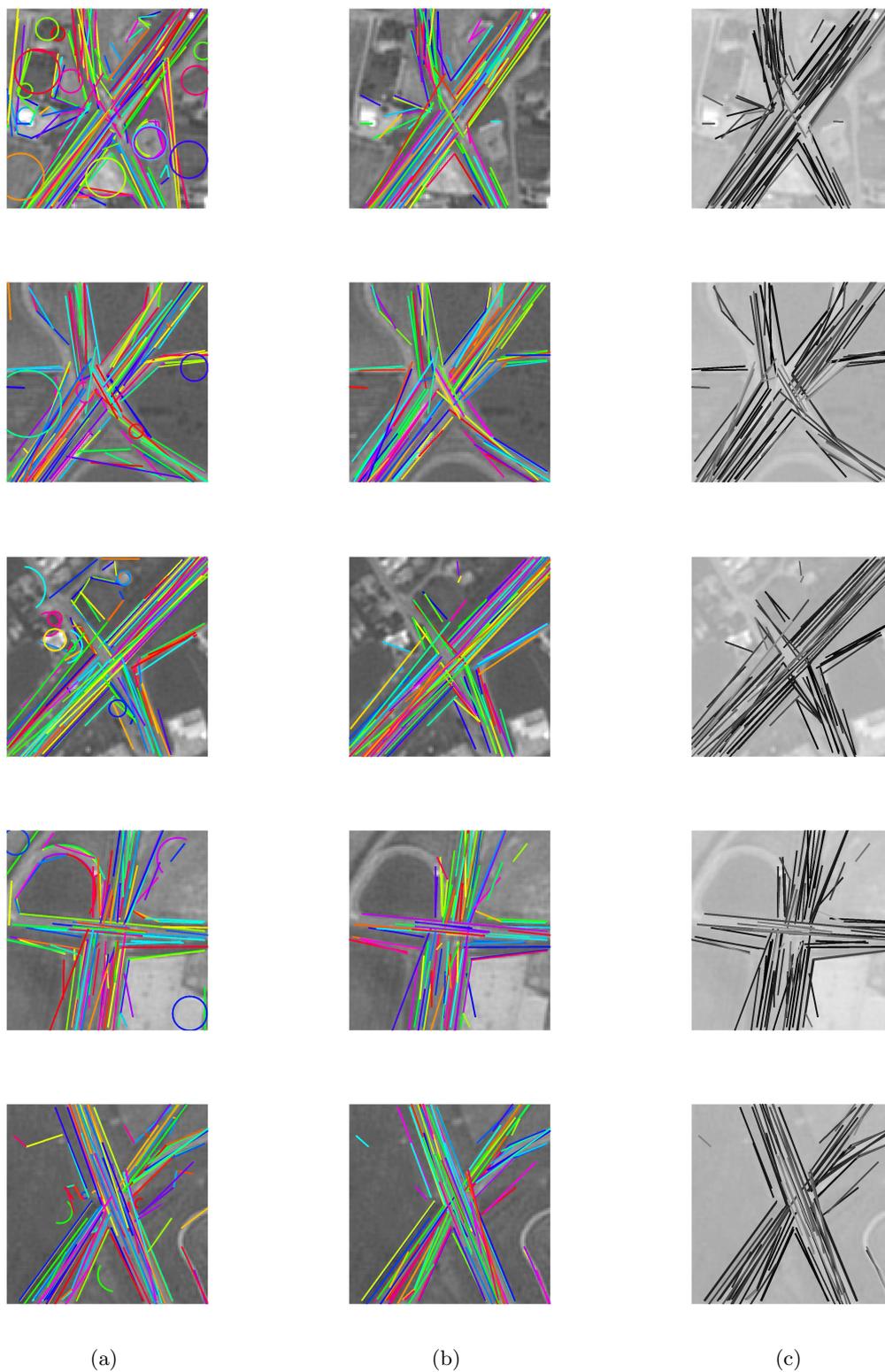


FIG. 6.9 – Les 5 images de ponts avec les scores les plus élevés d’après le score  $s_{PT}$ . (a). Toutes les primitives détectées, (b). Les primitives sélectionnées, (c). Les primitives sélectionnées, représentées par un niveau de gris correspondant à leurs scores d’appartenance.

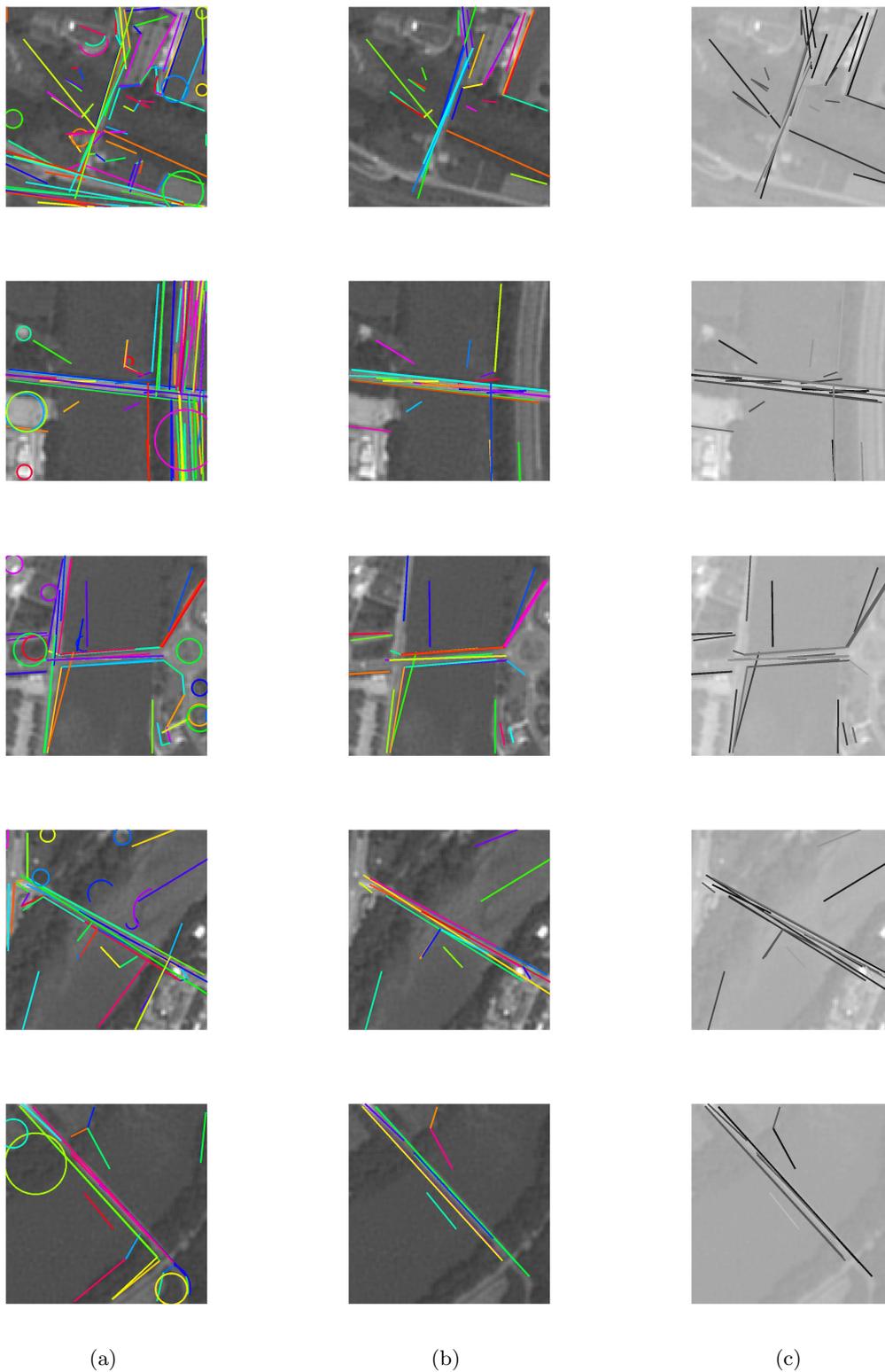


FIG. 6.10 – Les 5 images de ponts avec les scores les plus bas d'après le score  $s_{PT}$ . (a). Toutes les primitives détectées, (b). Les primitives sélectionnées, (c). Les primitives sélectionnées, représentées par un niveau de gris correspondant à leurs scores d'appartenance.

Les tableaux 6.5 et 6.6 montrent le nombre d’objets de chaque catégorie dans les  $n$  premières images où  $n$  correspond au nombre d’exemples positifs dans l’ensemble de test.

TAB. 6.5 – Le nombre d’objets de chaque catégorie dans les 36 premières images obtenues par le  $\mathcal{C}_{\mathcal{RP}}$ .

Catégorie d’objet	Rond-point	Bâti isolé	Carrefour routier
Nombre d’objet	30	4	2

TAB. 6.6 – Le nombre d’objets de chaque catégorie dans les 49 premières images obtenues par le  $\mathcal{C}_{\mathcal{PT}}$ .

Catégorie d’objet	Rond-point	Autoroute	Carrefour	Route moyenne	Voie ferrée
Nombre d’objet	31	9	5	3	1

## 6.8 Conclusion

Nous avons développé une méthode de classification fondée sur un modèle composants et structures pour reconnaître les ronds-points et les ponts. La principale contribution de ce chapitre est l’utilisation des propriétés géométriques des primitives et leurs organisations spatiales pour construire un modèle implicite d’objet en utilisant un algorithme d’apprentissage statistique.

On a obtenu une nette amélioration des résultats pour les ponts par rapport à la méthode précédente tout en ayant de très bons résultats pour les ronds-points.

Un des atouts de la méthode proposée est la possibilité de visualiser les primitives qui constituent les composants de l’objet. Cette propriété du modèle pourrait être utilisée pour la segmentation de l’objet ou pour l’apprentissage d’un modèle structurel d’objet.

Nous croyons que l’utilisation d’une approche structurelle est surtout essentielle pour résoudre les problèmes de reconnaissance observés dans les deux méthodes développées. Un modèle explicite qui correspond à la définition symbolique de l’objet (par exemple, “un pont est une route qui passe sur un objet cartographique”) serait capable de représenter les différents objets d’une même classe.

En effet, comme on peut voir sur la figure 6.11, les propriétés géométriques et l’organisation spatiale des primitives sélectionnées des objets qui ont obtenu les meilleurs scores correspondent à la structure du modèle explicite généré dans le chapitre 3. Nous pensons que c’est un rapprochement remarquable entre l’approche fondée sur l’apparence et l’approche structurelle. Nous envisageons comme perspective d’utiliser les résultats de ce chapitre comme entrée à la construction du modèle explicite d’objet. De cette façon on pourrait obtenir le modèle à partir d’images non-segmentées.

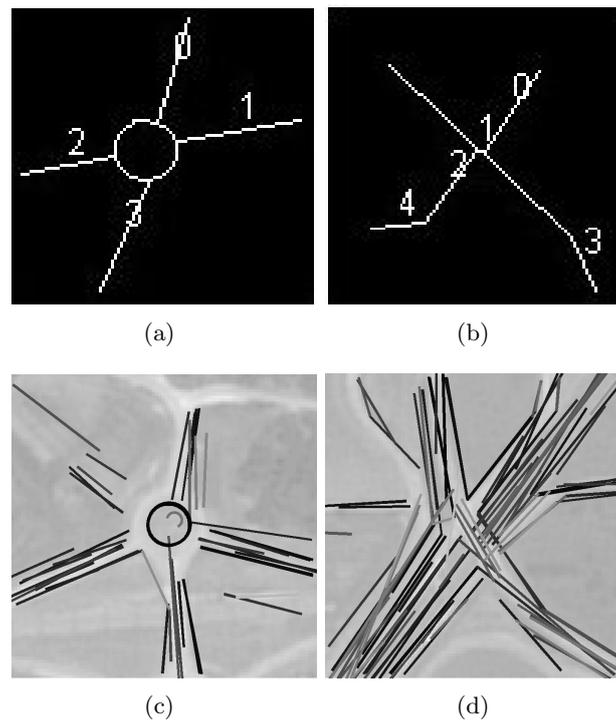


FIG. 6.11 – La correspondance entre la segmentation obtenu par l'approche composants et structure, et le modèle généré par l'approche structurelle. (a-b). Le minimal sur-graphe commun du modèle d'objet explicite de rond-point et de pont, (c-d). Les primitives sélectionnés d' une image de rond-point et d'une image de pont qui ont eu des scores de classification élevés à la classification par l'approche composants et structure.



## Quatrième partie

# Applications, évaluations, conclusions et perspectives

Cette partie présente dans un premier temps deux applications concrètes et leurs évaluations dans le chapitre 7. Nous avons participé à une compétition nationale dans la catégorie de la classification d'objets cartographiques compacts. Nous avons également développé une méthode de détection de ronds-points sur des images satellitaires de grande taille fournies par le CNES. Ces deux applications nous ont permis d'évaluer la performance de nos méthodes dans des problèmes concrets de reconnaissance.

Finalement, dans le chapitre 8 nous donnons les conclusions principales que nous avons retirées de nos travaux de recherche et indiquons les principales directions de recherche pour les travaux futurs.



## Chapitre 7

# Applications et Evaluations

### 7.1 Introduction

Nous avons présenté les méthodes de reconnaissance d’objets développées et leurs résultats dans les chapitres précédents. Nous voulons finalement démontrer l’utilisation de ces méthodes dans deux applications concrètes.

Dans la section 7.2 les résultats de la méthode de classification multi-classes que nous avons appliquée dans le cadre du projet *ROBIN* à **la compétition de classification d’objets cartographiques compacts** sont exposés.

La section 7.3 est consacrée à une méthode de détection de ronds-points dans les images satellitaires de grande taille.

### 7.2 La compétition *ROBIN*

Le projet *ROBIN* vise à mettre en place une structure d’évaluation d’algorithmes de reconnaissance d’objets sur des données d’intérêt opérationnel. Cela concerne la production des bases d’images, des vérités terrains, des compétitions, et des métriques et protocoles d’évaluation.

Le CNES a participé au projet avec une base d’images d’objets cartographiques. Cette base contient à peu près 150 images d’apprentissage et 150 images de test dans chacune de dix classes d’objets cartographiques. On a défini “une catégorie d’objets cartographiques compacts” qui se constitue des images de ronds-points (*RP*), des carrefours (*CR*), des bâtis isolés (*BI*) et des ponts (*PT*). La compétition de classification concernant cette catégorie consiste en la classification des images de test dans une des quatre classes d’objets. En pratique, cette étape de classification sera suivie d’une étape de localisation d’objets compacts sur une image satellitaire de grande taille.

Nous avons participé à cette compétition avec notre méthode de reconnaissance fondée sur le modèle implicite.

Notre méthode est initialement conçue pour les problèmes de classification binaire. Dans le cadre de la compétition, où on a un problème de classification multi-classe, nous avons proposé deux méthodes de votes pour classifier les quatre classes.

[Tax et Duin, 2002] a proposé un schéma de vote utilisant l’approche “un contre tous” ( $\mathcal{V}_1$ ) : ayant  $k$  classes  $\{c_1, \dots, c_k\}$ , pour chaque  $i$ , on considère chaque classe  $c_i$  comme la classe positive et les  $k - 1$  classes qui restent comme la classe négative. On construit un classifieur binaire  $\mathcal{C}_i$  entre ces deux classes. Une valeur de confiance  $conf_i$  est calculée pour chaque  $\mathcal{C}_i$  et l’image est assignée à la classe  $i$  pour laquelle  $conf_i$  est maximal.

La seconde alternative est une approche “un contre un” ( $\mathcal{V}_2$ ) où on construit un classifieur  $\mathcal{C}_{ij}$  pour chaque paire de classes  $\{i, j\}$ . On calcule les scores de confiance  $conf_{ij}$  pour chaque image de test et on assigne l’image à la classe  $i$  avec

$$score_i = \sum_j conf_{ij}$$

maximal. La confiance retournée par un classifieur binaire peut être utilisé en tant que telle où discrétisée dans l’ensemble  $(+1, -1)$ .

Dans l’approche  $\mathcal{V}_1$  on apprend 4 classifieurs, un pour chacune des classes d’objet à partir des images d’apprentissage. Dans l’approche  $\mathcal{V}_2$  on apprend 6 classifieurs, un pour chaque paire de classes. Pendant l’apprentissage on calcule pour chaque image de test un score correspondant à chaque classifieur et on calcule un score final par le mécanisme de vote utilisé pour déterminer la classe de l’image.

Nous avons utilisé trois ensembles de primitives pour la classification : les primitives régions ( $Prim_R$ ), les primitives contours ( $Prim_C$ ) et toutes les primitives ensemble ( $Prim_{Tous}$ ).

Les résultats sont évalués d’après les principes et les métriques de la compétition *ROBIN* proposés par Dangelo *et al.* [2006] pour observer la fiabilité du score de confiance assigné à une classification. La décision de la classe est le résultat d’un seuillage sur la valeur de confiance  $conf$ . Le seuil est initialement fixé à  $conf$  minimale. Quand on augmente le seuil  $s$ , les objets dont le  $conf < s$  sont classifiés comme “Ambigu”.

**La discrimination** dénote le taux d’objets correctement classifié en négligeant les objets “Ambigus” ;

**L’incertitude** dénote le taux d’objets ambigus.

Trois points d’opération spécifiques sont mesurés :

- le taux de discrimination à l’incertitude minimale :  $D$  ;
- le taux d’incertitude à la discrimination maximale :  $U$  ;
- le taux de discrimination et d’incertitude égales :  $EDU$ .

Le tableau 7.1 montre le taux de classification à l’incertitude minimale pour chaque cas par les deux méthodes  $\mathcal{V}_1$  et  $\mathcal{V}_2$ . On observe que par l’utilisation des primitives conjointement le taux

TAB. 7.1 – Les taux de classification à l’incertitude minimal.

	$Prim_R$	$Prim_C$	$Prim_{Tous}$
$\mathcal{V}_1$	0.8068	0.7486	<b>0.8087</b>
$\mathcal{V}_2$	0.6995	0.7432	0.7923

de classification correcte a significativement augmenté et nous avons obtenu un score supérieur à 80 %. Cela indique la complémentarité des deux types de primitives.

La figure 7.1 présente le graphe qui représente les valeurs des points d’opérations pour le meilleur classifieur ( $\mathcal{V}_1$  avec  $Prim_{Tous}$ ) qui a obtenu les scores  $D = 0.8087$ ,  $U = 0.8256$  et  $EDU = 0.8258$ .

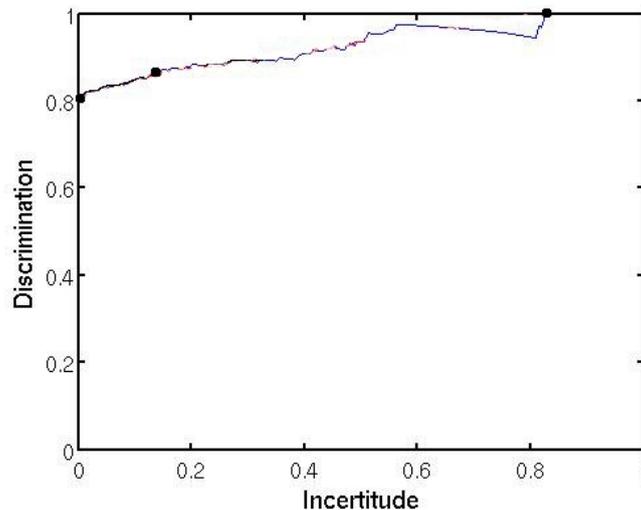


FIG. 7.1 – Les résultats de la classification. Les valeurs de  $D$ ,  $EDU$  et  $U$  sont respectivement marquées par des cercles de gauche à droite.

Avec ces résultats nous avons obtenu **le premier rang dans notre catégorie** dans la compétition *ROBIN*.

Nous avons effectué une analyse plus détaillée pour comprendre les raisons des erreurs de classification et pour voir s’il existe des motifs de confusion significatifs (Nous présentons les résultats seulement pour le meilleur classifieur dans l’intérêt de clarté).

La matrice de confusion est donnée au tableau 7.2. Le taux d’erreur le plus important est due à la confusion entre les classes  $CR$  et  $PT$ . Cela peut s’expliquer par la similarité structurelle élevée de ces deux classes. Elles sont toutes les deux composées des croisements des routes et il est difficile de les distinguer même manuellement sur un certain nombre d’images. Nous avons testé une classification à trois classes, en groupant les  $CR$  et  $PT$  dans une classe unique  $\{CR, PT\}$ . Nous avons obtenu un taux de classification globale de 89% avec 93% de classification correcte pour la classe  $\{CR, PT\}$

TAB. 7.2 – La matrice de confusion  $M$ .  $M(i, j)$  est le taux d'objets dans la classe  $i$  classifié comme  $j$ .

	RP	CR	BI	PT
RP	<b>0.8611</b>	0.0278	0.0833	0.0278
CR	0.0204	<b>0.6939</b>	0.0816	0.2041
BI	0.0612	0.1020	<b>0.8367</b>	0
PT	0	0.1429	0	<b>0.8571</b>

Quelques exemples de mauvaise classification sont montrés à la figure 7.2. Par une analyse qualitative des erreurs de classification, nous avons observé qu'ils se produisent pour la plupart du temps à cause des objets de petite taille (figure 7.2.(b),(c),(f)), ou avec les objets qui ont une forme irrégulière (figure 7.2.(a),(d),(e)). Cela est conforme à nos objectifs, car les résultats devraient s'améliorer pour les images à une meilleure résolution spatiale. Parmi les 7 ponts mal classifiés, 6 sont des ponts sur des voies fluviales (figure 7.2.(g)-(i)).

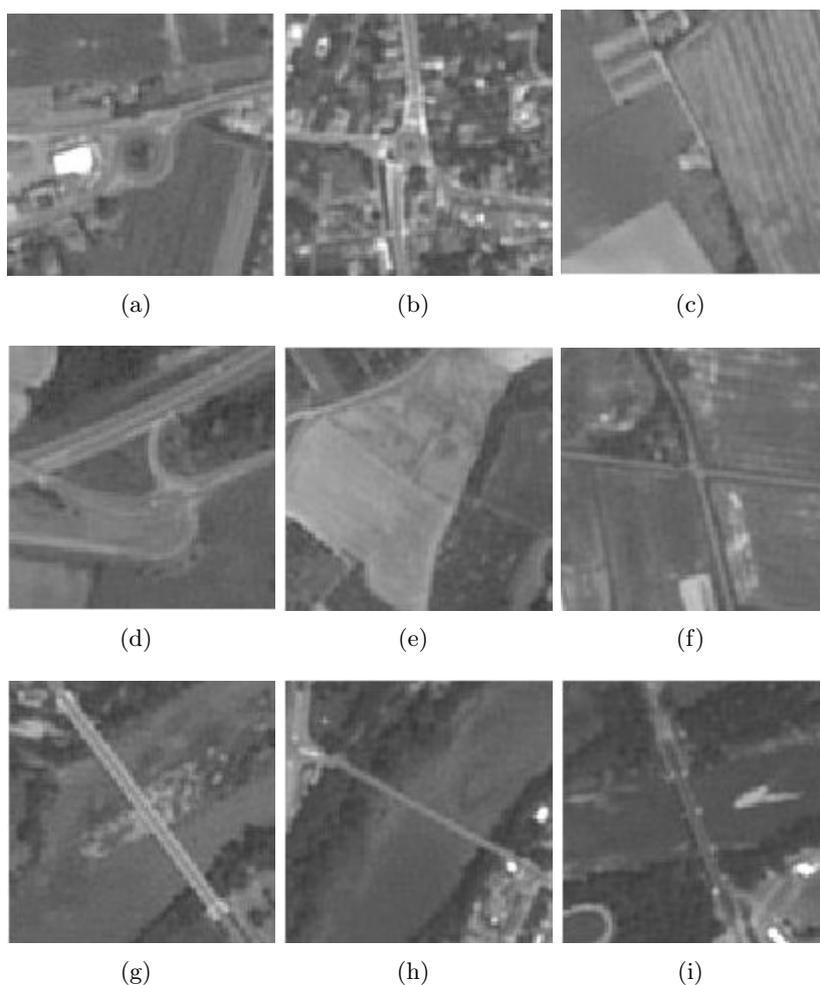


FIG. 7.2 – Exemples d'erreurs de classification. (a). RP classifié comme BI, (b). RP comme CR, (c). BI comme CR, (d)-(e). CR comme BI, (f). CR comme PT, (g)-(i). PT comme CR.

## 7.3 La détection des ronds-points sur une image satellitaire de grande taille

Le CNES nous a fournit récemment des images satellitaires de taille 24000x24000 pixels, accompagnées de coordonnées d'objets cartographiques détectés manuellement par un expert. Finalement, nous avons développé un système de détection sur ces images, qui pourrait constituer un bon exemple à l'utilisation éventuelle des méthodes proposées dans une application pratique.

Un système de détection de ronds-points est proposé parce qu'il est possible de développer sans trop de difficulté une première étape de focalisation qui réduit le temps de recherche considérablement. La généralisation aux autres objets cartographiques nécessite un travail supplémentaire qui n'a pas été encore abordé.

Nous avons testé notre méthode sur une image de taille 2500x1400 pixels, correspondant à la partie de l'image de grande taille avec la densité de ronds-points maximale. Sur cette image 8 ronds-points sont répertoriés. La figure 7.3 montre ces ronds-points.

Notre méthode est constituée de deux étapes principales :

- La focalisation : le but est de détecter les régions susceptibles de contenir des ronds-points. Le rond central est un indicateur bien approprié de l'existence d'un rond-point. Une approche fondée sur l'extraction des primitives est utilisée pour détecter les régions et les contours circulaires. On extrait tous les arcs dans l'image et les centres des objets sont considérés comme les centres des cercles correspondants.
- La classification : on découpe une imagerie de taille 100x100 pixels autour de chaque point détectés. On classe chaque imagerie dans une des classes "objet" ou "non-objet".

La première étape de la méthode proposée est la détection des régions susceptibles de contenir un rond-point afin d'éviter un temps de recherche très élevé. Cette recherche est effectuée par la méthode de détection des arcs de cercles proposée dans le chapitre 4.

La détection a donné 1756 arcs de cercles dans l'image. Pourtant la plupart de ces arcs sont des chaînes de contours courtes avec un angle d'ouverture très faible. Un filtrage par un seuil d'angle d'ouverture et un seuil de rayon a réduit ce nombre considérablement. Finalement nous avons obtenus 81 arcs de cercles (Figure 7.4). On a découpé une imagerie de taille 100x100 pixels autour du centre de chaque arc de cercle détecté et de cette façon on a obtenu un ensemble d'images de test.

L'ensemble de test est classifié en utilisant la méthode fondée sur les composants et la structure (Chapitre 6) en utilisant le classifieur appris pour les ronds points à partir d'ensemble d'images d'apprentissage. On a trié les scores par ordre décroissant et on a sélectionné un seuil de décision manuellement. Le tableau 7.3 montre les résultats de la classification.

La figure 7.5 présente les ronds-points détectés (la somme de  $TP$  et  $FN$  est égale à 10 au lieu de 8 parce qu'on a détecté deux ronds-points supplémentaires qui n'ont pas été répertoriés).



FIG. 7.3 – Les ronds-points détectés par un expert sur l'image satellitaire utilisée pour le test.

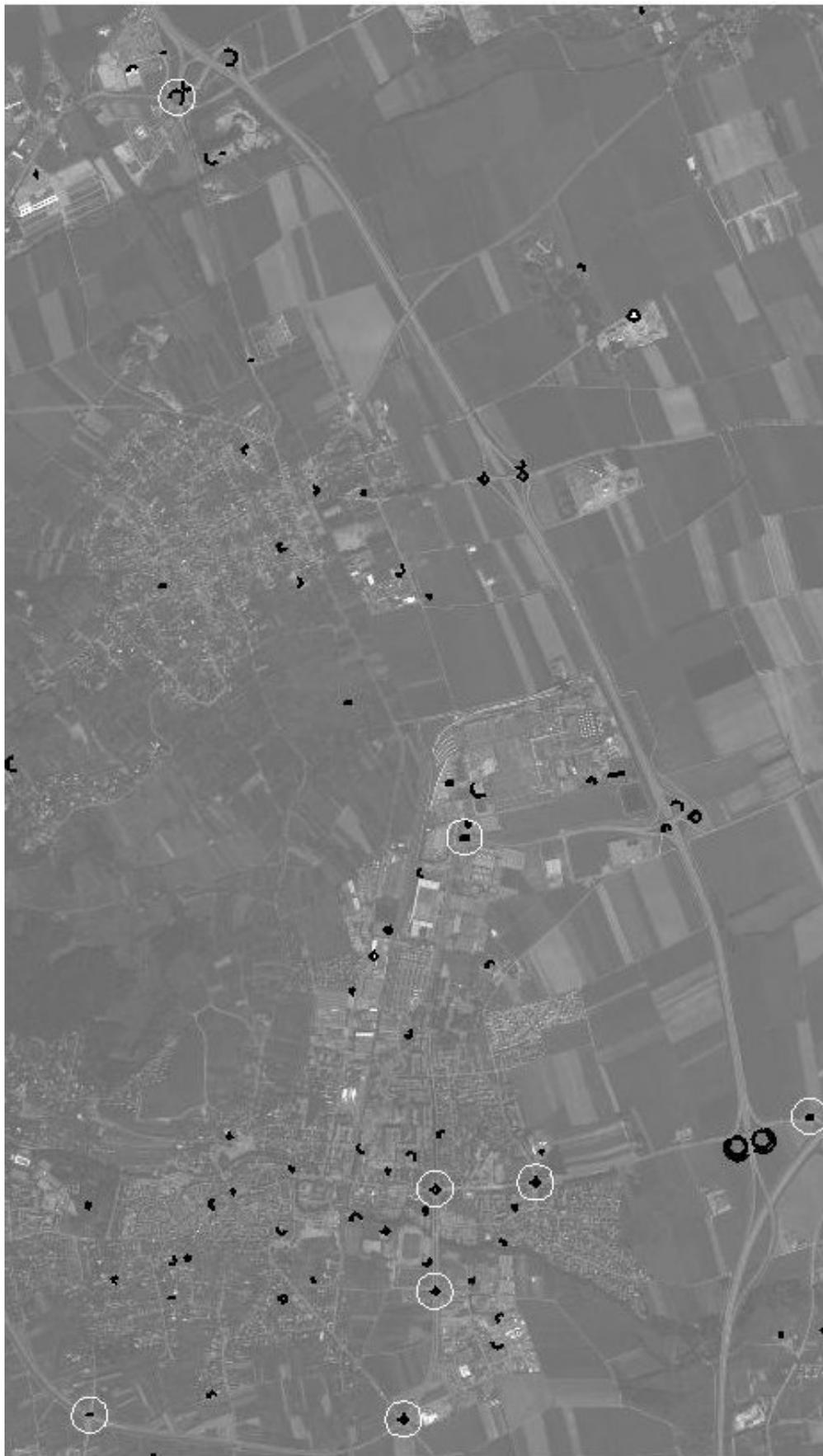


FIG. 7.4 – Les arcs de cercles détectés sur l'image satellitaire utilisée pour le test.

TAB. 7.3 – Les résultats de la classification par l’approche composants et structure (VP=vrais positifs, FP=faux positifs, FN=faux négatifs, VN=vrais négatifs).

<i>VP</i>	<i>FP</i>	<i>FN</i>	<i>VN</i>
7	1	3	80

Les résultats sont encourageant. L’application de la méthode sur l’ensemble d’images de grande taille et une analyse détaillée des résultats sont dans nos perspectives à court-terme.

## 7.4 Conclusion

Ce chapitre présente deux applications concrètes fondées sur les méthodes de reconnaissance d’objets cartographiques que nous avons développées (Chapitre 5 et chapitre 6).

La première application est un système de reconnaissance développé dans le cadre du projet *ROBIN*. Le projet réunit les acteurs industriels et institutionnels majeurs en imagerie pour la préparation des bases d’images, et les chercheurs pour le développement des algorithmes de reconnaissance d’objets. Le défi principal de ce projet est de refléter les besoins des utilisateurs réels. La participation à ce projet a été une très bonne occasion pour nous d’évaluer nos méthodes dans un cadre général et bien structuré. Nous avons obtenu des résultats très satisfaisants.

L’objectif de la deuxième application est de détecter les ronds-points dans les images cartographiques de grande taille. Nous sommes parvenus à détecter un grand nombre d’objets avec un taux de fausse détection faible.

Nous croyons que ces deux applications témoignent de l’utilité de nos méthodes pour la résolution des problèmes réels en imagerie satellitaire.



FIG. 7.5 – Les ronds-points détectés sur l'image satellitaire utilisée pour le test. (cercle blanc : objets répertoriés, cercle vert : bonne détection, cercle rouge : mauvaise détection, cercle jaune : bonne détection, mais l'objet n'est pas répertorié.



## Chapitre 8

# Conclusion et perspectives

Le sujet de notre thèse s'inscrit dans un contexte actuel. Les satellites de nouvelle génération sont équipés de systèmes optiques qui fournissent des images de la terre à une très haute résolution. Cette rapide avancée technologique rend nécessaire le développement de méthodes de traitement adaptées aux particularités de ces images. Nous avons travaillé sur un problème très appliqué, la reconnaissance d'objets cartographiques dans les images satellitaires à très haute résolution.

La principale difficulté de ce problème réside dans la variance intra-classe très importante des objets cartographiques qui sont souvent constitués de composants très similaires mais articulés suivant des organisations spatiales différentes. Les objets appartenant à une même catégorie ont des orientations, des tailles et des formes très variables.

Dès le début de nos travaux de recherche notre principale préoccupation a été d'explorer des méthodes de reconnaissance qui tirent profit de la très haute résolution des images satellitaires. A partir de cette problématique initiale, nous avons établi deux objectifs globaux : la construction d'**un modèle d'objet de haut niveau** qui est principalement fondé sur **les caractéristiques structurelles** des objets, et l'utilisation de ce modèle dans une application réelle pour la reconnaissance d'objets cartographiques dans les images satellitaires.

Notre ambition d'explorer les caractéristiques structurelles des objets nous a guidé avant tout au niveau de la sélection des descripteurs de bas niveau qu'on extrait à partir des images pour construire le modèle d'objet. Les objets construits par l'homme sont en général constitués de composants qui ont des formes géométriques simples, contrairement aux objets naturels qui ont des formes plus irrégulières. Suivant cette observation, un module d'extraction de **primitives géométriques simples** dans les images par deux approches complémentaires, une approche contours et une approche régions, a été proposé. Les étapes suivantes de construction de modèle ont été exclusivement fondées sur ces primitives géométriques.

Dans un premier temps nous nous sommes intéressés aux **méthodes d'apprentissage structurel**. Ces méthodes qui visent à construire **un modèle explicite d'objet** fondé sur une représentation symbolique nous semblent être convenables pour représenter les images sa-

tellitaires à un haut niveau d'abstraction.

Nous avons réussi à générer un modèle qui généralise les modèles structurels d'objets appartenant à une même catégorie, et dont la représentation visuelle est semblable à l'image de l'objet tracée manuellement.

Nous avons retiré deux conclusions principales de cette première approche structurelle. L'utilisation des caractéristiques géométriques est appropriée pour révéler la structure inhérente d'objets cartographiques. En revanche les méthodes structurelles ont une difficulté certaine d'applicabilité aux problèmes réels. Les modèles obtenus ne sont pas encore matures d'un point de vue théorique mais mériteront d'être mieux étudiés dans leur capacité à intégrer les relations spatiales entre les objets, un sujet émergent actuel.

Dans un deuxième temps nous nous sommes orientés vers l'apprentissage d'un modèle implicite plus souple par **les approches fondées sur l'apparence de l'objet**, un domaine de recherche relativement nouveau mais en plein développement, qui est appliqué sur plusieurs problèmes de reconnaissance d'objets actuels avec succès. L'idée principale est de représenter un objet par l'apparence des régions locales significatives qui sont observées dans plusieurs images de l'objet. Cette approche permet de poser le problème de reconnaissance comme un problème d'apprentissage supervisé et d'utiliser des méthodes statistiques pour l'apprentissage d'un modèle d'objet.

Les deux méthodes proposées sont testées sur un problème de classification où on vise à reconnaître les ronds-points et les ponts parmi les images appartenant aux dix catégories d'objets.

Nous avons d'abord développé une méthode de reconnaissance fondée sur **la fusion des classifieurs faibles par l'algorithme Adaboost**. Cet algorithme est utilisé récemment pour la construction d'un système de détection de visages très performant. L'idée principale est de fusionner des classifieurs faibles qui ne sont pas capables individuellement de bien classifier les exemples, mais qui peuvent ensemble construire un classifieur puissant. En effet, les primitives géométriques sont des bons candidats pour construire des classifieurs faibles qui servent à distinguer rapidement une certaine partie des objets cartographiques. Les classifieurs faibles sont obtenus en groupant les primitives dans des mots visuels d'après leurs propriétés géométriques et leurs organisations spatiales et en comptant le nombre des mots.

Cette méthode a donné un taux de classification élevé pour les ronds-points et moins élevé pour les ponts.

La deuxième méthode proposée est fondée sur **une approche "composants et structure"**, où on représente les composants de l'objet par des régions d'intérêt locales et la structure par les positions relatives de ces régions. Nous avons adapté la première étape d'apprentissage de composants à notre problème en utilisant les primitives géométriques comme les composants éventuels de l'objet. Pour représenter la structure de l'objet on a calculé les propriétés spatiales des primitives. Comme les objets **sont centrés** dans nos images, le centre de l'objet est utilisé comme un point de repère pour calculer les relations spatiales des composants. De façon générale,

nous pensons que cette notion de référentiel (points de repères centraux par exemple) dans le cadre de la manipulation des relations spatiales est une notion essentielle et constitue une des difficultés majeures de la mise en œuvre des algorithmes. Dans le cas satellitaire ceci est encore plus prononcé que dans le cadre de l'imagerie médicale où une certaine référence naturelle existe (droite, gauche, etc.). En clair, il manque un cadre théorique général pour ce genre d'études.

Cette méthode a donné de très bons résultats expérimentaux à la reconnaissance des ronds-points, et des résultats beaucoup plus élevés que la méthode précédente à la reconnaissance des ponts. Un autre atout de cette méthode est l'obtention d'une segmentation de l'objet par la sélection des primitives les plus significatives qui permet de visualiser les composants de l'objet.

La plupart des objets mal classifiés par les deux méthodes sont les mêmes. Ces objets ont en général des propriétés communes. On pourrait même les grouper dans des sous catégories d'objets. Ce sont les ronds-points qui ont une petite taille et les ponts qui passent sur des voies fluviales. Ces objets ne permettent pas d'extraire à partir de l'image des descripteurs géométriques nécessaires pour représenter leur structure.

Les contributions principales des deux méthodes proposées sont l'application des approches récentes de reconnaissance d'objets dans le domaine de l'imagerie satellitaire à haute résolution, l'utilisation d'une représentation fondée sur les propriétés structurelles des objets et l'obtention d'un système de classification d'objets cartographiques fonctionnel.

Les méthodes proposées sont utilisées dans deux applications de reconnaissance concrètes. Nous avons participé au projet ROBIN qui vise à mettre en place une structure d'évaluation d'algorithmes de reconnaissance d'objets sur des données d'intérêt opérationnel. Dans **la catégorie de compétition sur la détection d'objets cartographiques compacts** nous avons obtenu les meilleurs scores parmi les participants.

Finalement, nous avons développé **un système de détection de ronds-points dans les images satellites de grande taille** en utilisant les deux méthodes de classification, précédées d'une étape de focalisation sur des régions susceptibles de contenir un rond-point.

De nombreuses approches sont envisageables pour la reconnaissance d'objets cartographiques. Nous avons essayé d'en explorer certaines en nous concentrant principalement sur les caractéristiques structurelles des objets pour mieux bénéficier de l'apport de la haute résolution spatiale. En particulier les méthodes fondées sur l'apparence se sont révélées efficaces pour la reconnaissance. De nombreux points peuvent cependant être améliorés :

- **Le raisonnement spatial qualitatif** permet de manipuler l'information de nature spatiale d'une façon efficace. C'est un domaine de recherche qui fournit des formalismes théoriques pour représenter les relations entre les entités dans l'espace. Même si la mise en œuvre de ces formalismes est actuellement limitée à quelques applications simples en traitement d'images, leur utilisation serait essentielle pour franchir le fossé sémantique entre les caractéristiques de bas niveau de l'image et la représentation de haut niveau des objets. Nous avons très peu abordé les relations spatiales et nous les avons exprimées d'une ma-

nière simplifiée. L'intégration des outils du raisonnement spatial qualitatif dans le système de reconnaissance est une perspective d'une grande importance selon nous.

- Nous pensons que **l'utilisation des connaissances externes** pour interpréter les images cartographiques peut améliorer la reconnaissance d'objets. L'intégration **des règles sémantiques de** et **une interaction avec l'utilisateur final** peuvent permettre une meilleure résolution des conflits. Nous avons démarré un projet pour développer un système interactif de détection d'objets cartographiques en collaboration avec le Laboratoire de Traitement d'Images et Reconnaissance de Formes de l'université METU à Ankara, qui est en cours de développement.
- Une autre perspective à court terme est l'intégration de l'approche structurelle et l'approche fondée sur l'apparence de l'objet. Les composants géométriques détectés à la fin du chapitre 6 sont assez proches des primitives détectées à partir d'images segmentées pour construire le modèle explicite au chapitre 3. La construction du modèle en utilisant ces primitives peut permettre d'obtenir une représentation structurelle à partir des images non-segmentées.

Nous croyons que nos travaux de recherche apportent des éléments de réponse aux objectifs que nous avons posés et contribuent au développement de nouveaux systèmes de traitement d'images satellitaires à très haute résolution.

# Bibliographie

- A. GRUEN, E. P. B. et HENRICSSON, O., éditeurs (1998). *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*. Birkhauser Boston.
- ABRAMSON, Y., STEUX, B. et GHORAYEB, H. (2007). Yet even faster (yef) real-time object detection. *International Journal of Intelligent Systems Technologies and Applications*, 2(2/3):102–112.
- AGARWAL, S. et ROTH, D. (2002). Learning a sparse representation for object detection. In *ECCV '02 : Proceedings of the 7th European Conference on Computer Vision-Part IV*, pages 113–130, London, UK. Springer-Verlag.
- AMORES, J., SEBE, N. et RADEVA, P. (2005). Fast spatial pattern discovery integrating boosting with constellations of contextual descriptors. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005.*, volume 2, pages 769–774.
- ANDREW, A. M. (2000). Level set methods and fast marching methods : Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science. *Robotica*, 1:89–92.
- AUCLAIR, A., COHEN, L. D. et VINCENT, N. (2007). How to use SIFT vectors to analyze an image with database templates. In *In 5th International Workshop on Adaptive Multimedia Retrieval, Paris, France, July 5-6*.
- BARDINET, E., VIDAL, S., ARROYO, S., MALANDAIN, G. et BLANCA, N. (2000). Structural object matching. Rapport technique, Dept. of Computer Science and AI, University of Granada, Spain.
- BARNARD, K., DUYGULU, P., FORSYTH, D., de FREITAS, N., BLEI, D. M. et JORDAN, M. I. (2003). Matching words and pictures. *J. Mach. Learn. Res.*, 3:1107–1135.
- BEUCHER, S., BILODEAU, M. et YU, X. (1990). Road segmentation by watershed algorithms. In *Proc. of PROMETHEUS workshop, , Sophia-Antipolis, France*.
- BHATTACHARYA, A. (2008). *Indexation des images satellitaires en utilisant des informations structurelles*. Thèse de doctorat, TSI, Télécom Paris.
- BIEDERMAN, I. (1987). Recognition-by-components : a theory of human image understanding. *Psychol Rev*, 94(2):115–147.

- BIEDERMAN, I. (2007). Recent psychophysical and neural research in shape recognition. In N. OSAKA, I. Rentschler, . I. B., éditeur : *Object Recognition, Attention, and Action*, chapitre 5, pages 71–88. Springer.
- BLOCH, I., COLLIOT, O. et JR., R. M. C. (2006). Modélisation de la relation spatiale “entre” pour des objets d’extensions spatiales très différentes. In *RFIA 2006*.
- BORDES, G. (1997). *Interprétation d’images satellitaires à l’aide de données externes*. Thèse de doctorat, Université Marne-la-Vallée.
- BORENSTEIN, E. et ULLMAN, S. (2002). Class-specific, top-down segmentation. In *ECCV ’02 : Proceedings of the 7th European Conference on Computer Vision-Part II*, pages 109–124, London, UK. Springer-Verlag.
- BOUCHARD, G. et TRIGGS, B. (2005). Hierarchical part-based visual object categorization. In *CVPR ’05 : Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) - Volume 1*, pages 710–715, Washington, DC, USA. IEEE Computer Society.
- BRUNELLI, R. et POGGIO, T. (1997). Template matching : Matched spatial filters and beyond. *PR*, 30(5):751–768.
- BUNKE, H. et SHEARER, K. (1998). A graph distance metric based on the maximal common subgraph. *Pattern Recognition Letters*, 19.
- BURGES, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167.
- CANNY, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698.
- CARLSON, A., CUMBY, C., ROSEN, J. et ROTH, D. (1999). The snow learning architecture. Rapport technique UIUCDCS-R-99-2101, UIUC Computer Science Department.
- CARON, Y., MAKRIS, P. et VINCENT, N. (2002). A method for detecting artificial objects in natural environments. In *ICPR ’02 : Proceedings of the 16 th International Conference on Pattern Recognition (ICPR’02) Volume 1*, page 10600, Washington, DC, USA. IEEE Computer Society.
- CERELLA, J. (1986). Pigeons and perceptrons. *Pattern Recognition*, 19(6):431–438.
- CHEHATA, N. (2005). *Modélisation 3D de scènes urbaines à partir d’images satellitaires à très haute résolution*. Thèse de doctorat, Université Paris Descartes.
- CHENG, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(8):790–799.
- CIEPLINSKI, L. (2001). Mpeg-7 color descriptors and their applications. In *CAIP ’01 : Proceedings of the 9th International Conference on Computer Analysis of Images and Patterns*, pages 11–20, London, UK. Springer-Verlag.

- COLLIOT, O. (2003). *Représentation, évaluation et utilisation de relations spatiales pour l'interprétation d'images. Application à la reconnaissance de structures anatomiques en imagerie médicale*. Thèse de doctorat, Télécom Paris, ENST.
- COMANICIU, D. et MEER, P. (2002). Mean shift : A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619.
- CORDELLA, L. P., FOGGIA, P., SANSONE, C. et VENTO, M. (2002). Learning structural shape descriptions from examples. *Pattern Recognition Letters*, 23(12):1427–1437.
- CREVIER, D. et LEPAGE, R. (1997). Knowledge-based image understanding systems : a survey. *Comput. Vis. and Image Underst.*, 67(2):160–185.
- CSURKA, G., BRAY, C., DANCE, C. et FAN, L. (2004). Visual categorization with bags of keypoints. In *ECCV International Workshop on Statistical Learning in Computer Vision*, pages 1–22.
- DANGELO, E., HERBIN, S. et RATIÉVILLE, M. (2006). *ROBIN Challenge, Evaluation Principles and Metrics*. <http://robin.inrialpes.fr>.
- DEMPSTER, A. P., LAIRD, N. M. et RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.
- DESOLNEUX, A., MOISAN, L. et MOREL, J. (2000). Meaningful alignments. *Int. J. Comput. Vision*, 40(1):7–23.
- DEVERNAY, F. (1995). A non-maxima suppression method for edge detection with sub-pixel accuracy. Rapport technique RR-2724, INRIA, France.
- DI RUBERTO, C. (2004). Recognition of shapes by attributed skeletal graphs. *Pattern Recognition*, 37(1):21–31.
- DORKÓ, G. et SCHMID, C. (2003). Selection of scale-invariant parts for object class recognition. In *ICCV '03 : Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 634, Washington, DC, USA. IEEE Computer Society.
- DOUGLAS, D. et PEUCKER, T. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Canadian Cartographer*, 10:112–122.
- DUDA, R. O. et HART, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15.
- DUNHAM, J. (1986). Optimum uniform piecewise linear approximation of planar curves. *T-PAMI*, 8:67–75.
- EDISON (2003). *Edge Detection and Image SegmentatiON (EDISON) System*, <http://www.caip.rutgers.edu/riul/research/code.html>. Robust Image Understanding Laboratory, Rutgers University, New Brunswick, NJ.
- ERUS, G. et LOMÉNIE, N. (2005a). Apprentissage automatique des modèles structurels d'objets cartographiques. In PINSON, S. et VINCENT, N., éditeurs : *EGC*, volume RNTI-E-3 de *Revue des Nouvelles Technologies de l'Information*, pages 445–450. Cépaduès-Editions.

- ERUS, G. et LOMÉНИЕ, N. (2005b). Automatic learning of structural models of cartographic objects. In BRUN, L. et VENTO, M., éditeurs : *IAPR, Graph-Based Representations in Pattern Recognition*, volume 3434 de *Lecture Notes in Computer Science*, pages 273–280. Springer.
- ERUS, G. et LOMÉНИЕ, N. (2006). Extraction of cartographic objects in high resolution satellite images for object model generation. In *ICPR, 4th International Workshop on Pattern Recognition in Remote Sensing*.
- ERUS, G. et LOMÉНИЕ, N. (2007). Classification of structural cartographic objects using edge-based features. In BEBIS, G., éditeur : *ISVC (1)*, volume 4841 de *Lecture Notes in Computer Science*, pages 385–392. Springer.
- ERUS, G. et LOMÉНИЕ, N. (2008). Geometrical primitives for the classification of images containing structural cartographic objects. In *Fifth Canadian Conference on Computer and Robot Vision*.
- FELZENSZWALB, P. F. et HUTTENLOCHER, D. P. (2005). Pictorial structures for object recognition. *Int. J. Comput. Vision*, 61(1):55–79.
- FERGUS, R. (2005). *Visual Object Category Recognition*. Thèse de doctorat, Department of Engineering Science, University of Oxford.
- FERGUS, R., PERONA, P. et ZISSERMAN, A. (2003). Object class recognition by unsupervised scale-invariant learning. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 264–271.
- FERGUS, R., PERONA, P. et ZISSERMAN, A. (2005). A sparse object category model for efficient learning and exhaustive recognition. In *CVPR '05 : Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, pages 380–387, Washington, DC, USA. IEEE Computer Society.
- FERRARI, V., FEVRIER, L., JURIE, F. et SCHMID, C. (2008). Groups of adjacent contour segments for object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(1):36–51.
- FERRARI, V., TUYTELAARS, T. et GOOL, L. J. V. (2006). Object detection by contour segment networks. In LEONARDIS, A., BISCHOF, H. et PINZ, A., éditeurs : *ECCV (3)*, volume 3953 de *Lecture Notes in Computer Science*, pages 14–28. Springer.
- FISCHLER, M. A. et ELSCHLAGER, R. A. (1973). The representation and matching of pictorial structures. *IEEE Trans. Comput.*, 22(1):67–92.
- FORSSEN, P.-E. (1997). Detection of man-made objects in satellite images. Mémoire de D.E.A., Department of Electrical Engineering, Linköpings University, Sweden.
- FREUND, Y. et SCHAPIRE, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. In *EuroCOLT '95 : Proceedings of the Second European Conference on Computational Learning Theory*, pages 23–37, London, UK. Springer-Verlag.

- GALLAND, F., BERTAUX, N. et RÉFRÉGIER, P. (2003). Minimum Description Length Synthetic Aperture Radar image segmentation. *IEEE Transactions on Image Processing*, 12(9):995–1006.
- GARNESSON, P. et GIRAUDON, G. (1992). L'approximation polygonale : bilans et perspectives. Rapport technique RR-1621, INRIA - Sophia Antipolis, Equipe : PASTIS.
- HARALICK, R. (1979). Statistical and structural approaches to texture. *PIEEE*, 67(5):786–804.
- HARALICK, R. et SHAPIRO, L. (1985). Image segmentation techniques. *CVGIP*, 29(1):100–132.
- HARRIS, C. et STEPHENS, M. (1988). A Combined Corner and Edge Detector. In *4th ALVEY Vision Conference*, pages 147–151.
- HEISELE, B., SERRE, T., PONTIL, M., VETTER, T. et POGGIO, T. (2001). Categorization by learning and combining object parts. In DIETTERICH, T. G., BECKER, S. et GHAMRANI, Z., éditeurs : *NIPS*, pages 1239–1245. MIT Press.
- HONG, P. et HUANG, T. S. (2004). Spatial pattern discovery by learning a probabilistic parametric model from multiple attributed relational graphs. *Discrete Appl. Math.*, 139(1-3):113–135.
- HU, M.-K. (1962). Visual pattern recognition by moment invariants. *Information Theory, IEEE Transactions on*, 8(2):179–187.
- INGLADA, J. (2007). Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *PandRS*, 62(3): 236–248.
- IQBAL, Q. et AGGARWAL, J. (2002). Retrieval by classification of images containing large man-made objects using perceptual grouping. *Pattern Recognition Journal*, 35(7):1463–1479.
- JAIN, A. K., ZHONG, Y. et LAKSHMANAN, S. (1996). Object matching using deformable templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(3):267–278.
- KASS, M., WITKIN, A. et TERZOPOULOS, D. (1988). Snakes : Active contour models. *International Journal of Computer Vision*, V1(4):321–331.
- KAVALLIERATOU, E., FAKOTAKIS, N. et KOKKINAKIS, G. (2002). Handwritten character recognition based on structural characteristics. *icpr*, 03:30139.
- KHOTANZAD, A. et HONG, Y. H. (1990). Invariant image recognition by Zernike Moments. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(5):489–497.
- KOENDERINK, J. J. et van DOOM, A. J. (1987). Representation of local geometry in the visual system. *Biol. Cybern.*, 55(6):367–375.
- KOVESI, P. D. (2000). MATLAB and Octave functions for computer vision and image processing. School of Computer Science & Software Engineering, The University of Western Australia. Available from : <<http://www.csse.uwa.edu.au/~pk/research/matlabfns/>>.
- Le Petit Robert (1973). Déf. A.1, Dictionnaire Le Petit Robert, Société du nouveau littré.

- LEIBE, B., LEONARDIS, A. et SCHIELE, B. (2004). Combined object categorization and segmentation with an implicit shape model. *In ECCV'04 Workshop on Statistical Learning in Computer Vision*, Prague, Czech Republic.
- LEUNG, T. et MALIK, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. J. Comput. Vision*, 43(1):29–44.
- LI, F.-F. et PERONA, P. (2005). A bayesian hierarchical model for learning natural scene categories. *In CVPR '05 : Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 524–531, Washington, DC, USA. IEEE Computer Society.
- LINDBERG, T. (1998). Feature detection with automatic scale selection. *Int. J. Comput. Vision*, 30(2):79–116.
- LOMÉNIE, N., TRIAS-SANZ, R. et BARBEAU, J. (2003). Integrating textural and geometric information for an automatic bridge detection. *In Proc. of IGARSS'03*, Toulouse, France.
- LORETTE, A., DESCOMBES, X. et ZERUBIA, J. (2000). Texture analysis through a markovian modelling and fuzzy classification : Application to urban area extraction from satellite images. *International Journal of Computer Vision*, 36(3):221–236.
- LOWE, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110.
- MARR, D. et HILDRETH, E. (1980). Theory of edge detection. *In In Proc. Royal Soc. London*, volume 207 de *B*, pages 187–217.
- MATAS, J., CHUM, O., URBAN, M. et PAJDLA, T. (2004). Robust wide baseline stereo from maximally stable extremal regions. *IVC*, 22(10):761–767.
- MAURO, C. D., DILIGENTI, M., GORI, M. et MAGGINI, M. (2003). Similarity learning for graph-based image representations. *Pattern Recogn. Lett.*, 24(8):1115–1122.
- MCQUEEN, M. P. C. (1981). A generalization of template matching for recognition of real objects. *Pattern Recognition*, 13(2):139–145.
- MEYER, F. et BEUCHER, S. (1990). Morphological segmentation. *JVCIR*, 1(1):21–46.
- MOHAN, A., PAPAGEORGIOU, C. et POGGIO, T. (2001). Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):349–361.
- MURPHY, K. P., TORRALBA, A. B., EATON, D. et FREEMAN, W. T. (2006). Object detection and localization using local and global features. *In Toward Category-Level Object Recognition*, pages 382–400.
- OPELT, A. (2006). *Generic Object Recognition*. Thèse de doctorat, Graz University of Technology.
- OPELT, A., FUSSENEGGER, M., PINZ, A. et AUER, P. (2004). Weak hypotheses and boosting for generic object detection and recognition. *In PAJDLA, T. et MATAS, J., éditeurs : ECCV (2)*, volume 3022 de *Lecture Notes in Computer Science*, pages 71–84. Springer.

- OPELT, A., PINZ, A. et ZISSERMAN, A. (2006). A boundary-fragment-model for object detection. *In Proceedings of the European Conference on Computer Vision*.
- PERONA, P. et MALIK, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639.
- PETRAKIS, E. M. et FALOUTSOS, C. (1995). Similarity searching in large image databases. Rapport technique, Department of Computer Science, University of Maryland, College Park, MD, USA.
- POPE, A. et LOWE, D. (1993). Learning object recognition models from images. *In Proc. of ICCV93*, pages 296–301.
- PRICE, S. (1996). Interactive vision. Retrieved January, 2008, from the Edinburgh, Heriot Watt, and Napier Universities' Marble Project Web page : [http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\\_COPIES/MARBLE/high/pia/grouping.htm](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/MARBLE/high/pia/grouping.htm).
- RUSKONÉ, R. (1996). *Extraction automatique du réseau routier par interprétation locale du contexte : application à la production de données cartographiques*. Thèse de doctorat, Université de Marne-la-Vallée.
- SANGINETO, E. (2003). An abstract representation of geometric knowledge for object classification. *Pattern Recogn. Lett.*, 24(9-10):1241–1250.
- SANZ, R. T. et LOMÉNIE, N. (2003). Automatic bridge detection in high-resolution satellite images. *In Proc. 3rd International Conference on Computer Vision Systems (ICCVS'03)*, pages 172–181, Graz, Austria.
- SCHMID, C. et MOHR, R. (1997). Local grayvalue invariants for image retrieval. *IEEE Trans. Pattern Anal. Machine Intel.*, 19(5):530–534.
- SEBASTIAN, T. B., KLEIN, P. N. et KIMIA, B. B. (2004). Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):550–571.
- SERRA, J. (2003). Image segmentation. *In ICIP03*, pages 345–348.
- SHAO, Z. et KITTLER, J. (1999). Shape representation and recognition based on invariant unary and binary relations. *Image Vision Comput.*, 17(5-6):429–444.
- SHOTTON, J., BLAKE, A. et CIPOLIA, R. (2007). Multi-scale categorical object recognition using contour fragments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- SIDDIQI, K., SHOKOUFANDEH, A., DICKINSON, S. J. et ZUCKER, S. W. (1999). Shock graphs and shape matching. *Int. J. Comput. Vision*, 35(1):13–32.
- SIVIC, J., RUSSELL, B., EFROS, A. A., ZISSERMAN, A. et FREEMAN, B. (2005). Discovering objects and their location in images. *In International Conference on Computer Vision (ICCV 2005)*.
- TAX, D. et DUIN, R. (2002). Using two-class classifiers for multiclass classification. *In R. KASTURI, D. Laurendeau, C. S., éditeur : 16th International Conference on Pattern Recognition*

- (Quebec City, Canada, Aug.11-15), volume 2, pages 124–127, Los Alamitos. IEEE Computer Society Press.
- TORRALBA, A. B., MURPHY, K. P. et FREEMAN, W. T. (2006). Shared features for multiclass object detection. *In Toward Category-Level Object Recognition*, pages 345–361.
- TUCERYAN, M. et JAIN, A. K. (1998). *The Handbook of Pattern Recognition and Computer Vision*, chapitre Texture Analysis, pages 207–248. World Scientific Publishing Co.
- TUPIN, F., MAITRE, H., MANGIN, J.-F., NICOLAS, J.-M. et PECHERSKY, E. (1998). Detection of linear features in SAR images : Application to the road network extraction. *IEEE Trans. Geosci. Remote Sensing*, 36(2):434–453.
- TURK, M. A. et PENTLAND, A. P. (1991). Face recognition using eigenfaces. *In CVPR '91, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591.
- VAPNIK, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience.
- VENKATESWAR, V. et CHELLAPPA, R. (1992). Extraction of straight lines in aerial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(11):1111–1114.
- VIOLA, P. et JONES, M. J. (2004). Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154.
- VOGEL, J. et SCHIELE, B. (2002). On performance characterization and optimization for image retrieval. *In ECCV '02 : Proceedings of the 7th European Conference on Computer Vision-Part IV*, pages 49–66, London, UK. Springer-Verlag.
- WEBER, M., WELLING, M. et PERONA, P. (2000). Unsupervised learning of models for recognition. *In ECCV '00 : Proceedings of the 6th European Conference on Computer Vision-Part I*, pages 18–32, London, UK. Springer-Verlag.
- ZHANG, D. et LU, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19.
- ZHANG, J., MARSZALEK, M., LAZEBNIK, S. et SCHMID, C. (2007). Local features and kernels for classification of texture and object categories : A comprehensive study. *Int. J. Comput. Vision*, 73(2):213–238.
- ZHANG, T. Y. et SUEN, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Commun. ACM*, 27(3):236–239.