

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En habilitation conjointe avec l'Université de Rennes 1

Ecole Doctorale – MATISSE

Caractérisation et modélisation de la distribution spatiale de signatures locales dans les images : application à la classification d'image sonar de fonds marins.

Thèse de Doctorat

Mention : Traitement du signal et télécommunications

Présentée par **Huu Giao NGUYEN**

Département : Signal et Communications

Laboratoire : Pôle CID (Lab-STICC)

Directeur de thèse : Jean-Marc BOUCHER

Soutenue le 28 Octobre 2011

Composition du jury :

Avner BAR-HEN
Xavier DESCOMBES
Jean- Marc OGIER
Danielle DUCROT
Alain HILLION
Jean-Marc BOUCHER
Ronan FABLET
Axel Ehrhold

Professeur, Université Paris Descartes / président
Directeur de Recherche, INRIA Sophia Antipolis / rapporteur
Professeur, Université de la Rochelle / rapporteur
Maître de conférences, CESBIO Toulouse / examinateur
Professeur, Télécom Bretagne / examinateur
Professeur, Télécom Bretagne / directeur de thèse
Maître de conférences, Télécom Bretagne / directeur de thèse
Géologue, Ifremer - Plouzané / invitée

Caractérisation et modélisation de la
distribution spatiale de signatures locales
dans les images : application à la
classification d'image sonar de fonds marins.

Huu Giao NGUYEN

16 décembre 2011

Remerciements

Tout d'abord, je voudrais dire un énorme merci à Ronan et à Jean-Marc, pour votre encadrement, pour votre disponibilité, pour votre expertise scientifique et votre sympathie. Vous m'avez fait découvrir le monde de la recherche que je ne connaissais pas du tout. Vous m'avez aussi aidé à m'insérer dans le laboratoire et plus généralement dans la communauté statistique. Particulièrement, vous étiez toujours disponible pour soutenir, conseiller, faire avancer dans mes travaux et vous avez toujours mis en valeur mon travail.

Je voudrais remercier Xavier Descombes et Jean-Marc Ogier d'avoir accepté, malgré toutes vos occupations, d'être les rapporteurs de ma thèse. Je vous remercie pour votre lecture approfondie de ce manuscrit, pour vos remarques enrichissantes et pour la constitution du rapport. Je remercie également Avner Bar-Hen, Danielle Ducrot, Alain Hillion et Axel Ehrhold, qui me font le plaisir de faire partie du jury. Je vous remercie de l'intérêt que vous avez porté à mes travaux ainsi que vos suggestions qui me permettent de cibler dans mes perspectives de recherche, les plus intéressantes.

En espérant n'oublier personne, je remercie chaleureusement les membres (ou anciens membres) du département SC de TELECOM Bretagne à Brest pour votre sympathie, votre bonne humeur voire pour vos cours intéressants (dans le désordre et sans distinguer de statut particulier) : Ramesh, Monique, Abdeldjalil, Karine, Jean-Marc, Thierry, Annie, Frédéric, Sébastien, Christophe, Raphaël, Dominique, Samir, Arnaud, André, Si Mohamed, Quoc Anh, Hatim, Jia, Rongping, Moustapha, Zakaria, Yasser, Junyi, Souhaila, Elizabeth, Liyun, Quang Thang, François-Xavier, Hyunseuk, Abdelhakim, Omid, Sterenne, Sileye, Rabie.

Je remercie ma famille et mes amis pour votre soutien et particulièrement mes parents qui m'ont toujours laissé très libre dans mes choix et m'ont toujours encouragé dans ceux-là.

Cong cha nhu nui Thai Son
Nghia me nhu nuoc trong nguon chay ra

Enfin, je remercie ma femme Uyen, pour ton amour, ton soutien et pour tous ces événements riches en émotions que nous avons connus pendant ces trois ans.

Anh yeu em.

Table des matières

Remerciements	iii
Table des matières	vii
Notations	ix
Liste des figures	xiii
Liste des tableaux	xvi
Résumé	xvii
Abstract	xix
Introduction générale	1
 I Caractérisation et modélisation statistique d'ensemble de signatures locales dans les images	 7
1 Etat de l'art sur les descripteurs locaux d'images	9
1.1 Introduction	9
1.2 Points d'intérêt	10
1.2.1 Détection de points d'intérêt	10
1.2.2 Descriptions invariantes de points d'intérêt	13
1.3 Formes	15
1.3.1 Détection de formes	15
1.3.2 Description invariante de formes	18
1.4 Conclusion	21

2	Statistiques spatiales de points marqués	23
2.1	Introduction	23
2.2	Quantification des descripteurs et mots visuels	23
2.3	Cooccurrence statistique spatiale des points marqués : méthode proposée	25
2.3.1	Processus ponctuels spatiaux	25
2.3.2	Processus ponctuels marqués	27
2.3.3	Estimation de la fonction K de Ripley	27
2.3.4	Correction des effets de bord	29
2.3.5	Réduction de la complexité du vecteur descripteur	30
2.3.6	Discussion des propriétés d'invariance des descripteurs proposés	31
2.4	Conclusion	32
3	Modèles de Cox log-Gaussiens	33
3.1	Introduction	33
3.2	Généralités sur les modèles statistiques spatiaux	33
3.3	Modélisation statistique spatiale d'ensemble de signatures locales dans l'image par processus de Cox log-Gaussien	35
3.3.1	Processus ponctuels de Cox log-Gaussien dans le cas uni-varié .	35
3.3.2	Processus ponctuels de Cox log-Gaussien dans le cas multi-varié	36
3.3.3	Estimation de fonctions covariances des modèles de Cox log- Gaussien	36
3.3.4	Réduction de la complexité du vecteur descripteur	38
3.3.5	Discussion sur les propriétés d'invariances du descripteur proposé	38
3.4	Conclusion	39
4	Reconnaissance et classification d'images	41
4.1	Introduction	41
4.2	Méthodes de classification supervisée	42
4.2.1	k-plus proches voisins	42
4.2.2	Machine à vecteurs de support	42
4.2.3	Forêt aléatoire	45
4.3	Mesures de similarité	46
4.4	Applications à la classification des images texturées naturelles	47
4.4.1	UIUC	49
4.4.2	Brodatz	53
4.4.3	KTH-TIPs	55
4.5	Reconnaissance des catégories d'images de scène naturelle	57

4.6	Conclusion	59
II Application à la classification d'images sonars de fonds marins		61
5	Caractérisation et classification d' images sonars de fonds marins	63
5.1	Introduction	63
5.2	Construction de l'image sonar	64
5.2.1	Projet Rebent - Réseau de surveillance benthique	64
5.2.2	Principe d'un sonar à balayage latéral Edgetech DF1000	65
5.2.3	Acquisition et analyse des données	67
5.3	Base de données des textures sonars	69
5.3.1	La vase	69
5.3.2	La vase sableuse	69
5.3.3	Le maërl sablo-graveleux	70
5.3.4	Le sable fin à moyen	70
5.3.5	Les sédiments mélangés	70
5.3.6	Le substratum rocheux	70
5.4	Etat de l'art sur les méthodes de caractérisation des images sonars . . .	72
5.5	Résultats obtenus par les méthodes proposées basées sur les statistiques spatiales de points marqués	73
5.6	Conclusion	79
Conclusion générale et perspectives		81
Bibliographie		87

Notations

Abréviations

SSC	Statistiques Spatiales de Cooccurrences
LGCM	Modèles Cox log-Gaussien
SDM	Sac de Mots
kPP	k-plus proches voisins
ACP	Analyse en Composantes Principales
SVM	<i>Support Vector Machine</i>
RF	<i>Random Forest</i>
DoG	<i>Difference of Gaussians</i>
LoG	<i>Laplacian of Gaussian</i>
SIFT	(descripteur) <i>Scale-invariant feature transform</i>
SURF	(descripteur) <i>Speeded Up Robust Features</i>
BRIEF	(descripteur) <i>Binary Robust Independent Elementary Features</i>
CS-LBP	(descripteur) <i>Center-Symmetric Local Binary Pattern</i>
GLOH	(descripteur) <i>Gradient Location and Orientation Histogram</i>
GIH	(descripteur) Histogramme de l'Intensité Géodésique
Har-Lap	(détecteur) Harris-Laplacien
Hes-Lap	(détecteur) Hessienne-Laplacien
FH	(détecteur) <i>Fast-Hessien</i>
IDSC	<i>Inner-Distance Shape Context</i>
FLST	<i>Fast Level Set Transform</i>
CART	<i>Classification And Regression Tree</i>
BS	<i>Backscattering Strength</i>

Symboles

Ψ	Processus ponctuel marqué
X	Processus de Cox
$\tilde{X}$	Processus Cox log-Gaussien
λ	Fonction d'intensité
$g()$	Fonction de corrélation de paires
$c()$	Fonction de covariance
$\mu^{(p)}$	Moment d'ordre p du processus ponctuel
$\alpha^{(2)}$	Mesure de moment factoriel d'ordre 2
η	Facteur d'échelle

Liste des figures

1	Illustration de l'importance de la prise en compte de l'analyse statistique spatiale de la distribution de points d'intérêts dans une image texturée.	2
1.1	Illustration de l'estimation d'un pixel local extrême avec ses 26 voisins dans un espace d'échelles.	11
1.2	Exemples de points d'intérêt détectés dans une image par les différentes approches de détecteurs décrites dans la Section 1.2.1.	12
1.3	Illustration de la construction d'un descripteur SIFT. La distribution des orientations du gradient local dans 16 quadrants 4x4 est présentée dans la figure de gauche et d'un cercle représentant la gaussienne utilisée pour le lissage. La figure de droite schématise la forme du descripteur sous la forme d'histogrammes d'orientation de gradient suivant 8 orientations.	14
1.4	Illustration de la caractérisation de descripteurs locaux effectuée autour du point d'intérêt par différentes manières d'exploration : rectangulaire (a,b) ou circulaire (c,d,e) [1].	15
1.5	Illustration des principes de la ligne de partage des eaux (a) et des évolutions du contour actif (b).	16
1.6	Arbre de composantes connexes des ensembles de niveaux supérieurs dans une image simple.	18
1.7	Illustration de la représentation d'une image par les lignes de niveau : (a) image origine, (b) les formes détectées dans un niveau de gris.	19
1.8	Exemple de représentation de shockGraph. La référence de la figure est dans le papier original [2].	20
1.9	Le calcul de l'histogramme polaire sur un point de contour de forme (a). La distance minimale entre deux points dans l'espace intérieur de forme (b). Les figures de référence sont dans les papiers originaux [3, 4]	20
2.1	La visualisation de trois différents niveaux de la représentation pyramidale de mots visuels dans une image. Le niveau 0 illustre les sacs de mots. La référence de la figure est dans le papier original [5].	24
2.2	Le schéma des étapes principales des méthodes basées sur l'analyse des informations spatiales de mots visuels.	25

2.3	Illustration de l'analyse de statistiques spatiales de cooccurrence des points marqués dans les différents cercles d'étude.	28
2.4	La visualisation des cas d'intersection du cercle étude avec la frontière image.	30
2.5	Illustration des effets d'échelle pour le calcul des statistiques descriptives d'ordre 2 (SSC) : une image de référence I_1 (a) avec le facteur d'échelle $\eta_1 = 1$ et une autre image de même texture I_2 (b) avec le facteur d'échelle $\eta_2 \approx 2$ sont considérées. Sur le tracé, l'histogramme de SSC pour l'image I_1 est en bleu qui est calculé avec une boule de rayon $r = 20$. Les vecteurs descripteurs de SSC pour l'image I_2 sont en vert (où $r = 20$ sans la procédure d'adaptation d'échelle) et en orange (où $r = 40$ avec la procédure d'adaptation d'échelle). Une comparaison de deux distances de la mesure similarité entre l'histogramme de I_1 et avec deux histogrammes de I_2 montre l'intérêt d'appliquer la procédure d'adaptation d'échelle.	32
3.1	Exemple des effets d'échelle pour l'estimation des paramètres de la fonction covariance du modèle de Cox log-Gaussien appliqué à la réduction de la complexité. Nous considérons deux images suivantes : une image de référence I_1 (a) avec le facteur d'échelle $\eta_1 = 1$ et une autre image de même texture I_2 (b) avec le facteur d'échelle $\eta_2 \approx 2$. Sur les tracés, les histogrammes de paramètres de variances σ_u (c,d) et de paramètres d'échelle β_u (e,f) d'image I_1 sont en rouge. Les courbes en vert (res. en bleu) sont les histogrammes de paramètres d'image I_2 calculées avec (res. sans) la procédure d'adaptation d'échelle.	39
4.1	Illustration de la classification par la méthode des k-plus proches voisins. Considérant une base de données de 18 éléments de deux classes : cercles et carrés, un nouvel élément étoile est attribué à la classe des cercles si on considère la valeur de $k = 3$, parce que ses trois plus proche voisins sont constitués de deux cercles et d'un carré. Par contre, si on considère ses $k = 7$ voisins,, le nouvel élément est attribué à la classe des carrés car ses 7 premiers voisins sont constitués de 4 carrés pour 3 cercles. . .	43
4.2	Illustration de la principe des SVM : La transformation d'un problème de séparation non linéaire des données de l'espace d'entrée (a) ver un problème de séparation linéaire dans un espace de redescription (b). La détermination d'un hyperplan de séparation optimale qui maximise la marge dans l'espace de redescription (b).	43
4.3	Illustration de la partition en binaire et l'arbre CART associé.	46
4.4	Exemple d'images de chaque classe de texture de la base de donnée UIUC.	50
4.5	Exemples d'images de chaque classe de la base de données Brodatz.	54
4.6	Exemples d'images de chaque classe de la base de donnée KTH-TIPs.	56

4.7	Etapes principales de cette application : la détection des formes dans l'image (a), la détermination des mots visuels des formes (b), l'estimation du modèle statistique spatial des formes (c). Les catégories des formes et des points sont présentées par différentes couleurs.	58
4.8	Exemples d'images de 15 catégories de base de données de scène naturelle.	58
5.1	La carte des secteurs d'étude (en rouge) du REBENT Bretagne www.rebent.org	64
5.2	Illustration de la zone d'étude de projet Rebent dans la baie de Concarneau, Bretagne, France.	65
5.3	Exemple de base de données sonars (a) avec l'image extraite de la vidéo sous-marine (b) et l'échantillonnage biologique (c) de maërl sablo-graveleux obtenues dans les zones d'étude (d) de la baie de Concarneau (Rebent-IFREMER).	66
5.4	Poisson sonar latéral tracté en position de balayage au-dessus du fond (Rebent-IFREMER).	67
5.5	Image des sonars DF 1000 de la société Edgetech (Rebent-IFREMER).	67
5.6	(a) L'évolution de -85° à $+85^\circ$ de l'incidence de l'écho rétro-diffusé en fonction de l'angle de deux types de fond : le maërl sablo-graveleux (en rouge) et le sable fin à moyen (en vert) dans la Figure 5.3 et (b,c) les deux images sonars exemples du maërl sablo-graveleux en deux secteurs angulaires différentes (Rebent, IFREMER).	68
5.7	Exemples de textures sonars des différents types de fond marin (a) avec leurs images de l'échantillonnage biologique (b), de l'extraction de vidéo sousmarine (c) et de la localisation des zones d'étude (en bleu) de ce type de fond marin (d).	71
5.8	Exemples des différentes distributions spatiales de points d'intérêt dans les images sonars pour les différentes classes du fond marin : (a) Har-Lap détecteur, (b) DoG détecteur, (c) Hes-Lap détecteur, (d) FH détecteur.	74
5.9	Matrice de confusion de la classification supervisée de textures sonars utilisant l'approche LGCM. 5 échantillons d'apprentissage par classe sont utilisées pour cette expérimentation.	79

Liste des tableaux

2.1	La liste des conditions et des formules de calcul des angles intérieurs $\mathfrak{S}$ pour les différents cas d'intersection du cercle d'étude avec la frontière d'image.	30
3.1	Les différentes formes du modèle de paramétrisation $\mathbb{L}(\beta, r)$	37
4.1	Différentes formes des fonctions noyaux.	45
4.2	Comparaison des performances du descripteur LGCM pour la classification supervisée d'images de la base de données UIUCTex dans les cas des différents classifieurs (k-NN, SVM, RF) et des différentes mesures de similarité (Euclidienne, χ^2 , divergence de Jensen-Shannon), quand 20 images d'apprentissage sont considérées.	51
4.3	Comparaison des performances de reconnaissance d'images texturées UIUCTex pour différentes fonctions de covariance pour la définition du descripteur LGCM.	51
4.4	Comparaison des performances du descripteur LGCM pour la classification supervisée d'images texturées de la base UIUCTex dans les cas l'utilisation ou non des procédures d'adaptation d'échelle et de réduction de la taille du descripteur.	52
4.5	Comparaison des performances de reconnaissance d'images textures de la base UIUCTex pour différents types de points d'intérêt pour la définition du descripteur LGCM : DoG+Sift, FH+Surf, (Har-Lap)+(Sift-Spin), (Hes-Lap)+Daisy, FH+Brief. Le classifieur RF est considéré dans cette expérimentation.	52
4.6	Taux moyens et écarts types de bonne classification des approches proposées pour les textures UIUCTex en comparaison aux méthodes de l'état de l'art : Filtre de Gabor [6], Matrice de cooccurrence [7], sac de mots (SDM) [8], les statistiques de cooccurrence des paires des catégories de points d'intérêt [9], la description multifractale [10], la combinaison de différents types de points d'intérêt [11], statistiques descriptives spatiales de cooccurrence (SSC), modèles Cox log-Gaussien (LGCM).	52

4.7	Taux moyens et écarts types de bonne classification des approches proposées en comparaison à des techniques d'état de l'art pour la base de données d'images de l'album Brodatz.	54
4.8	Taux moyens et écarts types de bonne classification des approches en comparaison aux techniques d'état de l'art pour la base d'images KTH-Tips.	55
4.9	Taux moyens et écarts types de bonne classification de chaque classe pour une répétition de 20 fois du modèle d'apprentissage quand 100 images d'apprentissages sont considérées. Le moyen des performances en global est $82.9\% \pm 1.7$	59
5.1	Le pourcentage en moyenne des composants de chaque sédiment d'analyses granulométriques.	69
5.2	Le nombre moyen des différents types de points d'intérêt détectés pour chaque classe de la base de données des textures sonars.	73
5.3	Comparaison des performances de reconnaissance de textures sonars pour différents types de points d'intérêt utilisés pour la construction du descripteur LGCM : DoG+Sift, FH+Surf, (Har-Lap)+(Sift-Spin), (Hes-Lap)+Daisy, FH+Brief. Le classifieur RF est considéré dans cette expérience.	75
5.4	Comparaison des performances du descripteur LGCM pour la classification supervisée des textures sonars pour différents classifieurs (k-NN, SVM, RF) et des différentes mesures de similarité (Euclidienne, χ^2 , divergence de Jensen-Shannon).	76
5.5	Comparaison des performances du descripteur LGCM pour la classification de textures sonars dans les cas où on utilise ou non les procédures d'adaptation d'échelle et de réduction la taille de descripteur.	76
5.6	Comparaison des performances de reconnaissance de textures sonars pour différentes fonctions de covariance de l'approche LGCM.	77
5.7	Taux moyens et écarts types de bonne classification de nos approches pour les images texturées sonars comparées avec des techniques d'état de l'art : Filtre de Gabor [6], Matrice de cooccurrence [7], sac de mots (SDM) [8], les statistiques de cooccurrence des paires des catégories de points d'intérêt [9], la description multifractale [10], la combinaison des différents points d'intérêt [11], statistiques descriptives spatiales de cooccurrence (SSC), modèles de Cox log-Gaussien (LGCM).	78

Résumé

Cette étude traite de l'utilisation des processus ponctuels spatiaux marqués pour définir de nouveaux descripteurs invariants pour la caractérisation et l'interprétation du contenu des images, soit sous forme non-paramétrique par l'emploi des statistiques descriptives des cooccurrences, soit sous forme paramétrique par modèle de Cox log-Gaussien de signatures locales. Chaque signature locale de l'image est définie et est caractérisée par un vecteur descripteur des points d'intérêt ou des formes principales. Les statistiques considérées reposent sur la construction d'un dictionnaire des caractéristiques des signatures élémentaires. Considérant l'ensemble des différentes catégories des signatures locales dans l'image comme une réalisation de processus spatiaux ponctuels marqués, nous étudions les modèles probabilistes pour ces nouveaux descripteurs du contenu visuel des images, qui héritent des propriétés d'invariance des descripteurs de points d'intérêts ou de formes. Les applications à la reconnaissance d'images réelles de textures et de scène utilisant les classifieurs discriminants sont présentées pour différents cas d'étude.

Dans le contexte de la caractérisation des images sonars des fonds marins, nous introduisons une nouvelle base de données de textures sonars correspondant à six types de fonds marins obtenus dans différentes conditions d'acquisition des images sonar, où l'on note à la fois des variations du contraste et des modifications géométriques des structures observées. Les méthodes proposées améliorent la classification comparative à des descriptions plus classiques (matrice de cooccurrence, ondelettes), qui ne possèdent pas ces propriétés d'invariance.

Mots clés : Télédétection acoustique, image sonar, processus ponctuels spatiaux, modèle Cox log-Gaussien, point d'intérêt, carte topographique, sac de mots, méthode d'apprentissage supervisée, classification de texture, reconnaissance de la scène.

Abstract

This thesis is concerned with the use of marked spatial point processes to define new invariant descriptors for the characterization and interpretation of image content, either by the descriptive statistics of cooccurrences, or by the estimation of log-Gaussian Cox models. We are particularly interested in the spatial statistical analysis of elementary objects in images. Here, each elementary object is characterized by the vector descriptor of key-points or of shapes. The implementations of proposed methods rely on the construction of a dictionary of the visual signatures of elementary objects. Viewing a set of different categories of elementary objects as a realization of spatial marked point process, we study the probability models to introduce new image quality descriptors for the invariant analysis of visual image content. The proposed descriptors inherit the invariance properties of the descriptors used, as invariance to changes in contrast and affine transformations.

We discuss invariance properties of the proposed models for texture and scene recognition applications. These experiments include a comparison of the performance reached by the proposed models w.r.t. the state-of-the-art methods. In the context of the characterization of sonar images of the seabed, we introduce a sonar seabed texture database comprising six texture classes obtained from different acquisition conditions of sonar images, where different variations of contrast and geometric changes of the observed structures are considered. Recognition of sonar textures by statistical spatial analysis of key-points reaches the best performance compared classical descriptions, as cooccurrence matrices or wavelet coefficients.

Keywords : Acoustic remote sensing, sonar image, spatial point processes, log-Gaussian Cox model, keypoint, topographic map, bag of keypoints, supervisor learning method, texture classification, scene recognition.

Introduction générale

La cartographie des habitats marins constitue un enjeu majeur pour une gestion intégrée des zones côtières et la définition d'états de référence des écosystèmes marins. Les besoins d'inventaires et de surveillance de l'évolution des habitats benthiques côtiers, en particulier de leur biodiversité, de même que la nécessité d'approfondir la compréhension du fonctionnement des écosystèmes benthiques soumis aux fortes pressions anthropiques dans la frange côtière et aux effets des changements climatiques, conduisent à la mise en place d'actions à différentes échelles d'observations spatiales et temporelles. A titre d'exemples, nous pouvons citer plusieurs projets en cours [12, 13, 14, 15] dont le but est l'étude des habitats côtiers en vue d'aboutir à une meilleure connaissance des fonds marins pour permettre une cartographie précise de la faune et la flore en zone littorale. Cette thématique de cartographie des habitats marins constitue ici le cadre applicatif de ce travail de thèse. Les avancées réalisées dans le domaine du traitement de signaux acoustiques, qui de nos jours permettent un balayage de très haute résolution de larges surfaces, ont joué un rôle majeur dans l'émergence du sonar à balayage latéral comme un dispositif privilégié de télédétection pour l'observation des fonds marins. En effet, l'imagerie sonar permet de détecter sur les petits fonds et les limites de peuplements remarquables, mais également de définir de nouveaux états de référence à propos de la structuration et de la répartition d'organismes benthiques denses, vivant parfois en colonies sur le fond, tels que les haploids, les maërls et les crépidules.

Ce travail, proposé dans le cadre d'une collaboration étroite entre le laboratoire CNRS lab-STICC (Laboratoire en Sciences et Techniques de l'Information, de la Communication et de la Connaissance) à Telecom Bretagne et les départements NSE (Navires et Systèmes Embarqués) et DYNECO (Dynamique de l'Environnement Côtier) de l'Ifremer (Institut français de recherche pour l'exploitation de la mer), vise à développer des solutions innovantes et performantes pour l'automatisation de l'étude cartographique des peuplements marins benthiques. La reconnaissance des peuplements marins benthiques, s'effectue au préalable en analysant les images sonars acquises par différents capteurs qui portent une information de texture en lien avec les variations de la rugosité du fond due à la présence d'un certain nombre d'individus (parfois plusieurs milliers individu/ m^2). L'analyse de l'image doit également se porter sur les variations de formes liées à la structuration spécifique de ces peuplements.

Pour répondre à ces objectifs, nous avons étudié des méthodes de caractérisation et d'interprétation du contenu des images sonars. Dans ce but, nous avons proposé

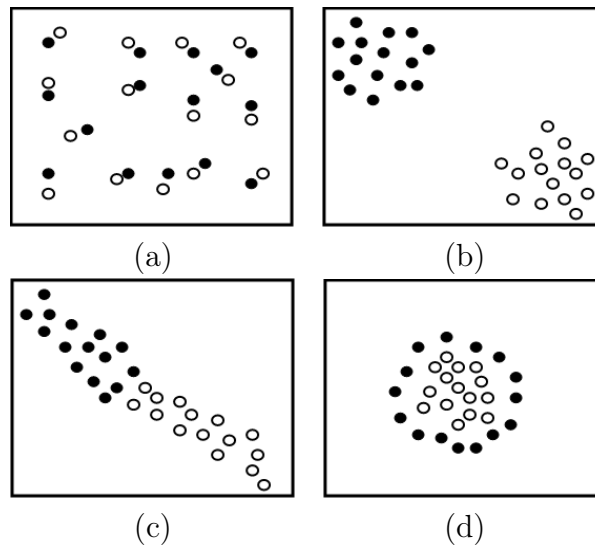


Figure 1 — Illustration de l'importance de la prise en compte de l'analyse statistique spatiale de la distribution de points d'intérêts dans une image texturée.

un nouveau descripteur invariant à des modifications des conditions de prise de vue, d'éclairement et de déformations géométriques (invariance aux transformations géométriques, photométriques, changement de contraste, etc. Les travaux initiaux dans le domaine du traitement d'images ont reposé sur des approches statistiques basées sur le calcul de distributions de réponses à des filtres, par exemple, matrices de co-occurrences et descripteurs associés [7, 16], descripteurs basés sur les transformées de Gabor et en ondelettes [6, 17], les décompositions multifractales [10], ou les cartes topographiques [18], etc. De récents travaux [19, 20, 21, 22] ont montré la pertinence des approches reposant sur l'extraction et la caractérisation de points d'intérêt. L'intérêt de ces approches réside notamment dans leurs propriétés d'invariance au changement de contraste et aux transformations affines des images. Ces nouvelles approches ont conduit à des améliorations significatives des performances de reconnaissance [11, 23]. Ces approches ont consisté à représenter une image par les statistiques d'un ensemble des points d'intérêt comme les descripteurs SIFT (Scale-invariant feature transform) [21] ou SURF (Speeded Up Robust Features) [19]. En particulier, la représentation par l'ensemble de mots visuels (bag of word) qui repose sur la définition d'un dictionnaire de points d'intérêt, a été proposée pour des applications de reconnaissance visuelles d'objets [8]. Toutefois, les statistiques considérées se limitent aux fréquences relatives des différents mots du dictionnaire, sans prise en compte de l'organisation spatiale des points d'intérêts. En réalité, la non prise en compte de l'information spatiale conduit à une perte importante d'information sachant que l'organisation de la texture est fondamentalement liée à la position spatiale des points d'intérêts. La Figure 1 illustre l'importance de l'analyse statistique spatiale de différentes distributions de points d'intérêt dans une image texturée. Ainsi, les approches récentes soulignent l'importance de la caractérisation de l'organisation spatiale des points d'intérêts et de leurs caractéristiques. Par exemple, [24, 25] ont proposés des descripteurs d'images qui peuvent hériter des propriétés d'invariance des signatures des points d'intérêts.

Dans ce travail, nous nous intéressons plus particulièrement à la caractérisation statistique des processus ponctuels spatiaux des signatures élémentaires dans l'image tels que les points d'intérêts, les objets et les formes pour des applications de classification et de reconnaissance d'images. Dans ce contexte, il faut également noter que la modélisation et caractérisation des processus ponctuels spatiaux ont connu des développements importants notamment en relation avec l'application dans des disciplines variées telles que l'épidémiologie [26, 27, 28], l'écologie, la biologie et l'exploitation forestière [29, 30, 31], la géographie [32], l'astronomie [33]. Récemment, ces techniques ont été introduites en vision par ordinateur pour résoudre des problèmes d'extraction de structures géométriques dans des images [34, 35]. Dans cette thèse, nous introduisons l'utilisation des processus ponctuels spatiaux multivariés pour la définition d'une représentation d'image reposant sur la caractérisation de l'organisation spatiale dans l'image des signatures élémentaires et de leurs caractéristiques. Nous allons exploiter les statistiques descriptives de cooccurrences et les modèles statistiques de processus spatiaux ponctuels marqués de Cox log-Gaussien. La définition de descripteurs proposés comprend principalement trois étapes :

- La détection d'un ensemble de signatures locales dans l'image et l'identification de leurs caractéristiques. Nous considérons ici plus particulièrement les points d'intérêts et les formes dans l'image.
- La construction d'un dictionnaire des signatures locales détectés qui consiste en l'application aux vecteurs de caractéristiques d'algorithmes de groupement de type de k-moyennes.
- L'estimation paramétrique et non-paramétrique de statistiques descriptives de cooccurrences ou des modèles statistiques de processus spatiaux ponctuels marqués de Cox log-Gaussien pour créer un descripteur d'images.

Les descripteurs proposés héritent des propriétés d'invariance des signatures locales des points d'intérêts et des formes, plus précisément de l'invariance aux transformations géométriques, aux changements de contraste et à l'occlusion partielle des objets. Des applications à la reconnaissance d'images basées sur des ensembles de classifieurs discriminatifs comme les k-plus-proches-voisins, les machines à vecteurs de support, les forêts aléatoires, sont présentées sur différentes bases de données.

Ce document décrit tour à tour ces différents aspects de nos travaux concernant la caractérisation et l'interprétation du contenu des images par l'analyse statistique spatiale des signatures élémentaires dans l'image et de leurs caractéristiques et ses applications à la classification et reconnaissance des images texturées (optiques et sonars) et des images optiques de scènes réelles. Il est organisé en deux parties de la manière suivante :

- Dans la première partie, nous présentons les méthodes que nous avons développées pour obtenir des descripteurs avec de bonnes propriétés d'invariance pour l'analyse du contenu visuel d'images basée sur la réalisation des processus ponctuels spatiaux multivariés de signatures locales dans les images. Tout d'abord, dans le Chapitre 1 nous décrivons les méthodes de détection de signatures locales dans une image et du calcul de leurs vecteurs de caractéristiques. Ensuite, nous introduisons différentes formes d'analyses statistiques spatiales de points marqués : les statistiques descriptives de cooccurrences sont présentées dans le

Chapitre 2 et les modèles statistiques de Cox log-Gaussien dans le Chapitre 3. Enfin, dans le Chapitre 4, nous exploitons ces méthodes dans des problèmes de classification et de reconnaissance d'images texturées. Trois bases de données sont considérées : les bases UIUCTex, Brodatz et KTH-Tips et des images optiques de scènes réelles.

- Dans la deuxième partie, nous présentons le problème de la dépendance angulaire des attributs de texture sonar et de la rétro-diffusion des fonds marins et nous décrivons notre base de données des images texturées sonars pour la classification des images sonars des fonds marin. Nous évoquons également les contributions des nos approches proposées par les comparaisons des performances de la classification des images sonars avec les techniques de l'état de l'art.
- La conclusion générale présente une synthèse de cette étude et quelques perspectives liées à ces travaux.

Publications

Les travaux présentés dans ce travail ont donné lieu aux publications suivantes :

- *Revues :*

1. Huu Giao Nguyen, R. Fablet, A. Ehrhold, J.-M. Boucher. "Keypoint-based analysis of sonar images : application to seabed recognition." IEEE Transaction on Geoscience and Remote Sensing, *TGRS*, à paraître.

- *Congrès internationaux :*

1. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Multivariate log-Gaussian Cox models of elementary shapes for recognizing natural scene categories". IEEE International Conference on Image Processing, *ICIP'2011*, Brussels, Belgium, Septembre 2011.
2. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Visual textures as realizations of multivariate log-Gaussian Cox processes." IEEE Conf. on Computer Vision and Pattern Recognition, *CVPR'2011*, Colorado Springs, US, Juin 2011, pp.2945-2952.
3. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Log Gaussian Cox Processes of visual keypoints for sonar texture recognition." Proc. 36th Int. Conf. on Acoustics, Speech and Signal Processing, *ICASSP'2011*, Prague, Czech Republic, Mai 2011, pp.1005-1008.
4. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Spatial statistics of visual keypoints for texture recognition." Proc. 11th European Conference on Computer Vision, *ECCV'2010*, Crete-Greece, Septembre 2010, Vol.6314, pp.764-777.
5. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Invariant descriptors of sonar textures from spatial statistics of local features." Proc. 35th Int. Conf. on Acoustics, Speech and Signal Processing, *ICASSP'2010*, Dallas-US, Mars 2010, pp. 1674-1677
6. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Invariant texture descriptors based on spatial statistics.", Workshop on Advances in Seafloor-Mapping Sonar 2009, Brest-France.

- *Congrès nationaux :*

1. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Statistiques spatiales de points d'intérêt pour la reconnaissance invariante de textures.", 5ème Congrès Fran-

cophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, *RFIA'2010*, Caen-France, Janvier 2010.

• *Séminaires :*

1. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Automatic cartography of marine benthos populations from multi-frequency sonar images and bathymetry data." Journée des projets Europôle Mer à l'Aber Wrac'h-Brest, 19 Novembre 2009.
2. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Invariant texture descriptors based on spatial statistics." Séminaires Lab-STICC à Telecom Bretagne, 28 Janvier 2010.
3. Huu Giao Nguyen, R. Fablet, J.-M. Boucher. "Log Gaussian Cox Processes of visual keypoints for sonar texture recognition." Assises Sciences et Techniques de la Mer à l'ENSTA Bretagne, 16 et 17 mai 2011.

Première partie

Caractérisation et modélisation statistique d'ensemble de signatures locales dans les images

Etat de l'art sur les descripteurs locaux d'images

1.1 Introduction

En vision par ordinateur, l'extraction de caractéristiques visuelles des images est une étape préliminaire pour de nombreux algorithmes de caractérisation et d'interprétation du contenu des images [36, 37, 19, 38]. Cette étape examine tous les pixels de l'image pour le but d'extraire certaines propriétés visuelles de l'image. Les caractéristiques visuelles permettent de décrire le contenu des images et de les associer avec la sémantique de l'image. Dans ce contexte, plusieurs méthodes d'extraction de caractéristiques structurelles visuelles ont été proposées [36, 37]. Ce chapitre décrit les connaissances actuelles sur les représentations locales des images par un ensemble de signatures élémentaires tel que chaque élément caractérise une information pertinente à propos du contenu visuel de l'image. L'approche locale offre un bon compromis avec une extraction de régions d'intérêt caractéristiques de l'image et une description locale stable et robuste. Dans ce travail, les signatures élémentaires considérées sont les points d'intérêt et les formes. Les méthodes de représentation locale d'image se focalisent sur des zones perceptuellement importantes de l'image ce qui permet de lutter contre des changements de contraste ou des variations de transformations géométriques et photométriques de l'image. Un vecteur signature est calculé par rapport à un certain nombre de critères dans chaque zone. L'intérêt de ce vecteur signature réside notamment dans ses propriétés de robustesse et dans son caractère distinctif ce qui permettra une correspondance entre deux signatures élémentaires.

Ce chapitre est organisé comme suit : nous commençons par l'étude sur les algorithmes de détection de zones d'intérêts pour l'obtention des points d'intérêt. Puis, différentes méthodes pour calculer un vecteur descripteur pour chaque point d'intérêt sont présentées. Ensuite, nous décrivons les méthodes de représentation d'image par les formes. En particulier, la méthode de carte topographique pour l'extraction d'un arbre des formes dans l'image sera détaillée. Enfin, nous décrirons notre méthode de sélection pour une description invariante des formes : le contexte de contour.

1.2 Points d'intérêt

1.2.1 Détection de points d'intérêt

Dans les applications d'analyse et de reconnaissance d'images, la détection de zones d'intérêt d'une image numérique tient une place essentielle. Elle consiste à mettre en évidence des zones qui ont des propriétés locales remarquables et invariantes aux conditions d'acquisition des images, par exemple : changement de contraste, transformations géométriques et photométriques [36, 37, 38]. Les algorithmes de détection de zones d'intérêt se focalisent en général sur des points, des courbes ou des régions particulières sélectionnées selon un critère précis pour constituer des descripteurs robustes.

- Parmi les nombreuses méthodes existantes dans la littérature, le détecteur de Harris [39] détermine les changements moyens de l'intensité de l'image dans un voisinage des points considérés. La matrice de Harris caractérisant les variations locales des courbures principales associées aux variations moyennes d'intensité comme suit :

$$M(x, y) = \sum_{u,v} G(u, v, \sigma) \begin{bmatrix} \left(\frac{\partial I(x,y)}{\partial x} \right)^2 & \frac{\partial I(x,y)}{\partial x} \frac{\partial I(x,y)}{\partial y} \\ \frac{\partial I(x,y)}{\partial y} \frac{\partial I(x,y)}{\partial x} & \left(\frac{\partial I(x,y)}{\partial y} \right)^2 \end{bmatrix} \quad (1.1)$$

où $I(x, y)$ est l'intensité de l'image I au pixel (x, y) . La fonction $G(u, v, \sigma)$ est une gaussienne isotrope et circulaire, définie par $G(u, v, \sigma) = \frac{1}{2\pi\sigma^2} \exp(-\frac{u^2+v^2}{2\sigma^2})$. Les valeurs propres positives de cette matrice permettent à caractériser un coin définie comme un pixel où l'intensité varie fortement dans toutes les directions. Le détecteur Harris est invariant aux rotations et changement de contraste, mais n'est pas invariant au changement d'échelle.

De même manière du calcul, la matrice hessienne H d'une image est donnée avec ses dérivées partielles secondes par :

$$H(x, y) = \begin{bmatrix} \frac{\partial^2 I(x,y)}{\partial x^2} & \frac{\partial I(x,y)}{\partial x} \frac{\partial I(x,y)}{\partial y} \\ \frac{\partial I(x,y)}{\partial y} \frac{\partial I(x,y)}{\partial x} & \frac{\partial^2 I(x,y)}{\partial y^2} \end{bmatrix} \quad (1.2)$$

- Contrairement à la méthode de Harris, la plupart des autres techniques de détection de points d'intérêt sont basées sur un espace d'échelles. Nous pouvons citer les méthodes basées sur l'analyse des différences de gaussiennes (difference of gaussians, DoG), laplacien de gaussien (Laplacian of Gaussian, LoG) [40]. Ces détecteurs sont basés sur le calcul de différents niveaux de la représentation échelle-espace d'image I par la convolution avec un filtre gaussien G de paramètre σ , soit :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1.3)$$

Considérant simultanément les minima/maxima locaux au pixel (x, y) avec ses 26 voisins dans le voisinage $3 \times 3 \times 3$ de la pyramide d'image (Fig.1.1) , le détecteur LoG consiste à détecter les extrêmes locaux de la norme de l'opérateur de laplacien défini comme suit :

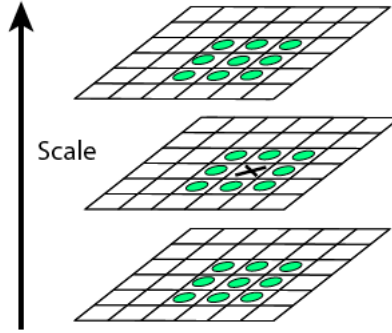


Figure 1.1 — Illustration de l'estimation d'un pixel local extrême avec ses 26 voisins dans un espace d'échelles.

$$\nabla_{norm}^2 L(x, y, \sigma) = \sigma \left(\frac{\partial^2}{\partial x^2} L(x, y, \sigma) + \frac{\partial^2}{\partial y^2} L(x, y, \sigma) \right) \quad (1.4)$$

De la manière, le DoG consiste en les opérateurs laplaciens de différents niveaux $(\sigma, k\sigma)$ qui est défini par :

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (1.5)$$

En respectant les espaces en échelle, ces méthodes permettent d'assurer l'invariance aux rotations, translations et changements d'échelle. Plus récemment, des combinaisons de ces méthodes ont été introduites pour renforcer les propriétés d'invariance du détecteur. Par exemple, le détecteur Harris-Laplacien (noté Har-Lap) [41] utilise la mesure de Harris et un filtrage gaussien. La matrice de Harris de l'Eq.1.1 est re-définie par :

$$M(x, y, \sigma_I, \sigma_D) = \sigma_D^2 G(x, y, \sigma_I) * \begin{bmatrix} \left(\frac{\partial L(x, y, \sigma_D)}{\partial x} \right)^2 & \frac{\partial L(x, y, \sigma_D)}{\partial x} \frac{\partial L(x, y, \sigma_D)}{\partial y} \\ \frac{\partial L(x, y, \sigma_D)}{\partial y} \frac{\partial L(x, y, \sigma_D)}{\partial x} & \left(\frac{\partial L(x, y, \sigma_D)}{\partial y} \right)^2 \end{bmatrix} \quad (1.6)$$

où σ_I, σ_D sont les facteurs d'échelle. Il peut s'agir des approches basées sur la matrice hessienne qui associent à l'opérateur Laplacien (Hessienne-Laplacien, noté par Hes-Lap) [41] ou aux échelles filtres dans les différentes fenêtres (Fast-Hessien, FH) [19]. De plus, la transformation affine est utilisée dans la littérature pour identifier les régions des images invariantes aux transformations affines par l'utilisation de la matrice de Harris (Harris-Affine) ou la matrice de Hessienne (Hessienne-Affine) [37].

- Il faut également souligner les autres catégories de détecteur dans la littérature, par exemple, le détecteur EBR (edge-based regions) qui exploite les bords des structures présentes dans l'image. Le détecteur IBR (intensity extrema-based regions), le détecteur MSER (maximally stable extremal region) et le détecteur SR (Salient region) qui évaluent des critères basés sur l'intensité à l'intérieur des régions en différents niveaux d'échelle [37].



(a) Détecteur de Harris : 323 points



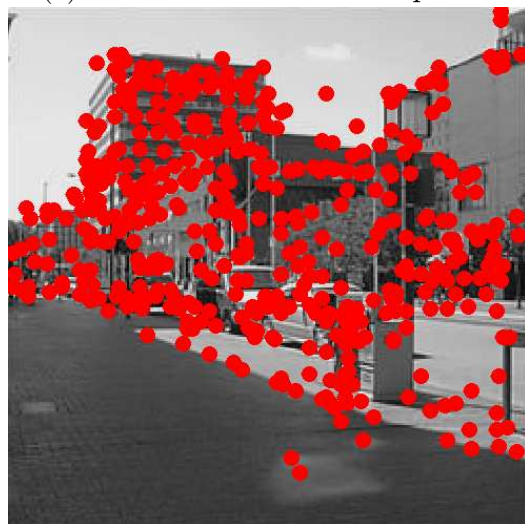
(b) Détecteur Har-Lap : 255 points



(c) Détecteur Har-Aff : 249 points



(d) Détecteur Hes-Lap : 427 points



(e) Détecteur DoG : 558 points



(f) Détecteur FH : 196 points

Figure 1.2 — Exemples de points d'intérêt détectés dans une image par les différentes approches de détecteurs décrites dans la Section 1.2.1.

Des comparaisons de performances entre les détecteurs de points d'intérêt sont présentées dans [36, 37, 19, 38]. Il est important de noter qu'il est difficile de choisir un algorithme de détecteur de points d'intérêts qui soit universellement meilleur que les autres dans une tâche de classification pour tous les types de conditions d'acquisition des images. Pourtant, les méthodes : Har-Lap, Hes-Lap, DoG et FH fournissent plus de régions que les autres détecteurs. La Figure 1.2 présente quelques exemples de calcul des différents détecteurs. C'est un critère important pour les approches basées sur les statistiques spatiales des signatures locales dans les images. Ces quatre méthodes de détection de point d'intérêt seront donc considérées dans cette thèse.

1.2.2 Descriptions invariantes de points d'intérêt

La description invariante d'image consiste en des transformations mathématiques permettant de caractériser et d'interpréter le contenu des images. C'est un enjeu majeur pour de nombreux domaines de la vision : reconnaissance des formes [9, 42, 3], reconnaissance des textures [36, 43, 10] et reconnaissance des scènes [44, 5], localisation de robot [24], reconstruction [45], etc. On recherche notamment des descripteurs invariants aux conditions d'acquisition des images, par exemple : changement de contraste, transformations géométriques, pour l'amélioration de la robustesse de la caractérisation des images. Les travaux initiaux reposent sur des approches basées sur des moments (Hu [46], Zernike [47]), des transformées intégrales (Fourier-Mellin [48], Gabor et ondelettes [17, 6]) ou sur le calcul de distributions de réponses à des filtres (matrice de cooccurrences [7]).

Des travaux récents [23] ont montré la pertinence des approches reposant sur l'extraction et la caractérisation de points d'intérêt. L'intérêt de ces approches réside notamment dans leurs propriétés d'invariance au changement de contraste et aux transformations affines des images. Etant donnée un l'ensemble des zones d'intérêt, différentes solutions ont été proposées pour calculer en chacune de ces zones un vecteur caractéristique [19, 49]. En général, ces mesures sont effectuées autour des points d'intérêt dans un domaine d'exploration, par exemple : rectangulaire, circulaire qui représente une région d'intérêt 1.4. Une amélioration significative des performances de reconnaissance a été obtenue [23, 11].

Parmi les nombreuses méthodes de description locale de points d'intérêt proposées, le descripteur SIFT (Scale-invariant feature transform) est le plus populaire. Il fournit de bonnes performances et se révèle stable et efficace pour la classification d'images [21, 23]. La première étape de l'algorithme est la détection des points d'intérêt. Dans cette thèse, nous utilisons le détecteur DoG. Chaque point d'intérêt est caractérisé par la distribution des orientations du gradient de l'intensité dans 16 quadrants 4x4 autour du point considéré (Figure 1.3). Les orientations sont quantifiées suivantes 8 valeurs. Le vecteur descripteur SIFT est finalement de taille $4 \times 4 \times 8 = 128$.

Dans la littérature, plusieurs extensions du descripteurs SIFT visant l'amélioration de ses propriétés de robustesse et de discrimination ont été proposées, par exemple : ACP-SIFT [50], "steerable" filtrages [51], moments invariants [52], filtrages complexes [53]. La Figure 1.4 fournit quelques exemples de la caractérisation de points d'intérêt

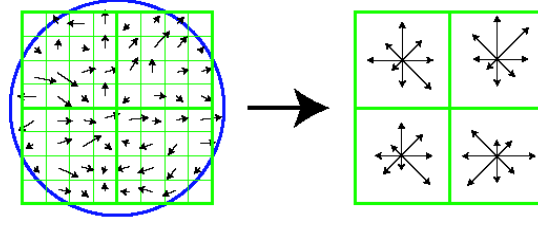


Figure 1.3 — Illustration de la construction d'un descripteur SIFT. La distribution des orientations du gradient local dans 16 quadrants 4x4 est présentée dans la figure de gauche et d'un cercle représentant la gaussienne utilisée pour le lissage. La figure de droite schématise la forme du descripteur sous la forme d'histogrammes d'orientation de gradient suivant 8 orientations.

qui sont similaires au concept de SIFT. Le SURF (Speeded Up Robust Features) [19] possède un domaine d'exploration rectangulaire. On peut citer également, le GLOH [23] et le CS-LBP [49], utilisant tous deux une géométrie circulaire pour leur masque d'analyse.

Pour construire le descripteur SURF (Speeded Up Robust Features) est initialement basé sur le détecteur Fast-Hessien pour détecter les points d'intérêt [19]. Ensuite, on calcule les réponses d'ondelettes de Haar dans les directions horizontale d_x et verticale d_y dans un cercle de rayon R centrée au point d'intérêt. Le problème d'invariance en échelle est géré par le choix de la valeur de R . La distribution des réponses d'ondelettes de Haar sous la forme $(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$ est estimée dans les 16 quadrants 4x4 autour du point considéré comme le SIFT. La signature locale est de dimension $4 \times 4 \times 4 = 64$ ce qui permet de réduire le temps de calcul des caractéristiques et de recherche des correspondances. Le descripteur SURF est un détecteur de points d'intérêts invariant aux changements d'échelle et aux rotations.

Comme celle de descripteur SIFT, le descripteur GLOH (Gradient Location and Orientation Histogram) [23] est aussi basé sur les points d'intérêt du détecteur DoG. Cependant, la caractérisation de son vecteur descripteur consiste en une grille de positions log-polaires avec trois boîtes en direction radiale (de rayons 6, 11 et 15) et huit en direction angulaire, soit 17 boîtes (bin en anglais) de position (Figure 1.4.c). La boîte centrale n'est pas divisée dans les directions angulaires. Les orientations des gradients sont quantifiées en 16 boîtes, d'où 272 histogrammes de boîtes. La taille du vecteur descripteur GLOH est ensuite réduite à 128 dimensions par l'algorithme ACP (l'analyse en composantes principales).

Plus récemment, de nouvelles méthodes d'analyse et de mesure des distributions spatiales d'intensités au point considéré pour caractériser de nouveaux descripteurs locaux qui sont à la fois robustes et discriminants ont été proposées. On peut citer les descripteurs suivants :

- Le descripteur GIH (histogramme de l'intensité géodésique) [4] est basée sur une déformation invariante. Une distribution conjointe distance géodésique et l'intensité des points d'intérêt forme le vecteur descripteur GIH. La taille de vecteur descripteur est $K \times N$ dimensions, où K est le nombre d'intervalles d'intensités et N est le nombre d'intervalles de distances géodésiques.

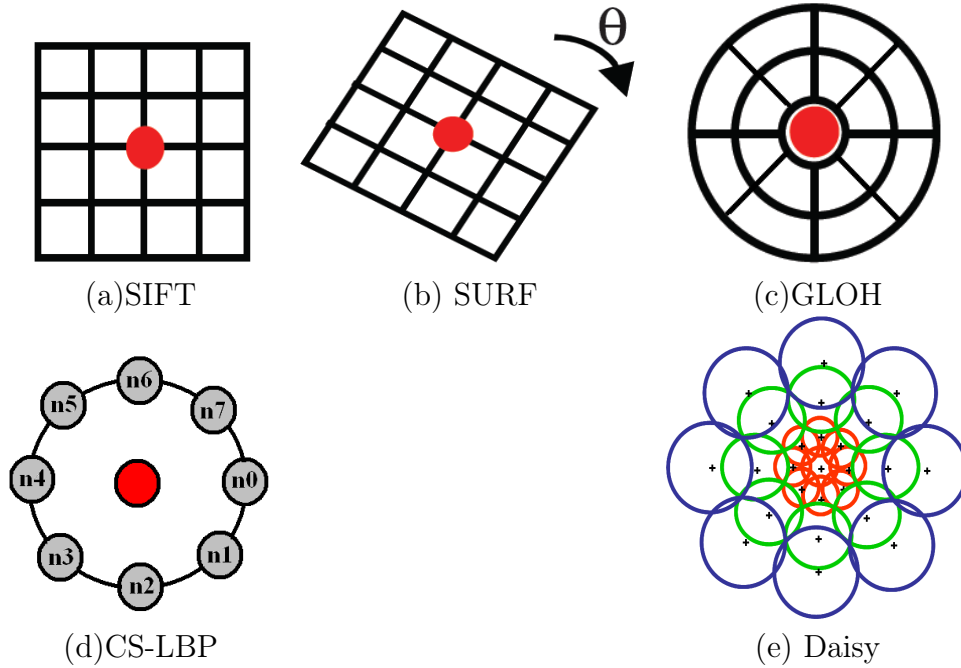


Figure 1.4 — Illustration de la caractérisation de descripteurs locaux effectuée autour du point d'intérêt par différentes manières d'exploration : rectangulaire (a,b) ou circulaire (c,d,e) [1].

- Le descripteur CS-LBP (center-symmetric local binary pattern) [49] calcule un vecteur descripteur local de points d'intérêt par l'utilisation d'une caractérisation des motifs binaires locaux (Figure 1.4.d) dans les cercles de rayon R . Le vecteur signature de CS-LBP est en $K^2 \times 2^{N/2}$ dimensions, où K^2 est la taille de grille considérée et N est le nombre de pixels dans le cercle de voisinage.
- Le descripteur DAISY [22] est constitué de cercles disposés suivant différentes échelles et orientés suivant la direction du gradient du point d'intérêt. La Figure 1.4.e présente une illustration de calcul de ce descripteur DAISY dans 25 voisinages autour du point d'intérêt avec 8 orientations. La taille du vecteur descripteur est un histogramme de 200 dimensions.
- Le descripteur BRIEF (Binary Robust Independent Elementary Features) [20] est un descripteur binaire défini par les comparaisons entre deux valeurs d'intensité dans un ensemble de N couples de tests. La taille du vecteur descripteur BRIEF est $N/8$. En conséquence, cette méthode est très efficace pour calculer et appliquer dans les applications temps-réel.

1.3 Formes

1.3.1 Détection de formes

L'identification et la détection de structures locales est une opération permettant de rassembler des pixels suivants des critères prédéfinis (couleur, texture ou intensité)

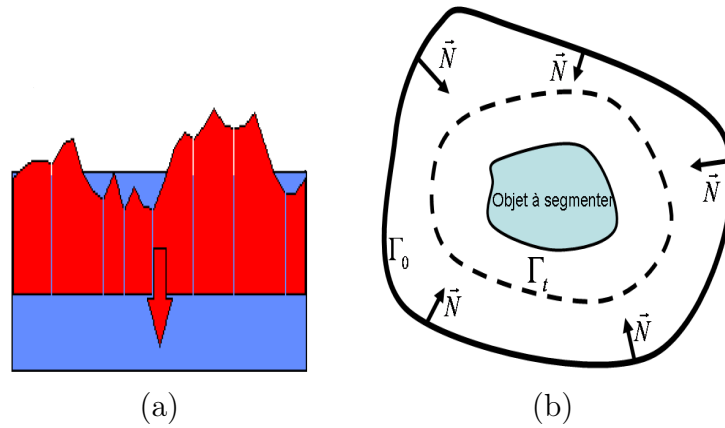


Figure 1.5 — Illustration des principes de la ligne de partage des eaux (a) et des évolutions du contour actif (b).

pour trouver les régions homogènes et leurs contours dans l'image. L'objectif principal est de simplifier la représentation de l'image par des objets élémentaires qui sont plus significatifs et plus faciles à analyser. L'identification de structures locales (ou la segmentation d'image) tient une place primordiale et essentielle dans de nombreuses applications en traitement d'image, par exemple : reconnaissance de visage [54], identification biométrique [55], traitement d'image médicale et satellite [56, 34], etc. La segmentation est un vaste sujet de recherche dans le domaine de vision par ordinateur. A ce titre les représentations des images reposant sur une segmentation spatiale en région peuvent présenter un intérêt. On peut grouper les méthodes connues en différentes catégories selon l'approche qui les régit :

- Les méthodes basées sur la détection du contour et des structures considèrent des variations d'intensité d'image pour déterminer les frontières séparant des régions de niveaux de gris différents et relativement homogènes, ou bien des régions de différents textures. Différentes solutions ont été proposées pour cette catégorie. Les opérateurs de Sobel [57], Canny [58], Shen [59] donnent les estimations de la dérivée directionnelle pour obtenir à la fois une bonne détection et une bonne localisation du contour. L'algorithme de la ligne de partage des eaux (watershed) [60] considère une image à niveaux de gris comme un relief topographique, dont on simule l'inondation (Figure 1.5a). Plus récemment, les modèles actifs ont été introduits. Le principe de ces approches par formes paramétriques est de faire évoluer un contour ou une surface vers les frontières de l'objet que l'on cherche à segmenter (Figure 1.5b), par exemple : contours actifs (snakes) [61], level-set géodésique [62], modèles actifs de forme et d'apparence (Active Shape (Appearance) models - ASM or AAM) [63, 64], etc.

- Il faut également souligner les méthodes basées sur des techniques de classification qui partitionnent les images en plusieurs classes d'ensembles connexes de pixels. On peut citer les approches de segmentation d'image basées sur les champs aléatoires Markoviens [65], sur les machines à vecteurs de support [66], sur les réseaux de neurones [67] (les classifications supervisées paramétriques et non-paramétriques) ou sur les algorithmes de k-moyennes, c-moyennes floues adaptatives [68] (les classifications non-supervisées).

• Il faut aussi remarquer les méthodes basées sur la théorie des graphes pour analyser les relations entre les pixels dans une image : Hypergraphes [69], Graph Cuts [70, 71]. Ces méthodes reposent sur l'exploitation des relations entre les noeuds de graphe pour résoudre le problème de minimisation de la fonction d'énergie entre les sommets ou la partition multi-échelle de l'hypergraphe de voisinage.

Néanmoins, ces méthodes de segmentation fournissent une liste de courbes non-fermées dont l'extraction des descriptions visuelles de l'image peut être complexe alors que nous nous intéressons ici les méthodes de segmentation en région qui nous fournissent une liste de formes (courbes fermées). Par rapport aux différentes techniques de segmentation région, il faut mettre en avant l'intérêt des ensembles de niveaux comme représentation complète de l'image (i.e., on peut passer des ensembles de niveaux à l'image et inversement).

La carte topographique

Dans ce travail, la représentation d'image par des formes invariantes élémentaires est ciblée, notamment le changement de contraste. Basée sur la morphologie mathématique, une représentation complète et invariante par changement de contraste de l'image a été proposée par Caselles et al. [18] en considérant les lignes de niveau sous-jacentes à une carte topographique. La définition de la carte topographique et de ses principales propriétés est rappelée dans cette section [72].

Une image (en niveaux de gris) est représentée par une fonction u d'un espace topologique connexe des représentations morphologiques X . Ces représentations topologiques reposent sur une segmentation de l'image comprise comme une partition en régions connexes. Les ensembles de niveau sont les éléments de base de la représentation morphologique. Elles sont hiérarchiques. On ne traite pas d'ensembles isoniveaux, mais de composantes connexes d'ensembles de différents niveaux, dont les trous sont bouchés. Cela permet de prendre en compte occlusions et transparences dans l'image. Le fait de boucher les trous permet de mélanger les ensembles de niveaux supérieurs et inférieurs dans un arbre de Kronrod [73], qui est orienté par l'inclusion. Cet arbre décrit la topologie de l'image (Figure 1.6) .

L'ensemble de niveau supérieur $\chi^{\geq \lambda}(u)$ de valeur $\lambda \in \mathbb{R}$ et l'ensemble de niveau inférieur $\chi_{\leq \mu}(u)$ de valeur $\mu \in \mathbb{R}$ sous-ensembles de X sont définis par :

$$\begin{cases} \chi^{\geq \lambda}(u) = \{x \in X, u(x) \geq \lambda\} \\ \chi_{\leq \mu}(u) = \{x \in X, u(x) \leq \mu\} \end{cases} \quad (1.7)$$

Il faut également noter la relation entre les ensembles inférieurs et les ensembles supérieurs : $X \setminus \chi^{\geq \lambda}(u) = \chi_{< \lambda}(u)$.

Une représentation d'image est complètement décrite par son ensemble de niveau supérieur ou inférieur (Figure 1.7). Des ensembles de niveau supérieur d'une image u , on peut retrouver u par la formule :

$$\forall x, u(x) = \sup\{\lambda, x \in \chi^{\geq \lambda}(u)\} \quad (1.8)$$

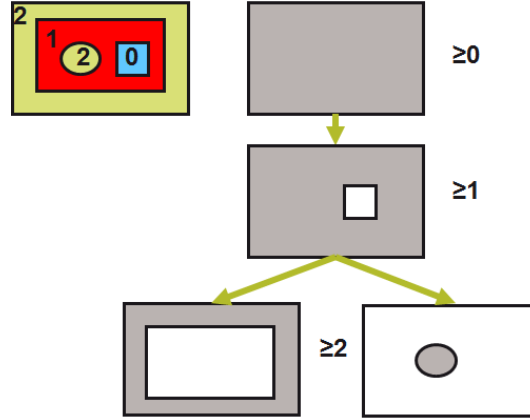


Figure 1.6 — Arbre de composantes connexes des ensembles de niveaux supérieurs dans une image simple.

et des ensembles de niveau inférieur par la formule :

$$\forall x, u(x) = \inf\{\mu, x \in \chi_{\leq \mu}(u)\} \quad (1.9)$$

La propriété importante de la carte topographique est l'invariance par changement de contraste global. On considère la fonction v qui se déduit de u par un changement de contraste g si $v = g(u)$. Cela signifie qu'il existe une fonction g strictement croissante telle que $v = g(u)$, les caractéristiques de v et u doivent être les mêmes. La famille des ensembles de niveau est une caractéristique morphologique de l'image car :

$$\begin{cases} \chi^{\geq g(\lambda)}(v) = \chi^{\geq g(\lambda)}(g(u)) & = \chi^{\geq \lambda}(u) \\ \chi_{< g(\lambda)}(v) = \chi_{< g(\lambda)}(g(u)) & = \chi_{< \lambda}(u) \end{cases} \quad (1.10)$$

Les formes d'une image sont construites à partir des composantes connexes d'ensembles de niveau. L'algorithme FLST (Fast Level Set Transform) pour extraire les formes d'une image en niveaux de gris et les représenter dans un arbre des formes est implémentée dans Megawave2¹. Les formes extraites par cet algorithme sont fermées et n'ont pas de trous. Les propriétés d'invariance par changement de contraste et par occultation partielle d'objet sont respectées.

1.3.2 Description invariante de formes

Après avoir détecté un ensemble de formes dans une image, l'un des principaux problèmes pour la reconnaissance d'image automatique consiste à trouver des schémas convenables de description de formes qui satisfont les propriétés d'invariance aux translations, rotations et changements d'échelle. Ces schémas doivent être capables de produire un vecteur de descripteurs pour chaque forme qui soit robuste aux variations pouvant affecter les éléments d'une même classe.

1. <http://megawave.cmla.ens-cachan.fr/>

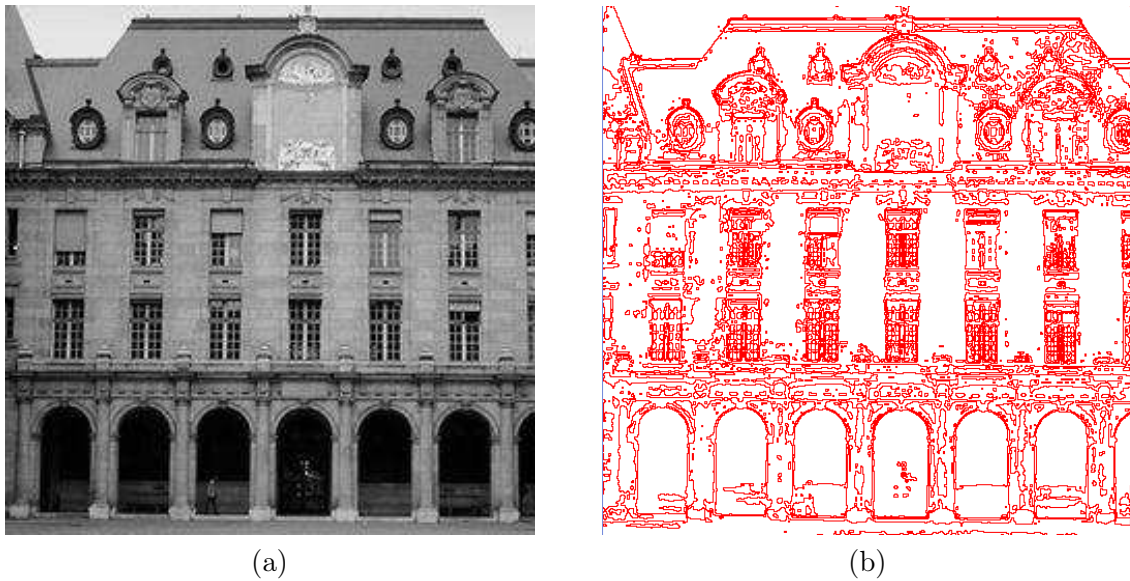


Figure 1.7 — Illustration de la représentation d’une image par les lignes de niveau :
 (a) image origine, (b) les formes détectées dans un niveau de gris.

De nombreuses méthodes de description de formes ont été proposées dans la littérature. Il peut s’agir de mesurer la similarité entre les formes par l’estimation des distributions des jonctions, des segments ou des points dans les contours des objets [74, 75]. Ces méthodes ont été appliquées avec succès dans des domaines tels que l’analyse de visage, du mouvement, etc [76, 77]. Une autre classe de techniques de description de formes repose sur la construction de squelette de formes dans un arbre ou un graphe. La méthode la plus connue de cette technique est le *shock-graph* [2, 78] qui est robuste pour décrire les articulations et les occlusions de formes. Etant donné une forme et son contour, les *shocks* sont définis par les composantes singulières du processus d’évolution de courbe pour extraire le squelette de forme. Le *shock-graph* est une organisation hiérarchique de shocks par un arbre acyclique (Figure 1.8). Les différentes méthodes pour mesurer des critères de shock-graph permettent de définir un vecteur descripteur de forme et mettent en correspondance deux formes distinctes [79, 76]. Néanmoins, les approches basées sur la théorie des graphes ont aussi des difficultés à caractériser les objets avec des contours bruités. Cette méthode ne sera pas considérée dans ce travail.

Il faut aussi remarquer les descriptions des formes qui se focalisent en général sur le calcul des contraintes de points sur le contour (shape landmark en anglais), par exemple : les transformations rigides [80], les déformations non-rigides [81] ou les autres contraintes spécifiques sur les landmarks [82, 83, 84]. Ici, nous nous intéressons à l’algorithme basé sur le *contexte de contour* (shape context) [3]. Comme celle du descripteur GLOH de points d’intérêt, une signature polaire d’une forme est donnée par l’estimation de la distribution grossière de la forme (la distance et l’orientation) par rapport à un point de référence sur le contour. La distribution est représentée par un histogramme polaire des points les plus proches dans le contour de la forme autour du point de référence. Le descripteur *contexte de contour* est l’ensemble des signatures polaires de la forme par rapport à chacun des points sur le contour de forme. La Figure 1.9a illustre le calcul d’un histogramme polaire. Pour un point x_i sur le contour de forme, un

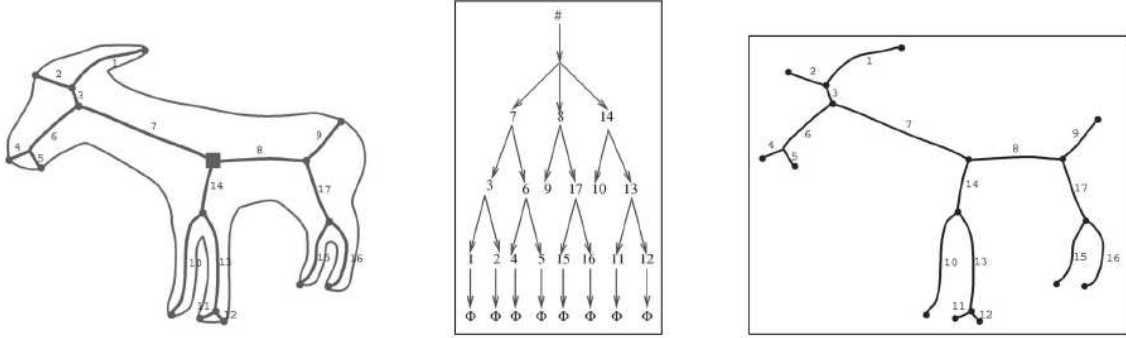


Figure 1.8 — Exemple de représentation de shockGraph. La référence de la figure est dans le papier original [2].

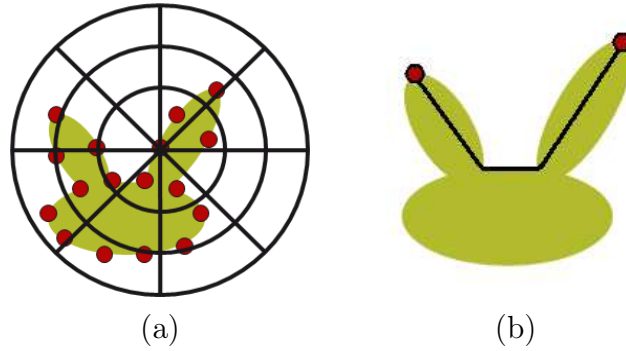


Figure 1.9 — Le calcul de l'histogramme polaire sur un point de contour de forme (a). La distance minimale entre deux points dans l'espace intérieur de forme (b). Les figures de référence sont dans les papiers originaux [3, 4]

histogramme polaire est donné, en utilisant les coordonnées relatives des autres points dans le contour, par :

$$h_i(k) = E_{j \neq i} \{x_j : x_j - x_i \in \text{bin}(k)\} \quad (1.11)$$

Notons que dans cette expression, E dénote l'espérance des points, k dénote un *bin* sur un secteur radial dans l'espace polaire ou log-polaire.

Plus récemment, Ling et al. [4] ont proposé une nouvelle approche, appelée *inner-distance shape context* (IDSC), qui est basée sur le calcul d'un histogramme polaire de *contexte de contour* et de la distance géodésique entre deux points dans un plan intérieur de forme (comme la construction du squelette de *shock graph*). Cette distance géodésique (inner-distance) est définie par la distance minimale entre deux points dans une forme (Figure 1.9b). L'utilisation de la distance et l'orientation assure l'invariance de la méthode aux translation et changement d'échelle. Par ailleurs, cette méthode est aussi robuste pour décrire les articulations et les occlusions de formes. Par contre, le IDSC n'est pas invariant à des transformations affines arbitraires mais l'utilisation de bins log-polaire assure une certaine invariance à des petites distorsions locales.

1.4 Conclusion

Dans ce chapitre nous avons décrit différentes alternatives de représentations locales d'images qui seront exploitées ensuite pour modéliser et caractériser le contenu des images basé sur la théorie des statistiques spatiales de points. Deux types de signatures élémentaires, points d'intérêt et formes, sont exploités en deux étapes : la détection des régions d'intérêt et la caractérisation des régions d'intérêt pour définir un vecteur signature. Les propriétés d'invariance aux conditions d'acquisition des images sont discutées, par exemple : changement de contraste, transformations géométriques et photométriques, occlusions, etc. Une image est représentée par un ensemble de signatures élémentaires. Chaque signatures élémentaire est représentée par un histogramme de descripteur et une position locale dans l'image. En réalité, le semis des signatures élémentaires dans une image est lié au contenu visuel de l'image, par exemple : l'organisation d'objets, texture. Dans le chapitre suivant, nous exploitons la caractérisation de l'organisation spatiale des signatures élémentaires et de leurs caractéristiques pour d'introduire les nouveaux descripteurs modélisés comme la réalisation d'un processus spatial ponctuel marqué.

2.1 Introduction

Ce chapitre focalise sur la définition d'une description statistique spatiale de points dans une image pour caractériser son contenu. Cette description pourra servir dans une tâche de classification d'image. Conjointement à l'emploi de descripteurs de points d'intérêt ou de formes, comme cela a été commenté dans le chapitre précédent (Chapitre 1) : il s'agit alors de processus ponctuels marqués. Dans le cas de points d'intérêt, la réalisation des processus ponctuels spatiaux est directement applicable. Pour les formes dans la carte topographique, le point considéré sera le centre de gravité de la forme, auquel nous associerons le vecteur descripteur défini dans le chapitre 1 (la méthode IDSC). Nous analysons la répartition spatiale des points notés par $\{s_i\}$ qui est associé à un vecteur descripteur et une position locale dans l'image. La prise en compte de la dimension spatiale de la distribution des points peut être envisagée de différentes manières. Ici nous souhaitons considérer la distribution statistique spatiale de cooccurrences des signatures élémentaires marqués dans les différentes boules d'étude où les marques des signatures élémentaires sont leurs descripteurs caractéristiques.

Ce chapitre est organisé comme suit : d'abord nous présentons la construction d'un vocabulaire de mot visuel obtenu par la quantification des vecteurs des descriptions des signatures locales. Ensuite, nous décrivons la réalisation d'un processus spatial ponctuel marqué pour construire un nouveau descripteur d'images basé sur les statistiques spatiales de cooccurrences des points marqués. Les discussions sur la complexité et les propriétés d'invariances de ces descripteurs sont présentées à la fin de ce chapitre.

2.2 Quantification des descripteurs et mots visuels

Considérant une représentation d'image par un ensemble de descripteurs locaux, différentes techniques d'apprentissage peuvent être appliquées directement sur ces signatures pour la recherche d'images à partir du contenu ou pour la classification d'images [85, 86, 87]. Toutefois, la rapidité de la recherche (efficacité du calcul et complexité en mémoire) est une condition indispensable pour un grand nombre de tâches d'indexation, d'annotation et de classification d'images basées sur un ensemble de points

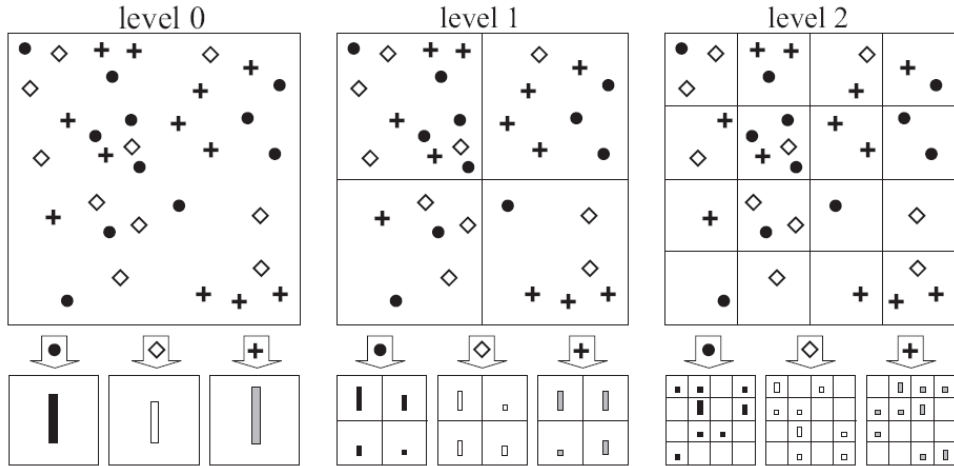


Figure 2.1 — La visualisation de trois différents niveaux de la représentation pyramidale de mots visuels dans une image. Le niveau 0 illustre les sacs de mots. La référence de la figure est dans le papier original [5].

d'intérêt. Sivic et Zisserman [88] proposent de regrouper les signatures locales similaires sous la forme d'un dictionnaire des points d'intérêt. La création des mots visuels est effectuée par une quantification des valeurs des descriptions visuelles. En général, cette quantification utilise un algorithme de clustering (k-moyennes [89]) qui prend comme entrée un échantillon des descriptions et crée des classes de points d'intérêt (*clusters*). Le centroïde de chaque classe représente un mot visuel. L'ensemble des mots visuels est appelé vocabulaire visuel. Après avoir construit un dictionnaire, chaque point d'intérêt s_i est décrit par une marque discrète correspondant à sa classe notée par $m_i = p$ où $p = \{1, \dots, k\}$, k est la taille de dictionnaire.

L'approche récente la plus populaire, le sac de mots (SDM) initialement proposée dans [8], repose sur un histogramme des fréquences d'apparition des mots visuels dans une image (Figure 2.1). Cette représentation nous permet de calculer les similarités entre images avec des gains significatifs en pertinence et en efficacité. Elle a été utilisée de plusieurs manières en reconnaissance de formes sur des images [90, 91, 92, 93]. Toutefois, l'approche du sac de mots se débarrasse de toutes les informations spatiales. Des approches plus récentes ont été également envisagées pour caractériser l'organisation spatiale des points d'intérêt et de leurs caractéristiques [24, 25]. La Figure 2.1 présente une représentation spatiale des mots visuels dans une image [5]. Le calcul à différents niveaux de la pyramide permet alors de décider pour chaque sous région la classe d'appartenance la plus probable. Le descripteur résultat hérite des propriétés d'invariance des signatures des points d'intérêt.

Le principe des méthodes basées sur l'analyse des informations spatiales de mots visuels (par exemple : sac de mots [8], pyramide de mots[5]) est présenté en Figure 2.2. Les étapes principales sont :

- La détection des points d'intérêt et leurs caractéristiques.
- La construction d'un dictionnaire des mots visuels basée sur l'algorithme de clustering de type de k-moyennes sur les caractéristiques des points d'intérêt.



Figure 2.2 — Le schéma des étapes principales des méthodes basées sur l'analyse des informations spatiales de mots visuels.

- L'estimation des différentes configurations spatiales de distributions des points.

C'est aussi le schéma des étapes principales de nos approches proposées dans ce travail, où la troisième étape est l'analyse statistique basée sur les processus ponctuels spatiaux.

2.3 Cooccurrence statistique spatiale des points marqués : méthode proposée

Après avoir procédé à la quantification vectorielle des descripteurs des points, rangeant chaque point dans une catégorie discrète, nous proposons ici un nouveau descripteur d'image reposant sur la caractérisation de l'organisation spatiale des signatures élémentaires d'une image et de ces caractéristiques. Nous nous intéressons plus particulièrement à la réalisation d'un processus spatial ponctuel marqué, la marque étant la catégorie des signatures élémentaires.

2.3.1 Processus ponctuels spatiaux

Les méthodes d'analyse spatiale connaissent un développement important dans de nombreux domaines d'application dans des disciplines variées telles que l'épidémiologie [26, 27, 28], l'écologie, la biologie et l'exploitation forestière [29, 30, 31], la géographie [32], l'astronomie [33], le traitement d'image [34, 94], etc. La statistique spatiale étudie des phénomènes dont l'observation aléatoire est la répartition spatiale $s = \{s_1, s_2, \dots, s_n\}$ des n points d'un ensemble spatial $B \subset \mathbb{R}^2$. Une réalisation s est dite localement finie si le nombre $N(s)$ de points de s est fini si B est fermée et bornée. Une discussion plus étendue des notions présentées dans cette section est disponible dans les articles de Moller et Waagepetersen [95]; de Gaetan et Guyon [96].

• **Définition** : Un processus ponctuel S sur B est la donnée d'une réalisation aléatoire $s = \{s_i\} \subset S$ localement finie telle que pour tout borélien borné B_i , le nombre de points $N(B_i) = N_S(B_i)$ de points de S tombant dans B_i est une variable aléatoire.

Les propriétés du processus ponctuel sont données par :

- La loi du processus ponctuel est une probabilité P sur (E, ϵ) , où E est l'espace des configurations des états du processus ponctuel et ϵ est la tribu sur E .
- Un processus ponctuel sur $\mathbb{R}^d$ est stationnaire si, pour tout $\epsilon \in \mathbb{R}^d$, la loi du processus ponctuel est invariante par translation : $S_\epsilon = \{S_i + \epsilon\}$.
- Un processus ponctuel est isotropique si de plus de la loi de ρS , S après ρ -rotation, est identique à celle de S pour toute rotation ρ . La propriété d'isotropie implique la stationnarité.

• **Moments d'un processus ponctuel** : Considérant une réalisation du processus ponctuel, les moments de la variable aléatoire sont pertinents pour les statistiques descriptives. Le moment d'ordre p du processus ponctuel S est défini dans le cas général par :

$$\mu^{(p)}(B_1 \times \dots \times B_p) = E\{\mathbb{N}(B_1) \dots \mathbb{N}(B_p)\} \quad (2.1)$$

où E est l'espérance mathématique, $\mathbb{N}(B_i)$ est le nombre des points aléatoires dans le borélien B_i .

La mesure d'intensité du processus ponctuel S est la mesure de moment d'ordre 1 ($p = 1$) :

$$\mu^{(1)}(B_i) = E\{\mathbb{N}(B_i)\} = E \sum_{s \in \mathbb{S}} \mathbb{I}_{B_i}(s) = \int_{B_i} \rho(s) ds \quad (2.2)$$

où $\mathbb{I}_{B_i}(s)$ est la fonction indicatrice qui explicite l'appartenance du point s au borélien B_i , définie par :

$$\mathbb{I}_{B_i}(s) = \begin{cases} 1 & \text{si } s \in B_i \\ 0 & \text{si } s \notin B_i \end{cases}$$

et $\rho(s)ds$ est la probabilité qu'il y ait un point de S tombant dans le petit volume infinitésimal ds voisin du point s . La normalisation du moment de premier ordre $\lambda = \mu^{(1)}(B_i)/|B_i|$ est la densité moyenne de points dans une unité de surface, avec $|B_i|$ qui est la surface d'un borélien B_i . Si S est stationnaire, λ est invariante par translation ν , $\lambda(B_i) = \tau\nu(B_i)$ où τ est l'intensité constante de S , nombre moyen de points par volume unité. Cette quantité suit une loi de Poisson et si B_i sont des boréliens disjoints, les points individuels d'une réalisation sont indépendants. En fait, la mesure de moment d'ordre 1 ne révèle pas l'organisation d'un processus spatial ponctuel stationnaire. Les moments d'ordre supérieur doivent être considérés en conséquence.

La structure de covariance de variables de comptage, les statistiques descriptives de la paire de points dans une collection aléatoire finie, peuvent être caractérisées à partir de la mesure de moment d'ordre 2 de S définie par :

$$\begin{aligned} \mu^{(2)}(B_1 \times B_2) &= E \sum_{s_1 \in \mathbb{S}} \sum_{s_2 \in \mathbb{S}} \mathbb{I}_{B_1}(s_1) \mathbb{I}_{B_2}(s_2) \\ &= \int_{B_1 \times B_2} \rho^{(2)}(s_1, s_2) ds_1 ds_2 \end{aligned} \quad (2.3)$$

où la densité d'intensité d'ordre 2 $\rho^{(2)}(s_1, s_2)ds_1ds_2$ est la probabilité que S possède un point dans le volume infinitésimal ds_1 autour de s_1 et un point dans le volume infinitésimal ds_2 autour de s_2 . Si S est isotropique, cette fonction de densité ne dépend que de la distance $\|s_1 - s_2\|$ entre les points s_1 et s_2 . Dans la théorie des processus ponctuels spatiaux [97, 98, 99], la mesure de moment factoriel $\alpha^{(2)}$ est fréquemment utilisée au lieu de la mesure de moment d'ordre 2 $\mu^{(2)}$ par :

$$\alpha^{(2)}(B_1 \times B_2) = E \sum_{s_1 \in \mathbb{S}} \sum_{\substack{s_2 \in \mathbb{S} \\ (s_2 \neq s_1)}} \mathbb{I}_{B_1}(s_1) \mathbb{I}_{B_2}(s_2) \quad (2.4)$$

où la relation entre la mesure de moment d'ordre 2 $\mu^{(2)}$ et la mesure de moment factoriel $\alpha^{(2)}$ est donnée par :

$$\alpha^{(2)}(B_1 \times B_2) = \mu^{(2)}(B_1 \times B_2) - \mu(B_1 \cap B_2) \quad (2.5)$$

2.3.2 Processus ponctuels marqués

Dans ce travail, nous appliquons la théorie des processus ponctuels spatiaux pour l'analyse de l'organisation des signatures élémentaires dans une image (points d'intérêt ou formes) qui ont les vecteurs descripteurs attachés. Comme pour l'approche sac de mot (SDM), nous construisons un dictionnaire des points et chaque point est décrit par une marque correspondant à sa classe. Considérant un ensemble de catégories des points $M = \{m_1, m_2, \dots, m_k\}$ dans une image, un processus ponctuel marqué Ψ sur $B \times M$ est un processus ponctuel sur B où la marque m_i est associée à chaque point s_i dans un borélien B . Ici, nous présentons la réalisation d'un processus ponctuel marqué sur les boules fermées $D(s_i, r)$ centrées en s_i et de rayon $r > 0$ (Figure 2.3). Les mesures des moments du processus ponctuel marqué se font de la même manière que pour le processus ponctuel. La statistique spatiale de cooccurrence des points de processus ponctuels marqués est caractérisée par la mesure de leur moment factoriel d'ordre 2 défini par :

$$\alpha_{i,j}^{(2)}(r) = E \left\{ \sum_h \sum_{l \neq h} \delta_i(m_h) \delta_j(m_l) \mathbb{I}(\|s_h - s_l\| \leq r) \right\} \quad (2.6)$$

où $\delta_i(m_h)$ est la fonction de Kronecker qui prend la valeur 1 si la marque m_h de point s_h est i , sinon elle prend la valeur 0.

2.3.3 Estimation de la fonction K de Ripley

La fonction K de Ripley [100] est la distribution cumulative des fréquences d'observation à une distance (ou classe de distance) donnée point par point. Formellement, la définition de la fonction K de Ripley est donnée pour le cas univariable par :

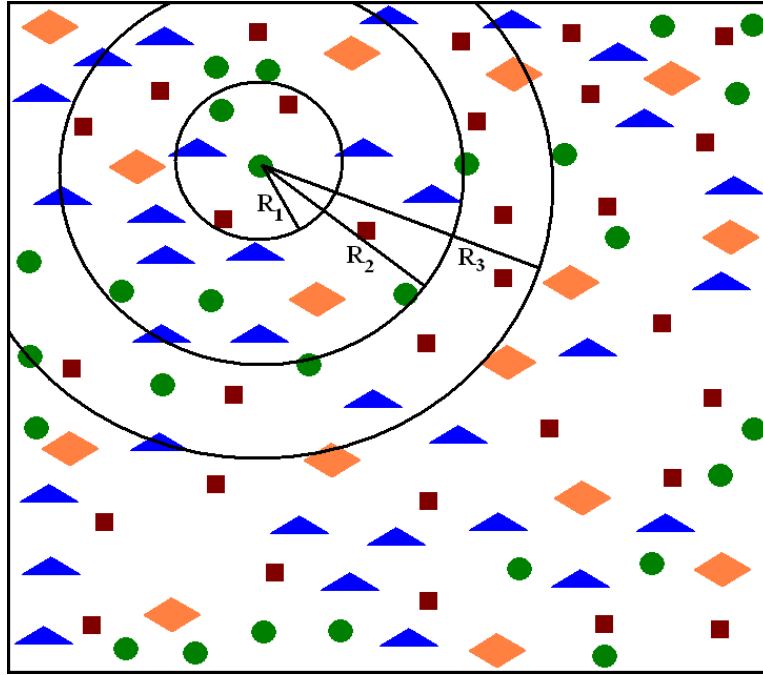


Figure 2.3 — Illustration de l'analyse de statistiques spatiales de cooccurrence des points marqués dans les différents cercles d'étude.

$$K(r) = \lambda^{-1} E(r) \quad (2.7)$$

où $E(r)$ est l'espérance du nombre de points à l'intérieur d'un cercle de rayon r centré autour d'un point quelconque du semis de points, λ est la densité moyenne de points dans une unité de surface. La statistique de Ripley permet de quantifier l'intensité du motif à de multiples échelles en changeant les différentes valeurs de r . La fonction K de Ripley est utilisée pour comparer le motif ponctuel observé à un motif ponctuel engendré par un processus ponctuel connu [101].

Pour mesurer le semis de points de type j dans une boule d'étude de rayon r centrée en un point de type i , nous considérons la version bivariable de la fonction K de Ripley donnée par :

$$K_{ij}(r) = \frac{1}{\lambda_i \lambda_j} \alpha_{ij}^{(2)}(r) \quad (2.8)$$

Ces statistiques d'ordre 2 ont démontré leur intérêt pour la caractérisation et la détection de structures spatiales complexes et offrent des outils efficaces d'interprétation de nuages de points. Nous proposons de considérer des statistiques de cooccurrence permettant d'étudier la répartition spatiale des points de type j des voisinages de tous les points de type i dans une image I :

$$\Gamma_{ij}^I(r) = \lambda_j K_{ij}(r) = \frac{1}{\lambda_i} \alpha_{ij}^{(2)}(r) \quad (2.9)$$

L'estimation des statistiques spatiales de cooccurrences pour tous les points marqués $\Gamma_{ij}^I(r)$ dans une image crée alors un vecteur descripteur de l'image. Etant donné un ensemble N_r de valeurs de rayon de voisinage et un dictionnaire de k mots visuels, une image est caractérisée par un vecteur de taille $N_r k^2$.

2.3.4 Correction des effets de bord

Regardant les différentes valeurs de rayon de voisinage pour la caractérisation de structures spatiales par statistiques de Ripley, l'estimation de la densité λ des points près du bord de l'image peut notablement dégrader les statistiques basées sur les distances (les figures 2.3 et 2.4). Donc le calcul robuste de la fonction K de Ripley doit être pondéré pour tenir compte des effets de bord [102, 103, 94]. La correction géométrique pour les points aux bords de l'image, relativement à la taille du voisinage considéré, est présentée sur la Figure 2.4. Dans ce travail, nous nous limitons au cas d'intersection avec un ou deux côtés de l'image. Les intersections avec trois ou quatre côtés d'image ne sont pas considérées, puisque la taille maximale du voisinage considéré est inférieur à la plus petite dimension de l'image. Nous notons e_1, e_2, e_3, e_4 les distances entre un point étudié et les quatre côtés d'image. Le Tableau 2.1 présente les formules des angles intérieurs du cercle pour chaque cas d'intersection avec la frontière d'image.

Dans ce contexte, le calcul des statistiques descriptives spatiales de cooccurrences consiste en une surface réelle A_i autour du point i qui est donnée alors par :

- Lorsque le rayon r du disque de voisinage est plus petit que toutes les distances e de son centre aux quatre côtés de l'image, il n'y a pas d'intersection avec le bord de l'image. La surface A_i est donnée alors :

$$A_i = \frac{1}{2} \mathfrak{S}_0 r^2 = \pi r^2 \quad (2.10)$$

- Dans le cas d'une intersection avec un côté de l'image (Tab.2.1, deuxième ligne), le calcul fait intervenir la surface du disque extérieur à l'image donnée par :

$$A_i = \frac{1}{2} \mathfrak{S}_1 r^2 + e_1 \sqrt{r^2 - e_1^2} \quad (2.11)$$

- Dans le cas de l'intersection avec deux côtés de l'image, deux différents cas correspondant à la position du coin de l'image avec le cercle d'étude (voir la Figure 2.4b,c.) sont obtenus :

- La troisième ligne du Tableau 2.1 décrit la formule d'angle intérieur dans le cas d'un coin (Figure 2.4b) correspondant à la condition $R^2 > e_1^2 + e_2^2$, où la surface A_i est calculée par :

$$A_i = \frac{1}{2} \mathfrak{S}_{21} r^2 + \frac{1}{2} e_1 \sqrt{r^2 - e_1^2} + \frac{1}{2} e_2 \sqrt{r^2 - e_2^2} + e_1 e_2 \quad (2.12)$$

- Par contre, avec la condition $R^2 < e_1^2 + e_2^2$, nous obtenons la formule dans la dernière ligne du Tableau 2.1 qui correspond à la position du coin (Figure 2.4c) , où la surface A_i est définie par :

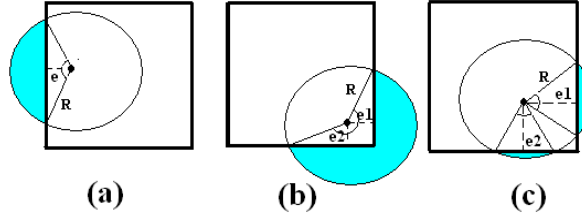


Figure 2.4 — La visualisation des cas d’intersection du cercle étude avec la frontière image.

Condition du rayon	Angles intérieurs du cercle étude
no-edge $r \leq e_1, e_2, e_3, e_4$	$\mathfrak{S}_0 = 2\pi$
one-edge $\begin{cases} r > e_1 \\ r \leq e_2, e_3, e_4 \end{cases}$	Figure 2.4(a) $\mathfrak{S}_1 = 2\pi - 2 \arccos(\frac{e_1}{r})$
two-edge1 $\begin{cases} e_1, e_2 < r \leq e_3, e_4 \\ r^2 > e_1^2 + e_2^2 \end{cases}$	Figure 2.4(b) $\mathfrak{S}_{21} = 2\pi - (\frac{\pi}{2} + \arccos(\frac{e_1}{r}) + \arccos(\frac{e_2}{r}))$
two-edge2 $\begin{cases} e_1, e_2 < r \leq e_3, e_4 \\ r^2 \leq e_1^2 + e_2^2 \end{cases}$	Figure 2.4(c) $\mathfrak{S}_{22} = 2\pi - (\arccos(\frac{e_1}{r}) + \arccos(\frac{e_2}{r}))$

Tableau 2.1 — La liste des conditions et des formules de calcul des angles intérieurs $\mathfrak{S}$ pour les différents cas d’intersection du cercle d’étude avec la frontière d’image.

$$A_i = \frac{1}{2} \mathfrak{S}_{22} r^2 + e_1 \sqrt{r^2 - e_1^2} + e_2 \sqrt{r^2 - e_2^2} \quad (2.13)$$

2.3.5 Réduction de la complexité du vecteur descripteur

La statistique descriptive ci-dessus repose sur la fréquence moyenne d’apparition des points d’une catégorie considérée dans les boules des rayons r centrées en un autre point d’une autre catégorie. La taille du descripteur, basé sur les processus ponctuels marqués, d’une image est une fonction de k et N_r soit : $N_r k^2$ à comparer avec k de l’algorithme SDM, où k est le nombre de catégories et N_r est un ensemble de valeurs de rayons. Il faut également noter que la grande taille de descripteurs a un effet sur la performance de la classification et de la reconnaissance des images. Par exemple, considérant les statistiques spatiales de cooccurrences de $k = 40$ classes de points dans $N_r = 20$ différentes valeurs de rayons des boules, un histogramme de 32000-dimensions est créé. Dans ce contexte, nous avons introduit une procédure pour réduire la complexité des caractéristiques de notre descripteur [104]. L’algorithme de clustering est appliqué une deuxième fois sur toutes les paires des points marqués pour déterminer les catégories

de paires des points. La distance entre deux paires des points marqués $P(c_1, c_2)$ et $P(c_3, c_4)$ est donnée par :

$$||P(c_1, c_2) - P(c_3, c_4)|| = \min(||c_1 - c_3|| + ||c_2 - c_4||, ||c_1 - c_4|| + ||c_2 - c_3||) \quad (2.14)$$

où c_i est le descripteur du centroïde de catégorie i , la distance $||c_i - c_j||$ est calculée par les mesures de similarité connues : distance euclidienne, χ^2 , etc.

Après avoir construit un dictionnaire des paires de points, une marque discrète $u = M(s_h, s_l)$, $u = \{1, \dots, k^*\}$ est attaché à la catégorie d'une paire de points marqués s_h et s_l , où k^* est la taille de dictionnaire. Puis, nous proposons d'estimer les fréquences d'apparition des catégories des paires de points dans les différentes boules de rayon r . Pour ce faire, la statistique spatiale d'ordre 2 dans l'équation 2.6 est donnée par :

$$\alpha_u^{(2)}(r) = E \left\{ \sum_h \sum_{l \neq h} \delta_u(M(s_h, s_l)) \mathbb{I}(\|s_h - s_l\| \leq r) \right\} \quad (2.15)$$

Après avoir appliqué cette procédure, la complexité du vecteur descripteur des statistiques descriptives est réduit de $N_r k^2$ à $N_r k^*$ dimensions.

2.3.6 Discussion des propriétés d'invariance des descripteurs proposés

Un nouveau descripteur basé sur la caractérisation de l'organisation spatiale des points par d'un processus ponctuel marqué est considéré. Ses statistiques reposent initialement sur la construction d'un dictionnaire des caractéristiques des signatures élémentaires dans une image. Le descripteur proposé hérite la propriété d'invariance au changement de contraste du descripteur des signatures élémentaires locales. En outre, l'invariance en rotation et en translation de ce descripteur est assurée par la théorie du processus ponctuel marqué. Ces statistiques spatiales de cooccurrences des points sont pourtant calculées avec différentes fonctions de voisinage. Les valeurs de rayon des boules étudiées peuvent être considérées comme un paramètre d'échelle. En fait, l'analyse de l'invariance en échelle est importante en pratique pour la définition robuste d'un descripteur d'image. Les effets d'échelle doivent donc être considérés. Une image dans l'ensemble des images est choisie comme une image de référence I_{ref} avec un facteur d'échelle $\eta_{ref} = 1$. Considérant une autre image, son facteur d'échelle est calculé par :

$$\eta = \frac{\lambda_{ref} \eta_{ref}}{\lambda} \quad (2.16)$$

où λ est la densité moyenne des points sur une unité de surface. Les valeurs réelles de rayons r pour le calcul de ces statistiques descriptives dépendent de son facteur d'échelle. Considérant deux images de même texture avec différents facteurs d'échelle, la Figure 2.5 montre l'intérêt de l'utilisation d'une procédure de calcul des effets d'échelle pour l'estimation des statistiques spatiales de cooccurrence.

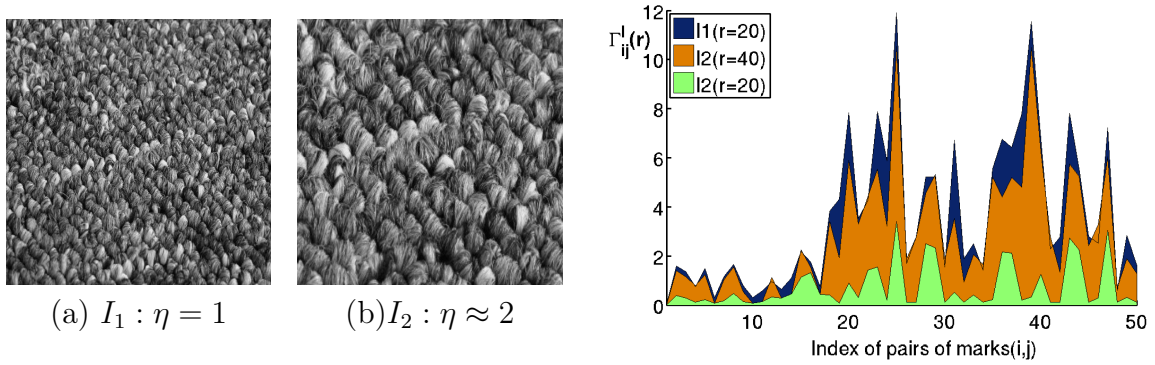


Figure 2.5 — Illustration des effets d'échelle pour le calcul des statistiques descriptives d'ordre 2 (SSC) : une image de référence I_1 (a) avec le facteur d'échelle $\eta_1 = 1$ et une autre image de même texture I_2 (b) avec le facteur d'échelle $\eta_2 \approx 2$ sont considérées. Sur le tracé, l'histogramme de SSC pour l'image I_1 est en bleu qui est calculé avec une boule de rayon $r = 20$. Les vecteurs descripteurs de SSC pour l'image I_2 sont en vert (où $r = 20$ sans la procédure d'adaptation d'échelle) et en orange (où $r = 40$ avec la procédure d'adaptation d'échelle). Une comparaison de deux distances de la mesure similarité entre l'histogramme de I_1 et avec deux histogrammes de I_2 montre l'intérêt d'appliquer la procédure d'adaptation d'échelle.

2.4 Conclusion

Nous avons proposé de caractériser une image en utilisant une modélisation non-paramétrique de la distribution conjointe des vecteurs descripteurs associés à des couples de points d'intérêt et de leur position spatiale. L'estimation applique la fonction K de Ripley à k^* catégories après la quantification vectorielle des couples de descripteurs, donnant aussi une complexité raisonnable de $N_r k^*$ pour le vecteur caractérisant de l'image. Les descripteurs proposés héritent des propriétés d'invariance des signatures des points d'intérêt ou des formes. La complexité de ces différents descripteurs est directement liée au nombre de catégories des mots visuels et aux valeurs de rayons des boules d'étude. Nous avons proposé des procédures pour réduire la taille de descripteurs et pour résoudre les variations en d'échelle des images. Dans la section expérimentale, la pertinence des descripteurs proposés pour les applications de classification d'images, ainsi que leur propriétés d'invariance seront discutées plus en détail sur des exemples concrets.

3.1 Introduction

Dans ce chapitre, nous cherchons à utiliser une modélisation statistique paramétriques des processus ponctuels spatiaux multivariés. L'intérêt en est une description plus précise et concise de l'image. Nous étudions des modèles formels spécifiques pour montrer la pertinence de ces modèles à exploiter les distributions des données de points multivariées. Dans cette thèse, nous considérons le processus de Cox log-Gaussien introduit par Moller et al. [105]. Ce modèle a prouvé sa flexibilité, et la pertinence d'analyse qui en découle pour les applications dans des disciplines particulières telles que l'épidémiologie [28] et l'agriculture [106]. Notre approche consiste en l'estimation de la structure de covariance des points marqués déterminée par le processus de Cox log-Gaussien. Les différents types de structure de covariance ont été analysés pour créer un vecteur descripteur d'image avec de bonnes propriétés d'invariance. Il faut également noter que les processus ponctuels spatiaux ont déjà été appliqués dans les applications d'analyse d'image, par exemple Lafarge et al. [34] ont appliqué le modèle de Gibbs pour détecter les structures géométriques dans des images texturées.

Ce chapitre est organisé comme suit. Dans la première section, nous présentons brièvement les principes des principaux modèles statistiques spatiaux. Dans la deuxième section, nous décrivons la réalisation d'un processus ponctuel de Cox log-Gaussien pour la caractérisation et la modélisation de la distribution spatiale d'ensemble de signatures locales dans l'image. Dans la troisième section, nous exploitons une estimation du modèle de Cox log-Gaussien pour définir une représentation d'image et les discussions sur la complexité et les propriétés d'invariance des descripteurs proposés sont présentées. Finalement, la dernière section donne des conclusions.

3.2 Généralités sur les modèles statistiques spatiaux

Les processus ponctuels connaissent un développement important dans de nombreux domaines d'application d'analyse spatiale. Nous donnons ci-dessous trois grandes classes de processus ponctuels :

- Les processus ponctuels de Poisson S de densité moyenne μ et d'intensité λ

[107, 108] qui modélisent une répartition spatiale de points indépendants sont caractérisés par les deux propriétés suivantes :

- Pour tous les boréliens bornés B_i de mesure $0 < \mu(B_i) < \infty$, $N(B_i)$ suit une loi de Poisson de paramètre $\mu(B_i)$.
 - Pour tous les boréliens bornés B_i , qui sont des boréliens disjoints, les variables aléatoires $N(B_i)$ sont indépendantes.
- La deuxième classe contient les processus ponctuels de Poisson conditionnels à un environnement aléatoire qui modélisent des répartitions spatiales de points présentant des agrégats [109, 110, 111, 112]. Il peut s’agir des processus ponctuels de Neyman-Scott [33], de *shot noise* [113], de Cox log-Gaussien [105], etc. Un processus de cette classe est un processus ponctuel de Poisson de densité aléatoire modélisant un environnement aléatoire. Les paramètres apparaissant dans ces modélisations sont typiquement les paramètres d’estimation des covariances, des variances des variables explicatives en chaque point considéré.
- Les modèles à noyau dur, par exemple : les processus ponctuels de Gibbs [114, 115], de Strauss [116], sont définis à partir d’une spécification conditionnelle pour modéliser des configurations spatiales régulières de points. La densité d’un processus de cette classe est caractérisée par la fonction d’énergie $U(\{x_i\})$ et la fonction de partition normalisante $\bar{Z}$:

$$f(\{x_i\}) = \exp\{-U(\{x_i\})/\bar{Z}\} \quad (3.1)$$

Les différentes spécifications de la fonction d’énergie $U(\{x_i\})$ qui s’écrivent sous forme de somme de potentiels d’interaction correspondent à différents modèles. Par exemple, l’énergie $U(\{x_i\})$ qui a été introduite par Descamps et al. [117] pour une application de détection d’objets dans l’image est considérée comme la somme d’un terme d’énergie a priori $U_p(\{x_i\})$ et un terme attache aux données $U_d(\{x_i\})$

$$U(\{x_i\}) = U_p(\{x_i\}) + U_d(\{x_i\}) \quad (3.2)$$

où l’énergie $U_p(\{x_i\})$ prend en compte des interactions entre les objets du processus et l’énergie $U_d(\{x_i\})$ prend en compte des valeurs de l’image analysée.

Comme les auteurs l’ont souligné dans [28, 34, 99], il est difficile de recommander a priori le meilleur modèle parmi les nombreux modèles trouvés dans la bibliographie pour l’analyse de la distribution spatiale des points marqués. Dans ce travail, nous nous intéressons à étudier la modélisation des distributions de points marqués basées sur le processus ponctuel de Cox log-Gaussien qui est caractérisé par les fonctions d’intensités log-linéaire aléatoires et de covariance des champs Gaussien et la fonction K de Ripley considérée dans le chapitre 2. Les détails sur ce type de processus ponctuels et leurs propriétés seront présentés par la suite. L’intérêt du modèle choisi sera aussi discuté en conclusion de ce chapitre.

3.3 Modélisation statistique spatiale d'ensemble de signatures locales dans l'image par processus de Cox log-Gaussien

3.3.1 Processus ponctuels de Cox log-Gaussien dans le cas univarié

Le processus de Cox X gouverné par une fonction d'intensité aléatoire $\Lambda = \{\Lambda(s) : s \in \mathbb{R}^2\}$ est une généralisation du processus ponctuel de Poisson doublement stochastique où l'intensité est aléatoire. C'est-à-dire que pour chaque réalisation, on tire dans un premier temps de façon aléatoire les valeurs de la densité locale sur le domaine, et dans un deuxième temps la position des points en fonction de cette densité locale [96, 97]. Si la densité Λ est stationnaire alors le processus X est aussi stationnaire. Le processus de Cox est donc un processus ponctuel homogène et ses propriétés sont les mêmes en tout point du domaine d'étude. Ce processus fournit un modèle utile qui est fréquemment appliqué pour modéliser les configurations de points présentant des agrégats [96, 97, 111].

Les processus de Cox log-Gaussien $\hat{X}$ introduit par Moller et al. [105] admettent une intensité log-linéaire avec effet aléatoire sur tout borélien B borné :

$$\Lambda = \exp\{Y\} \quad (3.3)$$

où $Y = \{Y(s) : s \in \mathbb{R}^2\}$ est un processus Gaussien, c'est-à-dire que la distribution jointe de chaque champ Gaussien $(Y(s_1), \dots, Y(s_n))$ est Gaussienne. Conditionnellement à Λ , la distribution du processus de Cox log-Gaussien $\hat{X}$ est un processus de Poisson avec la fonction d'intensité aléatoire Λ . Le processus de Cox log-Gaussien $\hat{X}$ peut être paramétrisé par sa fonction de moyenne $\mu = EY(s)$, sa fonction de variance $\sigma^2 = Var(Y(s))$ et sa fonction de covariance $c(r) = Cov(Y(s_1), Y(s_2))$, où $r = \|s_1 - s_2\|$ est la distance entre deux points s_1 et s_2 . Si Y est stationnaire (res. isotropique) alors $\hat{X}$ est stationnaire (res. isotropique). Pour un processus stationnaire de Cox log-Gaussien $\hat{X}$, les mesures de moments de ces processus sont définis dans le cas général par :

$$\begin{aligned} \rho^{(n)}(s_1, \dots, s_n) &= \exp \left\{ n\mu + \sigma^2 \left(\frac{n}{2} + \sum_{1 \leq i < j \leq n} \|s_i - s_j\| \right) \right\} \\ &= \rho^n \prod_{1 \leq i < j \leq n} g(s_i - s_j) \end{aligned} \quad (3.4)$$

Par conséquent, la fonction d'intensité λ et la fonction de corrélation de paires sont données par (voir l'Annexe 1) :

$$\begin{cases} \rho(s) = \lambda = \exp(\mu + \sigma^2/2), & \text{pour } n = 1 \\ \frac{\rho^{(2)}(s_1, s_2)}{\rho(s_1)\rho(s_2)} = g(r) = \exp(c(r)), & \text{pour } n = 2 \end{cases} \quad (3.5)$$

où la fonction de corrélation de paires répondérée $g(r) = \rho^{(2)}(s_1, s_2) / (\rho(s_1)\rho(s_2))$ introduit par Baddeley et al. [118] est un bon indicateur au second ordre de la dépendance spatiale (si $\rho(s_1)$ et $\rho(s_2) > 0$), ce qui conduit à l'interprétation :

- $g(r) = 1$ traduit une situation d'indépendance entre la position des points.
- $g(r) > 1$ traduit une tendance à l'accumulation des points.
- $g(r) < 1$ traduit une tendance à la dispersion des points.

3.3.2 Processus ponctuels de Cox log-Gaussien dans le cas multi-varié

Etant donné un ensemble de catégories des points, le modèle de Cox log-Gaussien peut être développé pour le cas multi-varié comme suit. Le processus de Cox log-Gaussien $\hat{X} = \{\hat{X}_i\}$ gouverné par les fonctions d'intensités aléatoires $\Lambda_i = \{\Lambda_i(s) = \exp(Y_i(s)) : s \in \mathbb{R}^2\}$ où $\{Y_i(s)\}$ est un processus Gaussien stationnaire et isotrope avec les fonctions de moyenne $\{\mu_i\} = E\{Y_i(s)\}$ et les fonctions de covariance $c_{ij}(r) = Cov(Y_i(s_1), Y_j(s_2))$, où $r = \|s_1 - s_2\|$ est la distance entre deux points s_1 et s_2 . En considérant la conditionnelle sur Y , $\hat{X}_i$ est un processus ponctuel de Poisson doublement stochastique où l'intensité Λ_i est aléatoire [105, 106, 119].

La fonction d'intensité λ et la fonction de corrélation de paires de $\hat{X}_i$ dans l'équation 3.5 sont alors données par :

$$\begin{cases} \lambda_i = \exp(\mu_i + \sigma_i^2/2) \\ g_{ii}(r) = \exp(c_{ii}(r)) \end{cases} \quad (3.6)$$

et la fonction mixte de corrélation de paires de $\hat{X}_i$ et $\hat{X}_j$ est définie par :

$$\bar{g}_{ij}(r) = E[\Lambda_i(s_1)\Lambda_j(s_2)] / (\rho_1\rho_2) = \exp(c_{ij}(r)) \quad (3.7)$$

3.3.3 Estimation de fonctions covariances des modèles de Cox log-Gaussien

Après avoir déterminé la distribution du processus de Cox log-Gaussien par la fonction d'intensité et la fonction de corrélation de paires, nous proposons une estimation statistique basée sur l'estimation des moments du modèle de Cox log-Gaussien pour exploiter la donnée des points de la manière suivante.

La fonction de corrélation de paires g et la fonction K de Ripley sont les deux propriétés importantes d'ordre 2 de la théorie des statistiques spatiales. La procédure proposée de l'estimation repose sur la relation entre la fonction g et la fonction K définie par [105, 99] :

$$K_{ij}(R) = 2\pi \int_0^R r g_{ij}(r) dr \quad (3.8)$$

Exponentiel	Cardinal sine	Hyperbolique
$\exp(-(r/\beta)^t)$	$\frac{\sin(r/\beta)}{(r/\beta)}$	$(1 + r/\beta)^{-1}$

Tableau 3.1 — Les différentes formes du modèle de paramétrisation $\mathbb{L}(\beta, r)$.

où R est la valeur de rayon déterminée par l'utilisateur. Dans le Chapitre 2, K_{ij} est définie comme une mesure du semis de points de type j dans une boule étude de rayon r centrée en un point de type i définie dans l'équation 2.8. En combinant les équations 2.6, 2.8 et 3.8, la fonction de corrélation de paires g_{ij} est définie par :

$$g_{ij}(r) = \frac{1}{2\pi r \lambda_i \lambda_j} \sum_h \sum_{l \neq h} \delta_i(m_h) \delta_j(m_l) \xi(\|s_h - s_l\|, r) b_{s_h} \quad (3.9)$$

où $\xi(\cdot)$ est un noyau (dans ce travail, nous choisissons le noyau Gaussien), λ_i est l'intensité des points de catégorie i , b_{s_h} est la proportion de la circonférence intérieure de la boucle considérée par rapport à l'image. En pratique, la procédure de la correction des effets de bord doit être prise en compte pour le calcul de $b_{s_h} = r\beta$, où β est l'angle intérieur du cercle d'étude présentée dans le Tableau 2.1 du chapitre 2. La fonction de corrélation de paires n'est donc pas symétrique pour les paires de i et j : $g_{ij} \neq g_{ji}$. L'estimation non-paramétrique de la fonction de covariance du modèle de Cox log-Gaussien dans les équations 3.6 et 3.7 est donnée par :

$$\begin{cases} c_{ii}(r) = \log(g_{ii}(r)) \\ c_{ij}(r) = \log(\bar{g}_{ij}(r)) = \log\left(\frac{\lambda_i g_{ij}(r) + \lambda_j g_{ji}(r)}{\lambda_i + \lambda_j}\right) \end{cases} \quad (3.10)$$

Par contre, la fonction de covariance c est symétrique pour les paires de i et j : $c_{ij} = c_{ji}$. Dans ce travail, nous souhaitons envisager les formes des paramètres de la fonction de covariance c à définir un modèle probabiliste compact pour une représentation d'image. Ce modèle est estimé par la minimisation du critère suivant :

$$\int_0^R \{\sigma_{ij}^2 \mathbb{L}(\beta_{ij}, r) - c_{ij}(r)\}^2 dr \quad (3.11)$$

où $\mathbb{L}(\beta, r)$ le modèle paramétrique de covariance. Une liste de modèle de paramétrisation est présentée dans le Tableau 3.1. En pratique, une procédure d'optimisation à base de gradient est appliquée pour résoudre la minimisation de ce critère. Les paramètres de variances σ_{ij} et les paramètres d'échelles β_{ij} sont déterminés.

Une représentation d'image basée sur l'estimation de la fonction de covariance du modèle de Cox log-Gaussien est caractérisée par une liste de paramètres $(\lambda_i, \sigma_{ij}, \beta_{ij})$. Le vecteur caractéristique d'image est finalement de taille $k(k+2)$, où k est le nombre de catégories des points dans l'image.

3.3.4 Réduction de la complexité du vecteur descripteur

Comparant la complexité entre les vecteurs descripteurs des différentes méthodes basées sur des mots visuels, la complexité $k(k+2)$ de l'estimation du modèle de Cox log-Gaussien est moindre que la complexité $N_r k^2$ des statistiques spatiales de cooccurrences (SSC) proposées dans le chapitre 2, où N_r est la liste de différentes valeurs de rayons des boules. Toutefois, comparativement au vecteur descripteur de k dimensions de l'algorithme SDM, notre vecteur descripteur proposé est plus compliqué. La taille du vecteur descripteur est un critère important qui intervient dans l'utilisation des algorithmes de classification et de reconnaissance d'images. Donc, il faut obtenir une procédure de la réduction de la complexité du vecteur descripteur. Ici, nous appliquons la même procédure que dans le chapitre 2 qui est basée sur la construction d'un dictionnaire des paires de points. L'estimation non-paramétrique de la fonction de covariance est donnée par :

$$c_u(r) = \log \left(\frac{1}{2\pi r \lambda_u} \sum_h \sum_{l \neq h} \delta_u(M(s_h, s_l)) \xi(\|s_h - s_l\|, r) b_{s_h} \right) \quad (3.12)$$

où $u = \{1, \dots, k^*\}$ est la catégorie attachée à une paire de points marqués s_h et s_l avec k^* la taille du dictionnaire. Par conséquent, la représentation d'une image qui est caractérisée par les paramètres de d'intensité λ_u , de variances σ_u et les paramètres de échelles β_u de chaque catégorie u des paires de points, sera calculée par la minimisation suivant :

$$\int_0^R \{ \sigma_u^2 \mathbb{L}(\beta_u, r) - c_u(r) \}^2 dr \quad (3.13)$$

Après avoir appliqué cette procédure, la complexité du vecteur descripteur des modèles de Cox log-Gaussien est réduite de $k(k+2)$ à $3k^*$ dimensions.

3.3.5 Discussion sur les propriétés d'invariances du descripteur proposé

L'intérêt de cette approche réside notamment dans ses propriétés d'invariance au changement de contraste, en rotation et en translation des images qui sont hérités naturellement du descripteur des signatures élémentaires locales (les points intérêts et les formes) et de l'analyse des statistiques spatiales basées sur le modèle de Cox.

En outre, l'image mise à l'échelle affecte clairement les moments d'ordre 2 des processus ponctuels spatiaux et ses modèles associés qui sont calculés dans les boucles étudiées, où les valeurs de ses rayons sont considérées comme les paramètres d'échelle. Donc, une adaptation des effets d'échelle doit être envisagée pour une définition robuste d'un vecteur descripteur basé sur l'estimation du modèle de Cox log-Gaussien. Ici, nous ré-appliquons la procédure de calcul le facteur d'échelle pour chaque image proposée

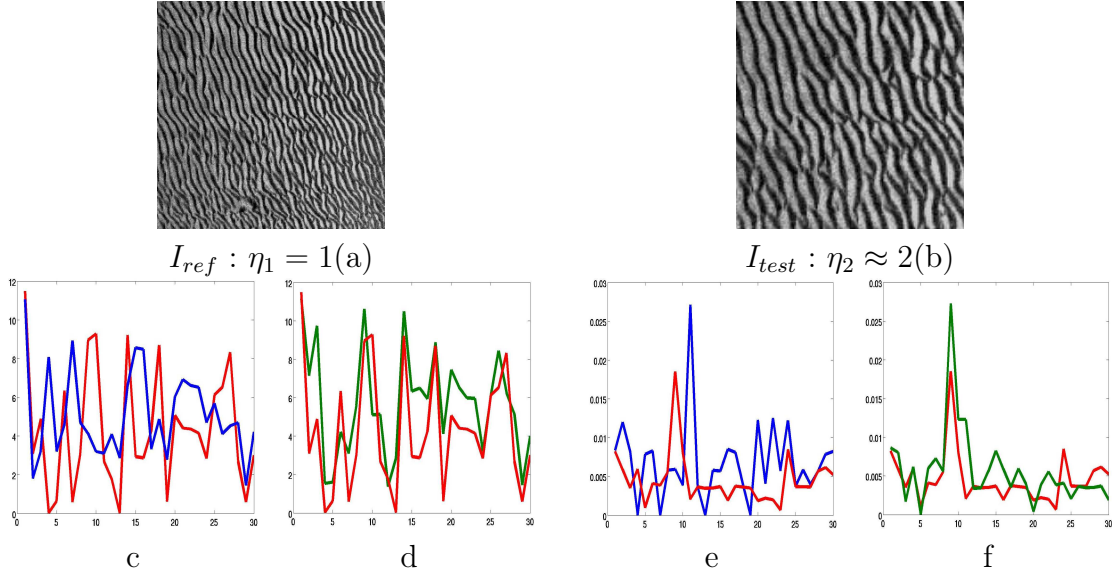


Figure 3.1 — Exemple des effets d'échelle pour l'estimation des paramètres de la fonction covariance du modèle de Cox log-Gaussien appliqué à la réduction de la complexité. Nous considérons deux images suivantes : une image de référence I_1 (a) avec le facteur d'échelle $\eta_1 = 1$ et une autre image de même texture I_2 (b) avec le facteur d'échelle $\eta_2 \approx 2$. Sur les tracés, les histogrammes de paramètres de variances σ_u (c,d) et de paramètres d'échelle β_u (e,f) d'image I_1 sont en rouge. Les courbes en vert (res. en bleu) sont les histogrammes de paramètres d'image I_2 calculées avec (res. sans) la procédure d'adaptation d'échelle.

dans le Chapitre 2. C'est-à-dire que facteur d'échelle η de l'image est calculé par :

$$\eta = \frac{\lambda_{ref} \eta_{ref}}{\lambda} \quad (3.14)$$

où λ est la densité moyenne des points sur une unité de surface, $\{\lambda_{ref}, \eta_{ref}\}$ est la densité moyenne et le facteur d'échelle de l'image de référence qui est choisie de manière aléatoire de l'ensemble des images d'apprentissage. Les valeurs réelles de rayons R des cercles pour l'estimation des paramètres du modèle de Cox log-Gaussien dans les équations 3.11 et 3.13 dépendent du facteur d'échelle η de l'image. Un exemple des effets d'échelle pour notre approche proposée est donné dans la Figure 3.1. Les distances de similarité entre les courbes rouges et vertes dans la Figure 3.1d (res. f) pour le cas d'utilisation d'adaptation en échelle sont plus faibles que les distances entre les courbes rouges et bleus dans la Figure 3.1c (res. e). Ces résultats montrent l'intérêt de notre procédure d'adaptation en échelle.

3.4 Conclusion

Dans ce chapitre, nous avons présenté les modèles statistiques spatiaux pour définir un nouveau descripteur d'analyse d'image présentant des qualités d'invariance basées sur la caractérisation de l'organisation spatiale des points d'intérêts (ou les formes) et

de leurs caractéristiques. Les modèles statistiques proposés consistent à considérer une image comme la réalisation des points marqués d'un processus de Cox log-Gaussien, la marque étant le descripteur du point d'intérêt (ou de la forme). Les propriétés d'invariance garantissent une représentation robuste de l'image par rapport au changement de contraste et aux transformations géométriques, comme dans le cas des méthodes utilisant la quantification des mots visuels [8, 25, 104].

Il faut également noter qu'il est difficile de choisir le meilleur modèle pour l'analyse statistique spatiale des points dans la littérature, par exemple : processus de Neyman-Scott, de Cox shot-noise ou de Gibbs. Ici, nous nous intéressons plus particulièrement aux modèles de Cox log-Gaussien multivariés qui possèdent les caractéristiques intéressantes suivantes :

- La distribution du processus de Cox log-Gaussien est complètement caractérisée par les fonctions d'intensités aléatoires et de covariance des champs de Gaussien. Cela conduit à une interprétation simple des modèles paramétriques et une estimation assez facile de ces modèles.
- Nous pouvons choisir les différentes formes paramétriques de la fonction de covariance. Ces modèles sont très flexibles pour couvrir une large gamme de structures de covariance.
- Le descripteur proposé ne dépend pas de la taille d'image et du problème des effets de bord.

En outre, ces modèles statistiques possèdent principalement les avantages suivants en comparaison avec les méthodes classiques des statistiques descriptives simples [25, 9, 104] :

- Les modèles de Cox log-Gaussien sont indépendants de la sélection des cercles d'étude. Le nombre de cercles et leur taille de rayon sont les paramètres critiques pour le calcul des statistiques descriptives.
- Les modèles proposés donnent une représentation compacte de l'image avec une taille de dimension $3k^*$ pour le vecteur descripteur vis-à-vis de la complexité de $N_r k^*$ pour les descripteurs issus de statistiques descriptives simples [25, 9, 104], où N_r est le nombre des cercles d'étude et k^* est la taille de dictionnaire des paires des points. Cette réduction de la complexité est un critère important en reconnaissance des images.

Dans les chapitres suivants, les représentations invariantes d'image basées sur les estimations du modèle de Cox log-Gaussien sont exploitées pour les applications à la classification de textures (UIUC, Brodatz, KTH-TIPs), d'images sonars texturées et la reconnaissance de scène.

4

Reconnaissance et classification d'images

4.1 Introduction

En vision par ordinateur, la classification d'images est un problème important pour l'organisation et l'utilisation des informations dans une base de données d'images. La classification d'une nouvelle observation est une méthode pour analyser les informations que l'image contient. Et puis, on va construire les relations de cette observation avec les individus existant afin de faciliter et d'effectuer des prédictions de sa classe. Le but de ce chapitre est de décrire les applications en classification d'images texturées et de reconnaissance de scène pour montrer l'intérêt des vecteurs descripteurs proposés dans les chapitres précédents en terme de robustesse vis-à-vis des méthodes de l'état de l'art. Ces applications sont basées initialement sur les algorithmes d'apprentissage supervisé et les distances de mesure de similarité pour pouvoir faire une comparaison de performances. Les définitions théoriques des classifieurs : k-plus proches voisins, machine à vecteurs de support, forêt aléatoire et des distances : Euclidienne, Chi2, la divergence de Jensen-Shannon sont présentées dans ce travail. Les comparaisons des performances de la classification avec les bases de données connues dans le domaine de la vision par ordinateur sont réalisées pour les différentes configurations des paramètres. Nous envisagerons les performances des approches proposées dans différents cas en considérant différentes variantes des méthodes d'extraction de points d'intérêt, des choix des classifieurs, des mesures de similarité et des fonctions de covariance des processus de Cox log-Gaussien.

Ce chapitre est organisé comme suit : nous présentons tout d'abord très brièvement l'état de l'art des méthodes des classifieurs supervisés dans la Section 4.2 et des distances de mesure de similarité dans la Section 4.3. Puis, l'application à la classification supervisée des textures réelles est présentée dans la Section 4.4 avec trois bases de données différentes : UIUC texture, Brodatz et KTH-Tips. Ensuite, la Section 4.5 présente les résultats expérimentaux obtenus pour l'application de reconnaissance de scène avec une base de données d'images des différentes scènes réelles. Finalement, la Section 4.6 donne des conclusions.

4.2 Méthodes de classification supervisée

Les algorithmes d'apprentissage supervisé tentent de trouver un classifieur, un modèle de prédiction ou une fonction mathématique, qui nous permettent d'expliquer le lien entre des données d'entrée et les classes existantes. De nombreuses méthodes de classifieurs existent dans la littérature [120, 121, 122, 123]. Dans cette sous-section, nous nous intéressons aux trois méthodes très populaires dans la communauté de la théorie de l'apprentissage : k-plus proches voisins, Machine à vecteurs de support, Forêt aléatoire. Les détails de ces méthodes sont exposés dans la suite.

4.2.1 k-plus proches voisins

L'algorithme de la classification supervisée des k-plus proches voisins (k-PP, k-nearest neighbors algorithm) est considéré dans ce travail en raison de sa simplicité et son efficacité de calcul [121]. Cet algorithme nécessite unique la définition d'une distance entre les éléments que l'on veut classifier. Nous considérons une base d'apprentissage de n éléments $z = \{(x_1, y_1), \dots, (x_n, y_n)\}$, où x_i est le vecteur descripteur de l'élément i et y_i est sa classe. Pour la classification d'une nouvelle observation x , nous calculons d'abord sa distance à tous les éléments dans la base d'apprentissage $d_i = \|x - x_i\|$. Les différentes mesures de similarité entre les éléments sont présentées dans la sous-section 4.3. Alors, nous obtenons donc n valeurs réelles et positives des distances d_i . De plus, nous trouvons les k plus petites valeurs parmi ces n valeurs d_i . Ces k valeurs correspondent aux k éléments $\{x_i, y_i\}$ de la base d'apprentissage qui sont les plus proches de la nouvelle observation x . En fin, la classe de x que nous souhaitons classifier est attribuée à la classe majoritaire y_i parmi ces k éléments. Si la taille de la base de données d'apprentissage est n , la complexité est alors en $O(n)$. La Figure 4.1 présente un exemple visuel de l'algorithme des k-plus proches voisins.

4.2.2 Machine à vecteurs de support

Une autre méthode de classification supervisée considérée dans ce travail est la méthode des machines à vecteurs support SVM (Support Vector Machine) introduit par Cortes et Vapnik [122]. Nous décidons d'utiliser les SVM dans ce travail parce qu'elles permettent une classification non linéaire et qu'elles ne dépendent pas de la dimension de l'espace de représentation des données. La construction d'un SVM, illustré dans la Figure 4.2, comporte deux étapes principales comme suit :

- La projection des données non linéaire de l'espace d'entrée vers un espace de redescription (espace de Hilbert) par une fonction noyau.
- La détermination d'un hyperplan de séparation optimale qui permet de maximiser la marge de séparation dans cet espace de redescription.

En ce sens, le détail des SVM est présenté comme suit. Nous considérons un couple (x_i, y_i) de variables aléatoires à valeurs dans $X \times Y$, où $X = \{x_i\}$ un espace des exemples d'entrée, $Y = \{-1, 1\}$. La formalisation présentée ici se limite à un problème de discrimination à deux classes pour classifier une nouvelle observation x . Cette formalisation

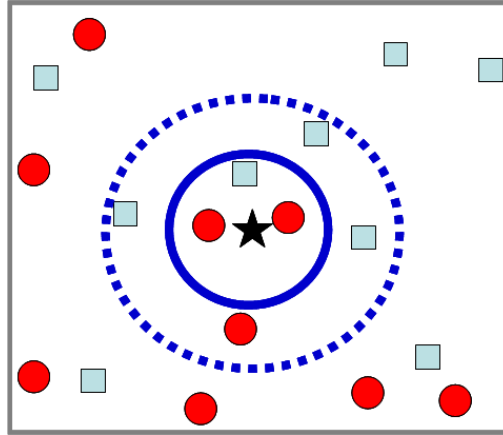


Figure 4.1 — Illustration de la classification par la méthode des k -plus proches voisins. Considérant une base de données de 18 éléments de deux classes : cercles et carrés, un nouvel élément étoile est attribué à la classe des cercles si on considère la valeur de $k = 3$, parce que ses trois plus proche voisins sont constitués de deux cercles et d'un carré. Par contre, si on considère ses $k = 7$ voisins, le nouvel élément est attribué à la classe des carrés car ses 7 premiers voisins sont constitués de 4 carrés pour 3 cercles.

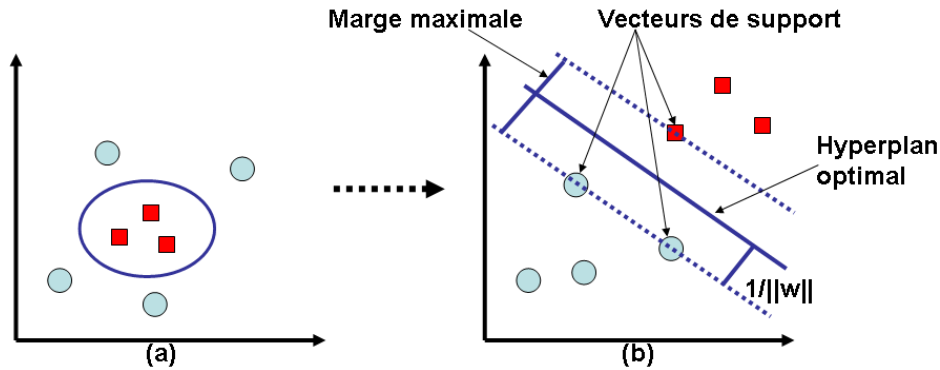


Figure 4.2 — Illustration du principe des SVM : La transformation d'un problème de séparation non linéaire des données de l'espace d'entrée (a) vers un problème de séparation linéaire dans un espace de redescription (b). La détermination d'un hyperplan de séparation optimale qui maximise la marge dans l'espace de redescription (b).

pourra aisément être généralisée au cas de multi-classes [124, 125]. Nous allons donc construire une fonction f qui, à toute entrée x_i , fait correspondre une classe $y_i = f(x_i)$. Une fonction discriminante h est obtenue par combinaison linéaire du vecteur d'entrée x par un vecteur de poids w et un biais b :

$$h(x) = w \cdot x + b \quad (4.1)$$

La classe est donnée par le signe de la fonction $h(x)$: $f(x) = \text{sign}(h(x))$. Sans perte de généralité nous posons que x est de classe 1 si $h(x) > 0$, sinon x est de classe -1 . La frontière de décision $h(x) = w \cdot x + b = 0$ est un *hyperplan*. Dans le cas où le problème de

classification admet une solution linéaire, il est à remarquer qu'il existe de nombreux hyperplans séparateurs, et *la marge* est la distance entre l'hyperplan de séparation et les échantillons les plus proches appelés *vecteurs support*. L'*hyperplan optimal* est celui qui maximise la marge. La distance entre un point quelconque x et l'hyperplan est donnée par :

$$d(x) = \frac{|w \cdot x + b|}{\|w\|} \quad (4.2)$$

Le problème de maximisation de la marge revient au problème de minimisation de $\|w\|$ sous contraintes. Pour la recherche de l'hyperplan optimal des SVM, deux types de formulations dit *primale* et *duale* des SVM s'expriment alors sous les formes suivantes :

- **La forme primale** dépend seulement de w et b .

$$\begin{cases} \min(\frac{1}{2}\|\mathbf{w}\|^2) \\ \forall i, y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \end{cases} \quad (4.3)$$

- **La forme duale** est une transformation de la formulation primale en utilisant les multiplicateurs de Lagrange optimaux α_i :

$$\begin{cases} \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \forall i, \alpha_i \geq 0 \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (4.4)$$

En utilisant les multiplicateurs de Lagrange, l'hyperplan optimal peut s'écrire sous la forme :

$$h(x) = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x}_j + b \quad (4.5)$$

Cependant, il n'existe pas en général des hyperplans permettant de séparer deux classes de données. Des solutions prenant en compte des fonctions de séparation non linéaires devront être trouvées. Dans le cadre des SVM, une solution est trouvée en projetant les données d'entrées dans un nouvel espace de dimension supérieure où une séparation linéaire des deux classes est possible. Pour ce faire, nous appliquons une transformation non linéaire Φ aux vecteurs d'entrée x_i tel que $x_i \in X$ et $\Phi(x_i) \in R$. La théorie montre qu'il existe des fonctions noyaux ξ l'espace de départ X , qui représentent les produits scalaires dans le nouvel espace de redescription :

$$\xi(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \quad (4.6)$$

La condition de représentation des produits scalaires de l'espace R par des noyaux de l'espace X est garantie par les conditions de Mercer [124, 125]. Le Tableau 4.1 donne des exemples de noyaux standards satisfaisant les conditions de Mercer. En pratique, il faudra choisir de manière empirique le noyau qui permet de représenter l'espace où peut se faire la classification optimale. Ceci constitue un inconvénient de l'utilisation des SVM. De plus, la complexité d'un algorithme SVM est cubique $O(dn^3)$, où n est le nombre de données d'apprentissage et d est la dimension des données.

Linéaire $\xi(\mathbf{x}_i \cdot \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$	Polynomial $\xi(\mathbf{x}_i \cdot \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^\sigma$ ou $(c + \mathbf{x}_i \cdot \mathbf{x}_j)^\sigma$
Gaussien $\xi(\mathbf{x}_i \cdot \mathbf{x}_j) = \exp\left(\frac{-\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{\sigma^2}\right)$	Sigmoïde $\xi(\mathbf{x}_i \cdot \mathbf{x}_j) = \tanh(\sigma(\mathbf{x}_i^T \mathbf{x}_j) + \sigma_0)$
Fourier régularisé $\xi(\mathbf{x}_i \cdot \mathbf{x}_j) = \cosh\left(\pi - \frac{\ \mathbf{x}_i - \mathbf{x}_j\ }{\sigma}\right)$	Hardy $\xi(\mathbf{x}_i \cdot \mathbf{x}_j) = \frac{1}{(\mathbf{x}_i^T \mathbf{x}_j + 1)^\sigma}$

Tableau 4.1 — Différentes formes des fonctions noyaux.

4.2.3 Forêt aléatoire

Les forêts aléatoires RF (anglais Random Forest) ont été introduites par Breiman en 2001 [123]. Le principe de la méthode des forêts aléatoires est une combinaison des méthodes d'arbre binaire avec des méthodes d'ensemble dit bagging sur des stratégies aléatoires. Ici, nous présentons une brève définition de la méthode d'ensemble : bagging et la construction des arbres binaires avant de décrire en détails la méthode des forêts aléatoires.

a) Bagging :

La méthode du Bagging (une contraction des mots "**B**ootstrap and **A**ggregating" en anglais) a été introduite par Breiman en 1996 [126]. Dans un contexte de classification, le principe du Bagging est de construire chaque classifieur à partir d'un tirage indépendant de plusieurs échantillons bootstrapés, d'appliquer la règle de base sur chacun d'eux pour obtenir une collection de prédicteurs, et enfin d'agréger ces prédicteurs de base. Soit Y une variable à expliquer, p variables explicatives $X = \{X_1, \dots, X_p\}$ et ϕ un modèle de prédiction sur un échantillon $z = \{(x_1, y_1), \dots, (x_n, y_n)\}$. On considère B échantillons bootstrapés $\{z_1, \dots, z_B\}$ d'individus étiquetés, ces échantillons sont issus de z par tirage aléatoire avec remise. Puis, on construit un modèle ϕ_{z_i} sur chacun des échantillons z_i . Enfin, on prédit Y en agrégeant les différentes décisions sur chacun des z_i en calculant la moyenne $\phi(X) = \frac{1}{B} \sum_{i=1}^B \phi_{z_i}(X)$ si Y est une variable quantitative, ou en faisant un vote à la majorité parmi les $\phi_{z_i}(X)$ si Y est une variable qualitative.

Dans ce contexte, on peut estimer assez naturellement l'erreur de prédiction par bootstrap "out-of-bag", ce qui permet de juger des propriétés de généralisation du modèle ou de prévenir d'un sur-ajustement [127].

b) Arbre binaire :

Ici, nous considérons une méthode d'arbre de décision dit CART (anglais Classification And Regression Tree) qui a été introduite par Breiman et al. en 1984 [128]. Le principe général des CART est de construire des prédicteurs par arbre à partir d'une partition binaire réursive de l'espace d'entrée, puis de déterminer une sous-partition optimale pour la prédiction.

La Figure 4.3 présente un exemple de la création d'un arbre décision binaire. Partant de la racine de l'arbre qui contient toutes les observations de l'échantillon d'apprentissage, la première découpe du partitionnement divise cette racine en deux noeuds fils

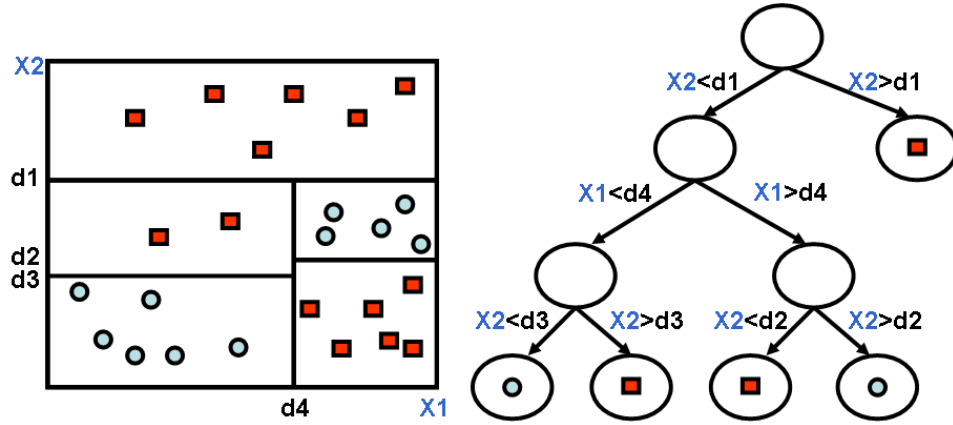


Figure 4.3 — Illustration de la partition en binaire et l'arbre CART associé.

et ainsi de suite. La forme de découpe est donnée par : $\{x_i \leq d\} \cup \{x_i > d\}$ qui signifie que tous les individus avec la valeur de la variable x_i plus petite que d vont dans un noeud fils, et tous ceux avec la valeur plus grande que d vont dans un autre noeud fils. Pour une application de classification, on cherche la meilleure découpe qui minimise l'indice de Gini des noeuds fils qui est défini par $\sum_{i=1}^L p_t^i(1 - p_t^i)$, où L est le nombre de classe et p_t^i est la proportion d'observations de classe i dans le noeud t .

c) Forêt aléatoire :

La méthode de forêt aléatoire est une amélioration d'un bagging dans le cas spécifique d'un arbre binaire, par l'ajout d'un aspect aléatoire, avec l'objectif de rendre les modèles des arbres construits plus indépendants entre eux. Cette propriété permet de rendre le vote plus efficace. Le principe de cette méthode est le suivant :

Soit $z = \{(x_1, y_1), \dots, (x_n, y_n)\}$ un échantillon d'apprentissage, où x_i sont les vecteurs explicatifs de dimensionnalité p et les y_i sont les variables à expliquer. Nous notons par B le nombre d'arbres formés dans la forêt aléatoire. Pour chaque $b = \{1, \dots, B\}$, on fait :

- Tirer un échantillon bootstrap aléatoire z_b avec remise parmi z .
- Estimer un arbre de décision binaire ϕ_b sur z_b avec choix aléatoire des variables.
- Limiter la construction de chaque noeud de chaque arbre, en tirant uniformément q variables parmi p pour former la décision associée au noeud.

En fin d'algorithme, la décision de la forêt se fait à la majorité des votes entre B arbres ϕ_b . Empiriquement, Breiman propose un choix optimal pour $q = \sqrt{p}$.

4.3 Mesures de similarité

Pour des applications en classification d'images, les choix de classifieurs SVM et k-PP se sont portés vers une mesure de similarité entre des vecteurs descripteurs d'images qui soient les plus appropriées. Dans la littérature, plusieurs méthodes statistiques et géométriques permettent d'exploiter des mesures de similarité pour résoudre des problèmes de classification, d'apprentissage ou de fusion d'informations [129, 130, 131].

On peut citer la distance de Malahanobis, la distance de Bhattacharya, la distance Euclidienne, la distance du Chi2, la divergence de Kullback-Leibler, la divergence de Jensen-Shannon, etc. Dans ce chapitre, dans un contexte de classification, pour mesurer la similarité des descripteurs nous considéreront les distances suivantes :

- La distance Euclidienne :

$$d_E(H, K) = \sum_i |h_i - k_i|^2 \quad (4.7)$$

- La distance du Chi2 χ^2 :

$$d_{\chi^2}(H, K) = \sum_i \frac{(h_i - m_i)^2}{m_i} \quad (4.8)$$

- La divergence de Jensen-Shannon :

$$d_J(H, K) = \frac{1}{2} \sum_i \left(h_i \log \frac{h_i}{m_i} + k_i \log \frac{k_i}{m_i} \right) \quad (4.9)$$

où $H = \{h_i\}, K = \{k_i\}$ sont les vecteurs descripteurs d'images donnés par des distributions de probabilité et $m_i = \frac{h_i + k_i}{2}$.

4.4 Applications à la classification des images texturées naturelles

Nous considérons des applications à la caractérisation et la classification supervisée des images texturées qui nous permettent d'évaluer et comparer les performances de nos approches à différentes techniques de l'état de l'art. Trois jeux de données sont choisis : UIUC texture¹, Brodatz texture² et KTH-TIPs texture³. Les étapes principales de ces applications sont :

- La détection des points d'intérêt et l'extraction de leurs caractéristiques (voir la Figure 1.2). Les différents types de points d'intérêt sont envisagés.
- La construction des mots visuels basée sur l'application d'algorithme de clustering de type des k-moyennes sur les caractéristiques des points d'intérêt (voir la Figure 2.1).
- L'estimation des paramètres du modèle de Cox log-Gaussien pour définir un vecteur descripteur d'image texturée (voir la Figure 2.3).
- La réalisation du modèle d'apprentissage supervisé.

L'apprentissage des modèles de classification est réalisé à partir d'un sous-ensemble d'images de chaque classe dans la base de textures. N_t images d'apprentissage sont

1. <http://www.cs.unc.edu/~lazechnik/>

2. <http://www.ux.uio.no/~tranden/brodatz.html>

3. <http://www.nada.kth.se/cvap/databases/kth-tips/documentation.html>

choisies aléatoirement parmi les N images texturées de chaque classe. Les $(N - N_t)$ images restantes par classe sont considérées comme les images de test. Cette étape d'apprentissage est répétée t fois pour évaluer les performances de la classification, en termes de taux moyens et d'écart types de bonne classification.

Etant donné un ensemble de vecteurs descripteurs des approches proposées, les méthodes de classification de texture que nous proposons sont comparées aux méthodes de l'état de l'art suivantes. Deux méthodes classiques : Filtre de Gabor [6, 17], matrice de cooccurrence [16, 7] sont choisies pour montrer l'effet des algorithmes reposant sur les signatures élémentaires invariantes de base dans les images. En outre, nous considérons trois autres méthodes : algorithme Sac de mots [8], algorithme de Xu [10], et de Zhang [11] pour évaluer la pertinence de notre contribution par rapport aux autres techniques de l'état de l'art basées sur les signatures élémentaires invariantes dans les images. De plus, nous comparons des résultats expérimentaux obtenus par nos approches aux résultats d'autres techniques existantes, qui consistent dans les analyses des informations spatiales des points d'intérêt : algorithme de Ling [9] et de Savarese [25]. Les détails de différentes configurations des paramètres de chaque méthode sont donnés par :

- **Filtre de Gabor** [6, 17] : Les attributs de Gabor sont extraits par les statistiques de la réponse à un espace d'échelle des filtres de Gabor. Pour ce type de filtre, nous considérons deux paramètres qui sont : la fréquence fondamentale $f_G = \{0, 4, 8\}$ et l'orientation $\theta_G = \{0, \pm\frac{\pi}{2}, \pi\}$.
- **Matrice de cooccurrence** [16, 7] : Les matrices de co-occurrence sont des histogrammes bidimensionnels qui représentent une estimation de la probabilité conjointe de deux pixels dans un voisinage. Nous considérons la distance $d_M = \{1, 2, 4\}$ qui permet d'évaluer la périodicité des structures photométriques et le paramètre angulaire $\theta_M = \{\frac{k\pi}{8}, k = 0 : 7\}$ qui sert à caractériser la direction.
- **Sac de mots** [8] : L'algorithme SDM exploite les statistiques d'occurrence des différentes catégories de mots visuels dans l'image. Les différents nombres de catégories de mots visuels $k = \{60, 120, 150\}$ sont envisagés.
- **Les algorithmes de Ling** [9] **et de Savarese** [25] : Les deux méthodes sont utilisées pour exploiter un histogramme de statistiques de cooccurrence des paires des catégories de points d'intérêt dans un voisinage étudié. Dans ce travail, nous calculons les histogrammes d'un ensemble des cercles étudiés avec les rayons $r = \frac{M}{2} \ln(x)$, où $M = \min(H, W)$ la valeur minimal de largeur et hauteur d'image ; x varié de 1 à $\exp(1)$ selon un pas linéaire de 0,05.
- **L'algorithme de Xu** [10] : Cet algorithme propose une description multifractale d'image texturées qui ont les propriétés invariantes du changement de contraste local, de la déformation non-rigides et de transformation affine. Nous envisageons les différentes configurations des paramètres de cet algorithme : le niveau de densité $ind = \{1, 8\}$, la dimension de MFS $f = \{16, 64\}$ et le niveau d'itération $ite = \{8, 10\}$.
- **L'algorithme de Zhang** [11] : C'est une combinaison de deux détecteurs de points d'intérêt (Har+Lap) et deux descripteurs locaux (SIFT+SPIN) pour une application de la classification des images texturées qui utilise un classifieur SVM et la distance EMD (Earth Move Distance). Les mêmes valeurs de nombres

de catégories de mots visuels $k = \{60, 120, 150\}$ que l'algorithme de SDM sont envisagés.

- **Les approches proposées : statistiques descriptives de cooccurrence (SSC) [104] , modèle de Cox log-Gaussien (LGCM) [119]** : Pour ces applications de classification d'images texturées, nous considérons les points d'intérêt comme l'information de base pour les approches que nous proposons. Comme les autres méthodes basées sur les points d'intérêt dans cette expérimentation, les nombres de catégories de mots visuels $k = \{60, 120, 150\}$ sont envisagés ici. En outre, le nombre de paires de points d'intérêt $k^* = \{60\}$ est considéré pour la procédure de réduction de la dimension de la taille de descripteur proposé. Les procédures d'adaptation d'échelle et de la correction du bord sont appliquées.

De plus, les configurations suivantes des classifieurs sont considérées par :

- Le classifieur k-PP est implémenté avec les différents valeurs de k qui dépends au nombre d'images d'apprentissage N_t .

$$\begin{cases} k = N_t & \text{if } N_t \leq 5 \\ k = 5 & \text{if } N_t > 5 \end{cases} \quad (4.10)$$

- La stratégie *un contre tous* (one-versus-all) est exploitée pour le classifieur SVM multi-classe avec un noyau Gaussien considéré :

$$G(H, K) = \exp\left(-\frac{d(H, K)}{2\sigma^2}\right) \quad (4.11)$$

- Nous utilisons le classifieur RF avec les paramètres suivants : $ntree = 500$ est le nombre de arbres de la forêt, et $mtry = 200$ est le nombre de variables tirées aléatoirement à chaque noeud.

4.4.1 UIUC

La première évaluation exploite la base de données des textures d'UIUCtex qui comporte 25 classes avec des textures différentes. Chaque classe est composée de $N = 40$ images de taille 640x480 pixels. Des exemples d'images de la base UIUCTex sont présentés dans la Figure 4.4. Chaque classe comporte des échantillons avec des variations de contraste, d'échelle et de rotation.

La performance moyenne de classification obtenue par les différentes méthodes sur la base de textures d'UIUC est évaluée par le choix aléatoire de $t = 50$ fois de $N_t = \{1, 5, 10, 15, 20\}$ images d'apprentissage parmi $N = 40$ images texturées d'UIUC. Tout d'abord, les évaluations expérimentales sont réalisées pour différentes configurations des paramètres du descripteur LGCM. LGCM sont étudiées en fonction du type de points d'intérêt, du choix des classifieurs et des mesures de similarité pour l'apprentissage supervisé et de la sélection des fonctions de covariance. On comparera ensuite le descripteur LGCM aux autres méthodes connues dans l'état de l'art.

Choix du classifieur et de la mesure de similarité

Les résultats en terme des taux moyens et d'écarts types de bonne classification obtenus par le descripteur LGCM pour différents modèles d'apprentissage supervisé

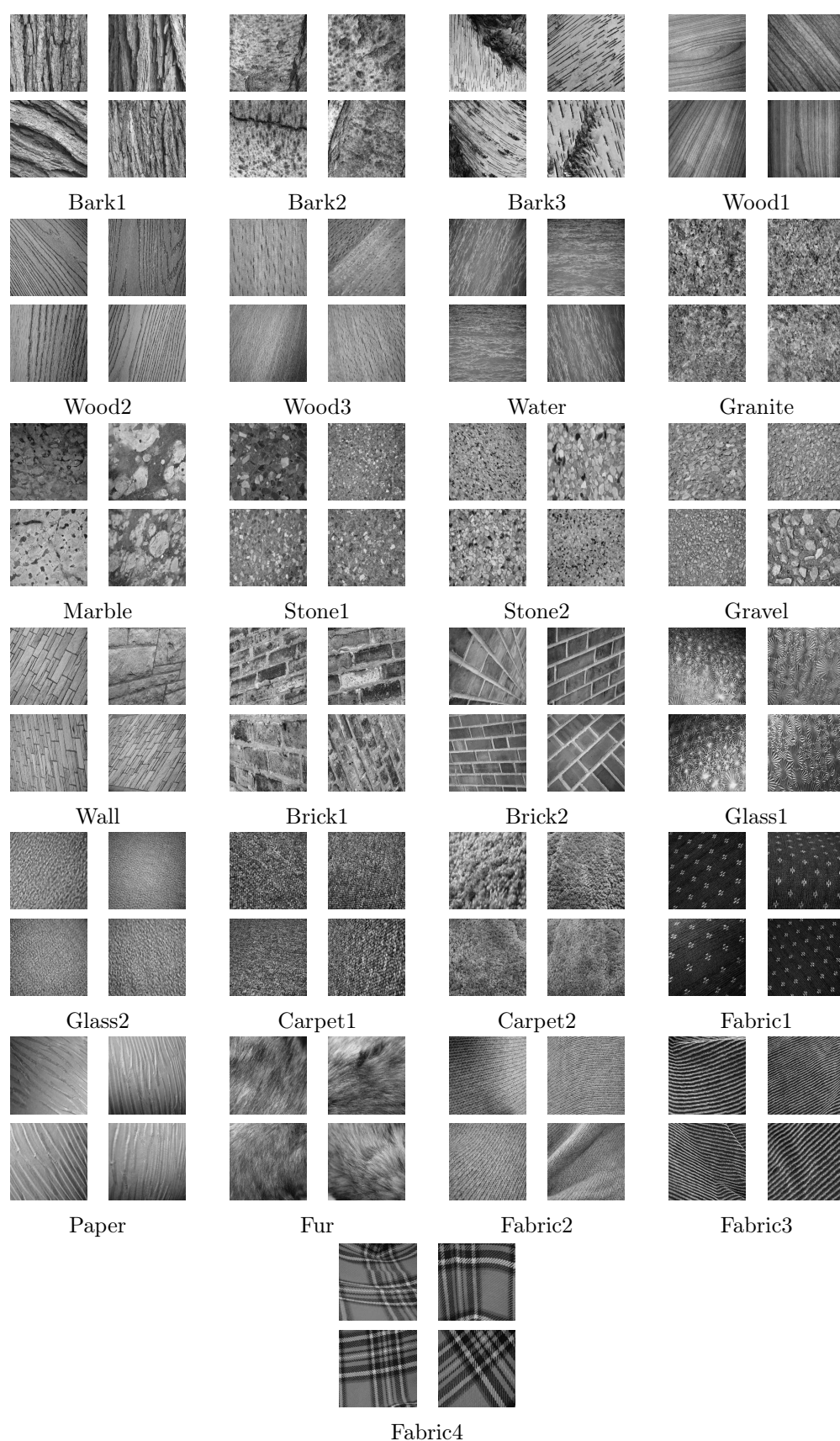


Figure 4.4 — Exemple d'images de chaque classe de texture de la base de donnée UIUC.

k-NN			SVM			RF
d_E	d_{χ^2}	d_J	d_E	d_{χ^2}	d_J	
95.95%	96.32%	97.21%	96.18%	96.40%	97.35%	97.84%

Tableau 4.2 — Comparaison des performances du descripteur LGCM pour la classification supervisée d’images de la base de données UIUCTex dans les cas des différents classifieurs (k-NN, SVM, RF) et des différentes mesures de similarité (Euclidienne, χ^2 , divergence de Jensen-Shannon), quand 20 images d’apprentissage sont considérées.

N_t	Gaussian	Cardinal sine	Hyperbolique
1	75.21±1.75	75.15±1.67	75.03±1.81
5	91.96±1.13	91.63±1.17	91.32±1.19
10	95.42±0.71	95.35±0.75	94.72±0.85
15	96.87±0.65	96.17±0.63	95.43±0.71
20	97.84±0.32	97.15±0.42	96.85±0.38

Tableau 4.3 — Comparaison des performances de reconnaissance d’images texturées UIUCTex pour différentes fonctions de covariance pour la définition du descripteur LGCM.

sont présentés dans le Tableau 4.2. Trois classifieurs : k-PP, SVM, RF et trois mesures de similarité : distance Euclidienne, distance de Chi2, divergence Jensen-Shannon sont envisagées dans cette évaluation. L’estimation des paramètres du modèle de Cox log-Gaussien avec une fonction de covariance Gaussienne et les points d’intérêt DoG+Sift est exploitée. En considérant les mesures de similarité, la divergence Jensen-Shannon améliore les performances de la classification avec un gain approximatif de plus de 1% par rapport à la distance χ^2 et la distance Euclidienne quand 20 images d’apprentissage sont considérées. Les différences dans les performances entre ces classifieurs ne sont pas évidentes pour cette base de données. Toutefois, le classifieur RF est légèrement plus robuste et plus stable que les classifieurs k-PP et SVM.

Choix du modèle de covariance du modèle LGCM

Ensuite, la deuxième évaluation des performances de notre descripteur LGCM dans différents cas d’évaluation des fonctions de covariance pour l’estimation des paramètres du modèle proposé est réalisée. La synthèse des résultats est donnée dans le Tableau 4.3. Les points d’intérêt DoG+Sift et le classifieur RF sont considérés dans cette évaluation. Peu de différences sur les performances obtenues entre les modèles de covariance, où le taux moyen et l’écart type de bonne classification du modèle de covariance Gaussien est de 97.84%±0.32 contre 97.15%±0.42 et 96.85%±0.38 respectivement pour les modèles de covariance hyperbolique et de sinus cardinal quand 20 images d’apprentissage sont considérées.

Invariance en échelle et réduction de dimension

Nous rapportons aussi les évaluations des performances du descripteur LGCM dans les cas d’application de méthodes pour résoudre de problèmes d’invariance en échelle et de grande taille de la dimension des descripteurs proposés. Les résultats des performances de classification dans le Tableau 4.4 mettent en évidence l’intérêt de la mé-

	sans procédure d'adaptation d'échelle	sans procédure de réduction de la taille	modèle complet
1	73.27±2.05	75.65± 1.82	75.21±1.75
5	89.12±1.27	91.67± 1.15	91.96±1.13
10	94.76±1.11	95.15±0.75	95.42±0.71
15	95.12±0.81	96.27±0.71	96.87±0.65
20	95.89±0.54	97.12±0.35	97.84±0.32

Tableau 4.4 — Comparaison des performances du descripteur LGCM pour la classification supervisée d'images texturées de la base UIUCTex dans les cas l'utilisation ou non des procédures d'adaptation d'échelle et de réduction de la taille du descripteur.

N_t	DoG +Sift[21]	FH +Surf[19]	FH +Brief[20]	(Hes-Lap) +Daisy[22]	(Har-Lap) +(Sift-Spin)[11]
1	75.21±1.75	75.05±1.94	75.43±1.71	75.18±1.69	74.85±1.87
5	91.96±1.13	90.73±1.11	91.42±1.23	92.13±1.19	91.15±1.41
10	95.42±0.71	95.15±0.91	95.22±0.85	95.47±1.08	95.23±0.72
15	96.87±0.65	96.14±0.63	96.43±0.51	96.75±0.58	96.37±0.61
20	97.84±0.32	96.75±0.41	97.25±0.34	97.67±0.35	97.14±0.37

Tableau 4.5 — Comparaison des performances de reconnaissance d'images textures de la base UIUCTex pour différents types de points d'intérêt pour la définition du descripteur LGCM : DoG+Sift, FH+Surf, (Har-Lap)+(Sift-Spin), (Hes-Lap)+Daisy, FH+Brief. Le classifieur RF est considéré dans cette expérimentation.

N_t	1	5	10	15	20
Filtre de Gabor [6]	31.22±3.14	45.14±2.54	57.37±1.93	61.25±1.52	67.78±1.28
Matrice de Cooc. [7]	45.33±3.03	61.58±2.14	70.67±1.72	73.85±1.34	80.12±1.30
SDM[8]	67.25±2.75	76.38±2.15	81.12±1.45	86.35±1.20	91.28±1.15
Ling[9]	67.62±2.93	78.42±2.33	84.14±1.72	86.38±1.25	91.87±1.38
Xu[10]	61.14±2.90	83.33±2.07	89.68±1.65	91.34±1.45	93.85±1.31
Zhang[11]	72.53±2.45	88.62±1.33	93.17±1.15	95.33±0.98	96.67±0.93
SSC[104]	75.66±1.65	91.67±0.93	94.33±0.78	96.54±0.53	97.34±0.25
LGCM	75.21±1.75	91.96±1.13	95.42±0.71	96.87±0.65	97.84±0.32

Tableau 4.6 — Taux moyens et écarts types de bonne classification des approches proposées pour les textures UIUCTex en comparaison aux méthodes de l'état de l'art : Filtre de Gabor [6], Matrice de cooccurrence [7], sac de mots (SDM) [8], les statistiques de cooccurrence des paires des catégories de points d'intérêt [9], la description multifractale [10], la combinaison de différents types de points d'intérêt [11], statistiques descriptives spatiales de cooccurrence (SSC), modèles Cox log-Gaussien (LGCM).

thode pour résoudre les problèmes d'invariance en échelle avec un gain de 2% obtenu (97.12% et 97.84% vs. 95.89%) quand 20 images d'apprentissage sont considérées. En outre, l'amélioration des performances de la classification de la procédure de réduction de la taille de la dimension des descripteurs est obtenue quand le nombre d'images d'apprentissage est supérieur 1.

Choix de méthodes d'extraction de points d'intérêt

Nous considérons aussi la réalisation du modèle de Cox log-Gaussien lors de l'utilisation des différentes méthodes d'extraction des points d'intérêt pour une application de classification des images texturées de la base de données de UIUCTex. Les résultats des performances de classification pour différents types de points d'intérêt sont présentés dans la Figure 4.5. Le plus efficace descripteur est le descripteur Dog+SIFT avec un léger gain de 0.15% à 1.2% de performance de classification par rapport aux autres méthodes quand 15 ou 20 images d'apprentissage sont considérées. Par contre, on peut trouver que le descripteur (Hes-Lap)+Daisy présente un comportement légèrement meilleur que les autres méthodes avec les taux moyens et les écarts types de bonne classification de $92.13\% \pm 1.19$ et $95.47\% \pm 1.08$ (respectivement) quand 5 et 10 images d'apprentissage sont considérées. Ces résultats pourraient s'expliquer par les plus grandes densités de points d'intérêt observées dans les images texturées de UIUCTex pour les deux méthodes Dog+Sift et (Hes-Lap)+Daisy.

Comparaison avec les méthodes de l'état de l'art

La dernière évaluation sur la base de données de UIUCTex est réalisée pour une comparaison des performances des descripteurs proposés avec différentes techniques de l'état de l'art. La synthèse des résultats de cette évaluation est donnée en Tableau 4.6. En considérant les résultats des approches classiques, tels que filtres de Gabor, matrice de cooccurrence et multifractal (méthode de Xu), nos approches donnent un gain significatif de performance de classification : respectivement, $67.78\% \pm 1.28$, $80.12\% \pm 1.30$, $93.85\% \pm 1.31$ vs. $97.84\% \pm 0.32$ avec notre méthode pour $N_t = 20$. Ces expériences démontrent clairement l'intérêt d'hériter de la robustesse de points d'intérêt visuels pour une application de la reconnaissance de textures en termes d'invariance des distorsions géométriques de l'image et de changement du contraste. En outre, les gains de plus de 6.5% et de 1.5% respectivement à comparer avec les méthodes l'état de l'art basées sur la détection des points d'intérêt : SDM et le schème de Zhang ont été obtenus avec $N_t = 20$. Ils montrent la pertinence des modèles statistiques pour d'analyser les informations d'organisation spatiale des points d'intérêt. D'autre part, comparativement aux performances des approches basées sur l'analyse spatiale des points d'intérêt : les méthodes de Ling [9], de Savarese [25] et le descripteur proposé SSC [104], le descripteur proposé LGCM se révèle légèrement plus robuste avec $97.84\% \pm 0.32$ vs. $91.87\% \pm 1.38$, $97.34\% \pm 0.25$ respectivement, quand 20 images d'apprentissage sont considérées.

4.4.2 Brodatz

Nous utilisons 111 images texturées de l'album Brodatz pour la deuxième évaluation des performances de classifications des descripteurs considérés. A partir de ces différences images texturées, nous créons une base de données de 111 différentes classes

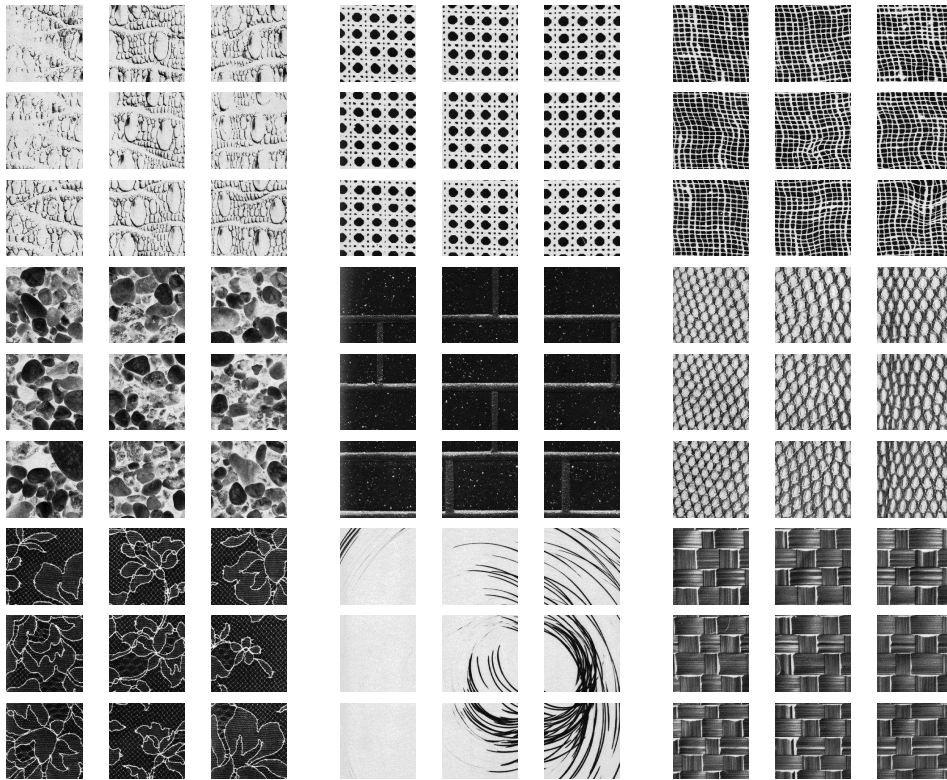


Figure 4.5 — Exemples d'images de chaque classe de la base de données Brodatz.

N_t	1	3
Filtre de Gabor [6]	78.52 ± 1.72	85.14 ± 1.41
Matrice de Cooc. [7]	75.42 ± 1.73	83.22 ± 1.04
SDM[8]	83.16 ± 1.50	92.78 ± 0.91
Ling[9]	84.33 ± 1.63	93.17 ± 0.87
Xu[10]	85.95 ± 0.91	93.41 ± 0.73
Zhang[11]	86.63 ± 1.05	94.34 ± 0.43
SSC[104]	87.67 ± 0.81	95.67 ± 0.33
LGCM	88.81 ± 0.92	96.14 ± 0.41

Tableau 4.7 — Taux moyens et écarts types de bonne classification des approches proposées en comparaison à des techniques d'état de l'art pour la base de données d'images de l'album Brodatz.

de textures. Chaque classe contient 9 images de la taille de 170x170 pixels qui sont extraites d'images dans l'album (voir la Figure 4.5). Il faut également noter que, dans cette base de données, les informations concernant la variance en échelle et changement d'illuminations ne sont pas connues.

Pour cette base de données, l'évaluation de performance est réalisée par une comparaison des performances de notre approche LGCM avec les mêmes méthodes de l'état de l'art que nous avons considérées pour la base de données de UIUCTex. L'apprentissage supervisé considère un sous-ensemble de 1 ou 3 images d'apprentissage sélectionnées aléatoirement avec une répétition de $f = 50$ fois pour chaque classe. Les points d'intérêt

N_t	5	10	20	30	40
Filtre de Gabor [6]	62.32 ± 2.41	74.67 ± 2.04	82.65 ± 1.69	87.75 ± 1.27	89.87 ± 0.86
Matrice de Cooc. [7]	62.83 ± 2.42	73.34 ± 2.22	80.45 ± 1.67	85.83 ± 1.30	88.95 ± 0.77
SDM[8]	64.42 ± 2.81	75.83 ± 2.12	81.12 ± 1.45	88.35 ± 1.20	90.18 ± 0.65
Ling[9]	71.07 ± 2.63	76.48 ± 2.27	83.47 ± 1.48	88.58 ± 1.17	91.15 ± 1.05
Xu[10]	72.63 ± 2.45	81.42 ± 1.95	87.18 ± 1.53	89.95 ± 1.35	91.33 ± 0.97
Zhang[11]	78.17 ± 2.35	85.42 ± 1.78	90.28 ± 1.31	92.15 ± 1.05	94.33 ± 0.67
SSC[104]	81.34 ± 1.93	87.38 ± 1.47	92.15 ± 1.26	93.67 ± 0.97	95.09 ± 0.41
LGCM	80.17 ± 2.15	86.96 ± 1.53	92.42 ± 1.11	94.33 ± 0.87	95.74 ± 0.45

Tableau 4.8 — Taux moyens et écarts types de bonne classification des approches en comparaison aux techniques d'état de l'art pour la base d'images KTH-Tips.

DoG+Sift, la fonction covariance de type Gaussien et le classifieur RF sont considérés dans cette évaluation. Les résultats des performances de classification des différentes méthodes sont rapportés dans le Tableau 4.7. Notre approche LGCM atteint jusqu'à $96.14\% \pm 0.41$ en terme de taux moyen et d'écart type de bonne classification. Une amélioration significative de 11% est obtenue d'une comparaison de notre descripteur LGCM avec les filtres de Gabor. Toutes les performances de classification des autres méthodes sont inférieures à 95,67%. Le descripteur proposé se révèle légèrement plus robuste et plus stable que les méthodes de Ling [9], de Savarese [25] et le descripteur proposé SSC [104], avec $88.81\% \pm 0.92$ vs. $86.63\% \pm 1.05$ et $87.67\% \pm 0.81$ où $N_t = 1$.

4.4.3 KTH-TIPs

La troisième évaluation des performances de classification exploite la base de données des textures KTH-TIPs. Cette base de données comporte 10 classes d'images de différents matériaux. Chaque classe contient 81 images de 200x200 pixels (sauf les exceptions de quelques images de deux classes : brown-bread et cracker). Les images texturées de cette base de données sont invariante aux directions d'illumination, du changement d'échelle et des transformations photométriques (voir la Figure 4.6).

Comme l'évaluation pour la base de données Brodatz, nous considérons ici une seule comparaison des performances de la classification de notre approche LGCM avec les autres approches de l'état de l'art dans le but de montrer la pertinence du descripteur proposé dans les différents types d'images texturées. Le modèle d'apprentissage supervisé est réalisé avec le classifieur RF et le nombre d'images d'apprentissage sélectionnées aléatoirement $N_t = \{5, 10, 20, 30, 40\}$. Cette étape d'apprentissage est répétée en $f = 50$ fois. Les points d'intérêt DoG+Sift et la fonction de covariance gaussienne sont considérés dans cette évaluation. Les résultats des performances de classification sont présentés dans le Tableau 4.8. L'approche proposée LGCM a obtenu un gain environ de 1% par rapport à la meilleure performance de toutes les autres approches, le descripteur proposé SSC [104], $95.74\% \pm 0.45$ vs. $95.09\% \pm 0.41$ quand 40 images d'apprentissage sont considérées.

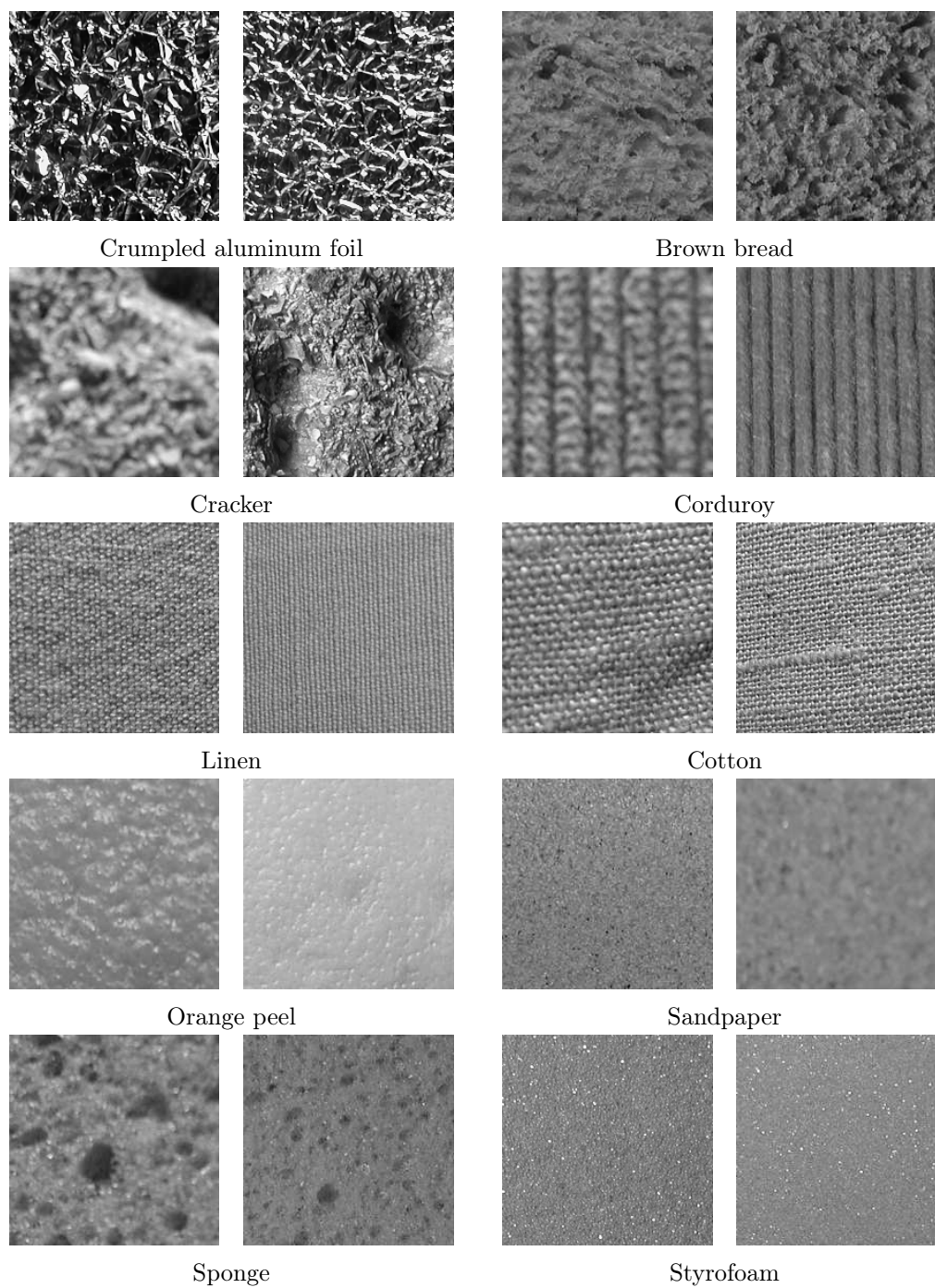


Figure 4.6 — Exemples d'images de chaque classe de la base de donnée KTH-TIPs.

4.5 Reconnaissance des catégories d'images de scène naturelle

L'analyse de la sémantique de l'image est une problématique importante pour de nombreux sujets de recherche en robotique [24] et en vision par ordinateur [132, 5]. Plusieurs approches ont été étudiées dans le but de chercher le lien entre les caractéristiques de l'image et sa description sémantique. Les travaux récents abordent de différentes manières l'analyse et la caractérisation des composantes locales de l'image, comme les points d'intérêt, les objets ou les formes. Elles ont démontré leur pertinence dans les applications à la reconnaissance d'images [5, 44, 133, 134]. Ici, nous étudions un nouveau descripteur d'analyse d'image de la scène présentant des qualités d'invariance basé sur la caractérisation des informations visuelles et spatiales d'un ensemble des formes dans l'image. Les étapes principales de cette application de la reconnaissance des catégories d'images de la scène naturelle sont données par :

- La détection d'un ensemble des formes dans l'image et le calcul d'un vecteur caractéristique pour chaque forme (voir les Figures 1.7 et 4.7a). Un arbre des formes et les descripteurs des formes sont extraits ici par l'algorithme FLST basée sur l'analyse des lignes de niveau et l'algorithme IDSC.
- La construction des mots visuels basée sur l'application d'algorithme de clustering de type de k-moyennes sur les caractéristiques des formes détectées (voir la Figure 4.7b).
- L'estimation paramétrique du modèle de Cox log-Gaussien pour une définition d'un vecteur descripteur d'image (voir la Figure 4.7c).
- La construction d'un modèle de classification par apprentissage supervisé.

Il faut également noter que les procédures de correction des effets de bord et de réduction de la complexité du vecteur descripteur pourraient s'appliquer de manière originale par rapport à la technique de l'analyse des points d'intérêt présentée dans les chapitres précédents. Par contre, on applique ici une méthode pour tenir compte de l'invariance en échelle en normalisation par un facteur η . Cette valeur η est calculée à partir de la moyenne des taux de surfaces des formes et de ceux de leurs parents à tous les niveaux sur la carte topographique comme suit :

$$\eta = \frac{1}{n_s} \sum_{i=1}^{n_s} \left(\frac{\varpi_i^l}{\frac{1}{l} \sum_{j=0}^{l-1} \varpi_i^j} \right) \quad (4.12)$$

où ϖ_i^l est la surface de la forme i à niveau de l d'arbre des formes ($l = 0$ est le niveau de la racine), n_s est le nombre total des formes dans l'image. Cette valeur η est considérée ici comme le facteur d'échelle associée à chaque image pour la détermination des valeurs réelles de rayons r des boucles d'études dans l'analyse nos statistiques spatiales proposées.

Nous considérons ici une base de données de 15 classes de scènes naturelles qui est fournie par Lazebnik et al. [5]. Cette base de données contient les différentes catégories

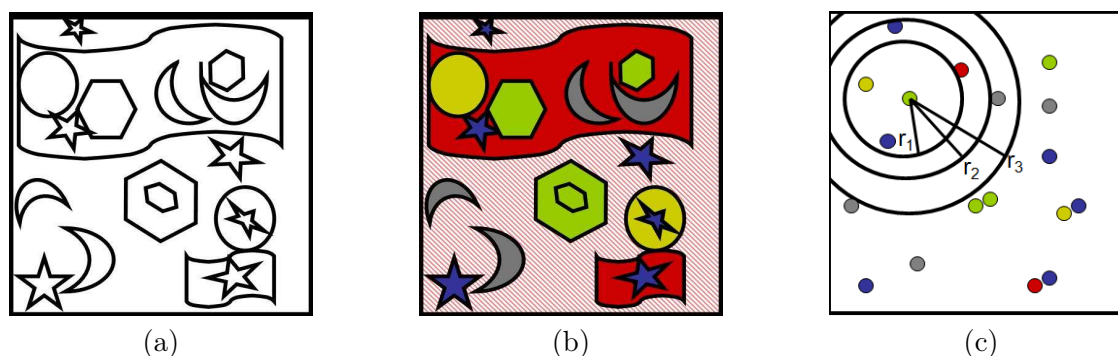


Figure 4.7 — Etapes principales de cette application : la détection des formes dans l'image (a), la détermination des mots visuels des formes (b), l'estimation du modèle statistique spatial des formes (c). Les catégories des formes et des points sont présentées par différentes couleurs.

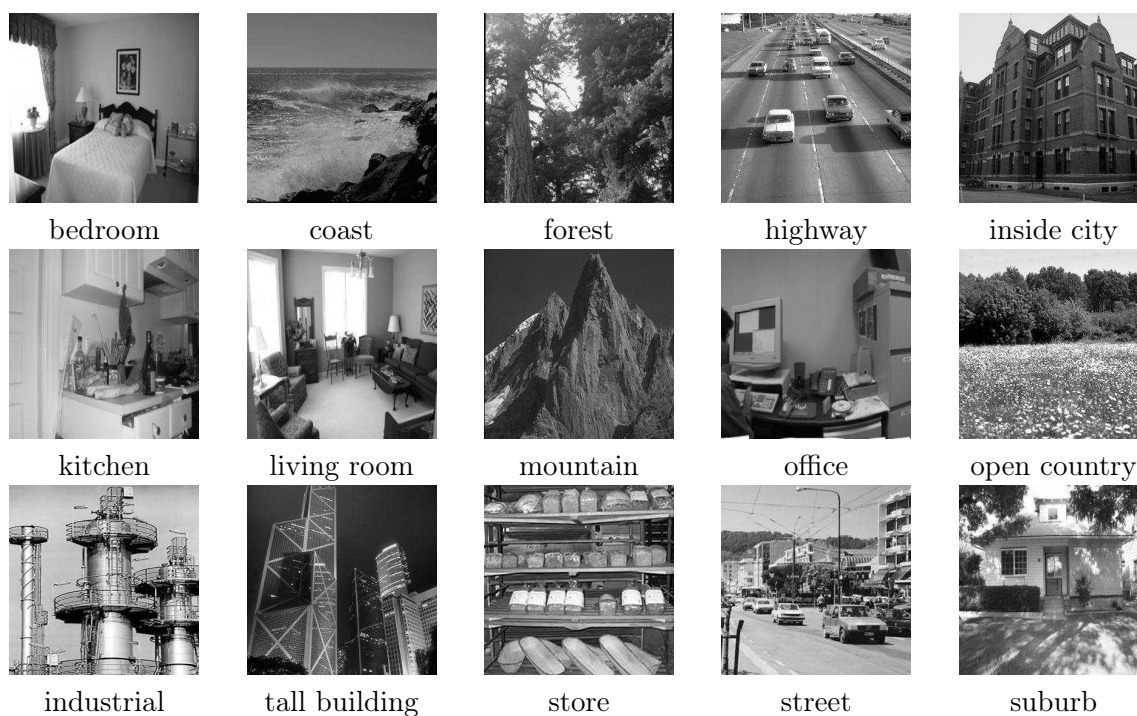


Figure 4.8 — Exemples d'images de 15 catégories de base de données de scène naturelle.

de scènes, y compris les deux types : "indoor" et "outdoor", qui sont obtenues de différentes sources : des photos personnelles, des collections de COREL, et des images résultats de recherche par Google. Chaque classe contient de 210 à 410 images. Les tailles d'image sont variantes, par exemple de 203 à 552 pixels de largeur et de 200 à 411 pixels en hauteur (voir la Figure 4.8).

Le paramétrage suivant est exploité dans cette expérience rapportée. Le modèle d'apprentissage supervisé est réalisé avec le classifieur RF. $N_t = 100$ images d'apprentissage de chaque classe sont choisies aléatoirement avec une répétition de $t = 20$ fois pour évaluer les performances en termes de taux moyens et d'écarts types de bonne reconnaissance. Les images restantes de chaque classe sont considérées comme les images

bedroom 75.6±2.1	coast 85.3±1.7	forest 91.1±1.2	highway 84.3±1.8	inside city 85.1±1.6
kitchen 76.0±1.5	living room 70.9±2.7	mountain 90.1±1.2	office 89.7±2.2	open country 76.7±1.8
industrial 66.8±2.3	tall building 88.9±1.3	store 82.2±2.0	street 87.6±1.5	suburb 93.3±1.4

Tableau 4.9 — Taux moyens et écarts types de bonne classification de chaque classe pour une répétition de 20 fois du modèle d'apprentissage quand 100 images d'apprentissages sont considérées. Le moyen des performances en global est $82.9\% \pm 1.7$.

de test. La surface minimum de chaque forme $\varpi_{min} = 100$ est considérée pour la segmentation des formes basée sur l'analyse des lignes de niveau. Le nombre de points sur le contour $n = 50$, les bins $n_d = 5$ et les orientations $n_\theta = 12$ sont considérés pour le calcul de l'histogramme log-polaire de l'algorithme IDSC. Le nombre de catégories de formes visuelles est fixé à $k = 50$. L'estimation du modèle statistique proposé est implémentée dans un ensemble des cercles étudiés avec les rayons de $r = 100\log(x)$, où x varié de 1 à $\exp(1)$ selon un pas linéaire de 0,05.

Les performances de reconnaissance de scènes naturelles pour les différentes classes sont rapportées dans le Tableau 4.9. La méthode proposée atteint jusqu'à $82,9\% \pm 1.7$ en terme de taux moyen et d'écart type de bonne classification en global quand 100 images d'apprentissage sont considérées. Une comparaison favorable des performances de reconnaissance de notre méthode avec des autres techniques de l'état de l'art est rapportée avec les performances de 74.8% pour le modèle hiérarchique bayésien [132], 81.4% pour le schème de pyramide spatiale des catégories de points d'intérêt [5] et 81.72% pour le concept de corrélogramme spatial [44]. Ces résultats démontrent la robustesse et la pertinence de l'approche proposée. Ils s'expliquent par les propriétés d'invariance aux changements de contraste, l'occlusion d'objet et la transformation de forme du descripteur proposée qui sont pleinement héritées de la caractérisation des formes visuelles par la segmentation basée sur les lignes de niveau et l'algorithme IDSC.

4.6 Conclusion

Dans ce chapitre, nous avons présenté les implémentations des modèles statistiques proposés pour les différentes applications de classification supervisées avec les images texturées réelles (UIUC, Brodatz, KTH-TIPs) et les images de scènes. Les gains obtenus par les comparaisons des performances de classification montrent la robustesse des modèles statistiques pour définir un nouveau descripteur d'analyse d'image présentant des qualités d'invariance basées sur la caractérisation de l'organisation spatiale des points d'intérêts (ou les formes) et de leurs caractéristiques. En outre, ces résultats expérimentaux démontrent que le modèle statistique proposé est particulièrement bien adapté pour les applications de classification et de reconnaissance d'images. Ils permettent d'envisager les applications des modèles proposés à la caractérisation et la classification d'images dans d'autres domaines comme les images satellites, les images sonars, etc.

Deuxième partie

Application à la classification d'images
sonars de fonds marins

Caractérisation et classification d'images sonars de fonds marins

5.1 Introduction

L'analyse actuelle des propriétés physiques du fond marin et de la cartographie des habitats des fonds marins est devenue de plus en plus importante notamment pour la gestion des ressources marines et la recherche scientifique [135, 136]. De nombreux projets ayant pour but le suivi des états des écosystèmes marins ont été initiés pour fournir une connaissance pertinente et cohérente des habitats benthiques côtiers et marins [12, 13, 14]. Parmi les études sur les écosystèmes marins, la cartographie des habitats marins constitue un enjeu majeur pour la gestion intégrée de la zone côtière. Les innovations dans le traitement des signaux acoustiques sous-marins ont joué un rôle important dans l'examen de larges surfaces du fond marin et obtenir des images acoustiques avec une grande résolution.

Dans ce travail, nous considérons une application à la caractérisation et la classification des images sonars des fonds marins. Ces images sonars, qui sont extraites du catalogue des données acoustiques du projet Rebent, nous permettent de prolonger dans les petits fonds, les limites de peuplements pour identifier la structuration et la répartition des organismes benthiques denses. Les propriétés physiques et biologiques des différents types du fond marin se remarquent généralement dans la texture et les contours des images. De plus, les caractéristiques texturées d'un type de fond marin varient et dépendent des conditions de prise de vue et de traitement. Pour répondre aux objectifs d'automatisation et d'amélioration des traitements cartographiques des habitats des fonds marins, nous appliquons ici les méthodes décrites dans la première partie avec des descripteurs invariants pour les images texturées sonars. Elles sont basées sur la caractérisation de l'organisation spatiale des points d'intérêt et de leurs caractéristiques par les analyses statistiques descriptives ou les estimations des paramètres du modèle de Cox log-Gaussien. Nous nous intéressons à montrer que nos approches sont des outils efficaces pour le dépouillement et l'analyse de ces données acoustiques.

Ce chapitre est organisé comme suit. La description du projet Rebent avec les informations sur les moyens de prospection acoustique, les techniques d'acquisition et d'analyse de données sonar du projet Rebent sont présentées dans la première section.

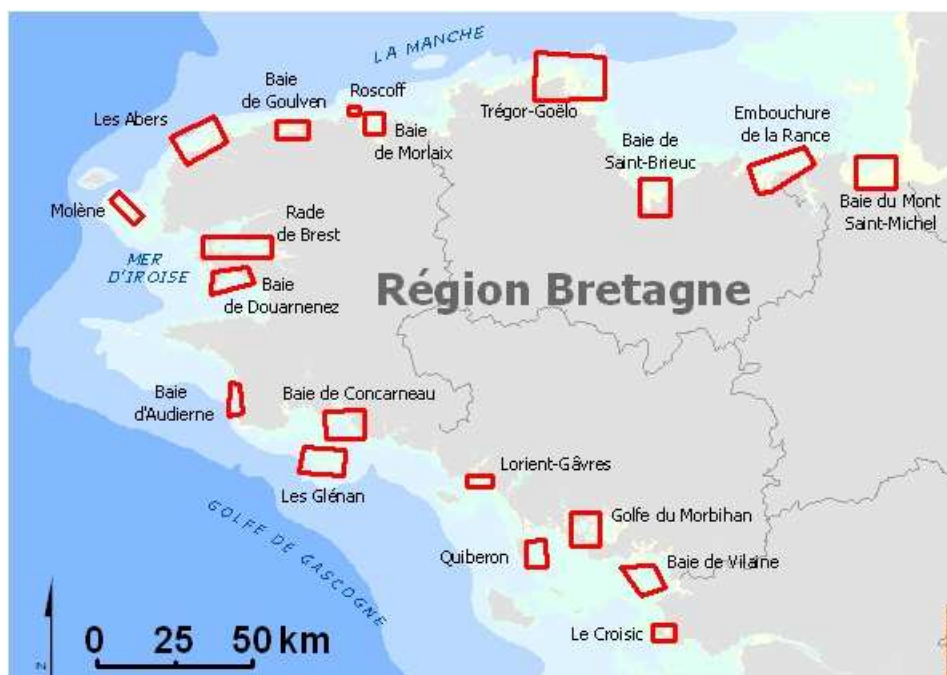


Figure 5.1 — La carte des secteurs d'étude (en rouge) du REBENT Bretagne www.rebent.org.

La deuxième section introduit notre base de données des images texturées sonars qui sont extraites d'un ensemble d'images acoustiques d'un levé d'un secteur du projet Rebent. Puis, les méthodes de l'état de l'art pour la caractérisation des images sonars sont décrites dans la troisième section. Ensuite, nous présentons les évaluations expérimentales de nos approches pour une application de reconnaissance de différents types du fond marin. Finalement, la Section 5.6 donne des conclusions.

5.2 Construction de l'image sonar

5.2.1 Projet Rebent - Réseau de surveillance benthique

Le projet REBENT (REseau de surveillance BENThique)¹, initié en décembre 1999 par le Ministère de l'Ecologie et du Développement Durable (MEDD) et coordonné par l'Ifremer, a pour objectifs d'acquérir une connaissance pertinente et cohérente des habitats benthiques côtiers afin de mettre à disposition des scientifiques, des gestionnaires concernés à la diversité biologique pour permettant de mieux connaître l'existant et de détecter les évolutions de ces habitats. Il vise à recueillir et à mettre en forme un ensemble de données relatives aux distributions spatio-temporelles des habitats et à leur composition faunistique et floristique dans une vingtaine de secteurs répartis autour de la Bretagne [137]. Les localisations des secteurs d'étude du REBENT Bretagne sont présentées dans la Figure 5.1.

1. <http://www.rebent.org>

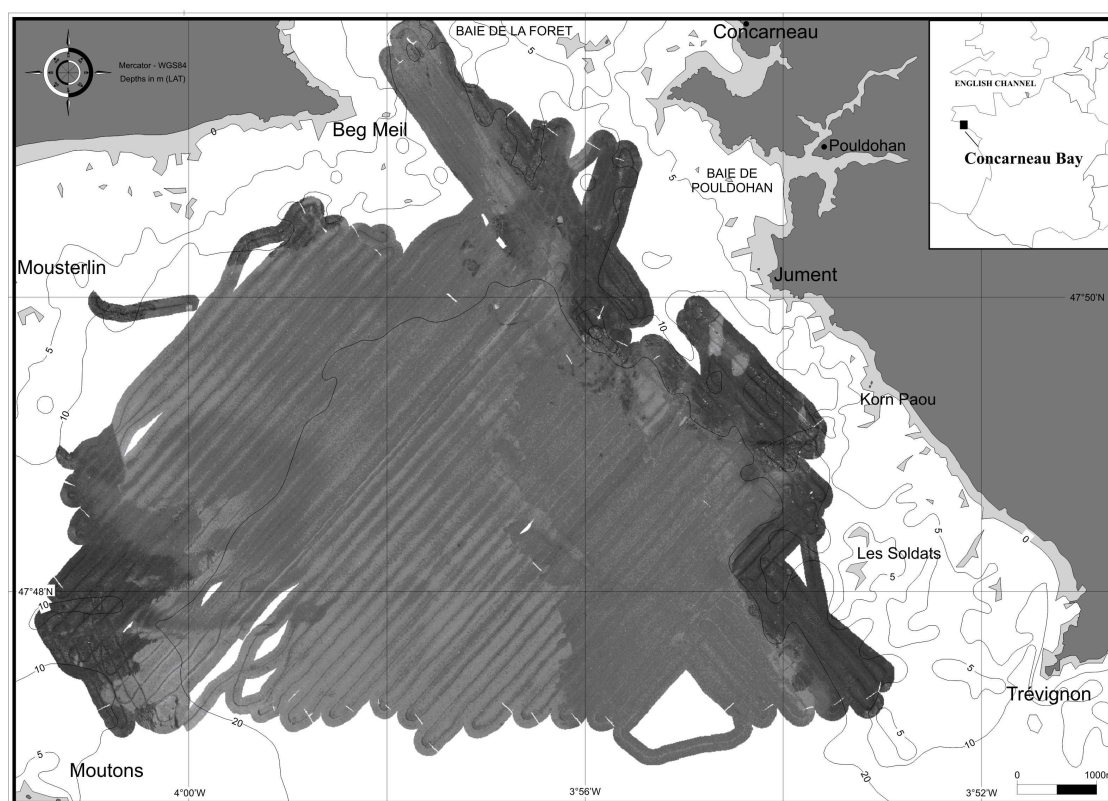


Figure 5.2 — Illustration de la zone d'étude de projet Rebut dans la baie de Concarneau, Bretagne, France.

Nous considérons les résultats de l'ensemble des travaux effectués sur 200km^2 du secteur de Concarneau en 2003 (Figure 5.2). Les levés géophysiques réalisés dans le cadre du REBENT reposent sur la combinaison de plusieurs capteurs acoustiques. Les jeux de données exploités comprennent les différents types de données sonars qui permettent d'accéder aux informations exhaustives de la bathymétrie et de la réflectivité du fond. Ici, les Sondeurs Multi-Faisceaux (EM1000 et Seabat8101) et le Sonar à Balayage latéral (DF1000) a été utilisé. Dans le cadre de notre travail, nous envisageons les images sonars DF1000, qui ont été déjà traité pour des applications de cartographie sédimentaire. Plus de détails sur ce type de sonar à balayage latéral est présentée dans la sous-section. De plus, au total 25 vidéos sous-marine et 93 échantillonnages biologiques ont été collectées avec pour but de mieux comprendre des signatures acoustiques. Ils constituent un maillage d'observations stratifiées conduisant à l'analyse taxonomique de la faune. Un exemple de base de données sonars avec l'image extraite de vidéo sous-marine et l'échantillonnage biologique de maërl sablo-graveleux obtenues sur le secteur de Concarneau est présenté dans la Figure 5.3.

5.2.2 Principe d'un sonar à balayage latéral Edgetech DF1000

Le sonar latéral est un émetteur-récepteur d'ondes acoustiques remorqué par un câble relié au navire (Figure 5.4). Un système complet de sonar à balayage latéral se

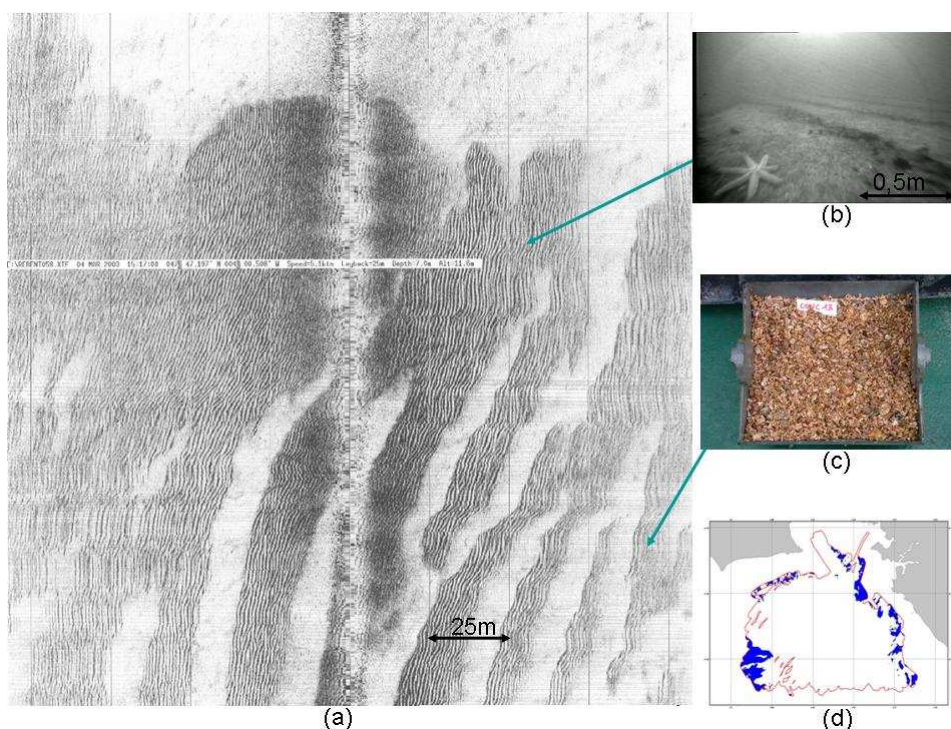


Figure 5.3 — Exemple de base de données sonars (a) avec l'image extraite de la vidéo sous-marine (b) et l'échantillonnage biologique (c) de maërl sablo-graveleux obtenues dans les zones d'étude (d) de la baie de Concarneau (Rebent-IFREMER).

compose :

- d'un poisson cylindrique de 1,5 m en longueur pour 10 à 15 cm de diamètre avec un câble porteur, un bras de remorque, un lest en zinc dans le nez et des ailerons pour la stabilisation et l'équilibre d'hydrodynamique. La Figure 5.5 présente la photo d'un poisson de la société Edgetech qui est utilisé dans ce projet.
- d'un enregistreur numérique situé à bord du navire qui lie au poisson par un câble électroporteur qui assure l'acquisition des données.
- occasionnellement, d'un dépresseur qui permet d'augmenter la profondeur d'immersion du poisson.

Le principe de fonctionnement du sonar latéral est : deux transducteurs latéraux émettent des faisceaux extrêmement fins en deux voies latérales. Ces faisceaux sont d'ouverture transversalement moins de 1° et large verticalement de plus de 20° . Les impulsions ultrasonores successives de courte durée (0,1 ms) et de haute fréquence (100 kHz ou 500 kHz) se propagent dans l'eau et sont réfléchies sur le fond. Seule l'onde spéculaire, celle empruntant la même trajectoire après réflexion, est reçue par le poisson du sonar latéral. L'amplitude de l'écho rétro-diffusé est ensuite enregistrée et corrigée pour la reconstitution d'une image acoustique du fond marin. Sur ces images acoustiques (sonogrammes), les types de fonds, les zones de roches ou des biocénoses spécifiques (maërl, moulières, crépidules, lanices, herbiers) sont différenciés et délimités avec précision. En effet, en fonction de la densité des individus sur le fond, les propriétés acoustiques du substrat sont modifiées et ces biocénoses peuvent être détectées

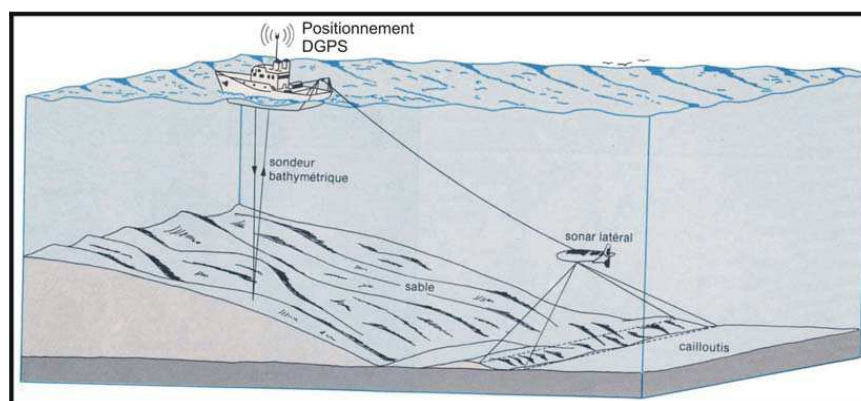


Figure 5.4 — Poisson sonar latéral tracté en position de balayage au-dessus du fond (Rebent-IFREMER).



Figure 5.5 — Image des sonars DF 1000 de la société Edgetech (Rebent-IFREMER).

par le sonar. Son principe repose donc sur une signature acoustique propre à chaque type de substrat (morphologie et composition granulométrique du fond). Les données brutes sont traitées en temps réel par un logiciel d'acquisition numérique qui corrige les distorsions géométriques liées à la vitesse du bateau et aux variations de profondeur [137, 138].

5.2.3 Acquisition et analyse des données

En pratique, l'acquisition des données des sonogrammes dépend de nombreuses caractéristiques de transfert des ondes acoustiques qui sont [139] :

- les pertes de propagation dues à la réflexion du signal acoustique en fonction du temps ou de la distance parcourue, par exemple : la diffusion de signaux émis et reçus, la perte par réflexion sur le fond, la diffusion sur la surface du fond marin, la réflexion sur le substratum induré.
- les variations de structuration spécifique des peuplements, des épaves et des avalanches sous-marines.

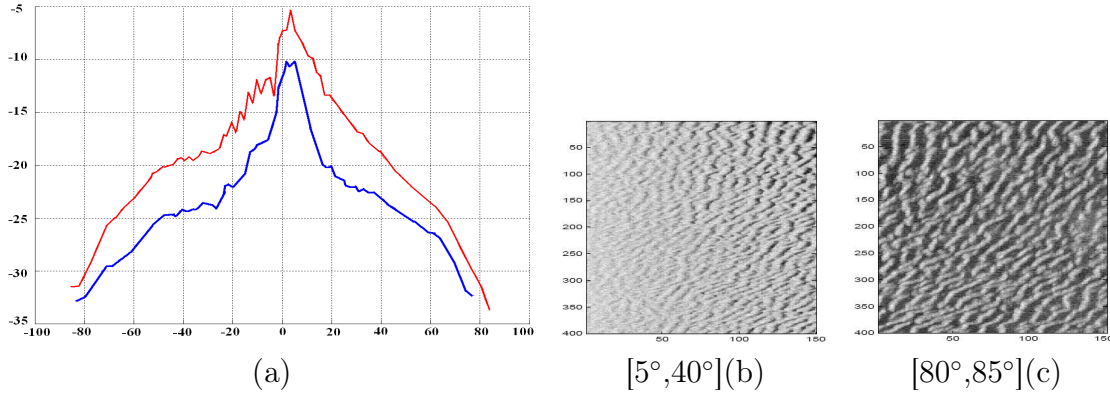


Figure 5.6 — (a) L'évolution de -85° à $+85^\circ$ de l'incidence de l'écho rétro-diffusé en fonction de l'angle de deux types de fond : le maërl sablo-graveleux (en rouge) et le sable fin à moyen (en vert) dans la Figure 5.3 et (b,c) les deux images sonars exemples du maërl sablo-graveleux en deux secteurs angulaires différentes (Rebent, IFREMER).

- l'intensité de l'écho rétro-diffusé (BS - Backscattering Strength) dépend de l'angle d'incidence θ_{inc} de l'onde acoustique avec le fond marin. [138, 140].

Les données de sonogrammes brutes sont corrigées au mieux en temps réel par le système d'acquisition numérique à bord du navire. Les distorsions géométriques sur les sonogrammes liées à la vitesse du bateau sont traitées aussi. Il reste cependant la variation angulaire d'incidence qui correspond à l'angle d'émission de la source de l'onde acoustique par action du fond marin $BS(\theta_{inc})$ (en dB). Les comportements angulaires des réflectivités des différents types de fond sont vraiment distinctifs. La Figure 5.6a présente un exemple des différentes courbes des réflectivités calculées en fonction de l'angle de deux types de fond : le maërl sablo-graveleux et le sable fin à moyen. Dans la littérature, plusieurs travaux se sont intéressés à ce comportement angulaire pour discriminer les différents types de fonds [141, 142, 143, 144].

Classiquement, la construction d'une image sonar résulte de la fusion des signaux acoustiques rétro-diffusés pour des fauchées successives. Un exemple d'image résultat de sonar à balayage latéral avec des angles d'incidence de -85° à $+85^\circ$ est rapporté dans la Figure 5.3a. Chaque colonne de cette image sonar correspond à une bande de 25m de largeur du fond marin. Les différents types de fond marin correspondent à différents types de texture. Les caractéristiques texturées d'un type de fond présentent des variations en fonction de l'incidence, en particulier pour des angles proches de la verticale. Par exemple, dans la Figure 5.6(b,c), deux observations de sable corpusculaire sont données. Ces observations correspondent à des incidences différentes, respectivement $[5^\circ, 40^\circ]$ et $[80^\circ, 85^\circ]$. On peut noter à la fois des variations du contraste et des modifications géométriques des structures observées. L'application de méthodes de caractérisation de texture invariantes au changement de contraste et à des transformations géométriques semble donc pertinente.

	Cailloutis	Graviers	Sables grossiers	Sables moyens	Sables fins	Sables très fins	Vases
la vase	0	1	2	2	3	9	83
la vase sableuse	0	3	5	3	3	13	73
le maërl sablo-graveleux	1	70	24	4	0	0	0
le sable fin à moyen	0	10	8	45	33	3	0
les sédiments mélangés	1	15	15	8	15	18	29

Tableau 5.1 — Le pourcentage en moyenne des composants de chaque sédiment d’analyses granulométriques.

5.3 Base de données des textures sonars

Dans le but de détecter les changements dans les habitats benthiques côtiers à différentes échelles de temps et d’espace, 50 sonogrammes de différentes caractéristiques physiques et biologiques de la zone étudiée ont été recueillies (Figure 5.3a). Parmi les différents types de fonds marins obtenus sur la zone de Concarneau, nous décidons de ne considérer que six classes privilégiées pour une application de reconnaissance des images texturées du fond marin. Les noms des six classes du fond marin sont : la vase, la vase sableuse, le maërl sablo-graveleux, le sable fin à moyen, le substratum rocheux et les sédiments mélangés. L’évaluation expérimentale réalisée exploite une base de 240 images sonars de taille 256 x 256 pixels avec des valeurs différentes d’angles d’incidence, d’échelles et de niveaux moyens d’écho rétrodiffusé qui sont extraites d’un ensemble de 50 sonogrammes. Chaque classe de fond marin contient 40 images texturées. Les détails d’analyses granulométriques (Tableau 5.1) et les exemples des échantillonnages (Figure 5.7) de chaque sédiment sont présentés ci-dessous.

5.3.1 La vase

Le faciès blanc homogène correspondant à une vase molle en surface est limité à de petites dépressions le long de la côte orientale de la baie de Concarneau et à de plus grandes nappes localisées au pied du talus littoral de $-20m$ à $-25m$ de profondeur, à côté de la Jument (voir les figures 5.2 et 5.7). Un vase contient au moins 75% de particules inférieures à $50\mu m$. Les prélèvements à la benne remontent une vase sableuse de consistance pâteuse, coiffée en surface d’une crème de vase fluide sur plusieurs centimètres.

5.3.2 La vase sableuse

Le pourcentage en vase de ce type du fond marin est le plus grand par rapport aux autres, mais ce taux ne dépasse jamais 75% du poids total du sédiment. Ce faciès

sédimentaire se trouve dans une grande zone au centre de la baie entre les pointes de la Jument et Trévignon. Sur les images acoustiques, ce type de faciès blanc nuageux contient les oeillets et les sillons qui correspondent à la formation de pockmarks (cratères) de plus petites tailles au sein même de pockmarks plus anciens. L'activité de chalutage et de dragage sur ce secteur, a laissé des traces sur les fonds colonisés par un véritable tapis de petits amphipodes (haploops), celles-ci encore visibles au sonar à balayage latéral sont parfois recoupées par des pockmarks plus récents.

5.3.3 Le maërl sablo-graveleux

Le maërl sableux et graveleux correspond à la signature du banc des Moutons. Le maërl (algue calcaire) qui constitue une biologie unique en Bretagne, est ici mélangé à des débris coquilliers, mais surtout à une fraction siliceuse plus importante, soit sous la forme de sable grossiers, soit sous la forme de petits graviers détritiques roulés. Les débris de maërl sont plus nombreux, en relation avec les conditions hydrodynamiques plus fortes qu'en baie orientale. Le banc est accroché en écharpe, en arrière (côté baie) du plateau des Moutons. Il est morphologiquement scindé en deux parties de part et d'autre d'un sillon de 5 m de profondeur. Le maërl occupe essentiellement le sommet et les flancs nord des deux bosses.

5.3.4 Le sable fin à moyen

Le sable fin à moyen se limite au sud-ouest de la baie de Concarneau, à 1,5 miles au nord-est du banc des Moutons, sur un fond plat homogène. La moyenne des analyses granulométriques fait ressortir les distributions centrées sur les sables moyens (45%) et les sables fins (33%). Le sédiment est localement plus coquillé ou un peu envasé. Les images texturées de ce sédiment sont tachetées.

5.3.5 Les sédiments mélangés

Le faciès gris sombre hétérogène des sédiments mélangés se trouve à proximité des nappes de maërl, surtout sur la côte Est ou à la sortie et dans le chenal d'accès au port de Concarneau. La vase est mélangée à une proportion de différentes particules constituées de fragments de maërl mort sur de la vase compacte, de petits graviers de quartz ou de vase épaisse sableuse relativement riche en éléments grossiers détritiques.

5.3.6 Le substratum rocheux

Ce faciès se trouve dans la frange littorale à partir de -20m de profondeur dans les zones entre la pointe de Mousterlin et celle de Trévignon. Sur les vidéo sous-marines, le substrat rocheux est constitué d'une accumulation de petits blocs et de dépôts grossiers le plus souvent orientés nord-sud et disposés sur des reliefs d'allure tabulaire d'un mètre de haut environ. Il s'agirait d'une surface d'abrasion recouverte d'accumulations détritiques littorales sous la forme de blocs et de galets constituant des paléocordons.

La vase :



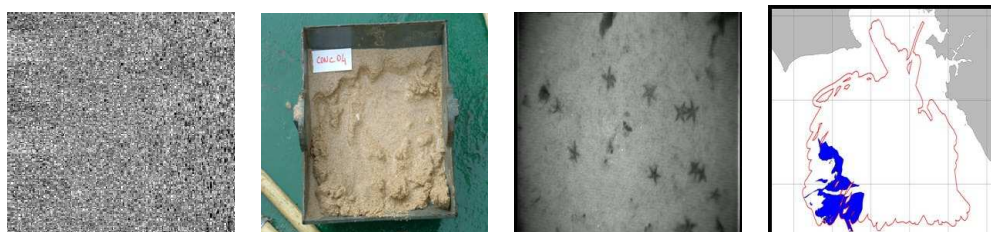
La vase sableuse :



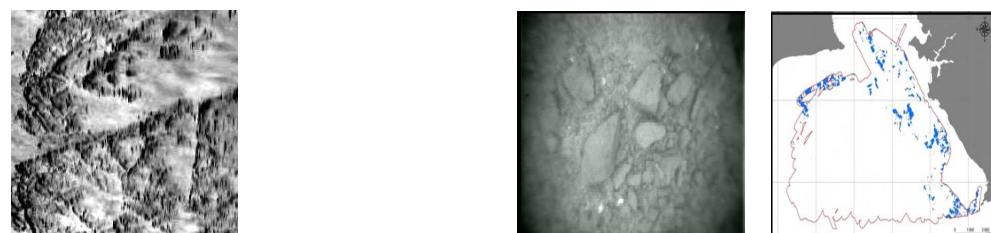
Le maërl sablo-graveleux :



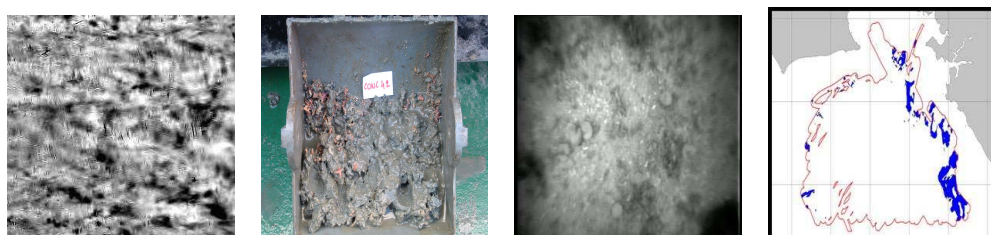
Le sable fin à moyen :



Le substratum rocheux :



Les sédiments mélangés :



(a)

(b)

(c)

(d)

Figure 5.7 — Exemples de textures sonars des différents types de fond marin (a) avec leurs images de l'échantillonnage biologique (b), de l'extraction de vidéo sousmarine (c) et de la localisation des zones d'étude (en bleu) de ce type de fond marin (d).

5.4 Etat de l'art sur les méthodes de caractérisation des images sonars

Le sonar à balayage latéral est une des plus pertinentes technologies de télédétection pour l'observation in-situ des fonds marins. La reconnaissance automatique et la segmentation des images sonars en termes de types de fonds marins [142, 94] des mosaïques sonars sont devenues des problématiques importantes pour la gestion des ressources marines et la recherche scientifique. Dans la littérature, de nombreux travaux se sont intéressés à trouver des modèles paramétriques qui approchent au mieux la distribution marginale de l'énergie rétro-diffusée pour discriminer les différents types de fonds. Parmi les modèles les plus utilisés, nous pouvons citer la loi K, la loi Γ , la loi de Weibull [145, 146, 147] pour envisager les variations des paramètres des modèles selon la variation de l'angle d'incidence de l'écho rétro-diffusé. De plus, il faut également noter les modèles bayésiens pour déterminer les propriétés intrinsèques du fond marin par le découplage des propriétés du système d'acquisition en considérant la variation en fonction de l'angle de l'intensité rétro-diffusé [140]. Cependant, aucune de ces études n'a réussi à réaliser une inversion précise des données du fait de la complexité des phénomènes physiques et des artefacts apportés par les capteurs [147, 148].

En outre, dans le domaine de la vision par ordinateur, plusieurs approches qui se sont révélées efficaces pour la reconnaissance des textures sont proposées. On recherche notamment de descripteurs de texture invariants aux conditions d'acquisition des images, par exemple le changement de contraste, la transformation géométrique. Les travaux initiaux ont reposé sur des approches statistiques basées sur le calcul de distributions de réponses à des filtres, par exemple la matrice de cooccurrences et les descripteurs associés [7, 16], les descripteurs basés sur les transformations de Gabor et en ondelettes [17, 6], ou les multifractales [10]. Cependant, ces méthodes ne peuvent pas résoudre intrinsèquement les problèmes dus aux variations d'angle d'incidence.

Des travaux récents [23, 50] ont montré la pertinence des approches reposant sur l'extraction et la caractérisation de points d'intérêt. Un descripteur invariant est défini par l'analyse de statistiques des points d'intérêt. L'intérêt de ces approches réside notamment dans leurs propriétés d'invariance au changement de contraste et aux transformations affines des images qui sont hérités des propriétés d'invariance de points d'intérêt. Plus récemment, l'évaluation de statistique des points d'intérêt, en particulier via le principe des sacs de mots visuels qui repose sur la définition de dictionnaire des caractéristiques des points d'intérêt, a été proposée [8] pour des applications à la reconnaissance des objets et des textures. Les statistiques descriptives sur les fréquences d'apparition et les informations spatiales des différents mots du dictionnaire sont proposées avec des améliorations significatives des performances de reconnaissance [25, 5].

Nous appliquons ici les descripteurs proposés [94, 149, 150] reposant sur les réalisations de processus spatiaux ponctuels marqués de points d'intérêt pour une application à reconnaissance d'images texturées du fond marin. Ces approches consistent à représenter une image texturée par un ensemble de descripteurs invariants qui héritent les propriétés d'invariance de points d'intérêt. Les différentes méthodes de l'extraction de

	DoG	FH	(Har-Lap)	(Hes-Lap)
La vase	334.85	35.125	20.925	182.325
La vase sableuse	670.5	166.65	230.95	534.65
Le maërl sablo-graveleux	1777.275	297.975	476.425	400.3
Le sable fin à moyen	377.525	143.825	205.55	368.5
Le substratum rocheux	1070.7	307.65	573.75	775.65
Les sédiments mélangés	1350.175	353.525	554.85	834.525
Moyen global	930	217.45	344	516

Tableau 5.2 — Le nombre moyen des différents types de points d'intérêt détectés pour chaque classe de la base de données des textures sonars.

points d'intérêt donnent les variations de nombres de points et leurs distributions spatiales qui sont liées à l'organisation de la texture de l'images sonars. Nous envisageons ici cinq types différents de points d'intérêt : DoG+Sift, FH+Surf, (Har-Lap)+(Sift-Spin), (Hes-Lap)+Daisy, FH+Brief pour le calcul du vecteur de descripteur des images sonars (Figure 5.8 et Tableau 5.2). Les comparaisons des performances du descripteur LGCM sont étudiées les types de points d'intérêt, le choix des classifieurs et les mesures de similarité pour l'apprentissage supervisée et les sélections des fonctions co-variances. Ensuite, une autre comparaison de notre descripteur LGCM sera comparée aux autres méthodes populaires et référencées. Par contre, nous avons fait l'extraction d'un arbre des formes reposant sur le changement de contraste global. Pourtant, les signatures locales obtenues par cette méthode ne sont pas discriminantes parmi les différentes images sonars de peuplements denses recouvrant des fonds de rugosité voisine, par exemple la classe du "sable fin à moyen" et la classe de la "vase". Donc, nous ne présentons pas les résultats de l'application de cette méthode.

5.5 Résultats obtenus par les méthodes proposées basées sur les statistiques spatiales de points marqués

Nous considérons ici une application à la caractérisation et la classification des images sonars des fonds marins. L'apprentissage des modèles de classification est réalisée à partir d'un sous-ensemble d'images de chaque classe choisies aléatoirement, pour N_t images d'apprentissage par classe. $(40 - N_t)$ images restantes par classe sont considérées comme images de test. Cette étape d'apprentissage est répétée 50 fois pour évaluer les performances de classification, en termes de taux moyens et d'écarts types de bonne classification. Nous présentons ici les détails de ces expérimentations pour le descripteur basé sur les modèles de Cox log-Gaussien. Les mêmes résultats ont été obtenus pour les statistiques descriptives de cooccurrence [104, 94].

Choix de méthodes d'extraction de points d'intérêt

Nous évaluons tout d'abord la pertinence des différents types de points d'intérêt pour la réalisation du modèle de Cox log-Gaussien. Le dictionnaire de $k = \{35, 50, 60\}$

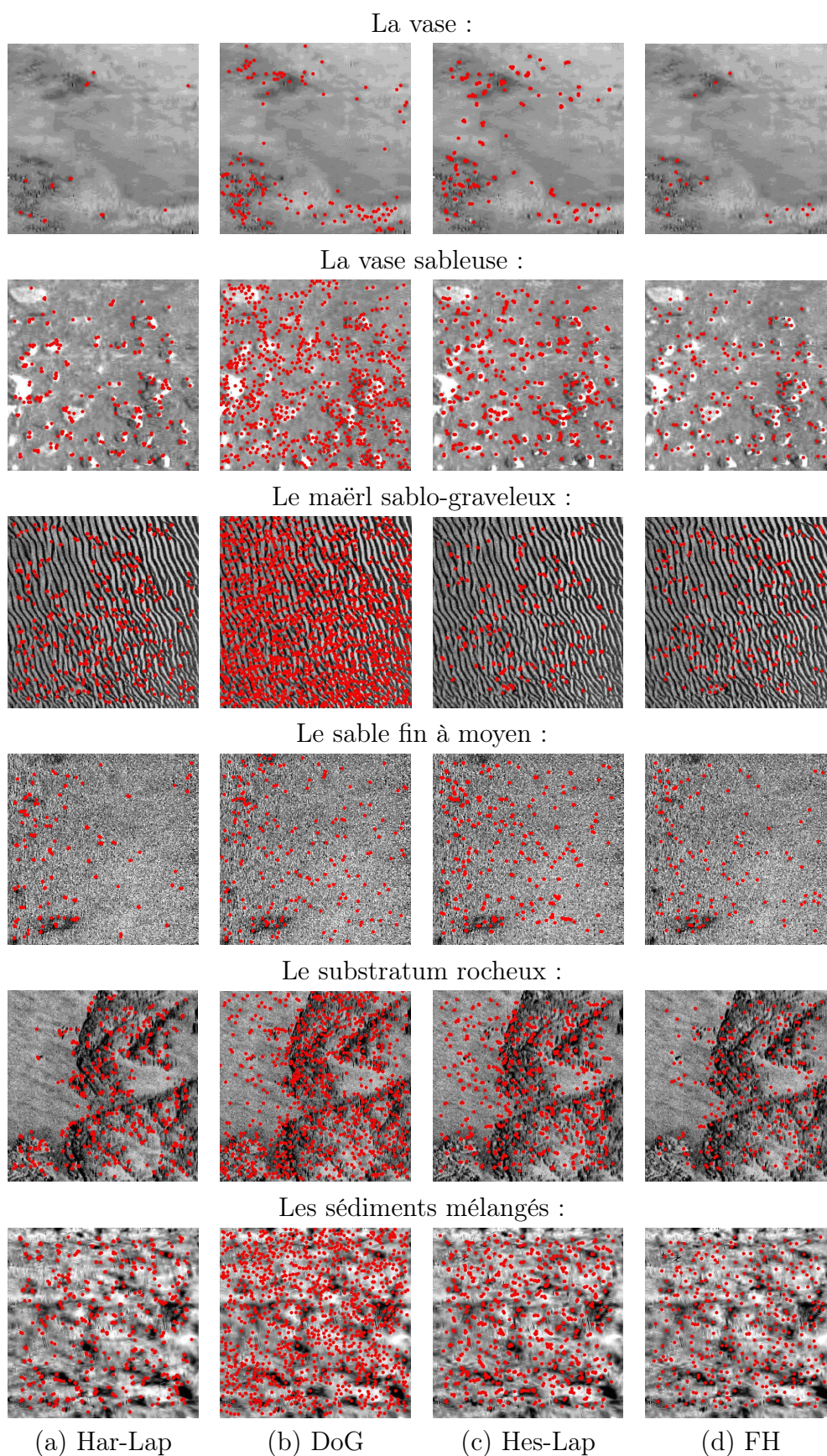


Figure 5.8 — Exemples des différentes distributions spatiales de points d'intérêt dans les images sonars pour les différentes classes du fond marin : (a) Har-Lap détecteur, (b) DoG détecteur, (c) Hes-Lap détecteur, (d) FH détecteur.

	DoG+Sift	FH+Surf	(Har-Lap) +(Sift-Spin)	(Hes-Lap) +Daisy	FH+Brief
1 image	73.85±1.75	71.15±1.91	72.67±1.95	73.67±1.52	74.85±1.65
3 images	82.95±1.13	79.95±1.41	82.31±1.55	82.35±1.23	81.62±1.28
5 images	91.34±0.72	90.17±1.37	91.05±1.31	91.24±0.81	90.84±1.13
8 images	93.87±0.61	92.08±0.82	92.12±0.95	92.75±0.73	92.07±1.11
10 images	97.14±0.37	95.67±0.51	96.12±0.62	96.63±0.47	95.92±0.59

Tableau 5.3 — Comparaison des performances de reconnaissance de textures sonars pour différents types de points d'intérêt utilisés pour la construction du descripteur LGCM : DoG+Sift, FH+Surf, (Har-Lap)+(Sift-Spin), (Hes-Lap)+Daisy, FH+Brief. Le classifieur RF est considéré dans cette expérience.

mots visuels est construit. La statistique spatiale des points d'intérêt est calculée sur l'ensemble des cercles étudiés avec les rayons de $r = 128 \ln(x)$, où x varié de 1 à $\exp(1)$ selon un pas linéaire de 0.05. La fonction covariance Gaussienne est utilisée pour l'estimation des paramètres du modèle de Cox log-Gaussien. Les procédures d'adaptation d'échelle et de réduction la taille de descripteur avec le nombre de paires de points d'intérêt $k^* = 30$ sont appliquées. Le classifieur RF est choisi pour le modèle d'apprentissage supervisée. Le taux moyen de bonne classification de chaque type de points d'intérêt est présenté dans le Tableau 5.3 comme une fonction du nombre d'images d'apprentissage avec $N_t = \{1, 3, 5, 8, 10\}$. La combinaison de détecteur-descripteur FH+Brief se révèle légèrement plus robuste uniquement quand une seule image d'apprentissage est considérée. Par contre, le Dog+Sift descripteur obtient un léger gain environ de 1% de performance de la classification par rapport à (Har-Lap)+(Sift-Spin), (Hes-Lap)+Daisy ou 2% par rapport aux méthodes : FH+Surf, FH+Brief. Ces résultats pourraient s'expliquer par le plus grand nombre de points d'intérêt obtenus en utilisant le schéma Dog+Sift contre les autres méthodes. Le nombre de signatures locales détectées est un critère important pour une estimation plus précise des statistiques spatiales ponctuels.

Choix du classifieur et de la mesure de similarité

La deuxième évaluation est réalisée pour une comparaison des performances de descripteur proposé LGCM pour les différents modèles d'apprentissages supervisées. Trois classifieurs : k-PP, SVM, RF et trois distances de mesure de la similarité : Euclidienne, χ^2 , Jensen-Shannon sont envisagées dans cette évaluation. La valeur k , nombre de voisinages considérés, du classifieur k-PP est choisi comme une fonction du nombre d'images d'apprentissage N_t (l'Equation 5.1). Le noyau Gaussien est choisi pour le classifieur SVM. Pour le RF, les paramètres $ntree = 500$ et $ntry = 100$ sont considérés. Le descripteur proposé LGCM est calculé par la caractérisation des points d'intérêt DoG+Sift qui a obtenu un légèrement meilleur résultat de classification. La même configuration de paramètres de la méthode LGCM est sélectionnée comme la première évaluation dans cette section, où le nombre de catégories de mots visuels est fixé à $k = 50$. Les résultats en termes des taux moyens et d'écarts types de bonne classification sont présentés dans le Tableau 5.4. La divergence de Jensen-Shannon réalise un gain de performance de 1,5%(3%) en moyenne à comparer avec la distance Chi2 χ^2 (respectivement la distance Euclidienne) pour les deux cas de classifieurs : k-PP

	k-NN			SVM			RF
	d_E	d_{χ^2}	d_J	d_E	d_{χ^2}	d_J	
5 images	87.45%	88.17%	89.56%	88.11%	88.95%	91.34%	91.34%
10 images	93.12%	94.83%	95.37%	94.18%	95.07%	96.95%	97.14%

Tableau 5.4 — Comparaison des performances du descripteur LGCM pour la classification supervisée des textures sonars pour différents classifieurs (k-NN, SVM, RF) et des différentes mesures de similarité (Euclidienne, χ^2 , divergence de Jensen-Shannon).

	sans procédure d'adaptation d'échelle	sans procédure de réduction la taille	modèle complet
1 image	72.63±1.75	74.41±1.62	73.85±1.75
3 images	82.27±1.53	83.11±1.25	82.95±1.13
5 images	89.85±1.11	91.05±0.95	91.34±0.72
8 images	91.72±0.85	92.37±0.62	93.87±0.61
10 images	95.15±0.55	96.28±0.45	97.14±0.37

Tableau 5.5 — Comparaison des performances du descripteur LGCM pour la classification de textures sonars dans les cas où on utilise ou non les procédures d'adaptation d'échelle et de réduction la taille de descripteur.

et SVM quand 5 et 10 images d'apprentissage sont considérées. Les similaires performances de classifieurs : SVM et RF sont obtenues pour la reconnaissance de notre jeu de données sonar. Comparativement au classifieur basé sur la distance Euclidienne, un gain de l'ordre de 2% en moyenne est obtenu.

$$\begin{cases} k = N_t \text{ if } N_t = 1 \text{ ou } 3 \\ k = 5 \text{ if } N_t = 5, 8 \text{ ou } 10 \end{cases} \quad (5.1)$$

Invariance en échelle et réduction de dimension

De plus, une autre évaluation sur les performances du descripteur LGCM est présentée dans le Tableau 5.5 dans le but de montrer la pertinence des procédures applicables à résoudre les problèmes d'invariance en échelle et la grande taille de la dimension des descripteurs proposés. Les mêmes configurations des paramètres de deux premières évaluations sont choisies. En considérant l'estimation des modèles de Cox log-Gaussien avec la procédure d'adaptation d'échelle, un gain de 2% est obtenu. En outre, le modèle de Cox log-Gaussien complet présente un comportement légèrement meilleur de 93.87% et de 97.14% (respectivement) avec la procédure de réduction la taille des descripteurs quand 5 et 10 images d'apprentissage sont considérées.

Choix du modèle de covariance du modèle LGCM

Nous comparons aussi les performances des descripteurs LGCM pour des modèles différents de fonctions covariances. Les résultats sont présentés dans le Tableau 5.6. Le modèle de covariance Gaussien obtient de légers meilleurs résultats par rapport aux modèles de covariance hyperbolique.

	Gaussien	Cardinal sine	Hyperbolique
1 image	73.85±1.75	73.60±1.79	73.42±1.91
3 images	82.95±1.13	82.47±1.21	82.15±1.23
5 images	91.34±0.72	90.84±0.67	90.52±0.81
8 image	93.87±0.61	93.11±0.65	92.57±0.75
10images	97.14±0.37	96.81±0.48	96.12±0.58

Tableau 5.6 — Comparaison des performances de reconnaissance de textures sonars pour différentes fonctions de covariance de l’approche LGCM.

Comparaison avec les méthodes de l’état de l’art

Enfin, les performances des descripteurs proposés pour la classification supervisée de texture sont comparées à différentes techniques de l’état de l’art pour notre jeu de données de textures sonar. Trois groupes de méthodes sont envisagés :

- Les méthodes classiques : Nous considérons les approches statistiques basées sur le calcul de distributions de réponses à des filtres : matrice de cooccurrences et descripteurs associés [7, 16], descripteurs basés sur les transformées de Gabor et en ondelettes [151, 152, 153], la description multifractale [10]. Les comparaisons des performances de classification avec ces méthodes montrent la pertinence des méthodes d’utilisation des points d’intérêt. Les configurations des paramètres de ces approches sont détaillées comme suit :
 - Gabor : la fréquence $f_G = \{0, 4, 8\}$ et l’orientation $\theta_G = \{0, \pm\frac{\pi}{2}, \pi\}$.
 - Matrice de cooccurrence : la distance $d_M = \{1, 2, 4\}$ et le paramètre angulaire $\theta_M = \{\frac{k\pi}{8}, k = 0 : 7\}$.
 - Multifractale : le niveau de densité $ind = \{1, 8\}$, la dimension de MFS $f = \{16, 64\}$ et le niveau d’itération $ite = \{8, 10\}$.
- Les méthodes basées sur la contribution statistique des mots visuels : Ce sont les méthodes qui consistent en la construction d’un dictionnaire des points d’intérêt et l’analyse statistiques se limitant uniquement aux fréquences relatives des différentes catégories de mots visuels dans l’image. Une comparaison de ce type de méthodes avec celles que nous proposons permet de mettre en évidence l’apport des statistiques spatiales des points d’intérêts. La méthode SDM [8] et la méthode de Zhang [11] sont envisagées avec le nombre de catégories de mots visuels $k = \{35, 50, 60\}$.
- Les méthodes basées sur l’analyse spatiale des points d’intérêt : L’exploitation d’un histogramme de statistiques de cooccurrence des paires des catégories de points d’intérêt sont considérées. Les méthodes de Ling [9], de Savarese [25] et notre méthode SSC sont considérées pour une comparaison des performances de classification avec notre méthode LGCM. Les paramètres sont $k = \{35, 50, 60\}$, $k^* = 30$, les rayons $r = 128\ln(x)$, où x varie de 1 à $\exp(1)$ selon un pas de 0.05.

La synthèse des résultats est donnée dans le Tableau 5.7. Comparativement aux performances de description multifractale (méthode de Xu [10]), de matrice de cooccurrence et de descripteur de Gabor, les gains de plus de 6%(respectivement, 16% et 27%) sont obtenus par lors de l’utilisation de la méthode LGCM. Elles mettent en évi-

N_t	1	3	5	8	10
Filtre de Gabor [6]	51.71±3.24	55.34±2.84	59.27±1.97	62.53±1.72	69.81±1.48
Matrice de Cooc. [7]	62.13±3.17	69.72±2.44	72.15±1.53	76.22±1.41	81.21±1.27
SDM[8]	67.15±2.55	76.27±2.08	85.42±1.56	90.35±1.42	92.03±1.21
Ling[9]	66.83±2.33	78.12±2.21	85.27±1.83	90.68±1.35	91.92±1.24
Xu[10]	67.54±2.49	79.33±2.11.	87.12±1.91	90.14±1.63	91.85±1.12
Zhang[11]	73.33±2.17	82.62±1.83	90.67±1.15	93.43±0.91	94.25±0.73
SSC[104]	74.57±1.69	83.23±1.19	91.17±1.08	93.25±0.58	96.67±0.35
LGCM	73.85±1.75	82.95±1.13	91.34±0.72	93.87±0.61	97.14±0.37

Tableau 5.7 — Taux moyens et écarts types de bonne classification de nos approches pour les images texturées sonars comparées avec des techniques d'état de l'art : Filtre de Gabor [6], Matrice de cooccurrence [7], sac de mots (SDM) [8], les statistiques de cooccurrence des paires des catégories de points d'intérêt [9], la description multifractale [10], la combinaison des différents points d'intérêt [11], statistiques descriptives spatiales de cooccurrence (SSC), modèles de Cox log-Gaussien (LGCM).

dence l'intérêt des propriétés d'invariance des techniques basées sur la caractérisation de texture à partir d'un ensemble de points d'intérêt vis-à-vis des approches statistiques classiques. Notre méthode LGCM obtient un gain de classification jusqu'à 11% (3%) par rapport à l'approche des fréquences des sacs de mots (respectivement, l'algorithme de Zhang [11]). Ceci démontre la pertinence de l'information apportée par l'organisation spatiale des points d'intérêt. Comme les observations des distributions des points d'intérêt dans les textures sonar (Figure 5.8), l'organisation spatiale des ensembles des points d'intérêt est également une information discriminative pour distinguer les types de fonds marins. En considérant les comparaisons des performances entre les différentes méthodes de l'analyse spatiale des points d'intérêt, notre approche obtient le meilleur résultat de 97,14% avec une amélioration proche de 6% par rapport aux méthodes de Ling [9] et de Savarese [25] quand 10 images d'apprentissage sont considérées. L'utilisation des procédures d'adaptation d'échelle et de réduction la taille des descripteurs et de correction des effets de bord explique ces gains des performances de classification. Par ailleurs, les deux méthodes proposées (méthode SSC et méthode LGCM) qui sont basées sur les statistiques spatiales ponctuelles marquées, réalisent des performances similaires. Cependant, le descripteur LGCM se révèle légèrement plus robuste que le descripteur SSC avec $97.14\% \pm 0.37$ vs. $96.67\% \pm 0.35$ quand 10 images d'apprentissage sont considérées. Il faut également souligner le gain en terme de complexité de la représentation des textures par le descripteur LGCM par rapport au descripteur SSC, avec $3k^*$ vs. $N_r k^*$, où N_r est le nombre des boucles étudiées et k^* est la taille de dictionnaire des paires des points.

Dans la Figure 5.9, nous présentons aussi une matrice de confusion de la classification supervisée des textures sonars utilisant l'approche LGCM quand 5 images d'apprentissage sont considérées. Les taux moyens de bonne classification des classes du fond marin varient de 87.35% à 93.65%. Les textures sonars du maërl sablo-graveleux sont les plus faciles types à reconnaître. La plus grande confusion parmi des classes est obtenue avec 5,75% des images de la classe du "sable fin à moyen" tombées dans la classe de la "vase sableuse".

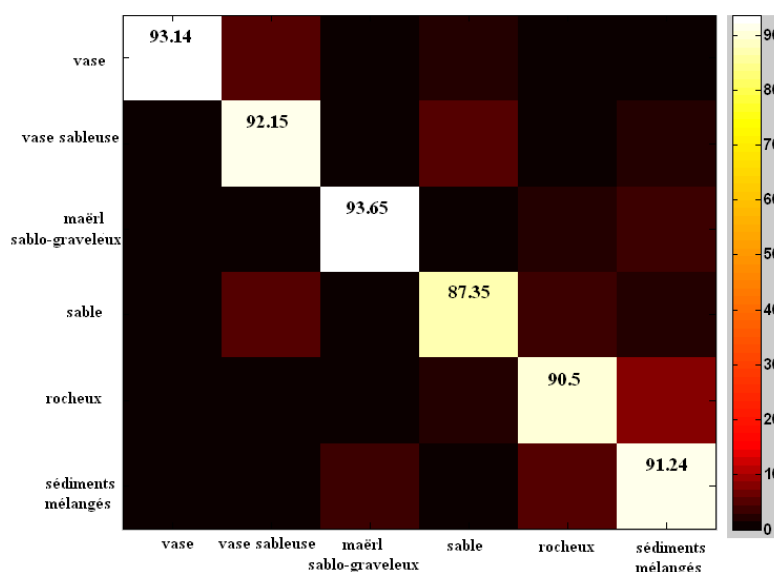


Figure 5.9 — Matrice de confusion de la classification supervisée de textures sonars utilisant l'approche LGCM. 5 échantillons d'apprentissage par classe sont utilisés pour cette expérimentation.

5.6 Conclusion

Nous avons présenté dans ce chapitre la problématique générale de l'acquisition et de l'analyse des données d'images acoustiques sous-marines. Nous avons en particulier introduit une nouvelle base de données des images texturées sonars comportant six différentes classes du fond marin : la vase, la vase sableuse, le maërl sablo-graveleux, le sable fin à moyen, le substratum rocheux et les sédiments mélangés. Nous croyons que cette étude pourrait fournir une base de données de référence pour les travaux futurs sur l'analyse et la classification des images texturées.

Nous avons en outre montré que les approches d'analyse statistique des distributions de points d'intérêt, développés pour des applications d'analyse de la texture visuelle, fournissent des signatures discriminantes pour les images sonars en considérant les changements de contraste local et les distorsions géométriques dans les observations sonar du fond marin. L'amélioration des performances de classification des fonds marins pour les modèles de Cox log-Gaussien par rapport aux autres techniques de l'état de l'art a été rapporté. Ceci démontre la pertinence de l'information apportée en termes d'utilisation d'un modèle statistique pour l'analyse d'organisation spatiale des points d'intérêt.

Conclusion générale et perspectives

Dans cette étude, nous avons considéré les aspects suivants :

- La réalisation du processus ponctuel marqué par les statistiques descriptives spatiales de cooccurrences basées sur la caractérisation de l'organisation spatiale des objets élémentaires d'image et de leurs caractéristiques pour une représentation d'image en qualité d'invariance.
- L'estimation paramétrique de la structure de covariance des points marqués déterminé par le processus de Cox log-Gaussien pour définir un vecteur descripteur invariant d'image.
- L'adaptation des méthodes proposées à la caractérisation et l'interprétation du contenu des images réelles dans de cas d'études : les textures réelles et les images de la scène.
- La reconnaissance de différents types de fond marin à partir de la caractérisation et la classification des textures sonars.

Les statistiques descriptives spatiales de cooccurrences

Nous avons appliqué la théorie des processus ponctuels spatiaux à l'analyse de l'organisation des signatures élémentaires (points d'intérêt ou formes) avec un vecteur descripteur attachée dans une image. La statistique spatiale de cooccurrence des points de processus ponctuels marqués est caractérisée par la mesure de moment factoriel d'ordre 2 et l'estimation de la fonction K de Ripley. De manière générale, les descripteurs proposés héritent des propriétés d'invariance des points d'intérêt et des formes. Les complexités de ces différents descripteurs sont directement liées au nombre de catégories des mots visuels et aux valeurs de rayon des boules d'étude.

Nous avons proposé des procédures pour réduire la taille de descripteurs et à résoudre les variations d'échelle des images.

L'estimation de modèles de Cox log-Gaussien

Nous avons présenté une estimation des paramètres d'un modèle de Cox log-Gaussien pour définir un nouveau descripteur invariant pour la caractérisation et l'interprétation du contenu des images basée sur l'analyse des distributions spatiales des signatures locales dans l'image. Nous avons étudié différentes formes des paramètres de la fonction de covariance pour montrer la pertinence de ce modèle à exploiter les distributions des données de points multivariées. Cela nous a permis de réaliser des

comparaisons effectives et quantitatives entre les différentes formes des modèles formels spécifiques. Nous avons également montré qu'il est possible de réduire la complexité des modèles et de résoudre le problème des variations d'échelles des images.

Les applications à la caractérisation des textures réelles et des images de scènes

Les applications sur les thèmes de classification des images texturées et de reconnaissance de scènes démontrent l'intérêt des approches proposées en terme de robustesse vis-à-vis des méthodes de l'état de l'art. Nous avons considéré les algorithmes d'apprentissage supervisés : kPPs, SVMs et RFs avec différentes distances de mesure de similarité : Euclidienne, Chi2, la divergence de Jensen-Shannon. Ici, nous avons envisagé les performances d'approches proposées dans des cas différents en fonction du choix des classifieurs et les distances de similarité pour l'apprentissage supervisée et de la sélection de la fonction de covariance de l'estimation du modèle de Cox log-Gaussien. Les différentes méthodes d'extraction de points d'intérêt et leurs descriptions ont été examinées : DoG+Sift, FH+Surf, (Har-Lap)+(Sift-Spin), (Hes-Lap)+Daisy, FH+Brief. En outre, l'algorithme FLST (Fast Level Set Transform) pour extraire un arbre des formes et l'algorithme IDSC (Inner-Distance Shape Context) ont été utilisés pour l'application à la reconnaissance de scènes.

L'application à la caractérisation des textures sonars

En considérant les images sonars du projet Rebut qui nous permettent de détecter dans les petits fonds, les limites de peuplements biologiques et fournir la structuration et la répartition des organismes benthiques denses, nous avons construit une base de données de 240 images sonars de 256 x 256 pixels acquises avec des valeurs différentes d'angles d'incidence, d'échelles et de niveaux moyens d'écho rétro-diffusé. Comparativement aux méthodes de l'état de l'art, nous avons également montré les améliorations de performances de classification des fonds marins réalisés avec les méthodes proposées, particulièrement avec les modèles de Cox log-Gaussien.

Perspectives :

Plusieurs perspectives sont envisageables pour prolonger ces travaux de thèse s'articulant autour de catégories de recherche. En premier lieu, il nous semble intéressant de poursuivre dans l'étude des statistiques spatiales des points marqués. D'autre part, des études complémentaires sur la recherche dans le domaine de la caractérisation et l'interprétation du contenu des images sonars pourraient être menées.

Les méthodes de la modélisation statistique spatiale de points marqués

D'un point de vue méthodologique, nous avons montré que l'analyse des processus ponctuels spatiaux avec les modèles de Cox log-Gaussien fournit une description mathématique formelle à l'analyse de la structure spatiale et aux signatures visuelles d'un ensemble des points marqués dans l'image. Les avantages des modèles paramétriques des processus ponctuels marqués ont été soulignés dans leur capacité à fournir une représentation plus compacte de l'information du contenu dans l'image. Il nous semble pertinent de poursuivre cet aspect. Les points suivants devraient être abordés :

- La définition d'une distance entre deux modèles proposés : statistique descriptive de cooccurrence et modèle de Cox log-Gaussien, à partir d'une distance entre deux

champs gaussiens. Il serait possible de s'inspirer d'outils statistiques, tels que les tests d'hypothèses ou les tests de goodness-of-fit, pour évaluer la signification des distances.

- Il existe dans la littérature d'autres modèles statistiques spatiaux de points, par exemple : processus de Neyman-Scott, de Cox shot-noise ou de Strauss [28, 34, 99]. Une comparaison pourra être considérée pour trouver le modèle le plus efficace et le plus flexible pour différents cas d'études.
- La combinaison de processus ponctuels marqués et de méthodes de segmentation avancée [142] est également parmi les problématiques importantes dans les travaux futurs.

La caractérisation et l'interprétation du contenu des images sonars

Les résultats obtenus dans ces travaux de thèse ouvrent une porte à d'autres applications de l'imagerie sonar basées sur l'analyse des processus ponctuels marqués et l'adaptation des schémas de points d'intérêt, y compris les AUV (Autonomous Underwater Vehicle) de navigation [154] et la détection des mines [155]. En outre, des méthodes automatiques pour la segmentation des fonds d'intérêt, de mines, d'un objet déposé sur le fond, ou la détection du changement de textures du fond marin tout en prenant en compte le facteur temporel pourraient également être testées. Particulièrement, une approche de segmentation variationnelle par la technique des ensembles de niveaux (Level Set) pour séparer les régions homogènes du fond marin pourrait être envisagée avec l'utilisation des descripteurs proposés dans ce travail [140].

Annexe 1 :

Preuve des résultats des fonctions : d'intensité et de corrélation de paires repondérée.

Nous considérons les processus de Cox log-Gaussien X gouverné par les fonctions d'intensités aléatoires $\Lambda_i = \{\Lambda_i(s) = \exp(Y_i(s)) : s \in \mathbb{R}^2\}$ sur tout borélien $B_i \subset \mathbb{R}^2$ et ses fonctions de moyenne $\{\mu_i\} = E\{Y_i(s)\}$, de variance $\sigma^2 = Var(Y_i(s))$ et de covariance $c_{ij}(r) = Cov(Y_i(s_1), Y_j(s_2))$, où $r = \|s_1 - s_2\|$ est la distance entre deux point s_1 et s_2 et $\{Y_i(s)\}$ est un processus Gaussien.

• La fonction d'intensité

$$\begin{aligned}\rho(s) &= E[\Lambda] = \int_B \exp(y) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_B \exp\left(-\frac{y^2 - 2y(\mu + \sigma^2) + \mu^2}{2\sigma^2}\right) dy \\ &= \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{(\mu + \sigma^2)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \int_B \exp\left(-\frac{(y - (\mu + \sigma^2))^2}{2\sigma^2}\right) dy \\ &= \exp(\mu + \sigma^2/2)\end{aligned}\tag{5.2}$$

$$\text{où } \frac{1}{\sqrt{2\pi\sigma^2}} \int_B \exp\left(-\frac{(y - (\mu + \sigma^2))^2}{2\sigma^2}\right) dy = 1.$$

• La fonction de corrélation de paires repondérée

Nous notons $\hat{r}$ le coefficient de corrélation entre variables aléatoires gaussiennes, la

mesure de moment d'ordre 2 de ces processus pour le cas de $\mu = 0$ est donnée par :

$$\begin{aligned}
 \rho^{(2)}(s_1, s_2) &= E[\Lambda_1, \Lambda_2] \\
 &= \int_{B_1 \times B_2} \exp(y_1 + y_2) \frac{1}{2\pi\sigma^2\sqrt{1-\hat{r}^2}} \exp\left(-\frac{1}{(1-\hat{r}^2)\sigma^2}(y_1^2 - \hat{r}y_1y_2 + y_2^2)\right) dy_1 dy_2 \\
 &= \int_{B_1 \times B_2} \frac{1}{2\pi\sigma^2\sqrt{1-\hat{r}^2}} \exp\left(-\frac{1}{(1-\hat{r}^2)\sigma^2}(y_1^2 - (2\hat{r}y_1y_2 + 2(1-\hat{r}^2)\sigma^2(y_1 + y_2)) + y_2^2)\right) dy_1 dy_2
 \end{aligned} \tag{5.3}$$

Considérons l'identification :

$$\begin{aligned}
 y_1^2 - (2\hat{r}y_1y_2 + 2(1-\hat{r}^2)\sigma^2(y_1 + y_2)) + y_2^2 \\
 &= (y_1 - \kappa)^2 - 2\hat{r}((y_1 - \kappa)(y_2 - \kappa)) + (y_2 - \kappa)^2 + A \\
 &= y_1^2 - (2\hat{r}y_1y_2 - (2\hat{r}\kappa - 2\kappa)(y_1 + y_2)) + y_2^2 + A + 2\kappa^2 - 2\hat{r}\kappa^2
 \end{aligned}$$

Nous obtenons :

$$\begin{cases} 2(1-\hat{r}^2)\sigma^2 = -2(\hat{r}-1)\kappa \\ A + 2\kappa^2 - 2\hat{r}\kappa^2 = 0 \end{cases} \tag{5.4}$$

$$\begin{cases} \kappa = \sigma^2(1+\hat{r}) \\ A = 2\kappa^2(\hat{r}-1) = 2\hat{r}^4(\hat{r}^2-1)(\hat{r}+1) \end{cases} \tag{5.5}$$

Remplaçant les valeurs de κ et A dans l'équation 5.3, nous avons :

$$\rho^{(2)}(s_1, s_2) = \exp\left(-\frac{A}{2(1-\hat{r}^2)\sigma^2}\right) = \exp((1+\hat{r})\sigma^2) \tag{5.6}$$

Alors la fonction de corrélation de paires repondérée est donnée par :

$$g(s_1, s_2) = \frac{\rho^{(2)}(s_1, s_2)}{\rho^{(1)}(s_1)\rho^{(1)}(s_2)} = \frac{\exp((1+\hat{r})\sigma^2)}{\exp(\sigma^2/2)\exp(\sigma^2/2)} = \exp(\hat{r}\sigma^2) = \exp(c(r)) \tag{5.7}$$

où $\hat{r}\sigma^2 = E[Y_1, Y_2] = c(r)$.

Bibliographie

- [1] M. Grand-brochier, C. Tilmant, and M. Dhome, "Descripteur local d'image invariant aux transformations affines. local affine-invariant image descriptor," *Congrès des jeunes chercheurs en vision par ordinateur ORASIS*, 2009.
- [2] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Recognition of shapes by editing their shock graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26(5), pp. 550–571, 2004.
- [3] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape context," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 509–522, 2002.
- [4] H. Ling and D. Jacobs, "Deformation invariant image matching," *International Conference on Computer Vision*, pp. 1466–1473, 2005.
- [5] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features : Spatial pyramid matching for recognizing natural scene categories," *Computer Vision and Pattern Recognition*, pp. 2169–2178, 2006.
- [6] J. Aujol, G. Aubert, and L. Blanc-Féraud, "Wavelet-based level set evolution for classification of textured images," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1634–1641, 2003.
- [7] R. Haralick, "Statistical and structural approaches to textures," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, May, 1979.
- [8] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," *European Conference on Computer Vision*, pp. 1–22, 2004.
- [9] H. Ling and S. Soatto, "Proximity distribution kernels for geometric context in category recognition." *International Conference on Computer Vision*, pp. 1–8, 2007.
- [10] Y. Xu, H. Ji, and C. Fermüller, "Viewpoint invariant texture description using fractal analysis," *International Journal of Computer Vision*, vol. 83(1), pp. 85–100, 2009.
- [11] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories : a comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, 2007.
- [12] J. Davies, J. Baxter, M. Bradley, D. Connor, J. Khan, E. Muray, W. Sanderson, C. Turnbull, and M. Vincent, "Marine monitoring handbook(uk marine sacs project)," 2001.

- [13] S. Degraer, V. van Lancker, G. Moerkerke, G. V. Hoey, M. V. P. Jacobs, and J.P.Henriet, "Intensive evaluation of the evolution of a protected benthic habitat : Habitat. final report of the federal office for scientific, technical and cultural affairs (ostc) e ministry of the flemish community, environment and infrastructure." *Waterways and Marine Affairs Administration, Coastal Waterways*, 2002.
- [14] A. Ehrhold, D.Hamon, and B. Guillaumont, "The rebent monitoring network, a spatially integrated, acoustic approach to surveying nearshore macrobenthic habitats : application to the bay of concarneau(france)." *ICES Journal of Marine Science*, vol. 63(9), pp. 1604–1615, 2006.
- [15] C. Hu, F. Muller-Karger, B. Murch, D. Myhre, J. Taylor, R. Luerssen, C. Moses, C. Zhang, L. Gramer, and J. Hendee, "Building an automated integrated observing system to detect sea surface temperature anomaly events in the florida keys," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47(6), pp. 1607–1620, 2009.
- [16] R. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.
- [17] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 842–848, 1996.
- [18] V. Caselles, B. Coll, and J.-M. Morel, "Topographic maps and local contrast changes in natural images," *International Journal of Computer Vision*, vol. 33(1), pp. 5–27, 1999.
- [19] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf : Speeded up robust features," *European Conference on Computer Vision*, vol. 1, pp. 404–417, 2006.
- [20] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief : Binary robust independent elementary features." *European Conference on Computer Vision*, vol. IV, pp. 778–792, 2010.
- [21] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60(2), pp. 91–110, 2004.
- [22] E. Tola, V. Lepetit, and P. Fua, "Daisy : An efficient dense descriptor applied to wide baseline stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, pp. 815–830, 2010.
- [23] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27(10), pp. 1615–1630, 2005.
- [24] M. Cummins and P. Newman, "FAB-MAP : Probabilistic Localization and Mapping in the Space of Appearance," *International Journal of Robotics Research*, vol. 27(6), pp. 647–665, 2008.
- [25] S. Savarese, J.-M. Winn, and A. Criminisi, "Discriminative object class models of appearance and shape by correlations," *Computer Vision and Pattern Recognition*, vol. 2, pp. 2033–2040, 2006.

- [26] L. Waller and C. Gotway, "Applied spatial statistics for public health data," *Wiley Newyork*, 2004.
- [27] A. Lawson, "Statistical methods in spatial epidemiology," *Wiley Newyork*, 2001.
- [28] P. Diggle, B. Rowlingson, and T. Su, "Point process methodology for on-line spatio-temporal disease surveillance," *Environmetrics*, vol. 16, no. 5, pp. 423–434, 2005.
- [29] J. Perry, A. Liebhold, M. Rosenberg, J. Dungan, M. Miriti, A. Jakomulska, and S. Citron-Pousty, "Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data," *Ecography*, vol. 25, pp. 578–600, 2002.
- [30] D. Stoyan and A. Penttinen, "Recent applications of point process methods in forestry statistics," *Statistical Science*, vol. 15, no. 1, pp. 61–78, 2000.
- [31] R. Law, J. Illian, D. Burslem, G. Gratzer, C. Gunatilleke, and I. Gunatilleke, "Ecological information from spatial patterns of plants : insights from point process theory," *Journal of Ecology*, vol. 97(4), pp. 616–628, 2009.
- [32] J. McGrew and C. Monroe, "An introduction to statistical problem-solving in geography," *Wm C Brooks, NewYork*, 1993.
- [33] J. Neyman and E. Scott, "Statistical approach to problems of cosmology," *Journal of the Royal Statistical Society*, vol. B(20), pp. 1–43, 1958.
- [34] F. Lafarge, G. Gimel'farb, and X. Descombes, "Geometric feature extraction by a multi-marked point process," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1597–1609, 2010.
- [35] L. Linnett, D. Carmichael, and S. Clarke, "Texture classification using a spatial-point process model," *IEE Vision, Image and Signal Processing*, vol. 142, no. 1, pp. 1–6, 1995.
- [36] S. Lazebnik, C. Schmid, and J. Ponce, "Sparse texture representation using affine-invariant neighborhoods," *Computer Vision and Pattern Recognition*, pp. 319–324, 2003.
- [37] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schafalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, pp. 43–72, 2005.
- [38] T. Tuytelaars and L. V. Gool, "Matching widely separated views based on affine invariant regions," *International Journal of Computer Vision*, vol. 59(1), pp. 61–85, 2004.
- [39] C. Harris and M. Stephens, "A combined corner and edge detector," *Proceedings of the Alvey Vision Conference*, pp. 147–151, 1988.
- [40] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30(2), pp. 79–116, 1998.
- [41] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," *International Conference on Computer Vision*, pp. 525–531, 2001.
- [42] S. Lazebnik, C. Schmid, and J. Ponce, "A maximum entropy framework for part-based texture and object recognition," *International Conference on Computer Vision*, vol. 1, pp. 832–838, 2005.

- [43] M. Varma and R. Garg, "Locally invariant fractal features for statistical texture classification," *International Conference on Computer Vision*, vol. 1, pp. 1–8, 2007.
- [44] J. Liu and M. Shah, "Scene modeling using co-clustering," *International Conference on Computer Vision*, pp. 1–7, 2007.
- [45] M. Brown and D. Lowe, "Recognising panoramas," *International Conference on Computer Vision*, pp. 1218–1227, 2003.
- [46] M. Hu, "Visual pattern recognition by moments invariants," *IEEE Transactions on Information Theory*, vol. 8, pp. 179–187, 1962.
- [47] A. Khotanzad and Y. Hong, "Invariant image recognition by zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12(5), pp. 289–497, 1990.
- [48] F. Ghorbel, "A complete invariant description for graylevel images by the harmonic analysis," *Pattern Recognition letters*, vol. 15, pp. 1043–1051, 1994.
- [49] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42(3), pp. 425–436, 2009.
- [50] Y. K. and R. Sukthankar, "Pca-sift : a more distinctive representation for local image descriptors," *Computer Vision and Pattern Recognition*, vol. 2, pp. 506–513, 2004.
- [51] W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13(9), pp. 891–906, 1991.
- [52] L. Gool, T. Moons, and D. Ungureanu, "Affine/ photometric invariants for planar intensity patterns," *European Conference on Computer Vision*, vol. 1, pp. 642–651, 1996.
- [53] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets," *European Conference on Computer Vision*, pp. 414–431, 2002.
- [54] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images : A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24(1), pp. 34–58, 2002.
- [55] A. Almansa and T. Lindeberg, "Fingerprint enhancement by shape adaptation of scale-space operators with automatic scale selection," *IEEE Transactions on Image Processing*, vol. 9(12), pp. 2027 – 2042, 2000.
- [56] S. Wang, W. Zhu, and Z.-P. Liang, "Shape deformation : Svm regression and application to medical image segmentation," *International Conference on Computer Vision*, vol. 2, pp. 209–216, 2001.
- [57] I. Sobel and G. Feldman, "A 3x3 isotropic gradient operator for image processing," *a talk at the Stanford Artificial Project*, 1968.
- [58] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–714, 1986.
- [59] J. Shen and S. Castan, "An optimal linear operator for edge detection," *Computer Vision and Pattern Recognition*, pp. 109–114, 1986.

- [60] L. Vincent and P. Soille, "Watershed in digital spaces, an efficient algorithm based on immersion simulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13(6), pp. 583–598, 1991.
- [61] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes : Active contour models," *International Journal of Computer Vision*, vol. 1(4), pp. 321–331, 1987.
- [62] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22(1), pp. 61–79, 1997.
- [63] T. Cootes and C. Taylor, "Active shape models - smart snakes," *British Machine Vision Conference*, pp. 266–275, 1992.
- [64] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23(6), pp. 681–685, 2001.
- [65] S. Ruan, D. Bloyet, M. Revenu, W. Dou, and Q. Liao, "Cerebral magnetic resonance image segmentation using fuzzy markov random fields," *ISBI*, pp. 237–240, 2002.
- [66] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. Resnick, and C. Davatzikos, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," *Neuroimage*, vol. 22(1), pp. 46–57, 2004.
- [67] R. Sammouda, N. Niki, and H. Nishitani, "A comparison of hopfield neural network and boltzmann machine in segmenting mr images of the brain," *TNS*, vol. 43(6), pp. 3361–3369, 1996.
- [68] J. Xue, S. Ruan, B. Moretti, M. Revenu, and D. Bloyet, "Knowledge-based segmentation and labeling of brain structures from mri images," *TNS*, vol. 22(3-4), pp. 395–405, 2001.
- [69] S. Rital, H. Cherifi, and S. Miguët, "Weighted adaptive neighborhood hypergraph partitioning for image segmentation," *International Conference on Advances in Pattern Recognition*, vol. 3687, pp. 522–531, 2005.
- [70] R. Zabih and V. Kolmogorov, "Spatially coherent clustering using graph cuts," *Computer Vision and Pattern Recognition*, vol. 2, pp. 437–444, 2004.
- [71] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-d images segmentation," *International Journal of Computer Vision*, vol. 70(2), pp. 109–131, 2006.
- [72] P. Monasse and F. Guichard, "Fast computation of a contrast invariant image representation," *IEEE Transactions on Image Processing*, vol. 9(5), pp. 860–872, 2000.
- [73] A. Kronrod, "On functions of two variables (en russe)," *Uspehi Mathematical Sciences*, vol. 5(35), pp. 24–134, 1950.
- [74] R. Basri, L. Costa, D. Geiger, and D. Jacobs, "Determining the similarity of deformable shapes," *Vision Research*, vol. 38, pp. 2365–2385, 1998.
- [75] L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt, "Shape representation and classification using the poisson equation," *Computer Vision and Pattern Recognition*, pp. 61–67, 2004.

- [76] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61(1), pp. 55–79, 2005.
- [77] H. Schneiderman and T. Kanade, "Object detection using the statistics of parts," *International Journal of Computer Vision*, vol. 56(3), pp. 151–177, 2004.
- [78] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker, "Shock graphs and shape matching," *International Journal of Computer Vision*, vol. 35(1), pp. 13–32, 1999.
- [79] P. Dimitrov, C. Phillips, and K. Siddiqi, "Robust and efficient skeletal graphs," *Computer Vision and Pattern Recognition*, pp. 417–423, 2000.
- [80] A. Diplaros and E. Milios, "Matching and retrieval of distorted and occluded shapes using dynamic programming," *IEEE Transactions on Pattern and Analysis and Machine Intelligence*, vol. 24(11), pp. 1501–1516, 2002.
- [81] C. Veltkamp and M. Hagedoorn, "State of the art in shape matching," *Principles of visual information retrieval*, vol. 33, pp. 87–119, 2001.
- [82] A. D. Bimbo and P. Pala, "Shape indexing by multiscale representation," *Image and Vision Computing*, vol. 17(3), pp. 245–261, 1999.
- [83] F. Mokhtarian and M. Bober, "Curvature scale space representation : theory, applications, and mpeg-7 standardization," *Kluwer Academic Publishers, Norwell, MA, USA*, 2003.
- [84] J. Zhang, X. Zhang, H. Krim, and G. Walter, "Object representation and recognition in shape spaces," *Pattern Recognition*, vol. 36(5), pp. 1143–1154, 2003.
- [85] K. Grauman and T. Darrell, "The pyramid match kernel : Discriminative classification with sets of image features," *International Conference on Computer Vision*, vol. 2, pp. 1458 – 1465, 2005.
- [86] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [87] M. Varma and A. Zisserman, "Classifying images of materials : achieving view-point and illumination independence," *European Conference on Computer Vision*, pp. 255–271, 2002.
- [88] J. Sivic and A. Zisserman, "Video google : a text retrieval approach to object matching in videos," *International Conference on Computer Vision*, vol. 2, pp. 1470 – 1477, 2003.
- [89] J. Macqueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Math, Statistics, and Probability*, vol. 1, pp. 281–297, 1967.
- [90] B. Leibe, K. Mikolajczyk, and B. Schiele, "Segmentation based multi-cue integration for object detection," *British Machine Vision Conference*, 2006.
- [91] S. Agarwal and A. Awan, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26(11), pp. 1475–1490, 2004.

- [92] A. Bosch, A. Zisserman, and X. Munoz, "Scene classification via plsa," *European Conference on Computer Vision*, pp. 517–530, 2006.
- [93] D. Larlus and F. Jurie, "Latent mixture vocabularies for object categorization," *British Machine Vision Conference*, 2006.
- [94] H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Invariant descriptors of sonar textures from spatial statistics of local features." *International Conference on Acoustics, Speech and Signal Processing*, pp. 1674–1677, 2010.
- [95] J. Møller and R. Waagepetersen, "Modern statistics for spatial point processes," *Scandinavian Journal of Statistics*, vol. 34, pp. 643–684, 2007.
- [96] C. Gaetan and X. Guyon, "Modélisation et statistique spatiales," *Springer Mathématiques et applications*, 2008.
- [97] P. Diggle, "Statistical analysis of spatial point patterns." *Academic Press, London*, 1983.
- [98] M. Schlather, "On the second-order characteristics of marked point processe," *Bernoulli*, vol. 7, pp. 99–117, 2001.
- [99] D. Stoyan and H. Stoyan, "Fractals, random shapes and point fields," *Wiley, Chichester*, 1994.
- [100] B. Ripley, "Spatial statistics," *J. Wiley, New York, NY*, 1981.
- [101] P. Dixon, "Ripley's k function," *Encyclopedia of Environmetrics*, vol. 3, pp. 1796–1803, 2002.
- [102] F. Goreaud and R. Péliissier, "On explicit formulas of edge effect correction for ripley's k-function," *Journal of Vegetation Science*, vol. 10, no. 3, pp. 433–438, 1999.
- [103] B. Ripley, "Edge effects in spatial stochastic processes," in *Bo Ranney ed : Statistics in Theory and Practice, Swedish University of Agricultural Sciences, Section of Forest Biometry, Umea, Sweden*, vol. 3, pp. 247–262, 1982.
- [104] H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Spatial statistics of visual keypoints for texture recognition." *European Conference on Computer Vision*, pp. 764–777, 2010.
- [105] J. Møller, A. Syversveen, and R. Waagepetersen, "Log gaussian cox processes," *Scandinavian Journal of Statistics*, vol. 25(3), pp. 451–482, 1998.
- [106] A. Brix and J. Møller, "Space-time multi type log gaussian cox processes with a view to modelling weeds," *Scandinavian Journal of Statistics*, vol. 28, pp. 471–488, 2001.
- [107] J. Kingman, "Poisson processes," *Clarendon Press, Oxford*, 1993.
- [108] N. Best, K. Ickstadt, and R. Wolpert, "Spatial poisson regression for health and exposure data measured at disparate resolutions," *Journal of the American Statistical Association*, vol. 95, pp. 1076–1088, 2000.
- [109] D. Cox, "Some statistical models related with series of events," *Journal of the Royal Statistical Society*, vol. B(17), pp. 129–164, 1955.
- [110] J. Grandell, "Mixed poisson processes," *Chapman and Hall, London*, 1997.

- [111] D. Daley and D. Vere-Jones, "An introduction to the theory of point processes," *Springer-Verlag, New York*, vol. I, 2003.
- [112] B. Matérn, "Doubly stochastic poisson processes in the plane," *Statistical Ecology (edited by G. P. Patil, E. C. Pielou and W. E. Waters)*, vol. 1, pp. 195–213, 1971.
- [113] J. Møller and G. Torrisi, "Generalised shot noise cox processes," *Advances in Applied Probability*, vol. 37), pp. 48–74, 2005.
- [114] C. Preston, "Random fields," *Lecture Notes in Mathematics 534, Springer-Verlag, Berlin*, 1976.
- [115] X. Nguyen and H. Zessin, "Integral and differential characterizations of gibbs processes," *Mathematische Nachrichten*, vol. 88, pp. 105–115, 1979.
- [116] D. Strauss, "A model for clustering," *Biometrika*, vol. 62(2), pp. 467–475, 1975.
- [117] S. Descamps, X. Descombes, A. Béchet, and J. Zerubia, "Détection de flamants roses par processus ponctuels marqués pour l'estimation de la taille des populations," *Traitement du Signal*, vol. 26(2), pp. 95–108, 2009.
- [118] A. Baddeley, J. Møller, and R. Waagepetersen, "Non- and semi-parametric estimation of interaction in inhomogeneous point patterns," *Statistica Neerlandica*, vol. 54, pp. 329–350, 2000.
- [119] H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Visual textures as realizations of multivariate log-gaussian cox processes." *Computer Vision and Pattern Recognition*, pp. 2945–2952, 2011.
- [120] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning," *Second edition. Springer*, 2009.
- [121] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" *International Conference on Database Theory*, pp. 217–235, 1999.
- [122] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [123] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.
- [124] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on neural networks*, vol. 13, pp. 415–425, 2002.
- [125] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
- [126] L. Breiman, "Bagging predictors," *Machine learning*, vol. 26(2), pp. 123–140, 1996.
- [127] P. Besse, "Data mining et statistique : Réponse aux intervenants," *Journal de la Société Française de Statistique*, vol. 142, pp. 89–95, 2001.
- [128] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees," *Chapman and Hall*, 1984.
- [129] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Processing*, vol. 18, pp. 349–369, 1989.

- [130] S. Aksoy and R. Haralick, "Probabilistic vs. geometric similarity measures for image retrieval," *Computer Vision and Pattern Recognition*, vol. 2, pp. 2357–2362, 2000.
- [131] N. Vasconcelos and A. Lippman, "A probabilistic architecture for content-based image retrieval," *Computer Vision and Pattern Recognition*, pp. 1216–1221, 2000.
- [132] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," *Computer Vision and Pattern Recognition*, pp. 524–531, 2005.
- [133] A. Oliva and A. Torralba, "Modeling the shape of the scene : a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, pp. 145–175, 2001.
- [134] J. Sivic, B. Russell, A. Efros, A. Zisserman, and B. Freeman, "Discovering objects and their location in images," *International Conference on Computer Vision*, vol. 1, pp. 370–377, 2005.
- [135] C. Brown, K. Cooper, W. Meadows, D. Limpenny, and H. Rees, "Small-scale mapping of sea-bed assemblages in the eastern english channel using sidescan sonar and remote sampling techniques." *Estuarine, Coastal and Shelf Science.*, vol. 54, pp. 263–278, 2002.
- [136] G. R. Cochran and K. Lafferty, "Use of acoustic classification of sidescan-sonar data for mapping benthic habitat in the northern channel islands, california." *Continental Shelf Research.*, vol. 22, pp. 683–690, 2002.
- [137] A. Ehrhold, A. Blanchet, D. Hamon, C. Chevalier, J. Gaffet, and A. Alix, "Réseau de surveillance benthique -région bretagne," *Rapport technique - IFREMER*, 2007.
- [138] X. Lurton, "An introduction to underwater acoustics : Principles and application," *Edition Springer - ISBN 3-540-42967-0*, 2002.
- [139] R. J. Urick, "Principles of underwater sound." *McGraw-Hill Book Company, New York*, p. 423, 1983.
- [140] I. Karoui, "Segmentation par méthodes markoviennes et variationnelles des images texturées : application à la caractérisation sonar des fonds marins." *Thèse de doctorat, Université de Rennes 1, mars 2007*, mars 2007.
- [141] G. L. Chenadec and J. Boucher, "Sonar image segmentation using the angular dependence of backscattering distributions." *Oceans*, vol. 1, pp. 147–152, 2005.
- [142] I. Karoui, R. Fablet, J.-M. Boucher, and J.-M. Augustin, "Seabed segmentation using optimized statistics of sonar textures," *IEEE Transaction on Geosciences and Remote Sensing*, vol. 47(6), pp. 1621–1631, 2009.
- [143] G. L. Chenadec, J.-M. Boucher, and X. Lurton, "Angular dependence of k-distributed sonar data," *IEEE Transaction on Geosciences and Remote Sensing*, vol. 45, pp. 1124–1235, 2007.
- [144] J. Hughes-Clarke, "Toward remote seafloor classification using the angular response of acoustic backscatter : a case study from multiple overlapping gloria data," *IEEE Journal of Oceanic Engineering*, vol. 19(1), pp. 112–127, 1994.

- [145] D. Jackson, D. Winebrenner, and A. Ishimaru, "Application of the composite roughness model to high frequency bottom backscattering." *Journal of the Acoustical Society of America*, pp. 1410–1422, 1986.
- [146] G. Chenadec, J. Boucher, X. Lurton, and J. Augustin, "Angular dependence of statistical distributions for backscattered signals : modeling and application to multibeam echosounder data." *Oceans*, vol. 2, pp. 897–903, 2003.
- [147] S. Dugelay, "Caractérisation des fonds marins à partir de données sondeur em12." *Thèse de doctorat, Département de Développement de Services Informatiques, Ifremer*, mars 1997.
- [148] C. Oliver, "A model for non-rayleigh scattering statistics," *Optica Acta*, vol. 31(6), pp. 701–722, 1984.
- [149] H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Log-gaussian cox processes of visual keypoints for sonar texture recognition." *International Conference on Acoustics, Speech and Signal Processing*, 2011.
- [150] —, "Keypoint-based analysis of sonar images : application to seabed recognition." *IEEE Transactions on Geoscience and Remote Sensing*, 2011.
- [151] Y. Wang, Z. Liu, E. Sang, and H. Ma, "Sonar image classification based on directional wavelet and fuzzy fractal dimension," *IEEE Conference on Industrial Electronics and Applications*, pp. 118 – 120, 2007.
- [152] N.Sun and T. Shim, "Sonar image classification based of seabed physiognomy based on the information fusion methods," *Congress on Image and Signal Processing*, pp. 813 – 819, 2008.
- [153] T. Randen and J. H. Husøy, "Filtering for texture classification : A comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 291–310, 1999.
- [154] S. Botelho, P. Junior, M. Figueiredo, C. Rocha, and G. Oliveira, "Appearance-based odometry and mapping with feature descriptors for underwater robots," *Journal of the Brazilian Computer Society*, vol. 15(3), pp. 47–54, 2009.
- [155] T. Wang, I. Gu, and T. Tjahjadi, "Enhanced landmine detection from low resolution image sequences," *Computer Analysis of Images and Patterns*, vol. 5702, pp. 1236–1244, 2009.