



Classification multi-vues d'un objet immergé à partir d'images sonar et de son ombre portée sur le fond

Isabelle Quidu

► To cite this version:

Isabelle Quidu. Classification multi-vues d'un objet immergé à partir d'images sonar et de son ombre portée sur le fond. Traitement du signal et de l'image [eess.SP]. Université de Bretagne occidentale - Brest, 2001. Français. NNT : . tel-00504902

HAL Id: tel-00504902

<https://theses.hal.science/tel-00504902>

Submitted on 21 Jul 2010

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

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

N° d'ordre :

THÈSE

présentée à

L'Université de Bretagne Occidentale

U.F.R. Sciences et Techniques

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

Mention : Électronique

par

Isabelle QUIDU

Classification Multi-vues d'un Objet Immergé à partir d'Images Sonar et de son Ombre Portée sur le Fond
--

soutenue le 11 décembre 2001 devant la Commission d'Examen composée de :

Rapporteurs

Joseph RONSIN
Christophe COLLET

Professeur à l'Institut National des Sciences Appliquées de Rennes
Professeur à l'Université Louis Pasteur, Strasbourg I

Examineurs

Jean-Marie NICOLAS

Professeur à l'École Nationale Supérieure des Télécommunications
de Paris

Benoît ZERR

Chef du Service d'Acoustique Haute Fréquence
Groupe d'Études Sous-Marines de l'Atlantique

Jacques MEUDEC

Directeur Affaires et Projets Guerre des Mines
Thales Underwater Systems S.A.S.

Gilles BUREL

Professeur à l'Université de Bretagne Occidentale

Pierre VILBÉ

Professeur à l'Université de Bretagne Occidentale

Jean-Philippe MALKASSE

Chef de groupe Traitement du Signal

Service Études Générales Sonar, Thales Underwater Systems S.A.S.

Travail faisant l'objet de la convention CIFRE n° 033/98 et effectué au sein du Service Études Générales Sonar de Thales Underwater Systems S.A.S. à Brest en liaison avec le Laboratoire d'Électronique et Systèmes de Télécommunications.

THALES



LEST FRE CNRS 2269

Remerciements

Ces travaux ont été réalisés dans le cadre d'une Convention Industrielle de Formation par la Recherche, bourse octroyée par l'Association Nationale de la Recherche Technique. En ce sens, je tiens en premier lieu à remercier mes partenaires à savoir la société Thales Underwater Systems S.A.S. à travers le Directeur Affaires et Projets Guerre des Mines, Monsieur Jacques Meudec, et le Laboratoire d'Électronique et de Systèmes de Télécommunications de l'Université de Bretagne Occidentale à travers le Directeur, Monsieur Michel Ney.

En second lieu, j'adresse mes remerciements à Monsieur Didier Billon et son successeur, Monsieur Franck Florin, pour m'avoir accueillie au sein de leur service Études Générales Sonar de Thales Underwater Systems S.A.S. à Brest en me permettant d'y effectuer mes recherches dans un contexte industriel privilégié. Cette position a fortement influé sur le caractère appliqué des recherches menées.

Je tiens particulièrement à remercier trois personnes qui ont participé à la réussite de cette thèse. J'ai, d'une part, bénéficié du soutien et de l'expérience de mes directeurs de thèse, Messieurs Pierre Vilbé et Gilles Burel, Professeurs à l'Université de Bretagne Occidentale. Qu'ils soient remerciés de leurs conseils avisés. J'exprime, d'autre part, ma gratitude à mon encadrant, Monsieur Jean-Philippe Malkasse, chef de groupe Traitement du Signal, pour son étroite collaboration, la pertinence de ses critiques et sa disponibilité.

Je remercie vivement Messieurs Joseph Ronsin, Professeur à l'Institut National des Sciences Appliquées de Rennes, et Christophe Collet, Professeur à l'Université de Strasbourg, pour avoir accepté le rôle de rapporteurs et consacré du temps à l'évaluation de l'ensemble de mes travaux. Je remercie également Messieurs Jean-Marie Nicolas, Professeur à l'ENST Paris et Benoît Zerr, Chef du Service d'Acoustique Haute Fréquence du GESMA de m'avoir fait l'honneur d'examiner mes travaux.

Une amicale pensée pour les bibliothécaires de la Bibliothèque Technique de Thales Underwater Systems S.A.S. à Brest, Mesdames Claudie Ollivier et Dominique Rousseau pour leur compétence et leur gentillesse.

Mes remerciements vont également à l'ensemble des personnes, amis et collègues, qui m'ont accompagnée de près ou de loin durant ces trois années, les ingénieurs du service et les différents stagiaires que nous avons accueillis, et aussi Jean-Bernard Henrion, Bruno Quellec, et tant d'autres du Service Développement Logiciel que j'ai eu le plaisir de retrouver en débutant cette thèse.

Je voudrais terminer en m'adressant à mes parents, mon frère et mes sœurs. Ce manuscrit leur est dédié en remerciement de la confiance qu'ils m'ont accordée et de leurs encouragements.

Sommaire

Liste des illustrations	13
Introduction générale	17
PARTIE I : Imagerie sonar : particularités et exploitation	21
I. NOTION DE SONAR [MARA-90]	23
II. PRINCIPE DE FORMATION DES IMAGES SONAR	23
II.1. Architecture du sonar	23
II.2. Caractérisation des signaux émis pour le sonar actif	24
II.3. Lien entre cellule de résolution du sonar et pixel de l'image	24
III. SPÉCIFICITÉ DE L'IMAGERIE SONAR HAUTE RÉOLUTION	26
III.1. Bruit en imagerie sonar [BOUV-92] [MART-94]	26
III.2. Visualisation d'une image sonar « haute résolution »	27
IV. GUERRE DES MINES : MENACE ET SOLUTIONS	28
IV.1. Menace	29
IV.2. Progrès technologiques	30
IV.3. Le principe de classification sur ombre	32
IV.4. Hypothèses	33
V. SIMULATION D'IMAGES SONAR	34
VI. MÉTHODOLOGIE EN TRAITEMENT D'IMAGES APPLIQUÉE À L'IMAGE SONAR	35

PARTIE II : Outils pour le traitement d'images sonar	37
INTRODUCTION DE LA PARTIE II	39
CHAPITRE II.1 : SEGMENTATION D'IMAGES SONAR	41
I. INTRODUCTION DU CHAPITRE II.1	41
II. CONNAISSANCES DU SONARISTE	43
II.1. Résolution	43
II.2. Attache aux données	44
III. ADAPTATION DE TECHNIQUES DE SEGMENTATION CLASSIQUES AU CONTEXTE SONAR	46
III.1. Modèles structurels pour une application sonar	46
III.1.1. Champs markoviens pour la segmentation	46
III.1.2. Contours actifs statistiques [CHES-99]	51
III.1.3. Comparaison des deux approches	54
III.2. Morphologie mathématique	55
IV. SPÉCIFICATION D'UN FILTRE SPATIAL POUR LE SEUILLAGE D'IMAGES SONAR	58
IV.1. Principe	58
IV.2. Optimisation du filtre spatial	58
IV.3. Mise en œuvre	60
IV.4. Résultats expérimentaux	63
CONCLUSION DU CHAPITRE II.1	67
CHAPITRE II.2 : EXTRACTION D'ATTRIBUTS CARACTÉRISTIQUES	69
I. INTRODUCTION DU CHAPITRE II.2	69
II. PRÉTRAITEMENTS	72
II.1. Normalisation de l'image sonar	72
II.2. Lissage du contour	74
III. ATTRIBUTS DESCRIPTIFS	76

III.1.	Décomposition en série de Fourier des coordonnées des points de contour	76
III.1.1.	Définition	76
III.1.2.	Propriétés	77
III.1.3.	Normalisation	78
III.2.	Les descripteurs de Fourier AW - invariants de Arbter [ARBT-89]	80
III.2.1.	Décomposition en série de Fourier et transformation affine	80
III.2.2.	Les A-invariants	82
III.2.3.	Les AW-Invariants	83
III.3.	Transformation de Fourier inverse	83
IV.	ATTRIBUTS SYNTHÉTIQUES	86
IV.1.	Paramètres intrinsèques [JAN-88] [PROK-92]	86
IV.1.1.	Paramètres géométriques	86
IV.1.2.	Paramètres topologiques	87
IV.2.	Attributs invariants vis-à-vis des transformations du plan	89
IV.2.1.	Invariance vis-à-vis des similitudes	89
IV.2.2.	Invariance vis-à-vis des transformations affines [FLUS-93]	90
V.	CONCLUSION DU CHAPITRE II.2	92
	CONCLUSION DE LA PARTIE II	93
	PARTIE III : Application mono-vue	95
	INTRODUCTION DE LA PARTIE III	97
	CHAPITRE III.1 : APPROCHE SÉQUENTIELLE	99
I.	INTRODUCTION DU CHAPITRE III.1	99
II.	OUTILS D'ÉVALUATION	100
II.1.	Classifieurs classiques [CELE-89]	100
II.2.	Analyse en composantes principales [VOLL-85] [CELE-89] [BOUR-98]	100
III.	EVALUATION DES ATTRIBUTS	102
IV.	HYBRIDATION [QUID-00A]	104
V.	CONCLUSION DU CHAPITRE III.1	106
	CHAPITRE III.2 : APPROCHES ALTERNATIVES	107

I. INTRODUCTION DU CHAPITRE III.2	107
II. APPARIEMENT DE SILHOUETTES	107
II.1. Introduction	107
II.2. Position du problème	108
II.3. Estimation des paramètres	108
II.3.1. Estimation de τ	108
II.3.2. Estimation de A	109
II.4. Mesure de similarité	110
II.5. Mise en œuvre pour les images sonar	112
II.6. Performances	114
II.7. Illustrations	116
II.8. Conclusion	118
III. SEGMENTATION ORIENTÉE CLASSIFICATION	120
III.1. Introduction	120
III.2. Algorithmes génétiques [DAVI-91] [MICH-94] [PAL-96]	122
III.2.1. Introduction	122
III.2.2. Formulation et terminologie	123
III.2.3. Application à l'estimation des descripteurs de Fourier	124
III.3. Segmentation orientée classification par optimisation génétique	124
III.3.1. Codage des individus	124
III.3.1.A. Description du contour	124
III.3.1.B. Initialisation du contour	124
III.3.1.C. Généalogie	125
III.3.1.D. Composition du chromosome	125
III.3.2. Population initiale	125
III.3.3. Critère d'optimisation pour la partition des données sonar	126
III.3.4. Description de l'algorithme génétique	127
III.3.4.A. Fonction de sélection	127
III.3.4.B. Opérateurs génétiques	128
III.3.4.C. Taux de croisements [PAL-96]	130
III.3.4.D. Fonction d'adaptation	131
III.3.4.E. Réinsertion	131
III.3.4.F. Critère d'arrêt	132
III.3.5. Module de classification	132
III.4. Expérimentations	133
III.4.1. Paramétrage	133
III.4.2. Performances	134
III.4.3. Illustrations	135
III.5. Conclusion	138
IV. CONCLUSION DU CHAPITRE III.2	138

CONCLUSION DE LA PARTIE III	139
PARTIE IV : Application multi-vues	141
INTRODUCTION DE LA PARTIE IV	143
CHAPITRE IV.1 : DE L'ANALYSE MONO-VUE À LA CLASSIFICATION MULTI-VUES	145
I. FUSION	145
I.1. Fusion de données en imagerie sonar	145
I.1.1. Modélisation	146
I.1.2. Choix du cadre théorique	147
I.2. Quelques fonctions décroissantes	148
I.3. Modèles flous	150
I.3.1. Par rapprochement à un élément idéal	151
I.3.1.A. Mesure vectorielle	151
I.3.1.B. Mesure multi-attributs	151
1) Distance indépendante de la distribution des données	152
2) Distance basée sur l'hypothèse d'une distribution gaussienne des données	152
3) Distance basée sur les histogrammes des attributs	153
4) Distance basée sur les histogrammes des attributs avec une prise en compte des recouvrements	155
I.3.2. Par rapprochement à l'ensemble flou	155
I.3.2.A. Distance basée sur un nombre limité de points représentatifs	155
I.3.2.B. Distance basée sur la densité des éléments au voisinage de l'élément courant	156
I.4. Opérateurs de fusion	158
I.5. Décision	160
I.6. Résultats expérimentaux	160
I.6.1. Quelques résultats obtenus par l'opérateur F_m	160
I.6.2. Quelques résultats obtenus par l'opérateur F_{mp}	164
I.7. Visualisation graphique	165
I.7.1. Opérateur F_m	165
I.7.2. Opérateur F_{mp}	166
I.8. Conclusion	167
II. CLASSIFICATION BASÉE SUR L'ÉVOLUTION DES ATTRIBUTS AU COURS D'UNE SÉQUENCE D'IMAGES	168
II.1. Conditions opérationnelles	168
II.2. Lien entre symétrie de l'objet et valeurs des attributs	170
II.3. Algorithme de classification multi-vues	170
II.4. Résultats expérimentaux	173

II.5.	Conclusion	175
CHAPITRE IV.2 : CARACTÉRISATION FLOUE DE LA SUPERPOSITION DES VUES		177
I.	NOTION DE RÉGION FLOUE [ROSE-84]	177
I.1.	Définition	177
I.2.	Attributs intrinsèques flous	178
II.	OBTENTION DE L'IMAGE MULTI-VUES	179
II.1.	A partir des images segmentées	180
II.2.	A partir des images brutes	181
II.3.	Fuzzification de l'image	183
III.	RÉSULTATS EXPÉRIMENTAUX	183
IV.	CONCLUSION	185
CONCLUSION DE LA PARTIE IV		187
Conclusion générale		189
Annexes		195
ANNEXE A : RAPPEL DES PRINCIPES THÉORIQUES DE QUELQUES TECHNIQUES CLASSIQUES DE SEGMENTATION		197
a)	<u>Segmentation markovienne</u>	197
i)	Principe théorique	197
ii)	Optimisation	199
b)	<u>les contours actifs statistiques [CHES-99]</u>	199
c)	<u>La Ligne de Partage des Eaux (LPE) [MEYE-90] [VACH-95]</u>	200
ANNEXE B : ANALYSE MULTIRÉSOLUTION PAR ONDELETES		203
ANNEXE C : EXTRACTION DU CONTOUR		205
1)	Algorithme de suivi de contour	205
2)	Equirépartition des points de contour	205

ANNEXE D : DÉCOMPOSITION EN SÉRIE DE FOURIER	207
1) Définition	207
2) Propriétés	207
3) Reconstruction	208
ANNEXE E : MOMENTS	209
1) Définitions [PROK-92]	209
a – Moments réguliers	209
b – Moments centrés	209
c – Moments centrés normalisés ou moments réduits	209
2) Transformations des moments	210
a – homothétie	210
b – translation	210
c – rotation	210
3) Propriétés des moments d'ordre inférieur ou égal à deux	211
a – Ordre zéro	211
b – Ordre un	211
c – Ordre deux	211
ANNEXE F : INVARIANTS DE MOMENTS	213
A) Invariants de moments centrés normalisés de Hu [HU-62]	213
B) Invariants de moments de Zernike exprimés en termes de moments centrés normalisés introduits par Teague [TEAG-80]	213
ANNEXE G : BASE D'IMAGES SIMULÉES	215
ANNEXE H : ALGORITHME FUZZY C-MEANS OU ISODATA FLOU [DUNN-74] [BEZD-81]	219
ANNEXE I : INTRODUCTION AU LOGICIEL NEUROCLASSE	221
1) NeuroClasse en pratique	221
2) Synoptique de la chaîne de traitement	222
ANNEXE J : CALCUL DE LA FONCTION D'ADAPTATION	223
Publications	225
Bibliographie	227

Liste des illustrations

Figures

Figure 1 - Cellule de résolution sonar	25
Figure 2 – Echo et ombres propre et portée	27
Figure 3 - Augmentation de l'ombre propre OPr au détriment de l'ombre portée OPo pour un objet de faible extension verticale	28
Figure 4 - Distorsions caractéristiques en imagerie sonar : l'élongation à gauche (mine Manta à 30m de distance pour une altitude du porteur de 15m) et le biais à droite (cylindre à 170m de distance pour une altitude du porteur de 50m)	28
Figure 5 - Exemples de mines : mines Manta, Sigeel et Rockan, sphère et cylindre	30
Figure 6 – Principe de la synthèse d'ouverture avec N récurrences	31
Figure 7 – Sonar de coque et PVDS	31
Figure 8 - Principe de classification sur ombre	32
Figure 9 – Secteur insonifié	33
Figure 10 - Images réelles obtenues par un traitement antenne synthétique: à gauche, mine Manta à 50m de distance et à droite mine sphérique à 76.5m de distance pour altitude du sonar de 10.9m	42
Figure 11 – Prise en compte de l'angle de rasance α	43
Figure 12 - Mélange des deux distributions de Rayleigh associées aux zones 'ombre' et 'fond'	44
Figure 13 – Synoptique des traitements de segmentation markovienne hiérarchique	49
Figure 14 – Etapes du traitement « contours actifs statistiques »	53
Figure 15 - Etapes du traitement « Ligne de Partage des Eaux »	55
Figure 16 - Lien entre allure du filtre spécifié et principe de formation de l'image sonar	59
Figure 17 - Principe de la méthode à pleine résolution	61
Figure 18 - Principe de chacune des méthodes proposées :	62
Figure 19 – Illustration du biais entre deux prises de vue du même objet	71
Figure 20 – Normalisation de l'image d'images binaires et brutes	73
Figure 21 – Normalisation en rasance	74
Figure 22 – Lissage de contour sur les images de la Figure 20	75
Figure 23 – Illustration de la paramétrisation proposée par Arbter dans le cas affine	81
Figure 24 – Exemples de reconstructions après une transformée de Fourier des coordonnées des points de contour	84
Figure 25 – Exemples de reconstructions par AW-invariants	85
Figure 26 – Evaluation de la distance entre deux pixels	86
Figure 27 - Lignes W_i et colonnes H_j caractérisant la zone d'ombre	87
Figure 28 – Evolution des deux rapports largeur sur hauteur en fonction de l'angle de rotation	88
Figure 29 – Exemples de configurations problématiques (multi-modalité à gauche et dispersion intra-classe à droite)	103
Figure 30 – Bonne partition dans le plan principal: bénéfice de l'hybridation	105
Figure 31 - Evolution de la similarité de part et d'autre de la bonne approximation (dixième image)	112
Figure 32 – Prototypes utilisés pour l'appariement (un prototype par classe)	112
Figure 33- Algorithme d'appariement de silhouettes	114
Figure 34 – Pour chaque individu (170 par classe), visualisation des mesures de similarité obtenues avec chacun des quatre prototypes	116
Figure 35 – <u>Appariements sur images simulées</u> : deux mines cylindriques (biais) et une mine Sigeel	117
Figure 36 - <u>Appariements sur images réelles</u> : mines Manta et sphérique	118
Figure 37 - Segmentation orientée classification	121
Figure 38 - Principe général des algorithmes génétiques	123
Figure 39 - Définition du chromosome	125
Figure 40 – Exemples d'individus (leur contour) de la population initiale	126
Figure 41 - Allures $P_i = f(r)$ pour différentes valeurs de q	127
Figure 42 – Echanges aléatoires entre deux chaînes chromosomiques	129
Figure 43 – Algorithme de classification du chromosome estimé	133

Figure 44 – Point de vue ambigu de la mine Rockan	135
Figure 45 – Résultats sur images réelles (mine Manta et mine sphérique)	136
Figure 46 – Résultats sur images simulées (2 mines cylindriques et une mine Sigeel)	136
Figure 47 – Résultats sur images simulées (une mine cylindrique et une mine Rockan) et le résultat de l'estimation des distributions des niveaux de gris des pixels	137
Figure 48 – Mesure d'une image courante I_n à une classe C_k	145
Figure 49 – Influence de la nature du fond sur la qualité des images	146
Figure 50 – Exemples de fonctions décroissantes utilisées pour le calcul de la fonction d'appartenance	149
Figure 51 – Exemples de distributions mono et multidimensionnelles des individus	150
Figure 52 – Distribution gaussienne	152
Figure 53 – Histogramme pour un attribut donné	153
Figure 54 – Fonction de répartition correspondante à l'histogramme de la Figure 53	154
Figure 55 – Individus à comptabiliser	154
Figure 56 – Différents cas pour le calcul de l'intervalle $\Delta_i^k(a_i)$	157
Figure 57 – Evolution du rapport largeur sur hauteur pour le cylindre le long d'une trajectoire semi-circulaire	159
Figure 58 – Evolution du rapport largeur sur hauteur pour la mine Rockan le long d'une trajectoire semi-circulaire	159
Figure 59 – Evolution des performances moyennes pour N de 3 à 10 vues: risque associé à la réduction d'un ensemble à un élément idéal	161
Figure 60 – Evolution des performances moyennes pour N de 3 à 10 vues: influence du seuillage	162
Figure 61 – Evolution des performances moyennes pour N de 3 à 10 vues: apport d'une analyse fine de la distribution des attributs	162
Figure 62 – Evolution des performances moyennes pour N de 3 à 10 vues: influence du nombre de vues	163
Figure 63 – Effet moyen de la pondération des points de vue	164
Figure 64 – Effet de la pondération des points de vue	164
Figure 65 – Mesures moyennes $\overline{M}_k$ pour les attributs hybrides	165
Figure 66 – Mesures moyennes $\overline{M}_k$ pour les descripteurs de Fourier	166
Figure 67 – Illustration du gain de F_{mp} sur F_m	167
Figure 68 – Lien entre propriétés géométriques et allure de l'évolution des attributs	169
Figure 69 – Courbes théoriques des valeurs d'attributs apprises par classe pour une séquence de prises de vue le long d'une trajectoire semi-circulaire	171
Figure 70 – Algorithme de classification multi-vues par caractérisation de l'évolution séquentielle de la valeur des attributs	172
Figure 71 – Taux moyen de bonne classification pour un nombre de vues N de 3 à 19	174
Figure 72 – Taux moyen de bonne classification par classe sur l'ensemble des séquences pour un nombre de vues N de 3 à 19	174
Figure 73 – Ecart-type associé au taux moyen de bonne classification sur l'ensemble des séquences pour un nombre de vues N de 3 à 19	175
Figure 74 – Exemples d'images I_s obtenues à partir de 19 prises de vues par pas angulaire de 10 degrés : cylindre, sphère, mines Manta, Sigeel et Rockan	181
Figure 75 – Exemples d'images I_b obtenues à partir de 19 prises de vues par pas angulaire de 10 degrés : cylindre, sphère, mines Manta, Sigeel et Rockan	182
Figure 76 – ACP (en haut) et cercle de corrélations associé (en bas) pour les images I_s (à gauche) et I_b (à droite)	184

Tableaux

Tableau 1 – Dimensions des mines simulées	30
Tableau 2 – Détermination des distances de corrélation suivant les deux axes de l'image	44
Tableau 3 – Comparaison des deux modèles structurels	54
Tableau 4 – Taux de bonne classification pour trois classifieurs <u>sans lissage de contour</u> : discrimination linéaire (ligne 1), nuées dynamiques (ligne 2) et k-ppv avec k=1 (ligne 3)	102
Tableau 5 – Taux de bonne classification pour trois classifieurs <u>avec lissage de contour</u> : discrimination linéaire (ligne 1), nuées dynamiques (ligne 2) et k-ppv avec k=1 (ligne 3)	103
Tableau 6 – Matrice de confusion par k-ppv avant ACP, i.e. un taux moyen de 99.0%	105
Tableau 7 – Matrice de confusion par k-ppv après ACP (2 composantes principales)	106
Tableau 8 – Comparaison espaces caractéristiques 2D et 4D	106
Tableau 9 – Matrice de confusion (taux de bonne classification moyen = 97.5%)	115
Tableau 10 – Synthèse des caractéristiques des trois approches énoncées	122
Tableau 11 – Matrice de confusion (taux moyen de bonne classification de 98.2%)	135
Tableau 12 – Informations à fusionner	147
Tableau 13 – Performances pour les attributs hybrides et N=3 vues	161
Tableau 14 – Performances pour les attributs hybrides et N=3 vues	163
Tableau 15 – Matrices de confusion pour M=N=19, i.e. un demi-cercle autour de la mine (resp. N=9, i.e. moins d'un quart de cercle autour de la mine)	174
Tableau 16 – Taux de bonne classification pour trois classifieurs : discrimination linéaire (ligne 1), nuées dynamiques (ligne 2) et k-ppv avec k=1 (ligne 3)	184
Tableau 17 – Taux de bonne classification pour trois classifieurs à partir des deux composantes principales : discrimination linéaire(ligne 1), nuées dynamiques (ligne 2) et k-ppv avec k=1 (ligne 3)	185

Introduction générale

Les progrès technologiques de ces dix dernières années permettent de faire de l'imagerie des fonds marins avec une très haute résolution. La guerre des mines a tout naturellement exploité les performances des sonars classificateurs haute fréquence pour saisir les petits détails et finement caractériser les contours des ombres portées des objets immergés.

Généralement, la destruction d'une mine se déroule en trois étapes successives. Caractérisé par une portée relativement grande et une résolution faible, le sonar détecteur permet de repérer l'objet par son écho. Grâce à une meilleure résolution et à une plus courte portée, le sonar classificateur fournit une image interprétable sur laquelle l'ombre portée de l'objet se détache. Si cet objet représente une menace potentielle, la neutralisation est effectuée à l'aide d'un véhicule sous-marin ou d'un plongeur démineur en déposant une charge explosive qui sera déclenchée à distance. Les travaux présentés s'articulent autour de la deuxième étape qui comporte deux points : la segmentation de l'ombre et la classification automatique de l'objet potentiellement dangereux. Les choix effectués au cours d'une première approche mono-vue du problème ont été conditionnés par la finalité de l'étude c'est-à-dire l'automatisation du processus de classification multi-vues qui devrait se généraliser du fait des études croissantes menées sur les véhicules sous-marins autonomes en chasse aux mines. Grâce aux progrès de la robotique, de nouvelles missions pourront ainsi leur être confiées dans les prochaines années. Parmi celles-ci, on peut citer les missions consacrées à l'exploration pour la chasse aux mines et la surveillance des fonds.

L'ombre acoustique est caractérisée par l'absence de réverbération de fond du fait de la présence de l'objet. La segmentation vise à extraire cette zone d'intérêt de l'image sonar pour la classification. Cette première étape est importante car elle conditionne la qualité de l'interprétation qui va suivre. Certains traitements classiques de segmentation donnent de bons résultats en imagerie sonar en autorisant l'adaptation de certains paramètres au contexte d'acquisition de l'image. Le risque de fausses alarmes est ainsi limité tout en garantissant la même précision le long des deux axes de l'image. Une approche originale par filtrage a par ailleurs été suggérée par la prise en compte de la corrélation des pixels inhérente au principe même de formation des images sonar.

La classification automatique de l'objet au sein de classes préétablies nécessite :

- de définir les classes d'objets,
- de caractériser ces objets,
- de définir un algorithme de classification automatique.

Les cinq types de mines retenus répondent à des critères de fréquence d'apparition, de variété de tailles, de furtivité et de complexité géométrique. Tout en étant homogènes, les classes doivent couvrir le plus grand nombre de cas possibles en termes de tailles et formes de l'ombre.

Pour pouvoir discriminer les objets ciblés, les ombres peuvent être caractérisées de deux manières :

- par codage réversible du contour

Une décomposition en série de Fourier du contour suivant lignes et colonnes est cohérente avec le principe d'acquisition des données sonar. Son intéressante propriété de linéarité peut en outre être diversement exploitée pour manipuler le contour.

- par extraction d'attributs caractérisant des propriétés géométriques ou statistiques propres de la forme.

Compte tenu des distorsions classiquement observées en imagerie sonar, l'ombre projetée de l'objet subit des transformations géométriques qu'il convient de considérer avant ou pendant l'extraction d'attributs caractéristiques. La normalisation en ranson de l'image permet de maintenir un taux de classification satisfaisant quelles que soient les conditions d'acquisition. Néanmoins, l'introduction d'invariants en reconnaissance de formes permet de s'affranchir de ce prétraitement s'ils sont construits de manière à ce que l'élongation n'ait aucun effet sur leurs mesures. En ce sens, l'invariance par rapport à toute transformation affine du plan est apparue particulièrement intéressante dans le contexte sonar.

Les attributs extraits sont ensuite introduits dans un processus de classification. La première solution consiste en une classification dite supervisée basée sur un mode d'apprentissage. Elle nécessite donc de disposer d'un grand nombre d'éléments représentatifs de chacune des classes. Au cours de cette phase, des relations sont établies entre les observations et les différents groupes ou classes auxquels elles appartiennent. Les classifieurs classiques exploitent ensuite ces données apprises en les rapprochant de l'élément inconnu.

Une autre solution consiste à réduire chacune des classes à un élément prototype qui à une transformation géométrique près est susceptible de coller à n'importe quel objet de même classe. Par ailleurs, une méthode novatrice orientée vers le résultat escompté c'est-à-dire la classification de la mine ciblée s'oppose à la procédure séquentielle classique « segmentation → extraction d'attributs → classification ». Les algorithmes génétiques offrent la possibilité de faire cohabiter un certain nombre de solutions possibles et de converger au fil des itérations ou générations vers la solution optimale. La solution recherchée est ici le contour qui partitionne au mieux les pixels de l'image. Sous un critère statistique basé sur la distribution des niveaux de gris des pixels de l'image sonar, l'algorithme génétique que nous proposons estime les paramètres descriptifs du contour. Celui-ci segmente l'image tout en permettant de le discriminer en finalité.

Les approches classiques mono-image ont montré leurs limites : environnement sous-marin source d'incertitudes, difficulté à traiter des mines plus complexes telles que la mine Rockan suédoise... Pour pallier l'insuffisance de ces techniques, l'acquisition de plusieurs images du même objet selon des points de vue différents est une solution envisagée. Plusieurs stratégies sont possibles suivant le nombre de vues et leurs conditions d'acquisition. Par exemple, si les données acquises sont incertaines (cas de contact avec un objet naturel ou de confusion avec le fond) et imprécises (points de vue indéterminés), la fusion des informations acquises suivant les différents points de vue est appropriée. Néanmoins, une bonne maîtrise de la navigation d'un véhicule à immersion variable permet d'acquérir les points de vue selon une trajectoire bien définie. Le problème dispose alors d'une dimension supplémentaire, la dimension séquentielle qui permet d'appréhender le problème de manière plus globale. Enfin, une solution intermédiaire consiste à former une seule image résultat de la superposition de

l'ensemble des images obtenues au cours de la séquence. Formée à partir des données brutes ou après binarisation de l'image, cette image présente une ombre dont les contours restent flous. Par extension du binaire à plusieurs niveaux de gris, des attributs géométriques flous sont calculés sur ces pixels d'ombre de manière à caractériser la dispersion des formes de l'ombre portée au cours de la séquence.

Un effort particulier a été apporté sur l'adaptation des algorithmes de traitements d'images et de reconnaissance de formes à la spécificité des images sonar tant au niveau des propriétés statistiques qu'au niveau des distorsions géométriques inhérentes aux conditions d'acquisition. Dans le passé en effet, ces aspects ont insuffisamment été exploités malgré leur importance et leur impact sur les traitements appliqués (robustesse, temps d'exécution...). De ce fait, l'étude d'outils de traitement d'images sonar a fait l'objet d'une part importante des travaux de thèse. Après les avoir testés en mono-vue, certains outils se sont avérés intéressants pour une application multi-vues. Pour cette étape de classification à partir de plusieurs points de vue, nous proposons plusieurs solutions fonction du contexte opérationnel posé (navigation du porteur). Nous insisterons en particulier sur l'importance du choix d'attributs pertinents dont les propriétés sont différemment exploitables. Le plan de la thèse est donc le suivant :

La première partie est une introduction à l'imagerie sonar.

La seconde partie discute la pertinence et l'adaptation d'outils de bas et haut niveaux pour l'analyse d'images sonar. La variété des attributs de formes introduits répond à différentes voies possibles pour la classification mono-vue dans un premier temps mais prévoit surtout son extension à plusieurs vues.

La troisième partie aborde la classification mono-vue en opposant l'approche séquentielle classique à deux approches alternatives offrant la possibilité de ne classifier qu'à partir d'un nombre initial limité de solutions potentielles.

La quatrième partie vise à exploiter un ensemble d'images présentant un objet vu selon des angles de vue différents pour effectuer une classification multi-vues. Trois techniques sont proposées : par fusion, par analyse de l'évolution des attributs au cours d'une séquence d'images et par caractérisation floue d'une superposition d'images.

PARTIE I :

Imagerie sonar :

particularités et exploitation

Après un bref rappel de la notion de **sonar**, cette première partie explicite le cadre opérationnel dans lequel on se situe et les contraintes associées en termes techniques. Dans notre étude, la **chasse aux mines** est appréhendée par le biais de la **classification sur ombre** à partir d'images sonar **haute résolution**. La spécificité du support sur lequel on travaille, à savoir **l'image sonar**, est mise en évidence par une description du principe d'acquisition de l'image et des caractéristiques qui en découlent. Ces différentes considérations nous ont permis de simuler des images répondant à un certain nombre de critères pour la validation des approches proposées dans les parties suivantes. Enfin, nous proposons une méthodologie en traitement d'images sonar.

I. Notion de sonar [MARA-90]

Les systèmes sonars utilisent les ondes acoustiques provoquées par la propagation d'une vibration mécanique pour transmettre et recevoir de l'information dans le milieu marin. Le terme SONAR pour *SOund Navigation and Ranging* désigne l'utilisation des ondes acoustiques pour la navigation et la transmission d'information (émission et réception). Il concerne donc l'ensemble des systèmes de connaissance du milieu (sondeurs bathymétriques, sondeurs multifaisceaux,...) et de localisation et/ou détection de cible (sonars actifs et passifs militaires, sonars de pêche,...).

Les applications militaires de l'acoustique sous-marine sont diverses et mettent en jeu des traitements de plus en plus complexes de manière à satisfaire de nouveaux besoins en terme de sécurité. Les progrès technologiques associés à des capacités de stockage et de traitement de plus en plus importantes ont favorisé le développement de techniques innovantes à tous les niveaux de la chaîne de traitement sonar (traitement antenne synthétique pour la synthèse d'image haute résolution, traitement multi-vues pour la classification...).

L'une des principales applications militaires concerne la détection, la localisation et la reconnaissance de cibles. Le sonar passif d'une part assure une surveillance passive en se bornant à « écouter ». Les sonars actifs d'autre part émettent un signal et la détection est souvent basée sur la réception de l'écho réfléchi sur la cible. Dans les deux cas, la réception s'effectue sur une antenne et les signaux captés passent dans la chaîne de traitement sonar pour former l'image sonar.

Dans le cas particulier de la **classification sur ombre**, le système sonar équipant les navires de chasse aux mines est composé d'un sonar détecteur et d'un sonar classificateur. Dans un premier temps, la recherche d'objets sur le fond ou dans le volume d'eau est effectuée par le sonar détecteur. Dans un second temps, le sonar classificateur fournit une image suffisamment résolue permettant de classer l'objet à partir de l'ombre acoustique portée sur le fond. Le développement des sonars classificateurs de coque débuté dans les années 70 se poursuit dans les années 80 par une nouvelle génération de sonars à immersion variable.

II. Principe de formation des images sonar

Les sonars d'imagerie sont apparus dans les années 50. L'exploration et la caractérisation des fonds marins ont fortement bénéficié de ces nouveaux équipements capables de fournir les détails naturels (nature et texture du fond...) et artificiels (objets manufacturés, mines marines...).

II.1. Architecture du sonar

Le système sonar se situe à l'interface entre monde acoustique et monde électrique. Les transducteurs électroacoustiques sous-marins effectuent la transformation d'énergie électrique en énergie acoustique et vice-versa en utilisant classiquement les propriétés piézoélectriques ou magnétostrictives de certains cristaux naturels ou céramiques. En émission, le transducteur se nomme projecteur ou base acoustique alors qu'en réception, il se nomme hydrophone dans le cas d'un récepteur ponctuel ou antenne pour un récepteur étendu c'est-à-dire constitué de

transducteurs élémentaires ou capteurs. En réception, la conversion des données acoustiques (pression de l'onde acoustique incidente) en données électriques (tension en sortie) est effectuée par un ou plusieurs capteurs ou hydrophones. A l'émission, les transducteurs émettent des impulsions dans une direction oblique vers le fond de la mer. De par leur conception, les antennes favorisent certaines directions d'arrivée du signal et par extension, localisent précisément la cible.

A fréquence donnée, la fonction de directivité $D(\theta, \varphi)$ d'un transducteur traduit la réponse énergétique en champ lointain en fonction de l'angle, normalisée par rapport à son maximum. Elle comporte en général un lobe principal que l'on caractérise par son ouverture à -3dB , dit lobe à $2\theta_3$, qui est la largeur du faisceau correspondant à $D(\theta, \varphi)=0.5$, i.e. une chute d'intensité de 3dB par rapport à l'intensité maximale émise ou reçue dans l'axe. Cette donnée traduit la résolution angulaire associée à l'antenne et est approximativement égale au rapport de la longueur d'onde λ sur la longueur de l'antenne L . De manière générale, une antenne est d'autant plus directive (lobe principal étroit) que le rapport dimension de l'antenne sur longueur d'onde est élevé. La diminution de l'ouverture (soit une meilleure résolution) peut être obtenue par une augmentation de la fréquence d'émission et/ou une antenne plus grande. Outre la présence indésirable de lobes secondaires, on est confronté dans le premier cas à l'accentuation de l'atténuation de l'onde acoustique, et dans le second cas, à une limitation en termes de coût et de fonctionnalité de la taille de l'antenne. Le traitement antenne synthétique pallie ce problème en construisant une antenne virtuellement plus longue à partir des déplacements successifs de l'antenne effective [BRUC-92] [BILL-98].

II.2. Caractérisation des signaux émis pour le sonar actif

Aux caractéristiques physiques de l'antenne d'émission s'ajoute le choix du signal ou pulse d'émission caractérisé par sa durée d'impulsion T et la bande passante B . Deux types de signaux sont couramment utilisés : l'impulsion monochromatique à fréquence pure (par convention, $BT=1$) et l'impulsion modulée linéairement en fréquence ($BT>1$). Pour une même durée d'impulsion T mais en utilisant des signaux à bande large B , une émission à fréquence modulée augmente le pouvoir de résolution en distance d'un facteur BT par rapport à une émission à fréquence pure. Ce phénomène est le résultat de la compression d'impulsion (de durée T au départ) en sortie du traitement cohérent (corrélation du signal reçu avec une copie du signal émis) [MARA-91].

II.3. Lien entre cellule de résolution du sonar et pixel de l'image

La **formation de l'image sonar** consiste à enregistrer et représenter les données sonar (ondes acoustiques réfléchies) dans le plan oblique. La dimension effective du pixel dépend à la fois des paramètres sonar et du traitement des données.

D'une part, les **paramètres sonar**, caractéristiques de l'antenne et du signal émis, fixent la dimension de la cellule de résolution, c'est-à-dire l'écart minimal suivant l'axe de propagation de l'onde (distance) et angulairement (gisement) devant séparer deux points pour qu'on puisse les distinguer [SOME-93] [BLON-97]. En utilisant les notations des paragraphes précédents, les résolutions en distance δR et en gisement $\delta \theta$ se définissent comme suit :

$$\delta R = c/2B$$

$$\delta\theta = 2\theta_3 = \lambda/L_a = c/f_0 L_a$$

où

c désigne la célérité du son dans l'eau

B est la bande passante du signal,

λ est la longueur d'onde pour la fréquence centrale du sonar f_0

L_a est la longueur de l'antenne de réception

$2\theta_3$ désigne l'ouverture de l'antenne

Dans le plan oblique à la distance R du sonar, on a donc une cellule de résolution de dimension $\delta R \times (R\delta\theta)$. La Figure 1 explicite ce principe de formation de l'image sonar [BRUC-92].

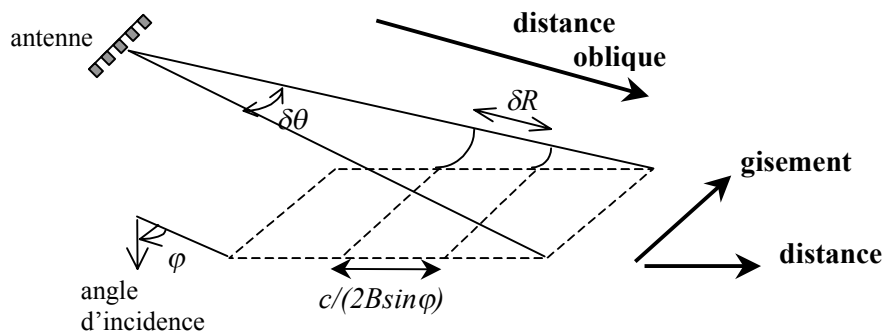


Figure 1 - Cellule de résolution sonar

Le **traitement des données** permet de passer des données sonar brutes à l'image en niveaux de gris illustrant la variation de la pression acoustique reçue par les capteurs de l'antenne et proportionnelle à la puissance du signal rétrodiffusé (c'est-à-dire en direction du sonar). Le niveau de gris d'un pixel de l'image représente ainsi l'amplitude de la somme des ondes réfléchies vers le sonar par les points du fond (diffuseurs) appartenant à la cellule de résolution. La juxtaposition des enregistrements obtenus dans chacune des voies pointées permet de former une carte bidimensionnelle de la zone insonifiée. Ce traitement est effectué à deux niveaux:

- au niveau des capteurs de l'antenne de réception, on effectue la conversion des données acoustiques en données électriques par préamplification, démodulation complexe, filtrage, conversion numérique et normalisation des signaux.
- au niveau des machines de traitement de signal se déroulent des algorithmes de compression d'impulsion, de formation de voies, de focalisation d'antenne ...

En général, la dimension du pixel est plus petite que celle de la cellule de résolution (dans le cas contraire, on perd de l'information). Elle dépend en effet de plusieurs paramètres tels que la fréquence d'échantillonnage temporel, l'échantillonnage spatial effectué lors de l'opération de formation de voies, la pondération de l'antenne,... Ces opérations induisent un suréchantillonnage plus ou moins important de la zone insonifiée qui n'apporte pas d'information supplémentaire mais améliore la visualisation. Diverses dégradations peuvent également intervenir tels que la calibration de l'antenne, le niveau des lobes secondaires...

III. Spécificité de l'imagerie sonar haute résolution

Qu'il s'agisse de traitements de bas-niveau comme le problème de segmentation ou de traitements de haut-niveau tels que la caractérisation des formes segmentées et leur classification, la prise en compte de phénomènes inhérents à l'imagerie sonar permet d'orienter les traitements de manière à les rendre plus robustes.

III.1. Bruit en imagerie sonar [BOUV-92] [MART-94]

Les bruits en jeu se partagent en deux groupes.

Le premier désigne les bruits perçus le long du canal marin soient :

- le bruit ambiant est ce qui est perçu par le système en l'absence de signal utile et du bruit propre du porteur : trafic maritime, agitation de la surface, pluie, agitation thermique
- la réverbération (propre au sonar actif) est générée par la somme des multiples contributions des diffusions de l'onde par les inhomogénéités (de l'eau : particules, bulles) ou irrégularités (relief du fond, aspect non-uniforme de la surface) du milieu : réverbération de volume, de fond et de surface
- les trajets multiples (dus au fond et à la surface)

Ces autres bruits constituent le second groupe :

- le bruit propre du porteur est généré par son déplacement, les machines,...
- le bruit électrique généré par l'électronique de réception du sonar

Le bruit ambiant et le bruit propre du porteur limitent les performances du sonar. Le bruit propre diminue en outre la discrétion du navire. Cependant, les sonars équipant des porteurs filoguidés ou autonomes sont concernés en moindre mesure par le bruit propre mais peuvent subir les bruits du navire dont ils dépendent.

En ce qui concerne l'imagerie haute résolution basée sur l'utilisation de **hautes fréquences**, on subit spécifiquement l'agitation de l'interface air-eau (dépendant des conditions météorologiques), le bruit électrique, le bruit thermique d'agitation moléculaire et les trajets multiples. La réverbération est par ailleurs source d'information en classification sur ombre (cf. §IV.3). D'une manière générale, l'onde réverbérée par le milieu (eau de mer, fond ou surface) est le résultat de la combinaison tantôt destructive (donnant une amplitude faible par rapport à l'amplitude moyenne) tantôt constructive (donnant une amplitude forte par rapport à l'amplitude moyenne) de toutes les ondes cohérentes mais déphasées aléatoirement rétrodiffusées par les diffuseurs (aspérités de dimension inférieure à la longueur d'onde de l'impulsion émise). D'après le théorème central limite qui établit que la somme d'un grand nombre de variables aléatoires indépendantes suit une distribution approximativement normale, la somme des ondes complexes reçues par un capteur s'exprime sous la forme d'un couple (somme des parties réelles et imaginaires des ondes) de variables gaussiennes indépendantes [ROSS-87]. L'amplitude et la phase résultantes suivent alors respectivement une loi de Rayleigh et une loi uniforme sur $[-\pi ; \pi]$ [PAPO-84]. En ce qui concerne la loi de Rayleigh, l'écart-type de la variable est proportionnel à l'espérance : ceci justifie le nom de bruit multiplicatif [MAIT-01]. Ce phénomène est commun à tous les

systèmes d'imagerie cohérente et apparaît lorsque la longueur d'onde incidente est grande devant la dimension des aspérités de la surface illuminée [GOOD-76].

L'aspect granulaire de la zone de fond est en particulier représentatif du bruit de speckle (terme donné à la réverbération) caractérisant l'amplitude de l'onde rétrodiffusée formée par la contribution d'un grand nombre de diffuseurs présents dans la cellule de résolution. Pendant l'insonification, les diffuseurs entraînent des fluctuations importantes du signal reçu autour de la valeur moyenne fonction de la rugosité du fond ou taille du grain des sédiments (vase, sable,...) et qui caractérise son pouvoir réflecteur. Un speckle de fort niveau constitue ainsi le signal utile pour la zone du fond alors que la zone d'ombre caractérisée par une absence réverbération de fond est soumise en contrepartie au bruit de réverbération de volume. Un phénomène dégradant intervient en outre au niveau de la frontière de la zone d'ombre. Une remontée du niveau de bruit dû à la présence de lobes secondaires entraîne une localisation imprécise et mauvaise définition des contours. Ces lobes secondaires sont associés à la fonction de directivité de l'antenne et pointent d'autres cellules de la zone de réverbération de fond (cf. §II.1).

Un certain nombre d'autres dégradations peuvent survenir lors de l'acquisition : mauvaise calibration de l'antenne, mouvements involontaires de l'antenne d'où une mauvaise estimation de la position des capteurs, électronique de la chaîne de traitement des données (quantification)...

III.2. Visualisation d'une image sonar « haute résolution »

En classification sur ombre, nous devons poser le problème de la délimitation exacte de la **région ombre** et de ses **distorsions**.

La dénomination 'ombre' désignant la zone sombre apparaissant derrière l'écho sous-entend l'**ombre portée** et l'**ombre propre** comme l'illustre la Figure 3. Elle correspond à une absence de signal sur la distance oblique, c'est-à-dire une occultation du fond par la mine. Suivant les cas, l'**écho** apparaît plus ou moins en avant de l'ombre puisqu'il est dû au contact entre le signal émis et une partie de l'objet (cf. Figure 2).

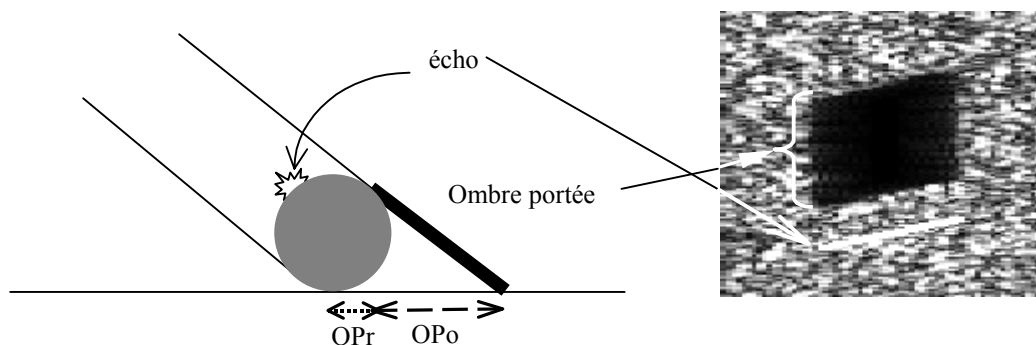


Figure 2 – Echo et ombres propre et portée

L'importance de l'ombre propre dépend à la fois de la rasant (qui tend à augmenter la proportion de l'ombre portée) et de la hauteur des mines [KERV-97b]. Dans le cas de mines de hauteur faible, on constate une augmentation des ombres propres (OPr) au détriment de l'ombre portée (OPo) (cf. Figure 3).

L'allure caractéristique des **ombres généralement très allongées** est due à la prise de vue parfois très rasante. En outre, ce phénomène est accentué par la différence de précision (résolution) en gisement et en distance. La solution consiste alors à normaliser l'image de manière à présenter une ombre représentative de l'objet.

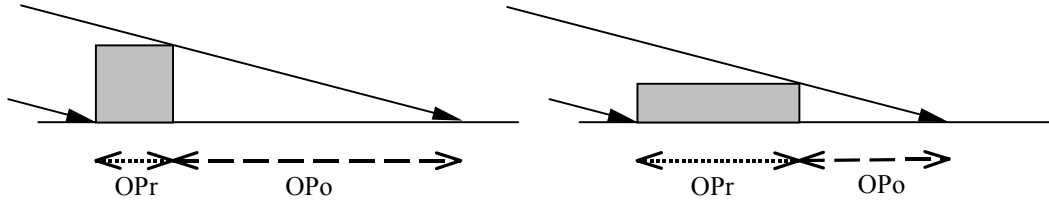


Figure 3 - Augmentation de l'ombre propre OPr au détriment de l'ombre portée OPo pour un objet de faible extension verticale

Le point de vue particulier du sonar peut ne pas être le plus caractéristique (pour le cylindre par exemple, cas où la ligne de visée est parallèle à la génératrice). Mis à part le cas des objets à symétrie radiale, l'ombre portée a des apparences variées suivant la position de l'objet vis-à-vis du sonar. Les distorsions caractéristiques de la prise de vue sont l'élongation augmentant avec la rasance et le biais dépendant de l'angle de vue (cf. Figure 4).

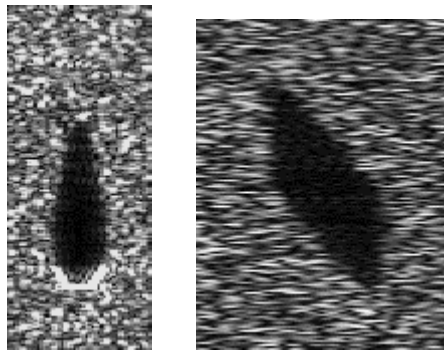


Figure 4 - Distorsions caractéristiques en imagerie sonar : l'élongation à gauche (mine Manta à 30m de distance pour une altitude du porteur de 15m) et le biais à droite (cylindre à 170m de distance pour une altitude du porteur de 50m)

IV. Guerre des mines : menace et solutions

La faiblesse des moyens militaires face aux mines marines de nouvelle génération apparut au début des années 90 lors des événements intervenus dans le golfe Persique : une modernisation des équipements vise depuis à faire face à ces mines dites furtives (petite taille, géométrie et structure étudiées présentant un index de cible réduit) [PUYH-00]. Ces dernières avancées justifient la recherche de procédés innovants pour la classification sur ombre.

IV.1. Menace

Les principales mines sont classées en deux catégories : les mines de fond et les mines à orin ([JANE], [KERV-97a]). Leur forme et leur charge explosive sont variables et dépendent de la hauteur d'eau. Une explosion de mine produit une onde de choc de nature à détruire un bâtiment de surface ou sous-marin.

Trois ensembles de mines nous intéressent :

- les mines de fond classiques de forte charge, utilisables pour des hauteurs d'eau moyennes (de quelques dizaines de mètres à 60m environ),
- les mines de fond de charge plus faible en général destinées à des hauteurs d'eau plus faibles,
- les mines à orin classiques sphériques ou oblongues (cylindre vertical) de pleine eau (pour des hauteurs d'eau allant jusque 80m).

Les premières sont généralement cylindriques, d'élongation horizontale. Les secondes d'apparition plus récentes ont une profondeur de mouillage inférieur à 20m. Ce sont des mines de formes très variées, furtives telles la mine Manta italienne. Ces mines d'influence ont une charge telle que le 'rayon de danger'¹ s'étend de 50 (pour les charges les plus faibles) à 170m environ. Enfin, les mines à orin de deux types, mines d'influence ou de contact, ont un rayon de danger qui varie de 0 (mine de contact) à 80m. Pour faciliter la simulation, les mines à orin ont été remplacées par d'hypothétiques mines de fond sphériques. Les ombres traitées seront donc plus courtes et de formes variées éloignées de la forme cylindrique à grand axe horizontal classique.

Pour chaque type de mines, on peut donc définir trois zones d'inspection pour un sonar classificateur de coque HMS (Hull-Mounted Sonar) : une zone interdite de risque maximal, une zone dangereuse à risque moyen (dépendant de la charge de la mine, du type de navire...) et une zone de relative sécurité où le sonar peut opérer avec un risque minimal. Quant au sonar de l'avant monté sur véhicule sous-marin filoguidé PVDS (Propulsed Variable Depth Sonar), sa conception originale lui permet de s'approcher très près de l'objet à classifier sans mettre en danger le navire qui le dirige (cf. §IV.2). Son utilisation se justifie notamment dans le contexte actuel de nouvelles menaces dues à l'apparition de mines de fond furtives et de taille réduite.

Les points énoncés ci-dessus nous ont amenés à élaborer la base de données avec les cibles suivantes dont les dimensions sont spécifiées dans le Tableau 1 :

- cibles de type mine classique (sphères et cylindres),
- cibles tronconiques de type mine furtive Manta et Sigeel,
- la mine furtive Rockan de forme aplatie.

Les charges explosives respectives [JANE] de ces mines donnent les rayons de danger suivants :

- la mine Manta porte 130kg de charge, d'où $R_d=54m$,

¹ Le **rayon de danger** [KERV-97a] synthétise la dangerosité de la mine. Sa valeur dépend du navire et de l'atteinte considérés. Dans notre cas, on se place dans les conditions suivantes :

- navire de type bâtiment de surface, atteinte du type "arrêt de mission" (facteur de choc $k = 0.3$),
- fond de type sable (coefficient de bourrage $C_f = 1.4$),
- explosif de type moderne (coefficient unitaire pratique $C_{up} = 1.4$), ce qui donne

$$R_d = \sqrt{(charge \text{ explosive en Kg}) \cdot C_f \cdot C_{up} / k} \approx 4.67 \sqrt{(charge \text{ explosive en Kg})}$$

- la mine Sigeel porte 400kg de charge, d'où $R_d=94m$,
- la mine Rockan porte 105kg de charge, d'où $R_d=48m$,
- les mines sphériques à orin d'influence portent une charge moyenne qui s'étend globalement de 80 à 300kg, d'où un rayon de danger R_d de 42 à 81m,
- les mines cylindriques de fond portent une charge moyenne qui s'étend globalement de 54 à 1350kg, d'où un rayon de danger R_d de 35 à 172m.

Ces dernières mines (sphériques et cylindriques) peuvent être classifiées par un sonar de coque tandis que les trois premières, du fait de leurs faibles rayons de danger et de leurs petites tailles, seront plus aisément classifiées par PVDS de par sa mobilité et une résolution plus fine.

Tableau 1 – Dimensions des mines simulées

D=diamètre, L=longueur, l=largeur, H=hauteur, PD=petit diamètre, GD=grand diamètre

<i>Forme</i>	<i>Dimensions en mm</i>
Cylindre	(D450, L1530) (D600, L2000) (D500, L2700)
mine Manta	(H370, PD475, GD980)
mine Sigeel	(H850, PD700, GD980)
Sphères	(D700) (D900) (D1100)
mine Rockan	(L1015, l800, H385)

Quelques illustrations sont données Figure 5.

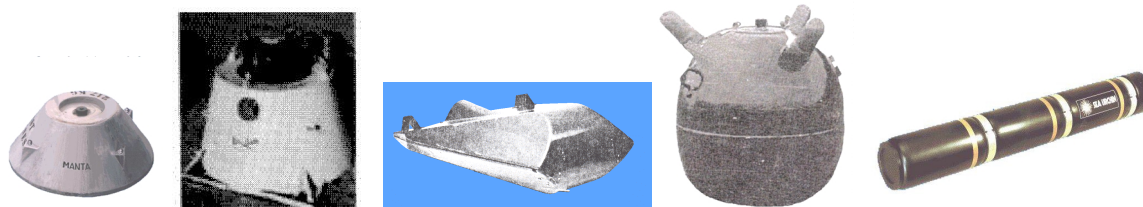


Figure 5 - Exemples de mines : mines Manta, Sigeel et Rockan, sphère et cylindre

IV.2. Progrès technologiques

L'utilisation de hautes fréquences (de l'ordre de 400kHz) vise à fournir suffisamment de détails alors que les mines actuelles sont de plus en plus petites. En contrepartie, on ne peut travailler à longues distances puisque l'amortissement du son (perte d'énergie de l'onde sonore dans l'eau) augmente en fonction de la fréquence.

Les techniques à synthèse d'ouverture permettent de réaliser le compromis entre la recherche d'une résolution suffisante pour affronter une menace furtive et l'encombrement de l'antenne. Le principe de « l'antenne synthétique » consiste à créer l'équivalent d'une grande antenne en exploitant le déplacement de l'antenne réelle de manière latérale (cf. Figure 6). Les signaux reçus lors des récurrences successives sont traités comme s'ils étaient issus d'une

même antenne (l'antenne synthétique) dont la longueur est déterminée par le déplacement de l'antenne réelle pendant la durée totale des récurrences [BRUC-92]. La résolution est par conséquent bien plus fine que celle de l'antenne réelle.

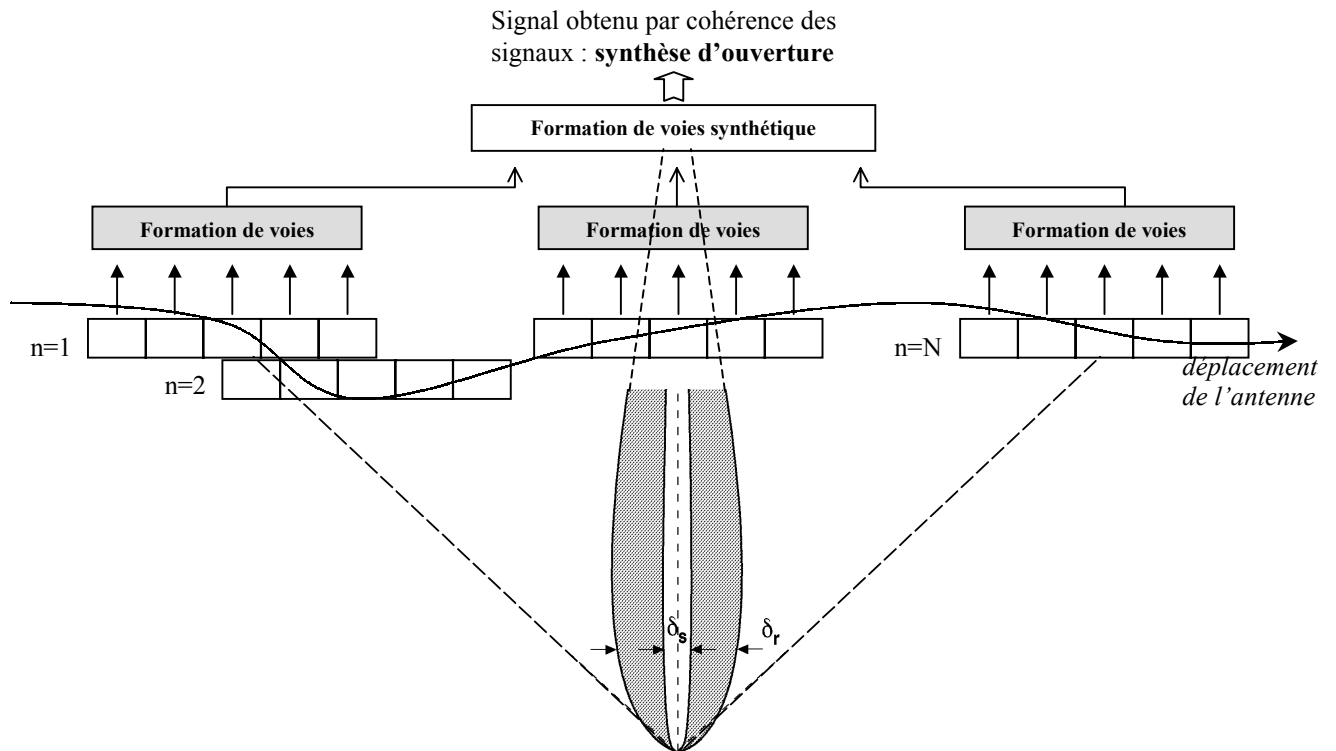


Figure 6 – Principe de la synthèse d'ouverture avec N récurrences

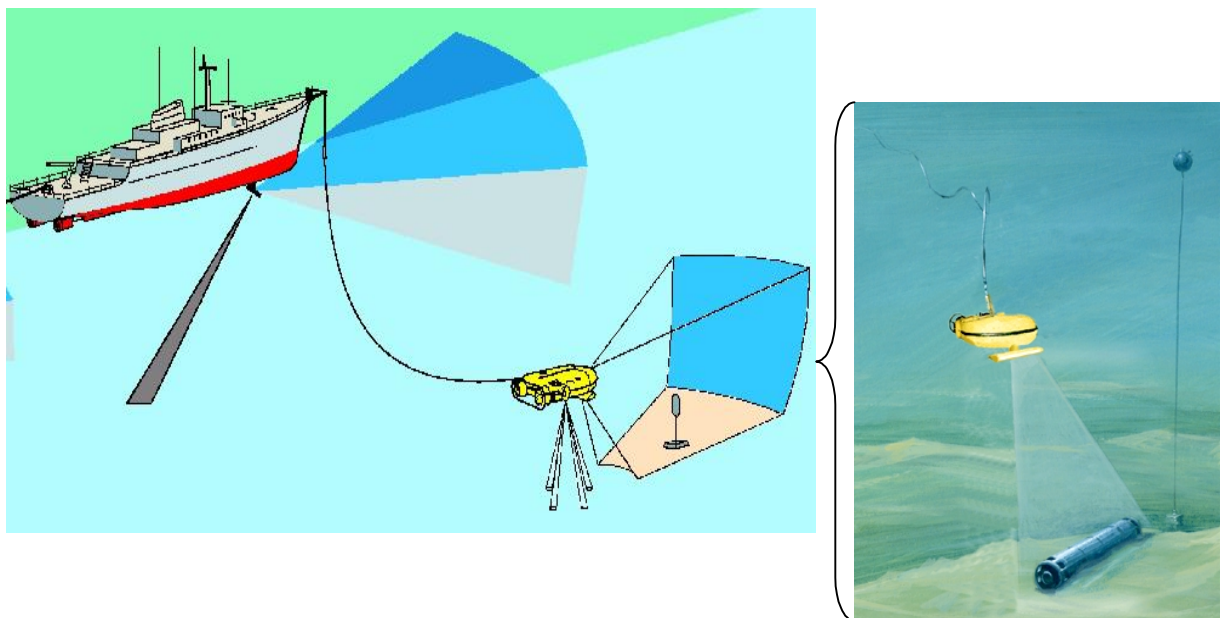


Figure 7 – Sonar de coque et PVDS

Alors que l'utilisation exclusive de sonars de coque ou latéraux nécessitait un compromis entre haute résolution et distance de sécurité (liée à la portée du sonar), Thales Underwater Systems a été le premier à développer le nouveau concept, particulièrement prometteur, du

sonar de l'avant: le PVDS (Propelled Variable Depth Sonar). Ce concept est la réponse adaptée à la nouvelle menace que font peser les mines futures sur les navires en permettant la récupération d'information précise sans que le navire soit menacé. Ce sonar monté sur véhicule est autopropulsé et peut aisément tourner autour de l'objet détecté et en fournir plusieurs vues sous des angles différents tout en maintenant le bâtiment à distance de sécurité. L'usage de ce véhicule offre de nouvelles possibilités d'exploitation des données sonar acquises par plusieurs vues et permet d'envisager une **classification multi-vues** de l'objet. La Figure 7 met en scène l'utilisation conjointe d'un sonar de coque et d'un PVDS depuis le chasseur de mines.

IV.3. Le principe de classification sur ombre

La caractérisation d'objets posés sur le fond marin est faite au moyen de **sonars classificateurs à haute résolution**, c'est-à-dire dont les dimensions de la cellule de résolution en distance et en gisement sont inférieures aux dimensions des objets recherchés. Sur l'image obtenue apparaît **l'ombre acoustique** que porte l'objet sur le fond, laquelle est souvent plus révélatrice de sa géométrie que son écho du fait de la complexité des lois de la diffusion acoustique qui régissent ce dernier et du revêtement anéchoïque de certaines mines furtives.

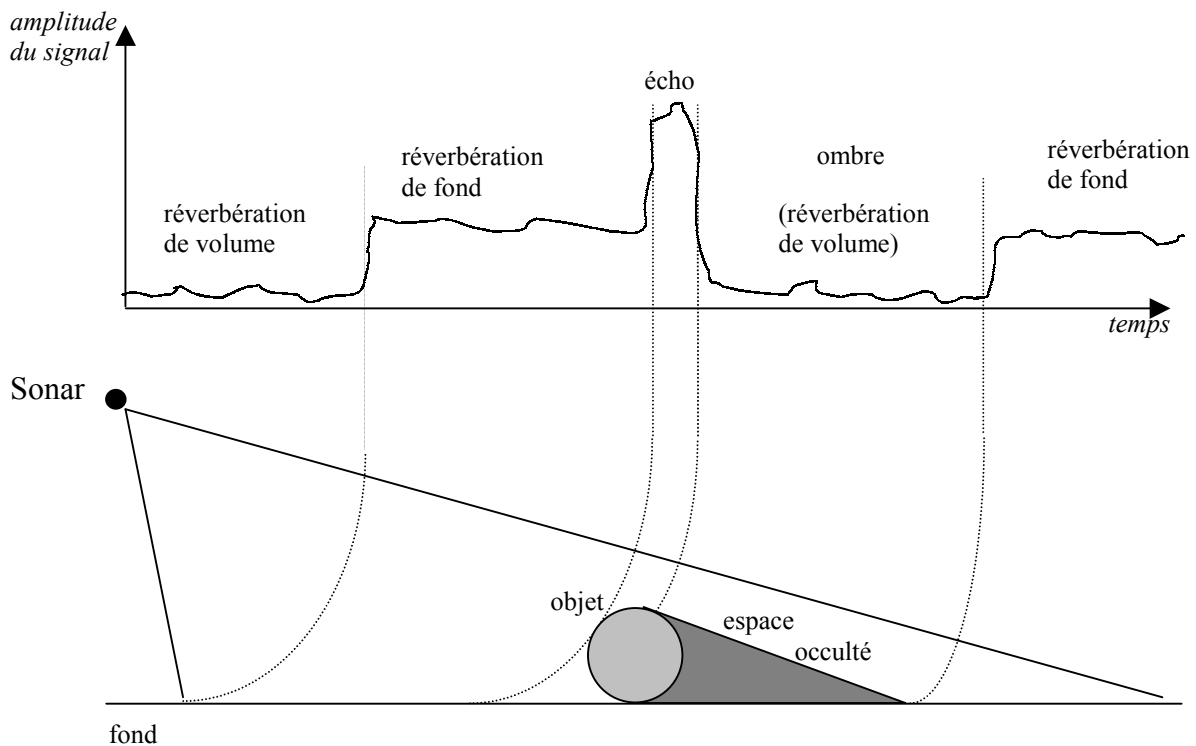


Figure 8 - Principe de classification sur ombre

La présence ou l'absence de réverbération de fond sert à délimiter la zone d'ombre associée à l'objet ciblé. Le principe de classification sur ombre est illustré Figure 8 où l'amplitude du signal reçu dans une voie donnée en fonction du temps évolue comme suit [BOUV-92] :

- un niveau faible dû à la réverbération de volume dans la colonne d'eau (trajet sonar-fond),
- un niveau moyen dû à la réverbération de fond,

- un fort niveau correspondant à l'écho de la cible,
- un niveau faible dû à la réverbération de volume, au bruit ambiant et aux trajets multiples, correspondant à la zone du fond occultée par l'objet et non insonifiée,
- enfin, à nouveau un niveau moyen dû à la réverbération de fond.

IV.4. Hypothèses

→ La dimension des rugosités du fond est faible devant la taille de la cellule de résolution sonar. Comme nous l'avons vu au paragraphe III.1, l'amplitude est par conséquent modélisée par une **loi de Rayleigh** traduisant le bruit multiplicatif appelé speckle [PAPO-84]. Dans le cas contraire, il conviendrait de prendre en compte un modèle plus général comme la loi de Weibull par exemple [MIGN-98].

→ L'image sonar traitée représente les données (distance, gisement) acquises dans le plan oblique : il s'agit d'une **image sectorielle** obtenue à l'aide d'un sonar sectoriel : l'étendue de la zone couverte est délimitée par sa portée R_{max} et le secteur angulaire $\Delta\theta$ qui peut être insonifié en une seule émission ou en plusieurs couvrant chacune un sous-secteur plus étroit (on dit alors que l'image est formée en plusieurs récurrences de durée égale au temps nécessaire à l'impulsion pour parcourir le trajet aller et retour jusqu'à R_{max}). En réception, ce secteur est couvert par plusieurs voies $\delta\theta$ (soit la résolution angulaire) formées par combinaison des signaux de capteurs (cf. Figure 9). La durée de la récurrence $T = 2R_{max}/c$ est

échantillonnée au rythme $T_e = \frac{1}{B}$ permettant de fixer la résolution en distance δR (cf. §II.3).

Notons toutefois que sous l'hypothèse de grandes distances sonar-objets, les traitements développés pour les images du sonar sectoriel sont également applicables au sonar latéral.

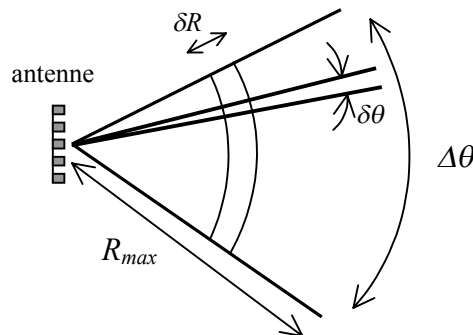


Figure 9 – Secteur insonifié

→ Le nombre de pixels d'écho n'est pas trop important : l'écho ne vient pas 'grignoter' la zone d'ombre et ne perturbe pas la statistique associée aux pixels de la zone de réverbération de fond. Ainsi, dans la plupart des algorithmes de binarisation de l'image sonar (ombre et reste de l'image), les pixels appartenant à l'écho sont affectés à la zone de réverbération de fond du fait de leur fort niveau de gris (forte amplitude du signal reçu).

→ Les mines ne sont pas enfouies.

V. Simulation d'images sonar

Les contraintes opérationnelles (lourdeur de la mise en œuvre) et militaires (confidentialité) associées à l'acquisition d'images sonar dans un environnement naturel rendent difficile le travail sur données réelles. L'élaboration et la validation d'algorithmes nécessitent pourtant de disposer d'images sonar illustrant des configurations (point de vue, nature du fond, relief...) diverses. La simulation vise à pallier ce manque d'images réelles et à obtenir dans des délais raisonnables une base de données répondant à des situations particulières induites par l'utilisation de nouveaux porteurs tels le PVDS (cf. §IV.2). De plus, les potentialités du simulateur permettent d'appréhender des problèmes de multiples manières en tentant d'être le plus exhaustif possible.

Dans le cadre de cette étude, on désire générer des images sonar haute résolution obtenues par un sonar de chasse aux mines fonctionnant donc en large bande (cf. §II.2) [JAN-88].

Les fonctions du simulateur SIMU tiennent compte des principaux aspects de la classification de mines :

a) Environnement marin

Outre les caractéristiques des éléments environnants (vent, salinité, température, pH), il est possible de générer un relief particulier. A chaque diffuseur discret du fond est associé un coefficient de réflexion (plusieurs lois de réverbération possibles).

b) La navigation du porteur du sonar et la sélection de la vue

Ce simulateur permet de simuler une séquence d'images en définissant précisément la trajectoire du porteur. La sélection de la vue dépend à la fois du type du sonar (frontal ou latéral) et de ses mouvements (position, vitesse, cap, lacet, roulis et tangage). Cette fonction est particulièrement utile pour la simulation de données suivant une trajectoire circulaire du porteur autour de la cible.

c) Définition et incrustation des cibles

Posées sur le fond ou entre deux eaux (mines à orin), les cibles sont repérées par les coordonnées du point de référence (centre de gravité pour le cylindre, centre de la sphère...) et orientées en tangage et en cap. L'importance et l'aspect de l'écho de cible dépendent de son type (constant ou points brillants) et de son index de réflexion.

d) Détermination des zones d'ombre

Les zones d'ombres sont les zones du fond non vues du sonar du fait du relief ou de la présence de cible(s).

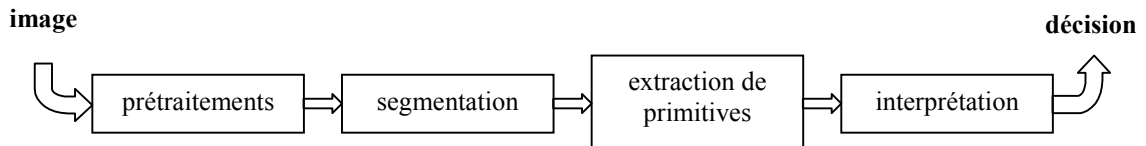
e) Traitement sonar

Le traitement sonar consiste à former l'image en définissant la dimension des pixels et leur niveau de gris. Il prend donc en compte les caractéristiques de l'instrument d'observation qu'est le sonar :

- nature du signal émis (fréquence centrale, durée d'impulsion et bande de fréquence)
- les paramètres d'émission (niveau sonore, lobes de directivité en gisement et en site)
- les paramètres de réception dont le diagramme de directivité en gisement de l'antenne de réception (demi-ouverture à 3dB, nombre de capteurs), les paramètres de la formation de voies (le calcul des voies sonar est effectué en convoluant les coefficients de réflexion par le diagramme de directivité en réception en gisement), l'échantillonnage en distance et en gisement qui définissent la dimension du pixel de l'image finale
- le contraste réverbération sur bruit au-dessous duquel on introduit le bruit.

VI. Méthodologie en traitement d'images appliquée à l'image sonar

L'image sonar est la mise en forme de données brutes acquises par le système sonar au cours de l'insonification. L'**analyse de l'image sonar** consiste à exploiter les données acquises puis discrétisées à l'échelle du pixel [MANT-87]. L'enchaînement des traitements illustré ci-dessous conduit à l'ultime étape d'interprétation permettant de prendre une décision.



Les **prétraitements** consistent à 'préparer' les données de l'image de manière à rendre les traitements ultérieurs plus robustes en corrigeant d'éventuels défauts lors de l'acquisition. La valeur du pixel est ainsi modifiée en fonction de sa localisation dans l'image, de son niveau de gris et/ou des niveaux des pixels voisins [COCQ-95].

La **segmentation** vise à regrouper les pixels de l'image en sous-ensembles appelés régions. Une région est un ensemble connexe de pixels ayant des propriétés communes (niveau de gris, texture,...). Bien qu'une multitude de techniques existent, seules quelques-unes sont utilisées voire adaptées (par un judicieux paramétrage) à l'imagerie sonar.

L'**extraction de primitives ou attributs caractéristiques** offre une description de l'image et de son contenu. En ce qui concerne la classification sur ombre, l'image binaire issue de l'étape de segmentation présente la forme ou silhouette de l'ombre portée de l'objet sur le fond. Selon la nature des données binaires, ensemble compact de pixels ou contour uniquement, deux orientations sont possibles. Nous pouvons manipuler les données avec un minimum de perte d'information et les traduire différemment par exemple, par un codage différent mais discriminant du contour. Une autre possibilité consiste à réduire les données à une grandeur discriminante, géométrique ou statistique par exemple. Dans tous les cas, il s'agit de restreindre la masse de données à un espace caractéristique moindre en extrayant des informations pertinentes.

La phase ultime de l'analyse d'image qu'est l'**interprétation** exploite les attributs précédemment extraits de manière à prendre une **décision**. La classification sur ombre sous-tend la reconnaissance de formes par une procédure de classification supervisée.

Nous montrerons que cette procédure séquentielle peut être diversement remise en cause et concurrencée par des méthodes alternatives (cf. chapitre III.2).

Par ailleurs, dans le cadre d'une classification multi-vues, cette approche classique peut être diversement exploitée. Ce traitement peut bien sûr être aisément effectué en parallèle sur plusieurs images (cf. chapitre IV.1). Il s'agira ensuite d'exploiter au mieux l'ensemble des attributs caractéristiques extraits. Une autre solution consiste à effectuer la même démarche séquentielle mais en l'appliquant à une image particulière résumant l'ensemble des vues (cf. chapitre IV.2).

Notons enfin que la mise en œuvre de tels traitements nécessite néanmoins une bonne appréhension du contexte sonar de manière à définir les outils adéquats.

PARTIE II :

Outils pour le traitement d'images sonar

Introduction de la partie II

Les traiteurs d'images ont à leur disposition de nombreux outils pour l'analyse d'images. La problématique posée dans le cadre de cette étude est la question de leur applicabilité à l'imagerie sonar. En d'autres termes, il s'agit de savoir dans quelle mesure tel ou tel traitement s'avère être adapté aux données observées.

Cette partie est consacrée à l'extraction d'informations pertinentes à deux niveaux. La segmentation, d'une part, fait intervenir des traitements de bas-niveau manipulant les données brutes. L'extraction d'attributs caractéristiques, d'autre part, fait appel à des traitements de haut-niveau manipulant les données binaires segmentées. Ces traitements se feront avec le souci permanent de conserver le lien étroit entre la nature des données traitées et le contexte sonar. Ceci se concrétise de manière différente à chacun des niveaux de traitement :

- A bas niveau, il s'agit de connaître précisément l'information codée dans le pixel. Pour y parvenir, on remonte à l'appareil d'acquisition des données sonar. Les différents traitements mis en œuvre pour permettre de visualiser l'image sonar induisent des effets en termes de taille et de corrélation des pixels. Il convient d'intégrer ces connaissances aux méthodes de segmentation.
- A haut niveau, il s'agit de traiter les distorsions géométriques caractéristiques qui affectent la forme de l'ombre portée de l'objet. Deux voies sont possibles suivant l'objectif visé : s'en affranchir ou s'en servir comme moyen de remonter aux positions relatives de l'objet ciblé et du sonar.

Chapitre II.1 :

Segmentation d'images sonar

I. Introduction du chapitre II.1

La segmentation est une étape initiale et par là même déterminante de la chaîne de traitement. Elle n'est pas une finalité en soi et ne doit donc pas être trop gourmande en temps de calcul. Néanmoins, elle est conditionnée par les traitements ultérieurs et en premier lieu, doit être suffisamment précise pour l'étape d'extraction d'attributs caractéristiques. De nombreuses méthodes sont apparues efficaces pour tel ou tel type d'images avec un paramétrage adapté et parfois l'introduction de connaissances a priori sur le résultat attendu. Les résolutions atteintes, grâce au traitement antenne synthétique notamment, ont entraîné l'application en imagerie sonar de ces méthodes auparavant réservées à d'autres types d'images. La régularisation du contour par exemple, est une donnée récemment prise en compte pour imposer un certain lissage et éviter les détails superflus.

Le sonariste a à sa disposition des connaissances que l'on va tenter d'exploiter pour adapter les traitements au contexte sonar. Ces connaissances sont de deux ordres.

Il s'agit d'une part d'introduire les paramètres d'acquisition de l'image sonar de manière à rendre les traitements indépendants de quelques distorsions géométriques typiques comme la forte élongation de l'ombre due à la rasance, la différence de précision suivant les deux axes de l'image, la corrélation des pixels... En s'adaptant aux conditions d'acquisition de l'image, les traitements gagnent en précision et robustesse voire en temps de calcul. Ces paramètres interviennent lors de la régularisation du contour ou lors de la définition d'éléments structurants en morphologie mathématique. L'exploitation du principe de formation de l'image a par ailleurs inspiré une méthode novatrice de segmentation : une analyse avisée des données corrélées permet de définir la taille minimale d'un filtre dont l'optimisation est basée sur un critère statistique.

La connaissance de la distribution des ondes réfléchies par le fond marin à l'intérieur de la cellule de résolution permet d'autre part de modéliser le bruit de speckle et de connaître ainsi l'allure de la distribution des niveaux de gris des pixels de l'image. Cette attache aux données permet en particulier de mesurer l'adéquation de la segmentation et des données sonar.

Ces connaissances seront exploitées de deux manières. Il s'agit tout d'abord de démontrer l'intérêt de leur intégration dans des trois méthodes classiques de segmentation. Les deux premières méthodes appartiennent aux approches structurelles et agissent sur la structure de l'image par l'intermédiaire d'un terme de régularisation que nous pouvons paramétrer. Ce terme vise à éviter l'instabilité d'un problème inverse basé sur l'attache aux données. La troisième relève de la morphologie mathématique et utilise un outil géométrique paramétrable appelé élément structurant. Outre ces méthodes classiques, nous montrerons comment l'exploitation du principe de formation de l'image sonar a permis de spécifier un filtre spatial pour le seuillage d'images sonar.

Parmi les techniques que nous présentons, deux approches se distinguent : 'région' et 'contour' [COCQ-95].

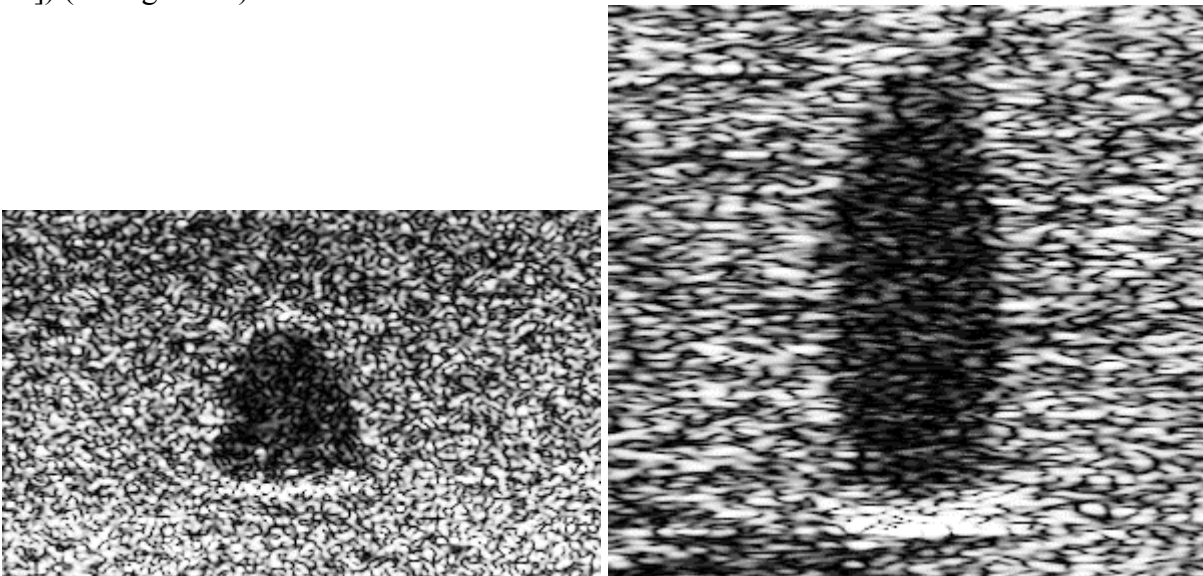
L'approche 'région' fait référence à des groupements de points ayant des propriétés (intensité, texture, ...) communes :

- Si les données le permettent et sous certaines conditions, un seuillage peut être effectué en sélectionnant un ou plusieurs minima locaux sur l'histogramme de l'image ou bien en évaluant le niveau moyen de la réverbération [JAN-87] [SAHO-88]
- Dans le cadre de l'estimation bayésienne par ailleurs, la modélisation markovienne de l'image fait coopérer au sein d'un même modèle statistique des connaissances a priori concernant les informations recherchées ainsi que le processus physique (modélisation du bruit) liant les étiquettes (image segmentée) aux observations (image observée) [GRAF-95] [THOU-96] [MIGN-98].

L'approche 'frontière' ou 'contour' est associée à une discontinuité entre deux ensembles connexes de points. Les méthodes de segmentation par contour visent à détecter les points de contraste important. Les détecteurs de contour classiques assimilent les contours aux points de fort gradient ou de dérivée seconde nulle. Ces opérateurs ne sont pas adaptés pour les images sonar pour lesquelles apparaît un nombre élevé de fausses alarmes. D'autres solutions sont envisageables :

- La morphologie mathématique en niveaux de gris voit l'image comme un relief. Le processus d'inondation appelé ligne de partage des eaux partitionne ainsi l'image en zones homogènes à partir du module du gradient morphologique [MEYE-90] [VACH-95].
- Les contours actifs mettent en œuvre un processus itératif de minimisation d'une fonctionnelle par lequel une courbe évolue sous contraintes [KASS-88] [BOSS-94] [CHES-99].

De manière à faciliter la comparaison, ce chapitre illustre l'application de chacune des méthodes aux mêmes images réelles (obtenues par un traitement antenne synthétique [BILL-98]) (cf. Figure 10).



**Figure 10 - Images réelles obtenues par un traitement antenne synthétique:
à gauche, mine Manta à 50m de distance et à droite mine sphérique à 76.5m de
distance pour altitude du sonar de 10.9m**

II. Connaissances du sonariste

II.1. Résolution

Il s'agit dans un premier temps de paramétrer des méthodes 'classiques' de segmentation dans le contexte propre de l'imagerie sonar. Dans un second temps, nous proposons d'analyser le principe d'acquisition des données sonar pour dériver une méthode originale de segmentation des images sonar.

La première connaissance que nous pouvons exploiter favorablement est liée à l'information relatée au niveau du pixel. Il s'agit d'une part de prendre en compte la taille du pixel en distance oblique r_i et en gisement r_j , et la rasance. On exploitera d'autre part le lien entre cellule de résolution du sonar et pixel de l'image (cf. §II.3 de la partie I).

De manière à effectuer les traitements avec la même précision suivant les deux axes de l'image, de nouvelles valeurs, la taille en cm en hauteur (c'est-à-dire sur les lignes) R_i et la taille en cm en largeur (c'est-à-dire sur les colonnes) R_j , sont obtenues par transformation de la taille du pixel en distance oblique en taille correspondante en hauteur. Pour effectuer ce passage, on prend en compte l'angle de rasance de la prise de vue α tel que : $R_i = r_i \times \sin \alpha$ (cf. Figure 11). Par contre, R_j est identique à r_j . La forte rasance est effectivement la cause d'une forte élongation de l'ombre le long de l'axe des distances. Notons toutefois que cette manipulation suppose cependant que la rasance soit la même sur l'ensemble de l'image. Ceci est légitime compte tenu de la faible étendue (on traite en fait des imagerie ciblées sur l'objet) et d'une distance à l'objet suffisamment grande.

Pour les deux images réelles étudiées, les tailles en distance et gisement sont $r_i = 3.14cm$ et $r_j = 1.25cm$ avec $\alpha = 0.22$ rad (resp. 0.143 rad) pour la mine Manta (sphérique).

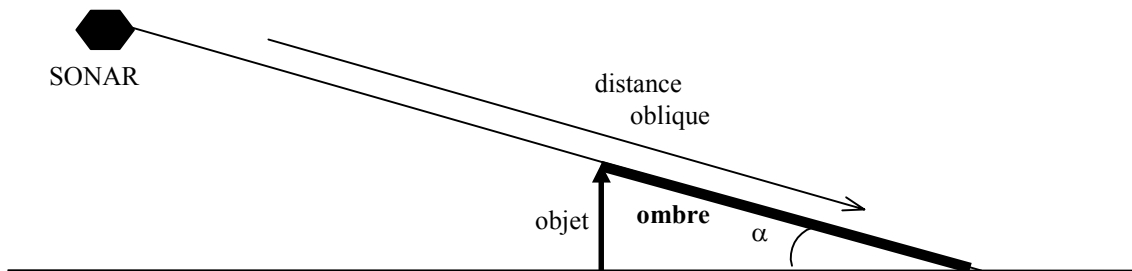


Figure 11 – Prise en compte de l'angle de rasance α

De manière générale, tout traitement agissant de manière localisée dans l'image gagne à être paramétrés par R_i et R_j . Ces grandeurs peuvent par exemple être introduites dans la définition de paramètres géométriques tels les éléments structurants en morphologie mathématique ou les cliques associées à la caractérisation d'un champ de Gibbs. Une autre possibilité consiste à effectuer une duplication de l'image de manière à la traiter indépendamment de la rasance. Nous illustrerons ces trois cas de figure en adaptant trois techniques de segmentation classiques. Deux d'entre elles mettent en œuvre des modèles dits structuraux dans le sens où une information a priori est prise en compte : la modélisation markovienne et les contours

actifs statistiques. La troisième relève de la morphologie mathématique en niveaux de gris : la Ligne de Partage des Eaux.

La taille du pixel dépend de deux facteurs, les paramètres du sonar et le suréchantillonnage. Ce suréchantillonnage n'apporte pas d'information supplémentaire et crée même de la redondance. La connaissance de la taille du pixel en distance oblique r_i et en gisement r_j d'un côté, et des paramètres sonar d'un autre côté, nous permet de calculer simplement les distances de corrélation suivant les deux axes de l'image. Cette connaissance nous permet de prendre en compte les pixels décorrélés qui, à eux seuls, portent l'information utile (pas de redondance). Le Tableau 2 illustre ce phénomène dans le cas de nos deux images réelles.

Tableau 2 – Détermination des distances de corrélation suivant les deux axes de l'image

Formation de l'image	Résolution sonar (en accord avec les paramètres sonar)		Taille du pixel (en accord avec le traitement des données)		Nombre d'échantillons corrélés par cellule de résolution	
	<i>distance</i>	<i>gisement</i>	<i>distance</i>	<i>gisement</i>	<i>distance</i>	<i>Gisement</i>
Image 1 (Manta)	6.25cm	3.7cm	3.15cm	1.25cm	≈ 2	≈ 3
Image 2 (Sphère)		9.5cm				≈ 8

II.2. Attache aux données

L'attache aux données sonar réfère ici à la distribution statistique des pixels. La nature particulière des images acoustiques est ainsi prise en compte via la modélisation par une loi de Rayleigh du phénomène de réverbération (cf. §III.1 de la partie I). Ce bruit de speckle est généré par la somme des multiples contributions des diffusions de l'onde par les inhomogénéités (de l'eau : particules, bulles) ou irrégularités (relief du fond, aspect non-uniforme de la surface) du milieu. Un speckle de fort niveau moyen est caractéristique de la réverbération de fond et contraste avec un speckle de faible niveau moyen pour la réverbération de volume prépondérante dans la zone d'ombre (cf. Figure 12).

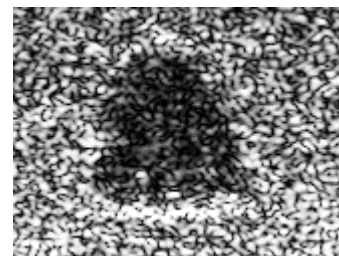
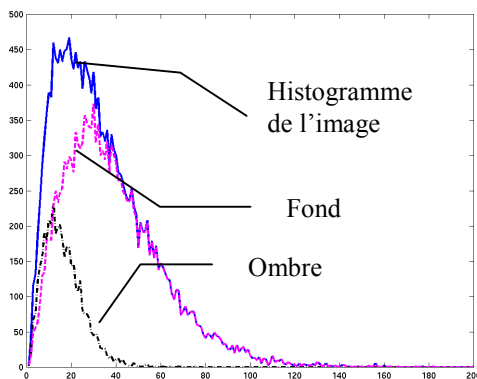


Figure 12 - Mélange des deux distributions de Rayleigh associées aux zones 'ombre' et 'fond'

Pour chacune des zones, les niveaux de gris ou intensités $i(x, y)$ des pixels de coordonnées (x, y) sont distribués selon la loi :

$$p(i(x, y)/\phi_{region}) = \frac{i(x, y)}{\alpha_{region}^2} \exp\left\{-\frac{[i(x, y)]^2}{2\alpha_{region}^2}\right\}$$

avec $\phi_{region} = [\alpha_{region}^2]$ où α_{region}^2 est le facteur d'échelle. *region* désigne, tantôt *O* pour la zone d'ombre, tantôt *R* pour la zone de réverbération de fond.

Deux hypothèses importantes concernent la distribution des pixels [FIGU-97] :

- ☞ les niveaux de gris des pixels de l'image sont indépendamment distribués
- ☞ les deux régions, l'une extérieure au contour recherché, i.e. la réverbération de fond, et l'autre intérieure, i.e. l'ombre, sont homogènes ce qui implique que la probabilité conditionnelle associée à chacun des pixels dépend uniquement de sa localisation dans l'une ou l'autre des régions

Ces hypothèses permettent d'une part d'écrire simplement la fonction de log-vraisemblance des données observées, conditionnellement à la classe à laquelle elles appartiennent, comme la somme des log-vraisemblances marginales, soit :

$$\ln(P(I_{region}/\phi_{region})) = \ln\left(\prod_{(x,y) \in region} p(i(x, y)/\phi_{region})\right) = \sum_{(x,y) \in region} \ln p(i(x, y)/\phi_{region})$$

$$\text{avec } \ln p(i(x, y)/\phi_{region}) = \ln(i(x, y)) - \ln \alpha_{region}^2 - \frac{[i(x, y)]^2}{2\alpha_{region}^2}$$

$$\text{soit, } \ln P(I_{region}/\phi_{region}) = \sum_{(x,y) \in region} \ln(i(x, y)) - N_{region} \ln \alpha_{region}^2 - \frac{1}{2\alpha_{region}^2} \sum_{(x,y) \in region} [i(x, y)]^2$$

si N_{region} désigne le nombre de pixels appartenant à ladite région.

La valeur α_{region} qui maximise cette fonction de vraisemblance est [EVAN-93] :

$$\alpha_{region}^2 = \frac{\sum_{(x,y) \in region} [i(x, y)]^2}{2N_{region}} \quad \text{estimée au sens du maximum de vraisemblance}$$

D'autre part, pour l'ensemble de l'image, il vient :

$$P(I/\phi) = P(I_O/\phi_O) P(I_R/\phi_R)$$

soit

$$\ln P(I/\phi) = \ln P(I_O/\phi_O) + \ln P(I_R/\phi_R)$$

$\ln P(I/\phi) = \sum_{(x,y) \in image} \ln(i(x, y)) - N_O \ln \alpha_O^2 - \frac{1}{2\alpha_O^2} \sum_{(x,y) \in O} [i(x, y)]^2 - N_R \ln \alpha_R^2 - \frac{1}{2\alpha_R^2} \sum_{(x,y) \in R} [i(x, y)]^2$

Maximiser la fonction de vraisemblance revient à maximiser la fonction de log-vraisemblance $\ln(p(I/\phi))$. Ce critère vise à fournir une partition des pixels au sens du maximum de vraisemblance (MV) et peut être vu comme une énergie. Son évaluation nécessite l'estimation des paramètres α_O et α_R au sens du MV et leur injection dans la fonction de log-vraisemblance.

Ces connaissances d'ordre géométrique ou statistique sont exploitables pour la mise en œuvre de méthodes de segmentation. En paramétrant judicieusement les traitements, la résolution peut être prise en compte de manière à fournir un résultat indépendant de celle-ci. Le contexte statistique des données sonar est par ailleurs favorable à la résolution de problèmes inverses basés sur l'attache aux données.

III. Adaptation de techniques de segmentation classiques au contexte sonar

III.1. Modèles structurels pour une application sonar

L'étude de la dégradation des observations recueillies et la seule modélisation de ces phénomènes physiques (attache aux données) ne suffit généralement pas. Hadamard a caractérisé ce type de problème en définissant la notion de problème *mal-posé* [HADA-23]. La recherche d'informations bien définies à partir des données observées ne peut se faire sans l'apport de contraintes additionnelles sous la forme d'un terme de régularisation. Cette prise en compte de contraintes ou informations a priori sur la structure du contour caractérisent les modèles structurels [COCQ-95]. Parmi l'ensemble des modèles structurels rencontrés dans la littérature, nous étudions un modèle stochastique (modélisation markovienne) et un modèle élastique (contours actifs statistiques). Le problème consiste à adapter les paramètres réglant l'importance de la régularisation au contexte sonar et en particulier, à la résolution. La régularisation doit en effet se faire de manière isotrope, c'est-à-dire quelle que soit la grille d'échantillonnage.

Ces deux modèles permettent à la fois d'introduire des informations a priori et de modéliser les phénomènes physiques affectant l'image. Leur principe théorique est détaillé dans l'annexe A. Pour chacune de ces méthodes, une approche bayésienne consiste à maximiser la probabilité a posteriori, i.e. à trouver la configuration des pixels (contour ou régions) qui partage les données sonar de manière optimale au sens du Maximum de vraisemblance A Posteriori (MAP). Il revient en fait à minimiser une énergie composée de deux termes : terme de régularisation et terme d'attache aux données (modèle de bruit). La segmentation markovienne d'une part, est une approche région largement discutée et appliquée à l'imagerie sonar [THOU-96] [MIGN-98]. La méthode des contours actifs statistiques d'autre part, est une approche variationnelle misant sur la recherche de la frontière (contour fermé) qui partage l'image de manière optimale.

III.1.1. Champs markoviens pour la segmentation

Dans l'approche markovienne, la segmentation est considérée comme l'estimation du champ des étiquettes X (image segmentée) à partir des observations Y (image observée) (cf. annexe A) [GEMA-84].

▪ Description du modèle

- ♦ $U(x)$ terme contextuel ou de régularisation qui introduit un effet régularisant sur le champ des étiquettes

Les contraintes ou informations a priori que l'on possède sur la solution attendue (propriétés géométriques, homogénéité, variabilités spatiales,...) s'expriment par l'intermédiaire des fonctions de potentiel $V_c(x)$. Nous avons alors opté pour le modèle autologistique réduit aux cliques binaires $c = \langle s, t \rangle \in C$, le site $s \in S$ appartenant au voisinage v_t , tel que

$$V_{\langle s, t \rangle}(x_s, x_t) = \begin{cases} -\beta_{\langle s, t \rangle} & \text{si } x_s = x_t \\ \beta_{\langle s, t \rangle} & \text{sinon} \end{cases}$$

où $\beta_{\langle s, t \rangle}$ paramètre spécifique à chaque type de cliques binaires (horizontale, verticale et diagonale) autorise le contrôle de la taille et de la direction des regroupements des pixels [CROS-83].

Les cliques binaires associées à un système de voisinage 8-connexité sont :



Le terme d'énergie contextuel s'écrit $U(x) = \sum_{c \in C} V_c(x)$.

- ♦ $V(x, y)$ est le terme d'attache aux données qui caractérise l'adéquation des étiquettes aux observations

La loi théorique retenue pour modéliser le phénomène de réverbération et l'histogramme des pixels des images sonar est la loi de Rayleigh

$$\mathfrak{R}(y_s, \alpha_{x_s}) = \frac{y_s}{\alpha_{x_s}^2} \exp\left\{-\frac{y_s^2}{2\alpha_{x_s}^2}\right\}$$

α_{x_s} est calculé à partir des N_{x_s} pixels d'étiquette x_s avec $x_s=0$ (resp. $x_s=1$) pour la zone d'ombre (resp. du fond).

On a alors, sous l'hypothèse que les observations sont des variables aléatoires indépendantes (cf. §II.2),

$$V(x, y) = \sum_{s \in S} \psi_s(x_s, y_s) = -\sum_{s \in S} \ln(\mathfrak{R}(y_s, \alpha_{x_s})) = \frac{1}{2\alpha_{x_s}^2} \sum_{s \in S} y_s^2 - \sum_{s \in S} \ln\left(\frac{y_s}{\alpha_{x_s}^2}\right)$$

▪ Paramétrage du modèle

La première étape du traitement consiste généralement à **estimer les paramètres du modèle**. Par souci de gain en temps de calcul, nous avons simplifié la procédure. Sous l'hypothèse d'une image binaire fournissant une carte initiale des étiquettes satisfaisante, nous avons écarté les méthodes d'estimation itératives des paramètres d'attache aux données telles que « Stochastic Expectation Maximisation (SEM) » ou « Estimation Conditionnelle Itérative (ECI) » qui alternent estimation de paramètres et optimisation d'une fonction d'énergie résultante [CELE-86] [PIEC-94] ou encore, la méthode d'estimation conjointe des paramètres comme « ECI Global » [MIGN-98].

D'une part, le poids des cliques est fixé en utilisant le principe de formation de l'image. D'autre part, les paramètres des lois traduisant l'attache aux données (paramètres du modèle de bruit) sont estimés au sens du MV. En somme, nous procédons comme suit :

- l'estimateur au sens du MV du paramètre d'échelle de la loi de Rayleigh associée aux N_{x_s}

sites d'intensité y_s est : $\hat{\alpha}_{x_s}^2 = \frac{\sum y_s^2}{2N_{x_s}}$

Ce paramètre est évalué pour chacune des zones d'ombre et de réverbération de fond initialisées par une image binaire issue d'un prétraitement.

- Pour fixer les valeurs des poids $\beta_{k,k \in \{1, \dots, 4\}}$ associés aux quatre cliques binaires du modèle autologistique, nous choisissons **d'adapter le poids des cliques en fonction de R_i la taille en hauteur (c'est-à-dire sur les lignes) et R_j la taille en largeur (c'est-à-dire sur les colonnes) dans l'image** (cf. §II.1) tel que :

$$\beta_1 = \frac{p}{R_j}; \quad \beta_2 = \frac{p}{R_i}; \quad \beta_3 = \beta_4 = \frac{p}{\sqrt{R_i^2 + R_j^2}}$$

En somme, plus la résolution est grossière (R . élevé), moins la régularisation est importante de manière à préserver les informations à l'échelle de la résolution. Inversement, si l'image est très bien résolue, la régularisation est importante de manière à éliminer les détails superflus. Par ailleurs, plus l'amplitude donnée au poids p est importante, plus le contour est régularisé.

▪ Mise en œuvre

L'analyse multirésolution semble la plus adaptée actuellement aux images bruitées par speckle. Elle a en outre l'avantage de diminuer les temps de calcul.

Pratiquement, on construit une pyramide des données à partir de l'image observée par décomposition en ondelettes (cf. annexe B). Ceci permet d'améliorer le rapport signal sur bruit au fur et à mesure que la résolution devient plus grossière ce qui facilite l'opération de segmentation. Dans le même temps, on constate aux faibles résolutions un fort lissage des zones homogènes lié à la perte de petits détails. En somme, alors que la certitude de l'existence d'une ombre de taille significative croît, la précision sur la position de son contour décroît avec le niveau de résolution. La décomposition est équivalente à une opération de filtrage passe-bas.

La configuration initiale du champ des étiquettes est obtenue par seuillage de la composante Basse Fréquence de la décomposition en ondelettes (jusqu'au niveau de résolution 2^2) où les zones d'ombre (zones basses fréquences) sont rehaussées. Le seuillage à -6dB au-dessous du niveau de réverbération correspondant à la valeur minimale entre les valeurs moyenne et médiane calculées sur l'ensemble des pixels [JAN-87] permet de détecter des ombres qui apparaissent plus compactes.

L'optimisation de la fonction d'énergie $U(x,y)$ du modèle markovien adopté pour la segmentation de l'image est effectuée par algorithme Iterated Conditional Modes (ICM) qui consiste en la recherche du MAP [BESA-86] (cf. annexe A). Ce choix est motivé par l'hypothèse que l'initialisation obtenue par analyse multirésolution des données est satisfaisante.

Deux alternatives résumées sur le diagramme de la Figure 13 se présentent :

- dupliquer deux fois l'image basse fréquence et effectuer une segmentation à pleine résolution par ICM
- effectuer une **segmentation multirésolution**, i.e. trois segmentations menées à trois échelles différentes de la résolution la plus grossière à la pleine résolution. Nous estimons ainsi une pyramide multirésolution d'étiquettes en choisissant, dans notre cas, à chaque niveau, un modèle markovien **adapté à l'échelle** de sorte que :

- les cliques sont pondérées par le facteur d'échelle de la décomposition en ondelettes, soit

$$\beta_k^m = \frac{\beta_k^0}{2^m} \text{ à la résolution } 2^m$$

- les paramètres d'attache aux données sont estimés au sens du MV à partir du champ des étiquettes obtenu par seuillage à la résolution la plus grossière ou bien, par duplication (d'un facteur 2 correspondant au facteur inter-échelle de la décomposition en ondelettes) de l'image segmentée obtenue à la résolution inférieure.

L'optimisation à chaque niveau est faite par ICM à partir de l'image binaire (ou champ des étiquettes).

Sur les résultats obtenus, outre la qualité de la pyramide des étiquettes estimée, on observe que la descente en résolution est profitable en augmentant le contraste réverbération sur ombre ce qui permet d'obtenir une bonne initialisation par seuillage. Cette initialisation est un point sensible de la méthode.

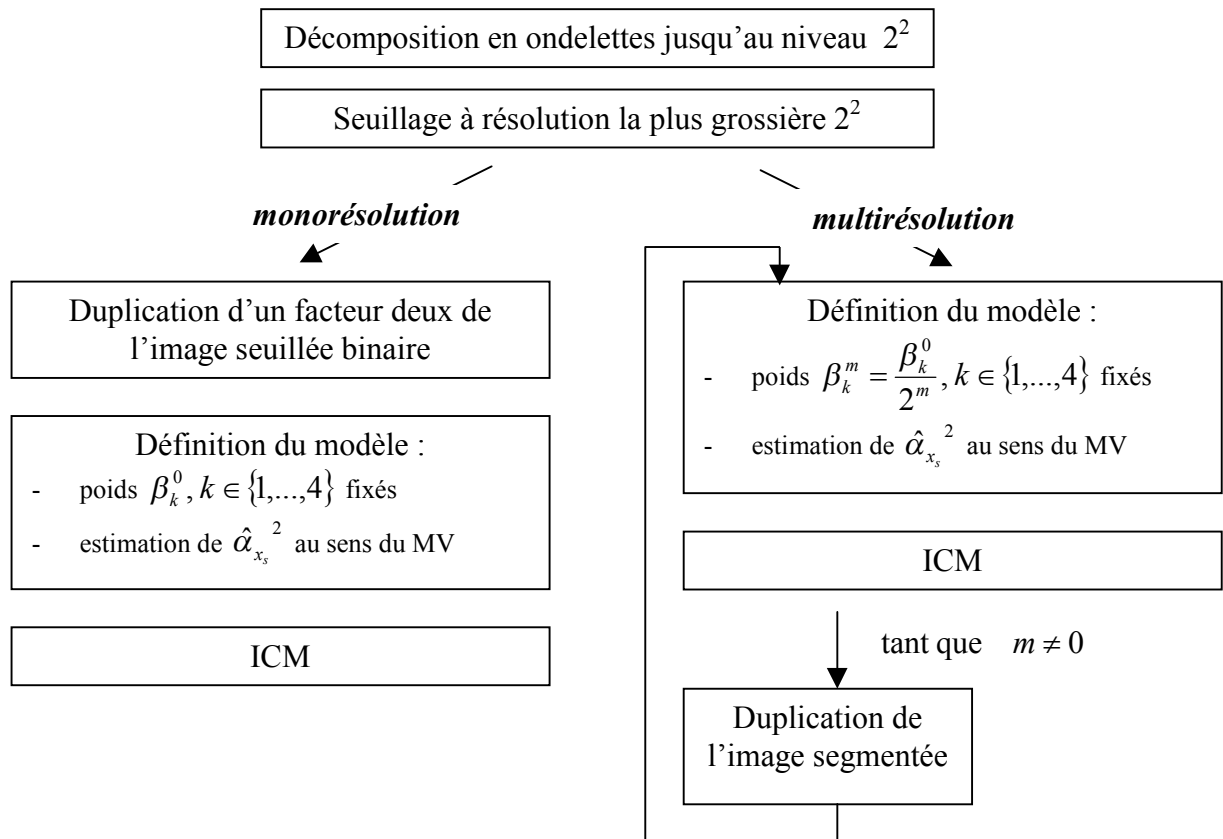


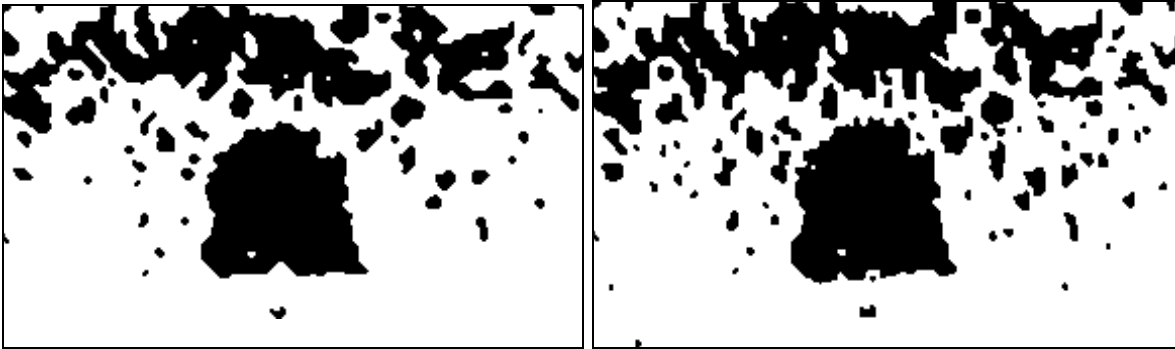
Figure 13 – Synoptique des traitements de segmentation markovienne hiérarchique

▪ Application

Compte tenu de la taille des pixels et des rasances données au §II.1, nous avons des cliques de valeurs suivantes :

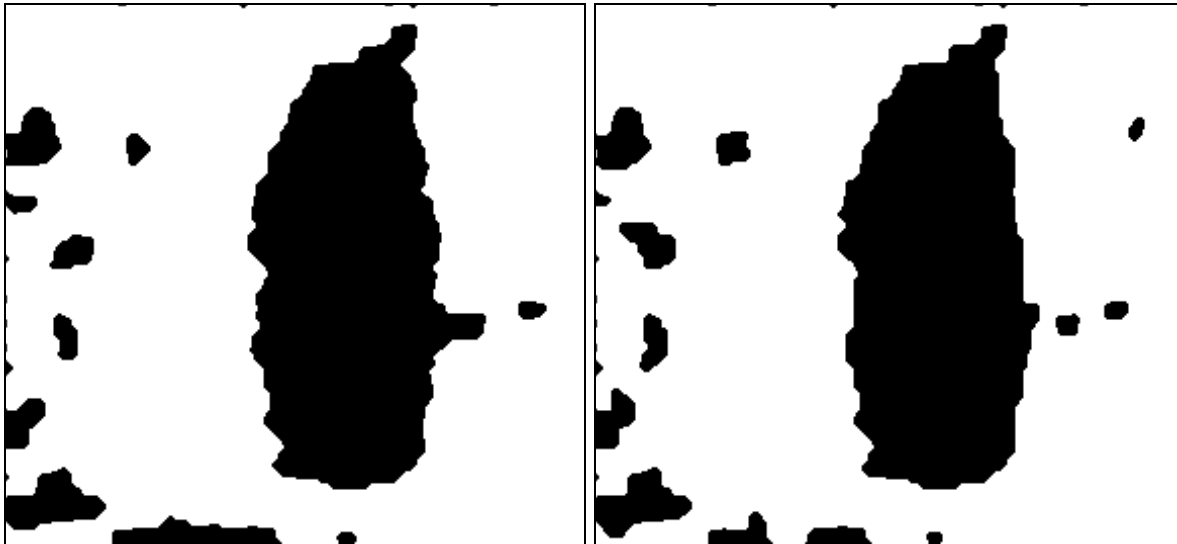
$$\begin{aligned} \beta_1 &= 0.80p, & \beta_2 &= 1.46p, & \beta_3 &= \beta_4 = 0.70p, & \text{pour la mine Manta,} \\ \beta_1 &= 0.80p, & \beta_2 &= 2.23p, & \beta_3 &= \beta_4 = 0.75p, & \text{pour la mine sphérique} \end{aligned}$$

Premier cas : en monorésolution, comparaison du modèle de bruit isotrope ($\beta_k = p, \forall k$) à gauche et du modèle adapté à la résolution proposé à droite avec $p=0.2$



Grâce à l'adaptation à la résolution, la base de l'ombre est mieux préservée sur la seconde image.

Second cas : en multirésolution, comparaison du modèle de bruit isotrope ($\beta_k^0 = p, \forall k$) à gauche et du modèle adapté à la résolution proposé à droite avec $p=1$



L'effet de lissage ou régularisation des côtés de la forme est plus important sur la seconde image. Ceci démontre bien qu'à une plus fine résolution (côtés de l'ombre mieux décrits grâce à une forte rasance) correspond un plus fort lissage.

III.1.2. Contours actifs statistiques [CHES-99]

Le principe du contour actif est de rechercher une position d'équilibre entre forces internes et forces externes (cf. annexe A). Les forces internes d'une part, sont liées à la courbure du contour. Les forces externes d'autre part, règlent l'adéquation entre le contour actif et les données observées (l'image).

▪ Description du modèle

- ♦ l'énergie interne E_{int} de régularisation du contour :

$$E_{int} = \sum_{k=1}^{\text{Nombre de sommets}} d_k^2$$

d_k est la distance reliant le sommet P_k au milieu du segment liant le sommet précédent P_{k-1} au sommet suivant P_{k+1} . $\{P_k, k=1 \dots \text{Nombre de sommets}\}$ est l'ensemble des sommets composant la ligne polygonale représentant le contour.

- ♦ l'énergie externe E_{ext} d'attache aux données de l'image

Ce terme est lié aux statistiques des zones intérieure et extérieure délimitées par le contour fermé recherché. Compte tenu des hypothèses formulées au §II.2, tous les pixels appartenant à la zone d'ombre (*resp.* de réverbération de fond) ont une distribution commune ϕ_{in} (*resp.* ϕ_{out}). Connaître $\phi = [\phi_{in}, \phi_{out}]$ revient à connaître la partition W de l'image (cf. annexe A) et il revient à maximiser la fonction de vraisemblance suivante [FIGU-97] :

$$P(I/W(\phi)) = P(I/\phi) = \left(\prod_{(x,y) \in O} p(i(x,y)/\phi_{in}) \right) \left(\prod_{(x,y) \in R} p(i(x,y)/\phi_{out}) \right), (1)$$

où $i(x,y)$ désigne l'intensité ou niveau de gris du pixel (x,y) , O (*resp.* R) la zone d'ombre (*resp.* de réverbération de fond).

Comme nous l'avons dit précédemment, dans le cas des images sonar, les niveaux de gris des pixels suivent une loi de Rayleigh.

Maximiser la fonction de vraisemblance (1) revient à maximiser la fonction de log-vraisemblance $\ln(p(I/\phi))$, i.e. minimiser l'énergie suivante :

$$E_{ext} = N_O \cdot \ln \left(\frac{1}{2N_O} \sum_{\text{pixels}(x,y) \in \text{Ombre}} i(x,y)^2 \right) + N_R \cdot \ln \left(\frac{1}{2N_R} \sum_{\text{pixels}(x,y) \in \text{Réverb}} i(x,y)^2 \right).$$

▪ Paramétrage du modèle

L'énergie totale à minimiser est paramétrée par λ qui gère le poids accordé à l'énergie interne de régularisation du contour [CHES-99] :

$$E_{tot} = (1 - \lambda) E_{ext} + \lambda E_{int}$$

Pour adapter cette méthode aux images sonar, **l'image est prétraitée** de manière à garantir une même régularisation quelle que soit la rasance et la résolution. Il s'agit de rééquilibrer les tailles en hauteur R_i et en gisement R_j des pixels de l'image (cf. §II.1). Nous effectuons alors

une simple **duplication** des lignes ou des colonnes qui rend le pixel « carré » de côté approximativement égal à la taille la plus fine (même précision suivant les deux directions).

Deux cas se présentent :

- si $R_j > R_i$, on duplique $E(R_j/R_i)$ fois les colonnes de l'image (gisement),
- si $R_i > R_j$, on duplique $E(R_i/R_j)$ fois les lignes de l'image (distance),

($E(.)$ désigne la fonction entier le plus proche).

▪ **Mise en œuvre**

Le modèle de contour actif se présente sous la forme d'une courbe fermée, dont l'initialisation est située à proximité du contour recherché et dont l'évolution s'effectue selon un processus itératif de déformation.

Un point délicat des contours actifs est l'initialisation du contour et notamment sa localisation. Le contour est initialisé par une liste chaînée de quatre sommets dont la largeur dépend du format de l'image dupliquée et centré sur le marqueur extrait selon la méthode suivante qui consiste en une analyse statistique multirésolution de l'image.

Pour cela, nous procédons itérativement en effectuant une descente en échelle de sorte que nous travaillons sur des imagerie de plus en plus petites. Le principe de base est la décomposition de l'image en M imagerie de taille identique. Pour chacune, on extrait deux grandeurs statistiques, soient la moyenne et la valeur minimale des niveaux de gris des pixels. Ces deux grandeurs sont les composantes du vecteur caractéristique dont on calcule le module. On sélectionne alors l'imagerie de faible module dont l'écart avec les modules des imagerie voisines est important. En somme, on cherche à sélectionner une imagerie dont les faibles grandeurs statistiques diffèrent de manière importante de ses voisines. Ce processus est réitéré sur l'imagerie sélectionnée : décomposition de cette imagerie *mère* en M imagerie *filles*, détermination des vecteurs caractéristiques et sélection de l'imagerie... Le critère d'arrêt suit deux options :

- lorsque la taille des imagerie est inférieure ou égale au sixième de la dimension de l'image, on stoppe la décomposition et le centre de la dernière imagerie sélectionnée constitue le marqueur recherché
- on constate une certaine homogénéité de l'imagerie mère, c'est-à-dire que deux ou plus de ses filles sont rivales, le centre de l'imagerie mère est prise pour marqueur.

L'algorithme itératif multirésolution consiste à augmenter progressivement le nombre de sommets et à calculer à chaque fois la nouvelle énergie. En ajoutant des sommets entre deux sommets de l'itération précédente, la précision du contour est plus fine par division de la distance entre sommets. Ces sommets intermédiaires sont alors déplacés pour minimiser l'énergie totale du contour. Le nombre de sommets ajoutés à chaque itération dépend à la fois du détail désiré de la segmentation et de l'itération.

La prise en compte de l'énergie interne augmente progressivement avec la résolution en partant de $\lambda = 0$ (initialisation du contour grossier à quatre sommets).

La Figure 14 expose brièvement le principe du traitement.

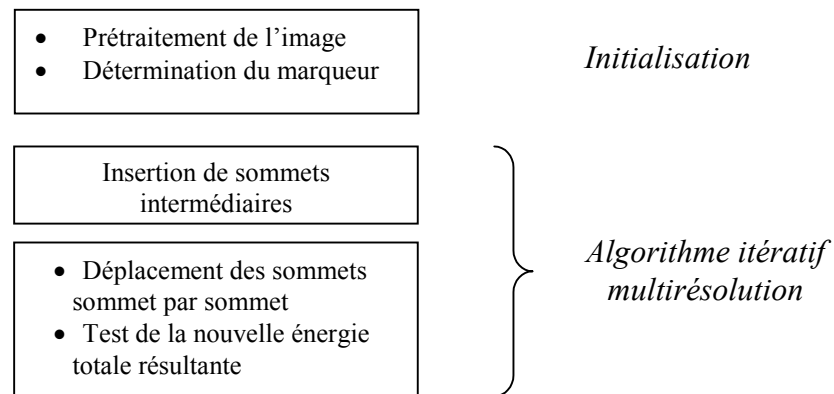
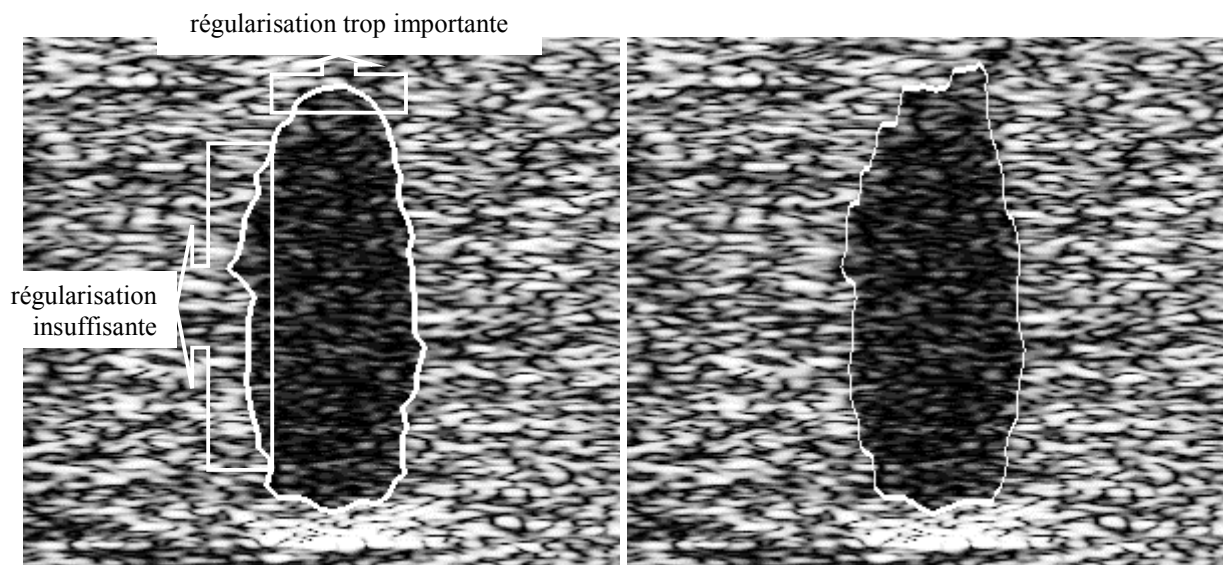


Figure 14 – Etapes du traitement « contours actifs statistiques »

▪ **Application**

Une comparaison est effectuée ci dessous pour une segmentation sans (à gauche) puis avec (à droite) prise en compte de la résolution. Le traitement est effectué en 6 itérations, λ évoluant de 0 à 0.2 par pas de 0.04.



III.1.3. Comparaison des deux approches

Le Tableau 3 affiche les ressemblances et les dissemblances des deux approches précédentes.

Tableau 3 – Comparaison des deux modèles structurels

	Segmentation markovienne	Contours actifs statistiques
Type de modèle	Modèle stochastique	Modèle élastique
Type d'approche	'Région' : Image considérée comme un champ de Markov	'Contour' : description polygonale du contour
Attache aux données	Modélisation du phénomène de réverbération par une loi de Rayleigh	
Régularisation	Valorisation des regroupements de pixels au moyen de cliques	Nombre de sommets de la ligne polygonale (courbure)
Compromis attache aux données / régularisation	Poids p associé aux cliques	Paramètre λ de pondération des énergies internes et externes
Action de la régularisation	Les cliques binaires correspondent au système des 8 plus proches voisins	λ agit indifféremment dans toutes les directions
Adaptation au contexte sonar	Les cliques binaires sont adaptées à la taille du pixel par l'intermédiaire de R_i et R_j	Image préalablement duplicquée pour garantir la même régularisation suivant les deux dimensions

III.2. Morphologie mathématique

L'élément structurant joue un rôle essentiel dans les transformations morphologiques. Sa forme géométrique vise à fournir des renseignements morphologiques sur l'objet bidimensionnel (ou forme) traité. Comme nous l'avons vu, la forme de l'ombre portée de l'objet ciblé par le sonar peut être très différente du fait de la rasance et de la résolution du sonar. Ces données sont à prendre en compte lors de l'opération morphologique et notamment par adaptation de la taille de l'élément structurant.

Nous illustrons ceci en morphologie mathématique en niveaux de gris en appliquant la technique de la Ligne de Partage des Eaux (LPE) [MEYE-90] [VACH-95]. Son principe théorique illustré Figure 15 est rappelé dans l'annexe A.

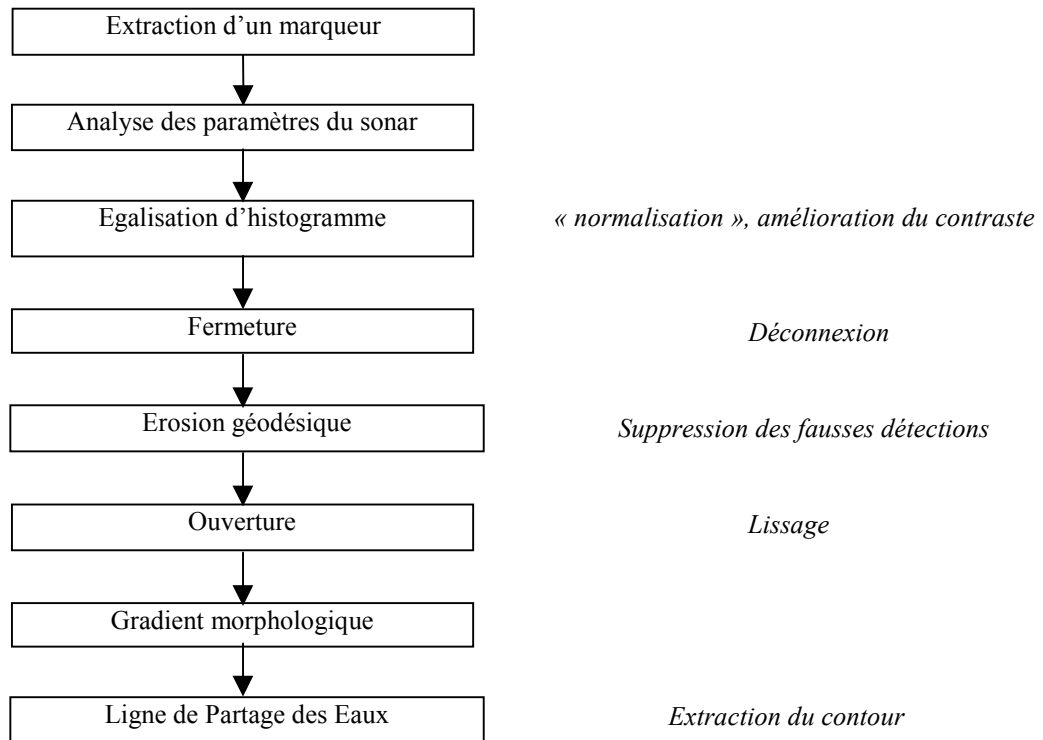


Figure 15 - Etapes du traitement « Ligne de Partage des Eaux »

■ Paramétrage

La forme rectangulaire de nos éléments structurants est paramétrée par la taille du pixel. Sa dimension est proportionnelle à $H \times L$, pour Hauteur suivant les lignes $\times$ Largeur suivant les colonnes. En transformant la taille en distance oblique en taille en hauteur comme nous l'avons proposé au paragraphe II.1, le nouveau rapport hauteur/largeur est

$$(r_i \times \sin \alpha) / r_j = R_i / R_j = RappHL$$

La fenêtre de traitement s'adapte à la résolution de sorte qu'à une fine résolution correspond une taille de fenêtre importante (davantage de pixels sont pris en compte) et inversement. Sa

taille est déterminée à partir du niveau de détail recherché (pratiquement, on prend 5cm sur l'objet réel) par

$$\begin{cases} L = \text{Détail}/R_j & \text{et} & H = L/R_{appHL} & \text{si } R_{appHL} \leq 1 \\ H = \text{Détail}/R_i & \text{et} & L = H \times R_{appHL} & \text{si } R_{appHL} > 1 \end{cases}$$

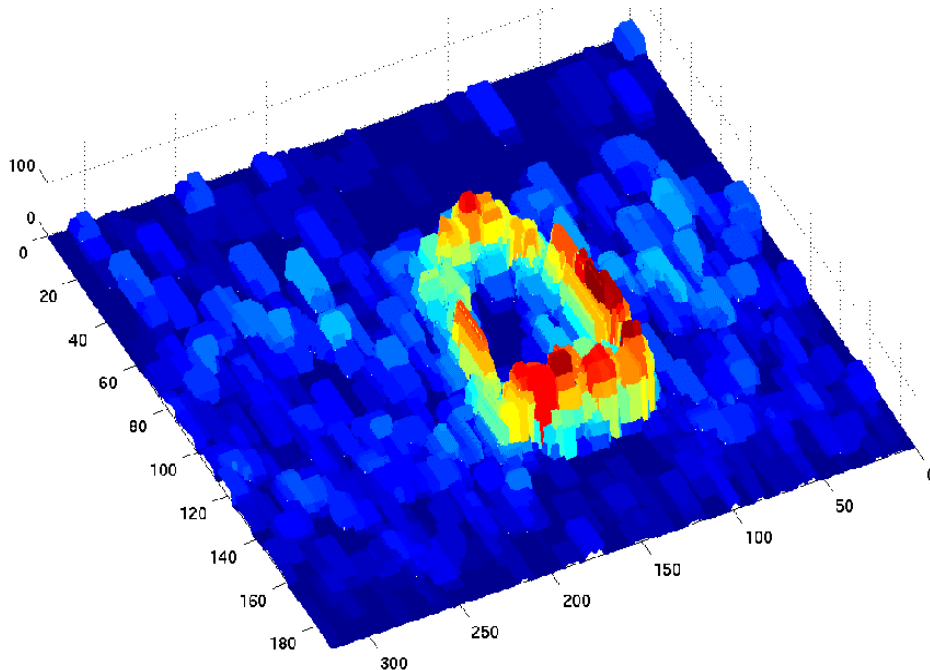
▪ Mise en œuvre

La LPE est obtenue à partir de l'inondation d'un relief concrétisé par le gradient morphologique. Avant de calculer ce gradient, l'image doit être débruitée. Pour ce faire et de manière à travailler sur tous les niveaux de l'image à la fois, on applique des opérations morphologiques en niveaux de gris telles qu'une fermeture, une ouverture et une érosion géodésique. La fermeture en niveaux de gris par élément structurant de taille $H \times L$ a l'effet de combler les vallées encaissées de l'image. Autrement dit, on élimine les petites structures sombres de l'image. Les zones foncées de l'image sont ainsi déconnectées, et les fausses détections pourront être éliminées. Par érosion géodésique infinie d'un marqueur dans l'image, on supprime les fausses détections. L'ouverture en niveaux de gris par élément structurant de taille $H \times L$ a pour effet de casser les pics de l'image. Autrement dit, on élimine les petites zones claires de l'image. On cherche ici à lisser les formes. C'est l'opération duale de la fermeture.

A ce stade, le gradient de l'image peut être calculé. Il est défini comme le résidu entre la dilatation et l'érosion avec un élément structurant de taille $m \times n$ définis par :

$$\begin{cases} m = \max(1.5 \times H, 3) \\ n = \max(1.5 \times L, 3) \end{cases}$$

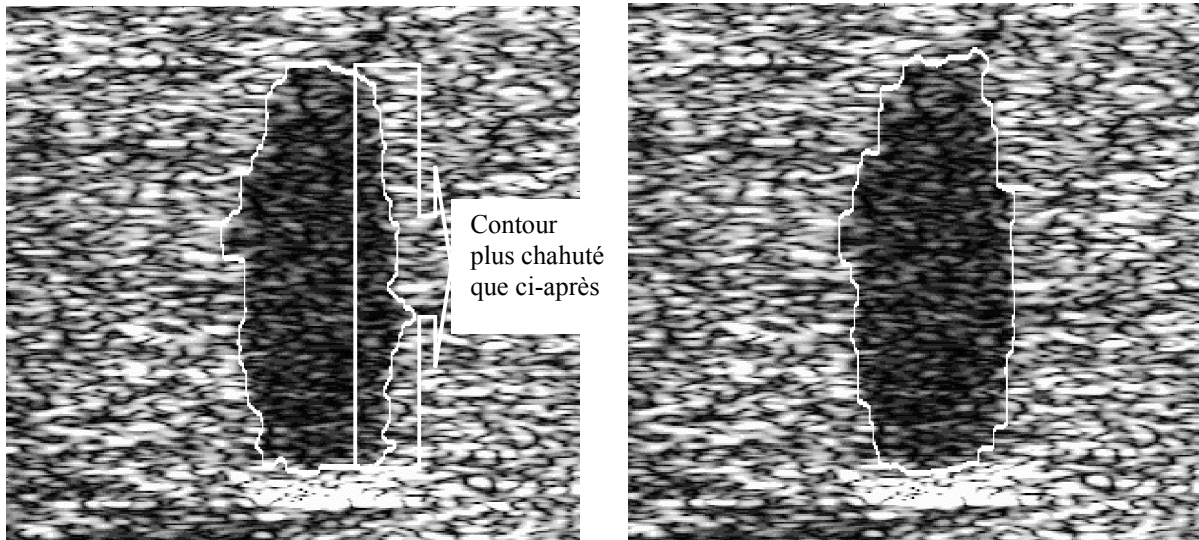
Les valeurs élevées du module du gradient correspondent aux zones à fort contraste de l'image.



▪ **Application**

D'après les valeurs données au §II.1 pour la mine sphérique, H est égal à 11 et L à 4. Par ailleurs, l'image est de dimension 286 par 320 pixels.

La première image présente un contour obtenu par une fenêtre de traitement carrée en forçant $H = L = \text{arrondi}\left(\frac{H+L}{2}\right)$, i.e. 8. A côté correspond le contour obtenu pour une fenêtre rectangulaire respectant la taille des pixels.



Globalement, on observe un nombre de fausses alarmes (pixels de réverbération affectés à la zone ombre) plus important si la résolution n'est pas prise en compte.

IV. Spécification d'un filtre spatial pour le seuillage d'images sonar

Cette méthode innovante basée sur la distribution des niveaux de gris de l'histogramme de l'image est spécifiquement adaptée à la segmentation d'images sonar [QUID-01b]. Elle fait en effet appel à une **double compétence** :

- ☞ Problématique du traiteur d'image : seuiller l'image de manière à la binariser.
- ☞ Connaissance du sonariste : principe de formation des images sonar.

IV.1. Principe

Une méthode classique de segmentation par région consiste à binariser l'image par seuillage en utilisant l'information globale de l'image comme l'histogramme ou les grandeurs statistiques des pixels [SAHO-88]. La zone d'intérêt se détache alors du fond de l'image. En imagerie sonar cependant, la sélection du seuil n'est pas triviale et s'applique sous certaines conditions. Par exemple, un seuil fixé par rapport au niveau moyen de réverbération exige un certain contraste réverbération sur ombre et une faible corrélation des pixels [JAN-87].

Si la zone d'ombre occupe une zone suffisamment importante de l'image, la répartition par niveau de gris des pixels de l'image doit vraisemblablement faciliter la sélection du seuil. Malheureusement, la distribution associée aux pixels de l'ombre est généralement *noyée* dans la distribution associée à la réverbération de fond. Ces distributions classiquement modélisées par une loi de Rayleigh du fait du bruit de réverbération (cf. §III.1 de la partie I) ne se distinguent pas l'une de l'autre. L'histogramme des niveaux de gris des pixels de l'image sonar est par conséquent unimodal et difficilement exploitable entraînant une faible séparabilité des deux classes.

Afin de rendre l'histogramme **bimodal** et rendre possible l'extraction d'un seuil entre les deux modes, on modifie par filtrage les caractéristiques statistiques des deux régions. Les modes de l'histogramme apparaissent en effet si les pixels de chacune des deux régions se concentrent autour de leurs valeurs moyennes respectives (dans le cas d'images à moyennes discriminantes). Sur l'histogramme de l'image filtrée, les pics correspondent aux valeurs moyennes des deux régions 'ombre' et 'fond' alors que la seule vallée correspond à la frontière c'est-à-dire le contour de l'ombre. Pour y parvenir, il revient à déterminer le **filtre optimal** qui minimise en sortie la **variance** des intensités des pixels de l'image. Si les pixels sont décorrélés, un filtre passe-bas suffisamment sélectif suffit. En imagerie sonar, les pixels sont cependant corrélés et ce du fait même du processus de formation des images (cf. §II.3 de la partie I). En tirant profit du principe d'acquisition de l'image et de la connaissance des paramètres sonar, on vise à spécifier le filtre pour l'image traitée.

IV.2. Optimisation du filtre spatial

L'optimisation du filtre 2D à partir de différentes zones de réverbération de fond réelles et simulées montre que son allure est directement reliée à la corrélation des pixels. Pour une image donnée, nous pouvons donc inversement exploiter nos connaissances sur le principe de formation de l'image sonar (les paramètres sonar et le traitement des données) pour synthétiser le **filtre optimal** $H = (h_{i,j})_{i,j \in [0,d_r] \times [0,d_g]}$ minimisant la variance et sous la contrainte

d'une somme des coefficients égale à 1 pour ne pas modifier la moyenne. Sa taille minimale correspond exactement à la distance de corrélation inter-pixels en distance et en gisement (d_r, d_g). L'optimisation s'exprime comme suit :

$$H = \underset{H}{\operatorname{Arg\,min}} \operatorname{VAR}(H * I),$$

où VAR représente la variance, $*$ est l'opérateur de convolution, et I est l'image observée,

$$\text{sous } \sum_{i=0}^{d_r} \sum_{j=0}^{d_g} h_{i,j} = 1$$

initialisé par $h_{i,j} = 1 / ((d_r + 1) \times (d_g + 1))$, $\forall i, j$, i.e. un filtre moyenneur.

Cette optimisation sous contrainte est effectuée par une méthode de programmation non linéaire appelée *Sequential Quadratic Programming method* et implémentée dans la boîte à outils *Optimization toolbox* de Matlab [FLET-80] [COLE-99].

En réalité, la fonction d'appareil du système d'acquisition est similaire à un opérateur de convolution des données brutes. En ce sens, nous sommes proches des problèmes de déconvolution étudiés en imagerie satellitaire ou médicale [KALI-99] [JAMM-99]. Dans notre cas, le traitement des données sonar est à l'origine de la redondance de l'information initialement captée par l'antenne. Le filtre spécifié pour ce type d'image présente alors une allure particulière avec une plus faible valeur des coefficients au centre du masque du filtre spatial : le filtrage a pour effet de moyennner des pixels décorrélés en minimisant l'information redondante. Le Tableau 2 du paragraphe II.1 et la Figure 16 explicitent le lien intrinsèque entre l'allure du filtre spécifié et les connaissances portant sur la formation de l'image sonar. Toutefois, d'autres dégradations (calibration de l'antenne, lobes secondaires de la fonction de directivité,...) peuvent avoir une influence non négligeable sur la dimension théorique de la cellule de résolution. On constate alors que le filtre à spécifier est plus grand.

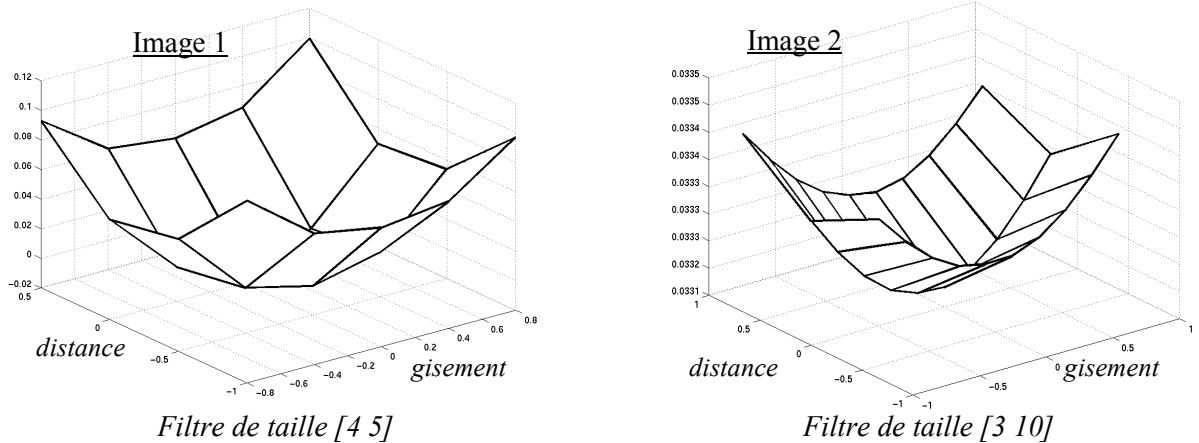


Figure 16 - Lien entre allure du filtre spécifié et principe de formation de l'image sonar

Compte tenu de ces observations, deux traitements peuvent être appliqués.

A pleine résolution, les coefficients du filtre de taille minimale donnée par la distance de corrélation inter-pixels sont optimisés afin de minimiser la variance des pixels en sortie de traitement comme nous l'avons expliqué plus haut.

Une autre solution consiste à décimer l'image de manière à supprimer toute redondance d'information. Par simple filtrage passe-bas de l'image décimée, on obtient alors la bimodalité de l'histogramme.

Notons toutefois que cette technique exige que l'image comporte un objet dont l'ombre portée sur le fond est suffisamment importante (en proportion de l'image). La bimodalité provient en effet du fait que les deux régions n'ont pas les mêmes caractéristiques statistiques (différence du niveau de gris moyen notamment).

Les deux modes de l'histogramme sont facilement identifiables du fait de leur niveau moyen respectif :

- le premier mode est relatif à la zone d'ombre avec de faibles niveaux de gris,
- le second mode est relatif à la zone de fond avec des niveaux de gris plus importants.

A ce stade, la partition des pixels en deux classes est obtenue par un simple **seuillage** avec une valeur du seuil issue de l'analyse de l'histogramme.

IV.3. Mise en œuvre

□ à pleine résolution

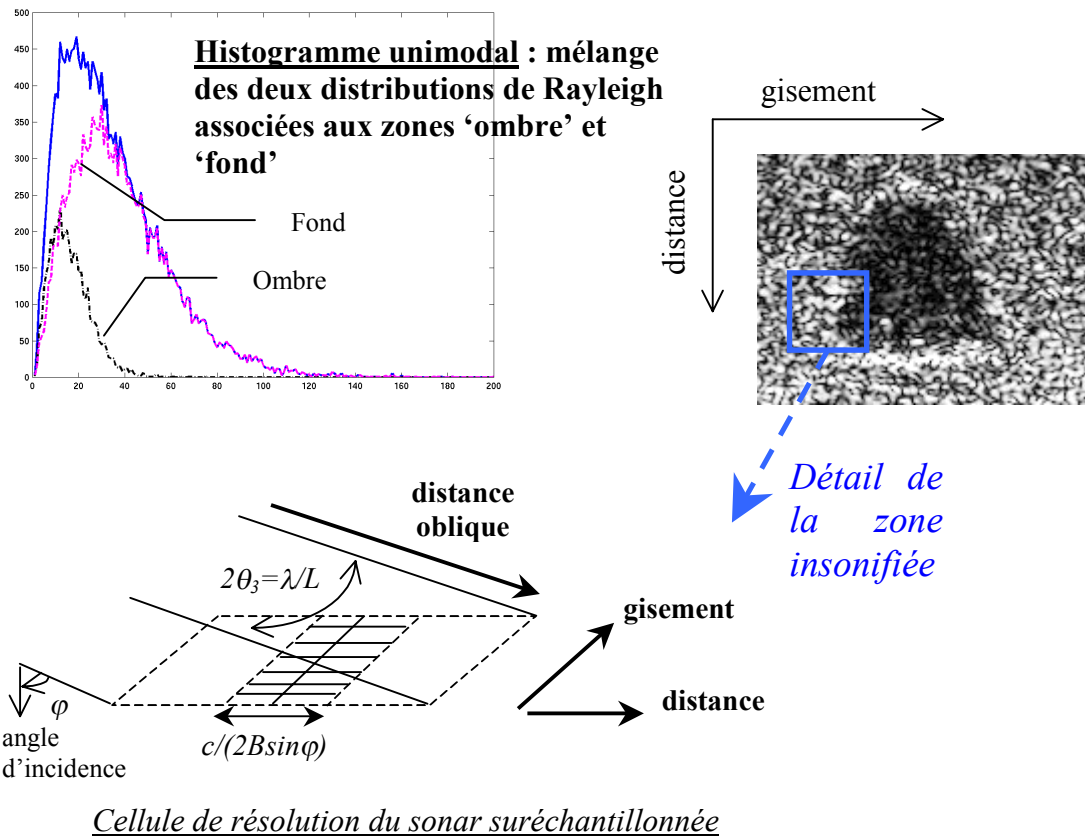
Sachant la distance de corrélation inter-pixels, le filtre optimal au sens de la minimisation de la variance est synthétisé (cf. §IV.2). En filtrant, on modifie la répartition des pixels de l'image de sorte que l'histogramme de l'image filtrée devient bimodal. La sélection du seuil par étude des modes (technique appelée 'mode method' [SAHO-88]) est alors possible pour séparer l'ombre du fond. Le principe est illustré Figure 17.

□ à basse résolution

L'alternative consiste à décimer l'image de manière à décorrélérer les pixels. Parmi l'étendue des filtres passe-bas classiquement utilisés, le filtre gaussien a été choisi car il effectue le compromis entre la minimisation de la variance des niveaux de gris des pixels et la localisation spatiale du contour [MARR-80]. L'ombre segmentée est alors inversement obtenue par duplication. Ce traitement est évidemment plus rapide mais l'image binaire est en contrepartie plus grossière.

Il convient de noter que, basée sur le principe de formation de l'image, cette décimation suffit à décorrélérer les pixels à condition que l'image n'ait pas subi des dégradations importantes. Dans le cas contraire, il faudra décimer davantage de manière à s'assurer de l'indépendance des pixels de la nouvelle image.

La Figure 18 expose brièvement les étapes de chacun des traitements.



*Synthèse du filtre optimal
lié à la corrélation des
pixels*

*Détail de la zone
insonifiée*

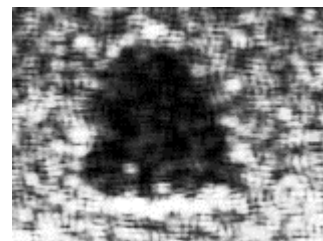
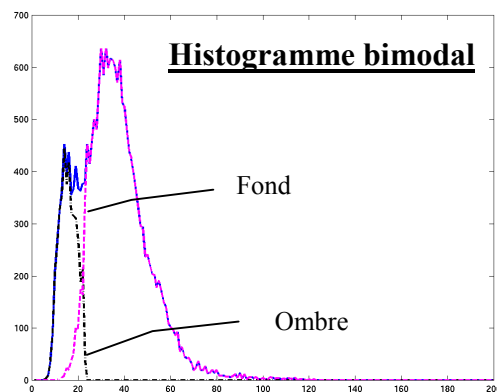
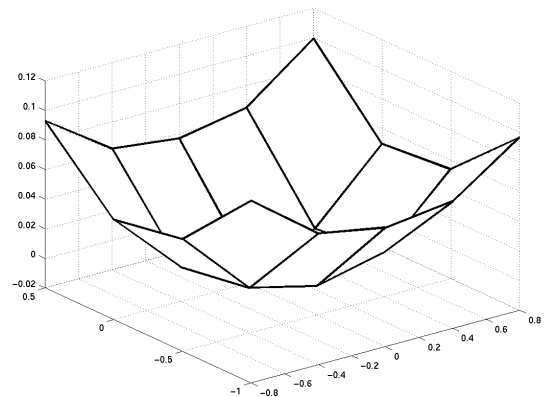
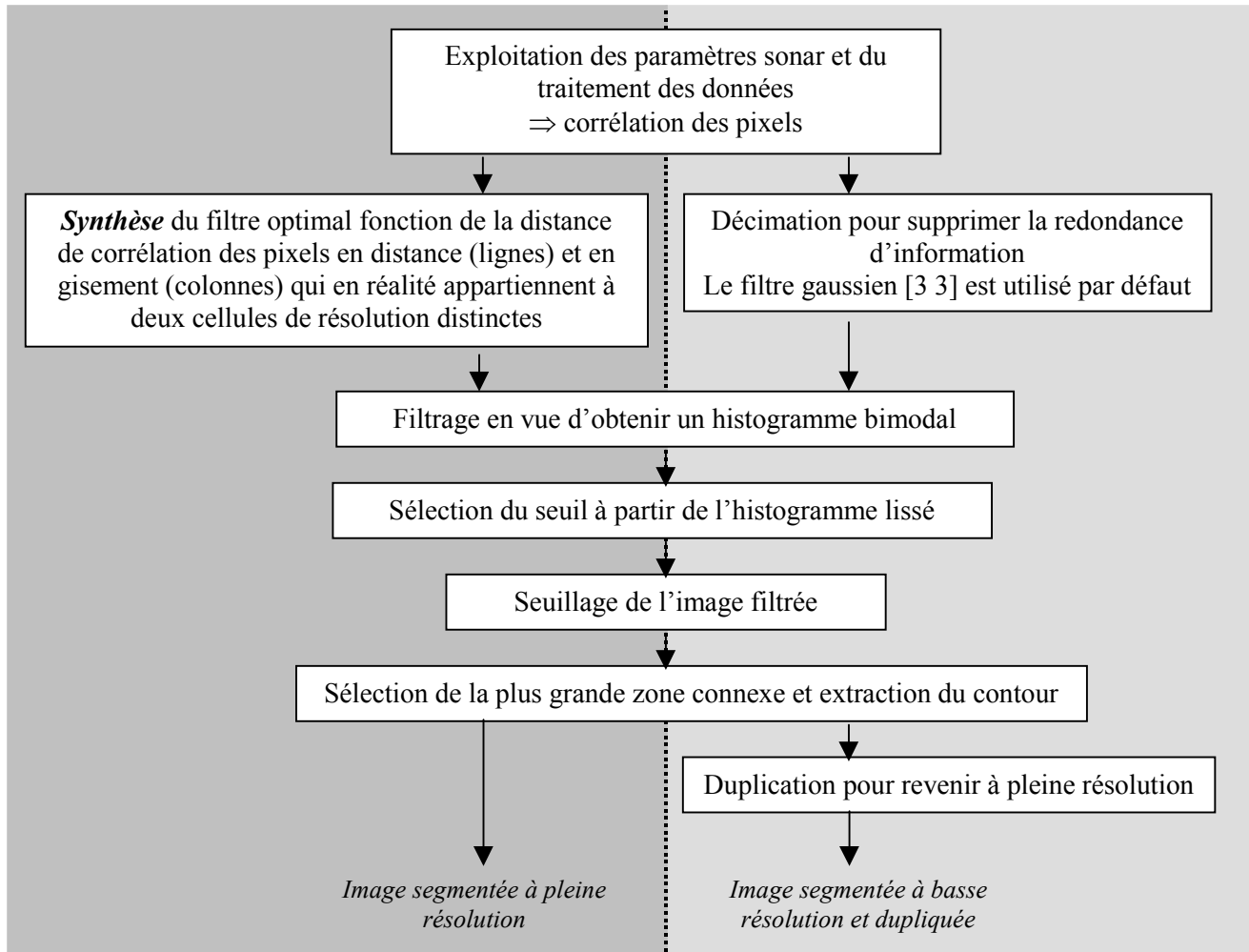


Figure 17 - Principe de la méthode à pleine résolution



**Figure 18 - Principe de chacune des méthodes proposées :
à pleine résolution (colonne de gauche), à basse résolution (colonne de droite)**

IV.4. Résultats expérimentaux

□ **A pleine résolution**

- image simulée (deux zones de Rayleigh – pixels **décorrélés**)

⇒ on utilise le filtre gaussien

Image et histogramme originaux

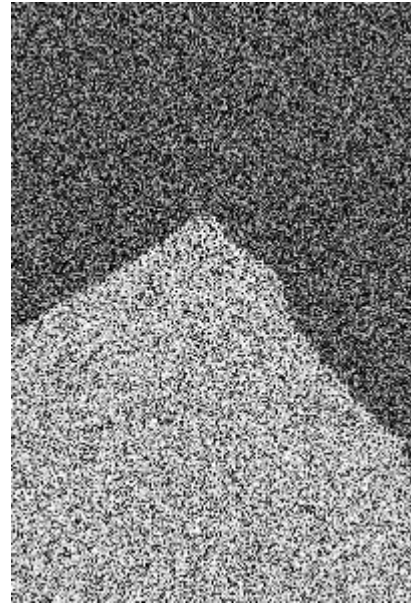
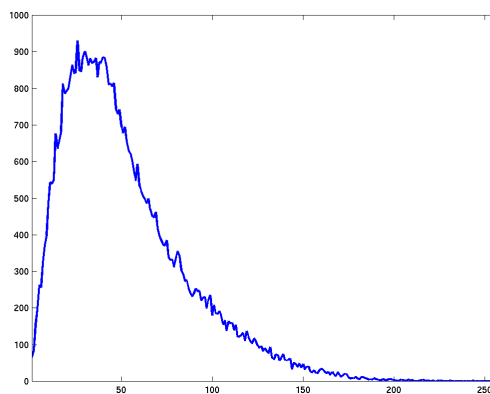
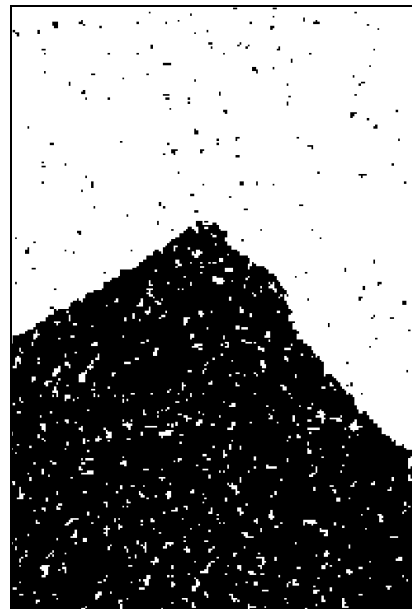
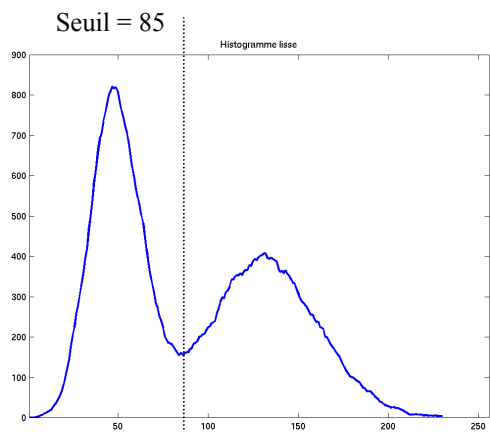
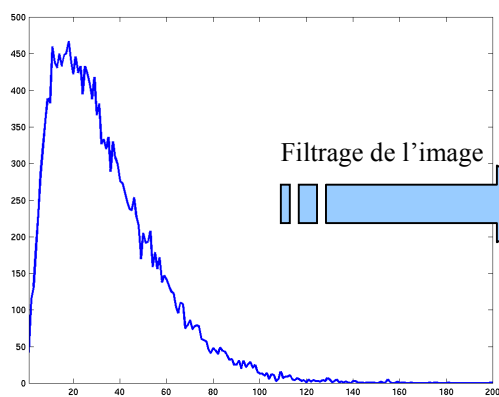


Image et histogramme après filtrage gaussien



- images réelles (pixels **corrélés** conformément aux données du Tableau 2 et de la Figure 16)

Image 1 (mine Manta)



Filtrage de l'image



Seuil = 38

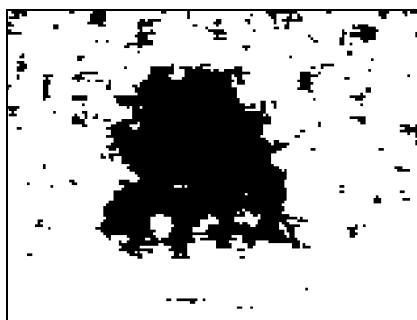
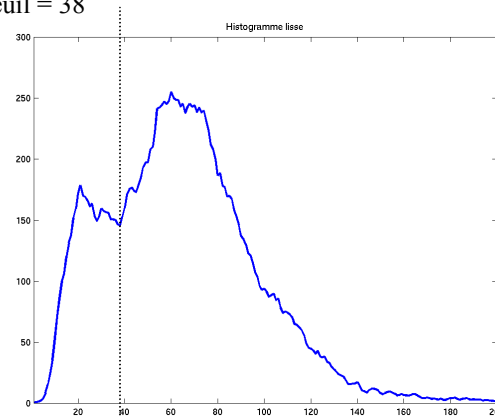
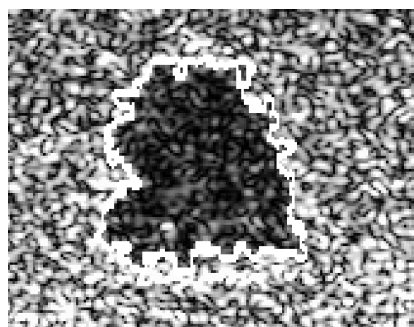
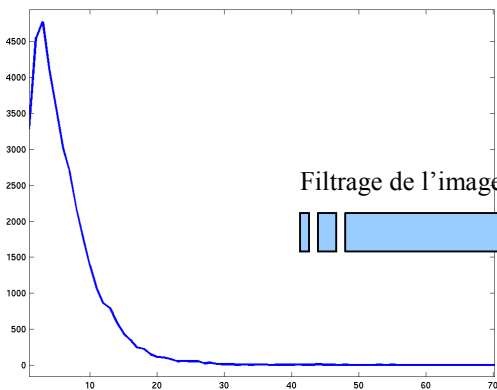


Image filtrée seuillée



Contour superposé à l'image observée

Image 2 (mine sphérique)



Filtrage de l'image



Seuil = 6

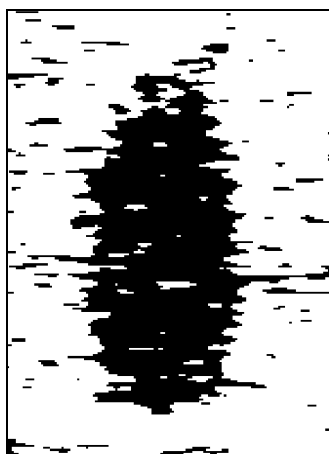
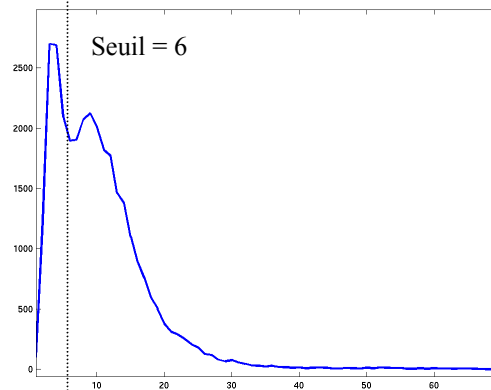
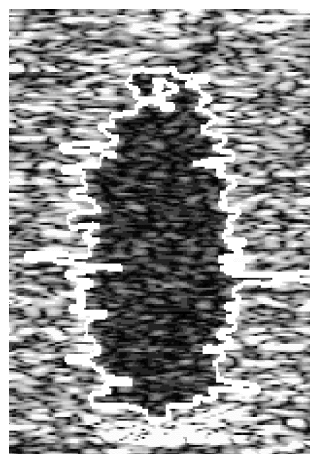


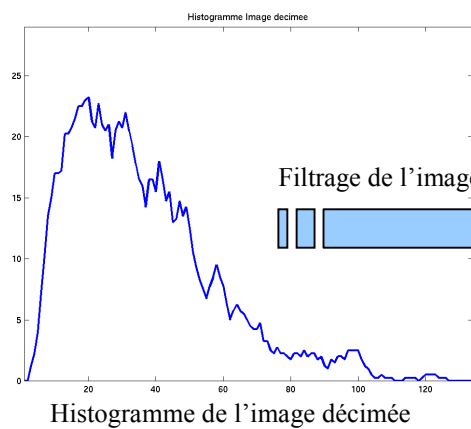
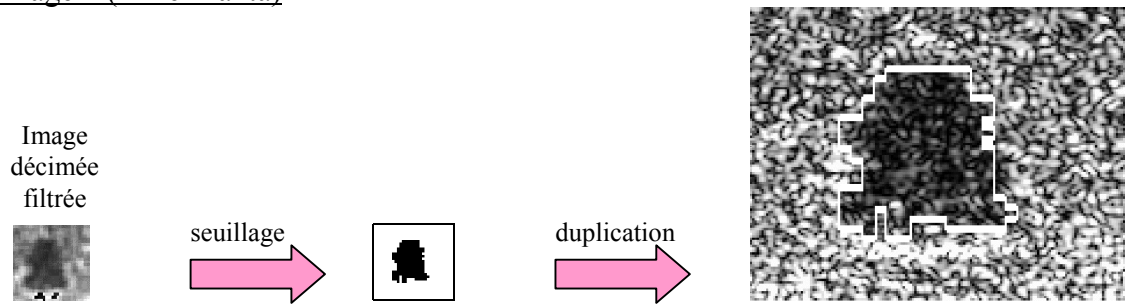
Image filtrée seuillée



Contour superposé à l'image observée

□ **A basse résolution**

Image 1 (mine Manta)



Filtrage de l'image

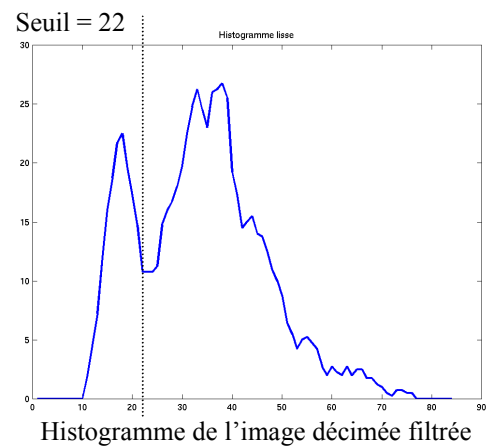
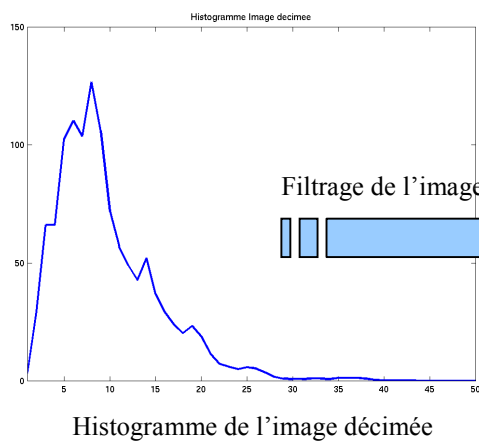
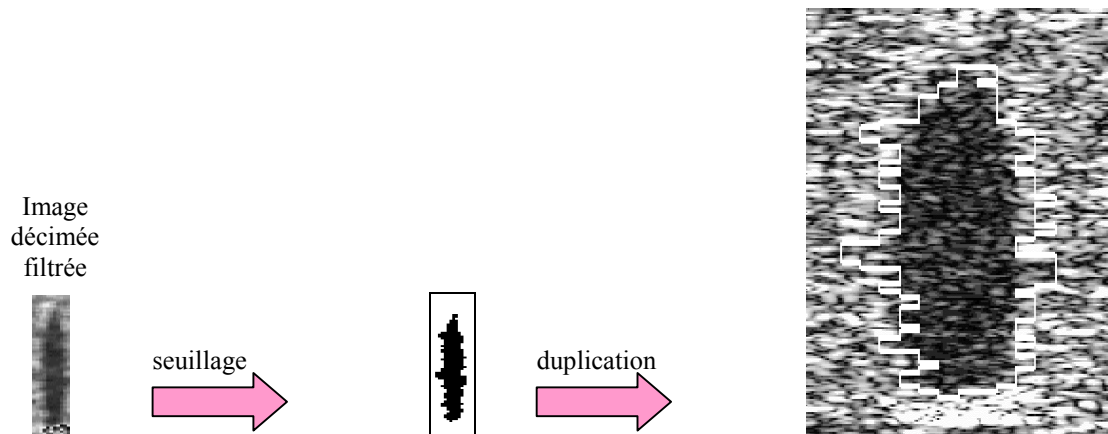
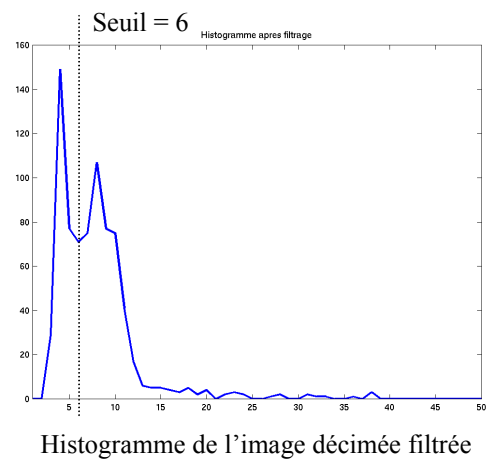


Image 2 (mine sphérique)



Filtrage de l'image



A pleine résolution, le contour extrait partage avec précision les données de l'image observée en deux régions homogènes. Seulement, cette méthode nécessite la synthèse d'un filtre spécifique dont le temps d'optimisation est proportionnel à sa taille (nombre de coefficients) ainsi qu'au format de l'image. Ce problème est cependant de moindre importance dans le cadre d'une classification multi-vues pour laquelle cette optimisation est effectivement faite pour une seule image et le filtre ainsi spécifié identique pour toutes les images de la séquence acquises dans des conditions similaires.

Comme prévu, les contours obtenus par la seconde méthode sont plus grossiers. Néanmoins, pourvu que l'information globale de la silhouette soit préservée (décimation pas trop importante), le calcul d'attributs caractéristiques peut tout à fait être effectué sur cette image binaire après une procédure de lissage.

Conclusion du chapitre II.1

Certaines informations spécifiques au support image traité peuvent être avantageusement introduites dans l'étape de segmentation. Le sonariste connaît en particulier le processus de formation de l'image sonar et notamment ses conséquences sur l'image formée (la nature du bruit, le lien entre cellule de résolution et pixel de l'image, la 'signification' du pixel vis-à-vis de l'objet...).

Les techniques présentées offrent la possibilité de prendre en compte les paramètres sonar et s'adaptent aux conditions d'acquisition de l'image sonar. Les paramètres sonar ont entre autres permis :

- ☞ de régulariser le contour indépendamment de la résolution sonar et de la rascasse
- ☞ d'adapter la taille d'éléments structurants pour les opérations de morphologie mathématique
- ☞ de choisir la taille du filtre à optimiser pour bimodaliser l'histogramme
- ☞ de guider les opérations de décimation ou duplication de l'image sonar

La suite des traitements appliqués à l'image segmentée, c'est-à-dire les caractéristiques des attributs à extraire et la stratégie de classification retenues, conditionnent le choix d'une segmentation appropriée parmi les méthodes que nous avons présentées. En particulier, les remarques suivantes s'imposent :

- A l'inverse des approches 'région' qui laissent des pixels étiquetés réverbération de fond à l'intérieur de l'ombre et inversement, les contours actifs statistiques et la ligne de partage des eaux fournissent par construction une ombre pleine et donc un contour fermé. Cette stratégie évite les fausses alarmes en extrayant la seule information réellement significative et utilisée pour les méthodes d'extraction d'attributs qui suivent. Pour obtenir le même résultat, les autres méthodes exigent de sélectionner la zone connexe la plus importante et d'effectuer un suivi de contour de l'ombre obtenue.
- Pour un traitement massif des données, on aura tendance à privilégier une méthode de segmentation robuste. Ainsi, les méthodes sensibles à l'initialisation (marqueur ou estimation du niveau moyen de la réverbération pour fixer le seuil) sont à éviter.
- La faiblesse de certaines méthodes n'est plus prépondérante si l'on étend le traitement à plusieurs images acquises dans les mêmes conditions (plusieurs vues d'un même objet par exemple). Ainsi, la spécification du filtre pour bimodaliser l'histogramme est réalisée sur une seule de ces images. De même dans une séquence d'images, la segmentation par contours actifs statistiques d'une image peut être initialisée par le résultat de l'image précédente [KASS-88]. Ceci peut permettre de réduire la durée de la phase d'optimisation.

Les précédentes remarques ont mené au choix suivant. La méthode de segmentation par **seuillage à partir de l'histogramme bimodal** a été retenue pour **l'application multi-vues** souhaitée. Ce traitement a l'avantage de s'appliquer massivement sur un ensemble d'images simulées dans des conditions similaires sans devoir réestimer les coefficients du filtre à appliquer. Il en résulte une étape de segmentation rapide avec une opération principale de filtrage spatial de l'image (ne nécessite ni optimisation, ni opération morphologique coûteuses en temps de calcul).

Chapitre II.2 : Extraction d'attributs caractéristiques

I. Introduction du chapitre II.2

La qualité de l'image résultant de la précédente étape de segmentation et notamment l'ensemble des pixels représentatifs de l'ombre portée constituent l'information utile en terme de caractérisation de l'ombre. Le problème de la **classification sur ombre** sous-tend la **reconnaissance d'une forme** représentée par une fonction f appartenant à un ensemble F des formes possibles. La notion de *forme* englobe la région intérieure de l'ombre segmentée et son contour, i.e. un nombre important de données élémentaires ou pixels qu'il convient de réduire. Le calcul d'un certain nombre d'**attributs caractéristiques** ou critères discriminants intervient de manière à faciliter l'étape finale de classification. De nombreux travaux en traitement d'images visent à améliorer la caractérisation et la discrimination d'une forme. Notre objectif est de rappeler les principaux paramètres rencontrés dans la littérature en justifiant leur utilisation dans le cadre spécifique de l'imagerie sonar.

Divers problèmes relèvent spécifiquement du contexte opératoire :

- Difficultés inhérentes aux conditions d'acquisition (bruit, distorsions géométriques...)
- Apparence changeante des objets suivant l'axe de visée
- Nouvelles mines à géométrie complexe dont l'ombre projetée est très différente selon les points de vue (la mine suédoise Rockan en est un exemple)

Nous distinguons deux types d'**attributs**. D'une part, les attributs descriptifs proviennent du codage de la forme ou de son contour. L'information ainsi transcodée permet de reconstruire la silhouette idéalement sans perte. De nombreux travaux portent sur les transformations mathématiques réversibles des points de contour [ZHAN-72] [PERS-86] [HENO-93] [GHOR-94]. D'autre part, les attributs synthétiques reflètent certaines propriétés géométriques et/ou statistiques de l'image. Deux aspects possibles : la caractérisation globale de l'ombre par des attributs intrinsèques visant à remonter à la forme de l'objet dont on observe l'ombre portée ou la recherche de paramètres invariants aux transformations du plan [HU-62] [TEAG-80] [JAN-87] [REEV-88] [ARBT-89] [KHOT-90] [PROK-92] [PEJN-92] [FLUS-93].

Compte tenu des formes parfois diverses de l'ombre en fonction des points de vue, il semble également intéressant de caractériser la forme de l'ombre par des grandeurs **invariantes vis-à-vis des transformations du plan**. Les transformations géométriques de base généralement prises en compte, soient la translation, l'homothétie et la rotation, appartiennent au groupe des similitudes. Dans le cas de l'imagerie sonar qui consiste à effectuer une projection perspective, les déformations de l'ombre portée de l'objet visé par le sonar sont plus généralement assimilées à des transformations affines du contour. Des

distorsions caractéristiques de l'imagerie sonar, à savoir élongation et biais, sont exclues du groupe des similitudes mais prises en compte par la transformation affine définie comme suit :

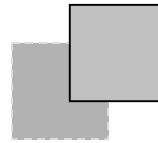
$$\begin{cases} u = a_0 + a_1x + a_2y \\ v = b_0 + b_1x + b_2y \end{cases} \text{ ou encore } \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} + t$$

Suivant les valeurs des coefficients $(a_i, b_i)_{i=0...2}$, elle regroupe l'ensemble des transformations suivantes que l'on illustre sur un carré prototype :



- D'une part, la translation au point (a_0, b_0) :

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}$$



- D'autre part, les autres transformations :

Elongation d'amplitude δ dans la première direction

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \delta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



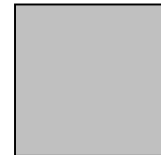
Elongation d'amplitude δ dans la deuxième direction

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



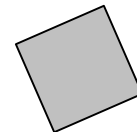
Homothétie d'amplitude ω

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \omega & 0 \\ 0 & \omega \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



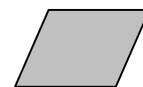
Rotation d'angle θ

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



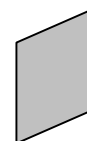
Biais d'amplitude l dans la première direction

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & l \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



Biais d'amplitude l dans la deuxième direction

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ l & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



Le biais est visible dans le cas d'objets n'ayant pas l'intéressante propriété de symétrie radiale tels que le cylindre par exemple. Les images de la Figure 19 illustrent bien le biais apparaissant entre deux prises de vue du même cylindre dans les mêmes conditions mais sous deux points de vue distincts, i.e. un écart de 60 degrés entre les prises.

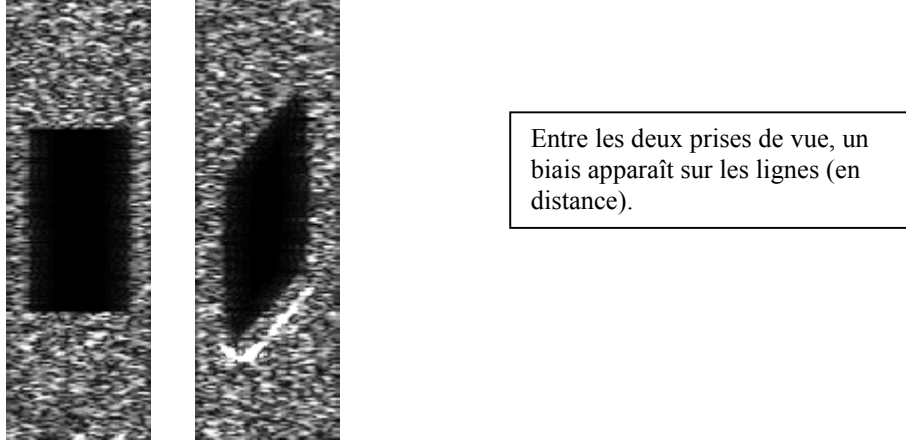


Figure 19 – Illustration du biais entre deux prises de vue du même objet

Dans notre étude, la recherche **d'invariants affines** semble donc justifiée. La notion d'invariance est à ce jour largement discutée dans la littérature. Soit c , la classe désirée de f , le problème de **reconnaissance de formes par invariants** consiste à chercher des grandeurs caractéristiques $A = \{a_i\}$ de la forme f , invariantes sous l'action d'un groupe de transformations G via une application Ω [GHOR-90]. Deux cas se présentent [WOOD-96] :

1. $a_i(\Omega f) = a_i(f) \quad \forall \Omega \in G, f \in F$ a_i est alors un *invariant*
2. $a_i(\Omega f) = \mu(\Omega) a_i(f) \quad \forall \Omega \in G, f \in F$ a_i est alors un *invariant relatif*

La recherche d'invariants vise à représenter les images dans un nouvel espace dans lequel la représentation de ces mêmes images consiste en un certain nombre de caractéristiques invariantes vis-à-vis des transformations du plan. Ces caractéristiques doivent former une famille complète afin d'éviter le cas où deux objets de formes différentes f et f^0 seraient caractérisés par les mêmes invariants. En somme, un ensemble $A = \{a_i\}$ de nombres réels ou complexes est appelé ensemble d'invariants si on a l'implication :

$$\forall k : a_i = a_i^0 \Leftarrow f \text{ et } f^0 \text{ se réfèrent à la même forme}$$

et cet ensemble est complet si on a l'équivalence :

$$\forall k : a_i = a_i^0 \Leftrightarrow f \text{ et } f^0 \text{ se réfèrent à la même forme}$$

Deux ensembles d'invariants affines nous ont intéressés : les invariants algébriques introduits par J. Flusser et T. Suk et les invariants de type intégrale de K. Arbter basés sur les descripteurs de Fourier [ARBT-89] [FLUS-93]. Il s'agit de deux approches différentes, l'une par région, l'autre par contour. La première apparaît moins sensible aux artéfacts qui affectent le contour (pourvu que la forme soit par ailleurs bien résolue) puisque les moments prennent en compte la totalité des pixels de la forme à classifier (pixels de la frontière et pixels intérieurs).

Pour cette étape d'extraction d'attributs, deux solutions sont envisagées :

- ☞ effectuer une normalisation de l'image de manière à rendre le calcul des attributs plus robuste
- ☞ calculer des attributs invariants vis-à-vis des transformations affines du plan (invariants affines).

II. Prétraitements

Si aucune invariance n'est intrinsèquement prise en compte dans le calcul des attributs, une étape préliminaire de **normalisation de l'image** est nécessaire de manière à limiter la dispersion intra-classe [JAN-87]. D'autre part, la plupart des attributs gagnent en robustesse s'ils sont calculés sur un **contour lissé** de manière à supprimer les aspérités dues au bruit de speckle. Le lissage s'avère notamment nécessaire après certaines méthodes de segmentation sans régularisation.

II.1. Normalisation de l'image sonar

Quelles que soient la hauteur de l'objet et sa distance au sonar, la forme de l'ombre doit permettre de discriminer l'objet. Afin de s'affranchir de la géométrie d'insonification, les images sont normalisées (cf. Figure 20).

La normalisation vise à pallier les fortes distorsions inhérentes à la prise de vue du sonar, c'est-à-dire la faible rasance et la faible précision en gisement entraînant des ombres généralement disproportionnées. Elle opère donc sur deux fronts conjoints :

1) rendre les pixels de l'image représentatifs de la surface d'insonification réelle :

Il existe une distorsion entre la taille du pixel (habituellement carré) de l'image et l'aire effective qu'il représente sur le sol. Ce phénomène provient du principe d'acquisition des systèmes sonar qui, pour des raisons technologiques, amènent à une plus grande précision en distance qu'en gisement. La solution proposée est de dupliquer les pixels le long de la dimension la moins résolue de l'image.

2) normalisation en rasance de façon à ce que la longueur des ombres soit indépendante de celle-ci :

L'angle choisi pour la normalisation en rasance doit permettre de comparer facilement des formes géométriques associées au même type d'objet en apportant de la robustesse à la classification par critères topologiques par exemple. La rasance choisie de 45 degrés entraîne une longueur de l'ombre visible sur l'image sonar (suivant la distance oblique) égale à la hauteur de l'objet multipliée par $\sqrt{2}$ [JAN-87]. Ainsi le contour de l'ombre associée à une sphère est une ellipse dont le petit axe est égal au diamètre de la sphère et le grand axe, le diamètre de la sphère que multiplie $\sqrt{2}$.

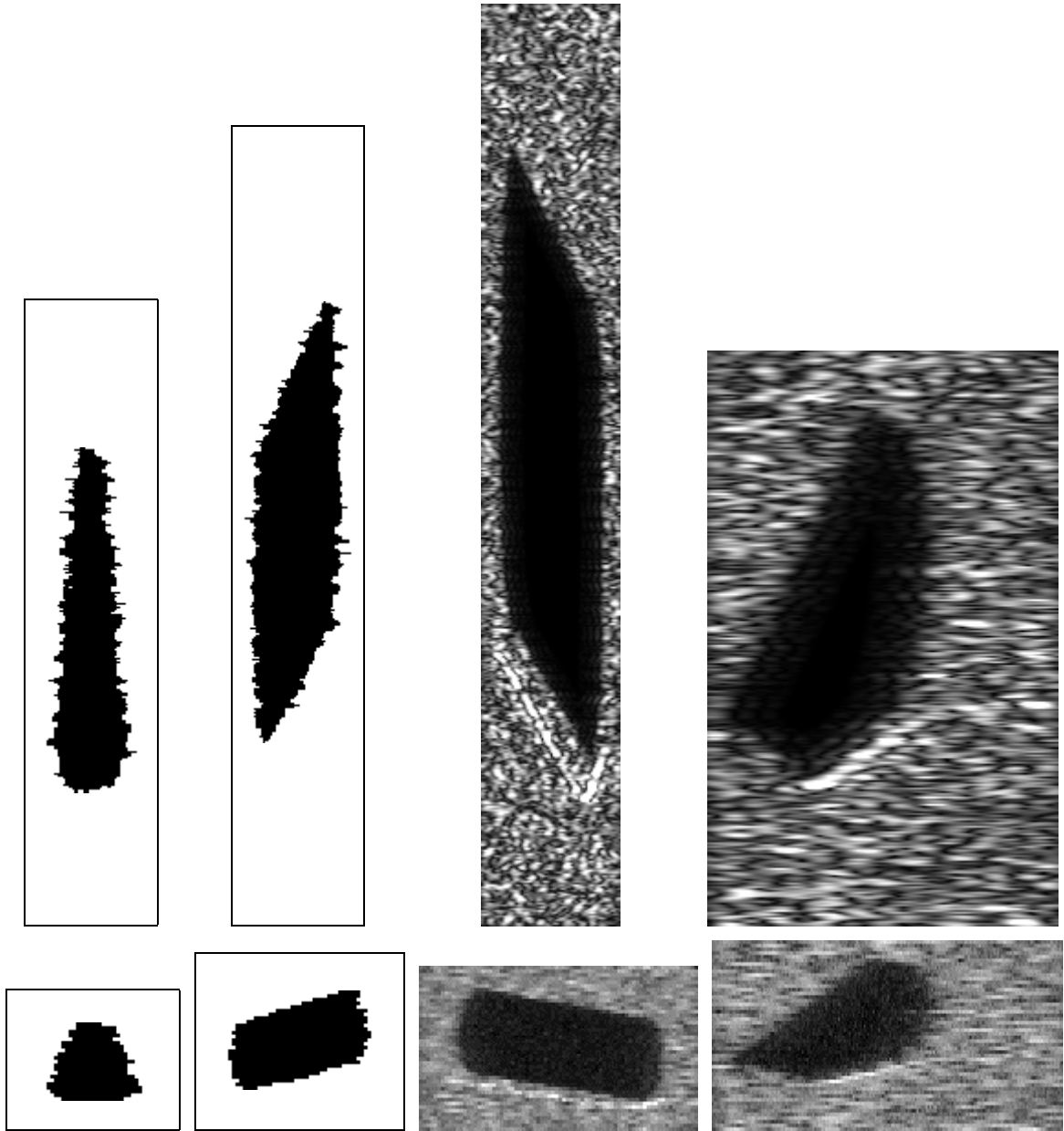


Figure 20 – Normalisation de l'image d'images binaires et brutes

La Figure 21 permet d'exprimer facilement les relations suivantes :

$$\sin \theta_0 = \frac{H_0}{D_0 + L_0} \text{ et } \sin \theta_1 = \frac{H_1}{D_1 + L_1},$$

$$\text{de même, } \frac{H_0}{h} = \frac{D_0 + L_0}{L_0} \text{ et } \frac{H_1}{h} = \frac{D_1 + L_1}{L_1},$$

$$\text{d'où, par combinaison, } \frac{\sin \theta_0}{\sin \theta_1} = \frac{H_0}{H_1} \frac{D_1 + L_1}{D_0 + L_0} = \frac{L_1}{L_0},$$

i.e. la distance normalisée sous rasance θ_0 , $L_0 = L_1 \frac{1}{\sin \theta_0} \frac{H_1}{D_1 + L_1}$.

Pour normaliser l'image à la rasance θ_0 , il revient donc à :

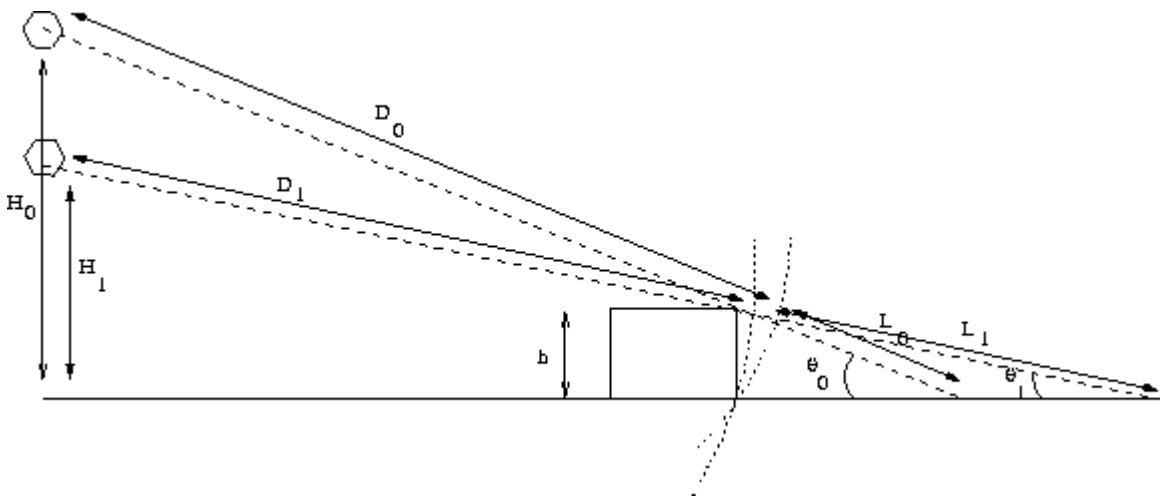
- regrouper les pixels en distance si $L_0 < L_1$ ($\theta_0 > \theta_1$)
- dupliquer les pixels en distance si $L_0 > L_1$ ($\theta_0 < \theta_1$)

En cas de regroupement, le niveau de gris affecté au pixel est la valeur médiane des niveaux de gris des pixels regroupés.

La taille du pixel dépend initialement des paramètres sonar mais est modifiée à l'étape de normalisation. La taille équivalente du pixel, identique suivant les deux axes, dépend ainsi à la fois des paramètres sonar et de la normalisation de l'image en rasance.

Remarques :

- ☞ Ce prétraitement est fait sous l'hypothèse d'un fond horizontal
- ☞ Les points 1) et 2) de la normalisation seront traités simultanément si la duplication nécessaire en 1) est effectuée en distance.



avec

- L_i : longueur de l'ombre sur le fond,
- D_i : distance oblique du sonar à l'objet,
- θ_i : angle de rasance,
- h : hauteur de l'objet,
- H_i : l'altitude du sonar,

Figure 21 – Normalisation en rasance

II.2. Lissage du contour

Le lissage peut s'obtenir de différentes manières (opérations morphologiques, décomposition du contour de l'ombre en série de Fourier...). Nous choisissons de reconstruire le contour à partir des $K=10$ premiers descripteurs de Fourier (positifs et négatifs, i.e. 20 au total) (cf. annexe D). On s'affranchit ainsi des détails du contour indésirables tout en conservant la forme globale utile pour la reconnaissance de formes (cf. Figure 22). Cette perte de détails peut être utile dans certains contextes comme nous le verrons plus loin.

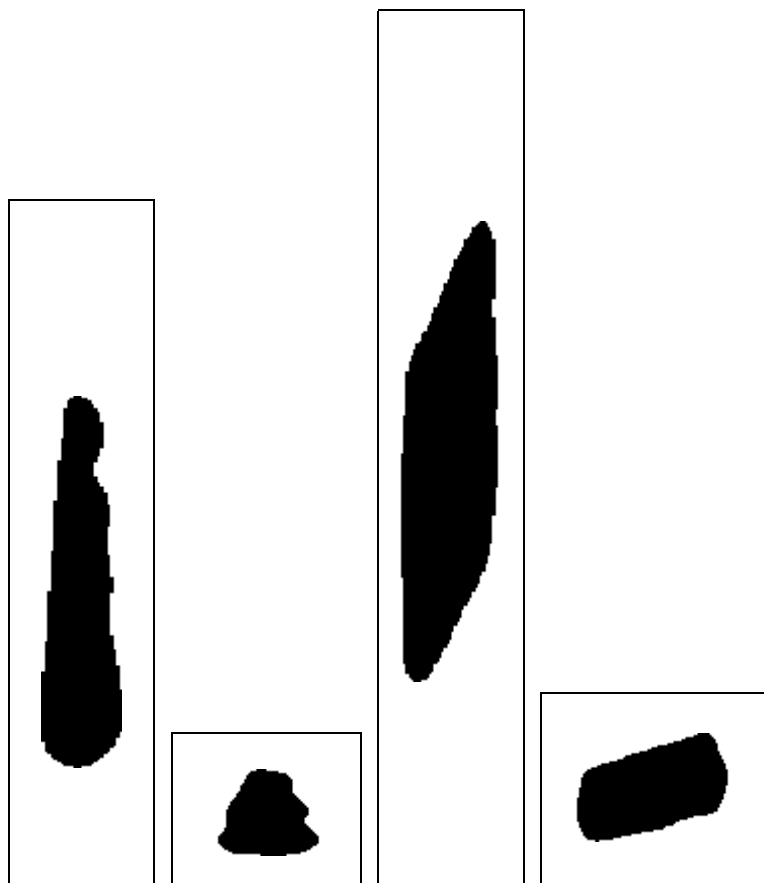


Figure 22 – Lissage de contour sur les images de la Figure 20

III. Attributs descriptifs

L'annexe D constitue un bref rappel de la décomposition en série de Fourier et de ses propriétés. De nombreux auteurs ont fait état de ces propriétés et proposé des méthodes de normalisation pour la reconnaissance de formes : reconnaissance de caractères manuscrits [GRAN-72] ou reconnaissance d'objets 3D à partir de leur projection 2D [WALL-80] [MITC-83] [HENO-93]. Généralement, la méthode de normalisation en échelle est commune et consiste simplement à diviser les descripteurs par le descripteur d'indice 1. Les méthodes de normalisation diffèrent toutefois en ce qui concerne la rotation et le point de départ.

A cette décomposition classique effectuée sur les coordonnées des points de contour représentées en nombres complexes $z(i) = z_i = x_i + jy_i$, nous opposons une double décomposition effectuée distinctement sur chacune des coordonnées x_i et y_i , mieux adaptée au principe de formation des images sonar. Une normalisation de ces descripteurs associée à une normalisation en rance de l'image permet de s'affranchir des principales transformations géométriques du plan.

Par les AW-invariants, on aborde ensuite le problème plus général de l'invariance vis-à-vis des transformations affines qui, en englobant les similitudes, prennent également en compte l'élongation et le biais.

Notons qu'un jeu de descripteurs invariants peut également être dérivé de la transformée de Fourier-Mellin analytique [GHORB-94]. Cette transformée s'applique cependant à une fonction exprimée en coordonnées polaires et s'avère difficile à mettre en œuvre sur des images [CARI-99].

III.1. Décomposition en série de Fourier des coordonnées des points de contour

Selon le principe de formation des images sonar, la projection de l'ombre sur le fond est représentative des positions relatives du sonar et de l'objet. Pourvu que l'image soit normalisée selon le paragraphe II.1, l'orientation du contour et de ses axes principaux codés dans les descripteurs fondamentaux est un critère discriminant à conserver. La décomposition du contour en série de Fourier effectuée distinctement sur chacune des coordonnées des points de contour est alors mieux adaptée à notre contexte. Elle permet en effet de conserver la double information sur les lignes et les colonnes correspondant à l'acquisition temporelle (distance) et angulaire (gisement) des données.

III.1.1. Définition

La transformée de Fourier est dans ce cas appliquée aux abscisses et aux ordonnées :

$$\begin{cases} X_k = \frac{1}{N} \sum_{i=0}^{N-1} x_i \exp\left(-j2\pi \frac{ki}{N}\right) \\ Y_k = \frac{1}{N} \sum_{i=0}^{N-1} y_i \exp\left(-j2\pi \frac{ki}{N}\right) \end{cases}$$

Remarque : $C_k = X_k + jY_k$ où C_k est défini dans l'annexe D

III.1.2. Propriétés

Outre les propriétés générales exposées dans l'annexe D pour la décomposition classique, cette double décomposition possède les intéressantes propriétés suivantes.

Les abscisses et les ordonnées étant réelles, leur transformée de Fourier est nécessairement symétrique, i.e. $X_{-k} = X_k^*$ et $Y_{-k} = Y_k^*$.

Le couple 'point moyen' (X_0, Y_0) correspond aux coordonnées moyennes des points de contour de la silhouette. Calculons les descripteurs des mêmes données centrées, c'est-à-dire lorsque le système d'axes est translaté au point moyen de la silhouette (on raisonne sur X_k) :

$$X_k^c = \frac{1}{N} \sum_{i=1}^N (x_i - X_0) \exp\left(-j2\pi \frac{ki}{N}\right) = \frac{1}{N} \sum_{i=1}^N x_i \exp\left(-j2\pi \frac{ki}{N}\right) - \frac{1}{N} \sum_{i=1}^N X_0 \exp\left(-j2\pi \frac{ki}{N}\right)$$

$$X_k^c = X_k - \frac{X_0}{N} \sum_{i=1}^N \exp\left(-j2\pi \frac{ki}{N}\right)$$

$$\text{soit } X_k^c = X_k - X_0 \frac{1 - \exp(-j2\pi k)}{1 - \exp(-j2\pi \frac{k}{N})} = X_k, \forall k \neq 0 \quad \text{et} \quad X_0^c = 0$$

En somme, l'information 'contour' de la silhouette est réellement contenue dans les descripteurs non nuls. Le point moyen (X_0, Y_0) permet uniquement de localiser la silhouette. La reconstruction s'effectue alors autour de ce point.

Ces deux points montrent que la seule connaissance des descripteurs d'indice strictement positif suffit donc à caractériser complètement le contour.

Cette décomposition permet de disposer d'une description de la silhouette selon les deux axes de l'image. Nous avons ainsi une caractérisation de l'ombre en distance (lignes) et en gisement (colonnes). Elle permet ainsi de conserver une information double cohérente avec les transformations géométriques classiquement observées en imagerie sonar (biais et/ou fortes elongations le long de l'axe des distances). A la nature propre de la forme s'ajoute aussi une information supplémentaire sur l'orientation des axes principaux caractéristiques de l'objet ciblé et de sa position vis-à-vis du sonar.

Une transformation affine des points de contour est enfin automatiquement retranscrite dans le domaine fréquentiel comme l'explique les lignes suivantes.

Soit la transformation affine suivante :

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = A \times \begin{bmatrix} x_i \\ y_i \end{bmatrix} + t = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix}$$

en passant dans le domaine de Fourier, on a

- pour $k \neq 0$,

$$\begin{cases} U_k = \frac{1}{N} \sum_{i=0}^{N-1} u_i \exp\left(-j2\pi \frac{ki}{N}\right) = \frac{1}{N} \sum_{i=0}^{N-1} (ax_i + by_i + e) \exp\left(-j2\pi \frac{ki}{N}\right) = aX_k + bY_k \\ V_k = \frac{1}{N} \sum_{i=0}^{N-1} v_i \exp\left(-j2\pi \frac{ki}{N}\right) = \frac{1}{N} \sum_{i=0}^{N-1} (cx_i + dy_i + f) \exp\left(-j2\pi \frac{ki}{N}\right) = cX_k + dY_k \end{cases}$$

$$\text{soit } \boxed{\begin{bmatrix} U_k \\ V_k \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} X_k \\ Y_k \end{bmatrix} = A \times \begin{bmatrix} X_k \\ Y_k \end{bmatrix}, \forall k \neq 0}$$

• pour $k=0$,

$$\begin{cases} U_0 = \frac{1}{N} \sum_{i=0}^{N-1} u_i = \frac{1}{N} \sum_{i=0}^{N-1} (ax_i + by_i + e) = aX_0 + bY_0 + e \\ V_0 = \frac{1}{N} \sum_{i=0}^{N-1} v_i = \frac{1}{N} \sum_{i=0}^{N-1} (cx_i + dy_i + f) = cX_0 + dY_0 + f \end{cases}$$

$$\text{soit } \boxed{\begin{bmatrix} U_0 \\ V_0 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} = A \times \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} + t}$$

La décomposition en série de Fourier des coordonnées est donc **linéaire**.

III.1.3. Normalisation

L'invariance par rapport au point de départ est obtenue à partir des descripteurs C_k . Arbter montre qu'on s'affranchit de la dépendance en τ (cf. formule (2) dans l'annexe D) en formant les invariants suivants [ARBT-89] :

$$\tilde{C}_k = |C_k| \Phi_k^{\lambda_0} \prod_{i=1}^n \Phi_{k_i}^{k_i \lambda_i}, \quad k_i = \pm 1, \pm 2, \dots \text{ avec } \Phi_k = \frac{C_k}{|C_k|}, \quad \Phi_{k_i} = \frac{C_{k_i}}{|C_{k_i}|} \text{ et } \lambda_0 + \sum_{i=1}^n k_i \lambda_i = 0, \quad \lambda_i \in \mathbb{Z}.$$

Soit en prenant $n=1$, $\tilde{C}_k = |C_k| \Phi_k^{\lambda_0} \Phi_{k_1}^{k_1 \lambda_1}$.

Si on prend $\lambda_0 = 1$, alors la contrainte $1 + k_1 \lambda_1 = 0$ revient par exemple à prendre $k_1 = 1$ et $\lambda_1 = -1$.

Il vient :

$$\tilde{C}_k = |C_k| \Phi_k \Phi_1^{-k} = |C_k| \frac{C_k}{|C_k|} \frac{C_1^{-k}}{|C_1|^{-k}} \quad \text{soit} \quad \boxed{\tilde{C}_k = C_k \frac{C_1^{-k}}{|C_1|^{-k}}}$$

Preuve

A titre de vérification, remplaçons C_k par $C_k^0 \tau^k$ et vérifions que $\tilde{C}_k = \tilde{C}_k^0$,

$$\tilde{C}_k = C_k \left(\frac{C_1}{|C_1|} \right)^{-k} = C_k^0 \tau^k \left(\frac{C_1^0 \tau^1}{|C_1^0|} \right)^{-k} = C_k^0 \left(\frac{C_1^0}{|C_1^0|} \right)^{-k} = \tilde{C}_k^0$$

Remarque : l'invariance par rapport au point de départ présentée ci-dessus utilise le descripteur de plus grand module. Il s'agit de C_l dans le cas où le contour est parcouru dans le sens inverse des aiguilles d'une montre, C_{-l} dans le cas contraire (cas de la paramétrisation

décrite dans l'annexe C). Notez que si le module de C_{-l} est plus grand que celui de C_l , le module de C_{-l} étant alors plus important, il vaut mieux prendre $k_1 = -1$ et $\lambda_1 = 1$.

On a alors $\boxed{\tilde{C}_k = C_k \frac{C_{-1}^k}{|C_{-1}|^k}}$.

De $C_k = X_k + jY_k$, on déduit les descripteurs normalisés pour les deux coordonnées :

$$\begin{cases} \tilde{X}_k = X_k \frac{C_1^{-k}}{|C_1|^{-k}} = X_k \frac{(X_1 + jY_1)^{-k}}{|X_1 + jY_1|^{-k}} \\ \tilde{Y}_k = Y_k \frac{C_1^{-k}}{|C_1|^{-k}} = Y_k \frac{(X_1 + jY_1)^{-k}}{|X_1 + jY_1|^{-k}} \end{cases}$$

Remarque : $\tilde{X}_0 = X_0$ et $\tilde{Y}_0 = Y_0$, i.e. aucune modification du point moyen

D'autre part, un changement d'échelle agit sur ces descripteurs normalisés comme suit :

$$\begin{aligned} \text{Homothétie} \quad z'_i = \rho z_i & \Leftrightarrow C'_k = \rho C_k \\ & \Leftrightarrow \tilde{C}'_k = \rho C_k \left(\frac{\rho C_1}{|\rho C_1|} \right)^{-k} = \rho C_k \left(\frac{C_1}{|C_1|} \right)^{-k} = \rho \tilde{C}_k \end{aligned}$$

L'**invariance en échelle** peut donc être obtenue en divisant chaque descripteur normalisé par le module de $\tilde{C}_1$ (resp. $\tilde{C}_{-1}$ si $|C_{-l}| > |C_l|$) donc sans modification de phase,

$$\tilde{\tilde{C}}_k = \frac{\tilde{C}_k}{|\tilde{C}_1|} \text{ (resp. } \tilde{\tilde{C}}_k = \frac{\tilde{C}_k}{|\tilde{C}_{-1}|} \text{)}$$

En somme, les descripteurs normalisés sont calculés en 4 temps :

- 1- transformée de Fourier des coordonnées des points du contour X_k et Y_k , $k=1, \dots, K$
- 2- normalisation par rapport au point de départ sachant $C_k = X_k + jY_k$,

$$\tilde{X}_k = X_k \frac{C_1^{-k}}{|C_1|^{-k}} \text{ et } \tilde{Y}_k = Y_k \frac{C_1^{-k}}{|C_1|^{-k}}, \text{ (resp. } \tilde{X}_k = X_k \frac{C_{-1}^k}{|C_{-1}|^k} \text{ et } \tilde{Y}_k = Y_k \frac{C_{-1}^k}{|C_{-1}|^k}, \text{ si } |C_{-l}| > |C_l| \text{)}$$

- 3- normalisation en échelle,

$$\tilde{\tilde{X}}_k = \frac{\tilde{X}_k}{|\tilde{C}_1|} \text{ et } \tilde{\tilde{Y}}_k = \frac{\tilde{Y}_k}{|\tilde{C}_1|}, \text{ (resp. } \tilde{\tilde{X}}_k = \frac{\tilde{X}_k}{|\tilde{C}_{-1}|} \text{ et } \tilde{\tilde{Y}}_k = \frac{\tilde{Y}_k}{|\tilde{C}_{-1}|} \text{)}$$

- 4- nous travaillons avec des grandeurs réelles, les descripteurs d'indices négatifs se déduisent donc simplement des descripteurs d'indices positifs, de sorte que

$$\tilde{\tilde{X}}_{-k} = \tilde{\tilde{X}}_k^* \text{ et } \tilde{\tilde{Y}}_{-k} = \tilde{\tilde{Y}}_k^*$$

III.2. Les descripteurs de Fourier AW - invariants de Arbter [ARBT-89]

Suivant l'application que l'on souhaite, la seule normalisation en échelle proposée ci-dessus peut suffire. En matière de reconnaissance de formes par invariance en revanche, on recherche plutôt des invariants vis-à-vis des transformations géométriques du plan. En ce sens, Arbter a introduit des descripteurs de Fourier invariants vis-à-vis des transformations affines du contour. Ces descripteurs sont calculés séparément sur les abscisses et les ordonnées des points de contour puis normalisés en deux temps. Dans un premier temps, on construit des invariants vis-à-vis des transformations affines : les A-invariants. Puis, dans un deuxième temps, on les rend invariants vis-à-vis du décalage à la paramétrisation du contour : les AW-invariants.

III.2.1. Décomposition en série de Fourier et transformation affine

NB : Le raisonnement est mené dans le cas continu par souci de clarté (expressions plus simples).

Soit C , une courbe fermée et $\gamma(t) = (x(t), y(t))^T$ paramétrée par t , la représentation paramétrique de C , où t est nul au point de départ et égal à T si la courbe est entièrement parcourue. La décomposition en série de Fourier effectuée distinctement sur les abscisses et les ordonnées donne :

$$\begin{cases} X_k = \frac{1}{T} \oint_C x(t) \exp(-j2\pi kt / T) dt \\ Y_k = \frac{1}{T} \oint_C y(t) \exp(-j2\pi kt / T) dt \end{cases}$$

L'invariance en translation est assurée en écartant les coefficients $[X_0, Y_0]$ qui ne contiennent aucune information sur la forme.

Comme nous l'avons montré au paragraphe précédent III.1.2, la décomposition en série de Fourier des coordonnées est linéaire de sorte que pour $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$,

$$\begin{bmatrix} u(t) \\ v(t) \end{bmatrix} = A \times \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} U_k \\ V_k \end{bmatrix} = A \times \begin{bmatrix} X_k \\ Y_k \end{bmatrix}$$

Cette relation n'est rigoureusement vérifiée que s'il n'existe pas de décalage du point de départ lors de la procédure de suivi de contour (cf. annexes C et D) entre le contour initial $\gamma^0(t^0) = (x(t^0), y(t^0))^T$ et le contour transformé $\gamma(t) = (u(t), v(t))^T$, c'est-à-dire si $t(t^0)$ est proportionnel à t^0 . Dans le cas contraire, la linéarité dans le domaine de Fourier est assurée seulement si $t(t^0)$ est également linéaire [ARBT-89].

Soit $t(t^0) = \mu(t^0 + \tau)$, μ constant, on a

$$\begin{bmatrix} U_k \\ V_k \end{bmatrix} = \exp(-2j\pi k\tau / T_0) \times A \times \begin{bmatrix} X_k \\ Y_k \end{bmatrix} = w^k A \begin{bmatrix} X_k \\ Y_k \end{bmatrix} \text{ avec } w = \exp(-2j\pi\tau / T_0)$$

Le décalage se concrétise donc sur chacun des descripteurs d'ordre k par un facteur multiplicatif $w(\tau)$ à la puissance k .

La paramétrisation du contour doit être adaptée à l'invariance recherchée [GHOR-90]. Plutôt qu'une paramétrisation de type équiaffine impliquant d'indésirables dérivations d'ordre élevé, Arbter propose la paramétrisation basée sur l'aire couverte entre deux points et le centre de gravité de la forme :

$$t = t(M) = \frac{1}{2} \int_{\eta=0}^{\eta=\eta(M)} |\det(\gamma(\eta) - G_c, \gamma'(\eta))| d\eta$$

où G_c désigne le centre de gravité de la courbe C et $\gamma(\eta)$ la représentation initiale de C au point M de la courbe (cf. Figure 23).

Les périodes T et T_0 sont reliées entre elles par la relation :

$$T = |\det(A)| T_0$$

En allant de $\eta=0$ à $\eta=\eta(M)$,
l'aire balayée est égale à :

$$\frac{1}{2} \left| \det \begin{bmatrix} x & dx \\ y & dy \end{bmatrix} \right|$$

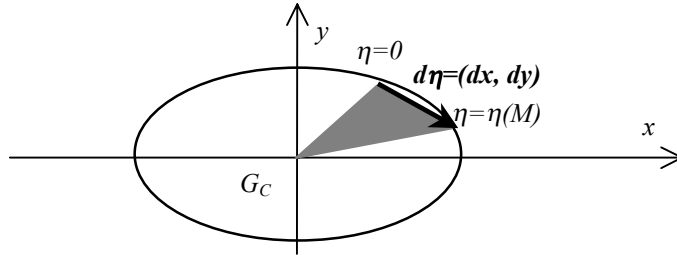


Figure 23 – Illustration de la paramétrisation proposée par Arbter dans le cas affine

Application au cas discret :

Soit C , un contour fermé décrit par ses N points de contour,

$$\gamma[i] = \begin{bmatrix} x[i] \\ y[i] \end{bmatrix} = \gamma_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, i = 0, \dots, N-1$$

La paramétrisation proposée par Arbter est donnée par :

$$\begin{cases} t_0 = 0; \\ t_{i+1} = t_i + \frac{1}{2} |(x'_i y'_{i+1} - x'_{i+1} y'_i)|, i = 0, 1, 2, \dots, N-1 \\ T = t_N \end{cases}$$

et la transformée de Fourier :

$$\Gamma_k = \begin{bmatrix} X_k \\ Y_k \end{bmatrix} = \frac{T}{(2\pi k)^2} \sum_{i=0}^{N-1} \left\{ \frac{1}{t_{i+1} - t_i} \begin{bmatrix} x_{i+1} - x_i \\ y_{i+1} - y_i \end{bmatrix} (\Phi_{k,i+1} - \Phi_{k,i}) \{1 - \delta(t_{i+1} - t_i)\} \right\} \\ + \frac{j}{2\pi k} \sum_{i=0}^{N-1} \left\{ \begin{bmatrix} x_{i+1} - x_i \\ y_{i+1} - y_i \end{bmatrix} \Phi_{k,i} \delta(t_{i+1} - t_i) \right\}$$

avec $\Phi_{k,i} = \exp\{-j2\pi kt_i/T\}$ et $\begin{pmatrix} x'_i \\ y'_i \end{pmatrix} = \begin{pmatrix} x_i - x_g \\ y_i - y_g \end{pmatrix}$
 où (x_g, y_g) désigne le centre de gravité de l'ombre.

Pour l'ensemble des descripteurs non nuls, en posant $\Gamma_k^0 = \begin{bmatrix} X_k \\ Y_k \end{bmatrix}$ et $\Gamma_k = \begin{bmatrix} U_k \\ V_k \end{bmatrix}$, il vient :

$$\Gamma_k = w^k A \Gamma_k^0, k \neq 0$$

et

$$\boxed{\Gamma = A \Gamma^0 W} \quad \text{où } W = \text{diag}(\dots, w^k, \dots)$$

$$\Gamma^0 = \begin{bmatrix} X_1 & \dots & X_k & \dots \\ Y_1 & \dots & Y_k & \dots \end{bmatrix} = \begin{bmatrix} \Gamma_1^0 & \dots & \Gamma_k^0 & \dots \end{bmatrix} \text{ et } \Gamma = \begin{bmatrix} U_1 & \dots & U_k & \dots \\ V_1 & \dots & V_k & \dots \end{bmatrix} = \begin{bmatrix} \Gamma_1 & \dots & \Gamma_k & \dots \end{bmatrix}$$

La construction d'invariants vise à éliminer A et W , i.e. **normaliser** les descripteurs de Fourier afin d'assurer leur invariance vis-à-vis :

- des transformations affines que peut subir la silhouette,
- du point de départ choisi lors de la paramétrisation du contour.

III.2.2. Les A-invariants

Les propriétés des descripteurs de Fourier définis plus haut permettent d'obtenir une famille d'invariants sur les silhouettes qu'ils décrivent.

D'une part, pour un décalage τ nul, grâce à la linéarité de la transformée de Fourier, on a :

$$\Gamma_k = A \Gamma_k^0, \quad k \neq 0$$

Prenons deux descripteurs distincts k et p , la matrice 2×2 $[\Gamma_k, \Gamma_p^*]$ peut s'écrire :

$$[\Gamma_k, \Gamma_p^*] = A [\Gamma_k^0, \Gamma_p^{0*}], \quad \text{où } * \text{ désigne le conjugué}$$

En prenant le déterminant de chaque membre de l'équation, on a :

$$\Delta_{k,p} = \Delta_k = \det[\Gamma_k, \Gamma_p^*] = \det(A) \det[\Gamma_k^0, \Gamma_p^{0*}] \quad \text{invariants scalaires relatifs}$$

Le nombre d'invariants peut être réduit en choisissant un indice p non nul et tel que $\Delta_p \neq 0$.

Des *invariants absolus* sont obtenus en éliminant $\det(A)$ de la manière suivante :

$$\boxed{Q_k = \frac{\Delta_k}{\Delta_p} = \frac{\det[\Gamma_k, \Gamma_p^*]}{\det[\Gamma_p, \Gamma_p^*]} = \frac{X_k Y_p^* - Y_k X_p^*}{X_p Y_p^* - Y_p X_p^*}, \Delta_p \neq 0.}$$

Dans le cas général de décalage τ non nul, $w = \exp(-2j\pi\tau/T_0) \neq 1$, on a alors :

$$[\Gamma_k, \Gamma_p^*] = A \times [\Gamma_k^0, \Gamma_p^{0*}] \times \text{diag}(w^k, w^{-p}), \text{ et}$$

$$\Delta_k = \det[\Gamma_k, \Gamma_p^*] = \det(A) \det[\Gamma_k^0, \Gamma_p^{0*}] w^{k-p}$$

$$\text{d'où } \boxed{Q_k = Q_k^0 w^{k-p}}$$

Il reste à ce stade à s'affranchir de la dépendance en w , i.e. assurer l'invariance vis-à-vis du point de départ.

III.2.3. Les AW-Invariants

La W-invariance est obtenue par normalisation des A-invariants Q_k . Les termes de phase sont définis par $\psi_{k_i} = \frac{Q_{k_i}}{|Q_{k_i}|}$ pour un sous-ensemble $\{Q_{k_1}, Q_{k_2}, \dots, Q_{k_n}\}$ dont les éléments ont tous un module non nul.

Arbter introduit une séquence $\{I_k\}_{k=\pm 1, \pm 2, \dots, \pm \infty}$ de grandeurs W-invariantes définies par :

$$I_k = |Q_k| \psi_k^{\rho_0} \prod_{i=1}^n \psi_{k_i}^{(k-p)\lambda_i}, \quad k=\pm 1, \pm 2, \dots, \pm \infty$$

où

ρ_0 est le plus grand commun diviseur (pgcd) de l'ensemble $\{k_1-p, k_2-p, \dots, k_n-p\}$

λ_i sont les entiers solutions de l'équation $\rho_0 + \sum_{i=1}^n \lambda_i (k_i - p) = 0$, $\lambda_i \in \mathbb{Z}$.

En définitive, on a les **descripteurs de Fourier affine-invariants** suivants pour les 3 premiers ordres:

$$\begin{aligned} n=0 &\rightarrow I_k = |Q_k| && \text{base incomplète} \\ n=1 &\rightarrow I_k = |Q_k| \psi_k^{q-p} \psi_q^{p-k} && \text{avec } |q-p|=1 \\ n=2 &\rightarrow I_k = |Q_k| \psi_k^{\text{pgcd}(q-p, r-p)} \psi_q^{(k-p)\lambda} \psi_r^{(k-p)\mu} && \text{avec } \text{pgcd}(q-p, r-p)=1 \end{aligned}$$

Sous l'hypothèse que les déformations 2D résultant du mouvement 3D de l'objet s'apparentent à une transformation affine de la silhouette, Arbter *et al.* montrèrent que les descripteurs de Fourier affine-invariants d'ordre zéro étaient adaptés à des applications 3D de reconnaissance d'objets [ARBT-90].

III.3. Transformation de Fourier inverse

Pour cette opération et compte tenu de la normalisation en échelle effectuée, deux paramètres sont à fixer : le paramètre K qui fixe le nombre de descripteurs de Fourier et un facteur échelle S qui fixe le format de la forme reconstruite du fait de la normalisation en échelle. La forme reconstruite est à traduire au centre d'une image de taille suffisante. Dans les exemples des Figure 24 et Figure 25, $K = 5$ et $S = 30$.

- **Double décomposition en série de Fourier** (cf. Figure 24)

$$\tilde{x}_i = \sum_{\substack{k=-K \\ k \neq 0}}^K S \cdot \tilde{X}_k \exp(j2\pi k i / N) \quad \text{et} \quad \tilde{y}_i = \sum_{\substack{k=-K \\ k \neq 0}}^K S \cdot \tilde{Y}_k \exp(j2\pi k i / N)$$

L'application de la transformation inverse fournit une forme identique à l'originale (avec un certain lissage dépendant du degré de décomposition K) à **une transformation géométrique près** composition d'une translation et d'une homothétie.

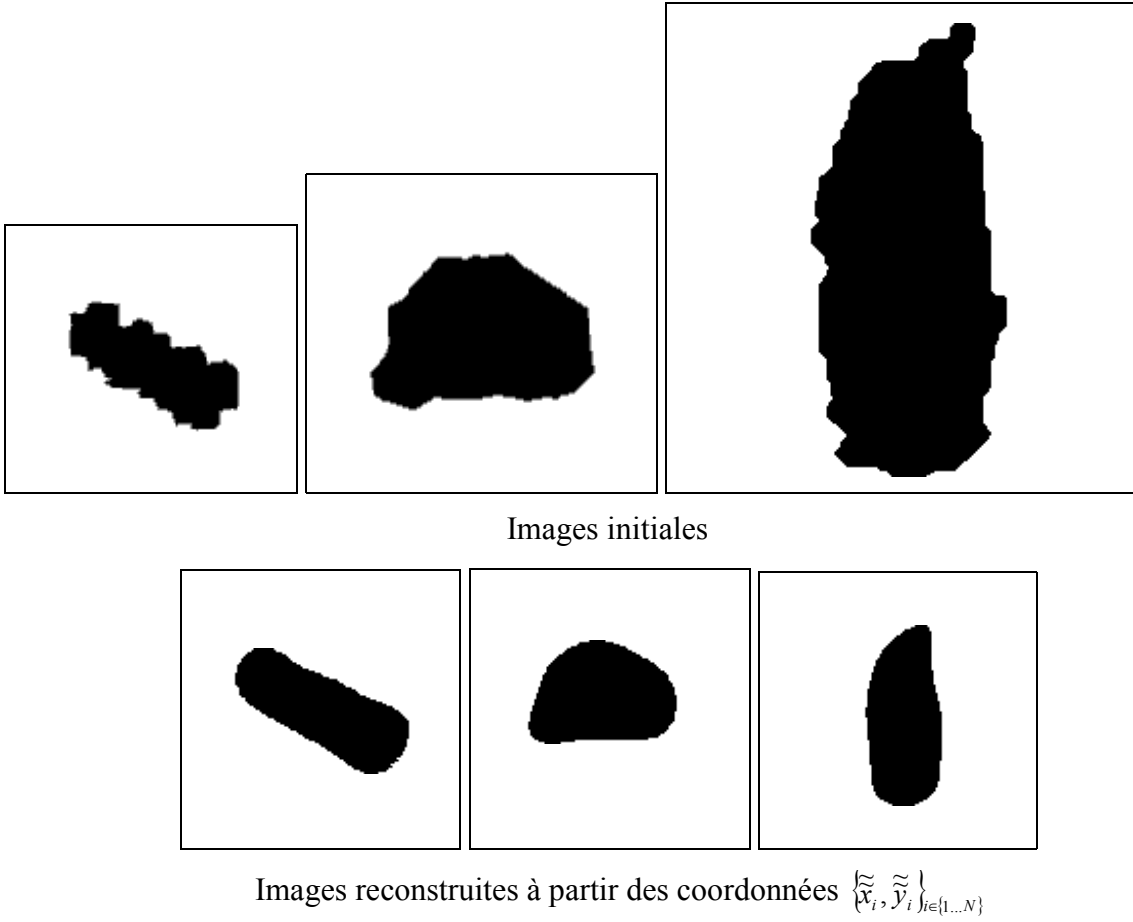


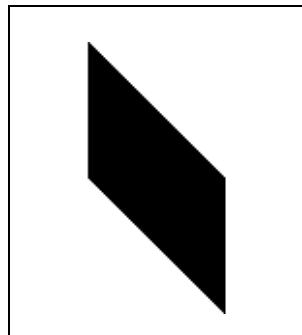
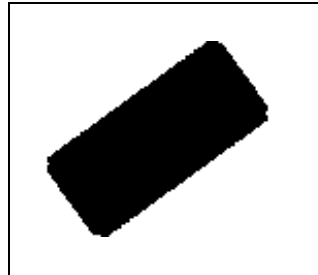
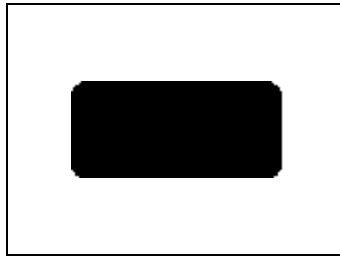
Figure 24 – Exemples de reconstructions après une transformée de Fourier des coordonnées des points de contour

- ***AW-invariants*** (cf. Figure 25)

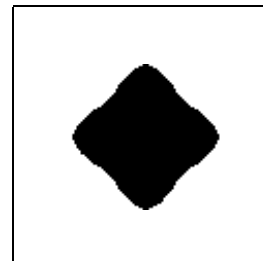
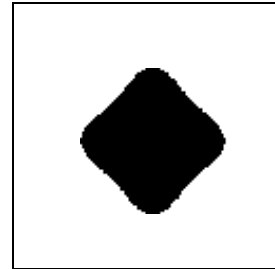
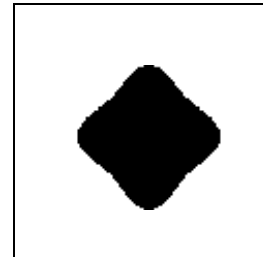
La reconstruction se fait par transformation de Fourier inverse à partir des *AW-invariants* $I_k = |Q_k|$ tel que

$$\tilde{\gamma}(t) = \begin{pmatrix} \tilde{x}(t) \\ \tilde{y}(t) \end{pmatrix} = \sum_{k=-K}^K S I_k \exp(j 2 \pi k t / T)$$

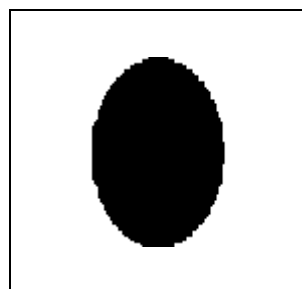
Notons que ces invariants d'ordre zéro (pour lesquels l'information contenue dans la phase est absente) ne sont pas complets (cf. §I), ils ne permettent généralement pas de reconstruire correctement la forme à la transformation affine près. Les exemples de la Figure 25 illustrent ce point tout en mettant en évidence l'invariance vis-à-vis de quelques transformations telles que rotation et biais. En outre, la normalisation vis-à-vis de la transformation affine ne permet pas de conserver le rapport des **longueurs des axes principaux** que nous observons sur la Figure 24. En ce sens, les *AW-invariants* $I_k = |Q_k|$ seront utilisés en tant qu'attributs synthétiques, notamment en ce qui concerne la méthode de classification.



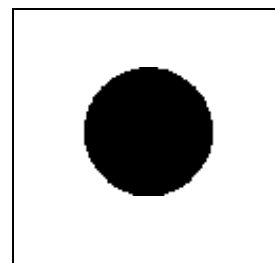
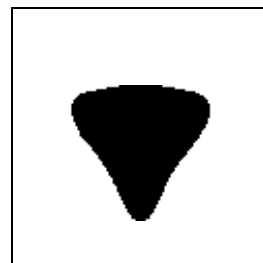
Silhouettes cylindriques initiales



Silhouettes reconstruites à partir des invariants



Silhouettes trapézoïdale et ellipsoïde initiales



Silhouettes reconstruites à partir des invariants

Figure 25 – Exemples de reconstructions par AW-invariants

IV. Attributs synthétiques

IV.1. Paramètres intrinsèques [JAN-88] [PROK-92]

Pour ce type de paramètres, une normalisation de l'image (selon §II.1) est indispensable. Les mesures métriques effectuées visent en effet à discriminer les objets selon leur morphologie. La normalisation permet de corriger les disproportions qui affectent la forme de l'ombre portée de manière à la rendre représentative de l'objet. Néanmoins, certaines de ces grandeurs ont le défaut de refléter franchement la disparité intraclasse des classes d'objets de tailles diverses comme cylindres ou sphères.

IV.1.1. Paramètres géométriques

Les paramètres quantitatifs tels que surface, périmètre, hauteur et largeur constituent les critères métriques de la forme.

La **surface S** est le nombre de pixels de la forme délimitée par le contour fermé. Elle correspond au moment d'ordre zéro (cf. annexe E).

Le **périmètre P** correspond à la longueur du contour fermé en tenant bien compte de la connexité des pixels. La distance entre deux pixels au sens d'un système de voisinage de la 4-connexité est 1.0 alors que la distance entre deux pixels au sens d'un système de voisinage de la 8-connexité est $\sqrt{2}$ (cf. Figure 26).

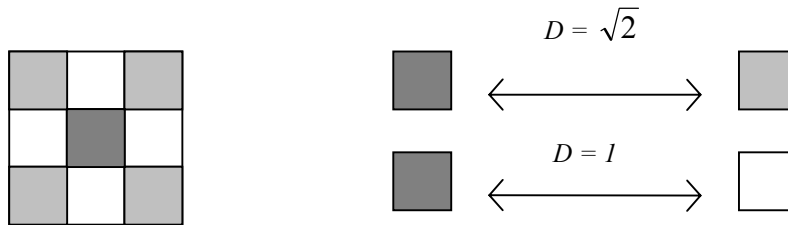


Figure 26 – Evaluation de la distance entre deux pixels

La **largeur W_s** et la **hauteur H_s** de l'ombre sont calculées à partir des lignes et colonnes voisines du centre de gravité de l'ombre et incluse dans la zone d'ombre (cf. W_i et H_j définis sur la Figure 27). La largeur (resp. hauteur) est la longueur moyenne des n lignes (resp. colonnes) voisines de celle qui passe par le centre de gravité, soient :

$$W_s = \frac{1}{n} \sum_{i=i_g - n/2}^{i_g + n/2} W_i$$

$$H_s = \frac{1}{n} \sum_{j=j_g - n/2}^{j_g + n/2} H_j$$

Le centre de gravité de l'ombre $G=(i_g, j_g)$ est calculé à partir des N points $(x_i, y_i)_{i \in \{0; N-1\}}$ de la zone d'ombre, $i_g = \frac{1}{N} \sum_{i=0}^{N-1} x_i$ et $j_g = \frac{1}{N} \sum_{i=0}^{N-1} y_i$.

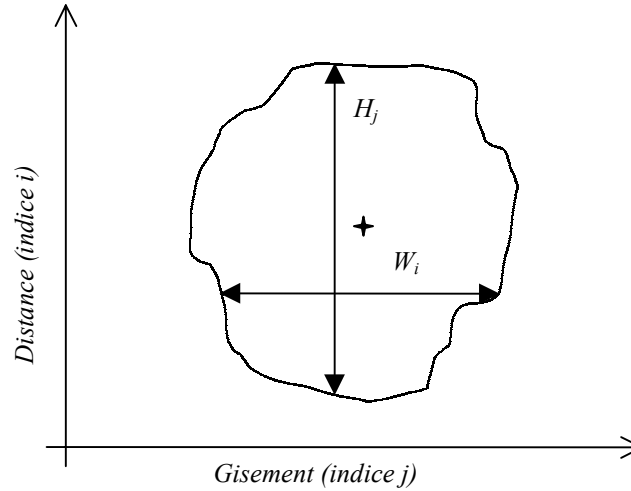


Figure 27 - Lignes W_i et colonnes H_j caractérisant la zone d'ombre

IV.1.2. Paramètres topologiques

Les paramètres topologiques caractérisent la forme globale de l'ombre.

- la **rondeur** d'une forme est définie par la relation

$$rondeur = \frac{P^2}{4\pi S},$$

où S et P désignent respectivement la surface et le périmètre précédemment définis. Ce paramètre est compris entre zéro et un quelle que soit la forme segmentée, et est égal à un pour le disque.

- l'**étendue** est définie à partir des moments centrés d'ordre deux (cf. annexe E),

$$etendue = \frac{\mu_{20} + \mu_{02}}{S^2}.$$

- l'**élongation** est obtenue en rapportant la différence des moments principaux d'inertie maximum et minimum à leur somme :

$$elongation = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{\sqrt{4\mu_{11}^2 + (\mu_{20} - \mu_{02})^2}}{\mu_{20} + \mu_{02}}.$$

Notons que ces derniers paramètres sont obtenus à partir de moments d'ordres faibles qui sont représentatifs de propriétés géométriques fondamentales de la forme. Suivant l'ordre choisi, on peut ainsi facilement connaître l'aire, les coordonnées du centre de gravité, l'orientation des axes principaux,...(cf. annexe E).

- Le **rapport hauteur sur largeur H/L** peut être calculé de deux manières.

Il est bien sûr égal au rapport des deux grandeurs définies ci-dessus, soit :

$$\left(\frac{H}{L}\right)_1 = \frac{H_s}{W_s}$$

Il peut être également évalué à partir des descripteurs de Fourier fondamentaux obtenus par une décomposition de chacune des coordonnées des points de contour en série de Fourier (cf. §III.1). En effet, les descripteurs d'indices 1 et -1 sont relatifs à l'ellipse qui approche le mieux la forme. En particulier, on peut récupérer une information qui concerne les axes principaux : l'orientation de l'axe principal majeur. Le rapport des modules de $\tilde{X}_1$ et $\tilde{Y}_1$ pour une paramétrisation dans le sens inverse des aiguilles d'une montre nous donnent une seconde définition du rapport hauteur sur largeur, soit :

$$\left(\frac{H}{L}\right)_2 = \frac{\tilde{X}_1}{\tilde{Y}_1}$$

Afin de démontrer la relation étroite entre ces deux définitions, on a fait tourner un cylindre autour de son centre de gravité de 0 à 360 degrés par pas de 5 degrés. La Figure 28 illustre l'évolution des deux rapports largeur sur hauteur. Il apparaît que les descripteurs de Fourier fondamentaux reflètent bien l'orientation des axes principaux de la forme. Cette propriété sera utilisée pour la classification multi-vues pour mettre en évidence les points de vue les plus pertinents lors d'une trajectoire circulaire du sonar autour de la cible.

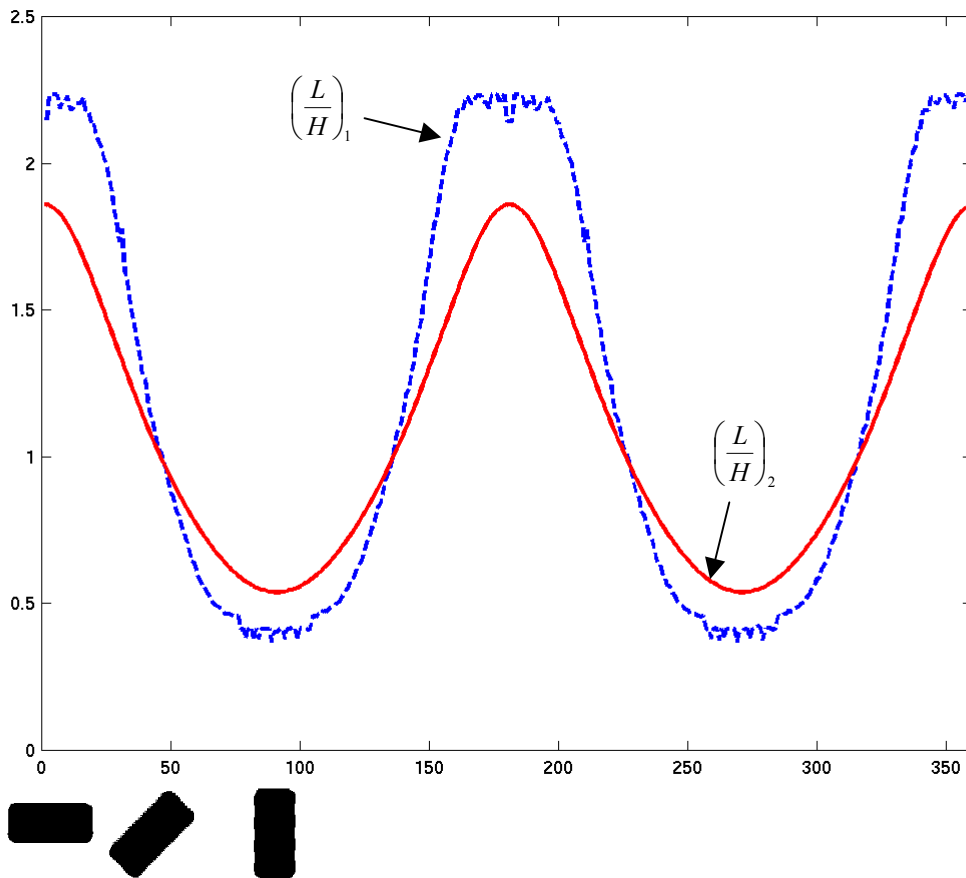


Figure 28 – Evolution des deux rapports largeur sur hauteur en fonction de l'angle de rotation

IV.2. Attributs invariants vis-à-vis des transformations du plan

Nous avons précédemment introduit des descripteurs de Fourier normalisés (cf. §III). Dans ce cas, nous pouvions alors reconstruire le contour et obtenir ainsi une approximation de la forme initiale. Nous introduisons cette fois un nombre plus réduit d'attributs qui synthétisent l'information initiale tout en ayant la propriété d'être invariant vis-à-vis des transformations du plan. La contrepartie d'une réduction du volume de données est l'impossibilité d'effectuer une opération inverse.

Les **invariants de moments** sont construits de manière à ce que les transformations du plan n'aient aucun effet sur les valeurs numériques des mesures de ces derniers. Ces invariants se présentent comme des **combinaisons de moments centrés normalisés**. D'une part, les moments centrés μ_{pq} déplacent le système d'axes au centre de gravité de l'ombre, c'est-à-dire, effectuent une invariance en translation. En les normalisant par $\mu_{00}^{1+\frac{p+q}{2}}$, on obtient d'autre part des moments centrés normalisés invariants en échelle μ'_{pq} (cf. annexe E). Il reste à assurer l'invariance vis-à-vis des autres transformations.

IV.2.1. Invariance vis-à-vis des similitudes

A partir de la théorie algébrique des invariants, Hu construit un ensemble de combinaisons non linéaires de moments centrés normalisés invariants par rotation [HU-62]. Ces sept grandeurs sont exprimées à partir des moments normalisés μ'_{pq} limités à l'ordre 3 (cf. annexe F). Un nombre plus important d'invariants de moments peut être dérivé d'une combinaison de moments de Zernike. Une approche simple développée par Teague permet de construire des invariants de n'importe quel ordre [TEAG-80].

Pour une fonction ou image $f(x, y) = f(r \cos \theta, r \sin \theta)$, les **moments complexes de Zernike** Z_{nl} sont définis par (* désigne le conjugué) :

$$Z_{nl} = \frac{(n+1)}{\pi} \int_0^{2\pi} \int_0^\infty V_{nl}(r, \theta)^* f(r \cos \theta, r \sin \theta) r dr d\theta$$

avec $V_{nl}(x, y)$ les **polynômes complexes de Zernike** d'ordre n définis par

$$V_{nl}(x, y) = e^{il\theta} \sum_{k=l}^n B_{nlk} r^k, \quad 0 \leq l \leq n, \quad n-l \text{ pair},$$

où les coefficients de Zernike B_{nlk} sont donnés par

$$B_{nlk} = (-1)^{(n-k)/2} \frac{[(n+k)/2]!}{[(n-k)/2]![(k+l)/2]![(k-l)/2]!} \quad \text{pour } n-k \text{ pair}.$$

Teague simplifie le calcul des moments de Zernike en établissant leur relation avec les moments centrés normalisés μ'_{pq} (cf. annexe E) :

$$Z_{nl} = \frac{(n+1)}{\pi} \sum_{k=l}^n \sum_{j=0}^q \sum_{m=0}^l (-i)^m \binom{q}{j} \binom{l}{m} B_{nlk} \mu'_{k-2j-l+m, 2j+l-m}, \text{ avec } q=(k-l)/2^2$$

Exprimés en termes de combinaisons de moments centrés normalisés, les moments de Zernike sont intrinsèquement invariants par translation et par changement d'échelle. Par ailleurs, la rotation provoque un simple décalage de phase des moments de Zernike. Ceci est à opposer à la formulation plus complexe obtenue pour les moments usuels (cf. annexe E), utilisée par Hu [HU-62]. Cette intéressante propriété fut exploitée par Teague pour la construction des invariants par combinaison de moments de Zernike. En se limitant à des moments d'ordre 4, on dispose de onze attributs $S_{i,j \in \{1...11\}}$ regroupés dans l'annexe F [PEJN-92].

IV.2.2. Invariance vis-à-vis des transformations affines [FLUS-93]

Flusser et Suk ont décomposé la transformation affine générale en six transformations ne dépendant chacune que d'un seul paramètre (présentées au §I) :

$$\begin{array}{lll} 1. \begin{cases} u = x + \alpha \\ v = y \end{cases} & 2. \begin{cases} u = x \\ v = y + \beta \end{cases} & 3. \begin{cases} u = \omega x \\ v = \omega y \end{cases} \\ 4. \begin{cases} u = \delta x \\ v = y \end{cases} & 5. \begin{cases} u = x + ty \\ v = y \end{cases} & 6. \begin{cases} u = x \\ v = t'x + y \end{cases} \end{array}$$

On y reconnaît la translation (1,2), l'homothétie (3), l'élongation (4) et le biais (5,6).

On recherche la fonction F qui permet de s'affranchir de ces quatre transformations élémentaires du plan :

☞ *invariance par translation*

Les moments centrés s'affranchissent des problèmes de translation (cf. annexe E). Toute combinaison F de ces moments est par conséquent invariante en translation :

$$F = \sum_i k_i \mu_{p_1(i)q_1(i)} \dots \mu_{p_{c(i)}(i)q_{c(i)}(i)}$$

où $c(i)$ est le nombre de moments constituant le $i^{\text{ième}}$ membre.

☞ *invariance par homothétie*

Dans ce cas, il convient de normaliser les facteurs de F tel que

$$F = \frac{\sum_i k_i \mu_{p_1(i)q_1(i)} \dots \mu_{p_{c(i)}(i)q_{c(i)}(i)}}{\mu_{00}^{z(i)}} \text{ où } z(i) = \frac{\left(\sum_{j=1}^{c(i)} (p_j(i) + q_j(i)) \right)}{2} + c(i)$$

² Rappel : $\binom{n}{p} = C_n^p = \frac{n!}{p!(n-p)!}$

En effet, par homothétie, les moments centrés deviennent $\mu'_{pq} = \omega^{p+q+2} \mu_{pq}$ (cf. transformation 3. ci-dessus), en particulier, $\mu'_{00} = \omega^2 \mu_{00}$ où μ_{00} représente la surface de l'ombre (cf. annexe E).

☞ *invariance par rapport à l'élongation*

Si on applique la transformation 4. ci-dessus, on obtient $\mu'_{pq} = \delta^p |\delta| \mu_{pq}$. Pour assurer l'invariance, on montre que la somme des indices p de F doit être égale à la somme des indices q c'est-à-dire $\sum_{j=1}^{c(\ell)} p_j(\ell) = \sum_{j=1}^{c(\ell)} q_j(\ell) = w$ où w est appelé poids de l'invariant.

☞ *invariance par rapport au biais*

L'invariance est obtenue en annulant les dérivées de F par rapport à t et t' (transformations 5. et 6.), soient deux équations différentielles :

$$DF = \frac{dF}{dt} = \sum_p \sum_q p \mu_{p-1, q+1} \frac{\partial F}{\partial \mu_{pq}} = 0$$

$$DF = \frac{dF}{dt'} = \sum_p \sum_q q \mu_{p+1, q-1} \frac{\partial F}{\partial \mu_{pq}} = 0$$

Grâce au théorème de Cayley-Sylvester, les auteurs obtiennent un nombre fini d'invariants de moments, indépendants, combinaisons de moments d'ordre 2 et 3 de poids w :

$$I_1 = \frac{1}{\mu_{00}^4} (\mu_{20} \mu_{02} - \mu_{11}^2), w = 2$$

$$I_2 = \frac{1}{\mu_{00}^{10}} (\mu_{30}^2 \mu_{03}^2 - 6 \mu_{30} \mu_{21} \mu_{12} \mu_{03} + 4 \mu_{30} \mu_{12}^3 + 4 \mu_{21}^3 \mu_{03} - 3 \mu_{21}^2 \mu_{12}^2), w = 6$$

$$I_3 = \frac{1}{\mu_{00}^7} [\mu_{20} (\mu_{21} \mu_{03} - \mu_{12}^2) - \mu_{11} (\mu_{30} \mu_{03} - \mu_{21} \mu_{12}) + \mu_{02} (\mu_{30} \mu_{12} - \mu_{21}^2)], w = 4$$

$$I_4 = \frac{1}{\mu_{00}^{11}} [\mu_{20}^2 \mu_{02}^2 - 6 \mu_{20}^2 \mu_{11} \mu_{12} \mu_{03} - 6 \mu_{20}^2 \mu_{02} \mu_{21} \mu_{03} + 9 \mu_{20}^2 \mu_{02} \mu_{12}^2$$

$$+ 12 \mu_{20} \mu_{11}^2 \mu_{21} \mu_{03} + 6 \mu_{20} \mu_{11} \mu_{02} \mu_{30} \mu_{03} - 18 \mu_{20} \mu_{11} \mu_{02} \mu_{21} \mu_{12}$$

$$- 8 \mu_{11}^3 \mu_{30} \mu_{03} - 6 \mu_{20} \mu_{02}^2 \mu_{30} \mu_{12} + 9 \mu_{20} \mu_{02}^2 \mu_{21}^2$$

$$+ 12 \mu_{11}^2 \mu_{02} \mu_{30} \mu_{12} - 6 \mu_{11} \mu_{02}^2 \mu_{30} \mu_{21} + \mu_{02}^3 \mu_{30}^2], w = 6$$

V. Conclusion du chapitre II.2

Les attributs classiques reflètent la géométrie de l'ombre. Malheureusement, des sous-groupes d'objets appartenant à la même classe mais de tailles différentes se distinguent en termes de mesures métriques. Ceci n'est pas acceptable dans le contexte de la chasse aux mines du fait des multiples dimensions des mines cylindriques et sphériques.

Outre ces attributs, nous avons introduit la notion d'**invariants** vis-à-vis des transformations du plan. Cette approche se justifie du fait des diverses formes possibles de l'ombre projetée en fonction de la prise de vue. La classification mono-vue suppose en effet que l'image observée soit caractéristique pour l'objet ciblé et ce, quel que soit le point de vue. Aussi, la classification mono-vue d'objets à géométrie complexe se trouve renforcée si l'on corrige les transformations géométriques qui affectent son contour. Deux stratégies sont possibles :

- normaliser l'image en rasance avant de calculer des invariants vis-à-vis de similitudes (translation, homothétie et rotation)
- utiliser des invariants affines qui corrigent également les distorsions inhérentes à l'imagerie à savoir, la forte elongation, le biais, et les disproportions dues à la différence de résolution selon les deux axes de l'image

Parmi ces invariants, deux approches se distinguent : une approche 'contour' dérivée d'une décomposition en série de Fourier et une approche 'région' via la théorie des moments sensibles à la distribution de masse dans l'image binaire (ombre sur fond).

D'un point de vue applicatif néanmoins, conserver une information distincte selon les lignes et selon les colonnes permet de rester cohérent avec le principe d'acquisition des images sonar en respectant l'orientation des formes relativement à l'axe de visée (pas d'invariance en rotation). La transformation de Fourier peut ainsi relier cette double information pourvu que la décomposition soit effectuée distinctement sur chacune des coordonnées des pixels de l'image.

L'exploitation de ces divers attributs dépend du caractère descriptif ou synthétique des attributs calculés.

D'une part, les attributs descriptifs tels que les descripteurs de Fourier permettent de reconstruire la forme. On verra que cette propriété est à l'origine de deux applications intéressantes : l'appariement de contours et une technique originale de segmentation orientée classification. L'avantage est de manipuler un faible nombre de solutions potentielles en exploitant au mieux les propriétés de ces attributs et leur flexibilité.

A l'inverse, les attributs synthétiques nécessitent une grande base de données pour l'apprentissage de l'algorithme de classification. Ils sont adaptés aux techniques classiques d'analyse de données [CELE-89] [BOUR-98] : Analyse par Composantes Principales (ACP), classification automatique par la méthode des k-plus-proches-voisins (k-ppv), nuées dynamiques supervisées, discrimination linéaire...

Cette analyse mono-vue fournit un ensemble de connaissances à intégrer dans le processus de classification multi-vues (partie III). Les différentes pistes proposées mèneront en particulier à l'utilisation judicieuse de différents jeux d'attributs en fonction des hypothèses et des algorithmes qui en découlent.

Conclusion de la partie II

Dans cette partie, nous nous sommes attachés à trouver les outils adaptés à l'analyse des images sonar.

L'étape initiale de la chaîne de traitement de l'image est la segmentation.

Dans un premier temps, des méthodes classiques de segmentation ont été paramétrées par deux grandeurs R_i et R_j . Etablies à partir de la taille du pixel de l'image observée et de la rasance du sonar, elles ont permis de mener les traitements de régularisation et les opérations morphologiques indépendamment des conditions d'acquisition de l'image.

Dans un second temps, une méthode originale de segmentation basée sur la minimisation d'un critère statistique a utilisé la corrélation des pixels pour adopter une taille de filtre adéquate.

Les traitements de haut-niveau, visant à extraire des attributs discriminants la forme de l'ombre portée, constituent la seconde étape de la chaîne de traitement. Deux avancées apparaissent par rapport à l'utilisation classique de critères géométriques.

La première avancée est la notion d'invariance vis-à-vis des transformations affines. Elle permet en particulier de corriger deux distorsions géométriques inhérentes à l'imagerie sonar : l'élongation et le biais.

La seconde avancée est l'information relatée par les descripteurs de Fourier calculés distinctement sur chacune des coordonnées des pixels. Outre ses qualités reconnues (reconstruction, lissage,...), cette transformation transpose dans le domaine fréquentiel une double information suivant lignes et colonnes. On peut alors récupérer une information approximative sur l'orientation des axes principaux ainsi que les paramètres de la transformation affine affectant le contour.

PARTIE III :

Application mono-vue

Introduction de la partie III

La classification des mines marines à partir d'images sonar haute résolution est basée sur l'analyse de la forme de l'ombre portée sur le fond.

Le processus classique de classification sur ombre d'un objet immergé consiste en trois étapes successives :

SEGMENTATION → EXTRACTION D'ATTRIBUTS → CLASSIFICATION

La segmentation extrait les régions d'intérêt dans l'image, i.e., dans le cas de la classification sur ombre d'images sonar, partage les pixels de l'image en deux classes : la classe 'ombre' comprenant les pixels de la zone d'ombre et la classe 'fond' comprenant à la fois les pixels de la zone de réverbération de fond et de l'écho. A partir de l'image binaire obtenue, des paramètres discriminants ou attributs caractéristiques sont calculés sur la forme de l'ombre ou uniquement son contour, et regroupés dans un vecteur caractéristique. Cette étape d'extraction d'attributs a un double objectif. D'une part, elle comprime l'information disponible en la réduisant à un nombre restreint de paramètres classifiants. D'autre part, le vecteur caractéristique introduit dans le module de classification doit permettre d'affecter en finalité l'objet à une classe d'objets. La classification est faite de manière automatique, i.e. sans intervention de l'opérateur, et de manière supervisée nécessitant un nombre conséquent d'éléments représentatifs de chacune des classes dont on peut rapprocher l'élément à classer.

A chacune des étapes du processus séquentiel ci-dessus sont associés des pré- et/ou post-traitements superflus au regard de l'objectif de classification final. Il s'agit par exemple du détail exigé sur le contour à l'étape de segmentation. Il s'agit aussi de préparer l'image pour l'étape qui suit et d'éviter qu'une erreur intervenant à un instant donné ne se répercute jusqu'au résultat final sans rétroaction possible. Deux approches alternatives sont proposées. La première consiste à apparier le contour segmenté à un contour prototype. Cette approche réduit la base d'apprentissage à son strict minimum, i.e. un prototype par classe. La seconde alternative, plus originale, réside en un processus dynamique par algorithme génétique qui permet d'envisager une « segmentation orientée classification ». En se concentrant sur l'objectif ultime de la chaîne de traitement, une segmentation discriminante du contour est effectuée à partir des données sonar brutes. Le résultat est double : tout en caractérisant le contour de l'ombre, on offre la possibilité de classer l'objet détecté à partir d'une population réduite d'individus.

Chapitre III.1 :

Approche séquentielle

I. Introduction du chapitre III.1

La classification d'un objet immergé revient à déterminer la classe associée à son ombre acoustique portée sur le fond. Par binarisation de l'image sonar, l'étape de segmentation a délimité le contour duquel nous avons extrait un certain nombre d'attributs caractéristiques. Ce processus est destiné à discriminer les différentes formes existantes.

Un attribut synthétique reflète un trait particulier mais non exclusif de la forme : une même valeur d'attribut peut être obtenue pour deux formes distinctes. Dès lors, une bonne reconnaissance de la forme nécessite d'y associer un vecteur caractéristique à L composantes (attributs caractéristiques) discriminant au mieux les différentes classes dans l'espace à L dimensions. Dans cet espace, on peut en outre recourir à une Analyse en Composantes Principales afin d'optimiser a posteriori la représentation des attributs en la rendant plus synthétique. Grâce à un apprentissage dit supervisé car toutes les classes (ou formes) sont connues, l'algorithme de classification tente alors d'affecter la forme observée inconnue à la classe souhaitée parmi toutes les classes apprises définies à l'apprentissage. La classification automatique consiste en une **analyse des données (attributs)** mise en œuvre dans des **classifieurs classiques** (k-plus-proches-voisins, discrimination linéaire...). Les performances de ces classifieurs dépendent de la configuration des données, c'est-à-dire une dispersion intra-classe, une distribution multi-modale des objets,...

La base de données utilisée est décrite dans l'annexe G. Un ensemble d'attributs est calculé après segmentation par seuillage selon la méthode du paragraphe IV du chap. II.1. Nous rappelons que celle-ci a été choisie pour l'application multi-vues. La présente analyse mono-vue doit nous aider à choisir les attributs appropriés à l'élaboration d'un processus de classification multi-vues de l'objet.

II. Outils d'évaluation

II.1. Classifieurs classiques [CELE-89]

Trois modules de classification implémentés dans le logiciel NeuroClasse (cf. annexe I) sont utilisés :

- le *k Nearest Neighbour (k-NN)* ou *k plus proches voisins (k-ppv)*

Cette méthode consiste, pour un pattern (= vecteur d'attributs) dont on cherche la classe, à réaliser un vote parmi les patterns qui sont les plus proches et dont on connaît la classe. Ce voisinage est dimensionné par un paramètre désignant le nombre de plus proches voisins pris en compte pour déterminer la classe. Ce paramètre est fixé à 1 par défaut.

Ce classifieur ne dépend pas de la distribution des éléments et reste donc performant dans le cas de dispersion intra-classe ou d'une distribution multi-modale. Son principal défaut est le temps de calcul nécessaire.

- L'algorithme des *nuées dynamiques supervisées*

Il consiste à caractériser la base d'apprentissage par un nombre fini de points (centroïdes) représentatifs des patterns d'apprentissage. Chaque nouveau pattern à tester est alors affecté au centroïde le plus proche.

Ce classifieur quantifie en quelque sorte le risque associé à la réduction de chaque classe à un seul élément représentatif, i.e. le centroïde.

- L'algorithme de *discrimination linéaire*

Il suppose que les densités de probabilité de chaque classe suivent une loi normale et que les matrices de covariance sont identiques. Après estimation de la matrice de covariance, l'approche bayésienne fournit alors explicitement les équations des hyperplans séparateurs.

Ce classifieur est par définition bien adapté dans le cas d'éléments normalement répartis.

II.2. Analyse en composantes principales [VOLL-85] [CELE-89] [BOUR-98]

Le but de l'analyse en composantes principales (ACP) est d'obtenir une représentation d'un nuage de points (objets ou individus) caractérisés par un certain nombre L de caractères (attributs) dans un espace de dimension réduite M de telle manière que l'inertie portée par cet espace soit la plus grande possible. L'intérêt de cette représentation, quantifiable par la part d'inertie du nuage d'individus qu'elle porte, n'est réel que si la substitution de ces M composantes aux attributs n'entraîne pas une perte d'information importante. En particulier, la réduction du nombre de caractères à deux composantes principales permet une représentation géométrique plane des individus.

Cette technique est appliquée ici à un nuage de points représentant les individus centrés et réduits c'est-à-dire que l'on a opéré sur chaque attribut la transformation linéaire ramenant sa moyenne à zéro et sa variance à l'unité. Cette opération permet de donner à chaque attribut la même importance que celle des autres attributs dans le calcul des distances entre individus. En somme, pour un attribut donné $a_i(l)$, on effectue l'opération suivante :

$$a_i^{new}(l) = (a_i^{old}(l) - \bar{a}_i) / \sigma_{a_i}$$

où $\bar{a}_i$ est la moyenne et σ_{a_i} est l'écart-type de cet attribut sur l'ensemble de la base.

Le bref rappel du principe de l'ACP suivant nous permet d'expliciter les notations utilisées.

Soit $A = (a_{ij})_{(i,j) \in \{1, \dots, L\} \times \{1, \dots, N\}}$, la matrice des N individus caractérisés par L attributs centrés réduits, l'ACP remplace les L attributs initiaux par des attributs non corrélés de variance maximale et d'importance décroissante. Ces attributs 'synthétiques' sont obtenus en combinant les attributs initiaux au moyen de facteurs que sont les vecteurs propres de la matrice de covariance :

$$C = \frac{1}{N} AA^T$$

Les axes principaux sont les L vecteurs propres $\vec{V}_i$ de C associés aux valeurs propres λ_i rangées en ordre décroissant. Soit U la matrice contenant les vecteurs propres,

$$U = \begin{bmatrix} \vdots & & \vdots \\ \vec{V}_1 & \dots & \vec{V}_L \\ \vdots & & \vdots \end{bmatrix}$$

Les composantes principales f_i sont obtenues par projection des N vecteurs caractéristiques (de composantes a_i) dans la base des vecteurs propres, soit :

$$F = U^T A$$

Chaque composante f_i est une nouvelle variable, combinaison linéaire des attributs initiaux, caractérisée par une moyenne nulle et une variance égale à la valeur propre associée λ_i .

Pour nos attributs centrés et réduits, l'inertie totale I de l'ensemble des individus égale à la somme des valeurs propres correspond au nombre d'attributs L . Le $i^{\text{ème}}$ axe principal a une part d'inertie égale à $\frac{\lambda_i}{I}$ et la part d'inertie cumulée des deux premiers axes est $\frac{\lambda_1 + \lambda_2}{I}$. Elle

mesure l'aplatissement du nuage sur le plan principal, i.e. la qualité de la représentation du nuage sur ce plan. Plus généralement, la réduction de l'espace des données est d'autant plus réalisable que la part d'inertie portée par les premières composantes est importante.

L'interprétation du lien entre composantes principales et attributs initiaux peut être effectuée par visualisation du cercle des corrélations. Dans ce cercle, les coordonnées de chaque point sont les corrélations de chaque attribut avec les composantes principales f_1 et f_2 . Ces dernières étant non corrélées, les attributs s'inscrivent dans un cercle de rayon 1. Un attribut est d'autant mieux représenté dans ce plan qu'il se situe au bord du cercle. Pour nos attributs centrés et réduits, le coefficient de corrélation entre f_i et a_l s'écrit simplement :

$$r(f_i, a_l) = V_{il} \times \sqrt{\lambda_i}$$

Sur ce cercle, des groupes formés d'attributs corrélés entre eux apparaissent. Un rapide examen du cercle permet de repérer les groupes liés entre eux ou opposés à condition que les points tracés soient proches de la circonférence.

III. Evaluation des attributs

En utilisant trois classifieurs différents, quelques comparaisons peuvent être faites entre différents types d'attributs introduits plus haut. Nous opposons plus particulièrement trois ensembles d'attributs (centrés et réduits) sans (cf. Tableau 4) puis avec lissage de contour (cf. Tableau 5) :

- les attributs intrinsèques classiquement utilisés qui nécessitent une normalisation de l'image : largeur, hauteur, rondeur, étendue et élongation (cf. §IV.1 du chap. II.2)
- les invariants de moments en rotation, translation et homothétie pour lesquels une normalisation en rasance est indispensable du fait de la forte élongation induite non prise en compte : les invariants de moments de Zernike (cf. §IV.2.1 du chap. II.2)
- les invariants affines pour lesquels aucune normalisation n'est effectuée : invariants de moments de Flusser et Suk et A-W invariants de Arbter (cf. §IV.2.2 et §III.2 du chap. II.2)

Tableau 4 – Taux de bonne classification pour trois classifieurs sans lissage de contour : discrimination linéaire (ligne 1), nuées dynamiques (ligne 2) et k-ppv avec k=1 (ligne 3)

Classes <i>Attributs</i> (nombre)	<i>Cylindres</i>	<i>Sphères</i>	<i>Mines Manta</i>	<i>Mines Sigeel</i>	<i>Mines Rockan</i>	moyenne (écart-type)
<i>Attributs intrinsèques (5)</i>	94.8	95.9	100	94.4	55.1	88.0 (18.5)
	84.5	95.8	100	97.4	62.8	88.1 (15.4)
	99.8	99.9	100	100	99	99.8 (0.4)
<i>Invariants de moments de Zernike (11)</i>	68.6	100	98.9	100	82.6	90.0 (14.1)
	67.4	100	100	100	61.9	85.9 (19.5)
	100	100	98.3	100	99.8	99.6 (0.7)
<i>Moments affines de Flusser et Suk (4)</i>	99	59.1	95.4	97.9	75.8	85.4 (17.5)
	62.7	74.7	70.4	99.1	58	73.0 (16.0)
	85.9	73.8	91.2	93.5	92.2	87.3 (8.1)
<i>AW-Invariants de Arbter (8)</i>	86.7	83.1	75.5	91.1	78	83.1 (6.1)
	73.7	92	76.9	90.2	80.2	82.6 (8.1)
	90.4	83.4	67.9	92.8	82	83.3 (9.7)

Parmi les différents ensembles d'attributs testés, les AW-invariants de Arbter donnent le plus faible écart-type (observé sur l'ensemble des résultats par classe) pour les deux premiers classifieurs. Ceci démontre que la configuration des attributs permet l'analogie avec la gaussienne faite par le classifieur « discrimination linéaire » ou encore la réduction à un centroïde comme le stipule le classifieur « nuées dynamiques ». Effectivement, quelle que soit la classe, les éléments sont caractérisés par des **attributs groupés autour d'une valeur moyenne** caractéristique avec une variance sensiblement identique. Ces invariants semblent le mieux répondre au problème posé qui consiste à s'affranchir des distorsions géométriques affectant le contour.

Les critères intrinsèques donnent les résultats les plus **médiocres pour la classe Rockan** excepté par le classifieur « k-ppv ». Par ailleurs, il est hasardeux de choisir un centroïde du fait des différentes tailles d'objets possibles pour cylindres et sphères. Ils en résultent **l'apparition de sous-groupes** (distribution multi-modale) comme l'illustre la figure de gauche de la Figure 29. En ce sens, notons que le fort pourcentage de 99.8% obtenu avec le classifieur k-ppv n'est obtenu qu'à la condition que la base d'apprentissage comporte les mêmes objets et surtout de mêmes tailles que ceux que l'on classe.

Les invariants de moments de Zernike donnent également des résultats moyens par les deux premiers classifieurs : l'écart-type associé aux résultats obtenus est effectivement important ce qui démontre que certaines classes sont pénalisées. Ces invariants reflètent en effet la **dispersion intra-classe** des classes 'cylindres' et 'mines Rockan' (cf. figure de droite de la Figure 29, classes 0 et 4). A nouveau, seule la méthode des plus proches voisins fournit un bon taux de classification.

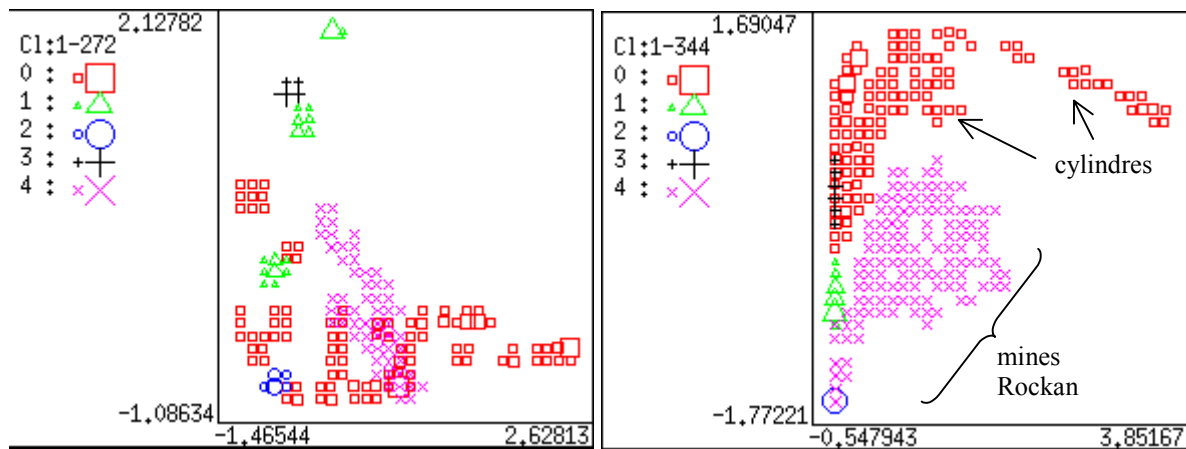


Figure 29 – Exemples de configurations problématiques (multi-modalité à gauche et dispersion intra-classe à droite)

Enfin, les moments affines de Flusser et Suk se sont révélés sensibles au bruit affectant les silhouettes de petits objets comme la mine Manta et Rockan (forte dispersion pour ces classes). Ce problème accentué par un contour mal résolu en gisement est moins critique après un **lissage de contour** comme le montre l'amélioration des résultats du Tableau 5.

Tableau 5 – Taux de bonne classification pour trois classifieurs **avec lissage de contour** : discrimination linéaire (ligne 1), nuées dynamiques (ligne 2) et k-ppv avec k=1 (ligne 3)

Classes Attributs (nombre)	<i>cylindres</i>	<i>Sphères</i>	<i>Mines Manta</i>	<i>Mines Sigeel</i>	<i>Mines Rockan</i>	Moyenne (écart-type)
<i>Invariants de moments de Zernike (11)</i>	99.4	100	100	100	80.2	95.9 (8.8)
	68.9	100	100	100	60.7	85.9 (19.5)
	100	100	97.4	100	99.2	99.3 (1.1)
<i>Moments affines de Flusser et Suk (4)</i>	90.5	99.1	96.1	97.7	73.8	91.5 (10.4)
	85.3	99.5	57.1	99.8	60.6	80.5 (20.6)
	95.9	92.9	86.6	95.8	90.7	92.4 (3.9)

Notons bien que les performances des attributs choisis, excepté pour les A-W invariants de Arbter, sont obtenues **sous la contrainte d'un prétraitement** de l'image, i.e. normalisation et/ou lissage du contour. En revanche, les A-W invariants de Arbter ont l'avantage de réaliser **par construction** ces prétraitements puisqu'il s'agit de la combinaison de descripteurs de Fourier (d'où le lissage) invariants vis-à-vis de la transformation affine (la forme est par conséquent normalisée). De même pour les combinaisons de moments, alors que les invariants de moments de Zernike nécessitent une normalisation de l'image, ceux de Flusser et Suk solutionnent le problème puisque la transformation affine inclut l'élongation.

Enfin, quels que soient les attributs testés, **la classe des mines Rockan reste problématique**. En effet, elle est caractérisée par une variance importante des attributs et ce, même pour les invariants affines, du fait de sa géométrie particulière : la différence observée entre deux points de vue n'est pas simplement due à une transformation affine du contour. **Une analyse multi-vues semble indispensable pour caractériser de manière robuste ce type particulier de mines et éviter les confusions.**

IV. Hybridation [QUID-00a]

Généralement, les auteurs travaillent sur un ensemble d'attributs de même nature (statistique, géométrique, une seule famille d'invariants,...). De manière à profiter des avantages de certains attributs tout en palliant leurs éventuels défauts, un **vecteur caractéristique hybride** formé d'attributs d'origines différentes, géométrique (topologie de la forme) et statistique (distribution spatiale des pixels), mais complémentaires a été construit (cf. §IV.1 du chap. II.2 de la partie II et §IV.2.2 du chap. II.2). Les moments basés sur la masse totale des pixels de l'ombre (y compris les pixels de la frontière et les pixels intérieurs) sont bien plus robustes au bruit affectant le contour que les techniques prenant en compte les seuls pixels de la frontière. Une combinaison judicieuse de ces moments effectuée par Flusser et Suk assure l'invariance vis-à-vis des transformations affines [FLUS-93]. Parmi les attributs topologiques d'autre part, l'étendue et l'élongation basés sur les moments d'ordre deux dits moments d'inertie ont été choisis du fait de leur indépendance vis-à-vis du changement d'échelle et de l'orientation. Finalement, le vecteur hybride comporte quatre composantes, i.e. les invariants de moments I_1 et I_2 d'une part, et l'étendue et l'élongation d'autre part. Evidemment, du fait de l'utilisation d'attributs géométriques, la normalisation préalable de l'image est indispensable ce qui rend plus robuste du même coup les attributs de Flusser et Suk en rééquilibrant les résolutions en distance et en gisement. Un lissage du contour est également effectué du fait des observations faites au paragraphe précédent.

Le bénéfice de l'hybridation est visualisé au moyen d'une ACP sur la Figure 30. La part d'inertie cumulée des deux premiers axes égale à 93,7 % montre que l'espace caractéristique peut être réduit à deux dimensions sans perte importante d'information ce que nous vérifions un peu plus loin via le taux de bonne classification.

Les première et cinquième classes correspondant aux mines cylindriques et Rockan sont bien entendu les plus dispersées du fait de la grande variété des apparences des ombres. Néanmoins, la comparaison des matrices de confusion obtenues par classifieur k-ppv ($k=1$)

avant (Tableau 6) et après ACP (Tableau 7) montrent qu'en réduisant l'espace à deux dimensions, on maintient un taux de bonne classification satisfaisant.

L'optimisation effectuée par l'ACP est aisément mise en évidence lorsqu'on compare différentes classifications avec un espace caractéristique à deux dimensions. Pour cela, nous avons effectué le test suivant : nous sélectionnons seulement deux composantes de nature différentes parmi les quatre initiales et confrontons leurs performances à celles que nous obtenons en prenant les deux premières composantes principales. Il apparaît que la réduction du nombre d'attributs à deux attributs synthétiques (les deux premières composantes principales) via l'ACP est meilleure que la sélection des deux meilleurs d'entre eux. Le Tableau 8 détaille ces résultats avec, à titre de comparaison, la classification obtenue dans l'espace caractéristique initial à quatre dimensions.

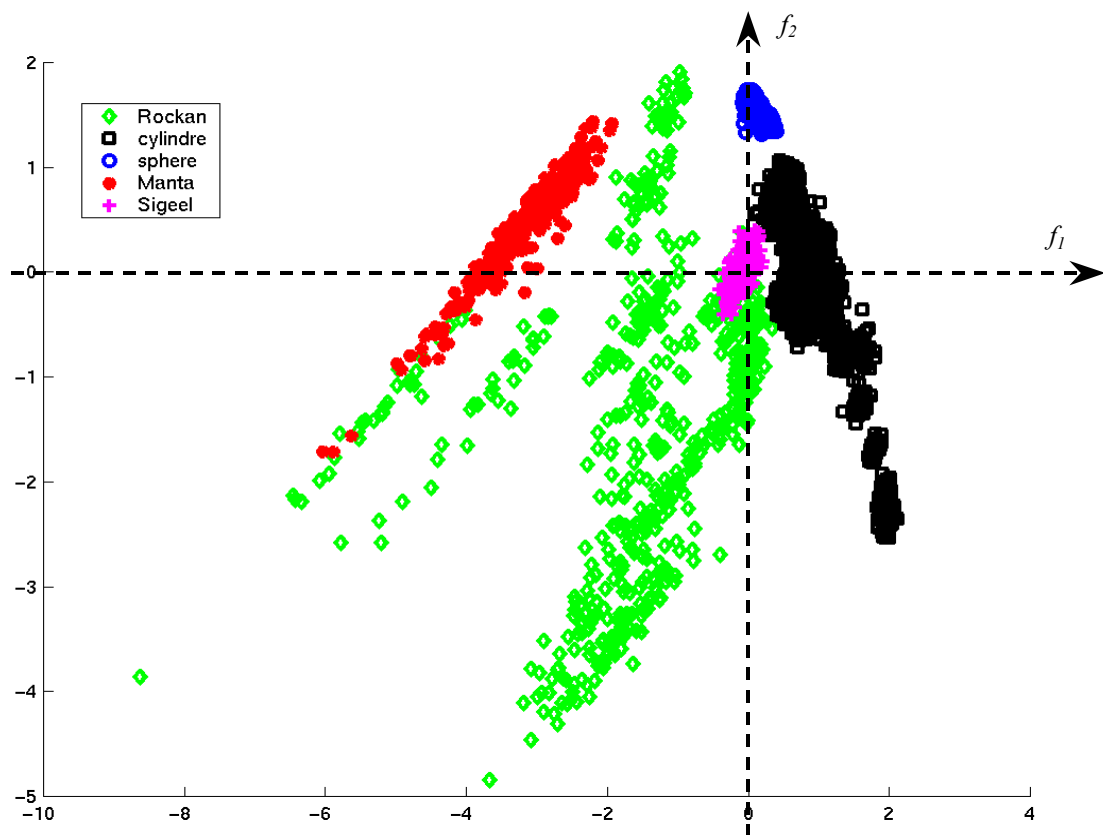


Figure 30 – Bonne partition dans le plan principal: bénéfice de l'hybridation

Tableau 6 – Matrice de confusion par k-ppv avant ACP, i.e. un taux moyen de 99.0%

classes exemples	cylindre	Sphère	mine Manta	mine Sigeeel	mine Rockan
cylindres	100	0	0	0	0
sphères	0	100	0	0	0
mines Manta	0	0	97.6	0	2.4
mines Sigeeel	0	0	0	99.3	0.7
mines Rockan	0	0	1.1	0.9	98.1

**Tableau 7 - Matrice de confusion par k-ppv après ACP (2 composantes principales)
soit un taux moyen de 96.6%**

classes exemples	cylindre	Sphère	mine Manta	mine Sigeel	mine Rockan
cylindres	100	0	0	0	0
sphères	0	100	0	0	0
mines Manta	0	0	95.6	0	4.4
mines Sigeel	0	0	0	92.2	7.8
mines Rockan	0.1	0	0.9	3.7	95.6

Tableau 8 – Comparaison espaces caractéristiques 2D et 4D

	I_1 , <i>Elongation</i>	f_1, f_2	I_1, I_2 , <i>étendue, élongation</i>
Cylindres	99.7	100	100
Sphères	100	100	100
Mines Manta	98.3	95.6	97.6
Mines Sigeel	89.7	92.2	99.3
Mines Rockan	91.3	95.3	98.1
<u>moyenne</u>	<u>95.8</u>	<u>96.6</u>	<u>99.0</u>
variance	4.9	3.4	1.1

Remarque : toutes ces classifications sont les classifications moyennes obtenues à partir de trois tirages aléatoires différents sur l'espace caractéristique initial avant division en base d'apprentissage et de généralisation.

V. Conclusion du chapitre III.1

Cette approche séquentielle pour l'application mono-vue permet de tirer quelques enseignements utiles à l'application multi-vues. La plupart des familles d'attributs testées prises de manière isolée admettent des faiblesses de différents types : sensibilité aux différentes tailles des objets dans une même classe (distribution multi-modale des attributs), dispersion des attributs, recouvrement de classes,... En ce sens, l'hybridation apporte une solution. Une ACP met effectivement en évidence, sur le plan principal, le fait que les faiblesses respectives des deux familles associées n'apparaissent plus. La multi-modalité a disparu et les recouvrements sont moins importants. Ce vecteur caractéristique d'attributs hybrides sera réutilisé pour une application multi-vues (cf. §I du chap. IV.1).

Chapitre III.2 :

Approches alternatives

I. Introduction du chapitre III.2

La principale faiblesse de l'approche classique séquentielle vient de son mode de classification. Les classifieurs classiques nécessitent en effet de disposer d'une grande base d'apprentissage. Alors qu'il est difficile d'être exhaustif, d'autres approches dites alternatives sont proposées.

II. Appariement de silhouettes

II.1. Introduction

En vue d'identifier la nature de l'objet détecté par le sonar, on propose d'apparier deux silhouettes, l'une correspondant à son ombre segmentée, l'autre au prototype. De nombreux travaux traitent de ce sujet en utilisant les descripteurs de Fourier mais en se limitant souvent aux similitudes (translation, rotation, homothétie) alors que l'élongation et le biais affectent fréquemment les silhouettes des ombres portées observées en imagerie sonar [PERS-86] [MARQ-97]. Le cas général d'une transformation affine a été récemment exposé mais en appariant un nombre donné de points équirépartis sur chacun des contours [JAIN-96] [MIGN-00].

L'algorithme proposé ici permet une estimation rapide des paramètres de la transformation affine au sens des moindres carrés. Plutôt que de raisonner directement à partir des points équirépartis sur chacun des contours, on choisit d'effectuer une décomposition du contour en série de Fourier en profitant de la propriété de linéarité des descripteurs de Fourier (cf. §III.1 du chap. II.2). En passant dans le domaine de Fourier, outre l'intéressant lissage obtenu, on réduit le nombre de données d'un facteur $N/4K$ en passant de N points de contour à $2K$ descripteurs (complexes) si on se limite à un degré de décomposition K . Néanmoins, l'appariement ne peut s'effectuer qu'à partir d'une description unique des contours d'où la nécessité de corriger le décalage du point de départ apparaissant lors de la procédure de suivi du contour. En réalité, ce décalage se concrétise par un simple facteur multiplicatif sur les descripteurs et peut être pris en compte dans le processus d'estimation.

Dans une ultime étape, la mesure de la similarité entre le contour prototype déformé conformément à la transformation affine estimée et le contour détecté permet alors de confirmer ou au contraire d'infirmer le choix du prototype initial.

II.2. Position du problème

La décomposition en série de Fourier des coordonnées des points de contour présente une pratique analyse de la forme selon les deux dimensions de l'image. Grâce à la linéarité de la décomposition, les transformations géométriques opérant sur les coordonnées des N points de contour $(x_i, y_i)_{i \in \{1, \dots, N\}}$ sont simultanément répercutées sur les descripteurs de Fourier associés $(X_k, Y_k)_{k \in \mathbb{Z}}$ (cf. §III.1 du chap. II.2). Néanmoins, le problème du décalage du point de départ à la paramétrisation doit impérativement précéder la procédure d'appariement de deux contours sous peine d'incohérence. En effet, dans le cas général, en utilisant une paramétrisation adaptée au cas affine, on a (cf. §III.2 du chap. II.2) :

$$\begin{bmatrix} U_k \\ V_k \end{bmatrix} = w^k A \begin{bmatrix} X_k \\ Y_k \end{bmatrix}$$

avec

$$w = \exp(-2j\pi\tau / T_0), \quad \tau \text{ représente le décalage}$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ la transformation géométrique appliquée au contour}$$

En somme, l'appariement nécessite l'estimation de cinq paramètres : τ , a , b , c et d .

II.3. Estimation des paramètres

Dans la suite, nous désignons par $\Gamma_k^0 = \begin{bmatrix} X_k \\ Y_k \end{bmatrix}_{k=1 \dots K}$ le contour prototype et par

$\Gamma_k = \begin{bmatrix} U_k \\ V_k \end{bmatrix}_{k=1 \dots K}$ le contour de l'ombre segmentée étant donné que les seuls coefficients d'indice positif suffisent à caractériser le contour.

II.3.1. Estimation de τ

S'affranchir du problème de décalage du point de contour est une condition nécessaire préalable à la procédure effective d'appariement. Cette dernière consiste à trouver, indépendamment du décalage, la transformation qui permet de passer du contour prototype au contour de l'ombre. Dans le cas général, il vient :

$$\Gamma_k = w^k \times A \times \Gamma_k^0 \quad \forall k > 0 \quad \text{où } w = \exp\left(-j2\pi \frac{\tau}{T_0}\right)$$

La dépendance en A peut être éliminée en passant par les A -invariants Q_k définis par Arbter [ARBT-89] (cf. §III.2 du chap. II.2). Au paragraphe III.2.3. du chap. II.2, on a établi une égalité simple entre les descripteurs affine-invariants :

$$Q_k = w^{k-p} Q_k^0$$

Finalement, l'estimation de τ revient à minimiser la somme $S(\tau) = \sum_k (Q_k - w^{k-p} Q_k^0)^2$ telle que :

$$\tau = \underset{\tau}{\operatorname{Arg\,min}} S(\tau) = \underset{\tau}{\operatorname{Arg\,min}} \sum_k (Q_k - w^{k-p} Q_k^0)^2$$

Notons que pour des contours bruités, il est recommandé de choisir p de sorte que $\det[\Gamma_p, \Gamma_p^*]$ soit aussi grand que possible (généralement, p est égal à un).

Rappelons que $w = -j2\pi \frac{\tau}{T_0}$ et que le stockage de T_0 est donc nécessaire lors du calcul des descripteurs. Pratiquement, en supposant que les paramétrisations des deux contours partent d'un point tel que le décalage est faible, on visualise la somme $S(\tau)$ pour τ allant de $-\alpha T_0$ à αT_0 où α peut dépendre du prototype utilisé. S'il existe plusieurs minima, on sélectionne le plus faible bien que cette règle ne donne pas toujours un résultat optimal. S'il n'existe aucun minimum, le décalage est fixé à zéro par défaut.

Remarques :

- Cette estimation n'est correcte que s'il existe une transformation affine entre les deux contours, en d'autres termes que ces contours sont de même nature.
- L'estimation proposée du décalage a un double avantage : d'une part, elle est robuste car elle est obtenue à partir de l'ensemble des descripteurs, d'autre part, elle est valable quelle que soit la transformation affine existant entre les deux contours. Cette technique est à confronter à quelques précédents travaux ne prenant pas en compte l'un ou l'autre de ces deux points tels que [PERS-86] [MARQ-97] [ELOI-99].

II.3.2. Estimation de A

Après s'être affranchi du problème de décalage du point de départ, il vient :

$$\Gamma_k = A \times \Gamma_k^0 \quad \text{soit} \quad \begin{bmatrix} U_k \\ V_k \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X_k \\ Y_k \end{bmatrix}, \quad \forall k > 0$$

On désire estimer les paramètres a , b , c et d de cette matrice de manière à **rapprocher le contour prototype du contour de la forme segmentée**. La relation ci-dessus lie des données complexes entre elles. Elle se décompose en deux autres égalités liant cette fois les parties réelles d'une part et imaginaires d'autre part :

$$\begin{bmatrix} \operatorname{Re}(U_k) \\ \operatorname{Re}(V_k) \end{bmatrix} = A \times \begin{bmatrix} \operatorname{Re}(X_k) \\ \operatorname{Re}(Y_k) \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} \operatorname{Im}(U_k) \\ \operatorname{Im}(V_k) \end{bmatrix} = A \times \begin{bmatrix} \operatorname{Im}(X_k) \\ \operatorname{Im}(Y_k) \end{bmatrix} \quad \forall k > 0$$

En prenant, l'ensemble des indices $k \leq K$, on établit l'égalité suivante liant les deux ensembles de descripteurs :

$$\begin{bmatrix} \operatorname{Re}(U_1) & \operatorname{Re}(U_2) & \cdots & \operatorname{Re}(U_K) & \operatorname{Im}(U_1) & \operatorname{Im}(U_2) & \cdots & \operatorname{Im}(U_K) \\ \operatorname{Re}(V_1) & \operatorname{Re}(V_2) & \cdots & \operatorname{Re}(V_K) & \operatorname{Im}(V_1) & \operatorname{Im}(V_2) & \cdots & \operatorname{Im}(V_K) \end{bmatrix} \\ = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} \operatorname{Re}(X_1) & \operatorname{Re}(X_2) & \cdots & \operatorname{Re}(X_K) & \operatorname{Im}(X_1) & \operatorname{Im}(X_2) & \cdots & \operatorname{Im}(X_K) \\ \operatorname{Re}(Y_1) & \operatorname{Re}(Y_2) & \cdots & \operatorname{Re}(Y_K) & \operatorname{Im}(Y_1) & \operatorname{Im}(Y_2) & \cdots & \operatorname{Im}(Y_K) \end{bmatrix}$$

$$\equiv N = A \times M$$

Ce système linéaire surdimensionné peut être résolu immédiatement par la *méthode pseudo-inverse* ou *méthode des moindres carrés* :

$$A = NM^T (MM^T)^{-1}$$

II.4. Mesure de similarité

Deux contours ne peuvent réellement être appariés que s'ils satisfont à un certain critère. Aussi, convient-il de quantifier la similarité entre les deux contours en y associant une certaine grandeur.

Rappelons que l'estimation du décalage τ est effectuée sous l'hypothèse d'une transformation affine **générale** (cf. §II.3.1). Cependant, nous le verrons au paragraphe suivant, l'estimation des paramètres a , b , c et d doit être contrainte dans le cas des objets à symétrie radiale sous peine de confusion entre objets de formes tridimensionnelles voisines (cf. § suivant). Ces deux étapes d'estimation peuvent donc ne pas être compatibles (au sens que la transformation sous-jacente à la première étape peut ne pas satisfaire les contraintes imposées à la deuxième étape) et rendent hasardeux le calcul de la distance euclidienne entre les vecteurs caractéristiques formés des descripteurs de Fourier. Par précaution, nous proposons une mesure indépendante des descripteurs basée sur l'étiquetage des pixels.

On note $\tilde{P}$ le prototype P transformé et F la forme segmentée. La similarité entre les formes $\tilde{P}$ et F est basée sur l'ensemble des pixels étiquetés 'ombre' des deux formes. On décide ainsi de comptabiliser le nombre de points étiquetés 'ombre' communs aux deux formes lorsqu'elles sont transposées dans un même repère dont l'origine coïncide avec leurs centres d'inertie.

Nous raisonnons pixel par pixel en pénalisant ceux qui n'ont pas la même étiquette (0 pour les pixels appartenant à l'ombre et 1 sinon) entre la forme segmentée *pleine* (c'est-à-dire pixels du contour et pixels intérieurs) et le prototype déformé *plein*.

Soit N_o , le nombre total de pixels étiquetés 'ombre' dans l'image contenant à la fois la forme segmentée et le prototype déformé centrés et soient e_p l'étiquette associée au prototype et e_f l'étiquette associée à la forme segmentée, la mesure de similarité est donnée par :

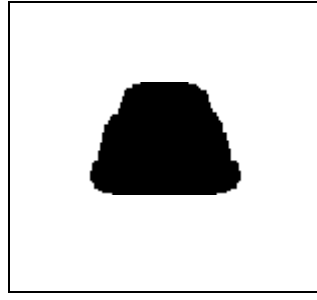
$$S(\tilde{P}, F) = \frac{1}{N_o} \sum_{N_o} \delta(e_p, e_f) = \frac{N_{pf}}{N_o}$$

où $\delta(.)$ est la fonction de Kronecker³

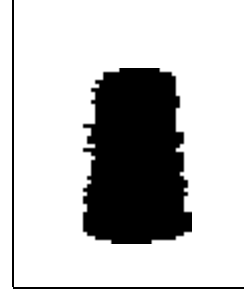
³ La fonction de Kronecker est définie par :

$$\begin{cases} \delta(e_p, e_f) = 1 & \text{si } e_p = e_f \\ \delta(e_p, e_f) = 0 & \text{si } e_p \neq e_f \end{cases}$$

Méthode : on compose une image à partir des deux images suivantes :

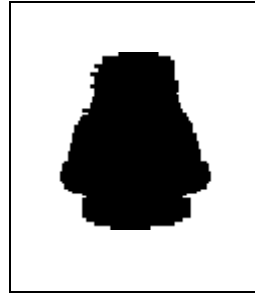


Prototype déformé I_1

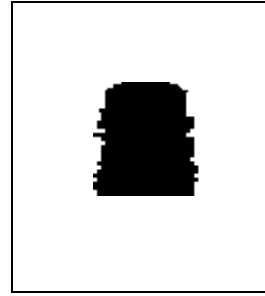


Ombre détectée I_2

Par coïncidence des centres d'inertie de ces ombres, on obtient l'image union. L'intersection des deux ombres donne alors l'aire que constituent les N_{pf} pixels communs,



*Image de l'union pour test sur les pixels
 $I_1 \cup I_2$*



*Image de l'intersection
 $I_1 \cap I_2$*

Pour illustration, nous avons visualisé l'évolution de S dans le cas de figure suivant (cf. Figure 31) :

A partir d'un appariement réussi, nous souhaitons voir comment évolue S de part et d'autre de cette bonne approximation (dixième image sur la Figure 31). Le format de l'ombre est maintenu constant ($inertie = N(forme)$). Le format du prototype évolue en appliquant une homothétie en plus de la transformation estimée. La transformation globale est donc la composition de A et d'une homothétie h de facteur échelle allant de 0.1 à 2 par pas de 0.1.

Observation :

Le nombre de pixels communs entre les deux formes c'est-à-dire $N_{pf} = \sum_{N_o} \delta(e_p, e_f)$ rapportés au nombre total N_o de pixels étiquetés 'ombre' est dissymétrique. Ceci illustre parfaitement les deux temps de l'expérience,

- un premier temps où le prototype est totalement inclus dans la forme segmentée, i.e. une évolution en $y = \frac{x}{N(forme)}$
- un deuxième temps où cette fois, le prototype englobe entièrement la forme segmentée, i.e. une évolution en $y = \frac{N(forme)}{x}$

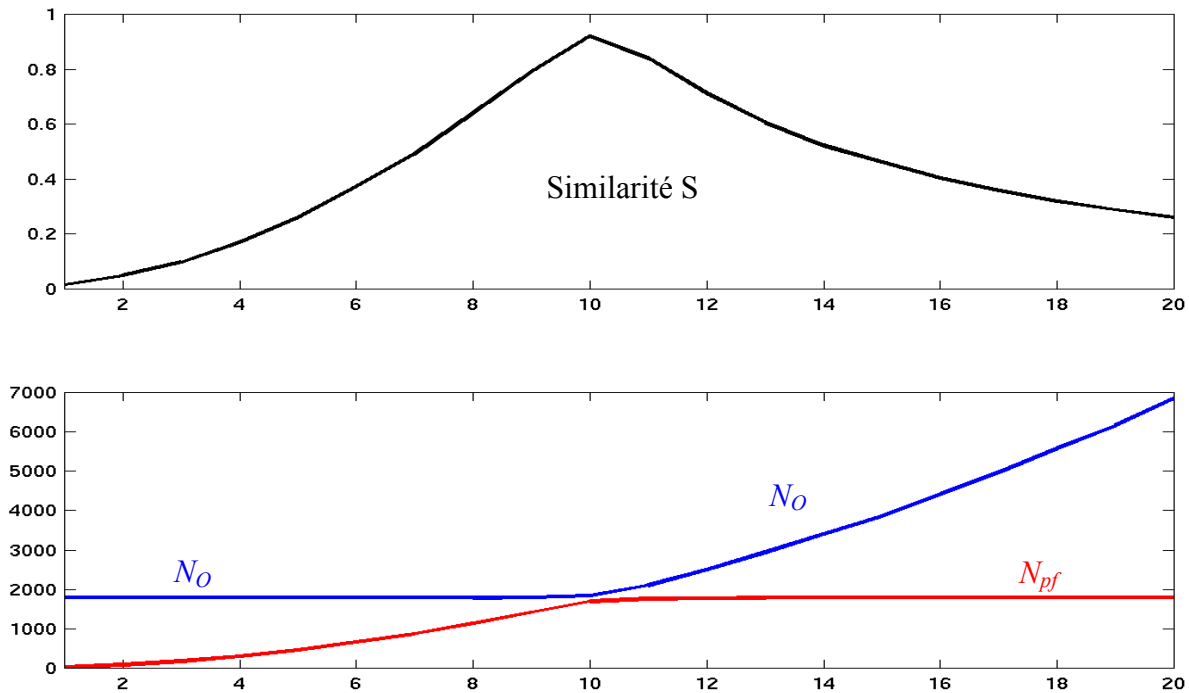


Figure 31 - Evolution de la similarité de part et d'autre de la bonne approximation (dixième image)

II.5. Mise en œuvre pour les images sonar

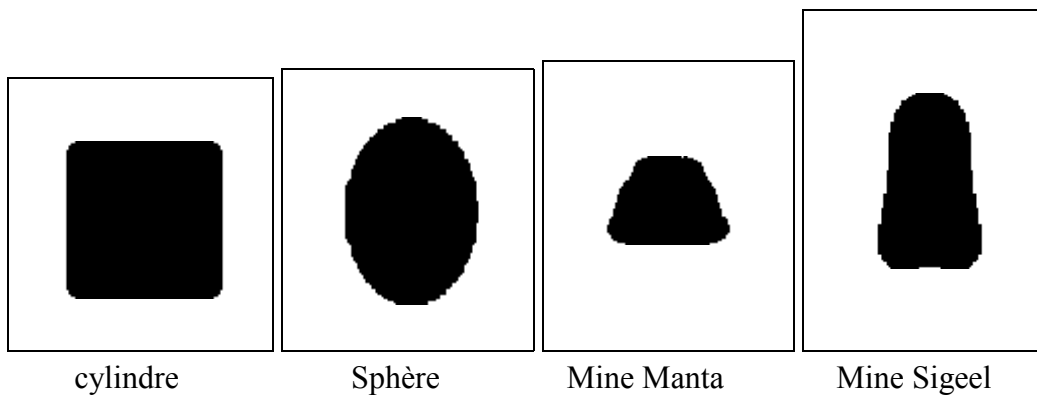


Figure 32 – Prototypes utilisés pour l'appariement (un prototype par classe)

Les prototypes utilisés sont illustrés Figure 32 :

- Cylindre et Sphère ont été simulés sous Matlab puis lissés pour avoir des contours plus doux
- Mines Manta et Sigeel ont été obtenues à partir de plusieurs exemples de la base d'images simulées (cf. annexe G).

- **Normalisation de l'image**

Les prototypes utilisés correspondent à une normalisation de l'image, comme explicité au §II.1. du chap. II.2, à l'exception du prototype cylindre. En effet, il s'agit de rester le plus fidèle possible, en proportion, aux ombres portées par les mines que nous souhaitons identifier. Les prototypes utilisés, à l'exception de la classe des cylindres, ont une élongation correspondant à une normalisation en rasant de 45° . Pour le cas du cylindre, l'ombre portée a une forme différente selon l'angle de vue et justifie le choix d'un prototype de forme carrée. Grâce à cette normalisation, l'estimation gagne en efficacité et en rapidité.

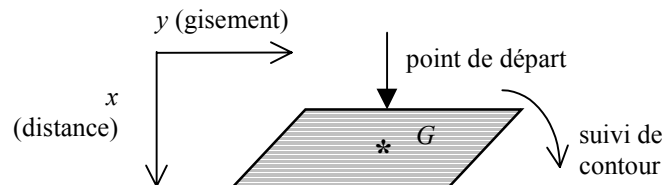
- **Prise en compte de la symétrie**

Des contraintes s'imposent si l'on ne veut pas confondre les deux mines tronconiques ayant toutes deux une ombre portée de forme trapézoïdale. Pour l'ensemble des objets à symétrie radiale, l'éventail des transformations géométriques admissibles est réduit à l'homothétie ($b=c=0$). Toutefois, afin de corriger les éventuels défauts survenus lors de l'étape de normalisation de l'image (défauts dus aux problèmes d'arrondis à l'entier), on autorise de faibles élongations. Ces nouvelles contraintes se traduisent par $|a-d| \leq 0.05$ (soit une élongation de 5% maximum).

Il revient à estimer un cas particulier de transformation affine c'est-à-dire la composition d'une homothétie et d'une faible élongation.

- **Précaution pour le suivi de contour**

Une estimation correcte de la transformation affine existant entre deux silhouettes n'est possible que s'il n'y pas de décalage du point de départ lors de la procédure de suivi de contour. Afin de limiter ce décalage à une valeur faible, le premier point sélectionné sur chacun des contours (prototype et contour de l'ombre segmentée) est celui d'abscisse y_G où $G=(x_G, y_G)$ désigne le centre de gravité de la forme pleine (points de contour et points intérieurs) comme l'explique la figure suivante :



Cette précaution permet d'autoriser une incertitude de valeur $\alpha \times (\text{longueur du contour})$ sur le point de départ (cf. §II.3.1). En particulier, on a choisi une faible valeur α égale à $1/15$ pour les objets à symétrie radiale et une valeur de $1/8$ pour le prototype cylindre du fait des transformations géométriques possibles (biais notamment).

- **Précision du contour**

Plus le degré de décomposition K en série de Fourier est élevé, plus le contour est précisément décrit en faisant apparaître les aspérités du contour. Cependant, pour limiter les temps de calcul et assurer un traitement robuste en cas de contours bruités, nous avons choisi K égal à 5.

Les principales étapes de l'algorithme implémenté sont résumées sur la Figure 33. Nous faisons le choix d'une segmentation markovienne (contour moyennement bruité grâce à la régularisation).

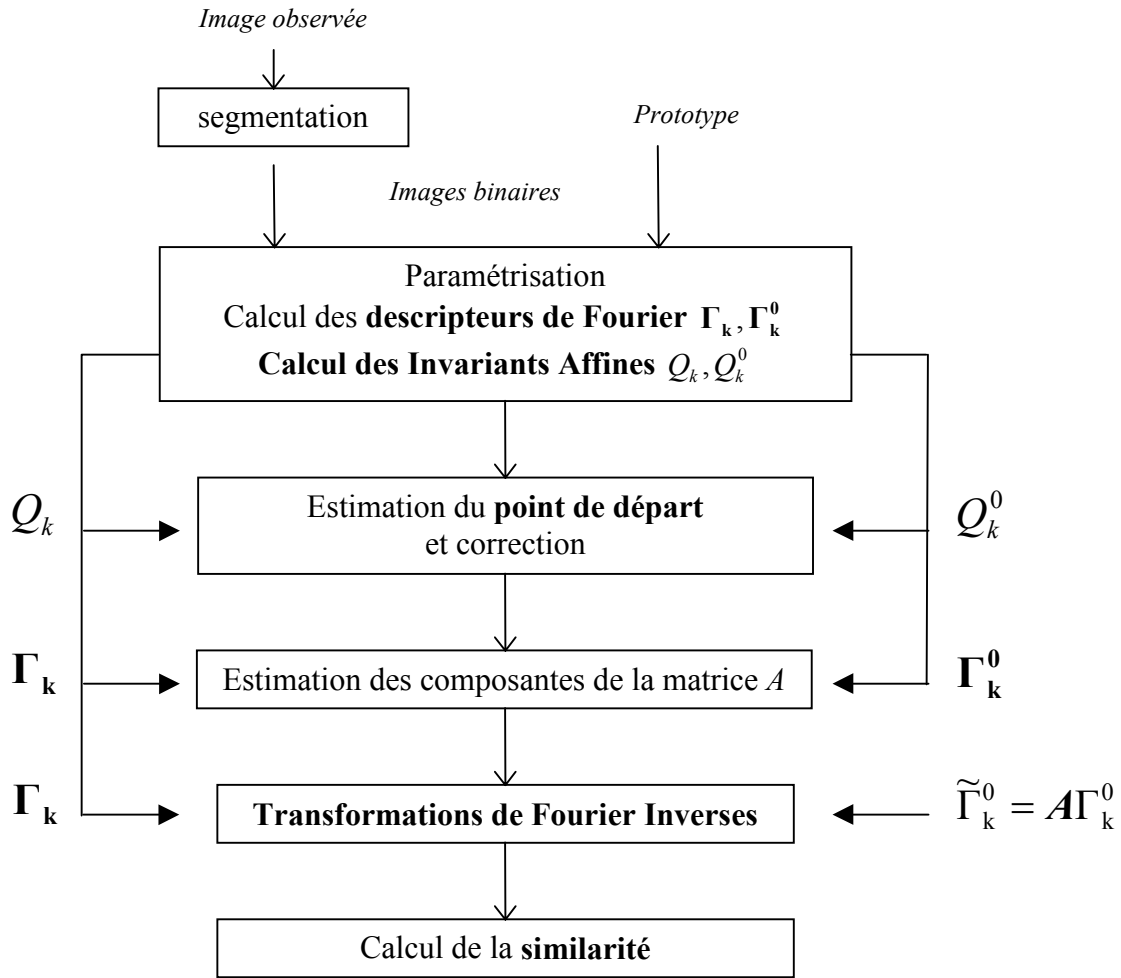


Figure 33- Algorithme d'appariement de silhouettes

II.6. Performances

Comme nous l'avons précisé au §II.4, la mesure de la similarité S entre les deux contours appariés vise à prendre une décision. En d'autres termes, la valeur de cette mesure conditionnellement à la nature du prototype choisi doit permettre, ou bien de le valider (l'objet ciblé est alors identifié), ou bien de le rejeter (l'objet ciblé n'est vraisemblablement pas de la même nature que le prototype).

Nous avons testé 170 images par classe issues de la base de données décrites dans l'annexe G. L'appariement est cependant plus difficile pour des mines à géométrie plus complexes comme la mine Rockan. Nous avons donc écarté cette classe d'objets dont les silhouettes des ombres portées ne peuvent se rapprocher d'une silhouette prototype unique à une transformation affine près.

Pour chaque image, on a recherché la transformation existant entre la forme segmentée et chacun des quatre prototypes puis calculé la mesure de similarité. Le Tableau 9 donne la matrice de confusion obtenue.

Tableau 9 – Matrice de confusion (taux de bonne classification moyen = 97.5%)

exemples \ classes	cylindre	sphère	mine Manta	mine Sigeel
cylindres	95.3	0.6	0	4.1
sphères	0.6	99.4	0	0
mines Manta	0	0	100	0
mines Sigeel	4.1	0.6	0	95.3

La visualisation sur la Figure 34 des mesures de similarités permet de mettre en évidence les ambiguïtés et les appariements erronés suite à une mauvaise estimation du décalage. Le premier graphe correspond à l'estimation sur mines cylindriques, la seconde sur mines sphériques, la troisième sur mines Manta et la quatrième sur mines Sigeel. Pour chacun de ces graphes, on notifie la courbe correspondant aux performances obtenues par estimation à partir du prototype exact (bonne classification) et la (les) courbe(s) correspondant au(x) prototype(s) portant ambiguïté lorsque le cas se produit.

L'observation de ces courbes permet de noter que, globalement, les formes segmentées sont correctement classifiées (mesure de similarité la plus forte).

Nous notons cependant que les valeurs maximales de S diffèrent selon les types d'objets et que la confusion entre les classes est fonction de la géométrie de l'objet. Ces remarques nous amènent à penser que si la décision est effectuée à partir d'un **seuil** sur la valeur de la similarité, celui-ci doit dépendre du type d'objet.

Pour les mines cylindriques, la confusion apparaît périodiquement (toutes les neuf images, i.e. un écart de neuf prises de vue par pas de 10 degrés autour de l'objet) avec les mines sphériques et Sigeel. Ceci correspond en effet à un point de vue non caractéristique du cylindre (le long de la génératrice).

Par ailleurs, la flexibilité du prototype cylindrique, induite par le fait que l'appariement s'effectue sans contrainte, entraîne quelques confusions pour les autres classes.

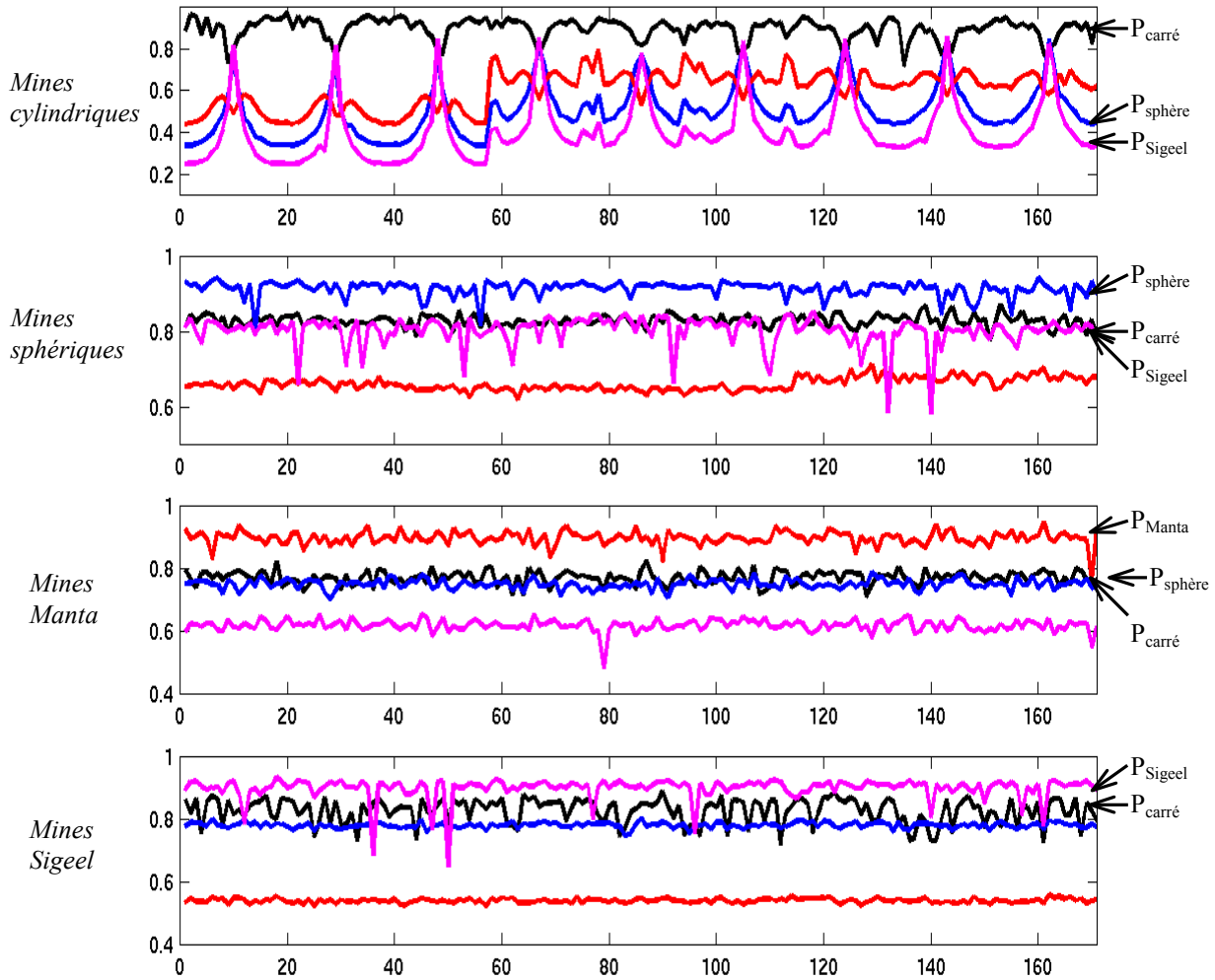
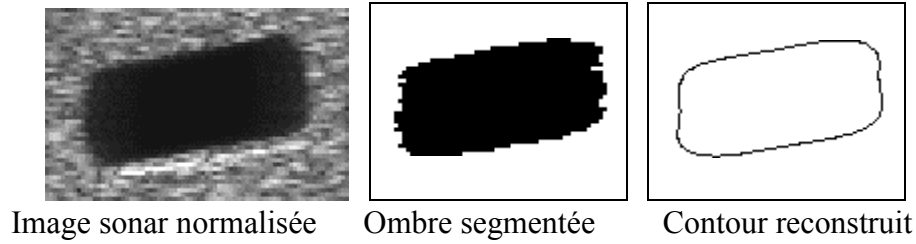


Figure 34 – Pour chaque individu (170 par classe), visualisation des mesures de similarité obtenues avec chacun des quatre prototypes

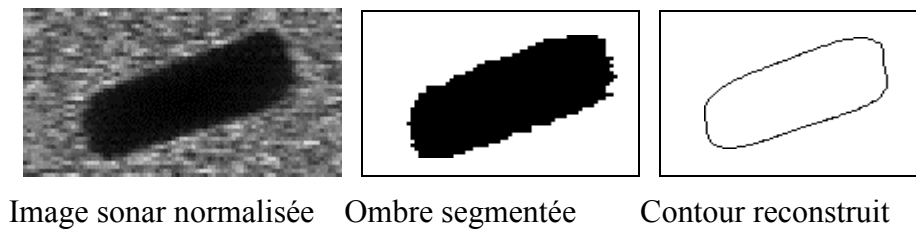
II.7. Illustrations

Figure 35 et Figure 36 montrent quelques appariements pour des images réelles et simulées. Chaque exemple est composé de quatre images illustrant les principales étapes : l'image sonar brute normalisée, l'image binaire correspondante, le contour reconstruit par transformée de Fourier inverse (cf. annexe D) et la superposition des deux contours appariés.



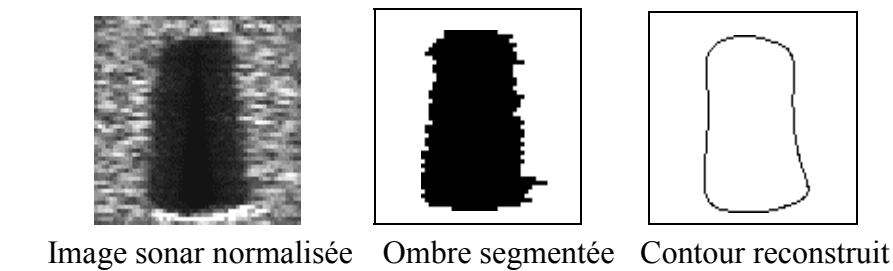
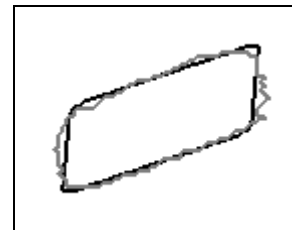
Contours appariés: le carré prototype déformé et l'ombre segmentée

avec $A = \begin{bmatrix} 0.666 & -1.189 \\ -0.053 & 1.392 \end{bmatrix}$ et $S = 0.916$



Contours appariés: le carré prototype déformé et l'ombre segmentée

avec $A = \begin{bmatrix} 0.628 & -0.455 \\ -0.057 & 1.352 \end{bmatrix}$ et $S = 0.891$

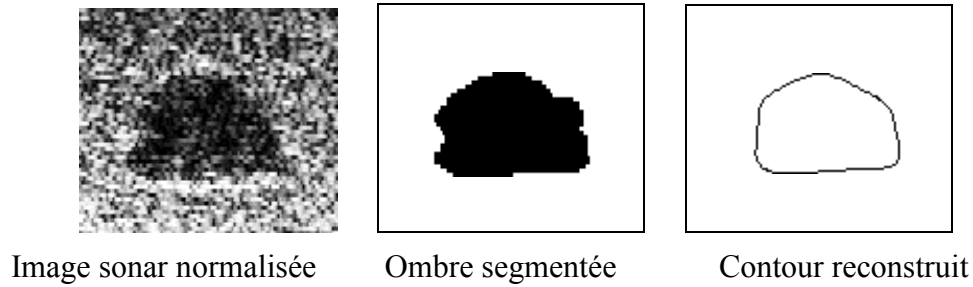


Contours appariés: le trapèze prototype déformé et l'ombre segmentée

avec $A = \begin{bmatrix} 0.855 & 0 \\ 0 & 0.905 \end{bmatrix}$ et $S = 0.883$

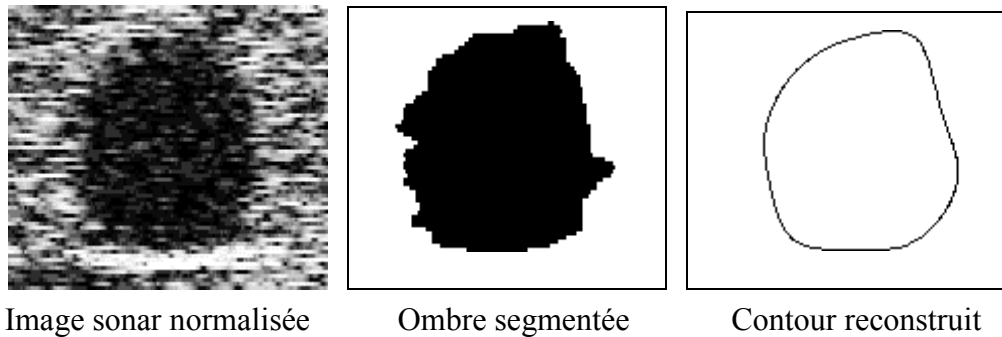


Figure 35 – Appariements sur images simulées: deux mines cylindriques (biais) et une mine Sigeel



Contours appariés: le trapèze prototype déformé et l'ombre segmentée

avec $A = \begin{bmatrix} 1.178 & 0 \\ 0 & 1.228 \end{bmatrix}$ et $S = 0.863$



Contours appariés: l'ellipse prototype déformé et l'ombre segmentée

avec $A = \begin{bmatrix} 1.271 & 0 \\ 0 & 1.321 \end{bmatrix}$ et $S = 0.860$

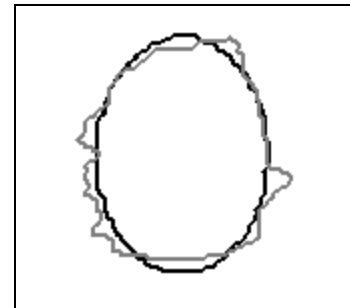


Figure 36 - Appariements sur images réelles : mines Manta et sphérique

II.8. Conclusion

Une méthode robuste d'appariement de silhouettes a été implémentée avec succès en offrant la possibilité de rapprocher la forme de l'ombre segmentée d'un modèle prototype.

Les propriétés de la décomposition en série de Fourier effectuée sur chacune des coordonnées des points de contour ont été avantageusement exploitées. D'une part, l'estimation des quatre paramètres de la transformation affine (translation exceptée) est facilement obtenue au sens des moindres carrés grâce à la linéarité des descripteurs. D'autre part, un point critique relatif au décalage du point de départ est initialement écarté par l'introduction d'invariants affines dérivés des mêmes descripteurs de Fourier.

L'estimation des différents paramètres permet de superposer les deux silhouettes appariées. Outre le résultat visuel, une mesure objective de similarité a été introduite de manière à quantifier la qualité de l'appariement. Nous disposons ainsi d'un moyen de classification d'un objet par rapprochement de la silhouette de son ombre portée et de l'un des prototypes proposés.

III. Segmentation orientée classification

III.1. Introduction

Les modèles déformables ont été diversement appliqués à la reconnaissance de formes, la segmentation, le suivi d'objet en mouvement,... [COHE-93] [KERV-96] [JAIN-96] [BOSS-97] [ZHON-98] [LAM-98] [CHES-99] [MIGN-00]. Ce procédé est utilisé dans deux approches largement discutées en traitement d'images : les modèles déformables paramétriques et les contours actifs.

D'une part, le principe des **modèles déformables paramétriques** est de disposer d'une base de modèles prototypes ou gabarits ('templates') décrits par un nombre donné de paramètres auxquels on associe un ensemble de déformations. Le choix de tels modèles est donc fortement conditionné par les connaissances a priori sur les objets susceptibles d'apparaître. Le modèle linéaire interagit de manière dynamique avec les particularités de l'image en faisant varier ses paramètres : son degré de liberté correspond au nombre de paramètres de déformations introduits. Leur estimation au sens du MAP est obtenue en lui associant une densité de probabilité a posteriori $P(\Theta/Y)$ conditionnellement aux observations (en fait la carte de segmentation en deux classes ombre/réverbération de fond). En utilisant la règle de Bayes, cette densité s'obtient en combinant deux termes, relatifs à l'énergie externe et interne [JAIN-96] :

$$P(\Theta/Y) = \frac{P(\Theta, Y)}{P(Y)} = \frac{P(Y/\Theta) \cdot P(\Theta)}{P(Y)} = \kappa_y \cdot P(Y/\Theta) \cdot P(\Theta)$$

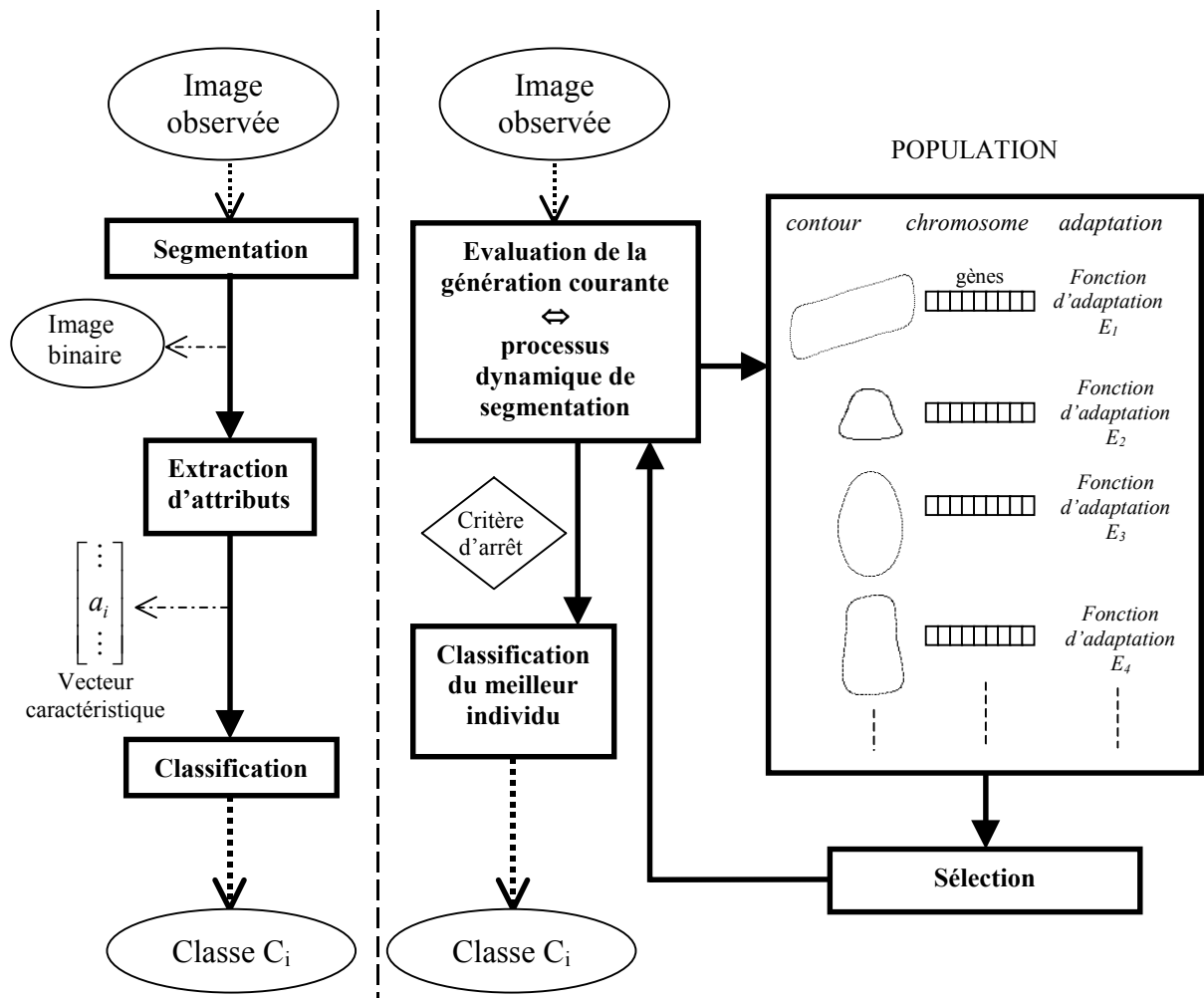
- $P(Y/\Theta)$ est le terme de *vraisemblance des observations conditionnellement aux déformations*. Il a pour effet d'attirer le modèle vers les traits remarquables de l'image.
- le second terme correspond aux *connaissances a priori* $P(\Theta)$ qui caractérisent le modèle probabiliste associé au modèle. Cette information a priori comprend la description du modèle et l'ensemble des déformations admissibles. Il est donc indépendant des données de l'image.

Le **contour actif** ou '*snake*' est d'autre part un modèle élastique dont un exemple a été donné au paragraphe III.1.2. du chap. II.1. Les déformations du contour sont dues d'une part aux contraintes internes et d'autre part aux interactions avec l'image [KASS-88]. Le critère selon lequel le processus de déformation fait évoluer le contour a été initialement associé à la recherche d'une minimisation d'une fonctionnelle généralement constituée de deux termes d'énergie qui doivent s'équilibrer : l'un est un terme de régularisation caractérisant le comportement mécanique de la courbe (énergie interne), l'autre attire la courbe vers les structures pertinentes de l'image que l'on recherche (énergie externe). Jain et al. montrèrent que, d'un point de vue bayésien, les énergies interne et externe peuvent respectivement s'apparenter à une densité de probabilité a priori et à une vraisemblance [JAIN-98].

La méthode proposée tire profit des avantages des deux approches décrites ci-dessus. Les modèles déformables paramétriques, d'une part, permettent de classifier la forme segmentée en l'approchant d'un prototype connu a priori. Les contours actifs, d'autre part, visent à segmenter la forme en extrayant son contour à partir de l'image brute en niveaux de gris. Nous proposons une technique originale utilisant aussi la déformation de modèles dans laquelle **la segmentation est orientée classification**. En d'autres termes, le domaine des solutions possibles est limité mais supposé représentatif des cas susceptibles de se présenter,

moyennant des déformations du même type que celles rencontrées pour les modèles déformables paramétriques. L'exploration de ce domaine, appelé « population initiale », a pour finalité de proposer la solution la mieux adaptée. Le contour à estimer est uniquement caractérisé par **K descripteurs de Fourier**. Tout en assurant un lissage du contour (on s'affranchit alors du terme de régularisation), ces paramètres permettent, par transformée inverse, de reconstruire le contour le mieux adapté, c'est-à-dire celui qui minimise, au sens du MV, une certaine énergie liée à la statistique de l'image (équivalente aux forces de l'image) [FIGU-97].

Nous manipulons ainsi des données brutes **sans devoir passer par une étape préliminaire et irréversible de segmentation** (partition des pixels de l'image en deux classes 'ombre' et 'réverbération de fond'). Cette approche s'oppose à l'approche classique séquentielle 'segmentation → extraction des attributs → classification' comme l'explique la Figure 37. L'ultime étape de classification intervient dans le rapprochement de ces descripteurs estimés avec une version étendue du domaine des solutions possibles initial (que l'on nomme « population de référence » en lien avec le contexte choisi de classification supervisée).



**Figure 37 - Segmentation orientée classification
opposée au processus de segmentation classique séquentiel**

La stratégie d'optimisation la plus appropriée à notre problème est le développement d'un **algorithme génétique** [DAVI-91] [MICH-94] [PAL-96]. Cette technique d'exploration

stochastique (indépendante de l'initialisation) a l'avantage d'optimiser massivement plusieurs chaînes de paramètres (individus) par le biais d'opérateurs génétiques à la différence de la technique du recuit simulé qui optimise une seule chaîne en évaluant ses configurations voisines. En outre, elle opère en un nombre d'itérations fonction du problème à optimiser et non d'un paramètre fixé a priori tel la température pour le recuit simulé [DAVI-87].

Les algorithmes génétiques ont déjà été utilisés pour partitionner l'image en plusieurs régions [SEET-91] [HASE-99]. A cette segmentation par région, nous préférons une approche par contour. Dans notre cas en effet, la segmentation, limitée à deux régions, n'est pas une fin en soi mais doit permettre en finalité de classifier l'objet ciblé par une extraction d'attributs discriminants de la frontière entre ces régions. Cette frontière correspond au contour de l'ombre acoustique portée et peut être décrite de manière unique par des **descripteurs de Fourier normalisés** estimés par l'algorithme génétique. L'information région constitue l'attache aux données prise en compte par l'intermédiaire du critère d'adaptation basé sur la **statistique de l'image**. A la différence de la classification des ombres portées en imagerie sonar également proposée par Mignotte et al., l'optimisation vise ici à estimer le contour et non les seules déformations qui l'affectent. Aucune hypothèse n'est faite a priori sur la nature de l'objet [MIGN-00].

Le Tableau 10 synthétise les caractéristiques des trois approches énoncées. Il apparaît bien que l'approche par algorithme génétique est *intermédiaire* entre les modèles déformables paramétriques et les contours actifs. Sans idée précise de la nature de l'objet ciblé (comme pour les contours actifs), on fait l'hypothèse qu'il appartient à un ensemble connu d'objets (équivalents des prototypes énoncés utilisés pour les modèles déformables paramétriques) dont on a établi un panorama des ombres portées pour différents points de vue soient N_p solutions ou individus. De plus, comme dans le cas des contours actifs, on travaille directement avec les données sonar brutes.

Tableau 10 - Synthèse des caractéristiques des trois approches énoncées

<i>Type de traitement</i>	<i>Modèles déformables paramétriques</i>	<i>Algorithme génétique</i>	<i>Contours actifs</i>
Données	image segmentée en deux classes : ombre / réverbération de fond	image brute	image brute
Espace des solutions possibles	un prototype par classe	N_p individus	espace infini
Critère d'évolution	fonction objective bayésienne	fonction d'adaptation	fonctionnelle
Objectif visé	classification	segmentation orientée classification	segmentation

III.2. Algorithmes génétiques [DAVI-91] [MICH-94] [PAL-96]

III.2.1. Introduction

Les algorithmes génétiques sont des méthodes d'exploration stochastique fondées sur les mécanismes de la sélection naturelle et de la génétique. Ils en tirent leurs principales

caractéristiques : principe de survie des individus les mieux adaptés (par évaluation de la 'performance' d'un descendant donné) et échanges d'informations pseudo-aléatoires entre individus (par opérations génétiques). Ainsi, à chaque génération, de nouveaux individus sont créés par le processus de sélection des parents compte tenu de leur adéquation au problème et de leur reproduction selon les principes généraux de la génétique que sont les croisements et les mutations.

III.2.2. Formulation et terminologie

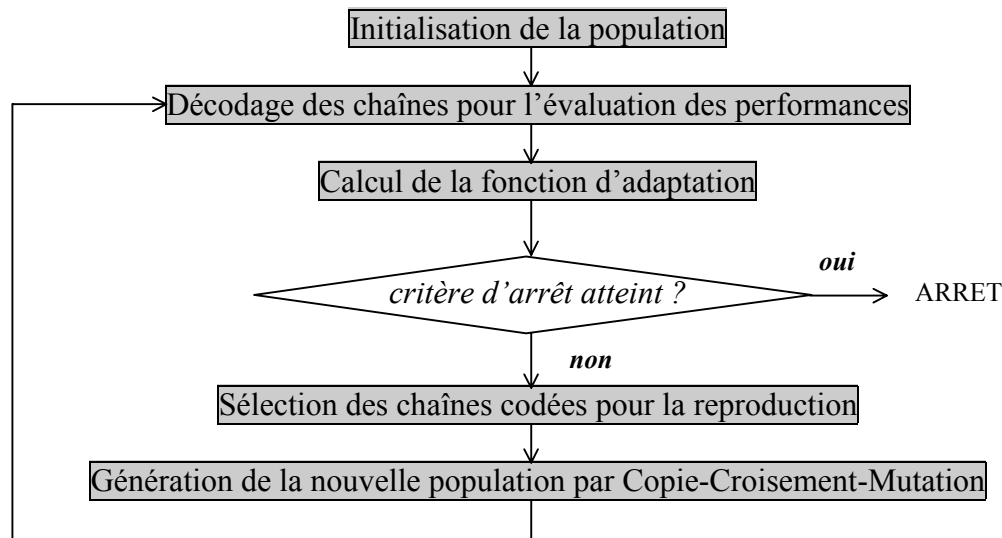


Figure 38 - Principe général des algorithmes génétiques

Les algorithmes génétiques procèdent de manière itérative à travers l'évolution d'une *population* de génération en génération. Chaque *individu* symbolise une solution de l'espace de recherche. Il est caractérisé par son *chromosome* constitué d'une liste finie de *gènes* (paramètres) pouvant prendre des valeurs dans un alphabet donné. A chaque génération (itération), une *fonction d'adaptation* juge de l'adéquation de chaque individu aux données. Cette fonction constitue le critère de *sélection* pour la reproduction. Les individus les mieux adaptés sont soumis à deux types d'opérations génétiques : les *croisements* et les *mutations*. Dans l'opération élémentaire de croisement, les chromosomes sont regroupés par paire et les chaînes ainsi appariées sont recombinaées. La mutation, accidentelle mais nécessaire afin de diversifier le matériel génétique, résulte classiquement d'une modification aléatoire de faible probabilité de la valeur d'un caractère de la chaîne. Dans notre cas, il nous a paru utile d'introduire en outre une mutation opérant une transformation globale du chromosome (cf. §III.3.4.B). La Figure 38 présente le principe général des algorithmes génétiques.

En somme, l'implémentation d'un algorithme génétique nécessite la détermination de sept points fondamentaux (cf. §III.3) :

- ↪ la représentation chromosomique
- ↪ la création de la population initiale
- ↪ la définition de la fonction d'adaptation
- ↪ le critère de sélection
- ↪ les opérateurs génétiques réglant la reproduction
- ↪ la réinsertion
- ↪ le critère d'arrêt

III.2.3. Application à l'estimation des descripteurs de Fourier

Conformément au principe des algorithmes génétiques, nous partons d'un ensemble d'individus ou contours prototypes constituant la population initiale. Chaque individu (ou contour) est caractérisé par 2K descripteurs de Fourier normalisés que constitue son chromosome (cf. §III.1 du chap. II.2).

La partition de l'image est effectivement réalisée par estimation du contour qui partage l'image de manière optimale au sens d'un critère statistique en deux zones dont les pixels obéissent à des lois indépendantes [FIGU-97] [CHES-99] (cf. §II.2 du chap. II.1). Grâce à ce critère énergétique, l'estimation du contour peut s'effectuer sur les données de l'image sonar brute (sans traitement préliminaire) présentant l'ombre portée d'un objet immergé. L'algorithme génétique vise à optimiser cette énergie (fonction d'adaptation), c'est-à-dire estimer les descripteurs de Fourier associés au contour qui correspond à une partition optimale des pixels (soit la segmentation désirée).

III.3. Segmentation orientée classification par optimisation génétique

L'implémentation algorithmique a été facilitée par l'utilisation de la boîte à outils nommée **GAOT** pour **Genetic Algorithms for Optimization Toolbox** [Houc-95]. Grâce à sa modularité, les modules de bases ont été enrichis de nouvelles routines spécifiques à notre problème.

III.3.1. Codage des individus

III.3.1.A. Description du contour

Comme nous l'avons déjà exprimé, les chromosomes que constituent les vecteurs solutions sont alors essentiellement constitués de descripteurs de Fourier. La transformée de Fourier effectuée distinctement sur les abscisses x_i et les ordonnées y_i des N points de contour équirépartis est la suivante (cf. §III.1 du chap. II.2) :

$$\left\{ \begin{array}{l} \tilde{X}_k = X_k \frac{C_1^{-k}}{|C_1|^{-k}} = X_k \frac{(X_1 + jY_1)^{-k}}{|X_1 + jY_1|^{-k}} \\ \tilde{Y}_k = Y_k \frac{C_1^{-k}}{|C_1|^{-k}} = Y_k \frac{(X_1 + jY_1)^{-k}}{|X_1 + jY_1|^{-k}} \end{array} \right. \quad \text{avec } C_k = X_k + jY_k$$

La seule connaissance des coefficients d'indices positifs suffit à définir complètement la transformée.

III.3.1.B. Initialisation du contour

La reconstruction du contour s'effectue autour du pixel dont les coordonnées sont les coordonnées moyennes des points de contour de l'ombre détectée c'est-à-dire (X_0, Y_0) . La recherche de ce point ou **marqueur** n'est pas triviale dans le sens où nous ne connaissons justement pas le contour exact de cette ombre. Pratiquement, la recherche de ce point est effectuée au voisinage d'un point initial (limité empiriquement à plus ou moins dix pixels

dans chacune des directions) et guidée par la valeur de l'évaluation obtenue pour chacune des positions du contour dans l'image.

A la première génération, les coordonnées du marqueur sont fixées par l'utilisateur. La position optimale du marqueur (au sens de l'évaluation choisie) est ensuite transmise et modifiée de génération en génération comme les autres gènes mais suivant un processus d'évolution notablement différent. L'individu a la capacité de s'auto-adapter : avant évaluation, il augmente ses chances de survie en jouant sur ces deux seuls gènes. En quelque sorte, ces deux gènes se réfèrent à l'acquis par opposition à l'inné contenu dans les autres gènes.

III.3.1.C. Généalogie

Outre la caractérisation intrinsèque de l'individu au moyen des descripteurs de Fourier qui lui sont associés, une information extrinsèque relative à l'arbre généalogique de l'individu peut être résumée dans C gènes particuliers relatifs aux C classes considérées. En effet, les descendants héritent du patrimoine génétique des parents lesquels sont eux-mêmes issus de la génération précédente. La nature de chacun des individus de la population initiale peut être récupérée à chaque génération par incrémentation d'un compteur pour chacune des classes (cf. §III.3.4.B).

III.3.1.D. Composition du chromosome

La Figure 39 montre un chromosome constitué de la manière suivante (coefficients pour $k \in \{1, \dots, K\}$) :

- ◆ 2×K gènes représentants 2×K réels (parties réelle et imaginaire) pour K descripteurs de Fourier complexes $\tilde{X}_k$ résultant de la transformée de Fourier appliquée aux abscisses
- ◆ 2×K gènes représentants 2×K réels (parties réelle et imaginaire) pour K descripteurs de Fourier complexes $\tilde{Y}_k$ résultant de la transformée de Fourier appliquée aux ordonnées
- ◆ 2 gènes permettant de localiser le contour, c'est-à-dire les réels X_0 et Y_0
- ◆ C gènes (pour C classes d'objets) relatifs à la généalogie de l'individu

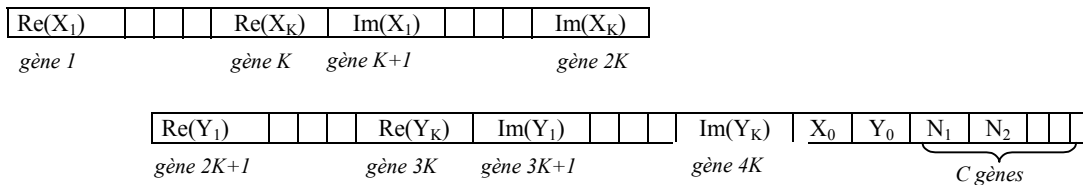


Figure 39 - Définition du chromosome

III.3.2. Population initiale

La population initiale, de taille N_p , doit couvrir l'ensemble des classes proposées dans l'annexe G : cylindres (classe C_1), sphères (classe C_2), mines Manta (classe C_3), mines Sigeel (classe C_4) et mines Rockan (classe C_5). Elle doit être représentative des différentes configurations susceptibles de se présenter. En particulier, du fait de la géométrie complexe des mines Rockan et cylindriques, la forme associée à l'ombre portée dépend du point de vue considéré. En outre, si les mines tronconiques et Rockan ont des dimensions fixes, il n'en est pas de même pour sphères et cylindres. Ces remarques montrent que la population initiale doit

comporter un nombre plus important d'exemplaires pour les objets complexes et/ou de dimensions multiples. Notons en outre qu'elle peut à tout moment être enrichie de nouveaux prototypes pour anticiper les menaces futures.

La Figure 40 montre quelques exemplaires 'prototypes' reconstruits à partir de leurs descripteurs de Fourier issus de cette population initiale.

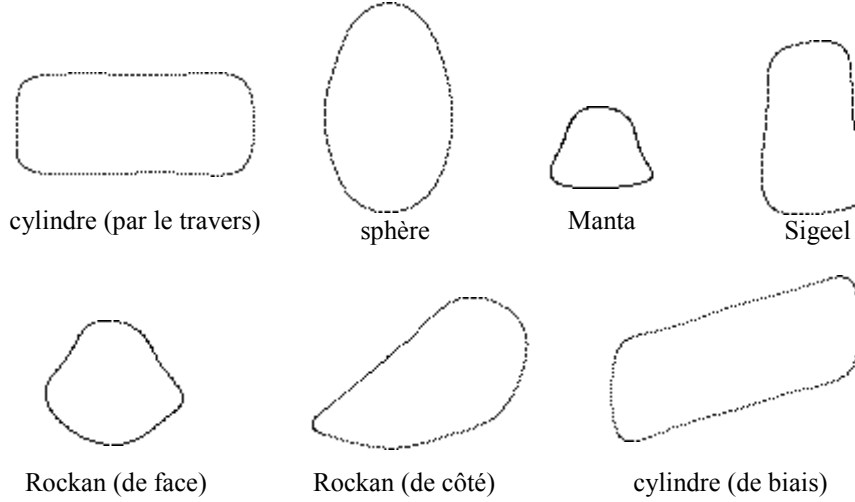


Figure 40 – Exemples d'individus (leur contour) de la population initiale

III.3.3. Critère d'optimisation pour la partition des données sonar

L'objectif visé est une description de contour optimale au sens d'un critère statistique déjà utilisé pour la segmentation par contours actifs statistiques (cf. §III.1.2 du chap. II.1). La segmentation **non supervisée** consiste à estimer, sans connaissance *a priori*, les paramètres $\theta_{(K)}$ (vecteur des descripteurs de Fourier) et $\phi = [\phi_{in}, \phi_{out}]$ (partition des pixels) à partir de l'image observée seule en maximisant la fonction de vraisemblance [FIGU-97][CHES-99] :

$$P(I/\theta_{(K)}, \phi) = \left(\prod_{(x,y) \in O} p(i(x,y)/\phi_{in}) \right) \left(\prod_{(x,y) \in R} p(i(x,y)/\phi_{out}) \right)$$

établie sous l'hypothèse de distributions des pixels indépendantes.

Au signe près (l'évaluation est codée ici comme un terme à maximiser), le terme à maximiser est identique au terme d'énergie externe des contours actifs, i.e. la maximisation de la fonction log-vraisemblance $\ln(P(I/\theta_{(K)}, \phi))$ (cf. §II.2 du chap. II.1) :

$$E = -N_O \cdot \ln \left(\frac{1}{2N_O} \sum_{(x,y) \in O} i(x,y)^2 \right) - N_R \cdot \ln \left(\frac{1}{2N_R} \sum_{(x,y) \in R} i(x,y)^2 \right)$$

$$\boxed{E = -N_O \ln \alpha_O^2 - N_R \ln \alpha_R^2}$$

où $i(x,y)$ désigne l'intensité ou niveau de gris du pixel (x,y) , O (resp. R) la zone d'ombre (resp. de réverbération de fond).

Cette estimation au sens du MV est à opposer à l'estimation au sens du MAP qui associe une probabilité a priori à la fonction de vraisemblance ci-dessus. En effet, les déformations du contour sont ici automatiquement pénalisées par limitation du degré K de décomposition en série de Fourier [FIGU-97].

III.3.4. Description de l'algorithme génétique

III.3.4.A. Fonction de sélection

L'étape de sélection consiste à choisir les parents de la génération suivante et se situe donc avant les opérations génétiques. Classiquement, une probabilité de sélection P_i est attribuée à chacun des individus, reflétant son adaptation au milieu. N_a nombres aléatoires $u_n(0,1)_{n=1\dots N_a}$ sont alors générés et comparés à la somme cumulée des probabilités $C_i = \sum_{j=1}^i P_j$, $i \in [1, N_a]$. Pour n de 1 à N_a , l'individu i est alors sélectionné et copié si $C_{i-1} < u_n(0,1) \leq C_i$.

Diverses réalisations de P_i sont possibles. Une des plus répandues, dite '**roue de loterie**', attribue à l'individu une probabilité proportionnelle à son adaptation au milieu. Son principal défaut est de supprimer à coup sûr les individus les plus faibles au risque de converger vers une solution sous-optimale. Par contre, les méthodes de tri et de sélection en fonction du rang, dites '**ranking methods**', sont plus souples et plus robustes [GOLD-91]. Ces méthodes attribuent une probabilité fonction du rang $r(i)$ de l'individu i lorsque toutes les adaptations ont été classées par ordre décroissant. Houck et Joines ont ainsi développé un algorithme appelé '**Normalized Geometric Ranking**' dans lequel P_i est défini par [HOUC-95] :

$$P_i = \frac{q}{1 - (1 - q)^{N_p}} (1 - q)^{r(i)-1}$$

Avec $r(i)$ le rang de l'individu où 1 désigne le meilleur,
 N_p la taille de la population,
 q la probabilité de sélection du meilleur individu.

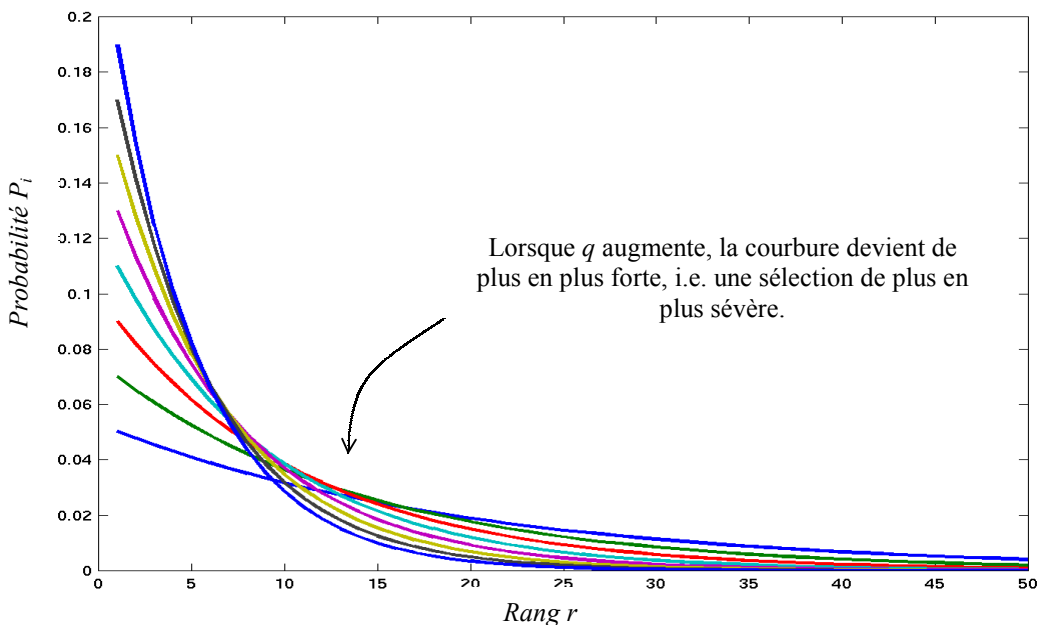


Figure 41 - Allures $P_i = f(r)$ pour différentes valeurs de q

Pour une population de 100 individus par exemple, nous visualisons les probabilités P_i obtenues en fonction du rang pour q allant de 0.05 à 0.2 par pas de 0.02, i.e. 8 courbes. La Figure 41 met en évidence une courbure de plus en plus forte, i.e. une sélection de plus en plus sévère lorsque q augmente. Il s'agit donc d'effectuer le compromis entre une sélection trop sévère (q trop important) qui mettrait en danger la diversité génétique et une sélection trop large (q trop faible) qui augmenterait les temps de calcul.

III.3.4.B. Opérateurs génétiques

Les parents sélectionnés vont se reproduire par croisement et subir des mutations. Nous avons implémenté des routines originales qui exploitent la nature des descripteurs de Fourier en utilisant leurs propriétés et l'interprétation qu'on peut en faire.

On note c_i (resp. p_i) les chromosomes des descendants obtenus (resp. des parents).

➤ croisement (entre deux parents) :

Par principe, l'individu issu du croisement génétique doit 'ressembler' à ses parents. En terme de contour, on doit donc visuellement rapprocher le nouveau contour créé des contours parentaux. Ceci impose des contraintes qui limitent les opérations admissibles entre les chaînes chromosomiques. Les trois routines suivantes sont basées sur la moyenne des descripteurs des deux parents.

Routine crois1

La moyenne concerne tous les descripteurs, le point moyen inclus, ainsi que l'information généalogique telle que

$$c_1(1, \dots, 4K + C) = \frac{p_1(1, \dots, 4K + C) + p_2(1, \dots, 4K + C)}{2}$$

Afin de maintenir le même nombre d'individus à chaque génération, un deuxième descendant copie conforme du parent le mieux adapté est formé.

Routine crois2

Une moyenne pondérée est effectuée de manière à ce que les descendants ressemblent plutôt à l'un ou l'autre des parents soit

$$\begin{cases} c_1(1, \dots, 4K + C) = \frac{p_1(1, \dots, 4K + C) + 0.5 \times p_2(1, \dots, 4K + C)}{1.5} \\ c_2(1, \dots, 4K + C) = \frac{p_2(1, \dots, 4K + C) + 0.5 \times p_1(1, \dots, 4K + C)}{1.5} \end{cases}$$

Parallèlement, l'information généalogique conformément au 'poids héréditaire' est transmise dans les mêmes proportions.

Routine crois3

Les seuls coefficients de degré 1 sont pris en compte. Ils caractérisent les axes principaux de l'ellipse approximant au mieux le contour. La même opération de moyenne que pour *crois1* est effectuée mais limitée à ce seul coefficient. Les autres coefficients conservent la même valeur.

Routine crois4

Pour garantir la diversité, une quatrième routine *crois4* permet de créer deux individus complémentaires dont les descripteurs de Fourier sont choisis par tirage aléatoire entre les chaînes chromosomiques des deux parents.

Prenons le cas $K=5$. Un exemple est illustré Figure 42 pour lequel le tirage aléatoire attribue les coefficients $k=2,3,5$ de p_1 au descendant c_1 et $k=1,4$ au descendant c_2 et inversement concernant p_2 .

Les points moyens retournés sont identiques pour c_1 et c_2 soit

$$c_i(4K+1, 4K+2) = \frac{p_1(4K+1, 4K+2) + p_2(4K+1, 4K+2)}{2}$$

et parallèlement, l'information généalogique est transmise par

$$c_i(4K+3, \dots, 4K+3+C) = \Pi_1^i \times p_1(4K+3, \dots, 4K+3+C) + \Pi_2^i \times p_2(4K+3, \dots, 4K+3+C)$$

où Π_1^i (resp. Π_2^i) représente la proportion de gènes reçus de p_1 (resp. p_2) par le descendant c_i .

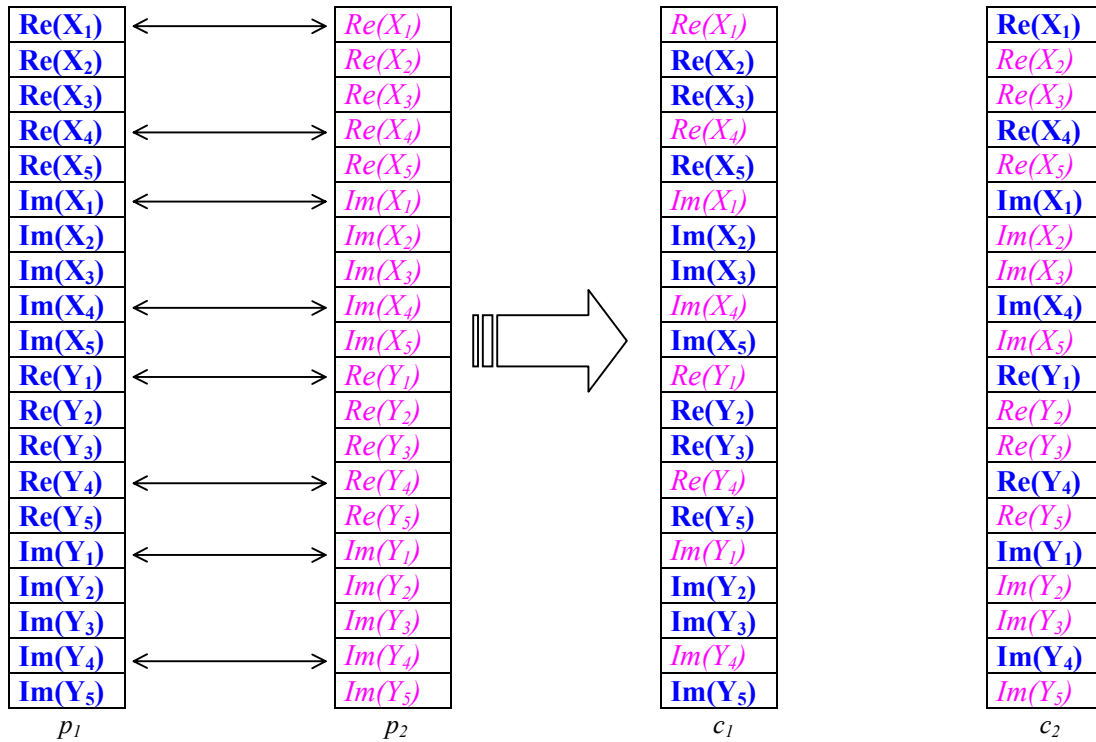


Figure 42 – Echanges aléatoires entre deux chaînes chromosomiques

➤ mutation (s'applique à un seul parent) :

Remarque : ces opérations ne concernent que les $4K$ premiers gènes c'est-à-dire les descripteurs de Fourier.

Routine mut1

Un bruit blanc gaussien est ajouté à l'un des descripteurs de Fourier. Ce bruit est de moyenne nulle et d'écart-type égal à la valeur maximale admissible (déterminée par apprentissage sur une base étendue d'individus) pour ce descripteur divisée par 10.

Routine mut2

Cette routine utilise la propriété de linéarité de la décomposition en série de Fourier (cf. §III.1 du chap. II.2). Une transformation géométrique des descripteurs obtenus par décomposition en série de Fourier des coordonnées des pixels de contour se traduit par une transformation identique du contour. On dispose donc d'un modèle paramétrique linéaire de degré 4

correspondant aux quatre paramètres de la matrice $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$:

$$\begin{bmatrix} U_k \\ V_k \end{bmatrix} = A \times \begin{bmatrix} X_k \\ Y_k \end{bmatrix}$$

Par mutation, on peut donc modifier le contour de manière globale en lui affectant une transformation affine aléatoire (la translation exceptée). Six déformations ou transformations géométriques affines de matrice A sont possibles (on utilise les notations introduites au paragraphe I du chap. II.2) :

- l'homothétie d'amplitude $\omega = 1 + \alpha$ avec $\alpha \in [-0.4; +0.4] \setminus \{0\}$ par tirage aléatoire
- l'élongation d'amplitude $\delta = 1 + \alpha$ sur les lignes ou les colonnes avec $\alpha \in [-0.4; +0.4] \setminus \{0\}$ par tirage aléatoire
- la rotation d'angle θ avec $\theta \in \frac{1}{iter} [-9^\circ; +9^\circ] \setminus \{0\}$ par tirage aléatoire
- le biais d'amplitude l sur les lignes ou les colonnes avec $l \in \frac{1}{iter} [-0.1; +0.1] \setminus \{0\}$ par tirage aléatoire

Pour les trois dernières possibilités, *iter* désigne l'itération de sorte que l'amplitude admissible des rotations et des biais diminue au fil des générations. L'amplitude des transformations a été fixée de manière empirique de manière à éviter des transformations trop importantes qui créeraient des individus non viables.

En modifiant ainsi la forme de manière globale (*mut2*), on approxime une déformation affine qui peut être due à l'angle de prise de vue, à la différence de taille entre la mine ciblée et la mine prototype, voire aux artéfacts de traitements préliminaires de formation et de normalisation en rasant de l'image sonar. La nature de la silhouette n'est donc pas altérée. A l'opposé, une modification partielle voire ponctuelle de la chaîne des descripteurs (*mut1*) modifie intrinsèquement la silhouette en modifiant sa nature. Par leur complémentarité, l'association des deux séries d'opérations génétiques apporte de la flexibilité et de la robustesse.

III.3.4.C. Taux de croisements [PAL-96]

Le choix des taux de croisements et de mutations ainsi que celui de la taille de la population initiale n'est pas trivial :

- accroître le taux de croisements revient à augmenter le nombre de recombinaisons mais dissocie en contrepartie les bons chromosomes,
- accroître le taux de mutations risque de transformer la recherche en une exploration aléatoire de l'espace des solutions, mais permet de réintroduire une partie du matériel génétique perdu (par *mut1*) ou de diversifier la population (grâce à *mut2*),

- augmenter la taille de la population accroît sa diversité mais augmente les temps de calcul.

Les taux que nous avons choisis dépendent des opérateurs et de la taille de la population. Ainsi, pour une petite population, on est amené à autoriser davantage d'opérations génétiques de manière à pallier un manque de diversité. Par ailleurs, la taille de notre population initiale est fortement conditionnée par l'objectif visé qui consiste à segmenter des ombres relatives à des objets connus que l'on souhaite classifier. Nous souhaitons ainsi rapprocher l'ombre détectée d'un ensemble de cas répertoriés et préalablement segmentés (cf §III.3.2).

III.3.4.D. Fonction d'adaptation

Le terme E défini au paragraphe III.3.3 correspond à une énergie (désignée par énergie externe ou terme d'attache aux données dans la théorie des contours actifs) et constitue pour notre application de l'algorithme génétique la **fonction d'évaluation** ou **d'adaptation**.

Pour calculer cette énergie, le traitement appliqué à chaque individu de la population à chaque génération est le suivant :

- 1) positionnement du marqueur (point moyen approché de l'ombre portée) sur l'image observée (cf. §III.3.1.B)
- 2) lecture des K descripteurs de Fourier, capital génétique de l'individu (cf. §III.3.1.A)
- 3) reconstruction de la silhouette associée à l'individu par Transformée de Fourier Inverse autour du point moyen telle que

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \sum_{\substack{k=-K \\ k \neq 0}}^K \begin{bmatrix} X_k \\ Y_k \end{bmatrix} \exp\left(2j\pi \frac{ki}{N}\right)$$

où $X_{-k} = X_k^*$, $Y_{-k} = Y_k^*$, et N désigne le nombre de points de contour.

- 4) calcul de la fonction d'adaptation E (cf. §III.3.3)

- si toutes les coordonnées du contour reconstruit sont inférieures aux dimensions de l'image :

$$E = -N_O \ln \alpha_O^2 - N_R \ln \alpha_R^2,$$

- sinon, on pénalise fortement un contour dont une partie sort du cadre de l'image en considérant l'image comme une seule région soit :

$$E = -N_{image} \ln \alpha_{image}^2$$

Remarque (cf. annexe J) : afin de diminuer le temps de calcul, nous utilisons l'histogramme de chaque région plutôt que l'ensemble des pixels de l'image. En effet, la contribution d'un pixel à la fonction d'adaptation dépend uniquement de son niveau et de la région à laquelle il appartient (et non de ses coordonnées). On passe ainsi d'une double somme sur les lignes et les colonnes à une simple somme sur les 256 niveaux de gris.

III.3.4.E. Réinsertion

L'étape de sélection réarrange la population de sorte que les individus les plus faibles sont éliminés au profit d'individus plus forts ayant une probabilité de sélection plus grande. Des copies des individus les plus forts sont réalisées jusqu'à ce que le nombre d'individus initial N_p soit atteint. Ensuite, les opérations de croisements prennent chaque fois deux parents pour renvoyer deux descendants. Les mutations, d'un autre côté, ne concernent qu'un individu à la

fois, remis en place après transformation. Quels que soient le choix des routines et leur occurrence (fonction des taux de croisement et de mutation choisis), à l'issue des opérations génétiques, N_p descendants doivent être transmis. Deux cas peuvent se présenter :

- Si le nombre de descendants est supérieur à N_p , on ne réinsère que les individus les mieux adaptés,
- Sinon, les descendants remplacent les individus de la génération précédente les moins bien adaptés.

III.3.4.F. Critère d'arrêt

La fonction de terminaison est appelée à la fin de chaque itération c'est-à-dire lorsque la nouvelle génération est obtenue. Nous choisissons d'arrêter l'optimisation lorsque l'un des cas suivants est rencontré :

- la population des descendants est homogène c'est-à-dire l'écart entre les évaluations des individus le mieux et le moins bien adaptés est inférieur à ϵ ,
- il n'y a pas eu d'amélioration pour trois itérations consécutives,
- le nombre maximum de générations admis est atteint.

III.3.5. Module de classification

Estimation faite des descripteurs de Fourier associés au contour le mieux adapté à l'image observée, il s'agit d'en déterminer la classe correspondante. Pour ce faire, nous décidons de rapprocher les descripteurs estimés avec les descripteurs caractérisant les N_p^c exemplaires associés à chacune des $C=5$ classes d'une population de référence, représentative des cas susceptibles d'apparaître, et plus étendue (c'est-à-dire plus diverse) que la population initiale (i.e. $N_p^c > N_p$).

Compte tenu de l'unicité des descripteurs de Fourier normalisés, nous introduisons un critère de classification basé sur le principe des k-plus-proches-voisins (k-ppv) qui permet d'attribuer au meilleur individu la classe majoritaire parmi les k voisins les plus proches au sens d'une distance définie ci-dessous.

La distance d_j peut être calculée à partir des gènes du meilleur individu et de ceux du $j^{\text{ième}}$ individu de la population de référence telle que :

$$d_j = \left(\sum_{i=1}^{4K} (g_i^j - \tilde{g}_i)^2 / 4K \right)^{\frac{1}{2}}$$

où

$(\tilde{g}_i)_{i=1 \dots 4K}$ sont les gènes estimés

$(g_i^j)_{i=1 \dots 4K}$, les gènes du chromosome du $j^{\text{ième}}$ individu de la population de référence

Chaque indice j réfère à un prototype spécifique dont on connaît la classe. L'indice $j = \text{Arg} \min_{j \in \{1 \dots N_p^c\}} d_j$ est en particulier relatif au plus proche prototype.

Les k-ppv consistent à trouver les k distances les plus faibles puis à déterminer la classe associée aux individus correspondants. Toutefois, il est intéressant de remonter à la valeur numérique de ces distances et de vérifier leur cohérence (détails sur la Figure 43). En ce cas d'incohérence du résultat de classification par k-ppv, l'information apportée par les C gènes $g_{c,c=1 \dots C}^{\text{gen}} = \tilde{g}_{i,i=4K+3 \dots 4K+2+C}$ relatifs à la généalogie de l'individu est prise en compte (cf. §III.3.1).

L'algorithme de classification est décrit sur la Figure 43. Deux cas se présentent :

- plusieurs classes sont concurrentes,
- une classe seule est majoritairement représentée parmi les k voisins.

- Trouver les k indices j pour lesquels les distances d_j sont les plus faibles

- Répartir les individus correspondants par classe $C_{c,c=1\dots C}$,

soit $N_{c,c=1\dots C}$ individus par classe tel que $\sum_{c=1\dots C} N_c = k$

- Calculer la distance moyenne associée à la classe $C_{c,c=1\dots C}$:

$$\bar{d}_c = \frac{1}{N_c} \sum_{j, j \in \{\text{indices des individus de } C_c\}} d_j$$

Deux cas possibles :

- Si plusieurs classes sont concurrentes, c'est-à-dire si elles bénéficient chacune du nombre maximal de représentants $\max_c N_{c,c=1\dots C}$ parmi les voisins

$c_1 = \text{Arg min}_{c \in \{1\dots C\}} \bar{d}_c$, la classe attribuée est celle dont la distance moyenne est la plus faible

- Sinon,

$c_1 = \text{Arg max}_{c \in \{1\dots C\}} N_c$, classe majoritaire parmi les k voisins les plus proches

Cohérence du résultat (n'existe-t-il pas une distance $\bar{d}_c$ bien plus faible que $\bar{d}_{c_1}$?) :

Soit le rapport R de $\bar{d}_{c_1}$ sur la valeur minimale des distances moyennes,

Si $R = \frac{\bar{d}_{c_1}}{\min_c \bar{d}_c} \geq 2$ « incohérence »

L'hérédité est prise en compte via $\text{Arg max}_{c \in \{1\dots C\}} g\tilde{n}_c$, classe majoritaire parmi les ascendants

Si $\text{Arg min}_{c \in \{1\dots C\}} \bar{d}_c = \text{Arg max}_{c \in \{1\dots C\}} g\tilde{n}_c$

alors la classe attribuée est $c_1 = \text{Arg min}_{c \in \{1\dots C\}} \bar{d}_c$

Figure 43 – Algorithme de classification du chromosome estimé

III.4. Expérimentations

III.4.1. Paramétrage

Une reconstruction lissée du contour peut être obtenue en fixant **K à 5** tout en préservant la forme globale de la silhouette qui permet donc de la discriminer.

La population initiale doit couvrir l'ensemble des $C=5$ classes (cf. §III.3.2). La proportion des exemplaires par classe dépend des propriétés de symétrie et de la taille (fixe ou multiple) des objets :

- les cylindres et les sphères ont des dimensions variées et doivent donc être représentés par un nombre d'exemplaires supérieur aux mines à taille fixe
- du fait de la géométrie des mines Rockan et cylindriques, la forme associée à l'ombre portée dépend du point de vue considéré : ces deux classes doivent comporter l'ensemble des configurations possibles.

Par simulation, différents points de vue sont acquis le long d'une trajectoire semi-circulaire du sonar autour de l'objet. La population initiale est de $N_p=41$ individus répartis entre les cinq classes en fonction de leur dispersion (diversité des tailles et des formes possibles de l'ombre) comme suit :

- $N_{cyl}=19$ en prenant 3 tailles de cylindres et 19 points de vue différents,
- $N_{sph}=10$ en prenant 3 tailles de sphères,
- $N_{man}=N_{sig}=3$, car les mines tronconiques sont de taille constante et ont la propriété de symétrie radiale,
- $N_{rock}=6$ en prenant 6 points de vue représentatifs.

La taille de la population initiale nous amène à fixer q à **0.25** pour la sélection des individus. Selon la formule et les considérations du paragraphe III.3.4.A, la probabilité de sélection P_i est alors inférieure à 0.001 (ce qui équivaut à une sélection improbable) pour la moitié des individus.

La population de référence pour la classification est environ trois fois plus importante pour pouvoir fixer le **nombre de voisins à $k=3$** dans le module de classification.

D'autre part, les images sont normalisées conformément à la technique présentée au §II.1 du chap. II.2.

La résolution du pixel dépend initialement des paramètres sonar mais est modifiée à l'étape de normalisation. La **taille équivalente du pixel** imposée en final, identique suivant les deux axes, dépend ainsi à la fois des paramètres sonar et de la normalisation de l'image. Cette donnée nous servira à adapter la taille des individus de la population initiale, simulés sous des conditions (paramètres sonar, position du sonar) propres qui peuvent différer de celles de l'image à segmenter.

Le **taux de croisements** est fixé à **10%** pour chaque opérateur de moyenne et à **20%** pour l'échange aléatoire entre chaînes chromosomiques. Pour la mutation par modification ponctuelle d'un gène, on effectue une seule opération par génération. Par contre, un **taux de 50%** a été choisi pour la **mutation par transformation affine** afin de multiplier les différentes configurations en termes de format et de point de vue des ombres potentielles associées aux individus sélectionnés. Ce taux diminue au fil des générations de sorte que pour une génération ou itération donnée $iter$, le taux est de $50/iter\%$ (cf. §III.3.4.B).

Enfin, le **nombre maximal d'itérations** est **10** et **epsilon** est fixé à **10^2** pour le critère d'arrêt.

III.4.2. Performances

Nous avons évalué notre algorithme sur une base d'images simulées décrite dans l'annexe G.

Le Tableau 11 donne la matrice de confusion obtenue.

Tableau 11 – Matrice de confusion (taux moyen de bonne classification de 98.2%)

	C_1	C_2	C_3	C_4	C_5
Cylindres	98.4	1.0	0.1	0	0.5
Sphères	0	99.3	0.3	0	0.3
Mines Manta	0	0.3	99.7	0	0
Mines Sigeel	0	0	0	100	0
Mines Rockan	0.2	0	6.3	0	93.5

La plus forte confusion intervient pour la mine Rockan dont l'un des points de vue est systématiquement associé à la classe des mines Manta de par la forme trapézoïdale de l'ombre portée (cf. Figure 44). Il convient de noter que ce problème pourrait vraisemblablement être résolu en augmentant le nombre de descripteurs (chromosome de longueur plus importante), c'est-à-dire en captant davantage de détails sur la silhouette de cet objet.

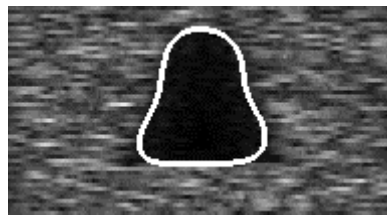


Figure 44 – Point de vue ambigu de la mine Rockan

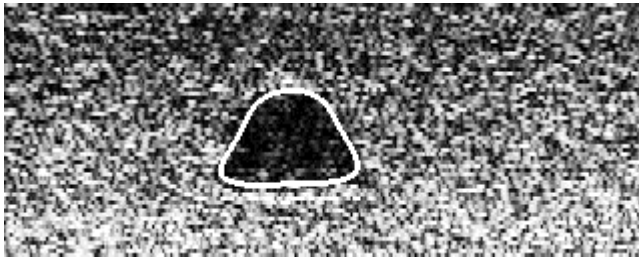
III.4.3. Illustrations

Les Figure 45, Figure 46 et Figure 47 illustrent les résultats obtenus sur images réelles acquises par un traitement 'antenne synthétique' et sur images simulées [BILL-98].

Pour chaque image testée, on affiche la classe C_c attribuée selon l'algorithme de la Figure 43 et le taux N_c/k associé. En particulier,

- ☞ un taux inférieur à 0.5 indique que plusieurs classes étaient concurrentes
- ☞ un taux nul indique que l'on a classifié à partir de la généalogie de l'individu et non des k-plus-proches-voisins

Sur la Figure 47, à l'histogramme de l'image observée, on superpose les histogrammes estimés des zones 'ombre' et 'réverbération de fond' résultant de la partition (segmentation) de l'image.



Objet classé 'mine Manta' par k-ppv avec un taux de 1

Généalogie :

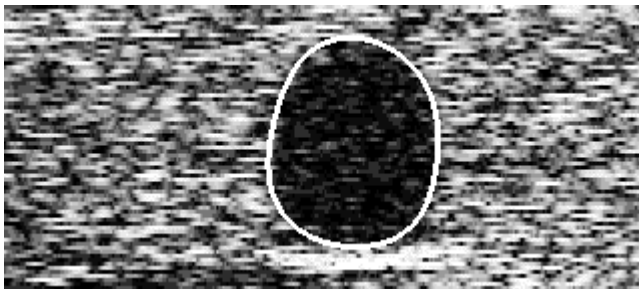
0% cylindre

0% sphère

88.2% mine Manta

0% mine Sigeel

11.8% mine Rockan



Objet classé 'sphère' par k-ppv avec un taux de 1

Généalogie :

14.6% cylindre

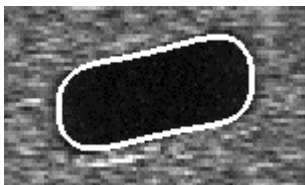
84.6% sphère

0% mine Manta

0.8% mine Sigeel

0% mine Rockan

Figure 45 – Résultats sur images réelles (mine Manta et mine sphérique)



Objet classé 'cylindre' par k-ppv avec un taux de 1

Généalogie :

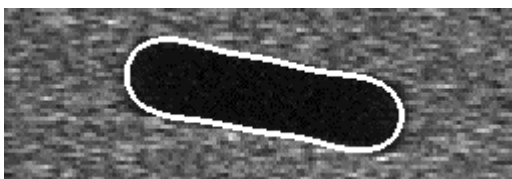
100% cylindre

0% sphère

0% mine Manta

0% mine Sigeel

0% mine Rockan



Objet classé 'cylindre' par k-ppv avec un taux de 1

Généalogie :

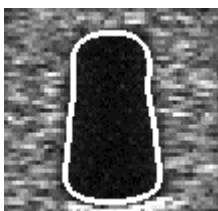
100% cylindre

0% sphère

0% mine Manta

0% mine Sigeel

0% mine Rockan



Objet classé 'mine Sigeel' par k-ppv avec un taux de 1

Généalogie :

0% cylindre

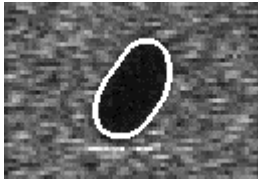
50% sphère

0% mine Manta

50% mine Sigeel

0% mine Rockan

Figure 46 – Résultats sur images simulées (2 mines cylindriques et une mine Sigeel)



Objet classé 'cylindre' par prise en compte de la généalogie (taux nul par k-ppv)

Généalogie :

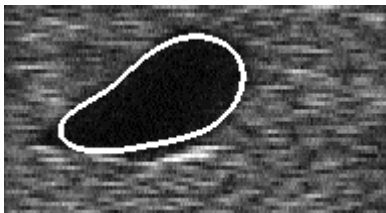
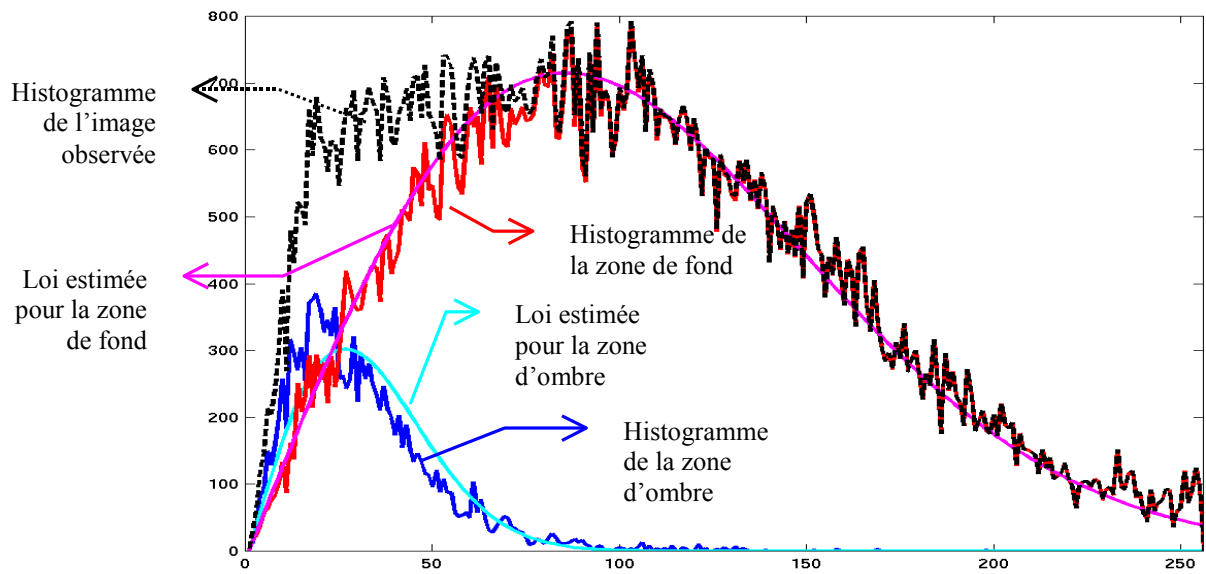
61.1% cylindre

0% sphère

0% mine Manta

0% mine Sigeeel

38.9% mine Rockan



Objet classé 'mine Rockan' par k-ppv avec un taux de 1

Généalogie :

0% cylindre

0% sphère

0% mine Manta

0% mine Sigeeel

100% mine Rockan

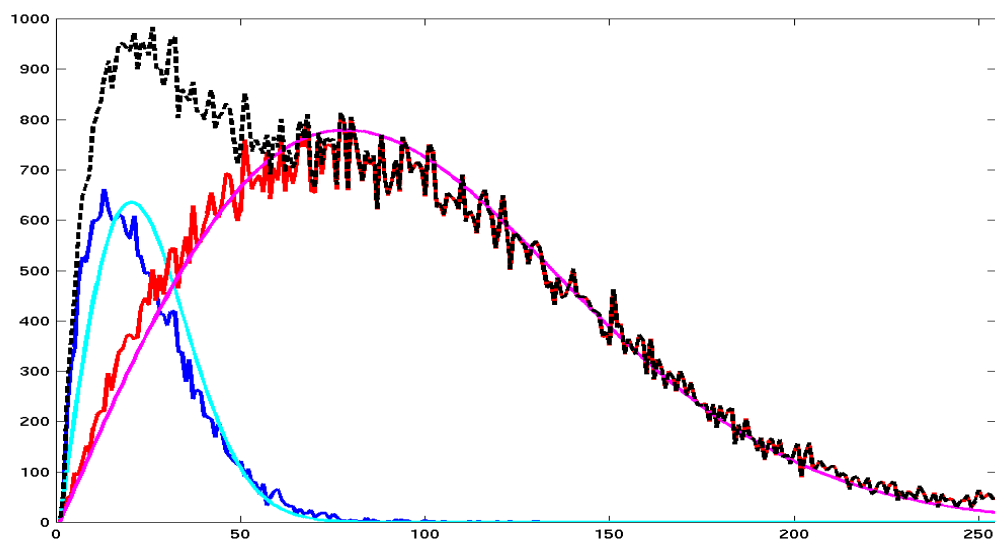


Figure 47 – Résultats sur images simulées (une mine cylindrique et une mine Rockan) et le résultat de l'estimation des distributions des niveaux de gris des pixels

III.5. Conclusion

Dans le contexte de la chasse aux mines, les informations que nous possédons sur les différents objets susceptibles de représenter une menace nous permettent d'accélérer les traitements en travaillant initialement avec un nombre réduit d'individus. La technique proposée consiste à explorer le domaine des solutions possibles par algorithme génétique. L'initialisation permet de se positionner dès le départ dans un contexte favorable, c'est-à-dire un ensemble d'individus dont les caractéristiques sont parfaitement connues. Au fil des générations, ces individus montrent une forte capacité d'adaptation à l'image observée en convergeant vers une solution optimale au sens d'un critère statistique d'évaluation. Des opérateurs génétiques ont été spécifiquement implémentés pour tirer profit des propriétés des descripteurs de Fourier codant les individus. Les tests effectués sur images réelles et simulées ont montré que, par sa flexibilité, l'algorithme génétique implémenté estimait correctement les descripteurs de Fourier décrivant au mieux le contour de l'ombre portée. En outre, on classe aisément l'objet ciblé en rapprochant la chaîne chromosomique estimée des chaînes des individus d'une population de référence.

En travaillant directement à partir des données sonar brutes, nous nous affranchissons des étapes intermédiaires (segmentation et extraction d'attributs caractéristiques) intervenant dans le traitement séquentiel classiquement mis en œuvre pour la reconnaissance de formes (cf. chapitre III.1). À travers les générations successives, l'ensemble des contours codés dans les chromosomes des individus évoluent suivant les lois de la génétique selon un processus dynamique de classification de l'ombre.

IV. Conclusion du chapitre III.2

Ces approches alternatives sont une intéressante mise en œuvre des attributs dits descriptifs c'est-à-dire qui permettent de reconstruire le contour. Nous avons pleinement profité de la propriété de linéarité des descripteurs de Fourier tant lors de l'estimation des paramètres de la transformation affine pour l'appariement de silhouette qu'à travers les opérations génétiques et la discrimination de contour pour une segmentation orientée classification.

Conclusion de la partie III

Cette partie fut consacrée à l'application mono-vue des outils présentés dans la partie II. Nous avons opposé l'approche séquentielle classique « segmentation - extraction d'attributs – classification » à deux approches alternatives offrant la possibilité de réduire considérablement la base d'apprentissage.

L'approche séquentielle a mis en jeu plusieurs familles d'attributs caractéristiques qui ont été comparés, par l'intermédiaire de trois classifieurs classiques, de manière à rendre compte de leurs atouts et de leurs faiblesses respectifs. A l'observation des résultats obtenus, nous avons voulu tester l'opportunité d'associer des attributs de différents types mais apparaissant comme complémentaires. Une hybridation « attributs intrinsèques – invariants de forme » a ainsi fourni des résultats probants (taux de bonne classification moyen de 99%) par la méthode des k-plus-proches-voisins, confirmés par une analyse en composantes principales offrant une bonne partition dans le plan principal.

Deux approches alternatives exposées au chapitre III.2 évitent la lourde contrainte de disposer d'une base d'images simulées importante. En mettant toutes deux en œuvre, de manière différente, les descripteurs de Fourier, elles profitent particulièrement de leur flexibilité liée à la propriété de linéarité. En outre, la seconde approche remet en cause le schéma classique séquentiel : l'algorithme génétique proposé est en effet orienté vers une description discriminante du contour en accomplissant ainsi une segmentation orientée classification.

Les résultats obtenus sont inférieurs à l'approche séquentielle appliquée aux attributs hybrides : un taux moyen de 97.5% par appariement de silhouettes et de 98.2% par estimation par algorithme génétique des descripteurs de Fourier décrivant le contour. Toutefois, une analyse plus fine, effectuée par classe, de ces chiffres montre qu'ils sont en particulier pénalisés par des ambiguïtés récurrentes dans la base d'images testées : elles correspondent à un angle de prise de vue non caractéristique de l'objet ciblé. Cette dernière remarque suggère que l'apport de connaissances a priori devrait notablement améliorer les performances.

PARTIE IV :

Application multi-vues

Introduction de la partie IV

Classier en mono-vue suppose que l'on se situe dans un cas favorable (fond horizontal, faible relief, ombre révélatrice de l'objet ciblé...). En réalité, certaines ambiguïtés demeurent :

- 1- le point de vue particulier de l'image peut ne pas être le plus caractéristique (exemple pour un cylindre la ligne de visée parallèle à la génératrice),
- 2- la forme du fond derrière l'objet a une influence sur l'ombre portée (dépression du terrain ou bosse),
- 3- l'environnement peut être source de confusion (rocher qui jouxte l'objet ciblé, rides profondes,...),
- 4- des objets ayant peu ou pas d'axe(s) de symétrie (cas de la mine furtive Rockan) présentant des ombres bien différentes selon l'angle de vue.

Le levé de ces incertitudes peut être réalisé en utilisant plusieurs images présentant les ombres acquises selon des points de vue distincts. Une caractérisation multi-vues est alors effectuée et le même objet est caractérisé par l'ensemble des attributs calculés sur chacune des images. La mise en œuvre de telles stratégies d'acquisition d'images est aujourd'hui confortée par les progrès majeurs concernant les techniques de chasse aux mines voyant la généralisation prochaine de véhicules filoguidés ou autonomes portant l'antenne sonar [JANE]. La classification automatique à bord de ces véhicules, est rendue possible par une meilleure résolution (sonar Haute Fréquence et proximité de la cible) associée à une grande latitude de mouvement autour de l'objet pour en extraire un plus grand nombre d'informations. Suivant le principe de classification supervisée, celles-ci sont alors comparées à une base de données de référence résultant d'un apprentissage supervisé.

Nous avons exploré deux axes d'étude :

- ☞ l'exploitation conjointe des informations résultant d'une analyse mono-vue effectuée image par image,
- ☞ la caractérisation d'une nouvelle image résultant de la superposition de l'ensemble des vues.

Le mode d'observation conditionne les points de vue, leur nombre, et le lien entre les informations qu'on en extrait. Ces dernières sont ainsi différemment exploitables.

Outre les incertitudes énoncées plus haut, les conditions d'acquisition des images sont parfois imprécises. Ce peut être le cas particulier de l'insonification de la même zone de fond, par plusieurs sonars, sans chronologie précise. Dans ce cas, la solution consiste à **fusionner les informations de haut-niveau** (attributs) tirées des différentes vues indépendamment de la trajectoire du sonar. Via les techniques de fusion, des mesures particulières relient les informations issues d'un point de vue donné à une décision potentielle sur la classe de l'objet ciblé. La décision intervient en finalité après combinaison de ces mesures en gérant incertitude et imprécision [BLOC-94]. Parmi les outils généralement utilisés pour définir des modèles de fusion, la théorie des ensembles flous est apparue adaptée au contexte opérationnel et aux informations extraites des images sonar segmentées [QUID-01c]. Ce formalisme de la logique floue proposé par Zadeh permet d'affecter à tout élément un degré d'appartenance à un ensemble flou représentatif d'une classe donnée [ZADE-65]. La classification est dite floue car elle est obtenue à partir d'un degré d'appartenance aux classes.

Sous la contrainte d'une trajectoire du sonar précisément suivie et calibrée, l'évolution de la forme de l'ombre d'image à image est limitée proportionnellement à l'écart entre les points de vue. Par le choix d'attributs appropriés, cette transformation de la forme est simultanément répercutée sur les attributs extraits et exploitable à l'échelle de toute la séquence d'images [QUID-01a]. Le problème s'identifie alors à l'analyse de la courbe **d'évolution des valeurs attributs en fonction des points de vue**. La contrepartie de cette approche est liée aux problèmes de navigation : une maîtrise de la trajectoire ou au mieux sa connaissance a posteriori sont indispensables.

Une autre voie semble prometteuse : superposer les images successives (sous la condition d'un recalage) pour **construire une image 'multi-vues'**. Inspirée des attributs rencontrés en mono-vue, une version floue d'attributs géométriques classiques est utilisée pour discriminer la cible. Celle-ci est dans ce cas caractérisée par une région floue résumant les différents aspects de l'ombre [ROSE-84]. Deux voies sont possibles :

- ☞ Si les images sont préalablement segmentées, le nombre de niveaux de gris de la nouvelle image est égal au nombre de vues. Le niveau de gris du pixel reflète sa fréquence d'apparition sur l'ensemble des images segmentées.
- ☞ Si on somme les données brutes, on extrait la zone d'ombre en seuillant la nouvelle image.

Chapitre IV.1 : De l'analyse mono-vue à la classification multi-vues

I. Fusion

Les ambiguïtés qui entachent le traitement mono-vue peuvent être gérées en fusionnant les informations ou attributs extraits de N images I_n présentant les ombres acquises selon des points de vue distincts.

I.1. Fusion de données en imagerie sonar

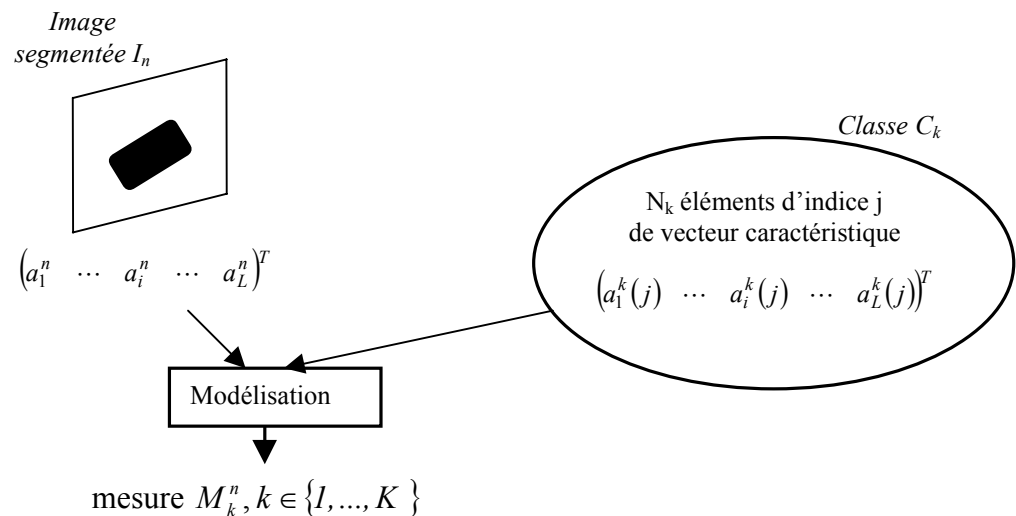


Figure 48 – Mesure d'une image courante I_n à une classe C_k

La fusion de données est l'utilisation conjointe des données pour l'aide à la prise de décision. L'analyse s'effectue en trois temps :

- définition des fonctions de confiance caractérisant l'information dans chaque image,
- combinaison de ces fonctions,
- prise de décision proprement dite.

La fusion de données vise à traiter des informations imprécises (défaut quantitatif de connaissance) et incertaines (conformité à la réalité) [BLOC-94] [BLOC-95]. Dans notre problème de classification multi-vues sur ombre, le principe de fusion apporte de la

robustesse et renforce la crédibilité des résultats en gérant l'imprécision liée aux prises de vue (angles de prise et nombre de vues) de même que les incertitudes liées à un environnement marin difficile (irrégularités et nature du fond et de l'eau) (cf. Figure 49).

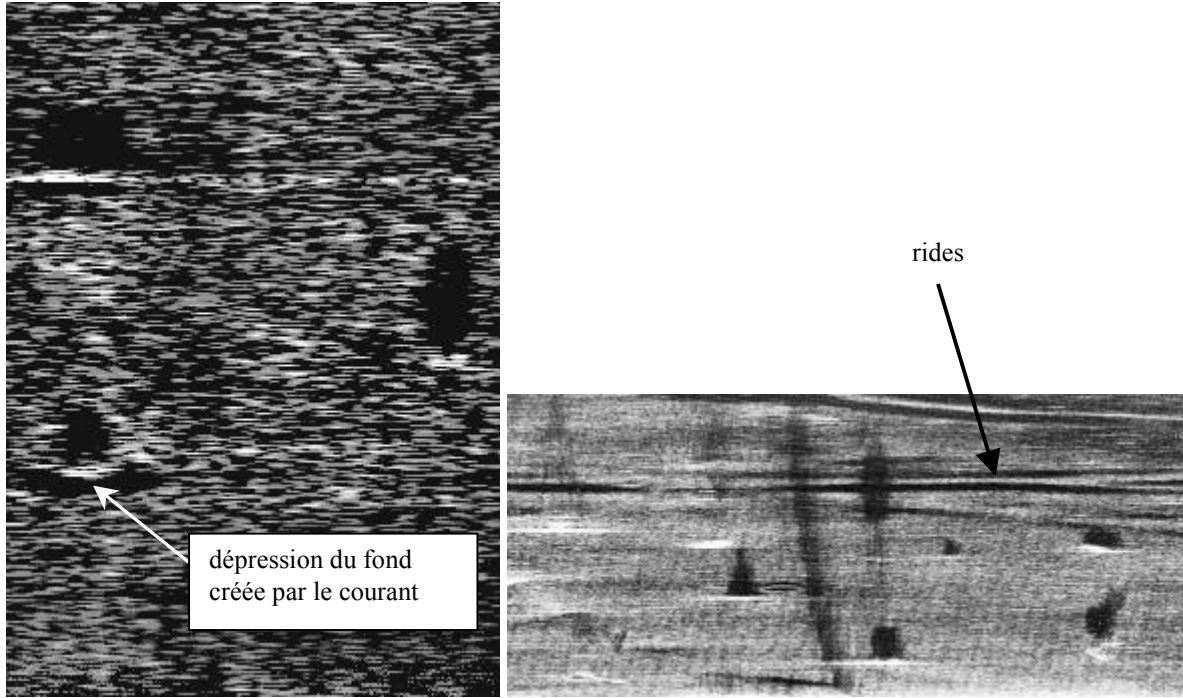


Figure 49 – Influence de la nature du fond sur la qualité des images

1.1.1. Modélisation

Nous disposons de N images représentant l'ombre portée du même objet avec des angles de vue distincts. On note f , l'ombre portée (forme) apparaissant sur l'image I_n sur laquelle porte la décision. La décision consiste à lui affecter la classe C_k parmi les $K=5$ classes d'objets recherchés, soient 'cylindre', 'sphère', 'mine Manta', 'mine Sigeel' et 'mine Rockan'. Selon le principe de classification supervisée, chaque classe est constituée de N_k exemples issus de la base d'apprentissage ou d'entraînement caractérisés par L attributs $a_i^k(j)_{(i,k,j) \in \{1, \dots, L\} \times \{1, \dots, K\} \times \{1, \dots, N_k\}}$. La décision sur f est prise à partir d'informations $a_i^n(f)_{i \in \{1, \dots, L\}}$ que constituent nos L attributs courants. La fusion s'appuie sur ces informations par l'intermédiaire de mesures $M_k^n(f)_{(k,n) \in \{1, \dots, K\} \times \{1, \dots, N\}}$, reliant les attributs sur f à une décision potentielle C_k sur f (cf. Figure 48). Le Tableau 12 résume les informations à fusionner pour les N vues.

Le modèle de fusion nécessite de disposer simultanément de toutes les informations recueillies à partir de N vues. On travaille alors sur chacune des lignes du Tableau 12 afin d'obtenir une mesure fusionnée relative à la décision C_k grâce à un **opérateur de fusion F** :

$$M_k(f) = F(M_k^n(f), 1 \leq n \leq N)$$

La **modélisation** consiste à passer des informations extraites des images à une représentation mathématique liée à une théorie particulière. Pour un modèle donné, la démarche vise à obtenir la mesure globale M_k sur les N vues : plus M_k se rapproche de 1, plus on accorde de confiance à la classe attribuée.

Tableau 12 – Informations à fusionner

Images \ Classes	I_1	I_2	...	I_N
$C_1 = \text{cylindre}$	M_1^1	M_1^2	...	M_1^N
$C_2 = \text{sphère}$	M_2^1	M_2^2	...	M_2^N
$C_3 = \text{mine Manta}$	M_3^1	M_3^2	...	M_3^N
$C_4 = \text{mine Sigeel}$	M_4^1	M_4^2	...	M_4^N
$C_5 = \text{mine Rockan}$	M_5^1	M_5^2	...	M_5^N

I.1.2. Choix du cadre théorique

Les informations obtenues suite à l'étape d'extraction d'attributs de l'ombre segmentée nous permettent de poser le problème de deux manières :

- en dénombrant les exemples en fonction de la valeur de leurs attributs caractéristiques, soit en faisant appel à la théorie des probabilités,
- en considérant que les exemples associés à chacune des classes constituent un sous-ensemble flou de la base, on se place alors dans la théorie des ensembles flous.

La probabilité d'occurrence de a_i^n conditionnellement à la classe C_k est égale au nombre d'exemples j de la base d'apprentissage pour lesquels $a_i^k(j)$ est voisin de a_i^n rapporté au nombre total d'exemples pour cette classe. Les modèles probabilistes exigent d'échantillonner au mieux l'espace des valeurs admissibles pour a_i . Cette donnée n'est pas triviale et influe sur les performances du modèle. D'autre part, l'opérateur de combinaison ou de fusion consiste en un produit de probabilités. Les mesures ainsi combinées de manière *conjonctive* (correspondant à un comportement sévère) « s'effondrent » [BLOC-94]. En somme, le produit tend très rapidement vers zéro si on multiplie le nombre de vues. La théorie probabiliste ne semble donc pas adaptée à notre problème. L'utilisation d'un opérateur de fusion dit *de compromis* est préférable dans notre cas.

Diverses mesures floues (degré d'appartenance floue, distribution de possibilités, fonction de masse de la théorie des croyances...) furent introduites pour la modélisation de raisonnement de type humain [BANO-81]. En particulier, aux règles probabilistes d'union ou d'additivité d'événements sont substituées d'autres règles de combinaison. Ceci ouvre le voie à des plusieurs combinaisons possibles des informations (répondant à divers comportements) moins contraintes que la combinaison probabiliste. La théorie des ensembles flous offre une grande variété d'opérateurs de fusion couvrant toute la gamme de comportements possibles [BLOC-96b]. Cette propriété l'oppose à la théorie des croyances de Dempster-Shafer, choisie par Stage et Zerr pour la fusion d'informations issues de plusieurs images sonar, pour laquelle l'opérateur de fusion est imposé [STAG-96]. La théorie des possibilités, dérivée des ensembles flous, fut également mise en œuvre par Daniel pour la reconnaissance d'objets sur le fond marin, mais en intégrant un facteur de fiabilité dans les valeurs de possibilités obtenues afin de surmonter les conflits entre sources [DANI-98]. Nous montrerons que la fiabilité des sources (points de vue dans notre cas) peut être prise en compte dès l'étape de fusion. L'opérateur de fusion agit alors de manière à pondérer défavorablement les points de vue non caractéristiques.

La **théorie des ensembles flous** permet de représenter les classes (caractérisées par les attributs) par des ensembles flous. Plutôt qu'une classification binaire $\{0,1\}$, la théorie des ensembles flous propose une graduation de l'appartenance à l'ensemble par le biais du degré d'appartenance à valeurs dans $[0,1]$. Quelle que soit la dimension du problème, cette donnée simple est aisément interprétable. Par définition, un ensemble flou (classe) A sur un espace de points (objets, observations ou éléments) $X = \{x\}$ est défini par une fonction d'appartenance $\mu_A(x)$ de sorte qu'à chaque élément $x \in X$ est associé un degré d'appartenance de x à A de valeur $\mu_A(x)$. A noter que, bien que la fonction d'appartenance s'apparente à une densité de probabilité, elles sont conceptuellement différentes avec des règles de combinaison et des propriétés propres. En effet, alors que la notion de probabilité est liée à l'occurrence d'un élément donné dans une population donnée, le degré d'appartenance mesure quant à lui la ressemblance de cet élément avec l'ensemble flou. Dans ce formalisme, la mesure M_k^n s'identifie à une fonction d'appartenance. L'intérêt des ensembles flous est de pouvoir assigner plusieurs classes à un individu donné avec des degrés d'appartenance différents et de repousser ainsi l'étape ultime de décision.

La théorie des ensembles flous fournit un cadre favorable pour la fusion multi-attributs de par la souplesse qu'elle offre à la définition de modèles. Vu la variété des attributs proposés dans la partie III, nous nous attacherons davantage à analyser correctement la caractérisation mono (attribut par attribut) ou multi-dimensionnelle (L attributs) des ensembles flous manipulés.

Les mesures floues proposées font intervenir une **fonction décroissante de la distance à l'ensemble flou**. Un modèle adapté à la distribution des éléments de l'ensemble maximise les chances de bonne décision.

1.2. Quelques fonctions décroissantes

Une fonction d'appartenance floue m_x est une fonction décroissante d'une distance d de $[0, \infty[$ dans $[0,1]$. Cette distance $d = d_k$ peut être calculée en L dimensions sur l'ensemble flou relatif à la classe C_k ou bien, attribut par attribut, i.e. $d = d_k|_{a_i}$. Cette fonction permet de donner à tout élément inconnu un degré d'appartenance compris entre zéro et un ('zéro' caractérise l'incompatibilité de l'élément testé avec l'ensemble flou).

La première fonction est une simple soustraction :

$$m_1(d) = \max\{0, 1 - d\}$$

Suivent deux autres fonctions strictement décroissantes [PAL-86] :

$$m_2(d) = \frac{1}{1 + d}$$

$$m_3(d) = \exp(-d)$$

Dans le cadre de la théorie des ensembles flous, de nombreuses fonctions d'appartenance ont été introduites pour représenter explicitement les informations imprécises. La S-fonction de Zadeh permet de réaliser un seuillage progressif suivant la valeur de l'attribut en posant, pour $x > 0$ [ZADE-75] :

$$S(x, \alpha, \beta, \gamma) = \begin{cases} 0.0 & \text{si } x < \alpha \\ 1.0 & \text{si } x \geq \gamma \\ 2 \frac{(x - \alpha)^2}{(\gamma - \alpha)^2} & \text{si } \alpha \leq x < \beta \\ 1 - 2 \frac{(x - \gamma)^2}{(\gamma - \alpha)^2} & \text{si } \beta \leq x < \gamma \end{cases}$$

où $\Delta\beta = \gamma - \alpha$ est la *zone floue* et β le *point de croisement*.

Afin de déterminer le degré de compatibilité ou d'appartenance, on utilise la fonction μ décroissante définie par $\mu(x) = 1 - S(x, \alpha, \beta, \gamma)$. Finalement, on a :

$$m_4(d) = \mu(d) = 1 - S(d, \alpha, \beta, \gamma)$$

Ces quatre premières fonctions sont illustrées sur la Figure 50 pour une distance allant de 0 à 1 par pas de 0,001.

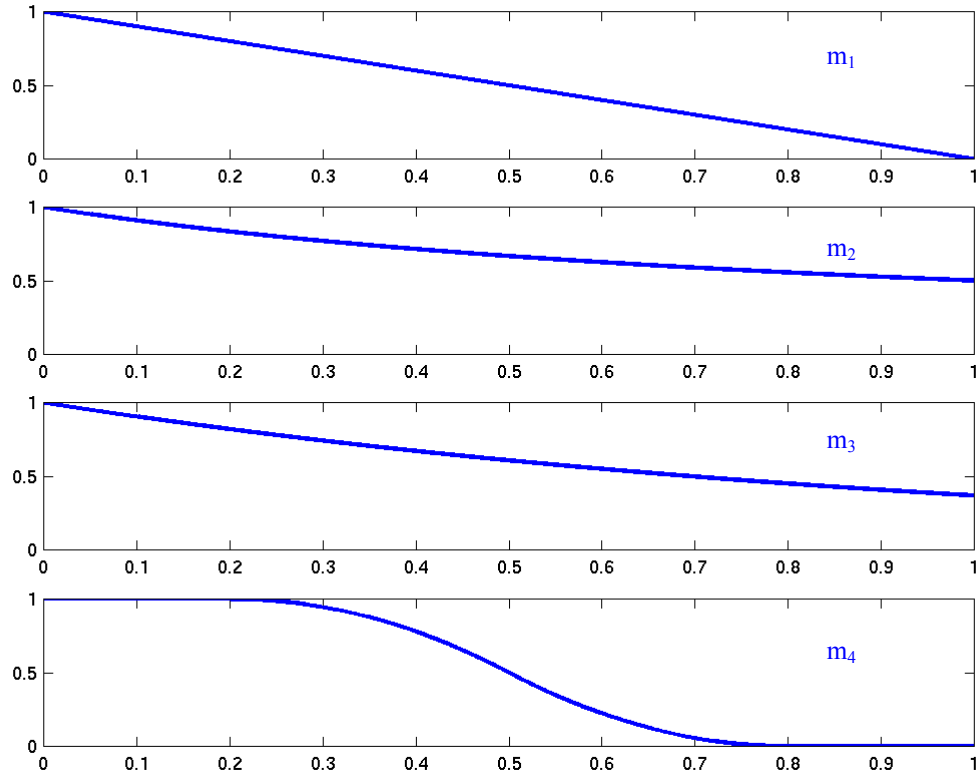


Figure 50 – Exemples de fonctions décroissantes utilisées pour le calcul de la fonction d'appartenance

Enfin, la mesure d'appartenance au cluster intervenant dans l'algorithme ISODATA flou ou fuzzy C-means permet de rapporter la distance à un ensemble donné à la somme des distances à chacun des ensembles [DUNN-74] [DANI-98a] :

$$m_5(d_k) = \frac{1/d_k}{\sum_{l \in \{1, \dots, K\}} 1/d_l}$$

1.3. Modèles flous

Globalement, un modèle flou est caractérisé par une distance à laquelle on associe une fonction décroissante de cette distance. Sa définition dépend de :

- 1) la dimension de l'espace caractéristique des éléments de l'ensemble flou
- 2) la configuration de l'ensemble flou
- 3) une distance d à l'ensemble flou

Le modèle permet d'affecter à chaque image courante I_n de la base de généralisation (élément inconnu à classifier) une mesure M_k^n à la classe C_k via une distance liant ses attributs a_i^n aux attributs a_i^k des exemples de la classe C_k . Le degré d'appartenance peut être calculé relativement à chaque attribut (espace mono-dimensionnel) ou bien sur l'ensemble des attributs (espace multi-dimensionnel). Le paragraphe ci-dessus propose différentes fonctions d'appartenance m_x . La mesure prend la forme suivante :

- mesure vectorielle $M_k^n = m_x(d_k)$ dans le cas d'un ensemble multi-dimensionnel
- mesure multi-attributs $M_k^n = \frac{1}{L} \sum_{i=1}^L M_k^n|_{a_i} = \frac{1}{L} \sum_{i=1}^L m_x(d_k|_{a_i})$ dans le cas d'un ensemble mono-dimensionnel

Elle est d'autre part étroitement liée à la caractérisation de l'ensemble flou suivant que sa configuration permet ou non de le réduire à un *élément idéal*.

Enfin, la formulation plus ou moins explicite d'une distance distingue chaque modèle. Sa pertinence dépend de la distribution des nuages de points des ensembles flous (mono ou multi-dimensionnels) pour l'application courante (cf. Figure 51). Les distances introduites répondent à un degré croissant de complexité, fonction des hypothèses de distributions des individus.

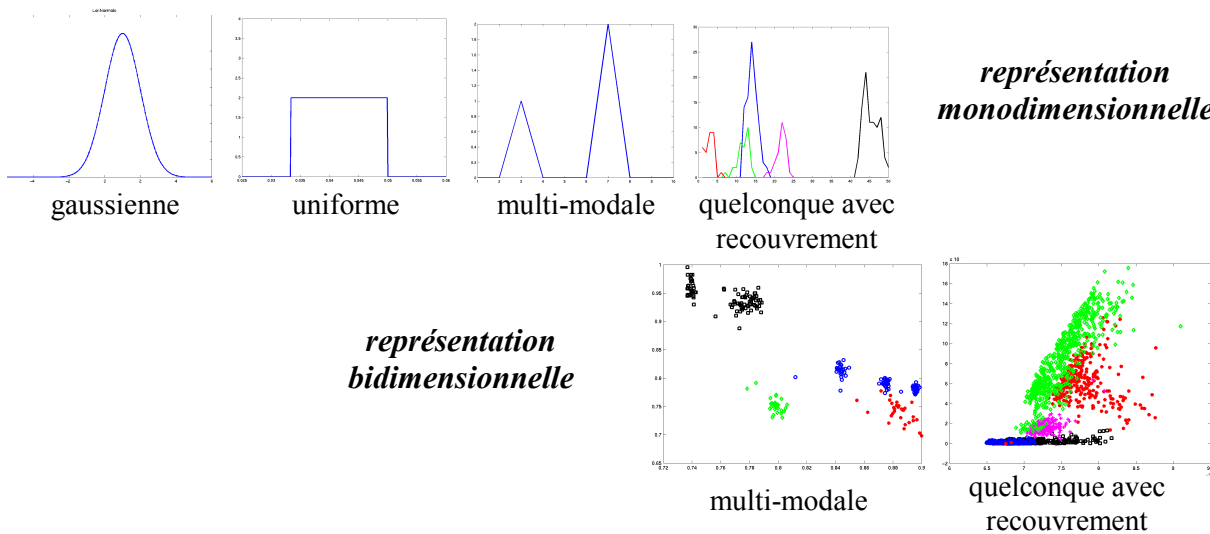


Figure 51 – Exemples de distributions mono et multidimensionnelles des individus

I.3.1. Par rapprochement à un élément idéal

Si l'on considère que la base est suffisamment importante, on peut en déduire, pour chaque classe, un élément idéal ou prototype. Le degré d'appartenance à la classe C_k est fonction de sa distance à l'élément idéal caractérisé par une ou plusieurs grandeurs statistiques calculées sur la base d'apprentissage.

I.3.1.A. Mesure vectorielle

Nous désirons rapprocher l'image courante I_n de vecteur caractéristique $V_n = (a_i^n)_{i \in \{1, \dots, L\}}^T$ de chacune des classes C_k pour $k \in \{1, \dots, K\}$ caractérisées par leur élément idéal $\bar{V}_k = (\bar{a}_i^k)_{i \in \{1, \dots, L\}}^T$ et la matrice de covariance Σ_k calculés à partir de la base d'apprentissage de N_k éléments. Le degré d'appartenance est alors inversement relié à la distance entre le vecteur caractéristique de l'image et celui de l'élément idéal [PAL-86].

La distance $D_0 = d_k = d(V_n, \bar{V}_k)$ entre l'image courante I_n et l'élément idéal de la classe C_k est la distance de Mahalanobis définie par :

$$D_0 = D_{Mal} = d_k = d(V_n, \bar{V}_k) = (V_n - \bar{V}_k) \Sigma_k^{-1} (V_n - \bar{V}_k)^T$$

avec la matrice de covariance

$$\Sigma_k = \left((\sigma_{ij}^k)^2 \right)_{i,j=1 \dots L}$$

avec

$$(\sigma_{ij}^k)^2 = \frac{1}{N_k} \sum_{l=1}^{N_k} (a_i^k(l) - \bar{a}_i^k) (a_j^k(l) - \bar{a}_j^k).$$

Les fonctions décroissantes m_2 et m_3 ont été utilisées par Pal et Dutta Majumder pour cette distance D_{Mal} [PAL-86].

Remarque : l'utilisation de la distance de Mahalanobis nécessite de centrer et réduire préalablement les attributs (cf. §II.2 du chap. III.1).

I.3.1.B. Mesure multi-attributs

Pour i donné, les attributs a_i calculés pour les objets d'une même classe C_k constituent un sous-ensemble flou dont l'élément idéal est caractérisé par un attribut représentatif pour cette classe C_k . Le degré d'appartenance de l'image courante à une classe est alors la combinaison de L degrés d'appartenance $M_k^n|_{a_i}$ calculés *relativement à chaque attribut* a_i . Nous souhaitons rapprocher chacun des attributs a_i^n caractérisant l'image courante I_n du sous-ensemble d'attributs $a_i^k(j)_{j \in \{1, \dots, N_k\}}$ des N_k éléments de la classe C_k . Les distances introduites ci-après font intervenir trois grandeurs statistiques : la valeur moyenne $\bar{a}_i^k$, la valeur médiane $med(a_i^k)$ et l'écart-type σ_i^k .

1) Distance indépendante de la distribution des données

$$D_1 = D_{seuil} = d_k|_{a_i} = \frac{|a_i^n - \bar{a}_i^k|}{\sigma_i^k}$$

Le lien entre frontière floue de l'ensemble et ses grandeurs statistiques associées apparaît clairement par le biais de cette distance. Le seuillage progressif effectué par la fonction m_4 traduit le comportement de cette frontière en autorisant une marge plus ou moins grande autour de la valeur moyenne. Cette marge correspond à la zone floue $\Delta\beta$ et dépend du paramétrage choisi (α, γ) . Trois cas possibles se présentent :

→ les deux cas extrêmes,

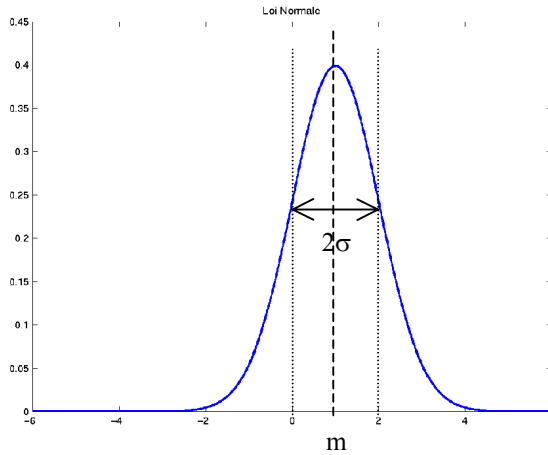
- cas où l'attribut courant a_i^n est incompatible avec la classe C_k car il s'écarte de l'attribut moyen de plus de γ fois l'écart-type,
- cas où il est pleinement compatible avec un écart avec l'attribut moyen inférieur à α fois l'écart-type,

→ le cas intermédiaire correspondant à la zone floue caractérisant l'imprécision.

Notons que la frontière sera nette si $\Delta\beta$ est nul, i.e. un modèle « tout ou rien ».

2) Distance basée sur l'hypothèse d'une distribution gaussienne des données

On suppose une distribution gaussienne des attributs relatifs à une classe donnée des images de la base d'apprentissage (cf. Figure 52).



Loi normale $N(m, \sigma)$, tel que $m = \sigma = 1$
de densité de probabilité

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(u-m)^2}{2\sigma^2}\right\}$$

Figure 52 - Distribution gaussienne

La surface totale comprise sous l'intégrale étant égale à un, la mesure recherchée est égale à cette surface moins la surface représentée par les images de la base d'apprentissage dont la valeur de l'attribut a_i^k est telle que :

$$\bar{a}_i^k - E_i(k, n) \leq a_i^k \leq \bar{a}_i^k + E_i(k, n)$$

avec $E_i(k, n) = |a_i^n - \bar{a}_i^k|$

En d'autres termes, on recherche la probabilité que la variable gaussienne U soit comprise dans l'intervalle $[m-x, m+x]$ c'est-à-dire :

$$\Pr(m-x \leq U \leq m+x) = \int_{m-x}^{m+x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(u-m)^2}{2\sigma^2}\right\} du = \operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma}\right)$$

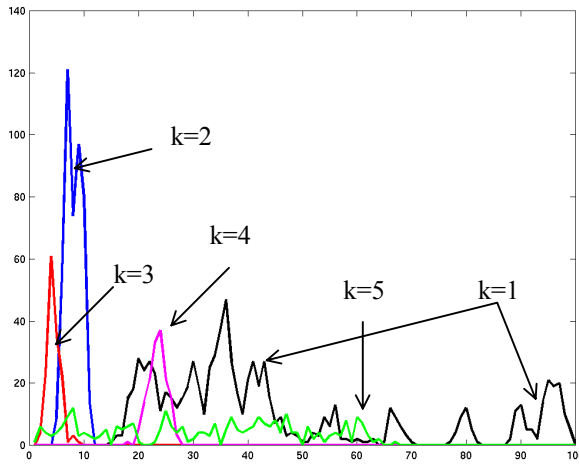
Par analogie, en prenant $U = a_i^k$, $m = \bar{a}_i^k$ et $x = E_i(k, n)$, alors :

$$D_2 = D_{gaus} = d_k|_{a_i} = \operatorname{erf}\left(\frac{E_i(k, n)}{\sqrt{2}\sigma_i^k}\right) = \operatorname{erf}\left(\frac{|a_i^n - \bar{a}_i^k|}{\sqrt{2}\sigma_i^k}\right)$$

3) Distance basée sur les histogrammes des attributs

Si l'hypothèse de distribution gaussienne n'est pas vérifiée, on peut raisonner directement sur les histogrammes. Soit l'histogramme de la Figure 53 des éléments de la base d'apprentissage pour un attribut a_i donné et $K=5$ classes, en rangeant les attributs a_i^k par ordre croissant, on peut connaître le nombre d'exemples N_b^k dont la valeur de l'attribut est inférieure ou égale à une valeur ponctuelle a_i^k (graphe de gauche de la Figure 54). En normalisant par rapport au nombre $N_{b_{tot}}^k$ d'exemples pour la classe C_k , on en déduit facilement la fonction de répartition F_H associée à l'histogramme telle que (graphe de droite de la Figure 54) :

$$F_H^k(x) = \frac{N_b^k(x)}{N_{b_{tot}}^k}$$



Histogramme pour l'attribut a_i et les images de la base d'apprentissage représentatives de $K=5$ classes d'objets. L'espace des valeurs admissibles pour cet attribut est divisé en 100 intervalles.

Figure 53 - Histogramme pour un attribut donné

$F_H^k(x)$ représente la part des individus dont la valeur de l'attribut a_i^k est inférieure à x . En particulier, pour chaque classe C_k , on a :

$$N_b^k(\operatorname{med}(a_i^k)) = N_{b_{tot}}^k / 2 \text{ ou encore } F_H^k(\operatorname{med}(a_i^k)) = 0.5$$

L'objectif étant de comptabiliser le nombre d'exemples N_k' de la base d'apprentissage pour la classe C_k dont la valeur de l'attribut a_i^k est plus proche de la valeur optimale $\operatorname{med}(a_i^k)$ que la valeur courante a_i^n , c'est-à-dire :

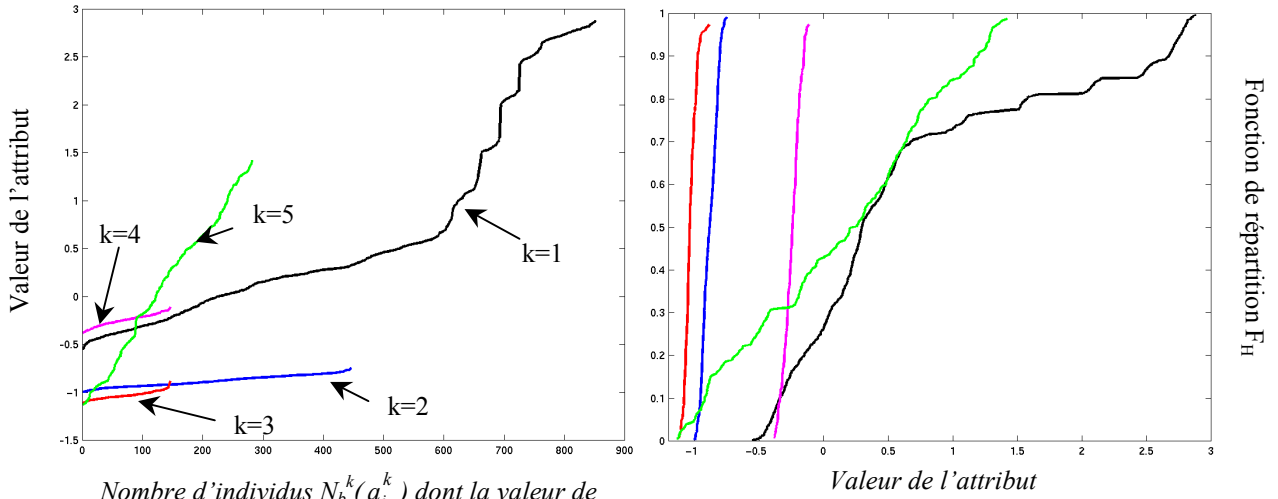
$$\bar{a}_i^k - E_i(k, n) \leq a_i^k \leq \bar{a}_i^k + E_i(k, n)$$

avec $E_i(k, n) = |a_i^n - \text{med}(a_i^k)|$

Ce nombre est égal à (cf. Figure 55) :

$$N'_k = 2 \times |F_H^k(a_i^n) - F_H^k(\text{med}(a_i^k))| \times N_{\text{btot}}^k = 2 \times |F_H^k(a_i^n) - 0.5| \times N_{\text{btot}}^k = |2F_H^k(a_i^n) - 1| \times N_{\text{btot}}^k$$

La Figure 55 illustre le cas où $a_i^n < \text{med}(a_i^k)$.



Nombre d'individus $N_b^k(a_i^k)$ dont la valeur de l'attribut est inférieure ou égale à l'ordonnée a_i^k

Figure 54 – Fonction de répartition correspondante à l'histogramme de la Figure 53

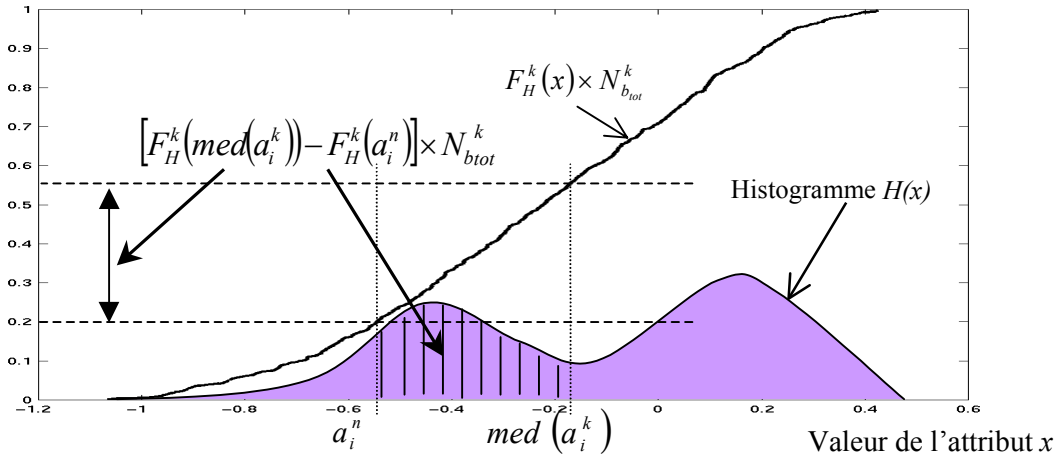


Figure 55 – Individus à comptabiliser

D'une manière générale, on peut donc définir la distance par :

$$D_3 = D_{\text{hist}} = d_k|_{a_i} = \frac{N'_k}{N_{\text{btot}}^k} = |2F_H^k(a_i^n) - 1|$$

4) Distance basée sur les histogrammes des attributs avec une prise en compte des recouvrements

L'histogramme donné plus haut montre que le recouvrement entre les classes peut être important. Nous avons donc introduit cette donnée et modifié la mesure précédente.

Pour ce faire, il nous faut connaître, pour chaque classe $C_k \neq C_k$, le nombre d'exemples N'_k dont la valeur de l'attribut est comprise entre les bornes a_i^n et $med(a_i^k)$ soit :

$$N'_k = |F_H^{k'}(a_i^n) - F_H^{k'}(med(a_i^k))| \times N_{btot}^{k'}$$

$$D_4 = d_k|_{a_i} = D_3 + \text{terme de prise en compte du recouvrement} = \frac{N'_k}{N_{btot}^k} + \frac{\sum_{k', k' \neq k} N'_k}{\sum_{k', k' \neq k} N_{btot}^{k'}}$$

$$D_4 = D_{histRcvl} = |1 - 2F_H^k(a_i^n)| + \frac{\sum_{k', k' \neq k} |F_H^{k'}(a_i^n) - F_H^{k'}(med(a_i^k))| \times N_{btot}^{k'}}{\sum_{k', k' \neq k} N_{btot}^{k'}}$$

En somme, si le recouvrement est trop important (distance importante) alors l'attribut courant n'est pas affecté à cette classe C_k . Cette propriété permet de pénaliser les attributs trop dispersés qui sont donc peu discriminants.

1.3.2. Par rapprochement à l'ensemble flou

Pour un ensemble non compact, il est plus prudent d'élaborer la mesure d'appartenance à partir de l'ensemble ou d'une partie des N_k exemples de chacune des classes de la base d'apprentissage. Les distances présentées font appel à une **analyse des données** qui s'effectue *hors ligne* c'est-à-dire une fois pour toutes en début de traitement.

D'une part, basée sur un partitionnement de l'ensemble des informations disponibles, une méthode métrique type 'clustering' réduit chaque classe à un nombre limité de points représentatifs. Deux méthodes statistiques d'autre part font appel à l'analyse de la répartition ou distribution locale des observations. Sans aucune connaissance a priori sur la distribution des données à analyser, on s'intéresse alors à la densité d'un nombre fixé de voisins autour du point courant.

1.3.2.A. Distance basée sur un nombre limité de points représentatifs

Chacun des ensembles flous de la base d'apprentissage associés aux K classes peut être représenté par un nombre limité de points particuliers représentatifs. Les techniques de 'clustering' permettent de partitionner l'ensemble d'éléments $V_k(j) = (a_i^k(j))^T$, $j \in \{1, \dots, N_k\}$ de la classe C_k en C sous-ensembles, les 'clusters', suivant un critère d'optimisation. Dans un même cluster, les différents points présentent alors un degré de ressemblance. La méthode des *fuzzy C-means* ou *ISODATA flou* d'abord introduite par Dunn et dont Bezdek montra la convergence consiste à minimiser le critère suivant [DUNN-74] [BEZD-81] :

$$J(U, v) = \sum_{j=1}^{N_k} \sum_{i=1}^C (u_{ij})^m (d_{ij})^2 \text{ avec } d_{ij}^2 = \|V_k(j) - v_i\|^2$$

où

U est la matrice de partition de dimension $N_k \times C$ dont l'élément u_{ij} désigne le degré d'appartenance floue (compris entre 0 et 1) du point j au cluster i représenté par son centroïde v_i

$m \geq 1$ définit la nature plus ou moins floue du clustering qui augmente avec m

Le détail de la méthode est donné dans l'annexe H.

En appliquant cette méthode à chacun des ensembles flous de la base d'apprentissage associés aux K classes, la distance d'un individu inconnu V_n à un ensemble donné est obtenue à partir des $C=10$ centroïdes obtenus au lieu des N_k éléments.

La distance à l'ensemble flou représentatif de la classe C_k devient :

$$D_5 = D_{clust} = d_k = \min_{i, i \in [1, C]} \|V_n - v_i^k\|$$

Distance euclidienne minimale entre l'élément inconnu et chacun des C clusters (i.e. leurs centroïdes v_i) de la classe C_k

Ce traitement évite de retenir un élément mal caractérisé de l'ensemble qui n'est donc pas représentatif.

1.3.2.B. Distance basée sur la densité des éléments au voisinage de l'élément courant

Ces distances ne requièrent aucune hypothèse sur la distribution des attributs. Cela permet une adaptation fine à des distributions différentes suivant les classes, voisine de l'approche proposée par Czesnalowicz pour l'estimation d'une densité de probabilité à partir des k plus proches observations [CZES-92]. Les méthodes suivantes dites non-paramétriques (les estimations se font sur la base d'un nombre fini d'échantillons) sont notamment adaptées à une distribution multi-modale des attributs.

a) A partir de la fonction de répartition associée à l'histogramme

A partir de la fonction de répartition associée à l'histogramme (cf. §1.3.1.B), on peut définir une nouvelle distance basée sur la configuration des exemples de l'ensemble flou au voisinage de l'attribut courant a_i^n . Ainsi, pour une classe C_k donnée et attribut par attribut, on s'intéresse à la densité ρ_i^k , ou inversement la dispersion, des individus au voisinage de a_i^n . Au sens d'une distance, l'élément inconnu sera d'autant plus proche de l'ensemble qu'il se situe dans une zone dense.

Soit $\Delta_{ex} = \alpha N_k$ la proportion des exemples recherchée autour de l'élément inconnu, la dispersion est égale à l'intervalle des valeurs $\Delta_i^k(a_i^n)$ occupé par les Δ_{ex} exemples au voisinage de l'élément inconnu caractérisé par a_i^n . Etant donné que la dynamique diffère d'un attribut à l'autre, la dispersion observée par attribut doit être rapportée à l'intervalle total Δ_i des valeurs occupé par l'ensemble des exemples de la base d'apprentissage. La distance calculée *relativement à chaque attribut* a_i mène au calcul d'une mesure multi-attributs et est définie par :

$$D_6 = D_{dens} = d_k|_{a_i} = \frac{1}{\rho_i^k(a_i^n)} = \frac{\Delta_i^k(a_i^n)}{\Delta_i}$$

La valeur de $\Delta_i^k(a_i^n)$ est facilement obtenue à partir de la fonction de répartition F_H^k définie au §I.3.1.B. En effet, la fonction $F_H^k \times N_k$ ci-dessous (en trait plein sur la Figure 56) correspond à la répartition exacte des exemples d'une classe C_k donnée en fonction de l'attribut a_i . Etant donnée une proportion α des exemples pour chaque classe C_k , soient Δ_{ex} exemples voisins de la valeur a_i^n , on a donc différents intervalles $\Delta_i^k(a_i^n)$ dont la largeur dépend de la dispersion des exemples voisins de l'élément inconnu. Dans les cas particuliers où a_i^n n'appartient pas à l'intervalle admissible des attributs de la classe C_k (cas de a_i^3), celui-ci est d'autant plus pénalisé qu'il en est éloigné.

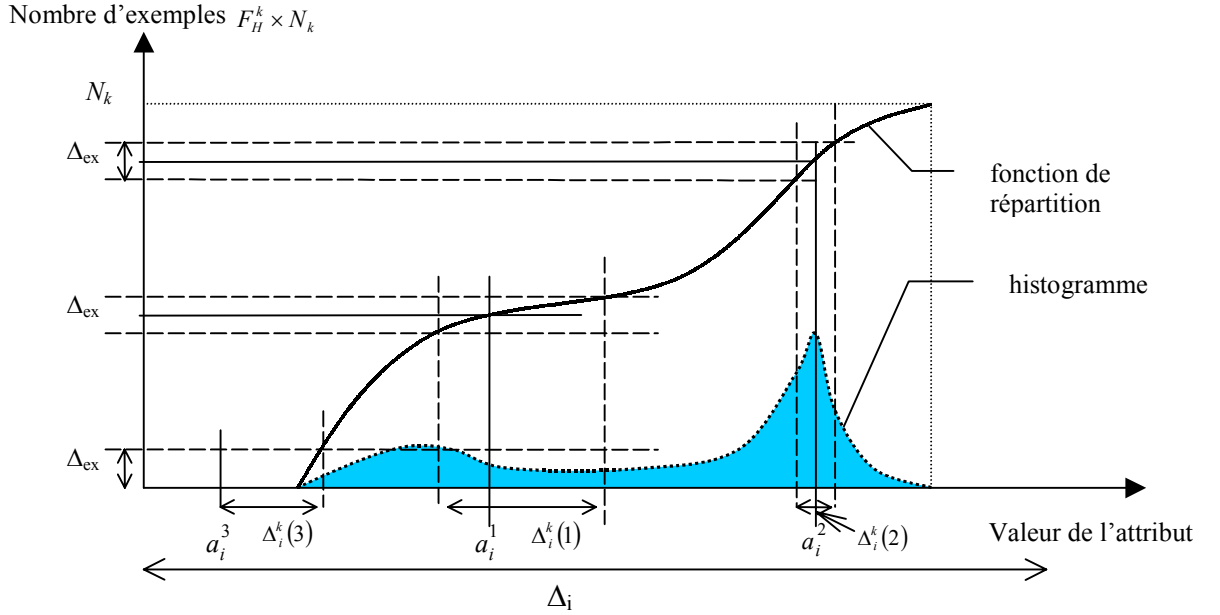


Figure 56 – Différents cas pour le calcul de l'intervalle $\Delta_i^k(a_i)$

Remarque : le paramétrage utilisé par la suite est $\alpha=0.1$.

b) A partir du potentiel

Un potentiel mesure la proximité d'un point à un ensemble ou classe. Par exemple, une estimation de la densité locale autour de ce point est inversement obtenue par le potentiel de Mizoguchi-Shimura défini comme la distance moyenne d'un point à ses k plus proches voisins [MIZO-80]. A cette première définition, Grenier ajouta un terme marquant l'éloignement à ce point [GREN-84]. Cette nouvelle fonction de potentiel s'apparente à un degré d'appartenance à l'ensemble (classe) qui atteint son maximum là où la densité d'observations (ou éléments) est maximale [CHAR-96]. La formulation est la suivante :

Chaque observation décrite par le vecteur caractéristique V_j d'une classe C_k induit une fonction définie en tout point de vecteur V de l'espace des observations par

$$\Pi_k(V, V_j) = P_k(V_j) + \frac{D(V_j) d^2(V, V_j)}{P_k^2(V_j)}$$

avec

$P_k(V_j)$ potentiel de Mizoguchi-Shimura basée sur k_{ppv} voisins V_{jm} dans C_k ,

$$P_k(V_j) = \frac{1}{k_{ppv}} \sum_{m=1}^{k_{ppv}} d(V_j, V_{jm})$$

$D(V_j)$ la variation de densité locale autour de V_j

$$D(V_j) = \frac{1}{k_{ppv}} \sum_{m=1}^{k_{ppv}} |P_k(V_j) - P_k(V_{jm})|$$

$\Pi_k(V, V_j)$ est minimale pour $V = V_j$ de sorte que la seule densité locale est prise en compte. Elle croît par ailleurs de manière parabolique avec la distance euclidienne $d(V, V_j)$

pondérée par le terme $\frac{D(V_j)}{P_k^2(V_j)}$ vu comme un gradient contrôlant la distribution des densités.

Pour un élément inconnu de vecteur caractéristique V_n , la distance à une classe est égale à la fonction de potentiel minimale induite par un élément de cette classe, soit :

$$D_\gamma = D_{pot} = d_k = \min_{j=1, \dots, N_k} [\Pi_k(V_n, V_j)]$$

Les trois modèles présentés font appel à une analyse des données qui résume avantageusement, en début de traitement, l'information relative à l'ensemble des éléments de chacune des classes de la base d'apprentissage.

Remarque : les distances D_{clus} et D_{pot} menant au calcul d'une mesure vectorielle nécessitent de centrer et réduire les attributs.

1.4. Opérateurs de fusion

Une fois les mesures M_n^k obtenues à partir des N vues (N sources), il nous faut les combiner par un opérateur de fusion [BLOC-96a]. Compte tenu des données à fusionner et de la procédure d'acquisition de ces données, notre choix se porte naturellement sur des opérateurs F de compromis correspondant à un comportement prudent. En l'occurrence, les **opérateurs de moyenne** semblent bien adaptés à notre problème. Nous distinguons deux opérateurs :

- un opérateur F_m traitant indifféremment les mesures obtenues : il agit donc indépendamment du contexte,
- un opérateur F_{mp} faisant intervenir la fiabilité des sources : il tient compte du contexte.

Premier cas : si le résultat de la fusion ne doit dépendre que des valeurs à combiner en ne faisant intervenir aucune autre information, on peut définir F_m par

$$F_m(M_k^n, 1 \leq n \leq N) = \frac{\sum_{n=1}^N M_k^n}{N}, \text{ i.e. effectuer une } \underline{\text{moyenne arithmétique}}$$

Second cas : on peut introduire une information supplémentaire relative à la fiabilité des sources, et en particulier à la pertinence du point de vue considéré. En effet, pour des objets à géométrie complexe, certains points de vue peuvent ne pas être caractéristiques et être sources

de confusion. Nous proposons donc de pondérer les mesures en fonction du point de vue, tel que

$$F_{mp}(M_k^n, 1 \leq n \leq N) = \frac{\sum_{n=1}^N \alpha_n M_k^n}{\sum_{n=1}^N \alpha_n}, \text{ i.e. effectuer une } \underline{\text{moyenne arithmétique pondérée}}$$

Une manière simple de détecter les points de vue non caractéristiques est d'évaluer le rapport $L/H = \text{largeur}/\text{hauteur}$ (cf. §IV.1 du chap. II.2). En effet, pour des images normalisées en rasant, plus ce rapport est faible moins le point de vue considéré est fiable. La mesure

relative à la $n^{\text{ième}}$ image peut être pondérée par : $\alpha_n = \frac{L_n}{H_n}$.

L'influence de ce rapport est particulièrement utile pour les mines cylindriques et Rockan dont on peut minorer l'influence des points de vue non caractéristiques voire indésirables car peu discriminants. En moyenne, pour une séquence de 19 images acquises le long d'une trajectoire de 180 degrés autour de l'objet cylindrique (resp. mine Rockan), α_n a l'allure obtenue sur la Figure 57 (resp. Figure 58).

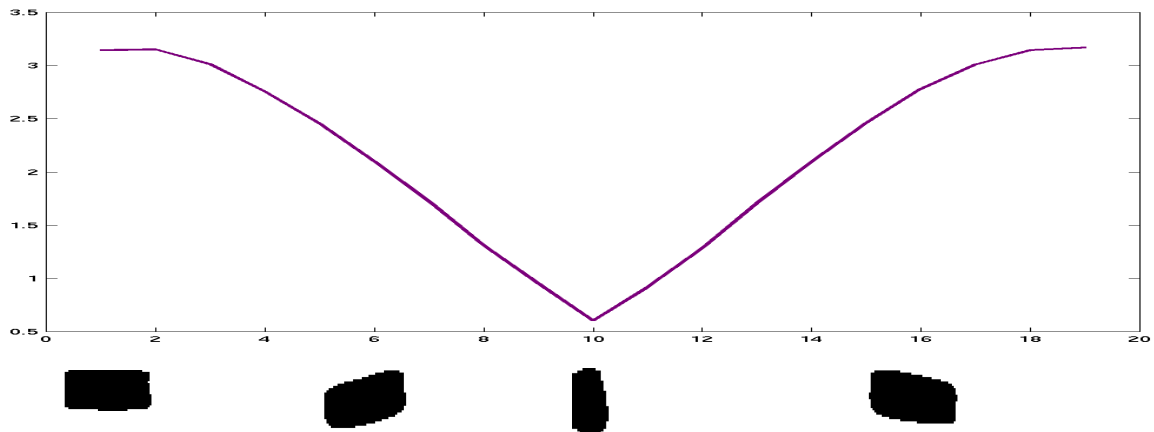


Figure 57 – Evolution du rapport largeur sur hauteur pour le cylindre le long d'une trajectoire semi-circulaire

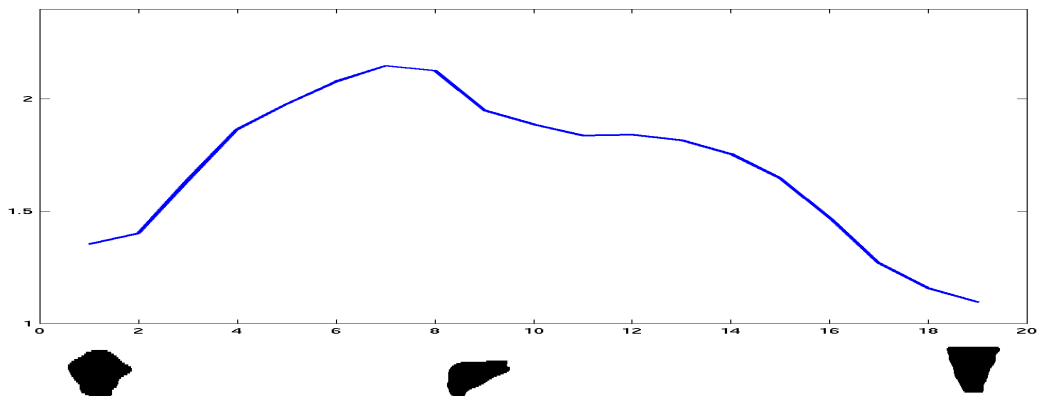


Figure 58- Evolution du rapport largeur sur hauteur pour la mine Rockan le long d'une trajectoire semi-circulaire

Il apparaît clairement qu'un point de vue indésirable tel que, pour le cylindre, la ligne de visée parallèle à la génératrice (10^{ème} point de vue sur la Figure 57) aura moins d'influence sur le résultat final de fusion de l'ensemble des N images quels que soient les points de vue si on prend en compte ce critère. Pour la mine Rockan, les points de vue indésirables correspondent à une vue de front ou par l'arrière (extrémités de la courbe de la Figure 58).

1.5. Décision

L'étape ultime consiste à prendre une décision à partir des mesures combinées :

$$M_l(f) = F(M_l^n(f), 1 \leq n \leq N)$$

L'objet ciblé de forme f est affecté à la classe C_k si :

$$M_k(f) = \max_{l \in \{1, \dots, K\}} \{M_l(f)\}$$

1.6. Résultats expérimentaux

Nous avons évalué ces modèles sur une base d'images simulées comptabilisant plus de 5000 prises de vues représentant cinq classes ou types de mines choisis (cf. annexe G et §IV et V de la partie I). Ces vues sont prises au cours d'une trajectoire circulaire du sonar autour de l'objet (même altitude, même distance). Nous proposons de classer les objets à partir de N vues dont les angles de prise de vue sont inconnus.

D'une part, on considère des **attributs hybrides** (L=4 attributs) : deux paramètres topologiques associés à deux invariants de moments (cf. §III.3.3 du chap. II.1). D'autre part, les modules des **descripteurs de Fourier** obtenus par une décomposition en série de Fourier des abscisses d'une part, et des ordonnées d'autre part, des points de contour de l'ombre puis normalisés par rapport au point de départ et en échelle $\left(\left| \tilde{X}_k \right|, \left| \tilde{Y}_k \right| \right)_{k \in \{1, \dots, 5\}}$, i.e. L=10 attributs (cf.

§III.1 du chap. II.2). De manière à utiliser cette base quel que soit le modèle choisi, nous effectuons d'emblée un centrage et une réduction des données. Après un mélange aléatoire de la base, elle est divisée en base d'apprentissage et base de généralisation. Après entraînement (détermination des caractéristiques de l'élément idéal ou analyse des données), on calcule, pour chacun des modèles, les mesures moyennes $\bar{M}_k = \sum_{objets} M_k = \sum_{objets} F(M_k^n, 1 \leq n \leq N)$

obtenues par sous-ensemble d'objets de la base de généralisation.

1.6.1. Quelques résultats obtenus par l'opérateur F_m

Nous rappelons que :

$$m_x \in \{m_1, \dots, m_5\} \text{ et } D_y \in \{D_{Mal}, D_{seuil}, D_{gaus}, D_{hist}, D_{histRcvt}, D_{clust}, D_{dens}, D_{pot}\}$$

L'ensemble des associations possibles (m_x, D_y) est donc trop important, et nous ne donnons que quelques résultats probants ou menant à d'intéressantes observations. Des tests effectués à partir des deux types d'attributs pour N=3 vues (quelconques entre 0 et 180 degrés) seulement nous ont permis de comparer quelques modèles dont l'approche est 'voisine'. Les meilleurs modèles ont de bonnes performances avec un taux de classification élevé (>90%) tout en

maintenant un faible écart-type sur les résultats (ceci garantit qu'une classe n'est pas défavorisée par rapport aux autres).

1) Modèles multi-dimensionnels

L'ensemble flou représentatif de la classe C_k peut être réduit à un élément idéal par la distance D_{Mal} ou partitionné suivant le principe de calcul de D_{clus} . Pour la même fonction m_2 , nous traçons sur la Figure 59 l'évolution des performances moyennes (toutes classes confondues) en pourcentage de bonne classification pour un nombre de vues allant de 3 à 10 prises seulement. Nous notons une nette différence des résultats pour les attributs hybrides. Sur le Tableau 13, le détail des taux de bonne classification par classe est donné dans le cas de $N=3$ prises de vue seulement pour les attributs hybrides : la nette amélioration des performances grâce à la partition souligne le risque associé à la réduction d'un ensemble à un élément idéal.

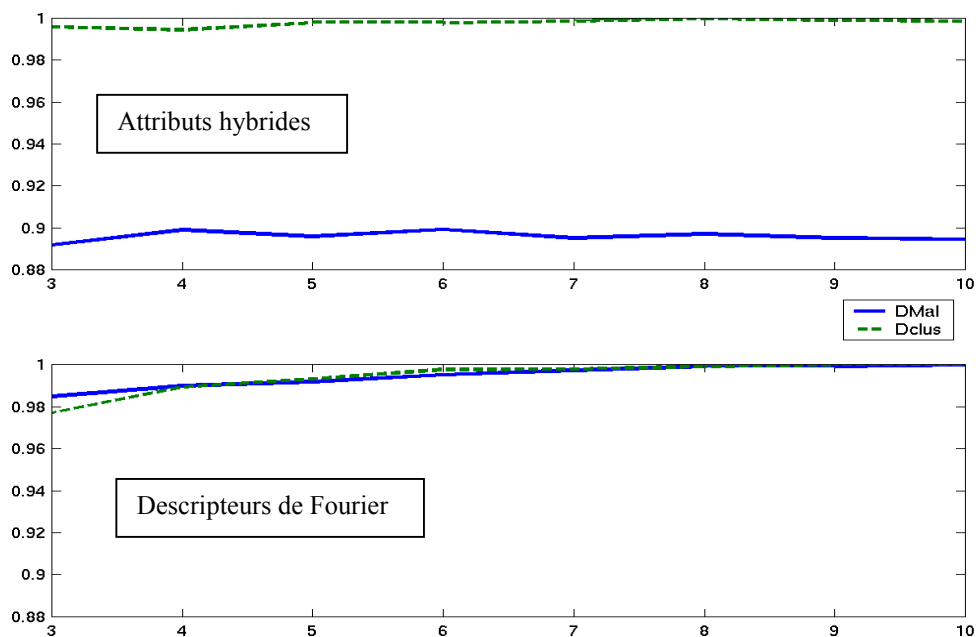


Figure 59 – Evolution des performances moyennes pour N de 3 à 10 vues: risque associé à la réduction d'un ensemble à un élément idéal

Tableau 13 – Performances pour les attributs hybrides et $N=3$ vues

	C_1 (cylindre)	C_2 (sphère)	C_3 (mine Manta)	C_4 (mine Sigeel)	C_5 (mine Rockan)	Moyenne (Ecart-type)
D_{Mal}	99	100	100	47	100	89.2 (23.6)
D_{clus}	100	100	100	100	98	99.6 (0.9)

2) Paramétrage du seuillage de la fonction m_4 associée à la distance D_{seuil}

Comme précédemment, la Figure 60 illustre l'évolution des performances pour seuillage net sur les frontières des ensembles flous, i.e. $\Delta\beta = 0$ et un seuillage flou, i.e. $\Delta\beta \neq 0$ avec

$(\alpha, \gamma) = (1, 1.5)$. La comparaison des courbes montre l'intérêt du second par rapport au premier quels que soient les attributs.

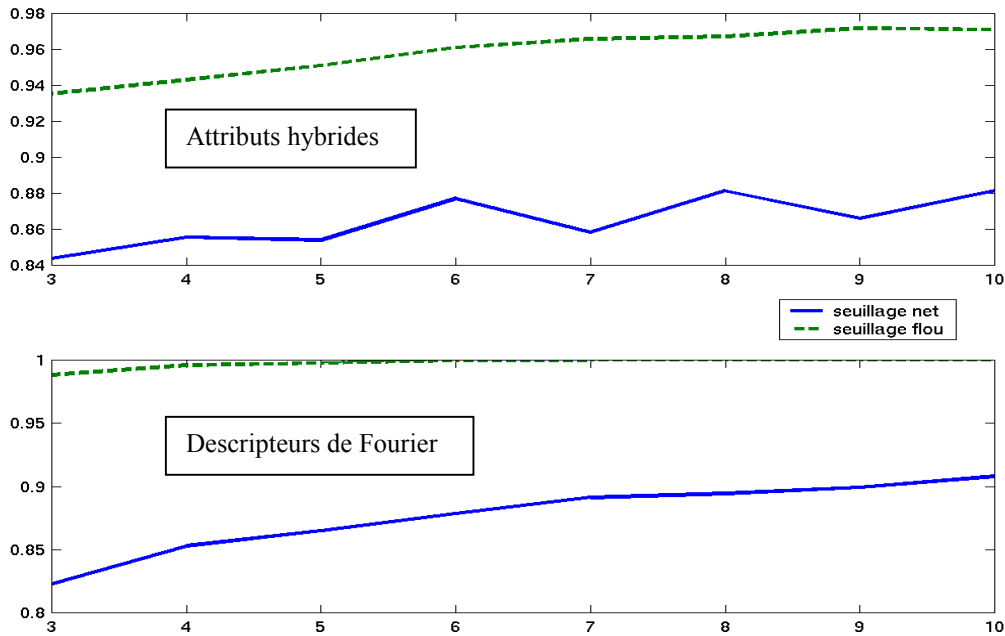


Figure 60 - Evolution des performances moyennes pour N de 3 à 10 vues: influence du seuillage

3) Importance de la distribution des attributs

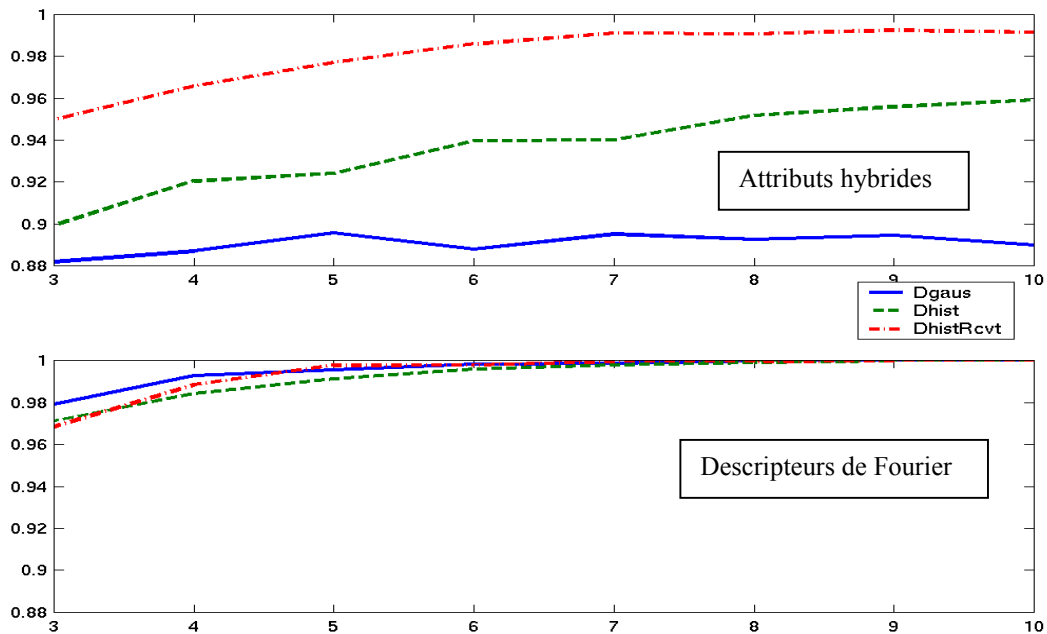


Figure 61 - Evolution des performances moyennes pour N de 3 à 10 vues: apport d'une analyse fine de la distribution des attributs

Pour le cas des attributs hybrides, une analyse plus ou moins fine de la distribution des données (distances D_{gaus} , D_{hist} et $D_{histRcvl}$ avec la même fonction m_1) influe sur les résultats

comme l'indiquent le Tableau 14 et la Figure 61. Lorsqu'on compare les distances D_{hist} et $D_{histRcv}$ toutes deux basées sur l'analyse des histogrammes des attributs, les performances sont bien meilleures pour $D_{histRcv}$ ce qui démontre l'intérêt de prendre en compte le recouvrement des histogrammes pour ces attributs. Les descripteurs de Fourier ont une distribution voisine de la distribution gaussienne ce qui explique les performances quasi-identiques des 3 modèles.

Tableau 14 - Performances pour les attributs hybrides et N=3 vues

	C ₁ (cylindre)	C ₂ (sphère)	C ₃ (mine Manta)	C ₄ (mine Sigeel)	C ₅ (mine Rockan)	Moyenne (Ecart-type)
D_{gaus}	99.7	99.7	96	49	96.7	88.2 (22.0)
D_{hist}	95.3	99.8	87	68.3	99.3	89.9 (13.1)
$D_{histRcv}$	95.7	100	94.3	89.3	95.7	95 (3.8)

4) Evolution des performances en fonction du nombre de vues

On cherche à présent à illustrer la dégradation plus ou moins marquée de quelques modèles performants (fort taux moyen de bonne classification avec un faible écart-type) en fonction du nombre de vues. Pour les attributs hybrides, la distance D_{pot} donne 100% de classification quel que soit le nombre de points de vue. Ce cas n'est donc pas illustré ici. Par type d'attributs (hybrides ou descripteurs de Fourier), deux modèles ont été retenus dans les deux cas de figure possibles c'est-à-dire caractérisation par un élément idéal ou prise en compte de l'ensemble dans sa globalité, soit :

- les modèles $(m_1, D_{histRcv})$ et (m_2, D_{dens}) dans le cas des attributs hybrides
- les modèles (m_4, D_{seuil}) (cas flou pour m_4) et (m_5, D_{clus}) dans le cas des descripteurs de Fourier

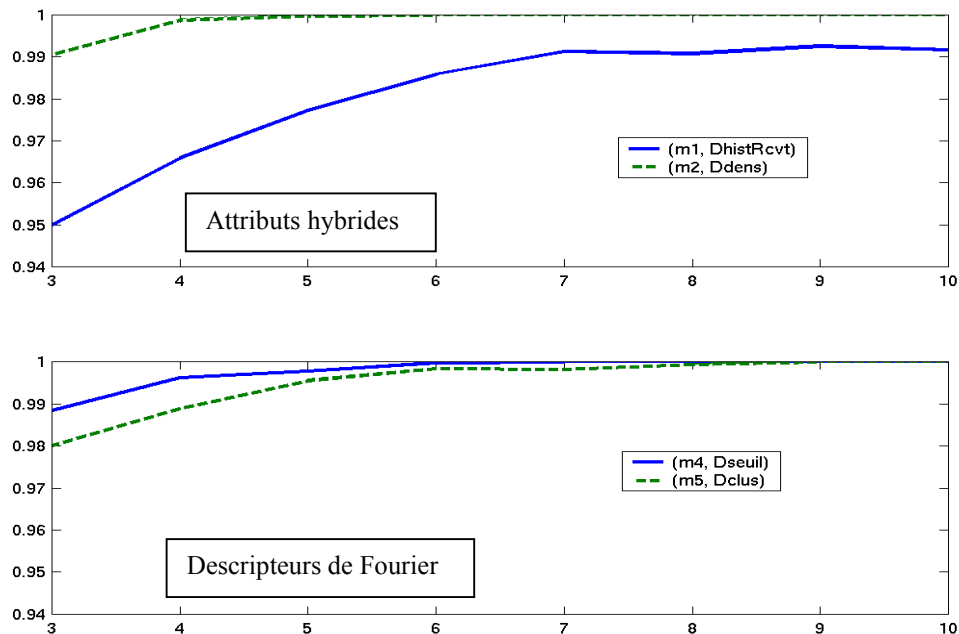


Figure 62 - Evolution des performances moyennes pour N de 3 à 10 vues: influence du nombre de vues

La Figure 62 montre que la dégradation n'intervient réellement que pour un nombre de vues strictement inférieur à 6.

I.6.2. Quelques résultats obtenus par l'opérateur F_{mp}

Pour deux ensembles de modèles, l'un constitué de quatre modèles performants, l'autre de quatre modèles peu performants, on visualise sur la Figure 63 l'effet de la pondération des points de vue sur le taux de classification moyen toutes classes confondues surtout lorsque l'on dispose de peu de prises de vue. La Figure 64 donne le détail pour les classes des mines cylindriques et Rockan les plus favorisées par cet opérateur de fusion.

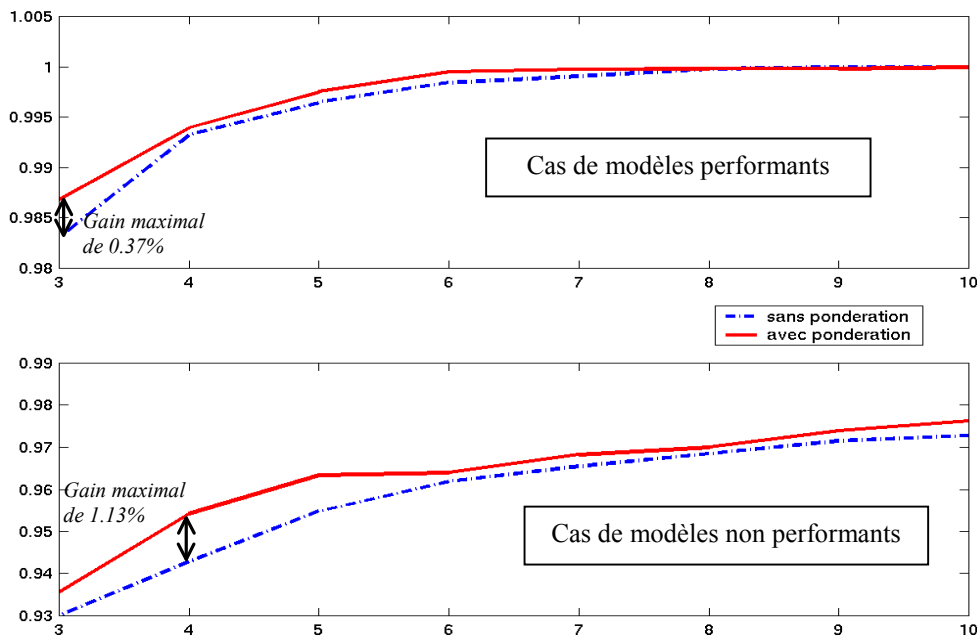


Figure 63 – Effet moyen de la pondération des points de vue

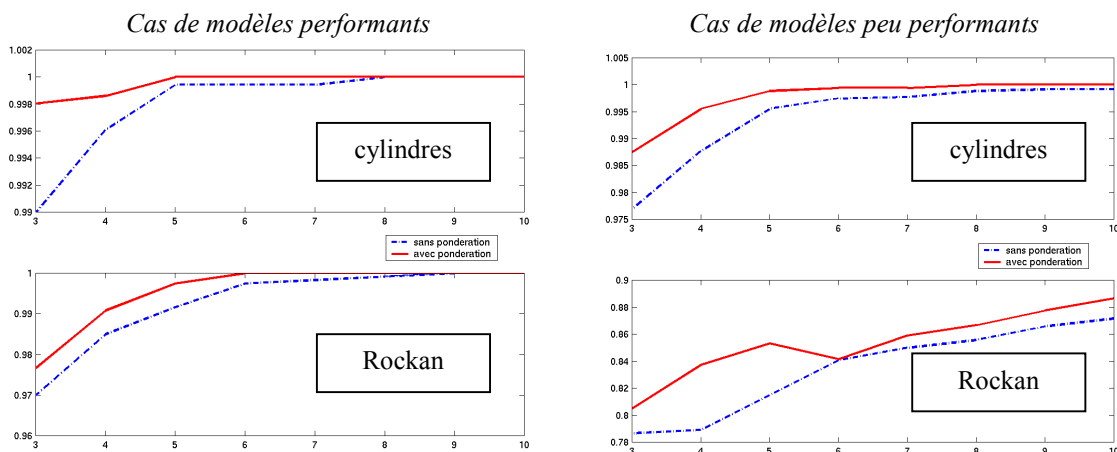


Figure 64 – Effet de la pondération des points de vue pour les mines cylindriques (en haut) et Rockan (en bas) pour quelques modèles performants (à gauche) et peu performants (à droite)

I.7. Visualisation graphique

I.7.1. Opérateur F_m

Figure 65 et Figure 66 illustrent les mesures $\bar{M}_k$ obtenues avec l'opérateur F_m pour chaque sous-ensemble d'objets (cylindres, sphères, mines Manta, Sigeel et Rockan) testés de la base de généralisation conditionnellement à la classe C_k en fonction des $K=5$ classes. Les graphiques représentent les mesures moyennes obtenues pour des modèles performants et seulement trois vues. Pour chaque sous-ensemble, un lot de mesures est représenté par des barres colorées dont la hauteur reflète le degré de compatibilité $\in [0,1]$ avec une classe donnée C_k :

Barre 1 (mauve)	$\Leftrightarrow \bar{M}_1 \Leftrightarrow$	Classe 1 $\Leftrightarrow$	'cylindre'
Barre 2 (bleue)	$\Leftrightarrow \bar{M}_2 \Leftrightarrow$	Classe 2 $\Leftrightarrow$	'sphère'
Barre 3 (verte)	$\Leftrightarrow \bar{M}_3 \Leftrightarrow$	Classe 3 $\Leftrightarrow$	'mine Manta'
Barre 4 (orange)	$\Leftrightarrow \bar{M}_4 \Leftrightarrow$	Classe 4 $\Leftrightarrow$	'mine Sigeel'
Barre 5 (marron)	$\Leftrightarrow \bar{M}_5 \Leftrightarrow$	Classe 5 $\Leftrightarrow$	'mine Rockan'

Les mesures obtenues font apparaître l'ambiguïté entre classes pour un sous-ensemble d'objets. Le résultat de bonne classification est d'autant plus robuste que, pour chaque type d'objets, la mesure de compatibilité la plus forte (menant à la prise de décision correcte dans les cas illustrés) est importante vis-à-vis des mesures de compatibilité avec les autres classes. Ce critère peut guider le choix de la fonction m_x pour une distance donnée.

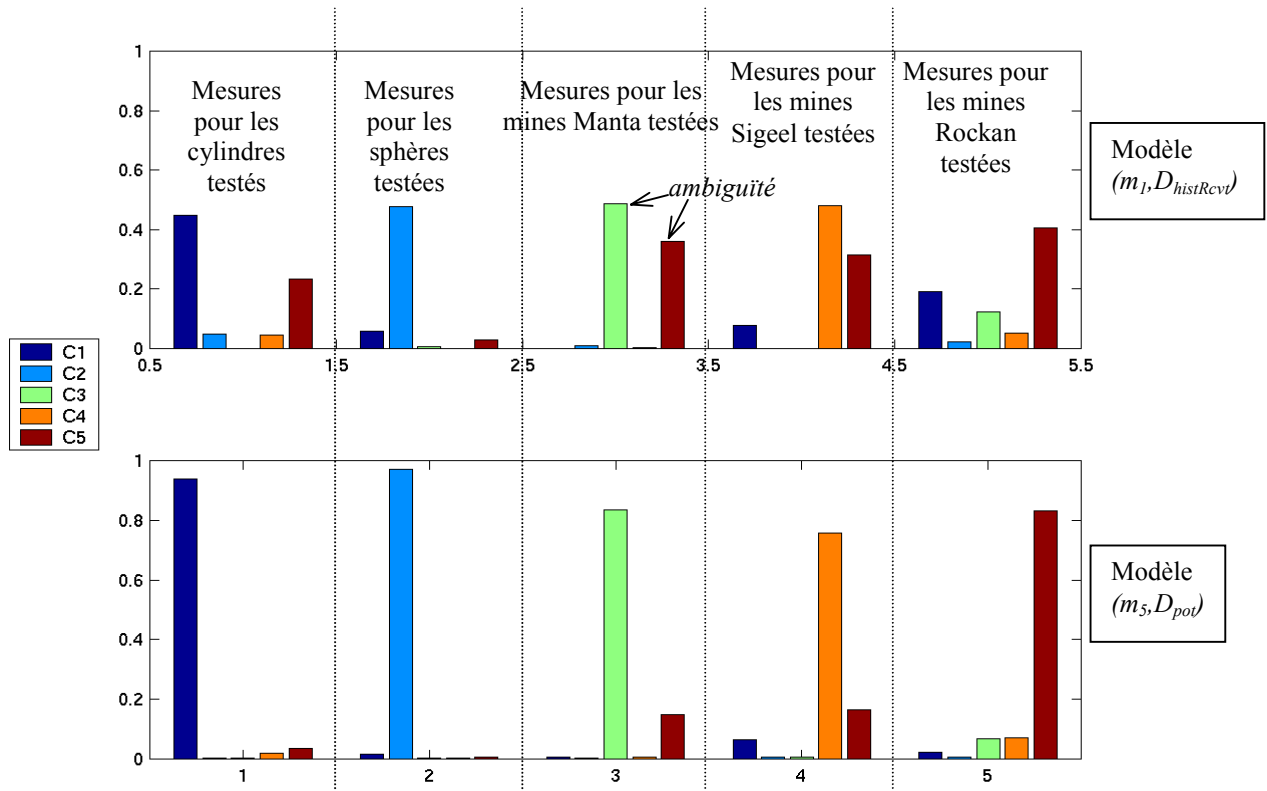


Figure 65 – Mesures moyennes $\bar{M}_k$ pour les attributs hybrides

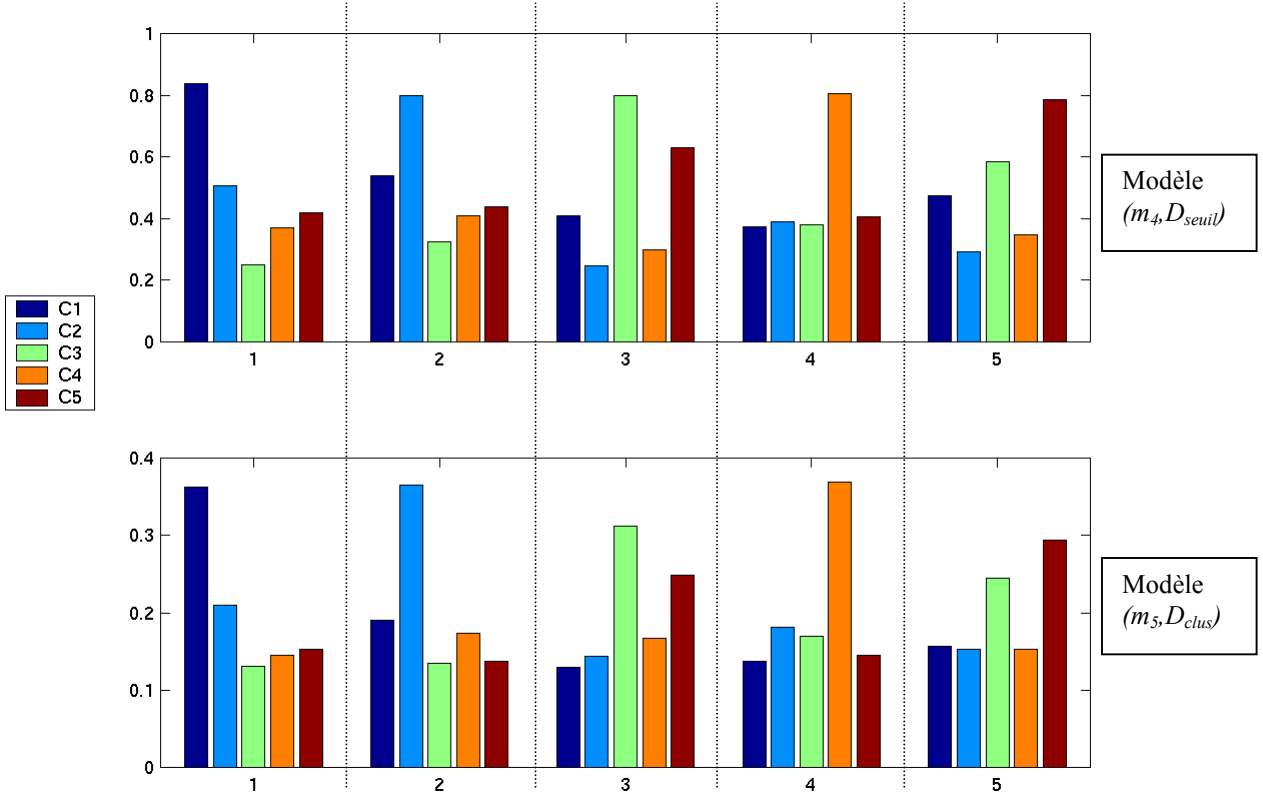


Figure 66– Mesures moyennes $\bar{M}_k$ pour les descripteurs de Fourier

Dans la plupart des modèles de fusion, la mesure relative au degré d'appartenance des individus avec l'élément idéal mine Rockan (classe 5) est non négligeable. Ceci démontre une confusion importante avec cette classe pour laquelle, en effet, la valeur de l'attribut idéal n'est pas caractéristique de la mine du fait de la dispersion des attributs (fort écart-type), conséquence de la géométrie complexe de cette mine.

Ce processus peut donc s'inscrire dans une chaîne de traitement semi-automatique où l'opérateur est amené à n'intervenir qu'en cas de litige entre deux classes. Cette intervention peut en outre être enrichie par l'expérience de sorte que la nature des classes concernées peut être prise en compte (cas de la classe C_5 des mines Rockan).

1.7.2. Opérateur F_{mp}

Le gain obtenu grâce à la pondération est notamment visible en termes de baisse de l'ambiguïté inter-classes.

Rappelons que si $\bar{M}_i$ est la valeur maximale obtenue pour un sous-ensemble d'objets donné, ces objets seront supposés appartenir à la classe i . Nous avons alors fait le calcul suivant pour chaque sous-ensemble d'objets :

Pour chaque opérateur F_j ,

Pour chaque classe C_k ,

$$d\bar{M}_{i,k}^j = \bar{M}_i - \bar{M}_k \text{ qui est forcément positif ou nul}$$

La différence $d\bar{M}_{i,k}^2 - d\bar{M}_{i,k}^1$ permet de visualiser, par sous-ensemble d'objets, ce qu'on gagne en termes de bonne différenciation inter-classes.

Le graphe de la Figure 67 montre la différence moyenne (le gain moyen) sur l'ensemble des deux modèles (m_4, D_{seuil}) et (m_5, D_{clus}) pour trois vues seulement. Nous observons une **amélioration pour les objets cylindriques et les mines Rockan** (premier et dernier sous-ensembles d'objets). Ceci était tout à fait prévisible compte tenu du fait que ces mines ont peu d'axes de symétrie et présentent donc des ombres portées bien différentes selon les points de vue. Grâce à F_{mp} , on tend à minorer l'influence des points de vue non caractéristiques qui portent ambiguïté et perturbent la classification si ces vues font partie des N vues disponibles. Du coup, l'écart entre la meilleure mesure (correspondant à la bonne classification) et les autres mesures se creuse et permet de lever certaines ambiguïtés en apportant de la robustesse.

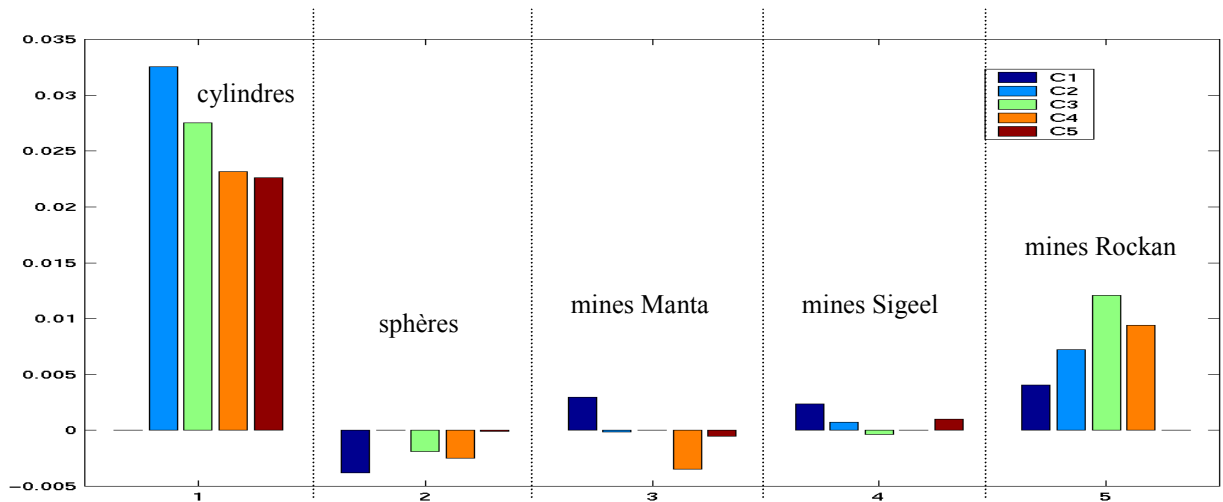


Figure 67 - Illustration du gain de F_{mp} sur F_m

1.8. Conclusion

Une analyse globale de N points de vue d'un même objet a été proposée par fusion des informations acquises image par image. **La modélisation par ensembles flous** est un outil intéressant pour l'aide à la classification : l'idée est de représenter les exemples associés à chaque classe (ou les attributs qui les caractérisent) par des ensembles flous associant à chaque objet **une mesure floue**.

La souplesse des fonctions d'appartenance nous a permis de définir deux groupes de modèles flous adaptés à la fusion multi-vues suivant que l'on réduit ou pas l'ensemble flou à un élément idéal. Un modèle flou se définit comme une fonction décroissante d'une distance calculée entre l'élément courant caractérisé par L attributs caractéristiques et l'ensemble flou. Le calcul est effectué dans un espace à L dimensions sur les vecteurs caractéristiques ou bien dans un espace monodimensionnel, i.e. attribut par attribut. Le choix d'un modèle est conditionné par la répartition statistique des éléments de l'ensemble flou : distribution gaussienne, distribution multi-modale, distribution différente suivant les classes, recouvrements des classes...

Par sa visibilité, la procédure de fusion est une aide à l'opérateur. En effet, la prise de décision est effectuée à partir de mesures dont les valeurs sont proportionnelles à la compatibilité de l'objet ciblé avec chacune des classes supposées. Cette classification floue est souple puisque la prise de décision (attribution d'une classe) se fait de manière éclairée en y associant un **degré de confiance** représenté par la **fusion des degrés d'appartenance** M_k de chacune des

N images à la classe C_k attribuée. Un modèle est d'autant plus fiable que ses mesures permettent d'éviter toute classification litigieuse. En d'autres termes, on préférera un modèle qui associe à un objet donné une mesure maximale (plus forte compatibilité à une classe donnée) bien supérieure aux mesures annexes (compatibilités avec les autres classes). Ainsi, les diagrammes exposés au §I.7.1 permettent à l'opérateur d'avoir une vision globale du résultat de la fusion de N points de vues de l'objet ciblé en disposant **de cinq mesures symbolisées par des barres de hauteur variable**. D'un point de vue opérationnel, ce processus a l'intérêt de pouvoir s'inscrire dans une chaîne de traitement semi-automatique où l'expertise de l'opérateur est alors sollicitée en cas de litige entre deux classes.

II. Classification basée sur l'évolution des attributs au cours d'une séquence d'images

Si l'on excepte le cas des objets à symétrie radiale, l'apparence de l'ombre portée d'un objet diffère généralement d'un point de vue à l'autre. Alors que cette propriété est source d'erreur de classification lorsqu'une seule vue est disponible, elle peut paradoxalement être avantageusement utilisée au cours d'une séquence d'images. Un ensemble d'attributs particulièrement sensibles (à l'inverse des attributs invariants précédemment discutés) à l'évolution de la forme de l'ombre en fonction des points de vue peut ainsi parfaitement discriminer un objet au cours d'une séquence d'images.

II.1. Conditions opérationnelles

Deux types d'attributs vont être associés : des attributs géométriques et des invariants de moments. Les attributs géométriques d'une part reflètent les caractéristiques globales de la forme de l'ombre (cf. §IV.1 du chap. II.2). Les invariants de moments d'autre part, caractérisent la forme en faisant abstraction des transformations géométriques qui l'affectent. En quelque sorte, ils se rapportent à la forme normalisée sous-jacente.

A chaque image I_n de la séquence est ainsi associé un vecteur caractéristique hybride à partir de trois attributs $V_n = (a_1^n \quad a_2^n \quad a_3^n)^T$: le rapport largeur sur hauteur, l'élongation et un invariant de moment II introduit par Flusser et Suk (cf. §IV.2.2 du chap. II.2) [FLUS-93]. Nous rappelons que l'on a précédemment utilisé l'influence du rapport largeur sur hauteur pour pénaliser certains points de vues non caractéristiques dans l'opération de fusion (cf. F_{mp} dans §I.4). Ces attributs sont calculés sur chacune des ombres segmentées de la séquence d'images. La classification de l'objet ciblé est basée sur le rapprochement des valeurs de ces attributs et de valeurs théoriques apprises sur l'ensemble de la base d'apprentissage. Pour rendre possible la comparaison, les données doivent être acquises suivant un scénario défini. La contrainte est en effet de rapprocher les données à classer selon une trajectoire donnée aux données correspondantes de la base d'apprentissage. L'invariance de cette base vis-à-vis de l'altitude et de la distance à la cible est assurée par normalisation de l'image. En revanche, les positions relatives du sonar et de l'objet qui définissent l'angle sous lequel le sonar voit l'objet constituent une donnée essentielle pour assurer la correspondance des points de vues.

Pour notre étude, une trajectoire circulaire du sonar est utilisée afin de disposer d'une séquence de points de vue distincts par pas de 10 degrés selon une trajectoire circulaire.

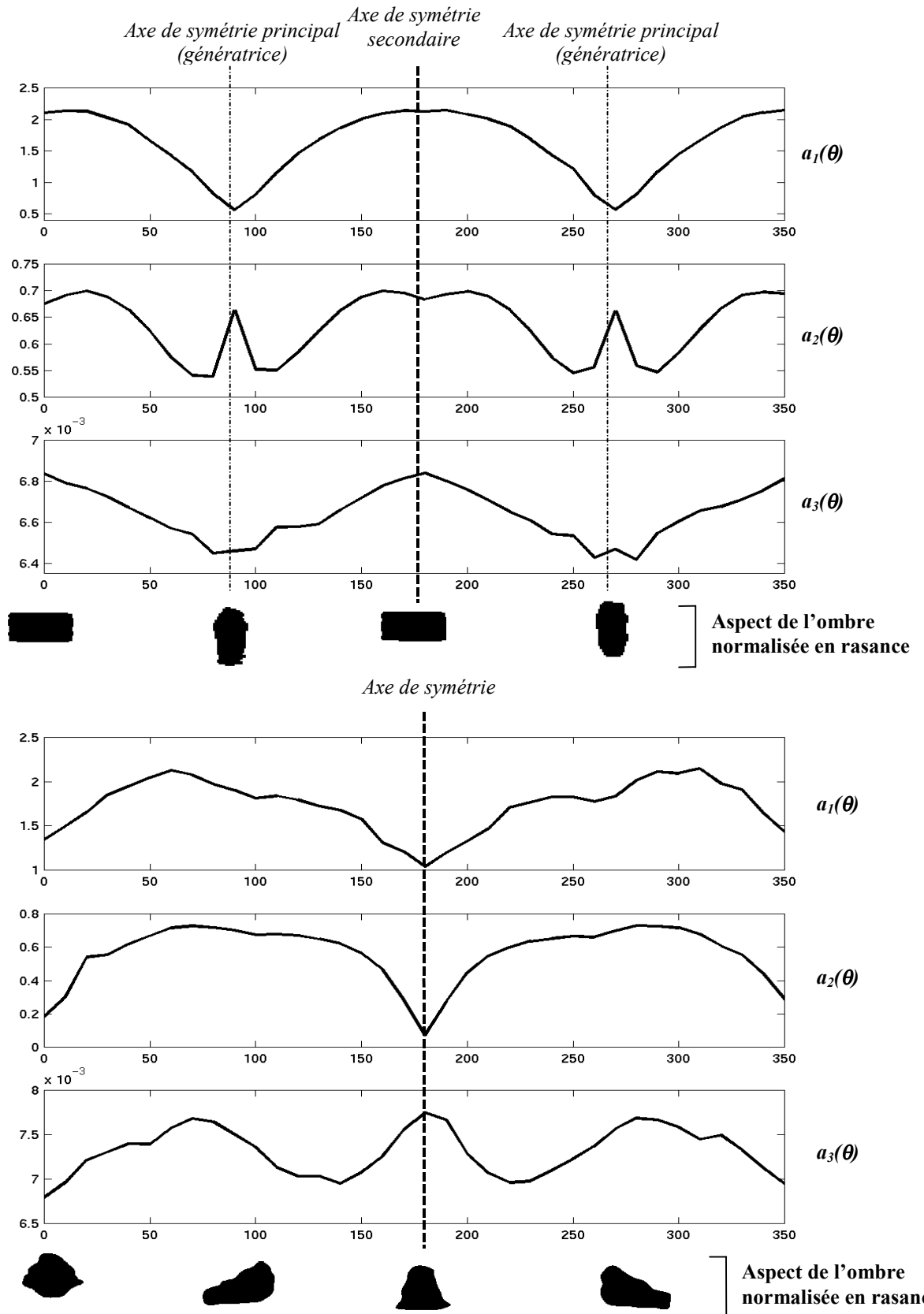


Figure 68 – Lien entre propriétés géométriques et allure de l'évolution des attributs (mine cylindrique et Rockan)

II.2. Lien entre symétrie de l'objet et valeurs des attributs

Suivant les propriétés de symétrie de l'objet ciblé, on note une répétition d'un motif, ou son symétrique, plus ou moins courte de l'évolution des attributs en fonction de l'angle de prise de vue (cf. Figure 68). En l'occurrence, plus l'objet possède de symétries par rotation, plus cette répétition est fréquente. En notant N_{ax} le nombre d'axes de symétrie de l'objet posé sur le fond et P_θ la durée angulaire associée au motif, on a alors la relation suivante : $P_\theta = 2\pi/2N_{ax}$.

Différents cas de figure se distinguent :

- objet à symétrie radiale (sphère, mine Manta, mine Sigeel)

Dans ce cas, les attributs sont constants quel que soit le point de vue. Le nombre d'axes de symétrie N_{ax} est infini et la durée P_θ nulle.

- objet à deux axes de symétrie (cylindre) $\Rightarrow P_\theta = 2\pi/2N_{ax} = \pi/2$.

Dans ce cas, un quota minimal de 9 points de vue espacés de 10 degrés est nécessaire pour caractériser la mine.

- objet à un seul axe de symétrie (mine Rockan) $\Rightarrow P_\theta = 2\pi/2N_{ax} = \pi$.

Cette fois, la mine n'est totalement caractérisée que si on dispose de 18 prises de vue espacées de 10 degrés.

Pour prendre en compte ces trois cas de figure, la base d'apprentissage doit donc comporter au minimum 18 prises consécutives de vue par objet.

Les tracés de la Figure 68 représentent l'évolution moyenne (sur 4 séquences) des attributs calculés pour 36 prises de vue lors d'une trajectoire circulaire autour de l'objet de 360 degrés avec une prise tous les 10 degrés.

II.3. Algorithme de classification multi-vues

D'après les observations ci-dessus, la classification revient à comparer les tracés obtenus pour chacun des attributs $a_i(k)$ avec les tracés théoriques $a_i^j(k)$ de chacune des classes pour $j = 1, \dots, K$. Ces tracés théoriques sont appris à partir de 15 séquences de la base d'objets simulés (cf. annexe G) et donnés Figure 69 : on vérifie que les valeurs des attributs sont indépendantes du point de vue pour les objets à symétrie radiale.

Les **opérateurs de corrélation** sont indiqués pour mesurer la similarité de deux courbes x et y de longueurs respectives M et N . Ces opérateurs sont préférés au calcul de la distance euclidienne pour mesurer la **ressemblance** entre les signaux. En effet, si ces deux critères permettent de comparer les signaux de manière quantitative, seule la corrélation prend en compte l'allure des courbes représentatives de l'évolution des attributs au cours de la séquence et constitue donc un critère qualitatif de comparaison. De plus, elle permet de régler les problèmes de décalage.

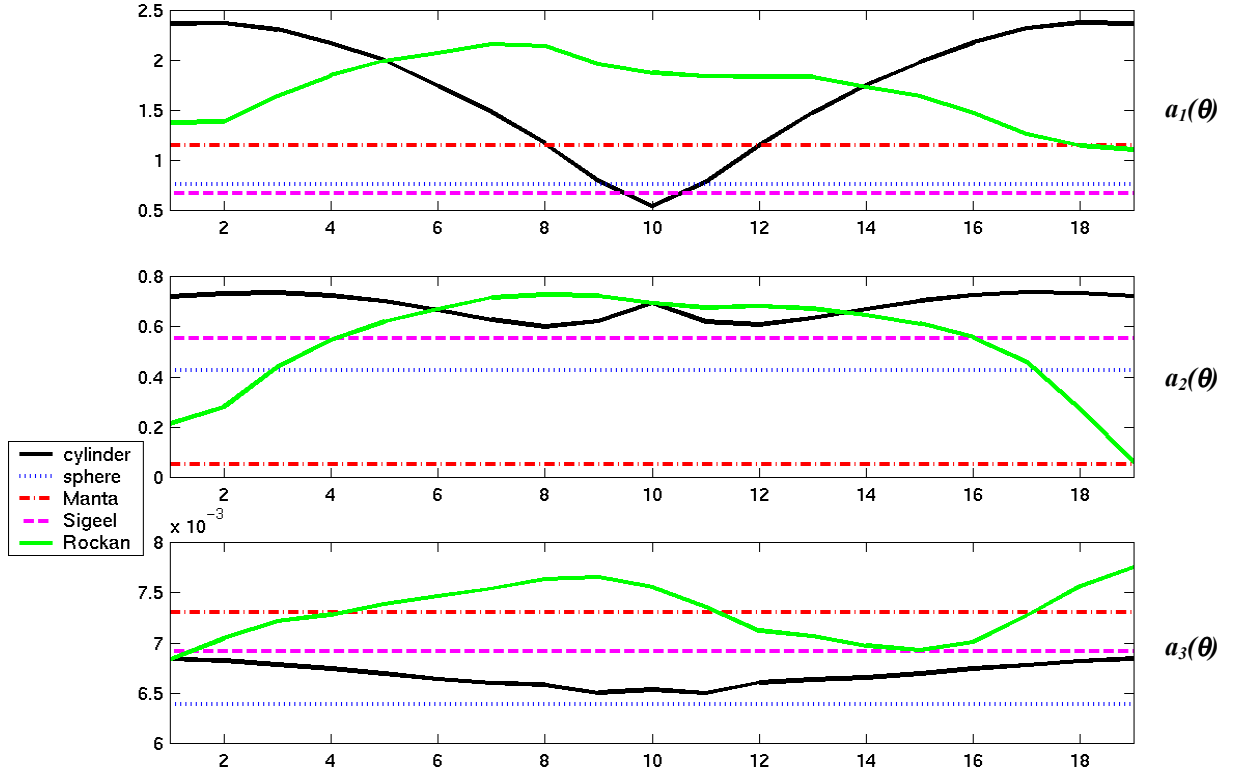


Figure 69 – Courbes théoriques des valeurs d'attributs apprises par classe pour une séquence de prises de vue le long d'une trajectoire semi-circulaire

La fonction de corrélation φ d'une part permet de connaître un éventuel décalage entre les courbes et est définie par :

$$\varphi_{xy}(k) = \frac{2}{M+N} \sum_{l=0}^{M-1} x(l)y(l+k)$$

Ce vecteur de longueur $M+N-1$ est maximal en $k_{\max}$, valeur désignant le décalage des courbes nécessaire pour obtenir la similarité des courbes la plus grande.

Le coefficient de corrélation ρ d'autre part mesure la similarité des courbes en normalisant la fonction précédente :

$$\rho_{xy}(k) = \frac{\varphi_{xy}(k)}{\sqrt{\varphi_{xx}(0)\varphi_{yy}(0)}} \leq 1$$

Ce coefficient est maximal et égal à 1 lorsque les courbes sont identiques (processus corrélés). Evidemment, si l'une des courbes est constante, ce coefficient ne fournit aucune information exploitable. Par exemple,

$$x(l) = x_0, \forall l \quad \text{entraîne} \quad \rho_{xy} = \rho_y \text{ indépendamment de } x$$

Dans ce cas particulier, une comparaison plus simple consiste à mesurer la différence entre x_0 et la valeur moyenne $\bar{y} = \frac{1}{N} \sum_{l=1}^N y(l)$. Ainsi, avant l'ultime module de classification, on distingue:

- les objets à symétrie radiale pour lesquels la valeur de l'attribut ne dépend pas du point de vue $\Rightarrow$ le coefficient de corrélation ne permet pas de prendre une décision (ambiguïté entre les classes d'objets à symétrie radiale) $\Rightarrow$ décision à partir des valeurs moyennes
- les objets complexes $\Rightarrow$ décision à partir des coefficients de corrélation.

Le diagramme de la Figure 70 donne les principales étapes de l'algorithme de classification implémenté. Dans la suite, Sf_i (resp. Sf_i^j) désigne les séquences expérimentales (resp. théoriques) relatives au $i^{\text{ème}}$ attribut a_i .

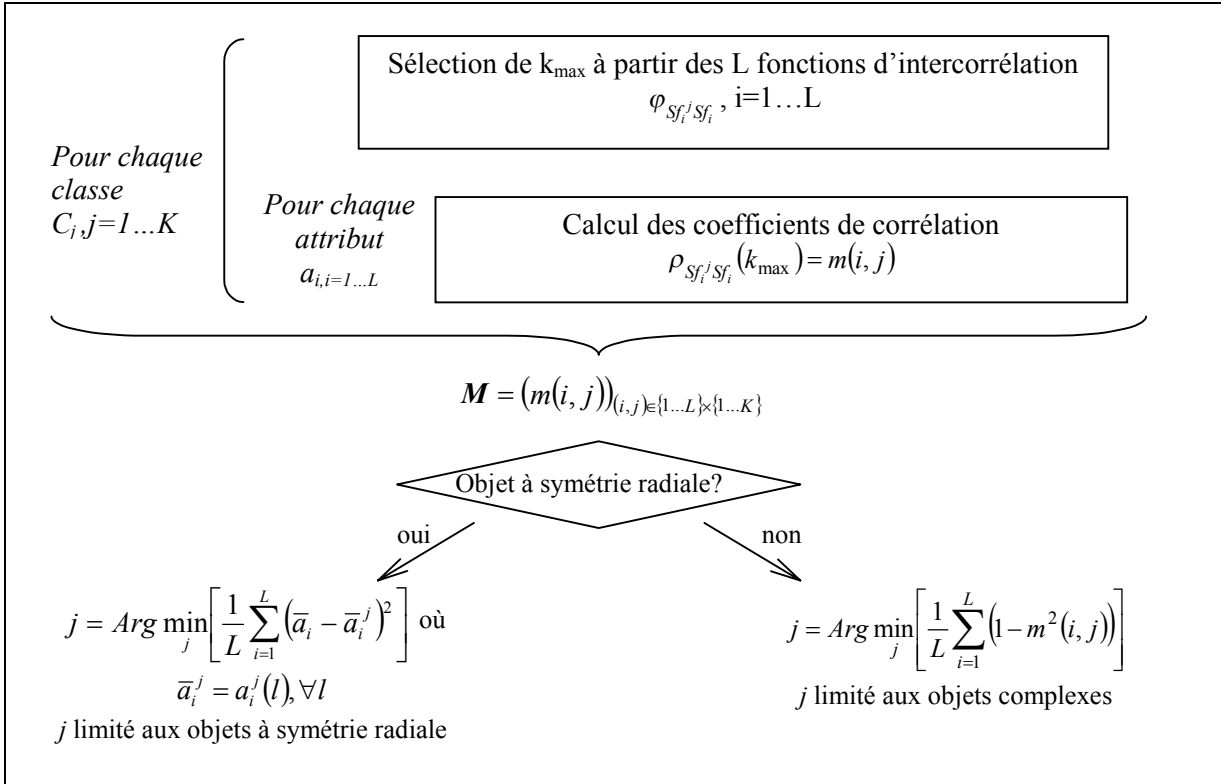


Figure 70 - Algorithme de classification multi-vues par caractérisation de l'évolution séquentielle de la valeur des attributs

Etape 1. Comparaison des séquences Sf_i et Sf_i^j au moyen des fonctions d'intercorrélation

$$\phi_{Sf_i^j Sf_i}.$$

Sélection de $k_{\max}$ tel que :

$$k_{\max} = \text{Arg max}_k \left[\sum_{i=1}^L \phi_{Sf_i^j Sf_i}(k) / L \right]$$

Etape 2. Calcul des coefficients de corrélation $\rho_{Sf_i^j Sf_i}(k_{\max}) = m(i, j)$.

La matrice obtenue $M = (m(i, j))_{(i,j) \in \{1...L\} \times \{1...K\}}$ de dimension $L \times K$ contient l'ensemble des coefficients de corrélation maximaux entre la séquence expérimentale et les séquences théoriques pour chacun des attributs.

Etape 3. Test sur la symétrie de l'objet ciblé

La classification (étape 4) n'est possible que si les grandeurs manipulées sont discriminantes pour l'objet ciblé. Or, comme nous l'avons déjà noté, dans le cas particulier d'un objet à symétrie radiale, le coefficient de corrélation n'est pas indiqué.

Néanmoins, cette propriété nous permet de décider si un objet est plutôt à symétrie radiale ou au contraire plus complexe. En effet, le résultat obtenu est identique pour l'ensemble des séquences théoriques relatives à ces types d'objets à savoir les classes « sphère », « mine Manta » et « mine Sigeel ». On peut alors effectuer le test suivant :

Par attribut i ,

on calcule le coefficient particulier ρ_{Sf_i}

on recherche la classe j pour laquelle le coefficient $m(i, j)$ est le plus grand soit $\max_j m(i, j)$,

si ce coefficient est égal ρ_{Sf_i} , cela signifie qu'il est indépendant de la classe

(classe d'objets à symétrie radiale) (il y aura en outre ambiguïté entre ces trois classes)

Si c'est le cas pour au moins la moitié des attributs, l'objet est vraisemblablement à symétrie radiale.

Etape 4. Décision à partir de l'ensemble des L attributs

☞ Premier cas : l'objet est supposé être à symétrie radiale, i.e. sphère, mine Manta ou Sigeel
La comparaison est basée sur les valeurs moyennes des attributs le long de la séquence et les valeurs théoriques (constantes puisqu'il s'agit des classes d'objets à symétrie radiale)

$$j = \text{Arg min}_j \left[\frac{1}{L} \sum_{i=1}^L (\bar{a}_i - \bar{a}_i^j)^2 \right] \text{ où } \bar{a}_i^j = a_i^j(l), \forall l$$

☞ Second cas : l'objet est complexe, i.e. cylindre ou mine Rockan

Les coefficients de corrélation sont les valeurs discriminantes utilisées,

$$j = \text{Arg min}_j \left[\frac{1}{L} \sum_{i=1}^L (1 - m^2(i, j)) \right]$$

II.4. Résultats expérimentaux

L'algorithme a été testé sur neuf mines différentes comptabilisant 30 séquences d'images simulées chacune : trois cylindres, trois sphères, les mines Manta, Sigeel et Rockan (cf. §IV et V de la partie I). Selon le processus de classification supervisée, la classification est effectuée sur la moitié des séquences, l'autre moitié servant à déterminer les séquences théoriques de la Figure 69.

Les séquences entières de 19 points de vues acquis par pas de 10 degrés le long d'une trajectoire circulaire du sonar donnent un taux de bonne classification moyen de 99.4%. Si l'on dispose de la moitié seulement de la séquence (9 points de vue consécutifs dont le premier est choisi aléatoirement dans la séquence), le taux descend à 86.2%. Les matrices de confusion correspondantes sont données dans le Tableau 15.

Tout en conservant le même échantillonnage angulaire, on augmente le secteur angulaire de $N=3$ à 19 vues. En somme, lorsque $N < 19$, on sélectionne N points de vue consécutifs parmi les 19 (le début de la séquence tronquée étant obtenu par tirage aléatoire). Tracée sur la Figure 71, l'évolution moyenne toutes classes confondues en fonction du nombre N de vues est supérieure à 80% à partir de $N=8$ vues et à 90% à partir de $N=10$ vues. Néanmoins, le détail des performances sur la Figure 72 et l'écart-type associé pour les 15 séquences

montrent que la classification reste discutable tant que N est inférieur à 14 corrélée à un écart-type important jusque 16% pour la mine Manta (cf. Figure 73).

Tableau 15 - Matrices de confusion pour $M=N=19$, i.e. un demi-cercle autour de la mine (resp. $N=9$, i.e. moins d'un quart de cercle autour de la mine)

	C1		C2		C3		C4		C5	
N	19	9	19	9	19	9	19	9	19	9
Cylindres	100	68	0	31	0	0.5	0	0	0	0.5
Sphères	0	0	100	100	0	0	0	0	0	0
Mines Manta	2	2	0	0	98	88	0	0	0	12
Mines Sigeeel	1	0	0	0	0	0	99	100	0	0
Mines Rockan	0	24	0	0	0	1	0	0	100	75

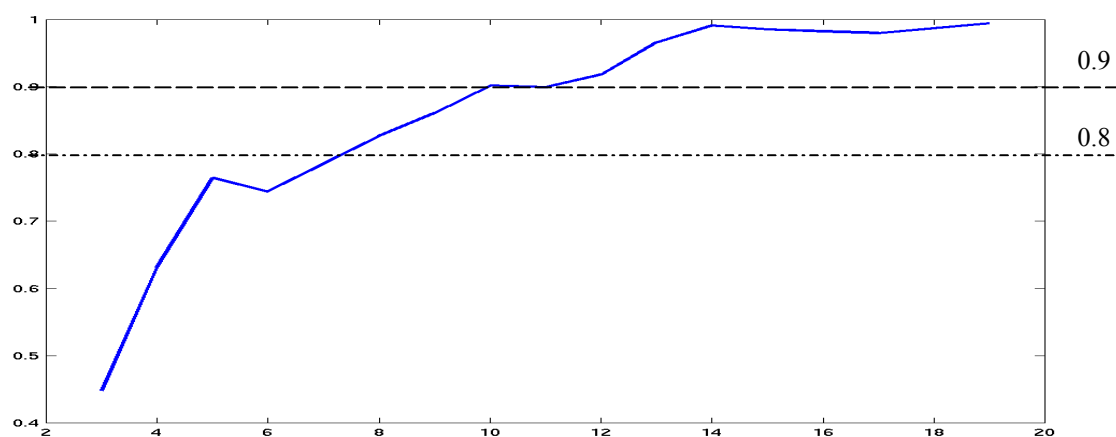


Figure 71 - Taux moyen de bonne classification pour un nombre de vues N de 3 à 19

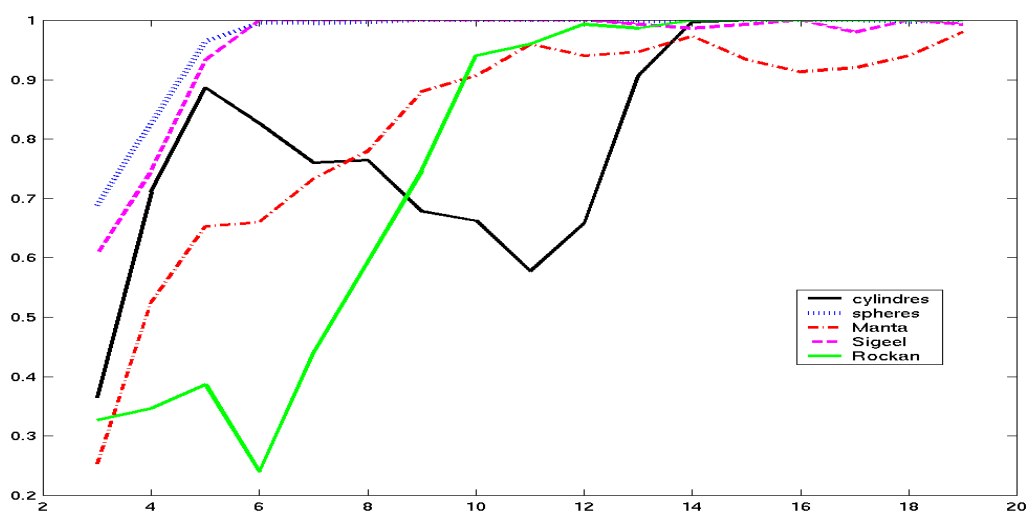


Figure 72 - Taux moyen de bonne classification par classe sur l'ensemble des séquences pour un nombre de vues N de 3 à 19

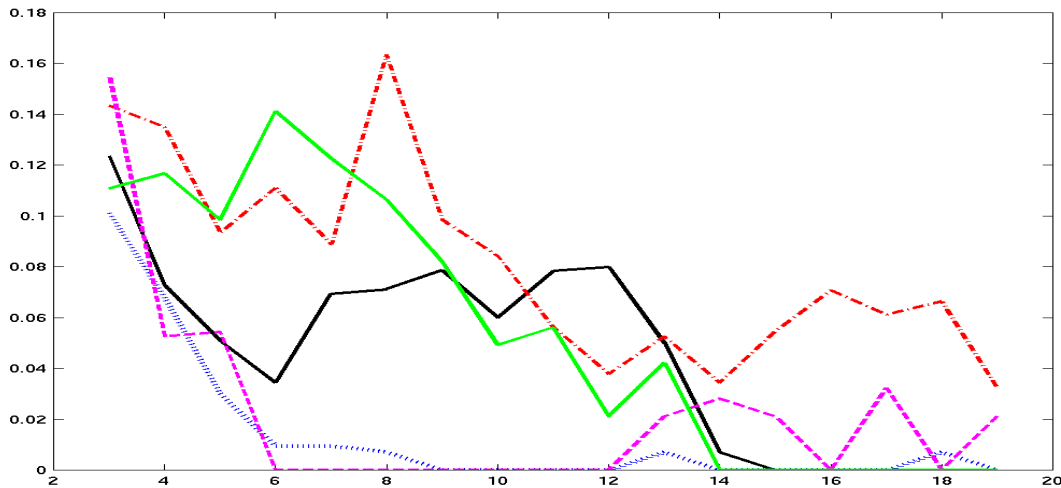


Figure 73 - Ecart-type associé au taux moyen de bonne classification sur l'ensemble des séquences pour un nombre de vues N de 3 à 19

II.5. Conclusion

Concrètement, l'algorithme proposé étudie l'évolution séquentielle des attributs via des opérations de corrélation. Nous caractérisons ainsi la ressemblance entre courbes théoriques relatives à chacune des classes et courbes expérimentales relatives à l'objet ciblé.

La caractérisation globale des objets sur toute une séquence d'images s'est avérée probante pour les objets à géométrie complexe. En effet, ces objets dont l'ombre portée est variable suivant l'angle de vue peuvent être avantageusement décrits par l'analyse de l'évolution progressive de la forme de l'ombre pour des points de vue acquis par échantillonnage angulaire déterminé le long d'une trajectoire précise du sonar. Cette caractérisation est plus complète puisqu'elle est effectuée en trois dimensions, i.e. le plan 2D de l'image en fonction de l'angle de vue. Ces informations 3D sont aujourd'hui facilement disponibles grâce à l'utilisation de véhicules propulsés à immersion variable bénéficiant d'une grande liberté de mouvement.

Chapitre IV.2 :

Caractérisation floue de la superposition des vues

L'idée développée dans ce chapitre est de résumer sur une seule image les données sonar acquises suivant différents points de vue, brutes ou traitées (segmentation), et d'extraire un certain nombre d'attributs caractérisant la zone d'intérêt. Contrairement à l'approche mono-image pour laquelle l'ombre est constituée de pixels de même poids, notre image « multi-vues » présente une ombre ou zone d'intérêt constituée de pixels de niveaux de gris inégaux fonction de leurs localisations successives (zone d'ombre ou zone de réverbération) au cours de la séquence. La géométrie floue vise à caractériser des régions vues comme des sous-ensembles flous de l'image. Cette approche est née de l'extension d'algorithmes standards de traitement d'images à la logique floue (géométrie floue, mesures floues de l'information de l'image, morphologie mathématique floue,...) [TIZH-98]. Dans ce formalisme, la pertinence de chaque pixel est proportionnelle à son niveau de gris vu comme un degré d'appartenance à ladite région. L'application de cette géométrie semble adaptée à la caractérisation de la région 'ombre' et notamment la dispersion des pixels qui est d'autant plus importante que l'ombre portée de l'objet diffère d'un point de vue à l'autre. En plus des grandeurs rencontrées dans la littérature, nous avons défini la notion de moment flou comme une extension des moments classiques au cas des images en niveaux de gris.

Un tel procédé est bien entendu tributaire d'un certain nombre de précautions. En particulier, les images doivent être acquises successivement par le sonar, à distance cible-sonar et altitude approximativement constantes.

I. Notion de région floue [ROSE-84]

Jusqu'ici, l'extraction d'attributs suivait une étape préliminaire de segmentation. Dans l'image binaire obtenue, les pixels étiquetés 'ombre' avaient le même niveau de gris. La géométrie floue est duale de la caractérisation géométrique effectuée à partir des images binaires (cf. §IV.1 du chap. II.2) mais concerne des régions formées de pixels ayant des niveaux de gris différents. Ces niveaux se rapportent à un degré d'appartenance à la région floue 'ombre'. Rosenfeld a généralisé plusieurs propriétés géométriques classiques en les étendant au cas flou [ROSE-84].

I.1. Définition

Un sous-ensemble flou d'un ensemble S de sites est une réalisation μ de S dans $[0, 1]$. Pour tout $p \in S$, $\mu(p)$ est appelé degré d'appartenance de p dans μ . Un sous-ensemble non flou ou net est un cas particulier des sous-ensembles flous pour lequel μ est dans $\{0, 1\}$.

Pour notre application, S est un plan ou une image. p désigne un pixel de coordonnées (i, j) . La région floue μ est bornée de telle sorte que $\mu = 0$ en dehors de la région.

Des grandeurs géométriques peuvent être calculées sur μ pour décrire la région dans laquelle la contribution de chacun des pixels est liée à son degré d'appartenance.

1.2. Attributs intrinsèques flous

Parmi les grandeurs d'abord introduites par Rosenfeld puis étendues par Pal et Ghosh, nous avons retenu (définitions dans le cas discret avec j indice pour les colonnes et i pour les lignes) [ROSE-84] [PAL-92] :

- **Aire a**

$$a(\mu) = \sum_i \sum_j \mu(i, j)$$

- **Largeur L**

$$l(\mu) = \sum_j \max_i \{\mu(i, j)\}$$

- **Longueur l**

$$L(\mu) = \sum_i \max_j \{\mu(i, j)\}$$

- **Largeur maximale l_m**

$$l_m(\mu) = \max_i \sum_j \mu(i, j)$$

- **Longueur maximale L_m**

$$L_m(\mu) = \max_j \sum_i \mu(i, j)$$

De ces quatre dernières grandeurs, on tire une information sur l'orientation de la forme :

- **Indice d'orientation horizontale IOH**

$$IOH(\mu) = \frac{l_m(\mu)}{l(\mu)}$$

- **Indice d'orientation verticale IOV**

$$IOV(\mu) = \frac{L_m(\mu)}{L(\mu)}$$

- **Indice de couverture de l'aire ICA**

$$ICA(\mu) = \frac{a(\mu)}{L_m(\mu) \times l_m(\mu)}$$

Cette grandeur est égale à 1 pour un rectangle dont les côtés sont parallèles aux axes, et de $\pi/4$ pour un cercle.

- **Densité d**

$$d(\mu) = \frac{a(\mu)}{N} \text{ où } N \text{ est le nombre de pixels de degré d'appartenance non nul à } \mu$$

Suivant le même principe, nous ajoutons des grandeurs dérivées de la théorie des moments. Pour ce faire, nous définissons un moment flou sur la région floue μ par :

$$\tilde{m}_{pq} = \sum_M \sum_N i^p j^q \mu(i, j)$$

d'où découlent les moments centrés et réduits flous $\tilde{\mu}_{pq}$ et $\tilde{\mu}'_{pq}$:

$$\tilde{\mu}_{pq} = \sum_M \sum_N (i - i_g)^p (j - j_g)^q \mu(i, j) \quad \text{et} \quad \tilde{\mu}'_{pq} = \frac{\tilde{\mu}_{pq}}{\tilde{\mu}_{00}^{\left[\frac{(p+q)}{2} + 1\right]}}$$

avec les coordonnées du centre de gravité de l'ombre (i_g, j_g) tel que $i_g = \frac{\tilde{m}_{10}}{\tilde{m}_{00}}$ et $j_g = \frac{\tilde{m}_{01}}{\tilde{m}_{00}}$.

Par extension au cas flou, nous introduisons les grandeurs suivantes :

- **Etendue Et**

$$Et(\mu) = \frac{\tilde{\mu}_{20} + \tilde{\mu}_{02}}{\tilde{\mu}_{00}^2}$$

- **Elongation El**

$$El_1(\mu) = \frac{\sqrt{4\tilde{\mu}_{11}^2 + (\tilde{\mu}_{20} - \tilde{\mu}_{02})^2}}{\tilde{\mu}_{20} + \tilde{\mu}_{02}} \quad \text{ou} \quad El_2(\mu) = \frac{\alpha}{\beta}$$

avec

$$\begin{cases} \alpha = \sqrt{\frac{2[\tilde{\mu}_{20} + \tilde{\mu}_{02} + \sqrt{(\tilde{\mu}_{20} - \tilde{\mu}_{02})^2 + 4\tilde{\mu}_{11}^2}]}{\mu_{00}}} \\ \beta = \sqrt{\frac{2[\tilde{\mu}_{20} + \tilde{\mu}_{02} - \sqrt{(\tilde{\mu}_{20} - \tilde{\mu}_{02})^2 + 4\tilde{\mu}_{11}^2}]}{\tilde{\mu}_{00}}} \end{cases}$$

- **Concentration elliptique $Ellipse$**

$$Ellipse(\mu) = \frac{\tilde{\mu}_{00}}{\pi\alpha\beta}$$

II. Obtention de l'image multi-vues

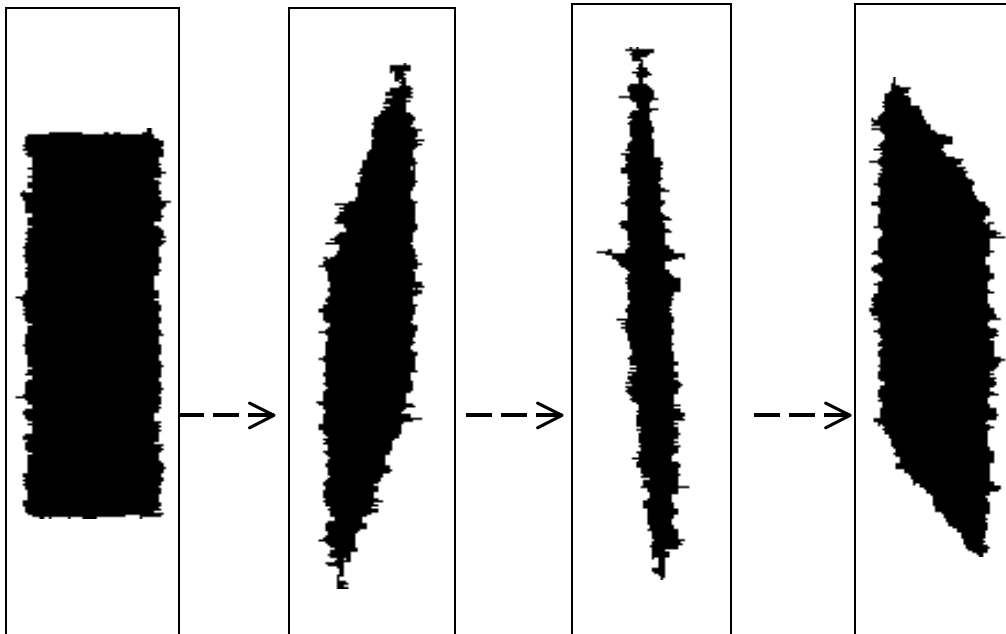
Nous travaillons avec des séquences de 19 points de vues acquis par pas de 10 degrés le long d'une trajectoire circulaire du sonar à altitude et distance sonar-objet constantes.

II.1. A partir des images segmentées

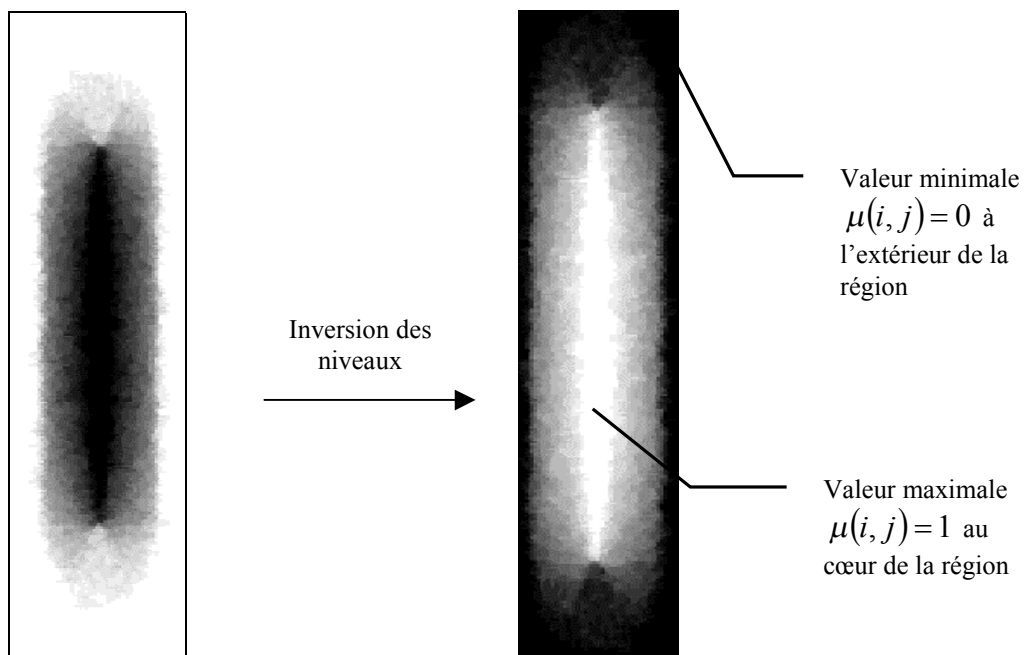
Une vision globale des différentes formes de l'ombre en fonction des points de vue est obtenue par **superposition des ombres segmentées** en faisant coïncider leur centre de gravité. Dans cette nouvelle image I_s , la valeur du niveau de gris est alors directement reliée à l'occurrence avec laquelle le pixel apparaît dans l'ombre de l'objet. Etant entendu que le nombre de prises de vue est largement inférieur au nombre de niveaux de gris usité, i.e. 256, le niveau de gris maximal de la nouvelle image sera égal au nombre de vues.

Voici ci-après illustrée la procédure d'obtention de l'image multi-vues I_s :

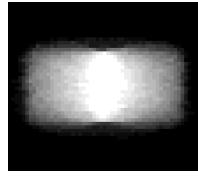
Etape 1. Segmentation des images (on n'illustre que 4 vues intermédiaires)



Etape 2. Superposition des $N=19$ images binaires et inversion des niveaux de gris



Etape 3. Normalisation de l'image



La Figure 74 donne un exemple pour chacune des classes.

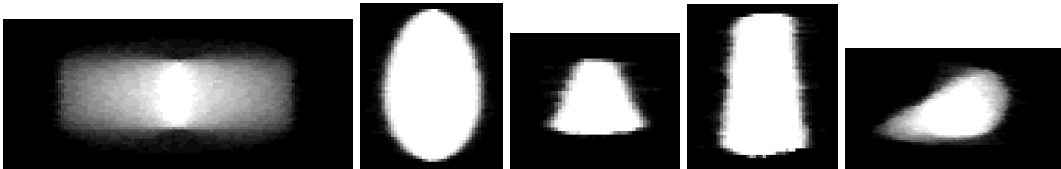


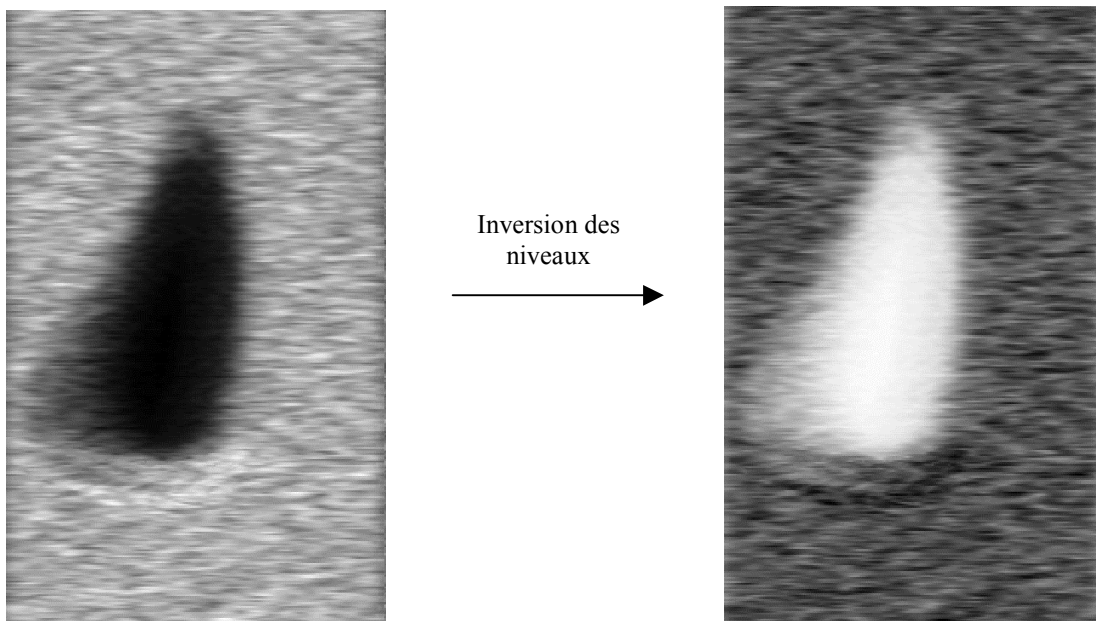
Figure 74 – Exemples d'images I_s obtenues à partir de 19 prises de vues par pas angulaire de 10 degrés : cylindre, sphère, mines Manta, Sigeel et Rockan

II.2. A partir des images brutes

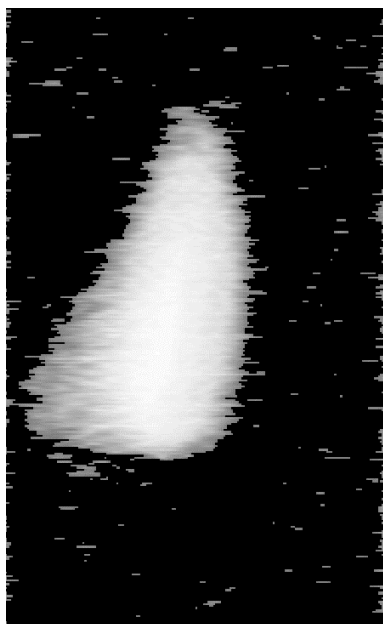
Le calcul d'attributs géométriques peut également s'effectuer sans traitement préalable (segmentation) mais par **addition des données brutes** (moyennant leur cohérence par recalage). Dans cette nouvelle image I_b , le niveau de gris du pixel à la zone ombre fait alors inversement référence à la difficulté à l'affecter à la zone 'ombre'. Nous repoussons ainsi l'étape de décision au niveau pixel, i.e. la segmentation proprement dite. L'accumulation des images permet de réduire les ambiguïtés en gérant la dispersion des pixels de frontières autour d'un noyau dur constitué des pixels de plus hauts niveaux de gris.

Voici ci-après illustrée la procédure d'obtention de l'image multi-vues I_b :

Etape 1. Superposition des $N=19$ images brutes et inversion des niveaux de gris après une quantification de l'image I_b sur 256 niveaux de gris

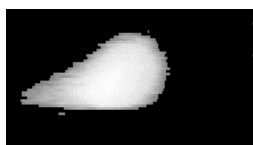


Etape 2. Seuillage dépendant du niveau moyen de réverbération (calculé sur une portion de la zone de fond) tel que l'on affecte un niveau nul aux pixels de niveau inférieur à $\max(128, \text{nivRevMoy} + 10)$.



Le but est de rendre les calculs indépendants de la proportion de pixels appartenant exclusivement à la zone de réverbération de fond.

Etape 3. Normalisation de l'image



Rappel : Cette normalisation revient à regrouper les pixels et à affecter au nouveau pixel un niveau de gris égal à la valeur médiane des niveaux de gris regroupés. Ceci a pour effet de supprimer les fausses alarmes de la zone de fond que l'on apercevait avant normalisation.

La Figure 75 donne un exemple pour chacune des classes.

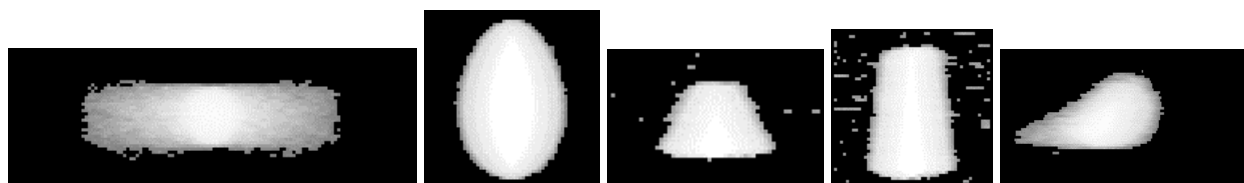


Figure 75 – Exemples d'images I_b obtenues à partir de 19 prises de vues par pas angulaire de 10 degrés : cylindre, sphère, mines Manta, Sigeel et Rockan

Pour les deux types d'images formées, la géométrie radiale des mines sphériques, Manta et Sigeel mène à une concentration des pixels d'ombre qui s'accumulent aux mêmes sites de l'image au fur et à mesure des points de vue. A l'inverse, la forte dispersion des pixels d'ombre pour le cylindre et la mine Rockan entraîne une zone frontière floue beaucoup plus importante.

II.3. Fuzzification de l'image

Le calcul des attributs flous nécessite d'avoir une valeur du degré d'appartenance associé aux pixels comprise entre zéro et un. Une *fuzzification* de l'image est donc préalablement réalisée en divisant chacun des niveaux de gris de l'image 'multi-vues' I_b ou I_s par le niveau de gris maximal.

III. Résultats expérimentaux

Mis à part l'indice de couverture de l'aire ICA , ces mesures métriques comme dans le cas des images binaires (cf. §IV.1 du chap. II.2), distinguent des groupes d'objets appartenant à la même classe mais de tailles différentes. Il en résulte une partition de la classe en plusieurs sous-groupes (multi-modalité). Cette difficulté est prise en compte par un classifieur k-ppv.

Le vecteur caractéristique a volontairement été choisi long de manière à ne pas effectuer de choix a priori sur les attributs mais à mettre en évidence leur pertinence par une ACP et l'examen du cercle des corrélations (cf. §II.2 du chap. III.1). Le vecteur caractéristique comporte donc 23 composantes $a_{i,i=1,...,23}$ parmi lesquelles coexistent 11 attributs intrinsèques flous et 12 moments flous soient $d, ICA, Et, El_2, Ellipse, IOH, IOV, l_m, l, L_m, L, \tilde{\mu}_{02}, \tilde{\mu}_{20}, \tilde{\mu}_{11}, \tilde{\mu}_{03}, \tilde{\mu}_{30}, \tilde{\mu}_{12}, \tilde{\mu}_{21}, \tilde{\mu}_{04}, \tilde{\mu}_{40}, \tilde{\mu}_{22}, \tilde{\mu}_{31}, \tilde{\mu}_{13}$.

La Figure 76 illustre les ACP effectuées à partir des attributs calculés sur les deux types d'images I_s et I_b et pour neuf mines différentes comptabilisant 30 séquences d'images simulées chacune : trois cylindres, trois sphères, les mines Manta, Sigeel et Rockan (cf. §IV et V de la partie I et annexe G).

Une rapide comparaison des deux ACP et de la part d'inertie cumulée des deux premiers axes montre que la classification à partir des images I_s sera plus facile qu'à partir des images traitées sans segmentation préalable I_b . Les groupements ou nuages de points sont en effet beaucoup plus compacts dans le premier cas avec une part d'inertie de 0.833 contre 0.718 dans le second cas. L'apparition de sous-groupes pour les classes 'cylindres' et 'sphères' était prévue du fait des différentes tailles possibles pour ces objets et les attributs choisis.

Le cercle des corrélations d'autre part permet d'écarter d'emblée certains attributs non pertinents du fait de leur éloignement à la circonférence, i.e. une faible distance :

$$\sqrt{r^2(f_1, a_i) + r^2(f_2, a_i)}$$

Ainsi, pour chacune des expériences réalisées, on peut s'affranchir d'environ la moitié des attributs et ne conserver que les attributs les plus proches de la circonférence.

Le Tableau 16 donne enfin les taux de classification moyens (pour trois mélanges aléatoires différents de la base) pour les deux séries d'images et trois classifieurs différents. Les mêmes tests sont effectués à partir des deux seules composantes principales f_1 et f_2 obtenus par ACP et donnés dans le Tableau 17. Nous notons que la réduction de l'espace de 23 à seulement 2 dimensions par ACP permet de maintenir des taux de classification honorables dans les deux cas.

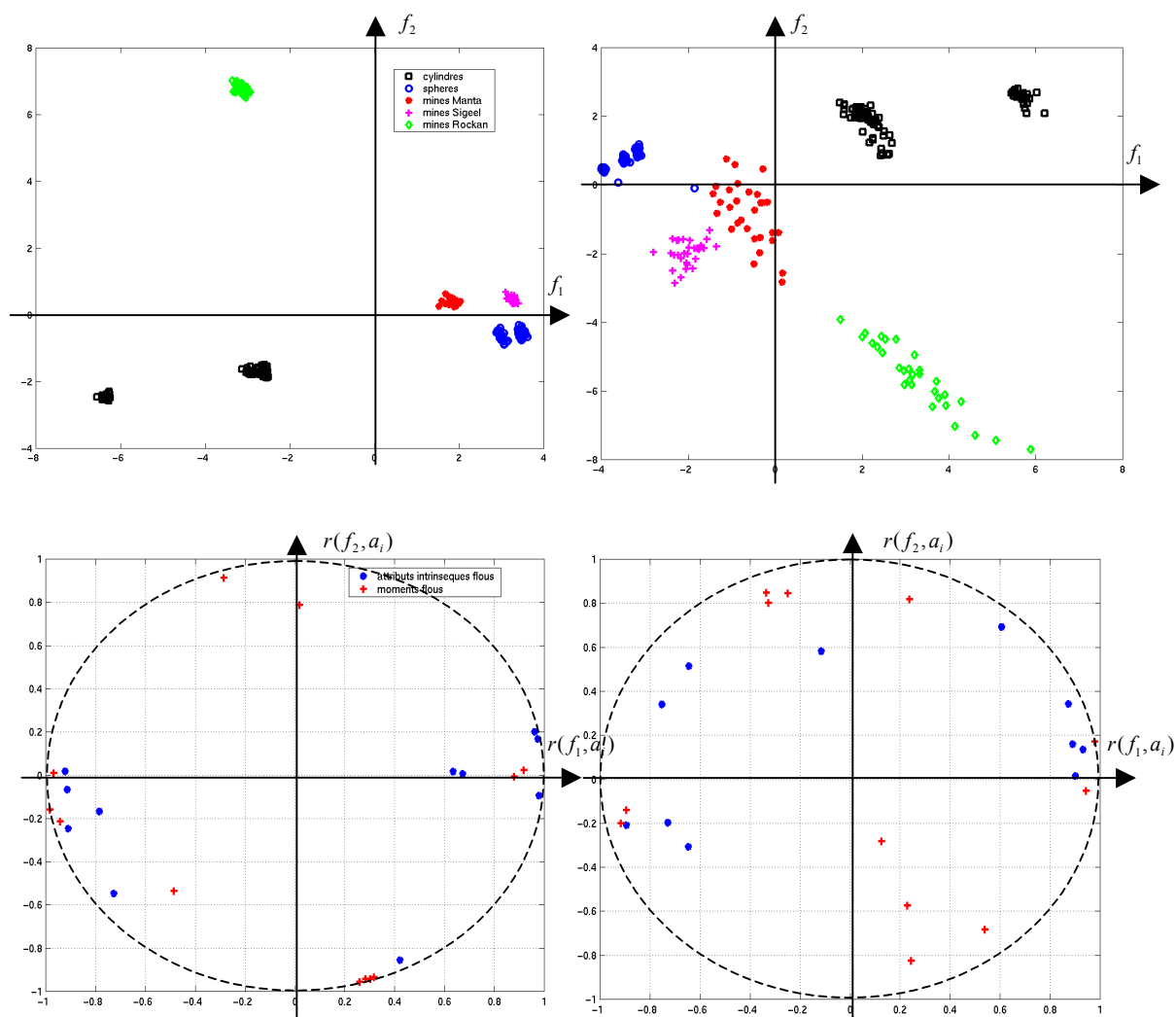


Figure 76 – ACP (en haut) et cercle de corrélations associé (en bas) pour les images I_s (à gauche) et I_b (à droite)

Tableau 16 – Taux de bonne classification pour trois classifieurs : discrimination linéaire (ligne 1), nuées dynamiques (ligne 2) et k-ppv avec k=1 (ligne 3)

Classes attributs	<i>cylindres</i>	<i>sphères</i>	<i>Mines Manta</i>	<i>Mines Sigeeel</i>	<i>Mines Rockan</i>	moyenne (écart-type)
<i>Images I_s</i>	100	100	100	100	100	100 (0)
	100	100	100	100	100	100 (0)
	100	100	100	100	100	100 (0)
<i>Images I_b</i>	100	99.2	100	100	100	99.4 (0.4)
	100	99.2	100	100	100	99.4 (0.4)
	100	99.2	100	100	100	99.4 (0.4)

Tableau 17 – Taux de bonne classification pour trois classifieurs à partir des deux composantes principales : discrimination linéaire(ligne 1), nuées dynamiques (ligne 2) et k-ppv avec k=1 (ligne 3)

Classes attributs	<i>cylindres</i>	<i>sphères</i>	<i>Mines Manta</i>	<i>Mines Sigeel</i>	<i>Mines Rockan</i>	moyenne (écart-type)
<i>Images I_s</i>	100	100	100	100	100	100 (0)
	100	100	100	92.7	100	98.5 (3.3)
	100	100	100	100	100	100 (0)
<i>Images I_b</i>	100	99.2	89.4	100	100	97.7 (4.7)
	100	100	79.1	100	100	95.8 (9.3)
	100	99.2	100	100	100	99.8 (0.4)

IV. Conclusion

Pourvu que les conditions d'observation (altitude et distance du sonar à l'objet) soient identiques, N images présentant des points de vue distincts du même objet peuvent être superposées avec ou sans segmentation préalable. La dispersion des pixels d'ombre est alors transcrite dans la nouvelle image par le biais des niveaux gris associés. La caractérisation de cette zone d'ombre aux frontières floues est alors réalisée de manière probante par l'introduction d'attributs flous. Ces derniers sont obtenus par extension d'attributs déjà utilisés en mono-vue sur les images binaires au cas des images en niveaux de gris.

Conclusion de la partie IV

Dans cette partie, nous avons étudié différentes approches pour la classification multi-vues d'un objet.

Le premier chapitre est consacré à l'exploitation des données acquises image par image selon les traitements classiquement effectués en mono-vue, i.e. segmentation et extraction d'attributs sur l'image binaire obtenue. Suivant la précision avec laquelle le sonar opère, deux voies sont explorées :

- La fusion des données extraites au fil des prises de vues est une solution en cas de défaut de maîtrise de la navigation. Dans ce contexte, les classes d'objets caractérisés par des données imprécises et incertaines que sont les attributs s'apparentent à des ensembles flous. Le formalisme flou a permis la définition de modèles adaptés à diverses configurations d'ensembles d'individus qu'il convient de connaître a priori. Testée sur deux ensembles d'attributs retenus lors de l'application mono-vue, certains modèles ont permis d'atteindre 100% de bonne classification pour un nombre de vues relativement faible (inférieur ou égal à 6 vues).
- A l'inverse, si les attributs extraits peuvent être considérés conditionnellement à l'angle de vue, l'objet peut être classifié à partir d'une caractérisation globale des transformations géométriques qui affectent son ombre au cours de la séquence d'images. Une telle stratégie induit le calcul d'attributs sensibles à ces transformations. Avec seulement trois attributs et une trajectoire semi-circulaire du sonar, nous avons obtenu 99.4% de bonne classification.

Le second chapitre propose une extension de la caractérisation mono-vue sur image binaire au cas d'une image 'multi-vues' en niveaux de gris. Dans cette nouvelle image, la région d'intérêt floue résulte de la somme des pixels d'ombre de chacune des vues. Sa frontière est par conséquent d'autant plus floue que les ombres superposées diffèrent. La géométrie floue introduite par Rosenfeld est l'outil que nous avons choisi pour calculer des attributs flous inspirés de ceux utilisés classiquement en reconnaissance de formes. Sans écarter des attributs a priori, nous avons effectué une classification à partir de 23 attributs avant et après analyse en composantes principales. Ces premiers résultats sont apparus encourageants et montrent l'intérêt de cette approche multi-vues originale.

Conclusion générale

Ce travail de thèse fut consacré à l'étude de la classification multi-vues d'un objet immergé à partir de son ombre et d'images sonar. Le problème fut décomposé de la manière suivante : nos efforts se sont tout d'abord concentrés sur la spécificité de l'image sonar haute résolution et son exploitation pour la classification sur ombre avant d'intégrer ces connaissances dans un processus plus complexe de classification multi-vues. En effet, il apparaît intuitivement qu'une bonne analyse des traitements mono-vue prépare la mise en œuvre d'algorithmes de classification multi-vues.

L'instrument de l'imagerie acoustique est le sonar actif dont les spécifications fixent le niveau de détail avec lequel les objets insonifiés sont décrits. En matière de lutte antimines, les progrès de la technologie profitent à la neutralisation de la menace actuelle (mines nouvelle génération) grâce aux sonars haute fréquence, au traitement antenne synthétique ou encore à l'usage de véhicules à immersion variable.

Le lien entre le pixel de l'image sonar et la cellule de résolution permet d'interpréter les distorsions caractéristiques des ombres, accentuées par une forte rasance, sur une image généralement mieux résolue en distance qu'en gisement. Les bruits intervenant à divers niveaux de la chaîne d'acquisition des données sonar sont généralement indésirables. Néanmoins, un phénomène inhérent à l'imagerie sonar appelé réverbération constitue une donnée majeure dont est tiré le principe même de la classification sur ombre. En effet, l'ombre acoustique est caractérisée par un faible niveau de gris moyen des pixels dû à une absence de réverbération de fond.

La menace actuelle comprend des mines dites furtives de petites tailles, de formes atypiques et difficiles à détecter, opposées aux mines cylindriques et sphériques classiques. Pour y faire face, outre les performances exigées en termes de résolution, de nouvelles stratégies de chasse aux mines doivent être proposées comme la classification multi-vues.

Dans un premier temps, l'analyse d'images de bas-niveau a bénéficié de **l'intégration de connaissances sonar**. Un certain nombre de techniques classiques de segmentation ont été adaptées à l'imagerie sonar notamment par la prise en compte de la différence parfois notable entre la résolution en distance et la résolution en gisement et le suréchantillonnage appliqué. Cette donnée associée à la connaissance du principe de formation des images sonar a également permis la synthèse d'un filtre spatial facilitant le seuillage de l'image. Dans certaines approches, la connaissance de la nature du bruit est introduite via un terme d'attache aux données : il s'agit de la fonction d'énergie d'un champ de Markov et de l'énergie externe d'un contour actif. La même énergie a par ailleurs été optimisée par algorithme génétique afin d'explorer une population d'individus caractérisés par leurs descripteurs de Fourier. Par transformée inverse, les descripteurs estimés sont associés au contour qui, de manière optimale, partage l'image sonar en deux régions en permettant dans le même temps de classer l'objet ciblé.

Les choix effectués au cours de ce travail de thèse sont liés au contexte sonar et aux applications futures de chasse aux mines. Les **déformations caractéristiques des ombres acoustiques** ont été étudiées de manière à envisager une stratégie de classification supervisée robuste aussi bien en mono-vue qu'en multi-vues. Deux modes de classification sont possibles fonction du type d'attributs extraits et notamment, suivant qu'ils autorisent ou non une reconstruction du contour.

Classiquement, l'image binaire issue de la première étape de segmentation sert de support à l'extraction d'un certain nombre d'attributs visant à réduire la quantité de données. Celles-ci sont ensuite présentées à l'entrée du classifieur dans le cas d'une classification mono-vue, ou manipulées au cours d'une procédure plus complexe de classification multi-vues.

Les paramètres intrinsèques ont souvent été utilisés pour caractériser les ombres bien qu'à différentes tailles d'un objet d'une même classe correspondent parfois différentes valeurs de ces paramètres. En revanche, la classification multi-vues peut exploiter favorablement ces attributs : d'une part, l'information apportée par le rapport largeur sur hauteur indique l'orientation de l'objet vis-à-vis du sonar et d'autre part, les valeurs d'attributs bien choisis évoluent au cours d'une séquence de prises de vue d'un objet en fonction de l'évolution progressive de la forme de l'ombre.

Du point de vue de la seule classification mono-vue, l'invariance vis-à-vis des transformations du plan permet de caractériser chacun des objets d'une même classe par une grandeur commune indépendamment de sa taille et de l'angle de vue. Dans l'espace caractéristique, la classe entière est idéalement représentée par un seul point où s'est agglutiné l'ensemble des individus. En réalité, on cherche à obtenir le noyau le plus étroit possible c'est-à-dire une faible variance des attributs pour une même classe. Deux familles d'invariants ont été comparées. D'un côté, l'invariance vis-à-vis des similitudes est une voie intéressante et largement étudiée dans la littérature. Cependant, elle nécessite d'effectuer une normalisation de l'image sonar puisque l'élongation n'est pas prise en compte. D'un autre côté, la forte rasance et la différence parfois notable entre résolution en distance et résolution en gisement sont intrinsèquement corrigées par les invariants affines. Parmi ces derniers, nous avons introduit des invariants de moments et des invariants par combinaison de descripteurs de Fourier. Outre leur intérêt commun évident pour s'affranchir des transformations caractéristiques des ombres, les seconds possèdent l'avantage supplémentaire d'effectuer un lissage du contour. Cette famille d'invariants est en effet formée après décomposition du contour en série de Fourier.

Enfin, la formation d'un vecteur caractéristique hybride de seulement quatre composantes a montré l'intérêt d'associer des attributs de natures différentes, à savoir paramètres intrinsèques et invariants affines. En les combinant par une analyse en composantes principales, leur complémentarité a été mise en évidence par une bonne partition des classes dans le plan principal. L'hybridation apparaît une parade efficace à l'évolution de la menace en diversifiant les attributs et leurs propriétés au sein d'un même vecteur caractéristique.

Les descripteurs de Fourier permettent de reconstruire le contour. S'ils sont calculés distinctement sur les abscisses et les ordonnées des pixels de la frontière, ils possèdent la propriété de linéarité. Cette propriété a été exploitée par deux techniques originales de classification supervisée qui ont, entre autres, l'avantage de ne pas nécessiter une importante base d'apprentissage généralement indispensable à l'entraînement d'un classifieur.

La première est une technique d'appariement de silhouettes entre un prototype représentatif d'une classe donnée et le contour de l'ombre dont la déformation s'apparente à une transformation affine. La base d'apprentissage est par conséquent réduite à un seul prototype par classe pour les objets de forme simple testé. Sa mise en œuvre sur des objets de forme plus complexe nécessiterait d'étendre le nombre de prototypes par classe. Traiter le cas

général de la transformation affine nécessite une paramétrisation adaptée. Un point critique de l'utilisation des descripteurs de Fourier est l'indispensable correction du décalage du point de départ garantissant une description unique des contours. Ce point est écarté par l'introduction d'invariants affines dérivés des mêmes descripteurs de Fourier. L'estimation des paramètres de la déformation sont ensuite effectuée au sens des moindres carrés.

La seconde approche fait appel aux notions de modèles déformables, modèles statistiques et algorithmes génétiques : elle consiste en une combinaison originale, coordonnée par algorithme génétique, des étapes de segmentation, d'extraction d'attributs et de classification dans un processus dynamique est opposée au processus séquentiel classique. Sans exiger une qualité de résultat à chacune de ces étapes, on fait l'économie des pré- ou post-traitements qui leur sont associés. Directement basée sur les données sonar brutes, cette deuxième technique met en scène un certain nombre de candidats potentiels (menace actuelle concrétisant la base d'apprentissage) qui, au fil des générations, convergent vers une solution optimale : il s'agit du contour le mieux adapté aux données de l'image (critère statistique d'adaptation de l'algorithme génétique). Dans ce contexte, la plus grande partie des gènes du chromosome sont relatifs aux descripteurs de Fourier permettant de reconstruire le contour. Les opérateurs génétiques proposés exploitent la nature des descripteurs de Fourier en utilisant la propriété de linéarité et de niveau de détail.

Des études menées en classification mono-vue sont issues **deux approches multi-vues conceptuellement différentes**.

La première est la fusion des informations (attributs) provenant des différentes vues. Cette technique paraît particulièrement indiquée dans des conditions de navigation mal maîtrisée, ou pour l'analyse de vues prises par différents sonars sans chronologie précise. Le guidage des véhicules autonomes pose effectivement un problème (courants marins...) qui peut justifier ce type de traitement. La multiplicité des vues et leur complémentarité permet de traiter les incertitudes et les imprécisions qui entachent chacune des vues prises séparément. Utiliser la théorie des ensembles flous pour l'aide à la décision présente l'avantage d'une grande souplesse à la modélisation. Ceci nous a permis de définir plusieurs modèles fonction de la configuration des individus par classe : la notion d'élément idéal et le problème de la distribution des attributs ont été traités en ce sens. Deux types d'attributs ont été testés : les attributs hybrides et les descripteurs de Fourier normalisés en échelle. Les propriétés de ces derniers ont offert la possibilité d'effectuer une combinaison pondérée des mesures. Les points de vue dits non caractéristiques ont en effet pu être minorés grâce à la prise en compte du rapport largeur sur hauteur directement relié aux descripteurs fondamentaux. Le point fort de la fusion proposée est de pouvoir combiner les informations indépendamment des points de vue et d'en extraire des mesures pour l'aide à la décision (laquelle incombe finalement à l'opérateur).

Contrairement à la première, la seconde approche se place sous l'hypothèse d'un contexte opératoire contraint. La classification de l'objet est, dans ce cas, basée sur l'évolution des attributs au cours d'une séquence d'images. Grâce à une vision globale des attributs extraits, la description de l'objet est plus complète que la précédente. Cette stratégie s'est avérée adaptée aux objets dont la forme de l'ombre portée varie en fonction des points de vue. La distinction entre deux objets est d'autant meilleure que les attributs reflètent franchement les transformations géométriques de l'ombre (attribut intrinsèque) ou encore sa forme affranchie de la transformation affine qui l'affecte (invariant affine). Les opérateurs de corrélation sont les outils privilégiés pour discriminer les objets par comparaison des courbes expérimentales et théoriques d'évolution des attributs. Cette approche semble réaliste compte tenu des

progrès technologiques qui permettent d'envisager une trajectoire circulaire d'un porteur filoguidé (véhicule à immersion variable PVDS) autour de la cible.

La dernière voie explorée entre dans le cadre des applications de la logique floue en traitement d'images. L'idée est de quantifier les différentes formes de l'ombre à partir d'une seule image. Effectivement, sous la condition d'un recalage adéquat, les ombres se superposent d'autant moins que leurs silhouettes se distinguent. **La zone d'ombre obtenue est une région floue** dont les pixels ont un niveau de gris fonction de leur pertinence d'appartenance à l'ombre. En étendant le calcul d'attributs sur images binaires aux images à plusieurs niveaux de gris, la notion d'attributs flous a été introduite afin de caractériser la dispersion des pixels. Outre les attributs flous géométriques, nous avons défini les moments flous qui entrent entre autres dans le calcul d'attributs topologiques. Des résultats probants par classifieur k-plus-proches voisins et une analyse en composantes principales justifient l'intérêt de cette analyse multi-vues originale. Néanmoins, ces premiers résultats sont obtenus dans des conditions de simulation favorables au recalage des vues et à leur superposition. Des tests supplémentaires portant sur la navigation du porteur, la résolution, et le nombre d'images entre autres, devront être menés pour juger de son applicabilité dans des conditions réelles.

Les approches multi-vues proposées répondent aux limites posées par l'approche mono-vue en atteignant d'excellentes performances en terme de bonne classification (>99%). La logique floue appliquée à la caractérisation d'une région ombre floue apparaît une voie prometteuse mais à approfondir. Les deux autres approches exploitent différemment les attributs extraits image par image. Sous réserve d'un contexte opérationnel favorable pour lequel la navigation est parfaitement connue, la caractérisation de l'évolution des attributs au cours de la séquence d'images est à privilégier. En effet, contrairement à la fusion qui 'écrase' une partie de l'information, chaque attribut est considéré conditionnellement à l'angle de prise de vue. A travers les attributs, c'est donc la diversité des formes de l'ombre qui transparaît. Cette approche anticipe par ailleurs la menace future qui comportera probablement de plus en plus de mines de forme géométrique complexe.

Les outils de traitements introduits (partie II) ont été choisis pour leur adaptation au contexte sonar (partie I) et en vue de leur mise en œuvre, voire leur extension, dans un processus de classification multi-vues (partie IV). La spécificité des images sonar mais aussi la menace actuelle en Guerre des Mines ont guidé nos travaux vers des techniques privilégiant la prise en compte de connaissances a priori, des méthodes de reconnaissance de formes qui intègrent les transformations géométriques de l'ombre (linéarité des descripteurs de Fourier) ou s'en affranchissent (invariants), des processus innovants de classification multi-vues obéissant à différentes stratégies d'acquisition des images par le porteur... Les résultats de classification encourageants justifient les choix effectués compte tenu des hypothèses posées. Les différentes approches sont résumées dans le tableau de la page suivante qui pointe les exploitations mono-vue ou multi-vues des attributs les plus judicieuses. Les approches multi-vues proposées demandent à présent à être validées grâce au PVDS par exemple, sur des images acquises en situations réelles, et diverses en termes de navigabilité, distance à l'objet, nature du fond...

	Approches	Attributs	Attributs hybrides	Descripteurs de Fourier	Attributs intrinsèques	Invariants
Mono-vue	<u>Classifieurs classiques</u> (p102 et 104)		✓	✓	✓	✓
	<u>Méthodes de classification alternatives</u> : 1. appariement de silhouettes (p107) 2. ‘segmentation orientée classification’ par algorithme génétique (p120)			✓✓		
Multi-vues	<u>Fusion des attributs mono-vue</u> provenant de plusieurs images (p145) <i>gère incertitude et imprécision</i>		✓✓ avec l’opérateur de fusion F_m (moyenne, p165)	✓✓ avec l’opérateur de fusion F_{mp} (moyenne pondérée $\Rightarrow$ fiabilité des sources, p166)		
	<u>Caractérisation de l’évolution des attributs au cours d’une séquence d’images</u> (p168) <i>en cas de navigation maîtrisée</i>				✓✓✓	
	<u>Caractérisation floue de la superposition des vues</u> (p177): ombre ~ région floue				✓✓	

Légende :

- ✓ désigne les attributs mis en œuvre
- ✓✓ désigne une mise en œuvre prometteuse de ces attributs à approfondir
- ✓✓✓ désigne une mise en œuvre réussie de ces attributs conditionnellement au cas mono-vue ou multi-vues considéré
- ✓✓✓✓ désigne la mise en œuvre conseillée de ces attributs dans le cas multi-vues

Annexes

ANNEXE A : Rappel des principes théoriques de quelques techniques classiques de segmentation

a) Segmentation markovienne

i) Principe théorique

Segmenter revient en somme à rechercher des informations cachées à partir des données observées. Réalisée dans le cadre de l'estimation bayésienne, la segmentation nécessite l'introduction de contraintes ou connaissances a priori sur les primitives recherchées puisque la seule modélisation des phénomènes physiques ayant permis le recueil des observations ne suffit généralement pas. Dans le cas particulier des images sonar, le formalisme markovien permet ainsi d'intégrer un modèle a priori fonction de la géométrie des ombres à segmenter (terme contextuel ou de régularisation) et exploite la nature particulière des images acoustiques par une modélisation adaptée du speckle (terme d'attache aux données).

Terminologie :

Lien entre termes spécifiques à la modélisation markovienne et les données sonar	
Champ des observations	image observée
Champ des étiquettes	image segmentée
Attache aux données	propriétés statistiques des intensités des pixels
régularisation	propriétés locales et globales du champ des étiquettes

La segmentation de l'image sonar consiste à **estimer le vecteur des variables aléatoires du champ des étiquettes** $X = \{X_s, s \in S\}$ à valeurs dans Λ_{et} en fonction des observations $Y = \{Y_s, s \in S\}$ à valeurs dans Λ_{obs} c'est-à-dire l'ensemble des pixels de l'image (et leurs niveaux de gris), et des paramètres du modèle markovien incluant le modèle de bruit (terme d'attache aux données) et le modèle a priori (terme contextuel). L'intérêt majeur de l'association d'une image à un **champ de Markov** est de pouvoir spécifier **localement** un modèle qui est appréhendé de façon globale. En d'autres termes, la probabilité d'observer l'événement $(X_s = x_s)$ ne dépend que de la réalisation du champ X dans le voisinage du site s d'un ensemble S comme l'explique la définition suivante :

Soit $X = \{X_s, s \in S\}$ un champ de variables aléatoires définies sur un ensemble fini de sites de S et à valeurs dans Λ_{et} , alors X est un **champ de Markov** si et seulement s'il existe un système de voisinage ν sur S tel que :

- Pour toute configuration x de X , $P(X = x) > 0$,
- Pour tout site s de S , $P(X_s = x_s \mid X_r = x_r, r \in S - \{s\}) = P(X_s = x_s \mid X_r = x_r, r \in \nu_s)$

Rappel des principes théoriques de quelques techniques classiques de segmentation

Dès lors, pour pouvoir utiliser les champs de Markov, il faut connaître les lois de probabilités pour chaque site. Le théorème d'Hammersley-Clifford permet d'établir l'équivalence entre champ de Gibbs et champ de Markov et de caractériser celui-ci par une densité de probabilité [BESA-74]. La **distribution de probabilité du champ markovien** X fait alors intervenir un ensemble C de cliques relativement au voisinage v et un ensemble $\{V_c | c \in C\}$ de potentiels d'interactions locales⁴. Elle s'écrit sous la forme suivante :

$$P_X(X = x) = \frac{1}{Z} \exp\{-U(x)\}$$

avec $Z = \sum_{x \in \Lambda_{et}} \exp\{-U(x)\}$ fonction de partition et $U(x) = \sum_{c \in C} V_c(x)$ fonction d'énergie.

En se plaçant dans le contexte statistique, segmenter revient à rechercher l'estimateur optimal $\hat{x}$ qui associe à une réalisation y du champ des observations une configuration d'étiquettes estimée avec un 'risque' minimal. La fonction de coût $c(x, \hat{x})$ mesurant la distance entre le champ des étiquettes et son estimation permet de caractériser le 'risque' associé à cet estimateur :

$$\hat{x}_{opt}(y) = \arg \min_{\hat{x}} \sum_{y \in \Lambda_{obs}} \sum_{x \in \Lambda_{et}} c(x, \hat{x}) P_{XY}(x, y) = \arg \min_{\hat{x}} \sum_{x \in \Lambda_{et}} c(x, \hat{x}) P_{X|Y}(x | y)$$

L'estimateur au sens du Maximum de vraisemblance A Posteriori (MAP) pour lequel $c(x, \hat{x}) = 1 - \delta(x, \hat{x})$ se traduit, en utilisant la règle de Bayes, par la maximisation d'une probabilité :

$$\hat{x}_{MAP} = \arg \max_{x \in \Lambda_{et}} P_{X|Y}(x | y) = \arg \max_{x \in \Lambda_{et}} P_{Y|X}(y | x) P_X(x)$$

La probabilité $P_{Y|X}(y|x)$ traduit la **vraisemblance** des observations conditionnellement aux étiquettes alors que $P_X(x)$ permet l'introduction de **connaissances a priori** sur les étiquettes.

En faisant l'hypothèse a priori markovienne sur le lien d'attache aux données, le produit des deux lois $P_{Y|X}(y|x)P_X(x)$ conservent une structure de champ de Gibbs. En notant $V(x, y) = -\ln P_{Y|X}(y|x)$, on a

$$\hat{x}_{MAP} = \arg \max_{x \in \Lambda_{et}} \left\{ \frac{1}{Z} \exp[-(V(x, y) + U(x))] \right\} = \arg \min_{x \in \Lambda_{et}} (V(x, y) + U(x))$$

La segmentation est alors faite en deux temps :

- **estimation des paramètres du modèle markovien** (paramètre des lois statistiques associées à l'ombre et à la réverbération de fond et définition des cliques agissant sur les regroupements de pixels)

⁴ Soit S un ensemble de sites s à cardinal fini N , on appelle clique, tout sous-ensemble c de S qui comporte :

- soit un seul pixel (clique singleton ou unaire) ;
- soit un ensemble de pixels dont les sites sont voisins deux à deux.

V_c est une application de $\Omega = \Lambda_{et}^S$ dans $\mathbb{R}$ qui à une configuration x du champ X associe un nombre $V_c(x)$ appelé potentiel d'interaction local ou fonction de potentiel. La valeur $V_c(x)$ ne dépend de la configuration x que par l'intermédiaire des valeurs d'états des sites de la clique c .

Rappel des principes théoriques de quelques techniques classiques de segmentation

- segmentation proprement dite, c'est-à-dire estimation du champ des étiquettes, par **optimisation de la fonction d'énergie** $V(x, y) + U(x)$.

ii) Optimisation

Pour satisfaire des contraintes de temps de calcul, l'optimisation de la fonction d'énergie $U(x, y)$ du modèle markovien adopté pour la segmentation de l'image peut être effectuée par l'algorithme Iterated Conditional Modes (ICM) qui consiste en la recherche du MAP [BESA-86] :

La démarche est globalement la suivante :

- choisir une configuration initiale aussi proche que possible d'une configuration optimale
- balayer l'ensemble des sites de manière aléatoire et, en chaque site, calculer l'ensemble des probabilités conditionnelles en affectant successivement l'étiquette $x_s = e_0 = 0$ et $x_s = e_1 = 1$. On compare alors son influence sur l'énergie globale $V(x, y) + U(x)$ en calculant l'énergie locale sachant qu'à une étiquette x_s correspond une unique observation y_s et que la seule connaissance du voisinage du site s suffit pour définir sa loi de probabilité.

On donne alors au site visité l'étiquette correspondant au minimum d'énergie locale (l'état qui correspond à la probabilité maximale)

- poursuivre les balayages pour un nombre déterminé d'itérations ou jusqu'à la réalisation d'un critère d'arrêt.

En utilisant cet algorithme, on ne doit cependant pas perdre de vue que le résultat final, c'est-à-dire l'image segmentée, est influencé par la stratégie de visite des sites choisie et reste fortement conditionné par l'initialisation. Cette dernière est rendue plus robuste en effectuant une analyse multirésolution des données, ou mieux encore en utilisant cette technique pour effectuer une segmentation multirésolution c'est-à-dire appliquer l'algorithme d'optimisation à chacune des résolutions [THOU-95].

b) les contours actifs statistiques [CHES-99]

D'une manière générale, le **contour actif** ou 'snake' est un modèle élastique se présentant sous la forme d'une courbe dont l'initialisation se situe à proximité du contour recherché. La courbe évolue sous contraintes selon un processus itératif de **minimisation d'une fonctionnelle** [KASS-88] [COCQ-95]. Cette méthode est bien adaptée à notre problème puisqu'elle mise sur la recherche de la frontière (contour fermé) qui partage l'image de manière optimale (au sens d'un critère statistique) en deux zones si l'on considère les distributions associées des niveaux de gris des pixels (obéissant à des lois de Rayleigh) indépendantes [MALK-00].

La formulation discrète du modèle identifie le contour à une suite de points discrets ou pixels représentant les sommets du contour associé à une ligne polygonale fermée.

Soient $I = \{(x, y) | (x, y) \in \text{image observée}\}$ et $W = \{w(x, y) | (x, y) \in \text{image binaire}\}$, le problème revient à trouver le contour C de W qui partage l'image en deux régions, intérieure $\phi_{in} = \{(x, y) | w(x, y) = 0\}$ et extérieure $\phi_{out} = \{(x, y) | w(x, y) = 1\}$ que sont idéalement les régions 'ombre' et 'réverbération de fond'. Comme pour l'approche par champs de Markov, il revient à maximiser la probabilité a posteriori :

$$P_{w|i}(w|i) \text{ qui est proportionnel à } P_{i|w}(i|w)P_w(w)$$

La probabilité $P_{i|w}(i|w)$ traduit la **vraisemblance** des observations conditionnellement au contour (attache aux données) alors que $P_w(w)$ permet l'introduction de **connaissances a priori** sur le contour (régularisation).

En choisissant une approche bayésienne, après estimation au sens du Maximum de Vraisemblance (MV) des paramètres définissant le modèle d'attache aux données (bruit), nous injectons ces estimées dans la formule de log-vraisemblance, i.e. $\ln(P_{i|w}(i|w))$ qui caractérise l'énergie externe E_{ext} .

L'énergie interne E_{int} est quant à elle reliée à la probabilité a priori $P_w(w)$ analogue à l'énergie élastique définie par Kass et al. pour leur modèle de contours actifs [KASS-88].

Finalement, le contour estimé au sens du MAP est :

$$\hat{w}_{MAP} = \arg \min_w E_{tot}$$

$$\text{avec } E_{tot} = (1 - \lambda).E_{ext} + \lambda.E_{int}$$

où λ est un paramètre qui gère le poids accordé à l'énergie interne de régularisation du contour.

c) La Ligne de Partage des Eaux (LPE) [MEYE-90] [VACH-95]

Dans ce concept, l'image est vue comme une surface en trois dimensions dans laquelle les zones homogènes sont des bassins versants associés à des minima que sont idéalement les marqueurs (cf. figure suivante).

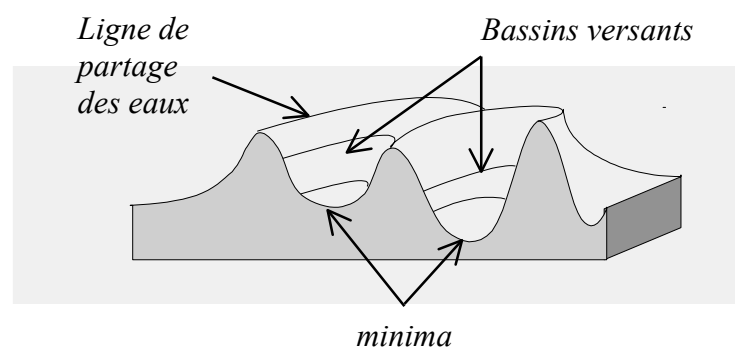


Figure – Caractères topographiques d'une image

La **ligne de partage des eaux** ou **LPE** s'apparente à une crête limitant deux bassins versants. La segmentation consiste précisément à partitionner l'image sonar en deux régions homogènes. Ces régions peuvent être caractérisées par un minimum régional et des bassins versants encadrant une vallée. Cette vue en relief de l'image est le résultat du calcul du **gradient morphologique**. Le module du gradient morphologique correspond à la notion physique de contraste. Si l'on peut extraire de chaque zone un pixel de gradient minimal

Rappel des principes théoriques de quelques techniques classiques de segmentation

(contraste faible), on peut alors partir de ce marqueur et « inonder » progressivement la zone. Lorsque les eaux provenant de deux marqueurs appartenant à des zones distinctes se rejoignent, les digues fictives qu'il faudrait ériger pour séparer ces deux plans d'eau constituent la « ligne de partage des eaux » du gradient.

Cette approche morphologique se décompose en deux principales étapes (accompagnées de pré et post-traitements) : une étape de marquage et une étape d'extraction des contours des régions sélectionnées.

En théorie, à tout minimum est associé le bassin versant d'où une goutte d'eau, suivant la ligne de plus grande pente, arrivera dans ce minimum. En pratique, nous ne rechercherons pas ce minimum mais nous bornerons à sélectionner un point, le marqueur, dans chacune des deux zones caractéristiques de l'image sonar : l'ombre et la réverbération de fond.

L'image est prétraitée par égalisation visant à améliorer le gradient sur la frontière entre l'ombre et le fond en renforçant le contraste. La *pseudo*-égalisation consiste à négliger les pixels les plus foncés et les plus clairs qui apportent une information indésirable dans le cas présent. Ce prétraitement est bienvenu ici du fait de la nature des traitements effectués ultérieurement qui n'exigent aucune grandeur statistique mais relèvent de morphologie mathématique en niveaux de gris. En l'occurrence, on tire naturellement profit d'une meilleure répartition des niveaux (tout en respectant la hiérarchisation initiale).

Le résultat de la segmentation par LPE dépend de la qualité de l'image gradient. Les contours des régions de l'image originale correspondent aux zones de forte transition d'intensité sur l'image et synthétisent les lignes de crêtes de l'image gradient. La moindre discontinuité locale d'intensité sera alors ponctuée d'un point de partage. La multiplication de ces lignes de partage insignifiantes et indésirables en terme de segmentation (sur-segmentation) est limitée par l'utilisation d'un élément structurant approprié. Néanmoins, il est bienvenu également de débruiter l'image avant de calculer le gradient morphologique. Pour ce faire, on applique des opérations morphologiques en niveaux de gris telles qu'une fermeture, une ouverture et une érosion géodésique. Le principe est de travailler sur tous les niveaux de l'image à la fois. Le calcul de la ligne de partage des eaux est appliqué sur l'image gradient : ainsi les points de partage des eaux correspondent aux crêtes du gradient autour de ces marqueurs (points de contour), c'est-à-dire aux lieux de fortes intensités. Les minima régionaux associés à chacune des régions doivent permettre d'initialiser **l'algorithme d'inondation** par lequel l'image est vue comme une surface percée au niveau des marqueurs. Cette surface est alors plongée progressivement dans l'eau. La LPE est l'ensemble des barrages qu'il faut ériger pour empêcher les eaux provenant de « lacs » différents de se rencontrer.

ANNEXE B : Analyse multirésolution par ondelettes

L'analyse multirésolution est particulièrement adaptée dans le traitement des images sonar fortement bruitées par speckle. Le pas de base de cette décomposition consiste à projeter l'image observée sur une échelle de plus basse résolution. En réitérant ce principe, on peut alors obtenir une succession d'approximations où chaque approximation est une version lissée de la précédente. A cette fin, **la décomposition en ondelettes** est une approche classique d'analyse multirésolution des signaux et des images telle que l'a introduite S. Mallat [MALL-89].

Dans un cas monodimensionnel, l'analyse multirésolution orthogonale dyadique (c'est-à-dire avec un facteur de résolution 2 entre deux échelles consécutives) est construite autour de deux fonctions intimement liées :

- La **fonction d'échelle** ϕ à partir de laquelle on construit par dilatations et translations les fonctions $\phi_{m,n}(x) = 2^{-m/2} \phi(2^{-m}x - n)$

L'ensemble des fonctions $\phi_{m,n}$ forme une famille de fonctions orthogonales, à m constant.

Le rôle de cette fonction est de fournir, par projection, une *approximation* du signal à la résolution 2^m .

- Les ondelettes $\varphi_{m,n}(x)$ sont générées à partir de l'**ondelette mère** ou **analysante** $\varphi(x)$ telles que $\varphi_{m,n}(x) = 2^{-m/2} \varphi(2^{-m}x - n)$

La fonction φ est choisie de telle sorte que $\varphi_{m,n}$ constitue une base orthonormale de $L^2(\mathbb{R})$. En décomposant le signal dans cette base, on récupère *l'information perdue* lors du passage de l'approximation à la résolution 2^{m-1} à l'approximation plus grossière de résolution 2^m .

On appelle l'approximation d'un signal à la résolution 2^m **résumé** tandis que l'information perdue est appelée **détail**.

Dans ces deux formulations, m est le paramètre échelle et n le paramètre de déviation.

Concrètement, la projection sur ϕ (resp. sur φ) est assimilée à un filtrage passe-bas (resp. passe-haut) du signal.

Les coefficients résultants de la décomposition en ondelettes orthogonales du signal f sont donc les suivants [MALL-89] :

$$\text{Les coefficients de détails } d_{m,n} = \langle f, \varphi_{m,n} \rangle = \int_{-\infty}^{+\infty} \varphi_{m,n}(x) \cdot f(x) \cdot dx$$

$$\text{Et les résumés } r_{m,n} = \langle f, \phi_{m,n} \rangle = \int_{-\infty}^{+\infty} \phi_{m,n}(x) \cdot f(x) \cdot dx$$

On peut les obtenir par convolution à partir de la résolution 2^{m-1} , avec les filtres h et g associés aux fonctions ϕ et φ :

$$r_{m,n} = \sum_k h(2n-k) r_{m-1,k} \quad \text{et} \quad d_{m,n} = \sum_k g(2n-k) r_{m-1,k}$$

$$\text{où } g(l) = h(1-l) \cdot (-1)^l \quad \text{et} \quad h(n) = \sqrt{2} \int_{-\infty}^{+\infty} \phi(x-n) \phi(2x) dx$$

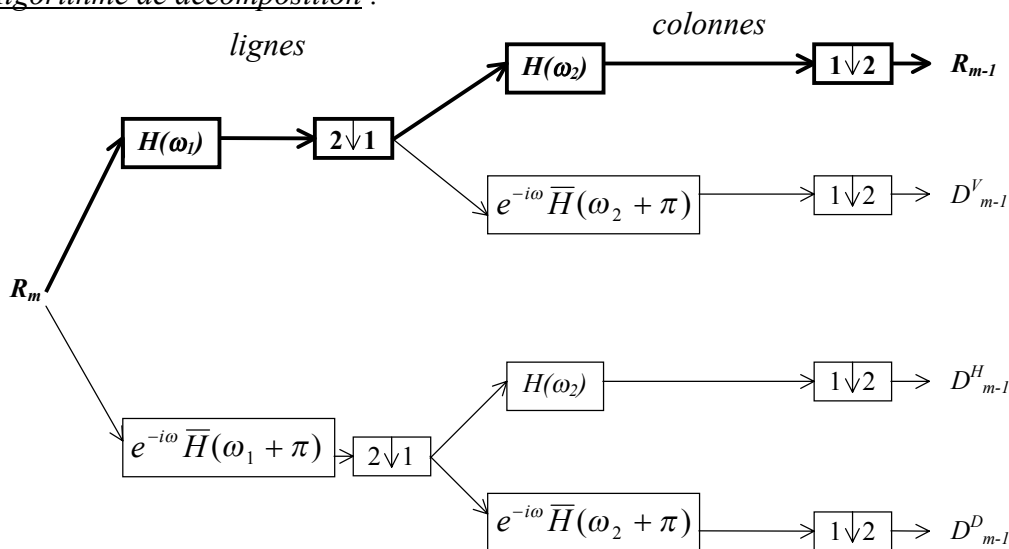
Spécification de l'ondelette :

Un choix important réside dans celui de la base d'ondelettes utilisée pour notre application. A partir de l'algorithme multirésolution de S. Mallat, I. Daubechies [DAUB-88] a construit une famille d'ondelettes à support compact pour l'analyse et la synthèse d'images dont la fonction de Haar est la plus rudimentaire. Ces ondelettes constituent une famille d'ondelettes orthogonales indexées par un entier N définissant leur régularité. La localisation précise de l'ombre étant essentielle, il faut travailler avec des fonctions de régularité faible. En effet, plus la régularité est élevée et plus le nombre de points définissant la réponse impulsionnelle du filtre associé est important.

La décomposition en ondelettes d'une image peut être représentée comme deux décompositions monodimensionnelles. Cette fois, la décomposition s'effectue en deux temps par une convolution des lignes par H et G puis des colonnes, H et G (G est le filtre miroir de H) étant les transformées de Fourier de h et g .

Le parcours en gras est celui que nous empruntons pour obtenir une **décomposition multi-échelle de l'image**.

Algorithme de décomposition :



H pour direction horizontale, V pour direction verticale et D pour direction diagonale.

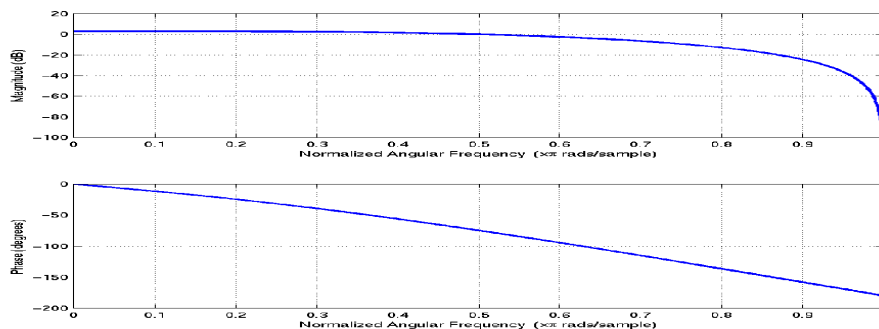
R_m résumé à la résolution 2^m , D_m détail à la résolution 2^m ,

$1 \downarrow 2$ supprime une ligne sur deux,

$2 \downarrow 1$ supprime une colonne sur deux.

Voici les coefficients du filtre $h_2(n)$ (de longueur $4=2N$) que nous utilisons :

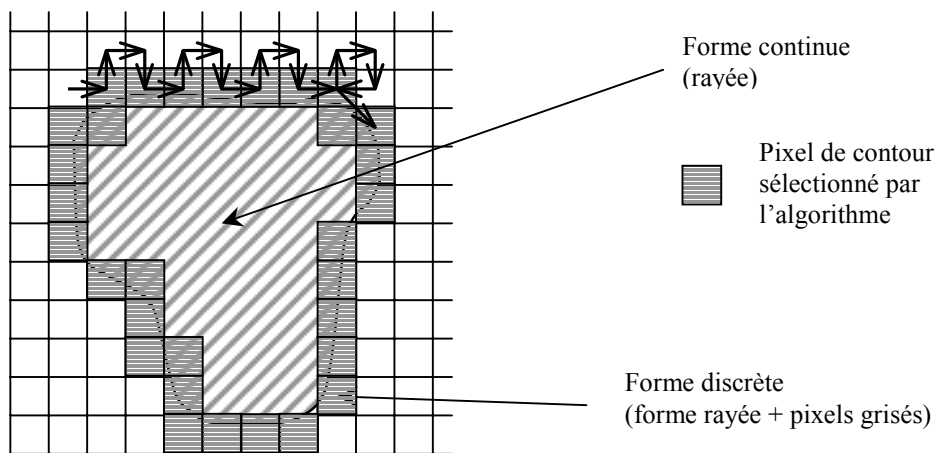
$$h_2(n) = \begin{bmatrix} \frac{1+\sqrt{3}}{4\sqrt{2}} & \frac{3+\sqrt{3}}{4\sqrt{2}} & \frac{3-\sqrt{3}}{4\sqrt{2}} & \frac{1-\sqrt{3}}{4\sqrt{2}} \end{bmatrix} \text{ de réponse fréquentielle :}$$



ANNEXE C : Extraction du contour

1) Algorithme de suivi de contour

Le premier pixel de l'ombre est généralement détecté lors d'un balayage ligne à ligne ou de telle manière que l'une de ses coordonnées coïncide avec l'une des coordonnées du centre de gravité. Pour les pixels suivants, on raisonne comme suit : si le pixel appartient à l'ombre, le pixel suivant est celui qui est à sa gauche (de manière à *sortir* de la forme) sinon on prend le pixel situé à sa droite (de manière à *rattraper* le contour) et ce jusqu'à la fermeture du contour. Ce faisant, nous circule dans le sens des aiguilles d'une montre.



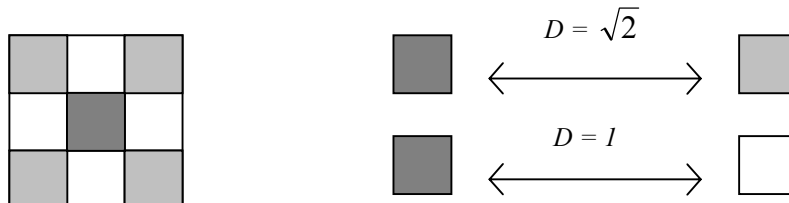
NB : quand on a un point double, c'est-à-dire lorsque deux points de contour consécutifs sont identiques, on néglige le second avant de poursuivre à partir du premier mais en procédant autrement : on sélectionne celui des 3 pixels de la ligne ou colonne incrémentée de 1 suivant le sens de parcours qui appartient à la forme tout en lui affectant le même sens de parcours.

2) Equirépartition des points de contour

Etant donné le contour d'une silhouette sous forme d'une liste de coordonnées de points discrets ou pixels, on construit un contour de vélocité uniforme de N_c points équidistants en rééchantillonnant le contour original.

Algorithme d'équirépartition :

- on calcule la longueur totale L du contour en tenant compte du fait que la distance entre deux pixels diffère selon leur position telle que :



- on calcule alors le pas l de parcours pour équirépartir N_equi points, i.e. $l = L/N_equi$,
- à partir du premier pixel puis en parcourant le contour **dans le sens des aiguilles d'une montre**, on répertorie tous les points équidistants de l en prenant en compte la connexité.

ANNEXE D : Décomposition en série de Fourier

Les descripteurs de Fourier et leurs propriétés ont été largement discutés dans la littérature [GRAN-72] [WALL-80] [MITC-83] [HENO-93].

1) Définition

Soit C un contour fermé d'une silhouette décrit sous la forme d'une liste de N coordonnées complexes de pixels $z(i) = z_i = x_i + jy_i, i=0 \dots N-1$ (suivant l'algorithme de suivi de contour et l'équirépartition de N points $(x_i, y_i)_{i=0 \dots N-1}$ proposés dans l'annexe C). Le contour peut être parcouru une infinité de fois et la fonction z_i est par conséquent considérée comme une fonction périodique de période N . En discret, la décomposition classique du contour en série de Fourier s'applique à la fonction complexe z_i telle que :

$$C_k = \frac{1}{N} \sum_{i=0}^{N-1} z_i \exp\left(-j2\pi \frac{ki}{N}\right) \quad \text{Transformation de Fourier discrète}$$

2) Propriétés

La décomposition du contour en série de Fourier permet de disposer d'une information à différents niveaux de détail. D'un côté, les basses fréquences (descripteurs d'indice faible) correspondent aux variations globales du contour. D'un autre côté, les hautes fréquences (descripteurs d'indice fort) codent les détails du contour. En limitant le nombre d'attributs, le lissage de contour est ainsi intrinsèquement réalisé. En particulier, la reconstruction à partir des seuls descripteurs fondamentaux C_1 et C_{-1} permet d'obtenir une approximation elliptique de la silhouette. Par ailleurs, le descripteur particulier C_0 sert uniquement à localiser la silhouette.

Afin de caractériser le contour de manière unique, il convient d'effectuer une normalisation des descripteurs de Fourier qui le caractérisent. En effet, un même contour sera décrit par des descripteurs de valeurs distinctes si un décalage est apparu dans la procédure préalable de suivi de contour. En d'autres termes, la paramétrisation du contour conditionne le résultat de la transformation de Fourier discrète de la manière suivante :

Soient deux descriptions du même contour décrit par N points équirépartis $z_0(i), i=0, \dots, N-1$,

$$C_k^0 = \frac{1}{N} \sum_{i=0}^{N-1} z_0(i) \exp\left(-j2\pi \frac{ki}{N}\right) \text{ et } C_k = \frac{1}{N} \sum_{i=0}^{N-1} z(i) \exp\left(-j2\pi \frac{ki}{N}\right)$$

Avec $z(i) = z_0(i+l)$, l désignant le décalage du point de départ à la paramétrisation, on a

$$C_k = \frac{1}{N} \sum_{i=0}^{N-1} z_0(i+l) \exp\left(-j2\pi \frac{ki}{N}\right) = \frac{1}{N} \sum_{m=l}^{N+l-1} z_0(m) \exp\left(-j2\pi \frac{k(m-l)}{N}\right) = C_k^0 \exp\left(j2\pi \frac{lk}{N}\right)$$

soit

$$C_k = C_k^0 \tau^k \quad \text{avec} \quad \tau = \exp\left(j2\pi \frac{l}{N}\right) \quad (2)$$

Pour rendre les descripteurs de Fourier discriminants, l'invariance par rapport au point de départ doit donc obligatoirement être assurée.

Les descripteurs de Fourier ont en outre l'avantage de répercuter simplement les transformations géométriques de base des points du contour à savoir la translation, la rotation et l'homothétie du contour :

<i>Translation</i>	$z'_i = z_i + T$	$\Leftrightarrow$	$C'_0 = C_0 + T$
<i>Homothétie</i>	$z'_i = \rho z_i$	$\Leftrightarrow$	$C'_k = \rho C_k$
<i>Rotation</i>	$z'_i = z_i \exp(j\phi)$	$\Leftrightarrow$	$C'_k = C_k \exp(j\phi)$

3) Reconstruction

La transformation de Fourier est une opération réversible et la reconstruction s'effectue par **transformation de Fourier inverse** à partir des descripteurs :

$$z_i = \sum_{k=-M}^M C_k \exp\left(j2\pi \frac{ki}{N}\right)$$

Le nombre M de descripteurs de Fourier est fixé en fonction du niveau de détail désiré. Une version lissée du contour est alors obtenue en s'affranchissant des hautes fréquences (relatives à des détails du contour indésirables).

ANNEXE E : Moments

1) Définitions [PROK-92]

a – Moments réguliers

Soit la fonction de distribution continue à deux variables $f(x,y)$, son moment d'ordre $(p+q)$, où p et q sont des entiers naturels, est défini par :

$$m_{pq} = \int \int_{-\infty}^{+\infty} x^p y^q f(x,y) dx dy, \quad p, q = 0, 1, 2, \dots$$

Cette formulation correspond à la projection de la fonction $f(x,y)$ sur le monôme $x^p y^q$.

Remarque : la base $x^p y^q$, $p, q \in \mathbb{N}^2$ n'est pas orthogonale.

On montre alors que si $f(x,y)$ est continue par parties et définie pour (x,y) appartenant à un ensemble fini du plan, alors les moments m_{pq} existent $\forall (p,q) \in \mathbb{N}^2$.

b – Moments centrés

Les moments centrés sont exprimés par :

$$\mu_{pq} = \int \int_{-\infty}^{+\infty} (x - \bar{x})^p (y - \bar{y})^q f(x,y) dx dy, \quad p, q = 0, 1, 2, \dots$$

$$\text{ou bien [PROK-92], } \mu_{pq} = \sum_{r=0}^p \sum_{s=0}^q C_p^r C_q^s \bar{x}^{p-r} \bar{y}^{q-s} m_{rs}, \text{ avec } C_n^p = \frac{n!}{p!(n-p)!}$$

où le centre de gravité est $(\bar{x}, \bar{y}) = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right)$, m_{00} est la masse totale de $f(x,y)$ c'est-à-dire

$$m_{00} = \int \int_{-\infty}^{+\infty} f(x,y) dx dy.$$

Les quantités μ_{pq} sont invariantes par rapport aux translations du plan qu'on peut effectuer sur $f(x,y)$.

c – Moments centrés normalisés ou moments réduits

On définit alors les moments centrés normalisés par $\mu'_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\left[\frac{(p+q)}{2} + 1 \right]}}$

Les quantités μ'_{pq} sont invariantes par rapport à la composition d'une translation et d'une homothétie du plan.

Application au cas image :

En numérique, pour une image binaire $M \times N$, l'intégrale devient une sommation sur les pixels de coordonnées (i,j) ,

$$m_{pq} = \sum_M \sum_N i^p j^q I(i,j), \text{ où } I(i,j)=1 \text{ si } (i,j) \in \text{ombre}, 0 \text{ sinon}$$

$$\mu_{pq} = \sum_M \sum_N (i - i_g)^p (j - j_g)^q I(i,j),$$

avec les coordonnées du centre de gravité de l'ombre (i_g, j_g) tel que $i_g = \frac{m_{10}}{m_{00}}$ et $j_g = \frac{m_{01}}{m_{00}}$.

2) Transformations des moments

a – homothétie

Soit le changement d'échelle d'un facteur α sur les abscisses et β sur les ordonnées, tel que l'on a la nouvelle fonction $g(x,y) = f\left(\frac{x}{\alpha}, \frac{y}{\beta}\right)$, les moments transformés m'_{pq} de $g(x,y)$ s'expriment simplement en fonction des moments m_{pq} de $f(x,y)$ tel que

$$\begin{aligned} m'_{pq} &= \alpha^{1+p+q} \beta^{1+p+q} m_{pq} & \alpha \neq \beta \\ m'_{pq} &= \alpha^{2+p+q} m_{pq} & \alpha = \beta \end{aligned}$$

b – translation

Soit la translation d'un facteur α sur les abscisses et β sur les ordonnées, tel que l'on a la nouvelle fonction $g(x,y) = f(x - \alpha, y - \beta)$, les moments transformés m'_{pq} de $g(x,y)$ s'expriment simplement en fonction des moments m_{pq} de $f(x,y)$ tel que

$$m'_{pq} = \sum_{r=0}^p \sum_{s=0}^q C_p^r C_q^s \alpha^{p-r} \beta^{q-s} m_{rs}$$

c – rotation

Soit une rotation d'angle θ autour de l'origine de $f(x,y)$ définie telle que $g(x,y) = f(x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta)$, les moments transformés m'_{pq} de $g(x,y)$ s'expriment simplement en fonction des moments m_{pq} de $f(x,y)$ tel que

$$m'_{pq} = \sum_{r=0}^p \sum_{s=0}^q C_p^r C_q^s (-1)^{q-s} (\cos \theta)^{p-r+s} (\sin \theta)^{q+r-s} m_{p+q-r-s, r+s}$$

3) Propriétés des moments d'ordre inférieur ou égal à deux

a – Ordre zéro

Le moment d'ordre zéro $m_{00} = \int \int_{-\infty}^{+\infty} f(x, y) dx dy = \mu_{00}$ représente la masse totale de la distribution $f(x, y)$. En numérique, pour une image binaire faisant apparaître une forme pleine, m_{00} représente la surface de cette forme (par convention, les pixels appartenant à la silhouette sont égaux à un tandis que les pixels du fond sont nuls).

b – Ordre un

Calculés sur une image binaire, $\{m_{10}, m_{01}\}$ permettent de localiser le centre de gravité de la forme segmentée tel que ses coordonnées soient données par $\left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right)$.

c – Ordre deux

Les moments d'ordre deux $\{m_{02}, m_{11}, m_{20}\}$ sont connus comme les *moments d'inertie* et sont utilisés pour déterminer un certain nombre d'attributs caractéristiques comme :

- les axes principaux

L'orientation de l'axe principal le plus proche de l'axe des x est défini par,

$$\phi = \frac{1}{2} \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) \text{ et tel que } -\frac{\pi}{4} \leq \phi \leq \frac{\pi}{4}.$$

- les longueurs des grand et petit axes (a, b) de l'ellipse approximant la forme,

$$\begin{cases} a = \sqrt{\frac{2[\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}]}{\mu_{00}}} \\ b = \sqrt{\frac{2[\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}]}{\mu_{00}}} \end{cases}$$

ANNEXE F : Invariants de Moments

A) Invariants de moments centrés normalisés de Hu [HU-62]

$$\begin{aligned}
 M_1 &= \mu'_{20} + \mu'_{02} \\
 M_2 &= (\mu'_{20} - \mu'_{02})^2 + 4(\mu'_{11})^2 \\
 M_3 &= (\mu'_{30} - 3\mu'_{12})^2 + (3\mu'_{21} - \mu'_{03})^2 \\
 M_4 &= (\mu'_{30} + \mu'_{12})^2 + (\mu'_{21} + \mu'_{03})^2 \\
 M_5 &= (\mu'_{30} - 3\mu'_{12})(\mu'_{30} + \mu'_{12})[(\mu'_{30} + \mu'_{12})^2 - 3(\mu'_{21} + \mu'_{03})^2] + \\
 &\quad (3\mu'_{21} - \mu'_{03})(\mu'_{21} + \mu'_{03})[3(\mu'_{30} + \mu'_{12})^2 - (\mu'_{21} + \mu'_{03})^2] \\
 M_6 &= (\mu'_{20} - \mu'_{02})[(\mu'_{30} + \mu'_{12})^2 - (\mu'_{21} + \mu'_{03})^2] + 4\mu'_{11}(\mu'_{30} + \mu'_{12})(\mu'_{21} + \mu'_{03}) \\
 M_7 &= (3\mu'_{21} - \mu'_{03})(\mu'_{30} + \mu'_{12})[(\mu'_{30} + \mu'_{12})^2 - 3(\mu'_{21} + \mu'_{03})^2] - \\
 &\quad (\mu'_{30} - 3\mu'_{12})(\mu'_{21} + \mu'_{03})[3(\mu'_{30} + \mu'_{12})^2 - (\mu'_{21} + \mu'_{03})^2]
 \end{aligned}$$

B) Invariants de moments de Zernike exprimés en termes de moments centrés normalisés introduits par Teague [TEAG-80]

Ordre deux :

$$\begin{aligned}
 S_1 &= Z_{20} = 3[2(\mu'_{20} + \mu'_{02}) - 1]/\pi \\
 S_2 &= |Z_{22}|^2 = 9[(\mu'_{20} - \mu'_{02})^2 + 4(\mu'_{11})^2]/\pi^2
 \end{aligned}$$

Ordre trois :

$$\begin{aligned}
 S_3 &= |Z_{33}|^2 = 16[(\mu'_{03} - 3\mu'_{21})^2 + (\mu'_{30} - 3\mu'_{12})^2]/\pi^2 \\
 S_4 &= |Z_{31}|^2 = 144[(\mu'_{03} + \mu'_{21})^2 + (\mu'_{30} + \mu'_{12})^2]/\pi^2 \\
 S_5 &= 2RE[(Z_{33})^*(Z_{31})^3] \\
 &= 13824[(\mu'_{03} - 3\mu'_{21})(\mu'_{03} + \mu'_{21})[(\mu'_{03} + \mu'_{21})^2 - 3(\mu'_{30} + \mu'_{12})^2] \\
 &\quad - (\mu'_{30} - 3\mu'_{12})(\mu'_{30} + \mu'_{12})[(\mu'_{30} + \mu'_{12})^2 - 3(\mu'_{03} + \mu'_{21})^2]]/\pi^3 \\
 S_6 &= 2RE[(Z_{31})^2(Z_{22})^*] \\
 &= 864[(\mu'_{02} - \mu'_{20})[(\mu'_{03} + \mu'_{21})^2 - (\mu'_{30} + \mu'_{12})^2] \\
 &\quad + 4\mu'_{11}(\mu'_{03} + \mu'_{21})(\mu'_{30} + \mu'_{12})]/\pi^3
 \end{aligned}$$

Ordre quatre :

$$S_7 = |Z_{44}|^2 = 25[(\mu'_{40} - 6\mu'_{22} + \mu'_{04})^2 + 16(\mu'_{31} - \mu'_{13})^2]/\pi^2$$

$$S_8 = |Z_{42}|^2 = 25 \left\{ [4(\mu'_{04} - \mu'_{40}) + 3(\mu'_{20} - \mu'_{02})]^2 + 4[4(\mu'_{31} + \mu'_{13}) - 3\mu'_{11}]^2 \right\} / \pi^2$$

$$S_9 = Z_{40} = 5[6(\mu'_{40} + 2\mu'_{22} + \mu'_{04}) - 6(\mu'_{20} + \mu'_{02}) + 1] / \pi$$

$$S_{10} = 2RE[(Z_{44})^*(Z_{42})^2]$$

$$= 250 \left\{ (\mu'_{40} - 6\mu'_{22} + \mu'_{04}) [4(\mu'_{04} - \mu'_{40}) + 3(\mu'_{20} - \mu'_{02})]^2 - 4[4(\mu'_{31} + \mu'_{13}) - 3\mu'_{11}]^2 \right. \\ \left. - 16[4(\mu'_{04} - \mu'_{40}) + 3(\mu'_{20} - \mu'_{02})][4(\mu'_{31} - \mu'_{13}) - 3\mu'_{11}](\mu'_{31} - \mu'_{13}) \right\} / \pi^3$$

$$S_{11} = 2RE[Z_{42}(Z_{22})^*] = 30 \left\{ [4(\mu'_{04} - \mu'_{40}) + 3(\mu'_{20} - \mu'_{02})](\mu'_{02} - \mu'_{20}) + 4\mu'_{11}[4(\mu'_{031} - \mu'_{13}) - 3\mu'_{11}] \right\} / \pi^2$$

ANNEXE G : Base d'images simulées

On dispose de 5 classes d'objets : cylindres (classe C_1), sphères (classe C_2), mines Manta (classe C_3), mines Sigeel (classe C_4) et mines Rockan (classe C_5) (cf. partie I, §IV). La base de données utilisée est constituée de séquences d'images acquises sous différents angles de vue suivant une trajectoire circulaire à 30m de la cible et avec une hauteur sonar de 5m. On utilise en outre les spécifications VHF du sonar classificateur TSM 2022MK3 en mode normal :

/***** Description de la base de données *****/

Simulateur : SIMU

Sonar : Sonar TSM 2022MK3 monté sur véhicule PVDS

Caractéristiques :

Distance cible-sonar R : 30m

Hauteur du sonar : 5m (sauf pour Sigeel 7m)

Cibles :

sphères diamètre 1.1m, 0.9m, 0.7m

cylindres (D 0.45m, L 1.53m)

(D 0.6m, L 2.0m)

(D 0.5m, L 2.7m)

Manta

Sigeel

Rockan

Trajectoire :

circulaire avec prise de vue tous les 10 degrés

Résolution sonar pour PVDS avec pas intervoie=0.2 degré :

en distance $0.03125m = c/2B$

en gisement $0.10472m = \text{pas intervoie} * R$

Dimension du pixel pour PVDS (suréchantillonnage) :

en distance $0.015625m = c/2B/2$

en gisement : résolution avec suréchantillonnage, i.e.

pour les cylindres: 0.032725m

pour les sphères: 0.032725m

pour Manta: 0.032725m

pour Sigeel: 0.032725m

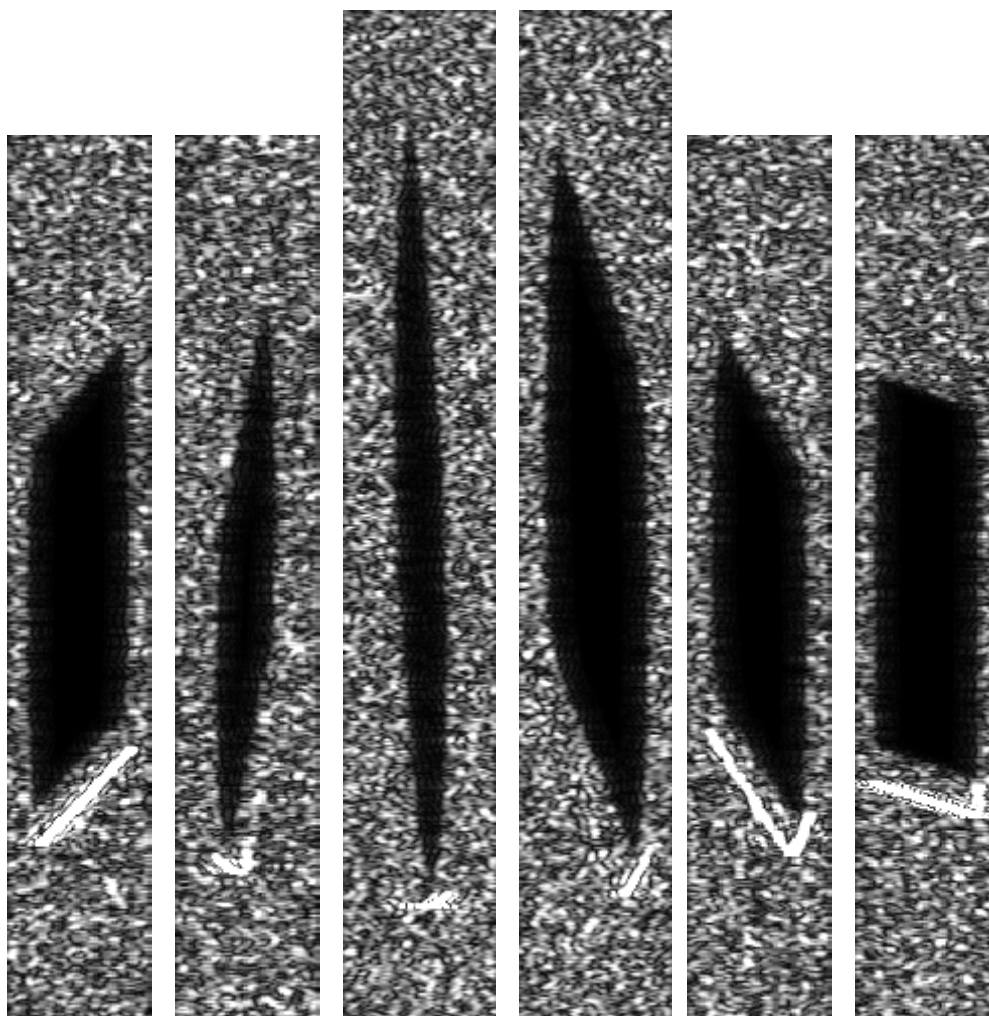
pour Rockan: 0.00942478m

Différentes réalisations (une trentaine) du bruit pour chacune des séquences (un objet donné et une hauteur sonar donnée) afin de disposer de 300 images environ par séquence d'objet à symétrie radiale et 570 images pour les cylindres et la mine Rockan.

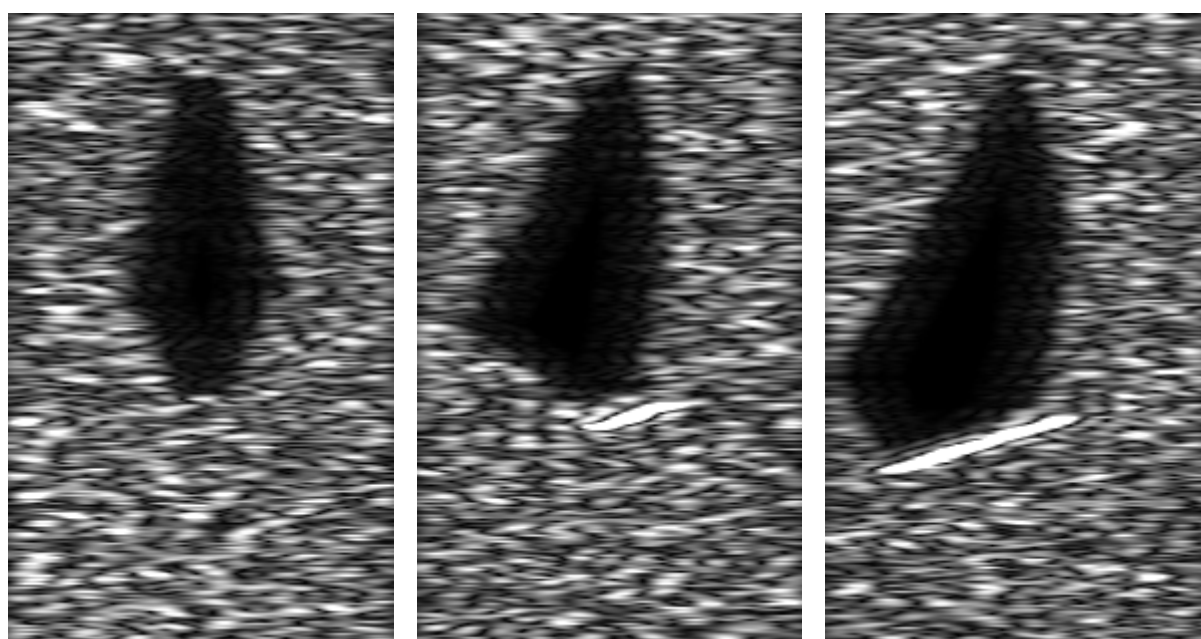
Les traitements appliqués concernent donc les séquences suivantes :

- 30 séquences de 19 images pour 3 tailles de cylindres soient $3*30*19$ images,
- 30 séquences de 10 images pour 3 tailles de sphères soient $3*30*10$ images,
- 30 séquences de 10 images pour la mine Manta soient $30*10$ images,
- 30 séquences de 10 images pour la mine Sigeel soient $30*10$ images,
- 30 séquences de 19 images pour la mine Rockan soient $30*19$ images.

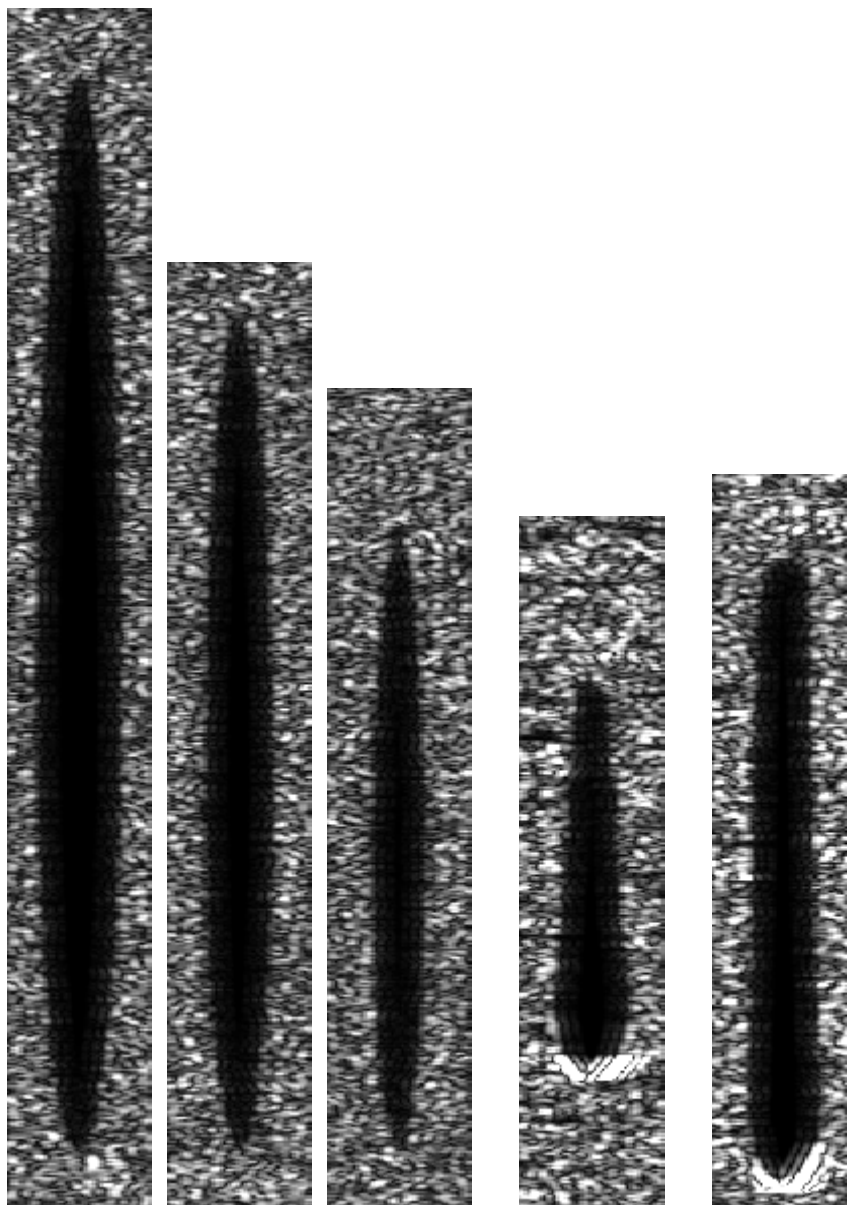
Les illustrations suivantes regroupent quelques exemples pour chacune des cinq classes.



cylindres (2 tailles)



mines Rockan



sphères (3 tailles), mine Manta, mine Sigee

ANNEXE H : Algorithme fuzzy C-means ou ISODATA flou [DUNN-74] [BEZD-81]

L'objectif est de partitionner un ensemble d'objets à classer en C classes par optimisation d'un critère.

Soit $\chi = \{X_1, X_2, \dots, X_N\}$, N objets de l'espace euclidien R^p . On cherche à assigner un degré d'appartenance μ_{ij} à chaque objet j de χ dans la classe ou cluster $c_{i, i \in \{1, \dots, C\}}$. Ce degré d'appartenance a les propriétés suivantes :

$$\mu_{ij} \in [0, 1], \quad \sum_{j=1}^N \mu_{ij} > 0 \quad \forall i, \quad \sum_{i=1}^C \mu_{ij} = 1 \quad \forall j$$

Le clustering flou diffère du clustering net du fait que μ_{ij} puisse prendre des valeurs dans l'intervalle $[0, 1]$.

L'algorithme fonctionne de manière itérative pour minimiser le critère suivant :

$$J(U, v) = \sum_{j=1}^N \sum_{i=1}^C (u_{ij})^m (d_{ij})^2 \quad \text{avec} \quad d_{ij}^2 = \|V_k(j) - v_i\|^2$$

où

U est la matrice de partition de dimension $N \times C$ dont l'élément u_{ij} désigne le degré d'appartenance du point j au cluster i représenté par son centroïde

$v = \{v_1, \dots, v_C\}$ est l'ensemble des C centroïdes des C clusters

$m \geq 1$ définit la nature plus ou moins floue du clustering qui augmente avec m

La distance choisie (euclidienne ou de Mahalanobis) dépend de la matrice A définie positive telle que :

$$d_{ij}^2 = \|V_k(j) - v_i\|^2 = (V_k(j) - v_i)^T A (V_k(j) - v_i)$$

A est la matrice identité dans le cas le plus simple.

L'algorithme est le suivant :

Initialisation aléatoire de la matrice U

A l'itération n (repérée par le symbole $^{(n)}$),

1. Calcul des P composantes des centroïdes :

$$v_{il} = \frac{\sum_{j=1}^N (u_{ij}^{(n)})^m x_{jl}}{\sum_{j=1}^N (u_{ij}^{(n)})^m} \quad l = 1, \dots, P$$

2. Mise à jour de $U^{(n)}$ en calculant $U^{(n+1)}$ comme suit :

Pour j de 1 à N,

Soit $I_j = \{i \text{ tels que } d_{ij} = 0\}$ et $\bar{I}_j = \{1, \dots, C\} - I_j$,

Si l'ensemble I_j vide,

$$u_{ij}^{(n+1)} = \frac{1}{\sum_{k=1}^C \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}$$

sinon,

$$u_{ij}^{(n+1)} = 0 \quad \forall i \in \bar{I}_j \quad \text{et} \quad \sum_{i \in I_j} u_{ij}^{(n+1)} = 1$$

3. Arrêt si $\|U^{(n)} - U^{(n+1)}\| < \varepsilon$ sinon retour à l'étape 1.

Les paramètres suivants sont utilisés dans le présent rapport :

m=2

$\varepsilon=10^{-4}$

C=10

$$\|U^{(n)} - U^{(n+1)}\| = \frac{1}{N \times C} |u_{ij}^{(n)} - u_{ij}^{(n+1)}|$$

ANNEXE I : Introduction au logiciel NeuroClasse

NeuroClasse est un logiciel créé par le Laboratoire Central de Recherche de Thomson-CSF (actuel groupe Thales) au début des années 1990. Il intègre les fonctionnalités de haut niveau nécessaires à la conception, la mise en œuvre et l'évaluation de classifieurs neuronaux ou classiques. Une quinzaine d'algorithmes de discrimination et de classification y sont développés [LCR-92a][LCR-92b]. Il fournit une analyse des données ainsi qu'une méthodologie pour qualifier notre application de classification en imagerie sonar.

1) NeuroClasse en pratique

Traitements effectués pour la création d'un 'modèle' associé à un algorithme de classification supervisée :

- chargement de la base de données d'entrée `base_1`,
- chargement d'un module de discrimination,
- activation de la base `base_1`,
- lancement de l'**apprentissage** par l'option '`train`'. On obtient en résultat une nouvelle base de données en sortie contenant les probabilités, estimées par l'algorithme, d'affectation des vecteurs de longueur P de la base d'entrée `base_1` aux différentes classes K. Par cette procédure d'apprentissage, le module de classification vient *d'apprendre un 'état'*. Il est maintenant configuré pour classer une nouvelle base de données (dont on cherche à estimer les probabilités d'appartenance aux classes).
- étape de **généralisation** (pour contrôler son pouvoir discriminant) avec une démarche similaire :
 - chargement d'une nouvelle base de données `base_2`,
 - activation de `base_2`, désactivation de `base_1`,
 - lancement de la généralisation par l'option '`use`'. On obtient en résultat une nouvelle base de données en sortie contenant les probabilités, estimées par l'algorithme, d'affectation des vecteurs de taille P de la base d'entrée `base_2` aux différentes classes K.

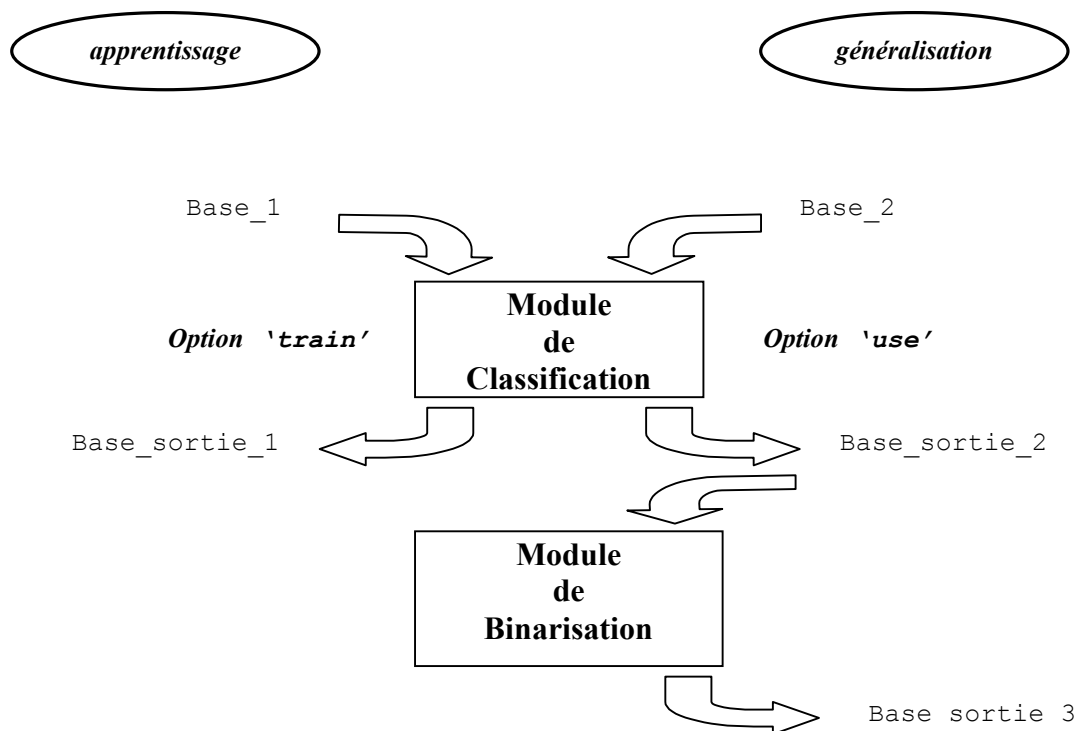
Pour entraîner le modèle représentatif du classifieur choisi, il est donc nécessaire de disposer de deux bases de données disjointes (l'une pour la phase d'apprentissage, l'autre pour la phase de généralisation) dont on connaît les probabilités d'affectation des patterns (ou individus, observations...) sur les classes. Ces deux bases sont appelées 'ensemble d'apprentissage' et 'ensemble de généralisation' et ont en commun le nombre P d'attributs des vecteurs d'entrée et le nombre de classes K. Ces deux ensembles sont particulièrement importants et doivent représenter le plus fidèlement possible les vecteurs (caractérisant les ombres dans notre cas) que nous cherchons à classer. D'autre part, de manière à donner à chaque attribut la même importance que celle des autres attributs dans le calcul des distances entre patterns, il convient d'effectuer l'une des opérations suivantes :

- centrer et réduire les données par l'utilitaire de pré-traitement `center`
- effectuer une analyse en composantes principales par l'utilitaire de pré-traitement PCA

Chiffrage des résultats : la matrice de confusion compare les valeurs estimées par le classifieur (probabilités d'appartenance aux classes) et les valeurs désirées définies dans la

base de données en entrée du classifieur. A noter qu'un post-traitement (binarisation par l'utilitaire de post-traitement `binarize`) est nécessaire sur les valeurs estimées de manière à prendre une décision sur les probabilités d'appartenance en affectant chaque pattern i de probabilité P_i à la classe la plus probable parmi les K classes possibles.

2) Synoptique de la chaîne de traitement



ANNEXE J : Calcul de la fonction d'adaptation

La fonction d'adaptation prend en compte sur l'ensemble des pixels des deux régions 'ombre' et 'réverbération' tel que :

$$E = -N_O \ln \alpha_O^2 - N_R \ln \alpha_R^2$$

avec

$$\hat{\alpha}_{region}^2 = \frac{\sum_{(x,y) \in region} [i_{region}(x,y)]^2}{2.N_{region}}, \text{ où } region \text{ désigne } O \text{ ou } R$$

Une manière d'accélérer le calcul, est de transformer ce problème à deux dimensions (lignes et colonnes) en un problème monodimensionnel. Pour ce faire, nous calculons l'histogramme des niveaux de gris des pixels $Hist(k)$, $k=0 \dots 255$, de chacune des régions. Nous avons alors les équivalences suivantes :

$$\sum_{(x,y) \in region} i(x,y) = \sum_x \sum_y i(x,y) \equiv \sum_{k=1}^{255} kHist(k)$$

de même,

$$\sum_{(x,y) \in region} i^2(x,y) \equiv \sum_{k=1}^{255} k^2 Hist(k)$$

et d'une manière générale, quelle que soit la fonction f appliquée à l'intensité ou niveau de gris $i(x,y)$,

$$\sum_{(x,y) \in region} f[i(x,y)] = \sum_x \sum_y f[i(x,y)] \equiv \sum_{k=1}^{255} f[k]Hist(k)$$

On peut en outre réduire le nombre d'opérations de l'expression ci-dessus en tabulant préalablement la fonction f .

Publications

Journaux

I. Quidu, J.P. Malkasse et G. Burel, « Classification de mines sous-marines à partir de l'image sonar brute : caractérisation du contour de l'ombre portée par algorithme génétique », *Traitement du Signal*, soumis.

Conférences

I. Quidu, J.P. Malkasse, G. Burel and P. Vilbé, « Mine classification using a hybrid set of descriptors », *Proc. OCEANS'2000 MTS/IEEE*, Providence, Rhode Island, Sept 2000.

I. Quidu, J.P. Malkasse, G. Burel and P. Vilbé, « Mine Classification based on raw sonar data: an approach combining Fourier Descriptors, Statistical Models and Genetic Algorithms », *Proc. OCEANS'2000 MTS/IEEE*, Providence, Rhode Island, Sept 2000.

I. Quidu, J.P. Malkasse, G. Burel and P. Vilbé, « Mine classification based on a multiview characterisation », *Undersea Defence Technology (UDT) Europe 2001*, Hamburg, Germany, 26-28 june 2001.

I. Quidu, J.P. Malkasse, G. Burel and P. Vilbé, « A 2-D Filter Specification for Sonar Image Thresholding », *Advanced Concepts for Intelligent Vision Systems (ACIVS'2001) conference*, Baden-Baden, Germany, July 30 – August 3, 2001.

I. Quidu, J.P. Malkasse, P. Vilbé and G. Burel, « Fusion multi-attributs d'images sonar », *colloque GRETSI'01*, Toulouse, 10-13 septembre 2001.

Rapports d'étude

I. Quidu, « Reconnaissance de formes en imagerie sonar », *Rapport TMS SAS 99/B/EGS/NC/087-IQ* du 13 juillet 1999.

I. Quidu, « Classification multi-vues d'un objet immergé à partir de son ombre portée sur le fond », *Rapport TMS SAS 00/B/EGS/RG/06-IQ* du 12 janvier 2000.

I. Quidu et J.P. Malkasse, « Classification automatique des ombres. Rapport d'étude – PEA sonar multicibles en guerre des mines », *Rapport TMS SAS 00/B/EGS/DR/104-IQ-JPM* du 12 septembre 2000.

Bibliographie

- [ARBT-89] K. Arbter, « Affine-Invariant Fourier Descriptors », in *From Pixels to Features*. Amsterdam, The Netherlands : Elsevier Science, 1989.
- [ARBT-90] K. Arbter, W. E. Snyder, H. Burkhardt and G. Hirzinger, « Application of Affine-Invariant Fourier Descriptors to Recognition of 3-D Objects », *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 12, n° 7, July 1990, pp. 640-647.
- [BANO-81] G. Banon, « Distinction between several subsets of fuzzy measures », *Fuzzy sets and Systems*, Vol. 5, pp. 291-305, North-Holland Publishing Company, 1981.
- [BESA-74] J. Besag, « Spatial Interaction and the Statistical Analysis of Lattice Systems », *Journal of the Royal Statistical Society*, Vol. 36, pp. 192-236, 1974.
- [BESA-86] J. Besag, « On the statistical analysis of dirty picture », *Journal of the Royal Statistical Society*, B-48, pp. 259-302, 1986.
- [BEZD-81] J.C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, 1981.
- [BILL-98] D. Billon, F. Fohanno, "Theoretical performance and experimental results for synthetic aperture sonar self-calibration", *Proc. OCEANS'98 MTS/IEEE*, Nice, France, 1998, pp. 965-970.
- [BLOC-94] I. Bloch, H. Maître, « Fusion de données en traitement d'images : modèles d'information et décisions », *Traitement du signal* 1994, Vol. 11, n° 6, pp. 435-446.
- [BLOC-95] I. Bloch, « Fusion de données, ensembles flous et morphologie mathématique en traitement d'images », *Rapport du département Images*, Télécom Paris, avril 1995.
- [BLOC-96a] I. Bloch, « Information Combination Operators for Data Fusion : A Comparative Review with Classification », *IEEE Trans. on Systems, Man, and Cybernetics – Part A : Systems and Humans*, Vol. 26, n° 1, pp. 52-67, January 1996.
- [BLOC-96b] I. Bloch, « Incertitude, imprécision et additivité en fusion de données : point de vue historique », *revue Traitement du Signal*, Vol. 13, n° 4, 1996.
- [BLON-97] P. Blondel, and B.J. Murton, *Handbook of Seafloor Sonar Imagery*, Chichester: John Wiley and Sons, 1997.
- [BOSS-97] P.L. Bossart, D. David, J.M. Dinten et J.M. Chassery, « Détection de contours réguliers dans des images bruitées et texturées : une approche par contours actifs multiéchelle », *Traitement du Signal*, vol. 14, n° 2, 1997.
- [BOUR-98] J.M. Bouroche et G. Saporta, *L'analyse de données*, Presse Universitaires de France, 7^{ième} édition, 1998.
- [BOUV-92] M. Bouvet, *Traitements des signaux pour les systèmes sonar*, Ed. Masson, Coll. Technique et scientifique des télécommunications, 1992.
- [BRUC-92] M.P. Bruce, « A Processing Requirement and Resolution Capability Comparison of Side-Scan and Synthetic-Aperture Sonars », *IEEE J. of Oceanic Engineering*, vol. 17, n° 1, pp. 106-117, Jan. 1992.
- [CARI-99] C. Cariou, J-M. Ogier, S. Adam, J. Gardes, R. Mullot et Y. Lecourtier, « Reconnaissance de formes en multi-orinetation et multi-échelle sur documents techniques », *colloque GRETSI'99*, Vannes, 13-17 septembre 1999.
- [CELE-86] G. Celeux and J. Diebolt, « L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités », *Revue Statistique Appliquée*, Vol. 34, n° 2, 1986.
- [CELE-89] G. Celeux, E. Diday, G. Govaert, Y. Lechevallier et H. Ralambondrainy, *Classification automatique des données*, Dunod informatique, 1989.

- [CHAR-96] B. Charroux, « Analyse d'images : coopération d'opérateurs de segmentation guidée par l'interprétation », *Thèse de doctorat*, Univ. de Paris XI Orsay, 1996.
- [CHES-99] C. Chesnaud, P. Réfrégier, « Snake region based segmentation for different physical noise models and fast algorithm implementation », *invited lecture of PSIP'99*, 1999.
- [COCQ-95] J.-P. Cocquerez, S. Philipp, *Analyse d'images : filtrage et segmentation*, Ed. Masson, 1995.
- [COHE-93] L.D. Cohen and I. Cohen, « Finite-Element Methods for Active Contour Models and Balloons for 2-D and 3-D Images », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, n° 11, Nov. 1993.
- [COLE-99] T. Coleman, M.A. Branch and A. Grace, *Optimization Toolbox for Use with Matlab*, The MathWorks, Inc., Natick, MA, 1999.
- [CROS-83] G.R. Cross and A.K. Jain, « Markov random field texture models », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 5, n° 1, pp. 25-39, Jan. 1983.
- [CZES-92] E. Czesnalowicz, « Application de l'estimateur non paramétrique des k plus proches voisins en classification automatique multidimensionnelle », *Thèse de doctorat, Autom. et Info. Indus*, Univ. des Sciences et Technologies de Lille, 1992.
- [DANI-98a] S. Daniel, « Fusion multisource appliquée à la reconnaissance d'objets dans le milieu sous-marin », *Thèse de doctorat*, ENST Bretagne, 1998.
- [DANI-98b] S. Daniel, B. Solaiman and E.P. Maillard; « Object recognition on the sea-bottom using possibility theory and sonar data fusion », *Proceedings of the International Conference on Multisource-Multisensor Information Fusion*, Las vegas, USA, Vol. 1, pp. 396-402, 1998.
- [DAUB-88] I. Daubechies, « Orthonormal Bases of Compactly Supported Wavelets », *Communications on Pure and Applied Mathematics*, Vol. XLI 909-996, 1988.
- [DAVI-87] L. Davis, (Editor), *Genetic Algorithms and Simulated Annealing*, Morgan Kaufmann Publishers, Los Altos, CA, 1987.
- [DAVI-91] L. Davis, (Editor), *Handbook of Genetic Algorithms*, Van Nostrand Reingold, New-York, 1991.
- [DUDA-77] S.A. Dudani, K.J. Breeding and R.B. McGhee, « Aircraft Identification by Moments Invariants », *IEEE Trans. on Computers*, Vol. c-26, n° 1, Jan 1977.
- [DUNN-74] J.C. Dunn, « A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters », *Journal of Cybernetics*, Vol. 3, pp. 32-57, 1974.
- [ELOI-99] A. El Oirrak, M. Daoudi, D. Aboutajdine, H. El Bouyakhf, F. Ghorbel, « Estimation des paramètres d'un mouvement affine utilisant les descripteurs de Fourier », *GRETSI 99*, Vannes, 13-17 Sept. 1999, pp. 191-194.
- [EVAN-93] M. EVANS, N. HASTINGS and B. PEACOCK, *Statistical Distributions*, Second Edition - New-York: Chichester: Brisbane: Toronto: Singapore: John Wiley & Sons, Inc., 1993.
- [FIGU-97] M.A.T. Figueiredo, J.M.N. Leitão, A.K. Jain, « Adaptive parametrically deformable contours », *Proceedings of Computer Vision and Pattern Recognition*, pp.35-50, 1997.
- [FLET-80] R. Fletcher, « Practical Methods of Optimization », *Vol.1, Unconstrained Optimization, and Vol.2, Constrained Optimization*, John Wiley and Sons, 1980.
- [FLUS-93] J. Flusser and T. Suk, « Pattern recognition by affine moment invariants », *Pattern Recognition*, Vol. 26, n° 1, pp.167-174, 1993.
- [GEMA-84] S. Geman and D. Geman, « Stochastic relaxation, Gibbs distributions and the Bayesian restoration of the images », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, n° 6, pp. 721-741, Nov. 1984.

- [GHOR-90] F. Ghorbel, « Vers une approche mathématique unifiée des aspects géométriques et statistiques de la reconnaissance de formes planes », *Thèse de doctorat*, Université de Rennes I, 1990.
- [GHOR-94] F. Ghorbel, « A complete invariant description for grey-level images by the harmonic analysis approach », *Pattern Recognition Letters*, Vol. 15, pp. 1043-1051, 1994.
- [GOLD-91] D.E. Goldberg, K. Deb, « A Comparative Analysis of Selection Schemes Used in Genetic Algorithms », *Foundations of Genetic Algorithms*, ed. Gregory J. E. RAWLINS, San Mateo, California, USA :Morgan Kaufmann Publishers, pp. 69-93, 1991.
- [GOOD-76] J.W. Goodman, « Some fundamental properties of speckle », *J. Opt. Soc. Am.*, Vol. 66, n° 11, pp. 1145-1150, 1976.
- [GRAF-95] C. Graffigne, F. Heitz, P. Pérez, F. Prêteux, M. Sigelle and J. Zerubia, « Hierarchical Markov random field models applied to image analysis : a review », *Proc. SPIE, vol.2568, Neural Morphological and Stochastic Methods in Image and Signal Processing*, San Diego, 1995, pp. 2-17.
- [GRAN-72] G.H. Granlund, « Fourier preprocessing for hand print character recognition », *IEEE Trans. on Computers*, Vol. c-21, pp. 195-201, Feb. 1972.
- [GREN-84] D. Grenier, « Méthode de détection d'évolution : application à l'instrumentation nucléaire », *Thèse de doctorat, Automatique et Traitement du Signal*, Univ. de Technologie de Compiègne, 1984.
- [HADA-23] J. Hadamard, *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, Yale University Press, New Haven, 1923.
- [HASE-99] M. Haseyama, M. Kumagai, H. Kitajima, « A genetic algorithm based image segmentation for image analysis », *Proc. Of IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, Arizona, 1999.
- [HENO-93] H. Hénocq, G. Burel, « Reconnaissance d'objets 3D par analyse de Fourier de la silhouette », *Revue technique Thomson-CSF*, Vol. 25, n° 1, Mars 1993.
- [HOUC-95] C.R. Houck, J.A. Joines, M.G. Kay, « A genetic algorithm for function optimization : A Matlab implementation », *NCSU-IE Technical Report 95-09*, 1995.
- [HU-62] M. K. Hu, « Visual pattern recognition by moment invariants », *IRE Trans. Inform. Theory*, Vol. 8, pp. 179-187, Feb 1962.
- [JAIN-96] A. K. Jain, Y. Zhong and S. Lakshmanan, «Object Matching Using Deformable Templates», *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 18, n° 3, pp.267-278, 1996.
- [JAIN-98] A. K. Jain, Y. Zhong and M.P. Dubuisson-Jolly, « Deformable template models : A review », *Signal Processing*, Vol. 71, pp. 109-129, 1998.
- [JAMM-99] G. Jammal and A. Bijaoui, « Denoising and Deconvolution in Nuclear Medicine », *Proceedings of 6th International Conference on Image Processing (ICIP'99)*, Kobe, Japan, 24-28 Oct. 1999, Vol. 3, pp. 896-900.
- [JANE] *Jane's Underwater Warfare Systems*, Ed. By Anthony J. Watts, 10th Edition 1998-1999.
- [JAN-87] D. Jan, « Traitement d'images numérique appliqué aux sonars à haute résolution », *GRETSI 87*, Nice, France, pp. 357-160, 1987.
- [JAN-88] D. Jan, « Simulateur d'images acoustiques large bande », *Rapport final ASM 88/B53/206 DJ/MHL du 30/12/88*, contrat DRET 87/015.05.
- [KALI-99] J. Kalifa, S. Mallat and B. Rougé, « Minimax Deconvolution in Mirror Wavelets Bases », *Rapport EP/CMA/RI 433*, Ecole Polytechnique, Centre de Mathématiques Appliquées, Palaiseau, France, 1999.
- [KASS-88] M. Kass, A. Witkins, D. Terzopoulos, « Snakes : active contour models », *International Journal of Computer Vision*, vol. 1, n° 4, pp 321-330, 1988.

- [KERV-96] C. Kervrann and F. Heitz, « Statistical model-based segmentation of deformable motion », 3rd IEEE International Conference on Image Processing, Lausanne, sept. 1996.
- [KERV-97a] G. Kervern, « Etat de l'art et perspectives en classification de mines sous-marines. Rapport 1 : les mines et leurs évolutions », *Rapport d'étude TMS SAS 97/B/SEGS/DR/154-GK*, 1997.
- [KERV-97b] G. Kervern, « Etat de l'art et perspectives en classification de mines sous-marines. Rapport 3 : les perspectives d'évolution en classification des mines sous marines », *Rapport d'étude TMS SAS 97/B/SEGS/DR/217-GK*, 1997.
- [KHOT-90] A. Khotanzad, Y.H. Hong, « Invariant Image Recognition by Zernike Moments », *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 12, n° 5, May 1990, pp. 489-497.
- [LAM-98] C.L. Lam and S.Y. Yuen, « An unbiased active contour algorithm for object tracking », *Pattern Recognition Letters*, Vol. 19, pp. 491-498, 1998
- [LCR-92a] *NeuroClasse - Manuel Utilisateur (2.0)*, Laboratoire Central de Recherches, Thomson-CSF, Décembre 1992.
- [LCR-92b] *NeuroClasse - Manuel de Référence (2.0)*, Laboratoire Central de Recherches, Thomson-CSF, Décembre 1992.
- [MAIT-01] H. Maître, *Traitement des images de RSO*, Hermès, 2001.
- [MALK-00] J.P. Malkasse, « Identification tool by sonar means with Synthetic Aperture Processing and 3D Reconstruction », *Undersea Defence Technology (UDT) Europe 2000*, Londres, 27-29 June 2000.
- [MALL-89] S.G. Mallat, « A Theory for Multiresolution Signal Decomposition : The Wavelet Representation », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, n° 7, July 1989, pp. 674-693.
- [MANT-87] J. Mantas, « Methodologies in pattern recognition and image analysis – a brief survey », *Pattern Recognition*, Vol. 20, n° 1, pp. 1-6, 1987.
- [MARA-90] J.P. Marage, *Sonars et Acoustique Sous-Marine - tome 1 : le milieu marin*, Valbonne: Thomson-Sintra ASM, 1990.
- [MARA-91] J.P. Marage, *Sonars et Acoustique Sous-Marine - tome 3 : la chaîne de traitement du sonar actif*, Valbonne: Thomson-Sintra ASM, 1991.
- [MARQ-97] J.S. Marques, A.J. Abrantes, « Shape alignment – Optimal initial point and pose estimation », *Pattern Recognition Letters*, Vol. 18, pp.49-53, 1997.
- [MARR-80] D. Marr and E. Hildreth, "Theory of edge detection", *Proc. Roy. Soc. Lond.*, B-207, pp. 187-217, 1980.
- [MART-94] R.L. Martin and R.W. Farwell, "High-Frequency Acoustic Modeling", *OCEANS'1994 MTS/IEEE*, Brest, France, Vol. 2, pp. 272-277, Sept 1994.
- [MEYE-90] F. Meyer, S. Beucher, « Morphological Segmentation », *JVCIR*, Vol. 11, n° 1, pp. 21-46, 1990.
- [MICH-94] *Genetic Algorithms + Data Structures = Evolution Programs*, AI Series, Springer-Verlag, New-York, 1994.
- [MIGN-98] M. Mignotte, « Segmentation d'images sonar par approche markovienne hiérarchique non supervisée et classification d'ombres portées par modèles statistiques », *Thèse de doctorat*, Université de Bretagne Occidentale, 1998.
- [MIGN-00] M. Mignotte, C. Collet, P. Pérez and P. Bouthemy, « Hybrid Genetic Optimization and Statistical Model-Based Approach for the Classification of Shadow in Sonar Imagery », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, n° 2, Feb. 2000.
- [MITC-83] O.R. Mitchell and T.A. Grogen, « Evaluation of Fourier Descriptors for Target Recognition in Digital Imagery », *Final Technical Report RADC-TR-83-33*, Purdue Univ., School of Electrical Engineering, Lafayette, IN, february 1983.

- [MIZO-80] R. Mizoguchi and M. Shimura, « A Nonparametric Algorithm for Detecting Clusters Using Hierarchical Structure », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 2, n° 4, July 1980.
- [PAL-86] S.K. Pal, D.K. Dutta Majumder, *Fuzzy mathematical approach to pattern recognition*, John Wiley/Halsted Press, New York, 1986.
- [PAL-92] S.K. Pal and A. Ghosh, « Fuzzy geometry in image analysis », *Fuzzy Sets and Systems*, Vol. 48, pp.23-40, 1992.
- [PAL-96] S.K. Pal and P.P. Wang, *Genetic Algorithms for Pattern Recognition*, CRC Press, Boca Raton, FL, 1996.
- [PAPO-84] A. Papoulis, *Probability Random, Variables, and Stochastic Processes*, Mc Graw-Hill, 1984.
- [PEJN-92] P. Pejnovic, Lj. Buturovic, Z. Stojiljkovic, « Object Recognition by Invariants », *Intern. Conference on Pattern Recognition*, La Hague, September 1992.
- [PERS-86] E. Persoon, K-S. Fu, « Shape discrimination using Fourier Descriptors », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, n° 3, pp 388-397, May 1986.
- [PIEC-94] W. Pieczynski, « Champs de Markov cachés et estimation conditionnelle itérative », *Revue Traitement du Signal*, Vol. 11, n° 2, pp. 141 :153, 1994.
- [PROK-92] R.J. Prokop, A.P. Reeves, « A Survey of Moment-Based Techniques for Unoccluded Object Representation and Recognition », in *CVGIP : Graphical Models and Image Processing*, Vol. 54, n° 5, pp.438-460, September 1992.
- [PUYH-00] P. Puyhabilier, « La modernisation des chasseurs de mines tripartites », *L'Armement*, n° 71, octobre 2000.
- [QUID-00a] I. Quidu, J.P. Malkasse, G. Burel and P. Vilbé, « Mine classification using a hybrid set of descriptors », *proc. OCEANS'2000 MTS/IEEE*, Providence, Rhode Island, Sept 2000.
- [QUID-00b] I. Quidu, J.P. Malkasse, G. Burel and P. Vilbé, « Mine Classification based on raw sonar data: an approach combining Fourier Descriptors, Statistical Models and Genetic Algorithms », *proc. OCEANS'2000 MTS/IEEE*, Providence, Rhode Island, Sept 2000.
- [QUID-01a] I. Quidu, J.P. Malkasse, G. Burel and P. Vilbé, « Mine classification based on a multiview characterisation », *Undersea Defence Technology (UDT) Europe 2001*, Hamburg, Germany, 26-28 June 2001.
- [QUID-01b] I. Quidu, J.P. Malkasse, G. Burel and P. Vilbé, « A 2-D Filter Specification for Sonar Image Thresholding », *Advanced Concepts for Intelligent Vision Systems (ACIVS'2001) conference*, Baden-Baden, Germany, July 30 – August 3, 2001.
- [QUID-01c] I. Quidu, J.P. Malkasse, P. Vilbé and G. Burel, « Fusion multi-attributs d'images sonar », *colloque GRETSI'01*, Toulouse, 10-13 septembre 2001.
- [REEV-88] A.P. Reeves, R.J. Prokop, S.E. Andrews and F.P. Kuhl, « Three-Dimensional Shape Analysis Using Moments and Fourier Descriptors », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10, n° 6, November 1988.
- [ROSE-84] A. Rosenfeld, « The fuzzy geometry of image subsets », *Pattern Recognition Letters*, Vol. 2, pp. 311-317, 1984.
- [ROSS-87] S.M. Ross, *Initiation aux probabilités*, Presses polytechniques romandes, 1987.
- [SAHO-88] P.K. Sahoo, S. Soltani, A.K.C. Wong and Y.C. Chen, « A survey of Thresholding Techniques », *Computer Vision Graphics Image Processing*, Vol. 41, pp. 233-260, 1988.
- [SEET-91] G. Seetharaman, A. Narasimhan, A. Sathe and L. Storc, « Image Segmentation with Genetic Algorithms : a Formulation and Implementation », *SPIE, vol. 1569, Stochastic and Neural Methods in Signal Processing, and Computer Vision*, pp. 269-273, 1991.

- [SOME-93] M.L. Somers, « Sonar Imaging of the Seabed: Techniques, Performances, Applications », in *Acoustic Signal Processing for Ocean Exploration*, J.M.F. Moura and I.M.G. Lourtie (eds.), Canadian Government (pp. 355-369), 1993.
- [STAG-96] B. Stage and B. Zerr, « Application of the Dempster-Shafer theory to combination of multiple view sonar images », *Undersea Defence Technology (UDT) 96*, London, UK, 2-4 july 1996.
- [TEAG-80] M. R. Teague, « Image analysis via the general theory of moments », *J. Opt. Soc. Am.*, Vol. 70, No 8, August 1980, pp. 920-930.
- [THOU-95] P. Thourel, C. Collet, P. Bouthemy et P. Pérez, « Multiresolution analysis and MRF modeling applied to the segmentation of shadows in sonar pictures », *ACCV'95 Second Asian Conference on Computer Vision*, Singapore, dec. 1995.
- [THOU-96] P. Thourel, « Segmentation d'images sonar par modélisation markovienne hiérarchique et analyse multirésolution », *Thèse de doctorat*, Université de Bretagne Occidentale, 1996.
- [TIZH-98] H.R. Tizhoosh, « Fuzzy Image Processing : Potentials and State of the Art », *Methodologies for the Conception, Design and Application of Soft Computing, proc. of IIZUKA'98*, pp. 321-324, 1998.
- [VACH-95] C. Vachier, « Extraction de caractéristiques, segmentation d'image et morphologie mathématique », *Thèse de doctorat*, Ecole des Mines de Paris, déc. 95.
- [VOLL-85] M. Volle, *Analyse de données*, Economica, 3^{ième} éd., 1985.
- [WALL-80] T.P. Wallace and P.A. Wintz, « An efficient three-dimensional aircraft recognition algorithm using normalized Fourier descriptors », *Computer Graphics and Image Processing*, Vol. 13, pp. 99-126, 1980.
- [WOOD-96] J. Wood, « Invariant Pattern Recognition : a Review », *Pattern Recognition*, Vol. 29, n° 1, pp.1-17, 1996.
- [ZADE-65] L.A. Zadeh, « Fuzzy Sets », *Information and Control*, Vol. 8, n° 3, pp. 338-353, June 1965.
- [ZHAN-72] C.T. Zhan and R.Z. Roskies, « Fourier descriptors for plane closed curves », *IEEE Trans. on computers*, Vol. c-21, n° 3, March 1972, pp. 269-281.
- [ZHON-98] Y. Zhong, A.K. Jain and M.P. Dubuisson, « Object Tracking Using Deformable Templates », *Int. Conf. On Computer Vision*, Bombay, India, pp. 440-445, 1998.

Résumé

La classification sur ombre de mines marines peut être effectuée à partir d'images sonar haute résolution. Cependant, la classification mono-vue admet des limites du fait de leur forme géométrique parfois complexe. Une manière de s'affranchir des ambiguïtés consiste à effectuer une classification multi-vues. Dans un premier temps, il convient d'assimiler les spécificités de l'imagerie sonar. D'un côté, les propriétés statistiques des pixels et le principe d'acquisition des images sont des connaissances avantageusement prises en compte lors de l'étape de segmentation. D'un autre côté, les distorsions géométriques qui affectent l'ombre portée du fait des performances du sonar et de la prise de vue doivent être considérées. En matière de reconnaissance de formes, les données sonar segmentées peuvent être diversement résumées suivant qu'on s'attache à définir des grandeurs indépendantes des positions relatives objet-sonar et/ou de la résolution du sonar ou bien, autorisant une reconstruction du contour et la conservation de l'orientation de la forme associée.

Ces considérations du problème mono-vue ont préparé l'élaboration de processus plus complexes de classification multi-vues.

D'une part, on manipule les attributs issus d'un traitement image par image. Deux cas se présentent : en exploitant l'évolution de la forme de l'ombre en fonction du trajet du sonar ou, au contraire, en s'affranchissant des transformations du plan. Dans le premier cas, on caractérise de manière globale l'ensemble des valeurs successives prises par des attributs sensibles aux diverses formes de l'ombre. Dans le second cas, sans connaissance précise des conditions d'acquisition, les attributs extraits sont fusionnés et présentés à l'opérateur sous forme de mesures pour l'aide à la décision.

Par la logique floue d'autre part, les outils de reconnaissance de formes calculés sur des données binaires ont été étendus au cas de données en niveaux de gris d'une nouvelle image pour sa caractérisation.

Mots-clefs : image sonar, classification sur ombre, reconnaissance de formes, multi-vues, appariement de silhouettes, algorithme génétique, fusion, logique floue.

Abstract

In the context of mine warfare, the classification of objects laying on the seabed is based on the characterization of their associated cast shadows. However, a single view may not be sufficient to classify new types of mines which are specific in shape. Multiple views can then be used to alleviate some ambiguities. Our study starts by considering sonar imagery particularities. On the one hand, pixels statistic properties as well as the principle of sonar image formation provide some useful knowledge to be integrated in the segmentation step. On the other hand, shadows undergo geometric distortions caused by sonar performances or under a particular view. This problem is differently tackled during the pattern recognition process designed to provide two types of features : whereas the first ones do not depend on relative sonar-object positions, the second ones allow contour reconstruction while keeping orientation information of the shape.

On account of the previous single view analysis, more complex processes have been derived to enable a robust multi-view classification.

At first, we deal with features extracted from each view. Two cases are presented : by using the shadow shapes evolution over the sonar trajectory or, on the contrary, by removing shapes deformations. In the first case, we use features sensitive to shadow changes and sequentially store their values to provide a complete multi-view characterization. In the second case, without precise requirements about the acquisition conditions, extracted features are fused to be presented to the operator as measures.

Secondly, we were interested in fuzzy logic which allows the extension of pattern recognition tools to the characterization of fuzzy geometry of image subsets. This has been applied to a new image created from the whole views of the sequence.

Keywords : sonar image, shadow classification, pattern recognition, multi-view, contour matching, genetic algorithm, fusion, fuzzy logic